

Reprogramming to Cancer Induced Pluripotent Stem Cells Elucidates the Contribution of Genetic and Epigenetic Alterations to Breast Carcinogenesis

By Leong Yeh Chwan, BSc (Hons)

Faculty of Medicine and Health Sciences
School of Veterinary Medicine & Science
University of Nottingham, Sutton Bonington Campus

Thesis submitted to The University of Nottingham for the degree of Doctor of Philosophy

February 2018

ABSTRACT

The induced pluripotent stem cells (iPSCs) technology has revolutionized disease modelling by enabling the generation of patient-specific pluripotent stem cells for the study of complex disorders such as cancer. Somatic cell reprogramming through iPSCs induces global epigenetic reconfiguration of the chromatin which converts cancer cells to an embryonic stem cell-like state with potential reversion of tumorigenicity. Therefore, reprogramming can be used to answer the question as to whether epigenetic alterations alone can be sufficient to induce carcinogenesis, independent of genetic defects. In addition, it can used to dissect the relative contribution of genetics and epigenetics and epigenetics to tumorigenicity. In this study, the triple negative breast cancer (TNBC) cell line BT-549 and oestrogen receptor positive (ER^+) cell line MCF7 were successfully reprogrammed by using the non-integrative episomal vectors expressing *OCT4, SOX2, L-MYC, KLF4, LIN28, EBNA1,* shRNA against *TP53*, and *microRNA-302/367* cluster together with treatment of sodium butyrate. Pluripotency of cancer-derived iPSCs was confirmed by RT-PCR, RT-qPCR and immunofluorescence staining for expression of pluripotency markers. Differentiation potential of iPSCs was also assessed by using *in vitro* differentiation either spontaneous or directed to the mammary lineage. Functional assays indicated potential loss of tumorigenicity in re-differentiated cells derived from cancer iPSCs. The same approach was applied to study an immortalised, non-malignant mammary epithelial cell line MCF10A and two of its derived isogenic lines harbouring the two most frequent mutations in breast cancer, $PIK3CA^{H1047R\ (+/-)}$ and $TP53^{(-/-)}$, created by using CRISPR-Cas9 gene editing.

Reprogramming induced a tumorigenic phenotype in iPSCs (*PIK3CA*$^{H1047R\ (+/-)}$ isogenic line only) and re-differentiated progenies (in both wild type MCF10A and *PIK3CA*$^{H1047R\ (+/-)}$ cell lines), suggesting the contribution of *PIK3CA* mutation in enhancing malignant transformation. Results in this study suggested that epigenetics alone and/or its interaction with genetic defects (e.g. *PIK3CA* mutation) has significant impact on breast cancer carcinogenesis. The dissection of the molecular mechanisms underlying the loss and gain of tumorigenicity using the iPSC models generated in this study could provide general understandings on breast carcinogenesis, which in turn could have important clinical implications.

ACKNOWLEDGEMENT

I can never be thankful enough for the support and love my parent and grandparent have been unconditionally channelling into my life, driving me forward whenever I was held back by doubts and hurdles. A heartfelt gratitude to my best friend Yap Kwai Ling for her invaluable advices and emotional support in both academic and personal life. Thank you to both my co-supervisors Ramiro Alberio and Anna Grabowska and all my colleagues, especially Ryan, Ahmad, Aouatif and Lola who had been so supportive and helpful and had created such a friendly and conducive study environment for the last three and a half years. A big thank you I owed to my caring senior Lina for generously sharing her experiences and offering help especially at the few beginning months of my PhD. Lastly, I am most grateful for the freedom, guidance, encouragement, motivation, patience, the list goes on, my supervisor Dr. Cinzia Allegrucci had given me throughout my PhD. Nothing would have been possible without the support of these people. Thank you very much.

DECLARATION

I hereby declare that the work present in this thesis has been written by myself and has not been previously submitted for any other degree. All procedures and work described were performed by myself.

LEONG YEH CHWAN

Date:

TABLE OF CONTENT

ABSTRACT ..... **ii**
ACKNOWLEDGEMENT ..... **iv**
DECLARATION ..... **v**
TABLE OF CONTENT ..... **vi**
LIST OF FIGURES ..... **x**
LIST OF TABLES ..... **xiv**
LIST OF ABBREVIATION ..... **xv**
1. INTRODUCTION 1
1.1 THESIS RATIONALE AND HYPOTHESIS 1
1.2 CANCER: AN OVERVIEW 5
1.2.1 Breast Cancer 8
1.2.2 Cancer genetics 28
1.2.3 Epigenetics in cancer 37
1.2.4 Breast cancer epigenetics 44
1.3 EPIGENETIC REPROGRAMMING OF BREAST CANCER 48
1.3.1 Epigenetic therapies (chemical approach) 48
1.3.2 Epigenetic reprogramming using biological approaches 49
1.3.3 Induced Pluripotent Stem Cell Technology 56
1.4 IPSCS AND CANCER 74
1.4.1 Cancer Cell Reprogramming using iPSC technology 75
1.5 AIM AND OBJECTIVES 82
2. METHODOLOGY 83
2.1 CELL CULTURE 83
2.1.1 Culture of continuous cell lines 83
2.1.2 Cell Counting 84
2.1.3 Cryopreservation of Cells 84
2.1.4 Thawing of Cells 84
2.1.5 Stem cells culture 85
2.2 IPSC REPROGRAMMING 86
2.2.1 X-tremeGEN 9 DNA Transfection Reagent (XG9) (Roche, 06365779001) 86
2.2.2 Amaxa® Cell line Nucleofector kits (Lonza) transfection 87

2.2.3 Fate-reprogramming & fate-maintenance media (FRM & FMM) 88
2.2.4 Small molecule treatment 89
2.3 *IN VITRO* DIFFERENTIATION 89
2.3.1 Embryoid body (EB) formation 89
2.3.2 Mammary organoid formation 89
2.4 SOFT AGAR COLONY FORMATION ASSAY 91
2.5 CHEMOTHERAPY DRUG SENSITIVITY TEST 91
2.6 IMAGING 92
2.7 IMMUNOLOGICAL TECHNIQUES 92
2.7.1 Fixation 92
2.7.2 Immunofluorescence (IF) staining 92
2.7.3 Fluorescence activated cell sorting (FACS) analysis 93
2.8 MOLECULAR TECHNIQUE AND ANALYSIS 94
2.8.1 DNA extraction 94
2.8.2 RNA extraction 94
2.8.3 cDNA synthesis 95
2.8.4 RT-PCR 95
2.8.5 RT-qPCR 96
2.8.6 Gel Electrophoresis 97
2.8.7 Growing & freezing bacteria 97
2.8.8 Plasmids purification 98
2.8.9 Restriction Enzyme Digest 98
2.8.10 Gel and PCR purification 99
2.9 CRISPR-CAS9 EXPERIMENTAL DESIGN 99
2.9.1 SgRNA and ssODN design 99
2.9.2 Cloning 100
2.9.3 SgRNA and Cas9-GFP plasmids transfection 102
2.9.4 DNA extraction 103
2.9.5 Screening 104
2.9.1 *TP53* mutation transfection 104
2.10 WESTERN BLOTTING 106
2.10.1 Protein extraction 106
2.10.2 Polyacrylamide Gel Electrophoresis of Protein Samples 106
2.10.3 Western Blotting Transfer 107
2.10.4 Membrane Blocking and Antibody Incubation 107

2.11 STATISTICAL ANALYSIS..... 108
3. REPROGRAMMING OF CONTROL CELL LINE MCF-10A..... 109
3.1 INTRODUCTION 109
3.2 RESULTS 113
3.2.1 Optimisation of transfection conditions 113
3.2.2 Episomal vector-mediated reprogramming of MCF10A 117
3.2.3 Characterization of the reprogrammed MCF-10A line (10AR) 120
3.3 DISCUSSION 128
3.3.1 Optimum cell density for reprogramming 130
3.3.2 Loss of episomal vectors..... 132
3.3.3 Reprogramming efficiency..... 133
3.3.4 Pluripotency of 10AR 139
4. REPROGRAMMING OF BREAST CANCER CELL LINES..... 149
4.1 INTRODUCTION 149
4.1.1 Epigenetic reprogramming of breast cancer**Error! Bookmark not defined.**
4.2 RESULTS 151
4.2.1 Optimisation of X-tremeGENE 9 DNA Transfection Reagent (XG9) transfection conditions..... 151
4.2.2 Reprogramming of BT-549, MCF7 and MDA-MB-231..... 159
4.2.3 Reprogramming of BT-549 using the FRM/FMM media system 165
4.2.4 Conversion of partially-reprogrammed iPSCs (pre-iPSCs) to *bone fide* iPSCs. 167
4.2.5 Reprogramming of BT-549 with five plasmids and sodium butyrate..... 169
4.2.6 Reprogramming of MCF7 with five plasmids and sodium butyrate treatment 177
4.2.7 Reprogramming of MDA-MB-231 with five plasmids and sodium butyrate..... 184
4.3 DISCUSSION 186
4.3.1 Reprogramming of cancer cells..... 187
5. REPROGRAMMING MCF10A MUTANTS HARBOURING PIK3CA$^{H1047R+/-}$ AND PIK3CA$^{H1047R+/-}$_TP53$^{-/-}$ MUTATIONS..... 201
5.1 INTRODUCTION 201
5.1.1 Breast cancer driver mutations..... 201
5.1.2 Generation of mutants harbouring driver mutations by CRISPR-Cas9 technology 204

5.2 RESULTS 205
5.2.1 Generation of PIK3CA$^{H1047(+/-)}$ mutant cell line with CRISPR-Cas9 205
5.2.2 CRISPR-Cas9 gene editing: *PIK3CA$^{H1047R(+/-)}$* 207
5.2.3 CRISPR-Cas9 gene editing: introduction of *TP53$^{(-/-)}$* mutation to MCF10A_PIK3CA$^{H1047R(+/-)}$ cell line 216
5.2.4 Reprogramming of mutant cell lines 224
5.3 DISCUSSION 230
5.3.1 Introduce of *PIK3CA$^{H1047R(+/-)}$* mutation via CRISPR-Cas9 system 231
5.3.2 Creation of cancer models using CRISPR-CAS9 system 233
6. DIFFERENTIATION POTENTIAL AND TUMORIGENICITY OF BREAST CANCER AND MUTANT MCF10A -DERIVED IPSC LINES 242
6.1 INTRODUCTION 242
6.1.1 Significance of cancer iPSC differentiation 244
6.1.2 Lineage-specificity of cancer mutations 245
6.2 RESULTS 247
6.2.1 Embryoid body (EB) formation 247
6.2.2 Directed differentiation to generate mammary organoids 256
6.2.3 CD24 and CD44 surface markers expression 264
6.2.4 Soft agar colony formation assay 272
6.2.5 Chemotherapy sensitivity and resistance assays 279
6.3 DISCUSSION 281
6.3.1 Differentiation potential of iPSC lines 283
6.3.2 CD24 and CD44 marker expression 286
6.3.3 Impact of iPSC reprogramming on breast cancer tumorigenicity 290
6.3.4 Chemotherapy sensitivity assay 293
7. GENERAL DISCUSSION 296
7.1 TUMOUR REVERSION THROUGH IPSC REPROGRAMMING 298
7.2 CANCER CELL REPROGAMMING AND RE-DIFFERENTIATION – *IN VITRO* CANCER MODELLING 299
7.3 LIMITATIONS OF THE STUDY 303
7.4 FUTURE WORK 306
7.5 CONCLUSION 310
APPENDIX **311**
REQUEST PERMISSIONS **344**

BIBLIOGRAPHY 352

LIST OF FIGURES

Figure 1.1 Models of tumour reversion 4
Figure 1.2 The accumulation of mutations throughout lifetime lead to cancer. 7
Figure 1.3 Schematic diagrams of human breast tissues..... 16
Figure 1.4 Schematic diagram of mammary differentiation hierarchy 18
Figure 1.5 Mammary differentiation hierarchy and cell-of-origin of breast cancers proposed by Lim et al. (2009) and Prat et al. (2013). ... 19
Figure 1.6 The clonal evolution model of breast cancer. 23
Figure 1.7 Schematic diagram of tumour heterogeneity in breast cancer. 28
Figure 1.8 The stages of iPSC generation process as proposed by Liang & Zhang (2013). 63
Figure 1.9 Cellular and molecular mechanisms underlying mesenchymal-to-epithelial Transition (MET) during early stage of reprogramming. 64
Figure 1.10 A model of pluripotency network activation. 66
Figure 1.11 Different kinetic models of iPSC generation. 69
Figure 1.12 The CRISPR-Cas9 gene editing. 81
Figure 3.1 Optimization of MCF-10A cell density with X-tremeGENE™ 9 DNA transfection reagent for reprogramming. 115
Figure 3.2 Representative images showing GFP^+ cells at different time points after transfection under light (top) and fluorescence microscopes (bottom). 116
Figure 3.3 Generation of iPSCs from MCF10A. 119
Figure 3.4 Episomal traces in reprogrammed MCF10A (10AR). 120
Figure 3.5 Representative RT-PCR of pluripotency markers in 10AR. 123
Figure 3.6 Quantification of pluripotency markers expression in 10AR by real-time-PCR (RT-qPCR). 125
Figure 3.7 Immunofluorescence staining of pluripotency markers in 10AR. 127
Figure 3.8 FACS analysis of pluripotency markers in 10AR. 128
Figure 4.1 Optimization of MCF7 cell density with X-tremeGENE™ 9 DNA transfection reagent for reprogramming. 152
Figure 4.2 Representative images of GFP^+ MCF7 cells seeded at different densities and different time points after transfection. 153
Figure 4.3 Optimization of BT-549 cell density with X-tremeGENE™ 9 DNA transfection reagent for reprogramming. 155
Figure 4.4 Representative images of GFP^+ BT-549 cells seeded at different densities at different time points after transfection..... 156

Figure 4.5 Optimization of MDA-MB-231 cell density with X-tremeGENE™ 9 DNA transfection reagent for reprogramming. 157
Figure 4.6 Representative images of GFP^{+} MDA-MB-231 cells seeded at different densities at different time points after transfection. 158
Figure 4.7 Reprogramming of breast cancer cell lines 161
Figure 4.8 Morphology of reprogrammed BT-549 cell lines cultured in Essential 8 medium. 162
Figure 4.9 Representative RT-PCR of pluripotency markers in reprogrammed MCF-7, BT-549 and MDA-MB-231. 164
Figure 4.10 Reprogramming of BT-549 using the FRM/FMM media system. 166
Figure 4.11 Conversion of FMM_BTR1 to *bona fide* iPSCs by 5-Azacytidine (AZA) treatment. 168
Figure 4.12 Reprogramming of BT-549 with five episomal plasmids and sodium butyrate (NaB). 171
Figure 4.13 Traces of episomal vectors in 5PLA_BTR5. 172
Figure 4.14 Representative RT-PCR of pluripotency markers in 5PLA_BTR5. 174
Figure 4.15 Quantification of pluripotency markers expression in 5PLA_BTR5 by real-time-PCR (RT-qPCR). 175
Figure 4.16 Immunofluorescence staining of SSEA-4, TRA-1-81 and NANOG in 5PLA_BTR5. 176
Figure 4.17 Reprogramming of MCF7 with five episomal plasmids in combination with sodium butyrate (NaB) treatment. 178
Figure 4.18 Traces of episomal vectors in 5PLA_MCF7R. 179
Figure 4.19 Representative RT-PCR of pluripotency markers expression in 5PLA_MCF7R. 181
Figure 4.20 Quantification of pluripotency markers expression in 5PLA_MCF7R by real-time PCR (RT-qPCR). 182
Figure 4.21 Immunofluorescence staining of SSEA-4, TRA-1-60 and NANOG in 5PLA_MCF7R. 183
Figure 4.22 Reprogramming of MDA-MB-231 with five episomal plasmids in combination with sodium butyrate (NaB) treatment. 184
Figure 4.23 Representative RT-PCR of pluripotency markers expression in 5PLA_MDAR1 and 5PLA_MDAR2. 185
Figure 5.1 sgRNA and single-stranded oligodeoxynucleotide (ssODN) template design for *PIK3CA*$^{H1047R(+/-)}$ mutation. 206
Figure 5.2 Plasmid maps of Addgene plasmid #41824 (a) and #44719 (b). 208
Figure 5.3 Cloning of sgRNA into sgRNA vector (Addgene #41824). 209
Figure 5.4 CRISPR-Cas9-mediated introduction of PIK3CA$^{H1047(+/-)}$ mutation in MCF10A. 210
Figure 5.5 BspH1 screening of pool GFP^{high} MCF10A cells. 212
Figure 5.6 Screening of *PIK3CA* mutants. 212
Figure 5.7 Sanger sequencing of four randomly selected GFP^{high} clones 213
Figure 5.8 CRISPR-induced mutations found in four GFP^{high} clones. 215

Figure 5.9 Sanger sequencing confirming the presence of CAT>CGT point mutation in MCF10A_PIK3CA$^{H1047R(+/-)}$ cell line from the Horizon Discovery. 215
Figure 5.10 sgRNA targeting exon 4 of *TP53* gene. 216
Figure 5.11 Schematic diagram of strategy for the identification of clones harbouring CRISPR-Cas9-induced mutations using PAGE assay. 218
Figure 5.12 Screening for *TP53* mutated clones. 219
Figure 5.13 Western blotting analysis of TP53 expression. 220
Figure 5.14 Alignment of cloned amplicon sequences from A3. 222
Figure 5.15 Alignment of cloned amplicon sequences from A8. 223
Figure 5.16 Reprogramming of MCF10A mutant cell lines MCF10A_PIK3CA$^{H1047R(+/-)}$ and MCF10A_PIK3CA$^{H1047R(+/-)}$/TP53$^{(-/-)}$..... 226
Figure 5.17 Representative RT-PCR of pluripotency markers in MCF10A_PIK3CA$^{H1047R(+/-)}$ and MCF10A_PIK3CA$^{H1047R(+/-)}$/TP53$^{(-/-)}$ and their respective iPSCs 227
Figure 5.18 Immunofluorescence staining of SSEA-4, TRA-1-60 and NANOG in 10A_PIK3CA$^{(-/+)}$ R4, 10A_PIK3CA$^{(-/+)}$/TP53$^{(-/-)}$ R1.0 and 10A_PIK3CA$^{(-/+)}$/TP53$^{(-/-)}$ R2.1.3. 229
Figure 5.19 The domain structure of Class IA p110α catalytic subunit. 234
Figure 6.1 H9 *in vitro* differentiation. 251
Figure 6.2 10AR *in vitro* differentiation. 252
Figure 6.3 5PLA_BTR5 *in vitro* differentiation. 253
Figure 6.4 5PLA_MCF7R *in vitro* differentiation. 254
Figure 6.5 10A_PIK3CA$^{(+/-)}$ R4 *in vitro* differentiation. 255
Figure 6.6 Timeline of mammary organoid generation. 256
Figure 6.7 Mammary organoid formation from H9. 257
Figure 6.8 Mammary organoid formation from 10AR. 258
Figure 6.9 Mammary organoid formation from 5PLA_BTR5. 258
Figure 6.10 Mammary organoid formation from 5PLA_MCF7R. 259
Figure 6.11 Mammary organoid formation from 10A_PIK3CA$^{(+/-)}$ R4.259
Figure 6.12 RT-PCR analysis for mammary markers expression in ESC- and iPSC-derived mammary organoids as compared to their parental and pluripotent counterparts. 263
Figure 6.13 Representative FACS analysis of CD24 and CD44 surface markers expression in MCF10A, iPSCs and differentiated derivatives. . 267
Figure 6.14 Representative FACS analysis of CD24 and CD44 surface markers expression in BT-549, iPSCs and differentiated derivatives. .. 268
Figure 6.15 Representative FACS analysis of CD24 and CD44 surface markers expression in MCF7, iPSCs and differentiated derivatives..... 269
Figure 6.16 Representative FACS analysis of CD24 and CD44 surface markers expression in MCF10A_PIK3CA$^{(+/-)}$, iPSCs and differentiated derivatives. 270
Figure 6.17 CD44/CD24 profiles of parental, iPSCs and re-differentiated cell lines. 271
Figure 6.18 Colony forming assay of MCF10A, iPSCs and differentiated derivatives. 274

Figure 6.19 Colony forming assay of MCF7, iPSCs and differentiated derivatives. 275
Figure 6.20 Colony forming assay of BT-549, iPSCs and differentiated derivatives. 276
Figure 6.21 Colony forming assay of MCF10A_PIK3CA(+/-), iPSCs and differentiated derivatives. 277
Figure 6.22 Potential effect of medium system, differentiation process and PIK3CA mutation on tumorigenicity. 278
Figure 6.23 IC_{50} value of paclitaxel for parental lines and differentiated derivatives. 280
Figure 7.1 The potential applications of iPSC technology in cancer research. 302

LIST OF TABLES

Table 1.1 Breast cancer molecular subtypes and their characteristics. 11
Table 1.2 CSC markers reported for different solid tumour types. Table derived from Medema (2013). 25
Table 1.3 Mechanisms and factors associated with cancer stem cells regulation (adapted from by Guo (2014)). 27
Table 1.4 Large-scale genome-wide studies of breast cancers. 34
Table 1.5 Different approaches used for epigenetic reprogramming of cancer cells. 51
Table 1.6 Different methods used for iPSC generation. Table was created based on Bayart and Cohen-Haguenauer (2013). 59
Table 2.1 Composition of fate-reprogramming (FRM) and fate-maintenance (FMM) media described by Valamehr et al. (2014) as compared to traditional hESC medium. 88
Table 2.2 Composition of media used for generation of mammary organoid (Qu et al., 2017). 90
Table 3.1 The reprogramming efficiency of episomal vector approach. 134
Table 5.1 Studies that generated PIK3CA mutation models using gene editing technology. 234
Table 6.1 Summary of iPSC lines isolated and characterisation performed 243
Table 6.2 Expression of three germ layer markers by all pluripotent stem cells and their embryoid bodies (at day 26) determined by RT-PCR. n = 1. 249
Table 6.3 Immunofluorescence staining for markers of three germ layers. n = 1. 250
Table 6.4 Expression of mammary markers by mammary organoids generated from different PSCs lines determined by RT-PCR. n = 1. 261
Table 6.5 IC_{50} value of paclitaxel for all cell lines. n = 3. 279
Table 6.6 A summary of results obtained in Chapter 6. 282
Table 6.7 Colony forming assay performed on different *PIK3CA* mutation studies 291

LIST OF ABBREVIATION

%	Percent
°C	Degrees Celsius
µL	Microliter
5-aza	5-azacytidine
BCSCs	Breast CSCs
bFGF	Basic fibroblast growth factor
bp	Base pairs
BSA	Bovine serum albumin
CD24	Cluster of differentiation 24
CD44	Cluster of differentiation 44
cDNA	Complementary DNA
CSCs	Cancer stem cells
Ct	Threshold cycle
DAPI	4',6-diamidino-2-phenylindole
Decitabine	5-aza-2'-deoxycytidine
DMEM	Dulbecco's (Modified Eagle's) minimal essential medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNMTi	DNA methytransferase inhibitor
dNTP	Deoxyribonucleotide triphosphate
E6	Essential 6 medium
E8	Essential 8 medium
EB	Embryoid body
EBNA1	Epstein–Barr virus nuclear antigen 1
EGFP	Enhanced green fluorescence protein
ER	Oestrogen receptor
FACS	Fluorescence activated cell sorting or flow cytometry
FBS	Foetal bovine serum
FMM	Fate maintenance medium
FRM	Fate reprogramming medium
GFP	Green fluorescence protein
h	hour
HDACi	Histone deacetylase inhibitor
HER2	Human epidermal growth factor 2
hESCs	Human embryonic stem cells
hiPSCs	Human iPSCs
hPSCs	Human PSCs
iPSCs	Induced pluripotent stem cells
Kb	Kilobase
KLF4	Kruppel like factor 4
M	Molarity
mESCs	Mouse embryonic stem cells

min	minute
miPSCs	Mouse iPSCs
miR	MicroRNA
mL	Millilitre
mM	Milimolar
mPSCs	Mouse PSCs
mRNA	Messenger ribonucleic acid
NaB	Sodium butyrate
ng	Nanogram
OCT3/4	Octamer-binding transcription factor 3/4
OSKM	OCT3/4, SOX2, KLF4, C-MYC or L-MYC
OSKML	OSKM and LIN28
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit α
PR	Progesterone receptor
PSCs	Human pluripotent stem cells
RNA	Ribonucleic acid
ROCKi	Rho-associated protein kinase (ROCK) inhibitor
RT	Room temperature
RT-PCR	Reverse transcriptase-polymerase chain reaction
RT-qPCR	Reverse transcriptase-quantitative polymerase chain reaction
SCNT	Somatin cell nuclear transfer
SD	Standard Deviation
shRNA	Short hairpin RNA
SOX2	SRY (sex determining region Y)-box 2
TCGA	The Cancer Genome Atlas
TF	Transcription factor
TNBC	Triple negative breast cancer
TP53	Tumour protein P53
UTR	Untranslated region
xg	Times gravity
XG9	X-tremeGEN 9 DNA (XG9) transfection reagent

1. INTRODUCTION

1.1 THESIS RATIONALE AND HYPOTHESIS

Epigenetic dysregulation has been associated with cancer (Berdasco and Esteller, 2010). In contrast to genetics, epigenetic landscape of cells is generally reversible and some of the stable epigenetic marks are inheritable mitotically and/or meiotically, thus making epigenome an interesting target for cancer pathological studies and anticancer therapies (Hatzimichael and Crook, 2013). The basis of epigenetic treatments is to effectively and permanently reprogram and remodel the epigenetic marks characteristic of cancer cells to normal somatic cells, reverting malignant properties of cancer cells and leading to complete cure of the disease (Blancafort et al., 2013).

Global-wide reconfiguration of cancer cell epigenome through epigenetic reprogramming represents a valuable opportunity to investigate the impacts of epigenetic alterations on malignant transformation independent of genetic aberrations (Semi et al., 2013). Through various approaches, such as somatic cell nuclear transfer (SCNT) (Gurdon, 1962), cell fusion (Harris and Watkins, 1965, Harris et al., 1969), exposure to ESC microenvironment (Postovit et al., 2006, Hendrix et al., 2007, Postovit et al., 2008) and oocyte extracts (Allegrucci et al., 2011) and the iPSC technology (Stricker et al., 2014), cells of different cancer types have been successfully reprogrammed. IPSC technology represents the most widely used method for cancer cell reprogramming due to the involvement of

defined transcription factors, high reproducibility as well as being less technical challenging. The majority of cancer cell reprogramming studies reported reduced or complete loss of malignancy, suggesting the importance of epigenetic alterations in cancer (Hendrix et al., 2007, Zhang et al., 2013). In addition, the use SCNT and iPSC technology have led to acquisition of pluripotency and restoration of differentiation potential in cancer cells, permitting the derivation and study of differentiated progenies from cancer-derived pluripotent stem cells. Therefore, these approaches allow questions pertaining cancer research such as cellular origin of cancer, mechanisms of cancer initiation and progression as well as the discovery and development of effective anticancer therapies to be addressed (Kim and Zaret, 2015).

In a comprehensive review, Powers and Pollack (2016) elaborated the idea of tumour reversion, suggesting malignant transformation is a bidirectional process (tumour progression and reversion) instead of unidirectional (tumour progression). Early evidence of tumour reversion includes ovarian teratomas, embryonal carcinoma and an epithelial tumour (Askanazy, 1907, Pierce and Dixon, 1959, Kleinsmith and Pierce, 1964), from which benign cells/tissues were generated through induction of differentiation. These studies suggested normal growth control can be restored in tumour cells despite the presence of transforming genetic and epigenetic alterations. Telerman and Amson (2009) proposed three models through which tumour cells re-acquire normalcy. First, a direct reversion through repair, inhibition or loss of major carcinogenic event/s. This is exemplified by the loss of oncogenic virus or alterations of the oncogenic

foreign material (at molecular or protein levels) in virally-transformed cells (Fischinger et al., 1972, Wyke et al., 1980). Imatinib (a BCR-ABL inhibitor) treatment targeting the BCR-ABL, a well-characterised driver of chronic myeloid leukaemia (CML), is also thought to follow this model (Druker et al., 1996, Ren, 2005). Second, tumour reversion takes an alternative path which comprises multistep and only intersects with the transformation pathway at some points. For example, Noda et al. (1989) reported the suppression of malignancy through *KREV1* (or *RAP1*) overexpression in *Ras*-mutant. Lastly, the third model suggests that events leading to malignant transformation, such as activation/loss of oncogenes/ tumour suppressor genes, operate in concert with other pathways involving genes, such as *SIAH1*, *STEAP3, PSEN1* and *TCTP*, to bring about tumour reversion. For example, *SIAH1* and *STEAP3* encode effectors of p53. Their overexpression in tumour revertants compensated the effect of p53 loss leading to suppressed malignancy (Tuynder et al., 2002, Telerman and Amson, 2009).

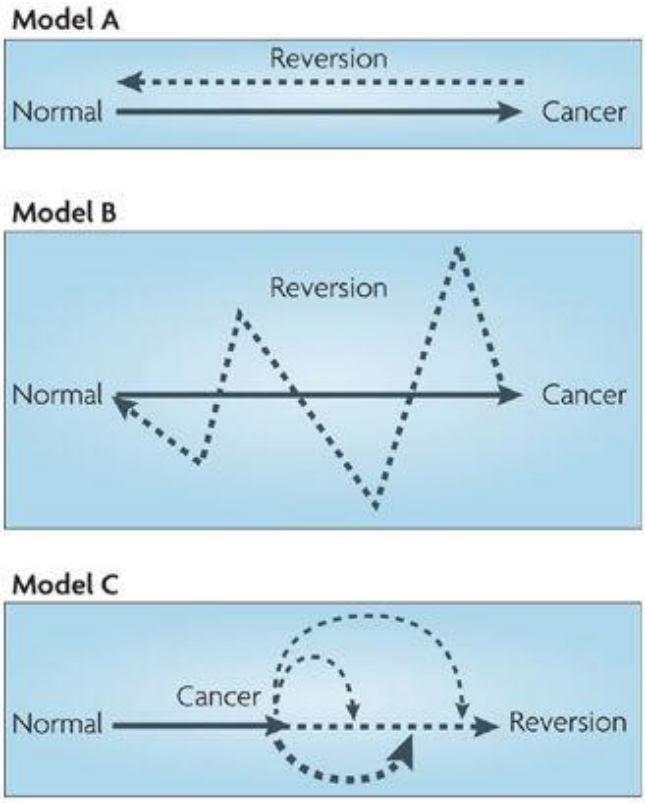


Figure 1.1 Models of tumour reversion

Model A: reversion can be achieved by direct targeting of the major event leading to malignant transformation. Model B: reversion takes an alternative path, which involves multiple steps and intercepts with the transformation pathway at some points. Model C: reversion occurs through a comprehensive process which involves interaction between various oncogenic events and pathways regulated by genes such as SIAH1, STEAP3, PSEN1 and TCTP. Figure derived from (Telerman and Amson, 2009) (Licence number: 4292611399315).

Based on the above findings, this project hypothesised that tumorigenicity of breast cancer cells can be reversed through epigenetic reprogramming by using the iPSC approach. In addition, this project aimed to evaluate the relative contribution of genetics and epigenetics in breast carcinogenesis by utilizing cancer-derived iPSCs as disease models.

1.2 CANCER: AN OVERVIEW

Despite continuous effort and investment of significant resources in cancer research, the disease remains a leading cause of death worldwide, with approximately 8.8 million deaths reported in year 2015 according to the World Health Organization (WHO) (WHO, 2017). Cancer has previously been regarded as a genetic disease characterized by the presence of mutations primarily in tumour suppressor genes, oncogenes and genes associated with the maintenance of genome integrity and DNA repair (Garraway and Lander, 2013, Shen and Laird, 2013). A variety of genetic mutation mechanisms have been identified, including aberrant nucleotide sequences (caused by point mutation, insertion and deletion), abnormal DNA and chromosomal copy number, DNA rearrangement and the integration of foreign DNA materials (Stratton et al., 2009). However, increasing evidence shows that epigenetic abnormalities play a crucial role in cancer initiation and development (Feinberg and Tycko, 2004, Esteller, 2007, Chin and Gray, 2008, Hanahan and Weinberg, 2011a). The dysregulation of epigenetic mechanisms, including DNA methylation, histone modification and non-coding RNA regulation has been associated with cancer initiation, maintenance, progression, treatment resistance as well as recurrence (Jones and Baylin, 2002, Jaenisch and Bird, 2003). Through extensive genome- and epigenome-wide studies, it is now clear that genetic and epigenetic alterations work complementarily (to enhance genes silencing/ activating effect) and/or cooperatively (at a stepwise manner to silent/ activate genes) to drive cancer (Brena and Costello, 2007, Shen and Laird, 2013).

Cancer is described as an evolutionary process (Greaves and Maley, 2012, Vogelstein et al., 2013). Cells accumulate mutations intrinsically (such as from errors during DNA replication) and extrinsically (via the exposure to mutagens) throughout lifetime (Stratton et al., 2009). Cells harbouring mutations are either repaired (through various biological mechanisms) or destroyed via apoptosis or activation of immune response if repair is not possible (Svejstrup, 2010). In rare occasion, cells bypass these safeguard mechanisms and continue to accumulate more damaging mutations, resulting in the expression of a mutator phenotype characterized by an exponential mutation rate (predominantly caused by the impairment of genome stability) (Loeb, 2011). The accumulative effects of multiple stably established mutations confer cells with growth advantages, disrupt normal cellular functions such as cell proliferation, apoptosis and differentiation, eventually give rise to cancer exhibiting the six biological hallmarks described by Hanahan and Weinberg (2011b), namely proliferative signalling maintenance, growth suppressors evasion, cell death resistance, replicative immortality, angiogenesis, and invasion and metastasis activation (Svejstrup, 2010, Hao et al., 2016) (Figure 1.2).

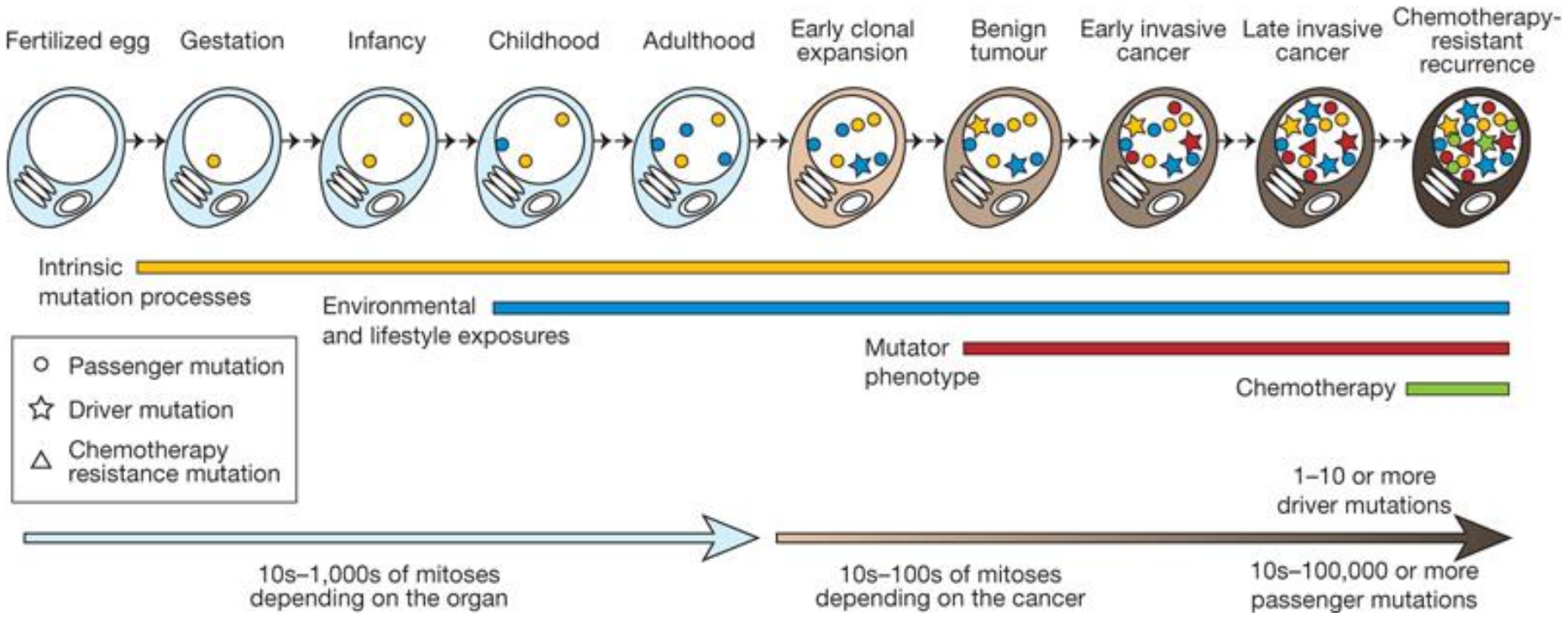


Figure 1.2 The accumulation of mutations throughout lifetime lead to cancer.

Intrinsic mutations resulted from normal cell division and exposures to exogenous mutagens contribute to mutations in cells. Passenger mutations do not alter cell normal phenotype. The first driver mutation/s, such that affects DNA repair, initiate the carcinogenesis process. Accumulation of more driver mutations confer cell with mutator phenotype characterised by remarkably high mutation rate. Chemotherapy may select resistant clones which go on to harbour more mutations leading to cancer relapse. Figure derived from (Stratton et al., 2009) (Licence number: 4292621042501)

1.2.1 Breast Cancer

1.2.1.1 Classification of breast cancer

Breast cancer is the fifth deadliest cancer worldwide contributing to 571000 deaths in year 2015 (WHO, 2017). It is the most common cancer in women with an estimated 1.67 million new cases diagnosed in 2012. Breast cancer is a heterogeneous disease whereby intra- and inter-tumoral heterogeneities contributed significantly to the complexity and failure of breast cancer treatment (Stingl and Caldas, 2007, Polyak, 2011, Schmitz et al., 2016). Therefore, the establishment of an in-depth classification system plays a crucial role to ensure accurate diagnosis, prognosis and drug response predictions as well as effective treatment of breast cancer (Ali et al., 2014).

Different approaches exist for human breast cancer classification. According to the World Health Organisation (WHO), breast cancers are commonly classified based on histopathological diagnosis complemented by epidemiology, macroscopy, genetics and clinical, prognostic and predictive factors (Sinn and Kreipe, 2013). Histopathological classification groups breast cancer according to various morphological features of the tumours. Tumour subtypes classified based on this system include invasive carcinoma of no special type (NST), special subtypes of invasive breast carcinoma, lobular neoplasia, intraductal proliferative lesions, microinvasive carcinomas and others (Lakhani, 2012). Each of these subtypes can be further classified into other breast cancer types. For example, special subtypes of invasive breast carcinoma consist of more

than ten other breast cancer types, including the invasive lobular carcinoma (ILC). While the identification of different minor subtypes has provided significant clinical implications, the refining power of the histopathological classification is considered low as a vast majority (~70 – 80%) of all breast cancers are classified as ILC and Invasive ductal carcinomas of NST (Viale, 2012). On the other hand, the staging and grading systems are also used to classify breast cancers. The tumour, node and metastases (TNM) system developed by the American Joint Committee on Cancer (AJCC) assign a stage (from 0 to IV) to breast cancer based on the size of the tumour, lymph node involvement, and metastases (Edge and Compton, 2010). One of the most commonly used grading systems, the Nottingham Histologic Scoring system, grades breast tumours (from 3 to 9) based on glandular/tubular differentiation, nuclear features and pleomorphism and mitotic (Frierson et al., 1995).

In recent years, gene expression profiling classification has become increasingly important to refine these stratification systems and better categorise breast cancers for effective treatments as well as providing insights into tumour heterogeneity and cellular origin of breast cancers (Schnitt, 2010, Prat et al., 2013, Yao et al., 2015, Russnes et al., 2017, Bayani et al., 2017, Sotiriou and Piccart, 2007). The molecular stratification system breaks breast cancer down into five "intrinsic" subtypes, namely luminal A, luminal B, basal-like, HER2-enriched and normal-like, based on global mRNA expression (Perou et al., 2000, Prat et al., 2013, Bayani et al., 2017). In addition, subsequent gene expression studies have also identified rare subtypes such as the claudin-low and

molecular apocrine subtypes (Herschkowitz et al., 2007, Lehmann-Che et al., 2013). Studies indicated different molecular subtypes exhibit distinct incidence, survival rates and responsiveness to therapies (Table 1.1) (Perou, 2011, Prat et al., 2013). In routine clinical practice, multigene predictors, such as PAM50 (Dowsett et al., 2013) and 70-gene signatures (Drukker et al., 2014) are extensively used to classify breast cancer, bypassing the prohibitive cost required for genome-wide analyses such as microarray and whole genome sequencing. Studies showed that the addition of multigene predictors to clinical risk prediction algorithms accurately classifies breast cancer molecular subtypes and significantly improves the prediction of risk of relapse, prognosis and probability for pathological response (Malhotra et al., 2010, Drukker et al., 2014), indicating the clinical relevance of gene expression–based classification.

Table 1.1 Breast cancer molecular subtypes and their characteristics.

Molecular subtypes	Frequency*	ER/PR/HER2 status (profile of the majority was shown)	Proliferation (as indicated by Ki67)	Histological grade	Prognosis
Luminal A	50-70%	ER+, PR+, HER2-	Low	Low	Excellent
Luminal B	10-20%	ER+/-, PR+/-, HER2-/+	High	Intermediate/ High	Intermediate/ Bad
HER2-enriched	5-15%	ER-, PR-, HER2+	High	High	Bad
Normal breast-like	1-10%	ER-/+, PR+/-, HER2-	Low	Low	Intermediate
Basal-like	10-20%	ER-, PR-, HER2-	High	High	Bad
Claudin-low	7-14%	ER-, PR-, HER2-	High	High	Bad

*Based on (Kim et al., 2006, Carey et al., 2007, Cheang et al., 2009, Banerji et al., 2012, Cancer Genome Atlas, 2012, Sweeney et al., 2014, Yersal and Barutca, 2014, Pereira et al., 2016b, Serrano-Gomez et al., 2016, Russnes et al., 2017). Variation exists between ethnic groups

Luminal A breast cancer accounts for 50-70% of all breast cancers (Fan et al., 2006, O'Brien et al., 2010, Voduc et al., 2010, Perou and Borresen-Dale, 2011). In general, luminal A breast cancer expresses both oestrogen receptor (ER) and progesterone receptor (PR) but not human epidermal growth factor receptor 2 (HER2), thus ER^+, PR^+ and $HER2^+$ tumours are susceptible to hormone therapies such as tamoxifen and aromatase inhibitors (Early Breast Cancer Trialists' Collaborative et al., 2011, Early Breast Cancer Trialists' Collaborative et al., 2015). Luminal A breast cancer has low histological grade, high survival rate and the best prognosis in comparison with other intrinsic subtypes (Perou et al., 2000, Sorlie et al., 2001, Blows et al., 2010, Perou and Borresen-Dale, 2011). However, luminal A is subdivided into five integrative cluster subtypes (IntClust 2, 3, 4, 7, and 8, a classification system based on copy number alterations), with each subtype harbouring distinct copy number driver and slightly different clinical outcomes, suggesting heterogeneity within the intrinsic subtype itself (Curtis et al., 2012). *PIK3CA* (38 – 58%) is consistently reported as the most frequently mutated gene in luminal A breast cancer (Cancer Genome Atlas, 2012, Zang et al., 2016, Pereira et al., 2016a).

In contrast, triple-negative (TNBC) breast cancer is a subtype that neither expresses ER, PR or HER2. It accounts for approximately 12-17% of all breast cancers (Shah et al., 2012, Nik-Zainal et al., 2016). Women of Hispanic and African descents are predisposed to TNBCs as compared to other ethnic groups (Carey et al., 2007). Absence of hormone receptors results in the unresponsiveness of TNBC to hormone therapies, making

systemic therapies, such as chemotherapy (very effective for early stage TNBC) and radiotherapy, the sole treatment options for TNBC (Koboldt et al., 2012). However, due to the high prevalence of *BRCA1/2* germ line mutations (~10 – 40%, associated with DNA repair defects and increased breast cancer susceptibility (Ford et al., 1994, Fong et al., 2009, Hartman et al., 2012)), TNBC is susceptible to PARP inhibitors treatment, a cytotoxic synthetic lethal therapeutic strategy targeting cancers with specific DNA-repair defects (Fong et al., 2009). Approximately 75% of TNBCs are basal-like (Perou, 2011). *TP53* (60 – 82%) and *PIK3CA* (10 – 22%) have consistently been reported as the most significantly mutated genes in TNBC (Perou, 2011, Cancer Genome Atlas, 2012, Koboldt et al., 2012, Shah et al., 2012, Pereira et al., 2016a). TNBC is often associated with early metastatic spread, high histological grade, high recurrence rate and poor prognosis (Perou, 2011, Tate et al., 2012).

1.2.1.2 Cell-of-origin of breast cancer

Stepwise malignant progression from normal, pre-malignant to malignant cancer cells has been extensively studied in attempt to dissect the molecular mechanisms underlying cancer initiation and progression as well as to identify the cellular origin of cancer. As discussed in Section 1.2, the acquisition of driver mutations provide cells with selective growth advantage and may cause pre-malignant lesion such as dysplasia and hyperplasia (Yokota, 2000, Stratton et al., 2009, Du and Elemento, 2015). Further accumulation of genetic and epigenetic alterations couple with aberrant cellular signalling responses and interaction with the microenvironment results in the formation of primary tumours and eventually highly malignant and invasive tumour capable of metastasis (Stratton et al., 2009).

It has been proposed that this multistage process commences with the cell-of-origin of cancer or cancer-initiating cell, which is defined as the normal cell that receives the first drive mutation/s and proceeds to form tumour capable of triggering diagnosis (Visvader, 2011, Blanpain, 2013, Rycaj and Tang, 2015). Therefore, the identification of the cell-of-origin can allow the early detection of cancer, better prediction of treatment outcome and the development of preventive therapies (Blanpain, 2013). It is worth mentioning that the cell of origin is different from a cancer stem cell (CSC) (also called cancer-propagating cell), which itself is a product of cell-of-origin capable of giving rise to a tumour during serial transplantation owing to its acquired self-renewal capability (White and

Lowry, 2015). The study of the cellular origin of cancer is closely related to the concept of lineage-specific variation in mutations, whereby cells of different lineages possess distinct prevalence to different driver mutations, harbour mutations at distinct locations/alleles of a given gene and are differentially affected by the same mutation (Chang et al., 2016b). One of the most common approaches used to identify cell-of-origin of cancer is genetic lineage tracing, whereby conditional activation/ silencing of oncogenes/ tumour suppressor gene is induced by the Cre-Lox system containing a lineage-specific promoter. Formation of tumour by a specific lineage upon the introduction of driver mutation/s indicate cancer-initiating potential (Blanpain, 2013, Rycaj and Tang, 2015). Through this approach, cells-of-origin have been reported in various cancers including skin, brain, gut, lung, prostate and breast (Blanpain, 2013).

Human breast tissue comprises of ducts and lobules organized in a branching network embedded in a stromal matrix (Figure 1.3a). The mammary epithelium is made up of epithelial cells derived from luminal (forming the inner layer) and basal/myoepithelial (the outer layer) lineages (Figure 1.3b & c) (Lim et al., 2009). Despite major efforts made to dissect the mammary epithelial differentiation hierarchy, the exact identity of the different cell populations present remains debatable. According to Skibinski and Kuperwasser (2015), the multipotent mammary stem cell (MaSC) sits at the top of the differentiation hierarchy and capable of giving rise to luminal stem cell and basal/myoepithelial stem cell, both of which subsequently differentiate into progenies making up the breast tissue (Figure 1.4).

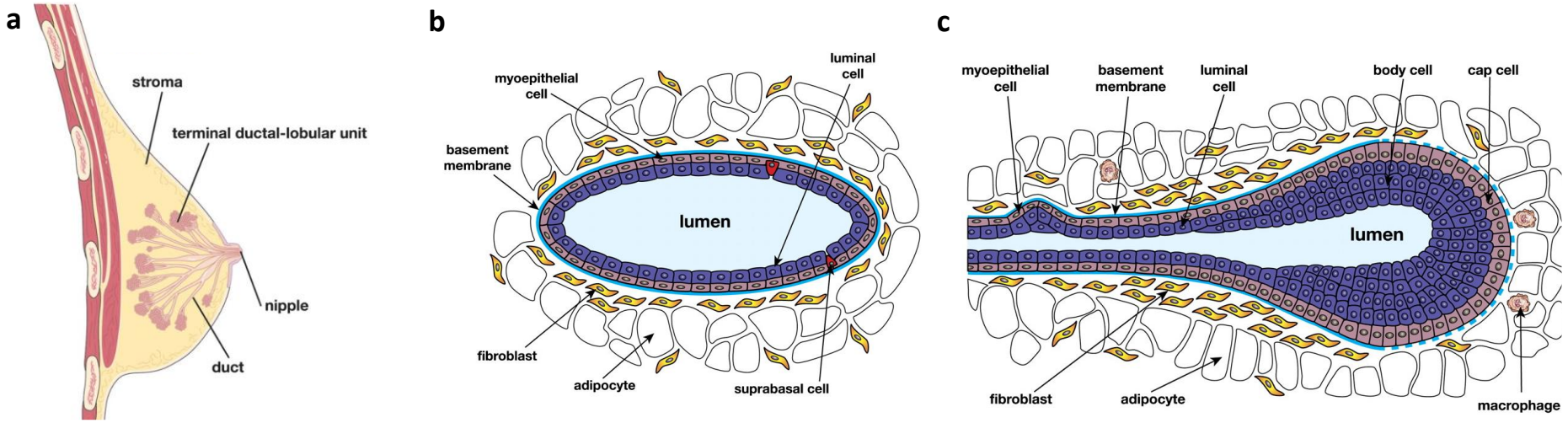


Figure 1.3 Schematic diagrams of human breast tissues.

(a) Human breast tissue structure. (b) Cross-sectional view of mammary duct. (c) Longitudinal view of lobule. Figure derived from Visvader (2009).

Lim et al. (2009) identified and functionally characterized four subpopulations, namely the mammary stem cells/ bipotent progenitors (MaSC/BiPs), luminal progenitors (pLs), mature luminal cells (mLs) and stromal cells in human mammary tissue via *in vivo* transplantation studies and *in vitro* cellular assays, from which an epithelial differentiation hierarchy (MaSC/BiP → pL → mL) was proposed (Lim et al., 2009, Prat et al., 2013) (Figure 1.5). It has been suggested that each of these distinct cell types could potentially initiate tumorigenesis (being the cell-of-origin), giving rise to breast tumours of different subtypes (Blanpain, 2013). Indeed, by comparing the expression profiles between these normal subpopulations to the existing six intrinsic molecular breast cancer subtypes, claudin-low and normal-like subtypes were found best resembled MaSC/BiP, basal-like subtype to pL, luminal A and B subtype to mL and HER2-enriched subptype was found between pL and mL and thought to be induced by HER2 amplification in a luminal lineage-committed cell (Prat et al., 2013) . In support to the above findings, by using Cre-LoxP system to induce conditional deletion of Brca1 and p53 in the mice basal/myoepithelial stem cells ($K14^+$) and luminal progenitor cells (beta lactoglobulin, Blg^+), Molyneux et al., 2010 demonstrated that tumours similar to human BRCA1 basal-like breast cancers and majority of the sporadic basal-like breast tumours were initiated by luminal progenitors and not basal/ myoepithelial stem cells (Molyneux et al., 2010). All the above findings have provided important insights into the cell-of-origin of breast cancer and the possible mechanisms underlying tumour heterogeneity.

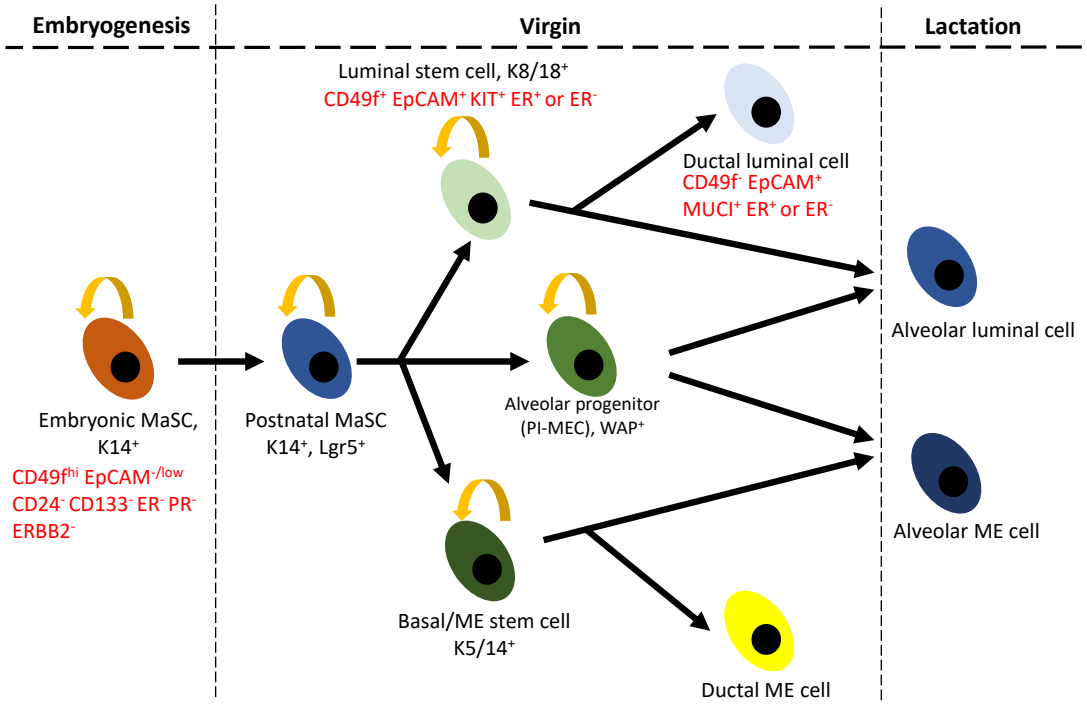


Figure 1.4 Schematic diagram of mammary differentiation hierarchy

Multipotent mammary stem cell (MaSC) tops the hierarchy and give rise to luminal stem cell and basal/myoepithelial stem cell, both of which could further differentiate to generate progenies making up the breast tissue. The position of alveolar progenitor in the hierarchy remains debatable. (Figure created based on Skibinski and Kuperwasser (2015)).

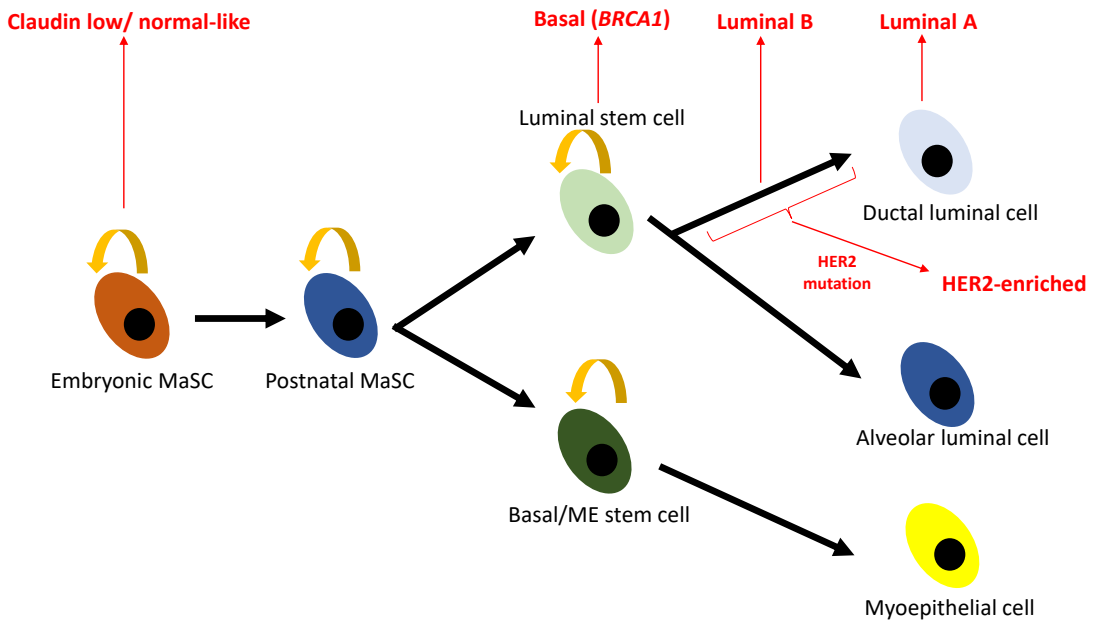


Figure 1.5 Mammary differentiation hierarchy and cell-of-origin of breast cancers proposed by Lim et al. (2009) and Prat et al. (2013).

Each subpopulation in the mammary epithelial differentiation hierarchy were associated with different intrinsic breast cancer subtypes based on gene expression profiles. Figure created based on Lim et al. (2009) and Prat et al. (2013).

1.2.1.3 Breast cancer heterogeneity

Tumour heterogeneity has a significant impact on diagnosis, prognosis, therapeutic resistance and treatment failure (Marusyk et al., 2012, Burrell et al., 2013). The presence of different cell populations within a tumour contributes to intra-tumour heterogeneity. These distinct tumour cell populations differ genetically, phenotypically and have different spatial and temporal existences (Marusyk et al., 2012). On the other hand, the concept of inter-tumour heterogeneity has also been proposed (Burrell et al., 2013), suggesting the existence of genetic and phenotypic variation between the same tumour type in different individuals. As mentioned in

Section 1.2.1.2, the existence of various subpopulations making up the breast tissues and their cancer-initiating potential might contribute to breast cancer's intra- and inter-tumour heterogeneity (Stingl and Caldas, 2007, Skibinski and Kuperwasser, 2015). Two models, namely the clonal evolution and cancer stem cells (CSCs) models, have been proposed to explain tumour heterogeneity.

Cells harbouring different mutations throughout the course of tumorigenesis are conferred with distinct phenotypes. The clonal evolution model proposes that the "fittest" (cancer cells with growth advantage as a result of mutation) is selected for by a Darwinian-like evolution process and constitutes a major part of the tumour (Greaves and Maley, 2012, Marusyk et al., 2012). Burrell et al. (2013) suggested that genomic instability contributes significantly to the overall diversity of subclones within a tumour. In addition to genetic and epigenetic (Meacham and Morrison, 2013) mutations, the tumour microenvironment also exerts significant selective pressure on tumour growth and progression, thus affecting the overall tumour phenotype (Marusyk et al., 2012). Furthermore, therapeutic intervention can induce selection against drug-sensitive clones and concurrently promote the expansion of resistant tumour cells. Such "clonal sweep" could result in a homogenous tumour cell population but render the subsequent treatments more challenging (Greaves and Maley, 2012).

In breast cancer, next generation sequencing of single tumour cells from two triple negative patient indicated similar copy number alternations (CNAs) profiles between primary and metastatic tumours. In addition, tumour progression was found adopted a "punctuated clonal evolution" (as opposed to the gradual clonal evolution model), whereby few/no intermediate populations were detected and each tumour subpopulation was genetically distinct from the original clone (Navin et al., 2011). A subsequent study using the same genome-wide study of single nuclei derived from an oestrogen receptor positive (ER^+) breast cancer and a triple-negative ductal carcinoma revealed each tumour cell was genetically unique (Wang et al., 2014). Consistently, chromosomal rearrangements (a source of copy number alteration (CNAs) was identified as an early event of tumorigenesis which remained largely unchanged throughout clonal expansion and adopted a punctuated evolution model. On the contrary, point mutations (a result of DNA repair or replication machinery defects) was found gradually evolved and intensified clonal diversity. Point mutations adopted gradual clonal evolution whereby increased mutation rates (a mutator phenotype) was conferred by previous mutations, leading to mutation accumulation and eventually genetic and phenotypic diversities in tumour subclones.

In an effort to elucidate the molecular history of 21 breast cancer tumours (starting from normal mammary epithelial cells to tumour capable of triggering diagnosis), Nik-Zainal et al. (2012b) reported the accumulation of thousands of mutations and minimal clonal expansion of normal cells before giving rise to the most recent common tumour ancestor

(with minimal clonal diversity), which subsequently underwent dramatic subclone diversification leading to a genetically-diverse tumour (Figure 1.6). The requirement for hundreds to thousands of mutations before extensive sublone expansion suggested a long-lived, quiescent cell lineage as the cell-of-origin of breast cancer. In addition, the most recent common ancestor appeared early (around 40% of the "molecular time") suggesting much time (the rest 60%) during tumorigenesis was spent on subclonal diversification, supporting the notion of breast cancer being a heterogeneous disease. Furthermore, a dominant subclone comprising more than 50% of tumour cells was consistently detected in each tumour, implicating the "fittest" population that survived the Darwinian-like selection (Nik-Zainal et al., 2012b).

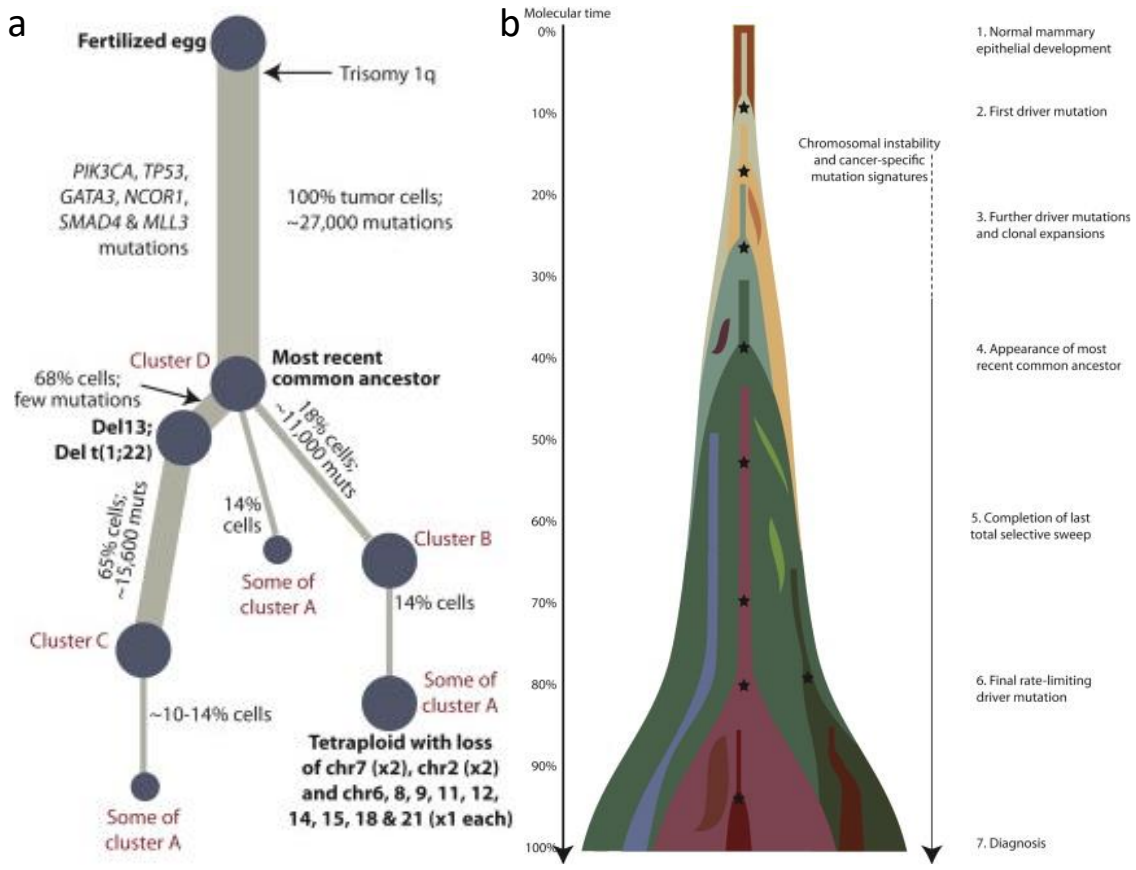


Figure 1.6 The clonal evolution model of breast cancer.
(a) A phylogenetic tree created from in-depth sequencing data of one tumour PD4120a. (b) A proposed clonal evolution model of breast cancer over molecular time. Figure derived from (Nik-Zainal et al., 2012b).

On the other hand, a different but not mutually exclusive model claims that the diverse cell types within a tumour are originated from cancer stem cells (CSCs) (Pardal et al., 2003, Reya et al., 2001, Lobo et al., 2007). CSCs are defined as a small subset of multipotent cancer cells reside within the tumour that possess indefinite self-renewal properties and can differentiate to generate various cancerous progenies (at different developmental stages) forming an entire heterogeneous tumour (Jordan et al., 2006, Pattabiraman and Weinberg, 2014). CSCs have been implicated in tumour initiation, progression, metastasis and recurrence

(Batlle and Clevers, 2017). In addition, increasing evidence shows that CSCs can evade standard chemo- and radio-therapies, by achieving quiescence, expressing ABC drug pumps and anti-apoptotic proteins or preventing DNA damage (Dean et al., 2005, Bao et al., 2006). Studies demonstrated that CSCs express high level of pluripotency-related factors commonly detected in pluripotent stem cells (PSCs), including *OCT3/4* and *SOX2* (Suva et al., 2014). In addition, CSCs and PSCs share various common signalling pathways, including Wnt/β-catenin signalling (Tetteh et al., 2016) and Notch signalling (Sato et al., 2011), important for the maintenance of their stem cells properties. Despite being largely elusive, the origin of CSCs is thought to be from stem cells niche, dedifferentiated progenitor cells or differentiated cells present in the organs affected by oncogenic mutations (Bjerkvig et al., 2005, Medema, 2013). Various CSC markers have been reported for effective identification of CSCs in different solid tumours (Table 1.2).

Table 1.2 CSC markers reported for different solid tumour types. Table derived from Medema (2013).

Breast	Colon	Glioma	Liver	Lung	Melanoma	Ovarian	Pancreatic	Prostate
ALDH1	ABCB5	CD15	CD13	ABCG2	ABCB5	CD24	ABCG2	ALDH1
CD24	ALDH1	CD90	CD24	ALDH1	ALDH1	CD44	ALDH1	CD44
CD44	β-catenin activity	CD133	CD44	CD90	CD20	CD117	CD24	CD133
CD90	CD24	α_6-integrin	CD90	CD117	CD133	CD133	CD44	CD166
CD133	CD26	nestin	CD133	CD133	CD271		CD133	$\alpha_2\beta_1$-integrin
Hedgehog-Gli activity	CD29						c-Met	α_6-integrin
α_6-integrin	CD44						CXCR4	Trop2
CD49f	CD133						nestin	
CD61	CD166						Nodal-Activin	
	LGR5							
Breast	**References**							
$CD44^+/CD24^{-/low}/Lin^-$	Al-Hajj et al. (2003a)							
$CD44^+/CD49f^{high}/CD133/2^{high}$	Meyer et al. (2010)							
$CD44^+/CD24^-/CD133$	Wright et al. (2008)							

The first identification of CSCs in solid tumour was in human breast cancer (Al-Hajj et al., 2003a, Al-Hajj and Clarke, 2004, Kreso and Dick, 2014), whereby a subset of $CD44^{+}/CD24^{-}$ tumour cells was shown capable of being serially passaged and generated a histologically heterogeneous tumour, mimicking the parental tumour in immune-deficient mice. Since then, other breast CSCs (BCSCs) markers have also been identified (Table 1.2) and different combinations of these markers, such as $CD44^{+}/CD24^{-/low}/ALDH^{high}$ have also been used to further enrich CSCs. Isolation by surface markers is commonly complemented by functional assays such as *in vitro* tumour spheres formation and *in vivo* transplantation assay to confirm the identity of CSCs (Charafe-Jauffret et al., 2009, Batlle and Clevers, 2017). The successful identification of CSCs has streamlined CSCs studies and Table 1.3 illustrates pathways and factors that have been associated with CSCs regulation.

Table 1.3 Mechanisms and factors associated with cancer stem cells regulation (adapted from by Guo (2014)).

Regulation mechanisms	Description
Epithelial-mesenchymal transition (EMT)	• Itself regulated by multiple pathways. • The activation of EMT by transforming growth factor β (TGF-β) was shown to promote formation of $CD44^{high}CD24^{low}$ cells and tumour initiating ability • The overexpression of *Slug* (or *Snail2*), a key EMT transcription factor, has also been associated with CSCs properties (Guo et al., 2012)
Wnt signalling	• Associated with cellular plasticity and stem cell maintenance (Reya and Clevers, 2005). • Loss of the epithelial adhesion molecule E-cadherin lead to redistribution or elimination of β-catenin and activation of Wnt signalling, which has been associated with cancer stem cell maintenance and tumour metastasis (Reya et al., 2001)
Hippo pathway	• Activation of the Hippo pathway effectors YAP/TAZ has been associated with enhanced mammosphere-forming ability, *in vivo* tumorigenicity, and metastasis (Cordenonsi et al., 2011, Chen et al., 2012).
microRNA	• MiR-200 family regulate EMT and stem cell properties. Inhibition of miR-200 increased breast CSCs (Iliopoulos et al., 2010)
SOX Family Transcription Factors	• Sox2 expression is associated with embryonic stem cell-like gene expression signatures in breast cancer and contribute to CSCs properties (Leis et al., 2012)
Cytokine Networks	• Colony-stimulating factor-1 (CSF-1) secreted by carcinoma cells promote cancer cell invasion • Cytokines, such as Interleukin -6 , -8, and -1β, promote BCSC self-renewal and survival (Korkaya et al., 2011)

At present, more studies are required to fully elucidate the cellular and molecular mechanisms underlying BCSCs generation and maintenance. Several limitations impeding BCSC studies should be circumvented, including inadequateness of samples as CSCs only make up a very small part of the tumour, difficulty in isolating homogenous CSCs population from clinical samples and the lack of optimal cell culture conditions *in vitro* for CSCs maintenance (Calvet et al., 2014, Nishi et al., 2014a). In addition, approaches such as, lineage tracing assay, tumour organoids and intravital imaging, should be exploited to investigate CSCs in their intact tumour microenvironments (Guo, 2014, Batlle and Clevers, 2017). Collectively,

these efforts could contribute to the development of selective therapeutic strategies against CSCs and thus cancer itself.

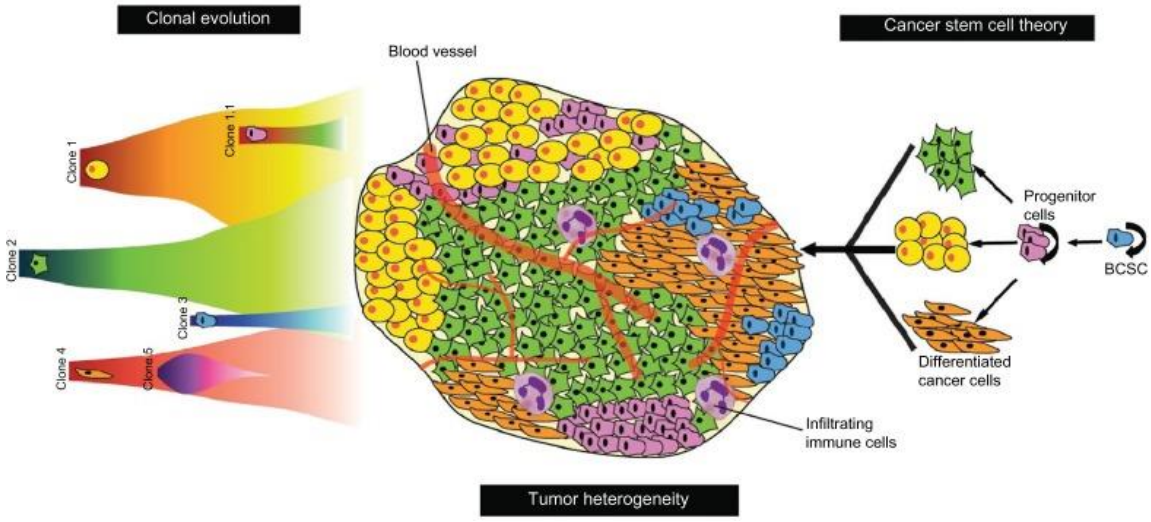


Figure 1.7 Schematic diagram of tumour heterogeneity in breast cancer.

The clonal evolution model proposes cancer cells conferred with growth advantages by mutations are selected for through a Darwinian-like evolution process and constitutes a major part of the tumour. Cancer stem cell model suggests the presence of a small subset of multipotent cancer cells capable of indefinite self-renewal within the tumour. The CSCs can differentiate to generate various cancerous progenies, representing different developmental stages, to form an entire heterogeneous tumour. BCSC = breast cancer stem cell. Figure derived from Shah and Allegrucci (2012) (Licence number: 4292630834137).

1.2.2 Cancer genetics

Cancer genetics has revolutionised oncology from providing general and in-depth understandings to effective treatment of cancers (Ciriello et al., 2013). Comprehensive pan-cancer sequencing studies involving large sample size and a variety of tumour types have provided important insights into commonality and variability across cancer types. These

findings are crucial to fully elucidate the molecular landscape of cancer, which in turn has significant impact on the development of targeted therapies (Zack et al., 2013). For example, common genetic alterations have been consistently reported in different studies suggesting the prospect of expansion of approved molecularly targeted therapies across different cancer types (Cancer Genome Atlas Research et al., 2013). A recent pan-cancer studies highlighted the importance of understanding cancer genetics for clinical applications (Zehir et al., 2017a).

By utilising a hybridization capture–based deep coverage NGS (MSK-IMPACT), Zehir et al. (2017a) analysed a cohort of 10,945 tumours from 10336 patients, encompassing a wide range of advanced (with 43% samples isolated from metastatic sites) solid tumour types (62 principle and 361 detailed tumour types). A comprehensive view on protein-coding mutations, copy number alterations (CNAs) and selected promoter mutations and structural rearrangements in 341 (2809 or 26% tumours) and 410 (8136 or 84% tumours) cancer-associated genes have been reported. A total of 78066 nonsynonymous mutations, 22989 CNAs and 1875 rearrangements were detected. MSK-IMPACT identified approximately 69% of somatic mutations not previously reported in the Catalog of Somatic Mutations in Cancer (COSMIC) database (v78) but were consistent with TCGA data for 16 common tumour types, including tumour identities and cancer mutations frequencies of different lineages. However, in the MSK-IMPACT cohort, more mutations were detected in significantly mutated genes previously reported in TCGA. For examples, *TP53* (associated with tumour aggressiveness) and genes associated with

hormone therapy resistance such as *AR* and *ESR1* in prostate and breast cancer, in line with the metastatic status of all specimens used in the study.

TP53 was reported as the most commonly mutated gene with 41% of patients or >10% incidence in 43 of the 62 principal tumour types (Zehir et al., 2017). It was followed by *KRAS* mutations (detected in 15% of patients) which also harboured the most frequently mutated codon (G12) contributing to 80% of all *KRAS* mutations and 12% of all tumours. Pancreatic adenocarcinoma (90%) and colon adenocarcinoma (44%) were most susceptible to *KRAS* mutations. $KRAS^{G12}$ was followed by $PIK3CA^{H1047}$, $PIK3CA^{E545}$ and $BRAF^{V600}$ as next few most frequently mutated codon, having prevalence in >20 principal tumour types. On the other hand, data also indicated lineage-specificity of cancer mutations (Chang et al., 2016b), whereby different mutational locations were detected within the same genes across different cancer types. For example, prevalence of mutations within extracellular N-terminal domain and kinase domain of *EGFR* were observed in glioma and lung cancer respectively. In addition, some genes were more frequently mutated in certain cancer types than the other such as *VHL, APC* and *IDH1* as well as rearrangements such as *TMPRSS2-ERG* in prostate cancer.

Genomic rearrangements were detected in 1597 or 15% of tumours and non-coding mutations in *TERT* promoter was also reported, both showing high cancer type-specificity. Based on base substitution data, the authors assigned tumours to a set of 30 mutational signatures reported

elsewhere (Alexandrov et al., 2013). It is far from surprising, distinct signatures were assigned to different tumour types, for example lung to smoking, melanoma to UV exposure and glioma to prior radiotherapy. Lastly, the clinical utility of MSK-IMPACT data was evaluated through identification of clinical actionability of mutations identified. Results indicated 36.7% or 3792 patients harboured at least one actionable mutation and gastrointestinal stromal tumour (76%), thyroid cancer (61%), breast cancer (61%) and melanoma (58%) were having the highest actionable alterations, suggesting the significance of tumour molecular profiling in the discovery and development of targeted therapies (Zehir et al., 2017a).

1.2.2.1 Genetic Dysregulation in Breast Cancer

Some breast cancers are hereditary. Germline mutation of cancer susceptibility genes significantly increase the risk of breast cancers (Meindl et al., 2010, Sun et al., 2017). *BRCA1* and *BRCA2* genes are two tumour suppressor genes associated with transcriptional regulation and DNA repair (Turner and Reis-Filho, 2006, Atchley et al., 2008). Women inherited with mutated *BRCA1* and *BRCA2* genes have a strong predisposition (as high as 80%) to breast cancer (Ford et al., 1994). In addition, *TP53* (Malkin et al., 1990), *PTEN, ATM* (Renwick et al., 2006), *BRIP1* (Seal et al., 2006), *CHEK2* (Walsh et al., 2006), *MAP3K1* (Glubb et al., 2015) and *ATM* (Goldgar et al., 2011) have also been identified as breast cancer susceptibility genes. A recent genome-wide association studies (GWAS) study have also implicated 19 known and 4 novel breast cancer risk loci, namely two loci

at 13q22 near *KLF5*, a 2p23.2 locus near *WDR43* and a 2q33 locus near *PPIL3* in ER^- breast cancer (Couch et al., 2016).

Large-scale genome-wide studies have provided important insights into the mutational landscapes of breast cancer. Major findings from several such studies are summarised in Table 1.4. In general, breast cancer has a relatively low somatic mutation prevalence of 0.19 per Mb in comparison with other cancer types, including lung, gastric and ovarian cancers (Greenman et al., 2007). A comprehensive study by The Cancer Genome Atlas in 2012 integrated exome sequencing with epigenome, transcriptome, proteome analyses and microRNA profiling of 510 tumours from 507 patients revealed a total 30626 somatic mutations encompassing 28319 point mutations, 4 dinucleotide mutations and 2302 indels (insertion and deletion) in four breast cancer subtypes. *PIK3CA, PTEN, AKT1, TP53, GATA3* and *MAP3K1*, genes previously implicated in breast cancer, along with other novel genes such as *TBX3, RUNX1, CBFB, AFF2, PIK3R1, PTPN22, PTPRD, NF1, SF3B1* and *CCND3* have been reported as significantly mutated genes (SGMs) (genes with somatic mutation rate higher than the stochastic rate and believed to be the drivers of cancer) (Cancer Genome Atlas, 2012, Goncalves et al., 2014). Breast cancer represents one of the most heterogeneous cancer with each subtype harbouring remarkably distinct somatic mutations, copy number alterations and epigenetic configurations. Indeed, only three genes (*TP53, PIK3CA* and *GATA3*) were identified as having >10% incidence across all subtypes. Furthermore, luminal A and B tumours were found harbouring more diverse and recurrent SMGs. Luminal A has the lowest overall mutation rate (but the most SMGs) and basal-like and HER2-enriched

subtypes the highest. *PIK3CA* is the most frequent SMG for luminal A (45%) and B (29%, same frequency as *TP53*). In stark contrast, 80% of basal-like tumours acquired *TP53* mutations and possesses SMGs that rarely, if ever, present in the luminal SMG list (except for *PIK3CA*). On the other hand, remarkably low SMG frequency but high prevalence of *TP53* (72%) and *PIK3CA* (39%) mutations were detected in the HER2-enriched subtype. In addition to distinct SMGs, different subtypes exhibit distinct mutation types, further highlighting breast cancer heterogeneity. For example, nonsense and frameshift mutations are most commonly found in *TP53* of basal-like tumour, while missense mutation are predominantly detected in luminal A and B tumours (Cancer Genome Atlas, 2012).

Table 1.4 Large-scale genome-wide studies of breast cancers.

References & scope of study	Mutational landscape	Possible driver genes	Other findings
Curtis et al. (2012) • Genomic & transcriptomic analyses of a discovery and validation set of 997 and 995 primary breast tumours	• Germline copy number variants (CNVs) • Acquired somatic copy number aberrations (CNAs) • Inherited CNVs and SNPs, and CNAs affected the expression of >39% genes (11198/28609), the latter has the most effect on expression profiles • Inherited CNVs and SNPs accounted for 2.9% and CNAs 85.2% • Total *Cis*-acting and *trans*-acting genes were 5942 and 5947.	• Possible drivers with *cis*-acting mutations in: *ZNF703, PTEN, MYC, CCND1, MDM2, ERBB2, CCNE1* and putative drivers *MDM1, MDM4, CDK3, CDK4, CAMK1D, PI4KB, NCOR1* • *Trans*-acting aberration hotspots (affecting >30 mRNAs) can be grouped into seven pathway modules: for example, two hotspots of T-cell-receptor (TCR) loci on chromosomes 7 (TRG) and 14 (TRA) modulated 381 and 153 mRNAs respectively and co-modulated 19 mRNAs, affecting T-cell activation and proliferation, dendritic cell presentation, and leukocyte activation	• Integration of copy number and gene expression data revealed 10 novel subgroups (IntClust 1–10) with distinct copy number profiles, pathway deregulation and clinical outcomes. • IntClust subgroups also correlated with intrinsic subgroups • ER-positive enriched with regions with heterozygous and homozygous deletions *PPP2R2A* (8p21, region 11), *MTAP* (9p21, region 15) and *MAP2K4* (17p11, region 33)
Banerji et al. (2012) • Exome sequencing (targeting 189980 exons) of 103 tumours 54 from Mexico and 49 from Vietnam patients	• Mutation rate was 1.66 per Mb (0.47–10.5) • Non-silent mutation rate was 1.27 per Mb (0.31–8.05) • 4985 candidate somatic mutations, including 3153 missense, 1157 silent, 242 nonsense, 97 splice site, 194 deletions, 110 insertions and 32 other mutations, in the protein-coding sequences and the adjacent splice sites • Mutations ranging from 14 to 307 per tumour	• Six recurrent somatic mutations in *PIK3CA* (27%), *TP53* (27%), *AKT1, GATA3* and *MAP3K1* and a novel gene *CBFB*	• Rearrangement rate: 30 in luminal A tumour (0–218), 237 and 246 in basal-like and Her2-enriched tumours respectively • MAGI3–AKT3 fusion resulted from rearrangement was enriched in triple-negative subtype leading to constitutive activation of AKT kinase
Stephens et al. (2012) • Analyses of somatic copy number changes, exon sequences of 21416 protein-coding genes and sequences	• 7241 somatic point mutations: 6964 single-base substitutions (4737 missense, 422 nonsense, 158 splice site, 8 stop codon read-through and 1637 silent changes).	• 40 cancer genes identified, including *AKT1, BRCA1, CDH1, GATA3, PIK3CA, PTEN, RB1* and *TP53* and new breast cancer genes including *AKT2, ARID1B, CASP8, CDKN1B,*	• Mutation number was associated with age of diagnosis and tumour histological grade • Multiple mutational signatures identified, such as mutations of

of 1664 microRNAs in 100 tumours (79 ER^+ and 21 ER^-)	• Two substitutions in microRNAs. • 277 small indels (231 frameshifts and 46 in-frame) • Copy number variations: 1712 homozygous deletions and 1751 regions amplification • Maximum 6 drivers per tumour with 20 tumours harbouring only one driver	*MAP3K1, MAP3K13, NCOR1, SMARCD1* and *TBX3* • *TP53, PIK3CA, ERBB2, MYC, FGFR1/ZNF703, GATA3* and *CCND1* had >10% incidence, contributing to 58% driver mutations • 73 different combinations of mutated cancer genes suggesting intertumoral heterogeneity	cytosine at TpC dinucleotides (in 10% of tumours)
Nik-Zainal et al. (2012a) • Whole-genome sequencing of 21 tumours integrated with mRNA expression data of 17 tumours	• 183916 somatically acquired base substitutions • In protein coding regions, there were 1372 missense, 117 nonsense, 2 stop-lost, 37 essential splice-site, and 521 silent mutations in protein coding regions • Of the 2869 indels identified, 2233 were deletions, 544 insertions and 92 complex • 21 coding indels, of which 15 were predicted to result in a translational frameshift and six were in-frame. In addition, 1192 structural variants (rearrangements), 16 homozygous deletions, and 14 regions of increased copy number (amplifications) were identified	• Cancer genes with possible driver substitutions and indels: *TP53, GATA3, PIK3CA, MAP2K4, SMAD4, MLL2, MLL3,* and *NCOR1*	• Five signature A-E based on base substitution • BRCA1 or BRCA2 mutations closely related to specific substitution mutation signatures and deletion profiles • Identified "kataegis", localised base-substitution hypermutation • Base substitutions in kataegis regions were predominantly cytosine at TpC dinucleotides and associated with somatic rearrangements • The APOBEC family of cytidine deaminases may be associated with kataegis
Pereira et al. (2016a) • Exome sequencing of 173 genes in 2,433 breast tumours	• 32476 somatic mutations: 13084 protein coding sequences, including 11006 SNVs (10,193 missense, 808 nonsense, 5 read-through) and 1635 indels (1315 frameshift & 320 in-frame) • Average 13 in each tumour • 131 tumours harbouring at least 30 mutations • 76 devoid of coding mutations • 38 tumours devoid of any mutations	• 40 mutation driver genes • Most frequently mutated genes: *PIK3CA* (40.1%), *TP53* (35.4%), *MUC16* (16.8%), *AHNAK2* (16.2%), *SYNE1* (12.0%), *KMT2C* (or MLL3, 11.4%) and *GATA3* (11.1%)	• Germ line mutations: BRCA1 and BRCA2 (1.36% and 1.64%), *CHEK2* (2.22%), TP53 (0.82%) • Major pathways affected by mutation driver genes: AKT signalling cell cycle regulation, chromatin function, DNA damage and apoptosis MAPK signalling tissues organisation and transcription regulation and ubiquitination

			• Significant intertumoral heterogeneity between IntClust groups was observed in terms of driver mutations distribution, clinical and pathological characteristics of tumours, prognosis and degree of intratumoral heterogeneity
Nik-Zainal et al. (2016) • Whole-genome sequences of 560 breast cancers	• 3,479652 somatic base substitutions, 371993 small indels and 77695 rearrangements	• 93 protein-coding drivers (31 dominant, 60 recessive, 2 uncertain) • Ten most frequently mutated genes accounted for 62% of all drivers: *TP53, PIK3CA, MYC, CCND1, PTEN, ERBB2, ZNF703/FGFR1* locus, *GATA3, RB1* and *MAP3K1* • Possible driver-substitutions and indels identified in several non-coding regions: promoter of *PLEKHS1, TBC1D12 WDR74* as well as two long ncRNAs *MALAT1* and *NEAT1*	• Five novel breast cancer genes: *MED23, FOXP1, MLLT4, XBP1, ZFP36L1* • Mutational signature identified: two indel signatures, twelve base substitution and six rearrangement • Kataegis (localised base-substitution hypermutation) observed in 49% breast cancers

*abbreviation: ER = oestrogen receptor, indels = insertion & deletion, ncRNA = non-coding RNA

1.2.3 Epigenetics in cancer

Epigenetics is defined as the study of the mitotically and/or meiotically heritable changes in gene expression which occur independent of DNA sequence alteration (Weinhold, 2006, Moutinho and Esteller, 2017). Despite being a relatively young branch of science, the pivotal role of epigenetics in all biological processes has been widely recognized (Waddington, 2012). Four epigenetic regulation mechanisms, namely DNA methylation (Jin et al., 2011), histone modification (Bannister and Kouzarides, 2011), chromatin remodelling (Narlikar et al., 2013) and non-coding RNA (ncRNA) regulation (Mattick and Makunin, 2006), have been extensively studied (Jang and Serra, 2014). These mechanisms are responsible for the formation of two chromatin states, heterochromatin (condensed and inactive areas) and euchromatin (loosely packed and actively transcribed areas), which regulate normal gene expression (Gilbert et al., 2004, Zakrzewski et al., 2011). Epigenetic dysregulation has been implicated in various human diseases, ranging from autism to cancer (Berdasco and Esteller, 2010, Portela and Esteller, 2010, Sandoval and Esteller, 2012).

1.2.3.1 DNA methylation

DNA methylation is a stable epigenetic mark responsible for biological processes such as gene silencing, imprinting and X-chromosome inactivation (Law and Jacobsen, 2010). During the methylation process a methyl group ($-CH_3$) is covalently added, from S-adenosyl-L-methionine (SAM), predominantly (~98%) to the 5th position of cytosine residues of cytosine-phosphate-guanine (5'-CpG-3') dinucleotides (Jin et al., 2011).

This process is mediated by three members (DNMT1, DNMT3a and DNMT3b) of the DNA methyltransferases (DNMTs) family. No methyltransferase activity has been reported in the other two of the five members, DNMT2 and DNMT3L (Jin et al., 2011, Lyko, 2017). Interestingly, DNA methylation can also occur in non-CpG sites as demonstrated in approximately 25% of the methylation sites in embryonic stem cells (ESCs), which have been identified as CHG and CHH sites (where H = A, C or T) (Lister et al., 2009).

CpG islands (CGIs) are areas (ranging from 0.2 to 3 kb) enriched with most commonly unmethylated CpG dinucleotides (Illingworth and Bird, 2009, Deaton and Bird, 2011). CGIs are found within majority (60%) of human gene promoters and are normally unmethylated (Portela and Esteller, 2010). Considering its location, it is not surprising that most, if not all, CGIs are responsible for transcriptional regulation (Deaton and Bird, 2011). In general, hyper- and hypo-methylation of CGIs suppresses and activates gene expression respectively (Portela and Esteller, 2010). In addition, the CpG island shores (a region located within ~2kb from the CGIs and characterized by a lower CpG density) is also a target of methylation. On the other hand, DNA methylation can also be detected at the gene body and repetitive sequence (Portela and Esteller, 2010).

Genome-wide hypomethylation (overall 5-methyl-cytosine reduced by 20–60%) and local hypermethylation of CpG islands are well-known epigenetic hallmarks of cancer (Rodriguez-Paredes and Esteller, 2011).

The methylation in gene bodies are associated with increased gene expression. Gene body DNA hypomethylation has been implicated in various cancers (Kulis et al., 2012). In addition, genome-wide hypomethylation re-activates repetitive sequences such as long interspersed element-1 (LINE-1) and contributes to genome instability, drug resistance and poor prognosis in different cancers (Daskalos et al., 2009, van Hoesel et al., 2012, Criscione et al., 2014, Burns, 2017). Besides, hypomethylation activation of cancer-related genes have also been identified in various cancers including prostate (Cytochrome *P450 1B1*), pancreatic (*MASPIN* and *S100P*), breast (*CDH3*) and ovary (*SNCG*) cancers (Wilson et al., 2007). On the contrary, the aberrant DNA hypermethylation in cancer occurs at specific CGIs and CpG island shores, predominantly affecting tumour suppressor genes, genes responsible for DNA repair and differentiation as well as microRNAs (Jones et al., 2016). The hypermethylation silencing of tumour suppressor genes was deemed to be cancer specific. A study investigating promoter hypermethylation changes in 12 genes (*p16INK4a, p15INK4b, p14^{ARF}, p73, APC5, BRCA1, hMLH1, GSTP1, MGMT, CDH1, TIMP3* and *DAPK*) from ~600 primary tumour samples encompassing 15 tumour types revealed that despite having several common gene alterations, each cancer type investigated exhibited a unique promoter hypermethylation profile. As a result, a panel of 3 – 4 markers was generated to accurately (70 – 90%) define abnormality in each cancer type, suggesting aberrant DNA methylation can serve as biomarker for the detection of different human cancer (Esteller et al., 2001). In addition, cancers with positive CpG island methylator phenotype (CIMP) are characterised by high degrees of CpG-rich promoters

hypermethylation silencing of tumour suppressor or tumour suppressor-related genes and epigenetic instability (Issa, 2004, Fu et al., 2016). The CIMP has also provided a valuable approach for diagnosis and prognosis of cancer patients based on DNA methylation (Sato et al., 2010, Flavahan et al., 2016).

1.2.3.2 Histone modifications

Chromatin is made up of nucleosome units, whose core particle consists of DNA wrapped around histone octamers (comprising two copies of the four different types of histones subunits: H2A, H2B, H3 and H4) (Lennartsson and Ekwall, 2009). Each nucleosomal core particle is wrapped around by approximately 146 base pair of DNA and is linked to the adjacent nucleosome by linker DNA (around 10-60 base pair) forming a "beads on a string" structure. The presence of the linker histone H1 binding to the core particle around the DNA entry and exit (~10 bp) sites forms a chromatosome. H1 functions to stabilise the nucleosomes and support higher-order chromatin structure (Fyodorov et al., 2017). The N-terminal tail of each histone (histone tails) protrudes from the surface of nucleosome and subject to at least 16 classes of post-translational modifications, including methylation, acetylation, phosphorylation and ubiquitination (Kouzarides, 2007, Bannister and Kouzarides, 2011, Dawson and Kouzarides, 2012). Distinct modifications act sequentially or in concert to generate a specific histone code, which affects different downstream processes through induction of structural changes (such as chromatin decondensation) or by recruiting/inhibiting the binding of effector proteins such as remodelling enzymes, chromatin modifiers and

transcription factors (Strahl and Allis, 2000, Bannister and Kouzarides, 2011, Fullgrabe et al., 2011, Lawrence et al., 2016). For instance, transcriptionally active gene promoters are normally associated with lysine acetylation (H3K12ac, H3K9ac and H3K18ac) and lysine trimethylation (H3K4me3), while lysine methylation such as H3K9me2, H3K8me3 and H4K20me2 are frequently detected in inactive promoter regions (Cao et al., 2015). On the other hand, different amino acid residues in histones are subject to different modifications. For instance, 13 modification sites for methylation, acetylation, phosphorylation and ubiquitination have been found in histone H2A (Lennartsson and Ekwall, 2009). Histone modifications are known to regulate chromatin compaction, nucleosome dynamics, gene expression, DNA repair, replication and recombination. Besides, it has also been shown crucial for cell differentiation and more recently somatic cell reprogramming (Apostolou and Hochedlinger, 2013, Theunissen and Jaenisch, 2014a, Lawrence et al., 2016).

Alteration of global histone modification patterns can compromise the structure and integrity of the genome, leading to aberrant gene expression, DNA repair, cell cycle checkpoints and eventually cancer (Esteller, 2007, Fullgrabe et al., 2011). Several histone marks, including histones H3 and H4 deacetylation, loss of histone H3 lysine 4 (H3K4) trimethylation and gain of H3K9 methylation and H3K27 trimethylation, have been associated with promoter hypermethylation in cancer (Esteller, 2007). In addition, DNA repetitive sequence hypomethylation in cancer has been associated with the global losses of K16 acetylation and K20 trimethylation in histone H4 (Fraga et al., 2005, Esteller, 2007). Two

independent studies on breast and liver cancers supported these findings and further emphasized the key roles of other modifications, such as H3K9me3, in tumorigenesis (Pogribny et al., 2006, Tryndyak et al., 2006). Increase in H3K36 methylation and a decrease in global K27 methylation have also been associated with open chromatin structure in multiple myeloma (Martinez-Garcia et al., 2011).

In addition to histone marks, tumours cells exhibited distinct histone-modifying enzymes expression patterns from their normal counterparts. These discrepancies could be used to distinguish different human solid cancers (Ozdag et al., 2006, Fullgrabe et al., 2011). On the other hand, the alteration of histone modifying enzymes, such as histone deacetylase (HDAC) 1 and 2 (HDAC1 and HDAC2), histone acetyltransferases (HATs such as hMOF and EP300) and histone methyltransferases (HMTs, such as EZH2) have been closely associated with cancers (Gayther et al., 2000, Song et al., 2005, Pfister et al., 2008, Simon et al., 2012). The disruptor of telomeric silencing 1 like (DOT1L) belongs to the family of lysine methyltransferases (KMTs) which catalyses the methylation of histone H3 lysine (K) 79 (H3K79) in mammals (Kim et al., 2012). Binding of DOL1T to MLL1 fusion proteins in acute leukaemia was shown to induce H3K79me2 hypermethylation of MLL1-regulated genes, increase their expression, leading to hematopoietic cell differentiation blockage (a hallmark of many blood cancers) (Okada et al., 2005, Mueller et al., 2007). In addition, reduced proliferation, self-renewal, malignancy were also reported through drug inhibition of DOL1T in breast cancer cells (Zhang et al., 2014a).

1.2.3.3 Non-coding RNAs (ncRNAs)

All biological functions and phenotypes in organisms were thought to be governed only by genes that encode for specific proteins. However, an effort to map functional elements across the human genome by using data collected from sequencing-based studies indicates that approximately 80.4% of the human DNA is biochemically functioning and only 2.94% could give rise to proteins (Consortium, 2012). Many of these non-coding regions, such as cis-regulatory elements (e.g. introns and intergenic sequences) and antisense sequences, are expressed, giving rise to different classes of small (20 – 30 nucleotide) and long (>200 nucleotide) non-coding RNAs (ncRNA) (Mattick and Makunin, 2006, Holoch and Moazed, 2015). These ncRNAs can be broadly classified into infrastructural and regulatory ncRNAs. Ribosomal, transfer, small nuclear and small molecular RNAs are constitutively expressed infrastructural ncRNAs. Regulatory ncRNAs can be further classified into different subgroups, namely microRNA (miRNAs), Piwi-associated small RNAs (piRNAs), small interfering RNAs (siRNAs), long ncRNAs (lncRNAs), transcription initiation RNAs, (tiRNAs), miRNA-offset RNAs (moRNAs), promoter-associated RNAs (PARs) and so on (Kaikkonen et al., 2011, Batista and Chang, 2013, Cech and Steitz, 2014). NcRNAs play key role in various biological processes, including gene regulation, chromosome remodelling, RNA editing, translational inhibition and mRNA destruction (Mattick and Makunin, 2006, Kaikkonen et al., 2011, Ling et al., 2017). MicroRNAs (miRs) is evolutionary conserved ncRNAs of 18-25 nucleotides in length (Garzon et al., 2010). Through its canonical pathway, miRs regulate gene expression via complementary binding between its seed sequence (position 2-8 from the 5' miRNA) with the 3'

untranslated region (3' UTR) of target mRNAs, leading to their destruction or translation inhibition depending on the degree of complementarity (Hayes et al., 2014, Romano et al., 2017).

MiRs have become a major focus in cancer research over the past decade (Hayes et al., 2014). The availability of genome-wide data has permitted the identification of miR signatures which facilitate cancer classification, detection, prognosis and drug response predictions (Hayes et al., 2014, Li et al., 2015a). Early study using bead-based flow cytometric miR expression profiling revealed a global downregulation of miRs in tumours as compared to normal tissues (Lu et al., 2005). These miR profiles more accurately classified poorly differentiated tumours in comparison to mRNA profiles, a notion also supported by subsequent studies (Gommans and Berezikov, 2012, Yang et al., 2012, Krell et al., 2013, Ng et al., 2013). Downregulation of miR has potentially compromised the cell division inhibition and promotion of termination differentiation abilities of miR leading to generation and maintenance of cancer stem cells (Lu et al., 2005). In agreement, many miRs targets have been associated with tumour growth, invasion and metastasis, angiogenesis, metabolic reprogramming, evasion of cell death and immune destruction (Kasinski and Slack, 2011, Stahlhut and Slack, 2013, Ling et al., 2017).

1.2.4 Breast cancer epigenetics

Epigenetic studies in breast cancer provided important insights into mechanisms, prevention, prognosis and treatment of the disease

(Stefansson and Esteller, 2013). Alteration of the three epigenetic regulation mechanisms mentioned earlier have been widely associated with breast cancer. Aberrant silencing by DNA methylation of genes responsible for DNA repair (*BRCA1*), cell-cycle regulation (*p16*) and other tumour suppressor genes (*APC* and *CDH1*) have been reported in breast cancer (Hoque et al., 2009, Chik and Szyf, 2011, Lustberg and Ramaswamy, 2011, Fleischer et al., 2017). Besides, dysregulations of post-translational histone modifications (such as H3K4 acetylation, H3K9 acetylation and H3K27 methylation) and non-coding RNA (such as miR-21, miRNA-497 and miR-155) have also been implicated in different breast cancer subtypes (Margueron and Reinberg, 2011, Li et al., 2015c, Judes et al., 2016, Karsli-Ceppioglu et al., 2017, Mekala et al., 2018).

On the other hand, a host of epigenetic biomarkers have been proposed by many studies to refine the detection, prognosis and treatment outcome prediction for breast cancer. For example, Hoque et al. (2009) reported methylation alterations in *APC, CDH1* and *CTNNB1* as highly specific and sensitive markers for the detection of early stages of breast cancer. In addition, the hypermethylation of *TWIST, RASSF1A, CCND2*, *HIN1*, *APC*, *CDH13*, *GSTP1 BIN1, BMP6, ESR2, CDKN2A, CDKN1A, TIMP3,* and *CST6* have all been demonstrated as potential diagnostic biomarkers of breast cancers in different studies (Toyooka et al., 2001, Fackler et al., 2004, Feng et al., 2007, Jeronimo et al., 2008, Sunami et al., 2008, Radpour et al., 2009). Furthermore, *RASSF1A* promoter methylation has been suggested as a good predictor for breast cancer hormone (oestrogen and progesterone receptors) status, which in turn inform breast cancer

prognosis (Kajabova et al., 2013). On the other hand, promoter hypermethylation silencing of *BRCA1* can predict breast tumour sensitivity to platinum-derived drugs (Stefansson et al., 2012) and poly(ADP-ribose) polymerase (PARP) inhibitors (Veeck et al., 2010). Widschwendter et al. (2004) also reported that high *ESR1* gene methylation is a good survival predictor of breast cancer patients after Tamoxifen treatment and Tamoxifen-treated patients with positive *CYP1B1* methylation are also predicted to have better survival. Recently, circulating cell-free DNA, mRNA and miRNA have come into the limelight as being valuable biomarkers for breast cancers. For example, cell-free DNA harbouring promoter hypermethylation at genes such as *ITIH5, DKK3,* and *RASSF1A, SFN, CDKN2A, MLH1, HOXD13, PCDHGB7* and *PITX2* have all been implicated in breast cancer (Gobel et al., 2011, Kloten et al., 2013, Jezkova et al., 2016, Shan et al., 2016, Hamam et al., 2017)

Considering the importance of epigenetic alterations in breast cancer, efforts have also been put forth to classify the disease into distinct epigenetic subtypes based on DNA methylation (Holm et al., 2010, Stefansson et al., 2015, Holm et al., 2016), histone modifications (Karsli-Ceppioglu et al., 2017) as well as miRNAs (Chuang and Jones, 2007) corresponding to the molecular intrinsic subtypes (Eroles et al., 2012). For example, two epigenetic subtypes, the Epi-LumB and Epi-basal, have been described by Stefansson et al. (2015). The Epi-LumB comprises all luminal B and rarely luminal A and HER2-enriched subtype while Epi-basal is made up of most, but not all, basal subtype. Interestingly, no unique DNA methylation patterns were detected in HER2 and Luminal A breast tumours,

indicating biological heterogeneity within these subtypes. When DNA methylation data was validated against The Cancer Genome Atlas (TCGA) cohort, 254 and 202 overlapped CpG's specific for luminal B and basal-like subtypes were identified and only one and none for HER2 and luminal A breast cancers, respectively. Finally, DNA sequence element analysis revealed that enriched CpG methylation in the promoter sequences and hypomethylation in gene body regions were detected predominantly in luminal B and basal-like. Despite the discrepancies in DNA methylation patterns, both epi-subtypes were associated with poor prognosis and survival (Stefansson et al., 2015).

In a separate study using whole-genome DNA methylation analysis, three distinct DNA methylation patterns were described between basal, luminal B and luminal A breast tumours, with the first two being the least and the most frequently methylated subtypes, respectively (Holm et al., 2010, Holm et al., 2016). Interestingly, no distinctive methylation patterns were detected in HER2-enriched and normal-like subtypes. The same study also reported seven epigenetic subtypes (epitypes) with distinct gene expression levels, chromatin states, genetic mutations, DNA copy number aberrations as well as clinicopathological features (Honeth et al., 2008). Therefore, the interesting correlations between the epigenetic and gene expression in breast cancer subtypes could help refining the existing classification system and providing insights into the biological basis underlying breast cancer heterogeneity.

1.3 EPIGENETIC REPROGRAMMING OF BREAST CANCER

1.3.1 Epigenetic therapies (chemical approach)

In contrast to genetic mutations, epigenetic alterations are reversible and thus represent an interesting target for anti-cancer therapies (Hatzimichael and Crook, 2013). The basis of epigenetic treatments is to effectively and permanently reprogram and remodel the epigenetic marks characteristic of cancer cells, reverting the malignant properties of cancer cells (Blancafort et al., 2013). Pfister and Ashworth (2017) described the four major roles of epigenetic therapies in cancer, namely inhibition (by small molecule inhibitors against gain-of-function mutations), synthetic lethality (against loss-of-function mutations), immune modulation (to enhance efficacy of immunotherapies) and reducing drug-resistance (by epigenetic inhibitors). In recent years, numerous FDA approved epigenetic drugs, in particular demethylating agents and histone deacetylase inhibitors (HDACi) have been used singly or in combination with other targeted treatment or systemic therapies to treat cancers (Dawson and Kouzarides, 2012, Pfister and Ashworth, 2017). Despite showing promising clinical and preclinical results against haematological malignancies, epigenetic therapies for solid tumours are relatively less efficient due to several factors, including hypoxia, drug penetration and tumour heterogeneity (Ramachandran et al., 2015, Pfister and Ashworth, 2017). In breast cancer, the DNA methyltransferase inhibitors (DNMTis) 5-azacytidine (5-aza) and 5-aza-2'-deoxycytidine (decitabine) are the two most commonly studied epigenetic drugs (Tsai et al., 2012). However, DNMTis alone have been shown insufficient to eradicate breast tumour

cells partly due to the concurrent re-activation of tumour suppressor genes and oncogenes through non-specific demethylation (Radpour et al., 2011). In addition, tumour cells within hypoxic region of tumour were shown harbouring distinct epigenetic profiles conferring tumour resistance to DNMTis (Ramachandran et al., 2015). These observations prompted the inclusion of other agents, such as HDACi (Connolly et al., 2011) and cytotoxic agents (Mataga et al., 2012) for enhanced specificity and efficacy (Byler et al., 2014). Indeed, recent studies have provided promising evidences on the efficacy of combination treatment of DNMTi with poly (ADP-ribose) polymerase (PARP) inhibitors (PARPi) (Muvarak et al., 2016), tranylcypromine (histone demethylase inhibitor) (Katz et al., 2014), butyrate (HDACi) (Pathania et al., 2016) and Trichostatin A (HDACi) (Al-Rayyan et al., 2014, Liao et al., 2014, Menschikowski et al., 2015). Interestingly, a phase II trial is currently testing the combination epigenetic therapy of 5-aza (DNMTi) and entinostat (HDACi) in patients with advanced hormone-resistant or triple-negative breast cancer (Li et al., 2015b, Huang and Davidson, 2016, Kala and Tollefsbol, 2016, Tomita et al., 2016, Connolly et al., 2017).

1.3.2 Epigenetic reprogramming using biological approaches

In addition to the use of epigenetic drugs mentioned earlier, different biological approaches have been used to reprogram epigenetic states in cancer. The latter often induces a more extensive (e.g. epigenome-wide) rather than mechanisms-specific (such as DNA methylation and histone modifications) changes. Biological approaches, including somatic cell

nuclear transfer (SCNT) (Gurdon, 1962), cell fusion (Harris and Watkins, 1965, Harris et al., 1969), exposure to ESC microenvironment (Postovit et al., 2006, Hendrix et al., 2007, Postovit et al., 2008) and oocyte extracts (Allegrucci et al., 2011) and the iPSC technology (Stricker et al., 2014) have been utilised to induce widespread epigenetic resetting in cancer cells in an attempt to suppress and eliminate tumorigenicity (Table 1.5).

Table 1.5 Different approaches used for epigenetic reprogramming of cancer cells.

Methods	Type of cancers	Major findings	References
SCNT	• Frog renal tumours	• Tumour nuclei became pluripotent and gave rise to swimming tadpoles with no evidence of malignancy	McKinnell et al. (1969)
SCNT	• Mouse medulloblastoma (primary culture, mouse, Ptc1 heterozygous)	• Medulloblastoma nuclei gave rise to embryos and underwent early stages of organogenesis • Tumorigenic properties were suppressed in recipient mice	Li et al. (2003)
SCNT	• Mouse $p53^{-/-}$ lymphoma, moloney murine leukaemia virus induced leukaemia, PML-RAR transgene-induced leukaemia, hypomethylated Chip/c lymphoma, $p53^{-/-}$ breast cancer cell line and $ink4a/Arf^{-/-}$, RAS-inducible melanoma cell line	• Leukaemia, lymphoma, and breast cancer cells failed to acquire pluripotency • A RAS-inducible melanoma nucleus generated blastocyst and ESCs capable of in vivo differentiation • Chimeras generated suffered from cancer of more aggressive form	Hochedlinger et al. (2004)
SCNT	• Embryonal carcinoma (EC)	• EC nuclei gave rise to blastocyst and ESCs with unchanged developmental and tumorigenic potential	Blelloch et al. (2004)
Embryonic environment	• Metastatic melanoma (C8161), breast cancer (MDA-MB-231)	• Presence of Lefty, an inhibitor of Nodal signalling, within the hESCs microenvironment suppressed tumorigenic phenotype	Postovit et al. (2008)
Oocyte extract	• Breast cancer cell lines (MCF-7 & HCC1954)	• Oocyte extracts, in particular from axolotl, induced DNA demethylation, histone mark remodelling, re-activation of tumour suppressor genes and thus malignancy suppression *in vitro* and *in vivo*	Allegrucci et al. (2011)
iPSC reprogramming			
Retrovirus	• Myeloproliferative disorders (MPDs) with JAK2-V617F somatic mutation: polycythemia vera (PV) and primary myelofibrosis (PMF)	• PV iPSCs-derived hematopoietic progenitor ($CD34^+CD45^+$) cells exhibited increased erythropoiesis (a disease hallmark) and expression of specific genes characteristic of the primary $CD34^+$ cells of PV patient	Ye et al. (2009)
Retrovirus	• R545 (aneuploid mouse melanoma cell line)	• Sox2 was dispensable to generate germline-competent iPSCs from melanoma cells • No evidence of malignancy was observed in chimeras	Utikal et al. (2009a)
Retrovirus	• Chronic myeloid leukaemia (CML)	• iPSCs lost addiction to BCR-ABL oncogene (a hallmark of the disease) and became resistant to BCR-ABL inhibitor imatinib.	Carette et al. (2010)

Retrovirus	• HuCC-T1 (colon), DLD-1 (colon), PLC (liver)	• Persistence of *c-MYC* expression was observed in differentiated derivatives, which exhibited increase proliferation and resistance to 5-fluorouracil	Nagai et al. (2010)
Retrovirus & lentivirus	• Gastrointestinal cancer cell lines	• Differentiated cells derived from iPSCs showed high sensitivity to anticancer drug and differentiation-inducing chemicals. • Suppression of malignancy was also observed *in vivo*	Miyoshi et al. (2010a)
Lentivirus	• Lung (A549 cell line)	• Hypoxia-inducible factor (HIF) enhanced reprogramming efficiency with OSLN factors of A549 (lung cancer cell line) • Partially-reprogrammed A549 generated highly aggressive tumours *in vivo* • HIF induces iPSC genes and hESC signature implicated in various cancers	Mathieu et al. (2011)
Episomal vector	• Chronic myeloid leukaemia (CML)	• CML iPSCs retained the complex karyotype with a 4-way translocation between chromosomes 1, 9, 22, and 11 and serves as a valuable tool for mechanistic study and drug screening	Hu et al. (2011)
Retrovirus & Sendai virus	• Chronic myeloid leukaemia (CML)	• Treatment with imatinib, a BCR-ABL oncogene inhibitor, did not affect CML-iPSCs and immature $CD34^{+}38^{-}90^{+}45^{+}$ cells, indicating resistance • Sensitivity recapitulated in more differentiated $CD34^{-}CD45^{+}$ cells, indicating cells of different epigenetic states exhibit distinct dependences on BCR-ABL	Kumano et al. (2012)
Lentivirus	• Non-small cell lung cancer cell lines (H358 and H460)	• Aberrantly methylated promoters and commonly upregulated genes were reversed in iPSCs • DNA demethylation-induced oncogenes and tumour suppressor genes expression changes contributed to reduced malignancy	Mahalingam et al. (2012)
Retrovirus	• Colorectal cancer cell line, HCT116 and its *TP53*$^{-/-}$ mutant	• IPSC in hypoxia have reduced proliferation, invasiveness and tumourigenicity, *TP53*$^{-/-}$ mutant iPSCs exhibited higher tumourigenicity • Hypoxia and loss of *TP53* increase reprogramming efficiency	Hoshino et al. (2012)
Lentivirus	• Sarcomas (SAOS, HOS, MG63, SW872 and SKNEP)	• Cells were reprogrammed to a mesenchymal stem cell differentiation stage, whereby terminal differentiation potential was restored to give rise bone and fat tissues with simultaneous loss of malignancy • Global DNA promoter methylation and gene expression analyses revealed significant epigenetic remodelling of oncogenes and tumour suppressors	Zhang et al. (2013)

PiggyBac transposase	• Glioblastoma (GBM) (brain tumour)	• Global epigenetic reconfiguration suppressed reduced infiltrative behaviour and reactivated tumour suppressors in cancer-iPSCs-derived non-neural cell types but not neural progenitors, suggesting an association between the epigenome of developmental lineage with manifestation of tumorigenicity	Stricker et al. (2013)
Retrovirus & lentivirus	• Non-small cell lung cancer (NSCLC): H3581 and H460	• iPSC generated lost the hallmark mutation: homozygous deletion of *TP53* and *CDKN2A* in H358 and H460 respectively, suggesting the selective reprogramming of cells from the earlier stages of tumorigenesis (which have yet experienced such damaging mutations) in the heterogeneous culture • These aneuploidy but "mutation-free" populations were more susceptible to reprogramming owing to the stability of genome, suggesting cancer cell reprogramming follows an Elite but not Stochastic model.	Lai et al. (2013)
Lentivirus	• Juvenile myelomonocytic leukaemia (JMML)	• Cancer-iPSCs-derived myeloid cells recapitulated phenotypes of primary JMML patient-derived cells, providing a patient-specific drug testing platform	Gandre-Babbe et al. (2013)
Lentivirus	• Gastrointestinal cancer (invasive pancreatic ductal adenocarcinoma)	• iPSCs exhibited differentiation propensity toward endodermal lineage, indicating the presence of epigenetic memory of its somatic origin • The teratoma-derived pancreatic ductal tissues recapitulated pancreatic intraepithelial neoplasia (PanIN), a precursor state of PDAC, allowing the generation of early stage cancer model for the identification of early stage biomarkers and investigation of cancer progression	Kim et al. (2013)
Retrovirus	• MCF7 breast cancer cell line	• Partially reprogrammed cells obtained, which overexpressed SOX2 and SSEA-4 • High proportion of $CD44^{+}$ and $ALDH^{bright}$ cells than parental cells, indicating CSC properties • The suppression of mTOR repressors was found responsible for the acquisition of CSC phenotypes	Corominas-Faja et al. (2013)
Lentivirus	• Hematopoietic cells of 'all-iPS' mice carrying four OSKM factors under the control of doxycycline (Dox) and harbouring human MLL-AF9 fusion gene	• iPSC contributed to chimera formation • Chimeric mice spontaneously reactivated MLL-AF9 and developed leukaemia • Whole-exome sequencing of iPSCs versus leukaemia cells did not reveal more genetic mutations, suggesting that epigenetics is	Liu et al. (2014b)

		responsible for the development of MLL leukaemia in chimeric mice	
TFs (POU3F2, SOX2, SALL2, and OLIG2) overexpression	• Glioblastoma (GBM)	• Reprogram differentiated GBM cells to ''induced'' tumour-initiating cells (TPCs), recapitulating the epigenetic landscape and phenotype of native TPCs	Suva et al. (2014)
Retrovirus	• Osteosarcoma cell lines (Saos-2, MG-63, G-292 and U-2 OS)	• Reprogrammed cell lines exhibited different reprogramming efficiency, differentiation potential and stability under long term culture owing to intrinsic differences	Choong et al. (2014)
Retrovirus	• Non-tumorigenic mammary epithelial cell line, MCF10A	• iPSC reprogramming and subsequent partial differentiation of MCF10A resulted in the generation of cancer stem cell (CSC)-like cell characterized by expression of CSC markers, significant tumour forming ability and sensitivity to CSC-specific drugs	Nishi et al. (2014b)
Sendai virus	• Pancreatic cancer cell line, PANC-1	• CSCs are more susceptible to reprogramming than non-CSCs	Noguchi et al. (2015)
Retrovirus	• Chronic myeloid leukaemia (CML) cell line K562 and KBM7 and primary bone marrow cells from a BCR-ABL+ CML patient	• The BCR-ABL fusion protein can initiate DNA methylation changes that could serve as a secondary event and lead to leukaemia • Both genetic and epigenetic mutations contribute to maintenance of leukaemia potential and thus demethylating agent could potentially be used as adjuvant treatment in leukaemia	Amabile et al. (2015)
Excisable lentivirus	• del(7q)-Myelodysplastic syndromes (MDS)	• Deletion of chr7q renders MDS phenotypes via dosage effect • Haploinsufficiency of *EZH2*, a gene located within a critical ~20 Mb region of the deleted chr7q, is partially responsible for the hematopoietic defects in MDS	Kotini et al. (2015)
Sendai virus	• Li-Fraumeni syndrome (LFS)	• Recapitulation of osteosarcoma (OS) characteristics and tumorigenicity in LFS iPSC-derived osteoblasts (OBs), which serves as a model for early stage of OS development • Various p53 mutants suppress the expression of imprinted gene H19 via an unknown mechanism and contribute to OS. • The activation of H19 and its downstream regulator DCN could prevent and/or provide treatment to the development of OS	Lee et al. (2015)
Retrovirus	• Human fibroblasts	• Generation of wild-type iPSCs and p53-knock down-iPSCs from healthy human fibroblasts.	Sancho-Martinez et al. (2016)

		• Differentiation of wild-type iPSCs and p53KD –iPSCs into neural progenitor cells (NPCs) and generation of different mutant NPC lines Ras/EGFR/SrciNPCs and p53KD-Ras/EGFR/SrciNPCs via lentivirus-mediated overexpression of Ras, EGFR and Src. • Mutant NPC lines contributed to the formation of tumour resembling glioblastomas and facilitated the screening of anti-cancer drugs	
Lentivirus	• HepG2 (hepatocellular carcinoma cell line)	• Xenograft transplanted iPSCs generated CSC-like cells with higher malignancy • Persistence of *OCT3/4* expression that formed positive feedback loop with *c-JUN* contributed to the CSC properties, including chemoresistance and high invasiveness.	Kuo et al. (2016)
Many different approaches tested	• MLL-rearranged human B cell acute leukaemia (ALL)	• Cells harbouring the mixed-lineage leukaemia 1 (*MLL1*) gene rearrangement (MLL-rearranged B-ALL) were refractory to iPSC reprogramming • Various methods and strategies to improve reprogramming efficiency were tested • Global transcriptome and DNA methylome analyses suggested that the intrinsic differentiation blockage of MLL-rearranged B-ALL resulted in its refractoriness to acquire pluripotency.	Munoz-Lopez et al. (2016)
Excisable lentivirus	• Bone marrow or peripheral blood mononuclear cells from patient carrying mutations specific for pre-leukaemic, low risk MDS, high risk MDS, MDS/AML stages.	• Generated a collection of patient-derived iPSCs recapitulating distinct stages of myeloid malignancy to layout a phenotypic roadmap of disease progression leading to leukaemia • Modelled disease progression and reversion by using CRISPR/Cas9-mediated gene mutation/ correction • Identified therapeutic effects of drugs in a stage-specific manner	Kotini et al. (2017)
In vivo reprogramming			
In vivo reprogramming	• Primary human colon cancer cells engrafted in mice	• Overexpression of OSKM in primary colon cancer cells, CC14 engrafted in NUDE mice enhanced metastatic properties of cancer cells accompanied by NANOG • High GLI2/ low AXIN2 was identified as prognostic of poor disease outcome.	Singovski et al. (2016)

1.3.3 Induced Pluripotent Stem Cell Technology

Induced pluripotent stem cell (iPSC) was first reported by (Takahashi and Yamanaka, 2006, Takahashi et al., 2007) as embryonic stem cells (ESCs)-like cells generated via epigenetic reprogramming of differentiated cells through ectopic expression of four transcription factors (*OCT3/4, SOX2, C-MYC* and *KLF4*, or OSKM). The first report successfully derived iPSCs from adult mouse tail-tip fibroblasts through retrovirus-mediated transduction of OSKM (Takahashi and Yamanaka, 2006). A year later, derivation of iPSC was reported from human fibroblast using OSKM (Takahashi et al., 2007) and OS + *NANOG* + *LIN28* (Yu et al., 2007b) by two different groups. IPSCs exhibit strong similarities to ESCs in numerous aspects including morphology, proliferation, transcriptional and epigenetic profiles as well as differentiation potential (Takahashi et al., 2007, Wernig et al., 2007, Mikkelsen et al., 2008b). This Noble prize-winning breakthrough circumvents several limitations inherent to the other reprogramming approaches mentioned in Table 1.5, including the use of defined factors, ease of experimental procedure and free from ethical concern (Ma et al., 2014). Since its discovery, the iPSC technology has been utilised in a wide range of research areas, including embryonic development, cell differentiation, drug discovery and development, disease modelling, and other medical applications (such as regenerative and cell-based therapies) (Yamanaka, 2012a, Ma et al., 2014).

1.3.3.1 Similarities between ESCs and iPSCs

iPSC exhibits high degree of similarity to ESCs and capable of passing all tests commonly used to characterise PSCs including pluripotency markers expression, *in vitro* differentiation, teratoma formation (the most stringent test for pluripotency so far for human PSCs) and chimera formation (gold standard for mouse PSCs) (Okita et al., 2007a, D'Antonio et al., 2017b). However, several studies have identified a series of intrinsic and functional differences between ESCs and iPSCs. Chin et al. (2009) suggested iPSCs as a unique subtype of pluripotent cells after comparing five iPSC lines with three human ESC lines using high-resolution array profiling and uncovered hundreds of differentially expressed genes and miRNA as compared to ESCs. Besides, residual donor cell gene expression was detected through global gene expression study and iPSCs of different origins were shown to exhibit varied "closeness" to ESCs (eg. fibroblast-derived iPSCs displayed highest similarity to ESC followed by adipose-, neonatal fibroblast- and keratinocyte-iPSCs) (Ghosh et al., 2010). With respect to epigenetics, DNA methylation differences between ESCs and iPSCs were reported by Deng et al. (2009) (higher global methylation was observed in iPSCs) and Doi et al. (2009) (71 differentially methylated regions were identified, including 51 hypermethylation and 20 hypomethylation in iPSCs). In addition, Kim et al. (2011) and Ohi et al. (2011) identified DNA methylation discrepancies and varying differentiation potential among iPSC lines of different somatic origins, partly due to somatic epigenetic memory retention, and thus distancing themselves from ESCs.

Despite the discrepancies, based on a reference map generated from DNA methylation and gene expression data, Bock et al. (2011) suggested ESCs and iPSCs should not be considered as two entities but two partially overlapping point clouds in the cellular space, as at least one iPSC line was found indistinguishable from ESC lines. In addition, the presence of cell line specific signatures was detected among iPSCs as well as among ESC lines (an observation also supported by (Osafune et al., 2008), although the former was more prevalent. Furthermore, the authors argued that similarity to ESCs does not fully represents the utility of iPSC lines, warranting a thorough characterisation and suggesting the importance to select the best cell line well-suited for specific application. In agreement, Boulting et al. (2011) demonstrated despite showing efficiency discrepancies among iPSC lines, iPSCs exhibited average differentiation potential as good as ESCs. Lastly, several studies suggested the quality differences between iPSC lines are not due to reprogramming process itself but various technical variabilities, including starting cell types, transcription factor combinations, gene delivery methods and culture conditions (Carey et al., 2011).

1.3.3.2 Different reprogramming approaches

The original iPSC reprogramming approach rely on retrovirus-mediated transduction of OSKM TFs into somatic cells (Takahashi and Yamanaka, 2006). Several caveats of this approach have limited its wider use, which include low efficiency and kinetic, risk of insertional mutagenesis and potential risk of tumorigenesis due to re-expression of TFs (each of which has been individually implicated in different cancers) (Bilousova and Roop,

2014). To circumvent these limitations, a wide range of reprogramming methods have been established since the discovery of iPSCs (Gonzalez et al., 2011). These methods can be broadly categorised into three groups, namely integrative (require the integration of transgene into host genome), non-integrative (require only transient expression of transgenes) and TF-independent (do not require the overexpression of OSKM TFs) (Table 1.6).

Table 1.6 Different methods used for iPSC generation. Table was created based on Bayart and Cohen-Haguenauer (2013).

Methods	Advantages	Disadvantages	Citations
Integrative			
Retroviruses	• Stable integration of transgene into host genome. • Transgenes delivered silenced via naturally occurring mechanism when pluripotency is achieved. • Gammaretroviral Mo-MLV (Moloney murine Leukemia Virus)-based vectors such as pMXs or pMSCV provide a significantly high transgene delivery efficiency (up to 90% in actively dividing cells)	• Lower efficiency in slowly- or non-dividing cells such as neurons • Insertional mutagenesis causing genomic aberrations (aneuploidy, and trisomy 8) and epigenetic abnormalities	(Takahashi and Yamanaka, 2006)
Lentivirus	• Lower risk of insertional mutagenesis • Effective for both dividing and non-dividing cells. • Broader cloning capacity and higher efficiency than retrovirus • Doxycycline inducible lentivectors could prevent transgene "leakage" after reprogramming	• Repression of vector occur to a lesser extent at the end of reprogramming resulted in unwanted expression of the TFs	(Blelloch et al., 2007, Maherali et al., 2007, Yu et al., 2007a)
Polycistronic retro-/lentivirus	• Utilizing a combination of 2A peptide and internal ribosome entry site (IRES) technology • Enhance chances of cells acquire all transgenes • Reduced vector copy number integration per cell expected to reduce risk of insertional mutagenesis	• To a lesser extent but same of described for retroviruses and lentiviruses	(Carey et al., 2009)
loxP-Cre recombinase	• *LoxP* sites flanking integrated transgene cassette allow later excision by targeted and transient expression of Cre recombinase.	• Possesses risk of non-specific excision and can result in integration site sequence modifications after removal from the	(Soldner et al., 2009)

	•	genome (termed genomic scar)	
Piggyback transposon	• High efficiency (up to 2.5%) • Precise excision without genomic scars	• Induces microdeletion of genomic DNA (Santostefano et al., 2014)	(Kaji et al., 2009)
Non-integrative			
Adenovirus	• Non-integrating adenoviruses transiently expressing OSKM • Provide high transgene expression for a limited amount of time;	• Reduced with each cell division results in low efficiency rates • Not known to work on human cells (Bayart & Cohen-Haguenauer, 2013) • Formation of tetraploid cells reported	(Stadtfeld et al., 2008b)
F-deficient Sendai viral vector	• A negative-sense, single stranded RNA virus that replicates in the cytoplasm without DNA intermediates or stable integration into genome • Easy removal of viruses after reprogramming either by exposure to high temperature, treatment with siRNA against L-gene of virus or by microRNA-302L naturally produced by pluripotent cells developed	• Expensive and requires stringent biosafety containment measures.	(Fusaki et al., 2009)
Episomal vector	• oriP/Epstein-Barr nuclear antigen-1-based (oriP/EBNA1) episomal vector capable of extrachromosomal expression of transgene • Loss of episomes (~5% per cell cycle) owing to defects in plasmid synthesis and partitioning • Single transfection of polycistronic episomal vector improves overall efficiency (Chou et al., 2011)	• Multiple transfections required to sustain reprogramming factors expression.	(Yu et al., 2009, Okita et al., 2011)
Minicircles	• Reduced size episomes devoid of bacteria backbone sequence • Longer factors expression by reducing potential silencing through methylation	• Efficiency still low and the production and purification methodology is complex and time-consuming	(Narsinh et al., 2011)
Synthesized RNAs	• Continuous, transient transfection of modified mRNAs, which translated by host machinery • Modified mRNA and the use of interferon inhibitor significantly reduce cytotoxicity • Easy manipulation of factors expression levels by altering mRNA dose.	• Multiple transfections required. But can be improved by using polycistronic self-replicating Venezuelan equine encephalitis (VEE) virus RNA replicon. • Severe cytotoxicity reported but can be reduced by modified mRNA and use of interferon inhibitor	(Yoshioka et al., 2013)
Proteins	• Direct transduction of recombinant proteins or cell extracts containing proteins tagged with cell-penetrant polyarginine sequence	• Slow kinetic, inefficient, low reproducibility and high cost • Multiple transfections required	(Kim et al., 2009a)

	• Easy manipulation of factors expression levels by altering protein dose.		
TFs-independent			
Small molecules	• Generation of mouse iPSCs with A combination of seven small molecules forskolin (FSK), valproic acid sodium salt (VPA), CHIR99021, 616452, Tranylcypromine (Tranyl), 3-deazaneplanocin A (DZnep) and TTNPB • Highly controllable • Easy synthesis, preservation and standardization	• Low efficiency (improved by later study) • Mechanisms are largely unknown	(Hou et al., 2013)
MicroRNA (miRNA)	• Direct transfection of mature microRNA from mir-200c, mir-302s and mir369s family	• Mechanisms are not well understood	(Lin et al., 2011, Miyoshi et al., 2011)

1.3.3.3 Mechanisms of iPSC reprogramming – as revealed by genetic studies

Acquisition of a specific cell identity is an epigenetic process, whereby cells of different lineages obtain specific epigenetic configurations during embryonic development enabling cell-specific gene expression/ suppression (Liang and Zhang, 2013a, Theunissen and Jaenisch, 2014a). The conversion of fibroblasts to myoblasts through *MyoD* (a mammalian transcription factor) overexpression is one of the classical examples which demonstrates cell fate is determined and can be induced by master regulator/s (Davis et al., 1987). Likewise, iPSC generation is defined as an epigenetic process involving a series of epigenetic events, leading to the abolishment of somatic epigenetic marks and formation of new epigenetic configuration and transcription profile characteristic of PSCs via ectopic expression of pluripotency factors OSKM (Brambrink et al., 2008, Stadtfeld et al., 2008a, Theunissen and Jaenisch, 2014b). In agreement, accumulating evidence indicates reprogramming efficiency is significantly affected by various epigenetic modulators (Liang and Zhang, 2013a)

Liang and Zhang (2013a) described iPSC generation as two sequential events (mesenchymal-epithelial transition or MET and activation of pluripotency circuitry) followed by a maturation/correction stage involving progression of nascent to *bona fide* iPSCs. Each event involves a series of epigenetic barriers. Somatic cells procure epithelial properties through MET at the early stage of reprogramming (Li et al., 2010). This process involves upregulation and downregulation of epithelial genes (including *Cdh1, Epcam*) and mesenchymal genes (*Snai1/2* and *Zeb1/2*), respectively (Sridharan et al., 2009, Samavarchi-Tehrani et al., 2010) Factors promoting MET (TGF-β inhibitors, BMPs, miR-200 and miR-302/367 and *Cdh1*) have been shown to significantly enhance reprogramming efficiency. The acquisition of epithelial features has been widely used to predict the iPSC-destined cells (Smith et al., 2010).

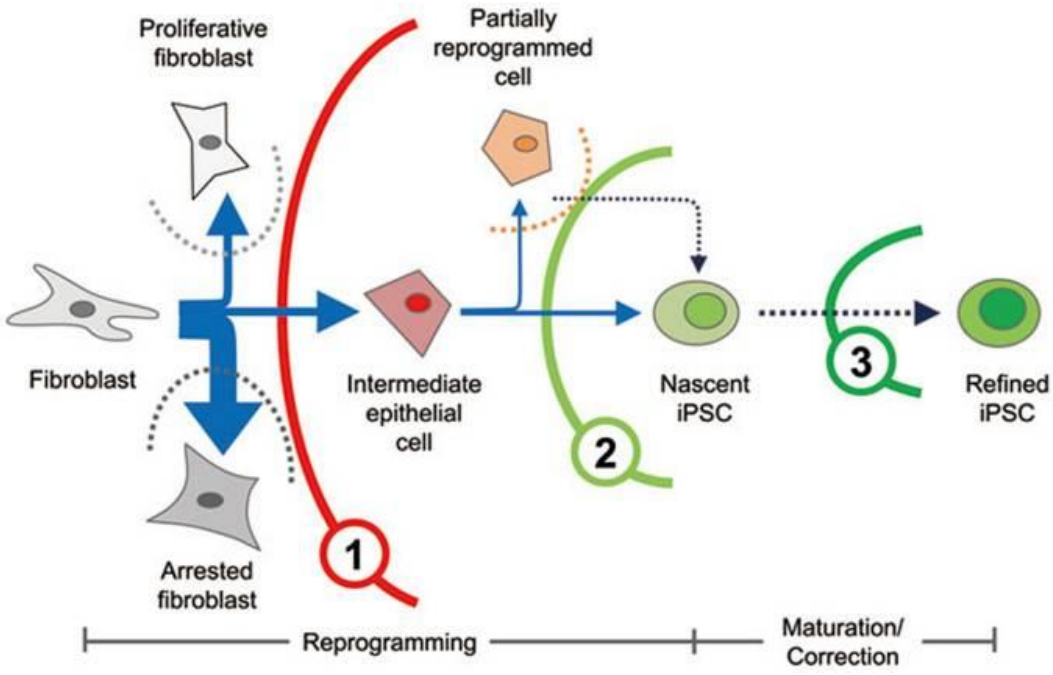


Figure 1.8 The stages of iPSC generation process as proposed by Liang & Zhang (2013).

(1) Early stage involves mesenchymal to epithelial transition (MET). (2) Activation of pluripotency circuitry allows reprogrammed cell to express pluripotency genes independent of exogenous factors. (3) Several barriers, including elimination of epigenetic memory of somatic origin, need to be overcome before generation of *bona fide* iPSCs. Figure derived from Liang and Zhang (2013a) (licence number: 4292631332964).

The molecular mechanisms leading to MET remain elusive. Samavarchi-Tehrani et al. (2010) suggested OSKM induces BMP signalling-dependent induction of miR-205 and miR-200 family (key regulators of MET), which in turn stimulate MET (Figure 1.9). In support to this finding, Hu et al. (2014) demonstrated Tet and TDG induce DNA demethylation and thus expression of the miR-200 clusters during MET. On the other hand, actin cytoskeleton reorganization is prerequisite for MET and reprogrammed cells which successfully complete MET have reduced cell size and tight cell-cell interaction (Smith et al., 2010). By knocking down

two kinases linked to integrin signalling (tesk1 and limk2), Sakurai et al. (2014) reported destabilization of actin cytoskeleton (through decreased cofilin phosphorylation), which potentially induce alteration of biophysical environment and thus histone acetylation and methylation (Downing et al., 2013), eventually leading to enhanced reprogramming efficiency (Sakurai et al., 2014). In the same study, ILK and TGF-β pathway crosstalk was found responsible for cytoskeletal remodelling. In a thorough review on EMT, Lamouille et al. (2014) acknowledged the cooperative effective of multiple signalling pathways (regulated by diverse extracellular signals) and highlighted TGF-β family proteins (such as SMADs) as major players in EMT regulation. In agreement, small-molecule inhibitor of TGF-β signalling can replace *Sox2* in iPSC reprogramming (Ichida et al., 2009).

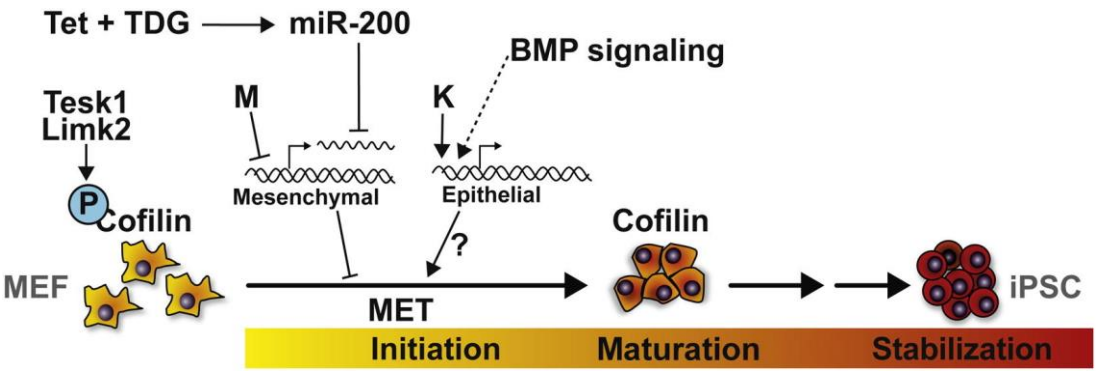


Figure 1.9 Cellular and molecular mechanisms underlying mesenchymal-to-epithelial Transition (MET) during early stage of reprogramming.

Various regulators have been identified in different studies responsible for MET regulation. MET represents initial phase of reprogramming followed by iPSC maturation and stabilisation. Figure derived from David and Hirsch (2014) (licence number: 4292631509455)

Cascades of escape from cell cycle arrest, increased proliferation rate, resistance to apoptosis and senescence have been detected simultaneously with epithelial properties acquisition. Repression of pathways (p53/p21 pathway) and factors promoting cell proliferation (*Ink4a/Arf* locus) has been shown to promote iPSC generation (Hong et al., 2009b, Li et al., 2009). In addition, accelerated proliferation is also accompanied by transition from oxidative phosphorylation to glycolysis-dependent energy metabolism which is characteristic of pluripotent stem cells (Panopoulos et al., 2012a).

Next, the second event described by Liang and Zhang (2013a) involves pluripotency circuitry activation. Recent studies demonstrated the up-regulation of various pluripotency genes after epithelial gene activation, suggesting these genes may act as pioneer factors in initiating the pluripotency circuitry (Ichida et al., 2009, Onder et al., 2012). Through single-cell study of 48 genes in approximately 7000 cells, Buganim et al. (2012) reported a specific set of pluripotency genes including *Esrrb, Utf1, Lin28* and *Dppa2* were expressed prior to pluripotency circuitry activation in iPSC-destined cells. Results suggested the presence of an activation hierarchy of pluripotency factors. Based on this observation, Buganim et al. (2012) proposed an activation hierarchy of gene activation comprises an early stochastic probabilistic phase and a late deterministic hierarchical phase (Figure 1.10).

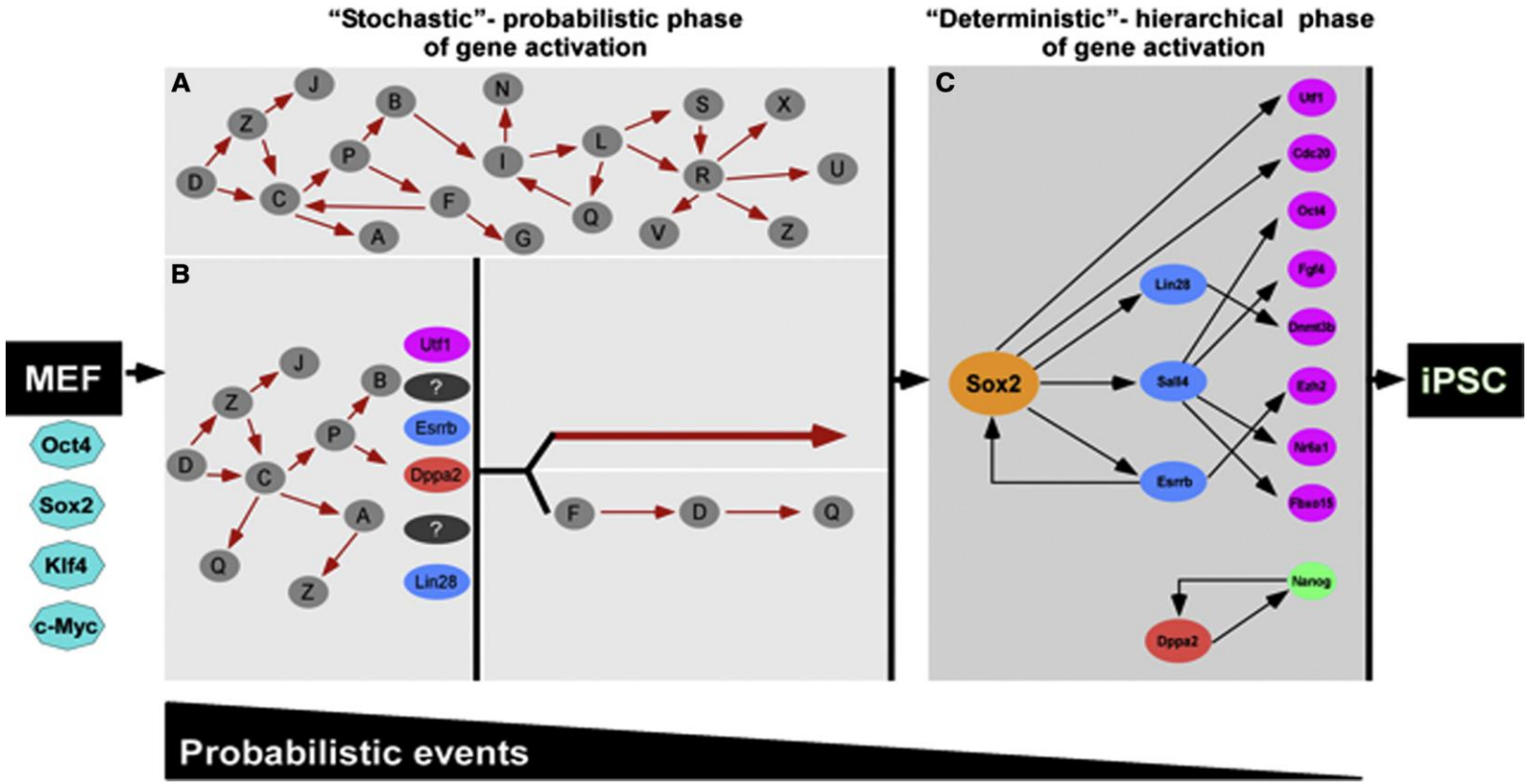


Figure 1.10 A model of pluripotency network activation.

The activation process involves two stages, a stochastic (A & B) and deterministic phase (C). The former is subdivided into two paths, one complete probabilistic (A) and another initial probabilistic late sequential (B). Both lead to the activation of *Sox2*, whose expression marks the beginning of the latter. Figure derived from Buganim et al. (2012) (Licence number: 4292640171745).

The early stochastic phase can be explained by two sub-models, a complete probabilistic model or an initial probabilistic late sequential gene activation model, both lead to the activation of the Sox2 locus. During the stochastic phase, great gene expression variation between cells is observed. *Esrrb, Utf1, Lin28* and *Dppa2* are among the genes capable of predicting reprogramming success. Besides, the activation hierarchy also suggests generation of iPSC can be achieved through different combinations of reprogramming factors (such as a combination of *Lin28, Sall4, Esrrb* and *Dppa2*), implicating pluripotency circuitry can be turned on at multiple entry points (Buganim et al., 2012). Indeed, various combinations of reprogramming factors have been proposed and demonstrated efficacy on driving iPSC generation (Yu et al., 2007b, Feng et al., 2009, Heng et al., 2010, Ho et al., 2011, Festuccia et al., 2012). The probabilistic phase is followed by a deterministic phase that begins with the activation of endogenous *Sox2*, which subsequently induces activation of downstream factors, including *Sall4, Esrrb* and *Lin28*, and eventually the whole pluripotency circuitry. Results suggested *Sox2* is important to generate permissive chromatin state, which in conjunction with exogenous factors, allows the activation of the pluripotency network. In agreement, the expression of *Nanog* (an important marker for pluripotency) has been shown to occur after radical chromatin reorganization (Mattout et al., 2011). However, the molecular determinants that drive the epigenetic changes during the two phases remain largely elusive. These epigenetic barriers might be the non-permissive global and/or the inherent chromatin configuration at the pioneer pluripotency genes (Buganim et al., 2012). After MET and pluripotency circuitry activation, nascent iPSCs progress to

achieve a *bona fide* iPSC state. At this stage, despite their capability for exogenous factor-independent self-renewal, the reprogrammed cells carry remnant of somatic memory in the form of gene expression and/or epigenetic configuration (Polo et al., 2010, Kim et al., 2011). The removal of somatic memory is prerequisite to acquire ESCs-like pluripotency and it can be achieved/promoted through cell passaging and chemical-mediated DNA methylation and histone deacetylation inhibitions (Kim et al., 2011, Liang and Zhang, 2013b).

In a comprehensive review on reprogramming molecular mechanisms, different kinetic models have been proposed in addition to the activation hierarchy (an early stochastic and late deterministic model) described by Buganim et al. (2012) (Theunissen and Jaenisch, 2014a) (Figure 1.11). Hanna et al. (2009) suggested reprogramming is a stochastic event whereby reprogramming cells overcome one or more rate-limiting steps before acquiring pluripotency, contributing to the presence of different latencies. However, reprogramming can be accelerated by cell division rate-dependent (such as inhibition of p53-p21 pathway) or -independent modifications (overexpression of *Nanog*). On the other hand, a deterministic model defines reprogramming as a process that involves a series of specific events and progress with fixed latency (Di Stefano et al., 2014, Guo et al., 2014, Rais et al., 2015). Lastly, the biphasic model advocates reprogramming consists of two waves of gene activity sandwiching a predominantly stochastic phase (Polo et al., 2012b).

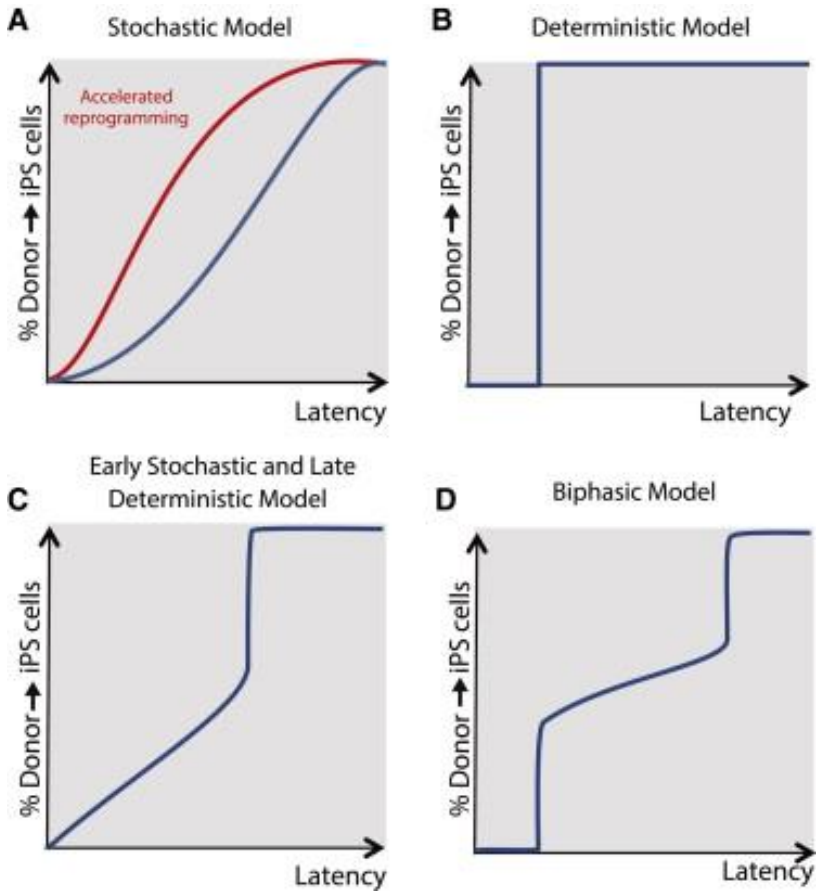


Figure 1.11 Different kinetic models of iPSC generation.
Kinetic models described by (a) Hanna et al. (2009), (B) Di Stefano et al. (2014), (C) Buganim et al. (2012) and (D) Polo et al. (2012). Figure derived from Theunissen and Jaenisch (2014a) (Licence number: 4292640348079).

1.3.3.4 Mechanisms of iPSC reprogramming – as revealed by epigenetic studies

While various genetic mechanisms have been elucidated, the epigenetic mechanisms underlying iPSC generation remains largely elusive (Polo et al., 2012b, Theunissen and Jaenisch, 2014a). According to Polo et al. (2012b), early stages of reprogramming are associated with histone modification changes, while DNA methylation alterations occur at later stage of reprogramming indicated by up-regulation of DNA methylation and demethylation genes (*Dnmt3a* and *Dnmt3L*, *Apobec* and *Tet1*). In addition, global redistribution of H3K4me2, characteristic of eurochromatin, at

numerous gene loci is also observed at the early stage (Koche et al., 2011). Repression of fibroblast genes is associated with the presence of H3K27me3 (a silencing mark) marks. During reprogramming, localized depletion of H3K27me3 has been reported at promoters that acquired pre-existing H3K4me3 (an activation marker). In addition, the loss of H3K27me3 and gain of H3K4me3 have also been associated with pluripotency genes activation (Polo et al., 2012b).

Recent efforts focusing on transcription factors (TFs) binding and chromatin remodelling have provided important insights into epigenetic events underlying iPSC reprogramming (Soufi et al., 2015, Chen et al., 2016, Chronis et al., 2017, Knaupp et al., 2017, Li et al., 2017). Soufi et al. (2012) reported OSK act as pioneer factors for c-Myc and bind to distal elements (e.g. enhancer) with closed chromatin features, independent of pre-existing histone modifications. Initial binding of OSK is facilitated, although not necessary, by c-Myc and allows c-Myc to gain access to its target sites subsequently. However, failure of initial OSKM binding to megabase-scale chromatin domains harbouring H3K9me3 marks which encompass many pluripotency genes was observed, warranting further study to investigate loss of H3K9me3 marks during reprogramming. In the same study, downregulation of enzymes catalysing H3K9me3 deposition was shown to enhance reprogramming efficiency. A subsequent study by the same group suggested by using different classes of DNA binding domains, Oct4 (POU domain), Sox2 (HMG box) and Klf4 (zinc finger) recognize and bind partial motifs exposed on nucleosomes. This allows OSK to gain access to their target sites in closed/silent chromatin. Based on the above findings, both

studies proposed a pioneer factor model stating OSK are the first/pioneer factors to initiate cell-fate switch from somatic to pluripotent stem cells (Soufi et al., 2012, Soufi et al., 2015).

Knaupp et al. (2017) described reprogramming as three phases, an early global chromatin remodelling, a middle stage with minimal remodelling activity and a second wave of chromatin reconfiguration prior to pluripotency acquisition. The early stage involves binding of Oct4/Sox2 (acting as pioneer factors) predominantly to methylated/inaccessible regions. In addition, Oct4/Sox2 is recruited by Klf4 to open and unmethylate chromatin (~30% of sites bound by Oct4/Sox2) in MEFs. Failure to replace somatic TFs by Oct4/Sox2 in constitutively open chromatin results in partial maintenance of MEF chromatin pattern as observed in refractory cells. The second widespread reconfiguration increases enhancer chromatin accessibility induced by OSK triggering the pluripotency program. In addition to loss of MEF identity (as indicated by early accessible chromatin regions in MEFs which become progressively inaccessible during reprogramming) and gain of iPSC characteristics (early inaccessible regions which acquire accessibility later), (Knaupp et al., 2017) described the importance of a group of transiently accessible regions (TARs) which become accessible immediately after OSKM induction and revert back to somatic state (inaccessible) after iPSC state is retained. These TARs are opened by somatic TFs binding recruited by Oct4/Sox2, through a process called somatic TF redistribution (Knaupp et al., 2017).

On the other hand, using assay for transposase-accessible chromatin sequencing (ATAC-seq) profiling, (Li et al., 2017) described two major events occur during iPSC reprogramming, namely dynamic chromatin state shifts from open to closed (OC) and closed to open (CO), occurring at early and later stage of reprogramming respectively. The former consists predominantly of genes associated with somatic state and the latter pluripotency. Enrichment of OSK motifs in CO loci and not OC was reported suggesting chromatin closing occurs through alternative mechanisms instead of OSK binding. Indeed, Sap30 (a Sin3A corepressor complex component) was found responsible for downregulation of H3K27ac deposition at OC loci leading to OC shift (Li et al., 2017). In addition, overexpression of factors governing somatic cell fate, including c-JUN, ETS, TEAD, RUNX and MAD families, were found negatively associated with iPSC generation primarily by affecting OC shift. On the other hand, Chronis et al. (2017) proposed that at early stage of reprogramming, OSK bind primarily to open somatic enhancers leading to immediate silencing via three different mechanisms, namely induction of somatic TFs redistribution (from somatic enhancers to other sites bound by OSK), recruitment of Hdac1 and repressing Fra1 (together with Jun protein make up the transcription factor AP-1 responsible for cell differentiation). Activation of pluripotency enhancers begins early and culminates at later stage mediated by OSK and other pluripotency TFs. Collectively, interactions among OSK and between stage-specific TFs facilitate somatic enhancer inactivation and pluripotency enhancer activation during reprogramming. Both of these studies highlighted the involvement of other factors during early stage of reprogramming, Sap30 (Li et al., 2017) and stage-specific TFs (Chronis et

al., 2017), thus rejecting the pioneer factor model proposed and supported by Soufi et al. (2015) and Knaupp et al. (2017), respectively.

1.4 IPSCS AND CANCER

As mentioned in Section 1.3.3.3 & 1.3.3.4, epigenetic reprogramming results in the establishment of pluripotency via global epigenetic reconfiguration, an incidence also observed in cancer (Ramos-Mejia et al., 2012). Cancer cell reprogramming represents a valuable approach to investigate the impacts of epigenetic alterations on malignant transformation (Semi et al., 2013). As illustrated in Table 1.5, generally, not all cancer cells are amendable to reprogramming as some are found refractory to the reprogramming process. Successful reprogramming often results in the restoration of ESC-like proliferation and differentiation potential (chimera and/or teratoma formation) and most notably, reduced or complete loss of malignant properties (Dewi et al., 2012, Kim and Zaret, 2015).

Comparisons have long been made and astounding similarities reported between the iPSC reprogramming and the process of malignant transformation. Among these similarities, the global reconfiguration of the epigenome (and thus transcriptional profile), acquisition of self-renewal and unlimited proliferation (Ben-Porath et al., 2008, Ohnishi et al., 2014) and metabolic switch (Folmes et al., 2011, Panopoulos et al., 2012b, Kida et al., 2015, Mathieu and Ruohola-Baker, 2017, Nishimura et al., 2017) have been highlighted. Indeed, several studies have identified numerous common pathways in oncogenesis and somatic cell reprogramming, suggesting mechanisms involved in both processes might be overlapping (Ben-Porath et al., 2008). In addition to the implications of OSKM factors in cancer

(discussed in detail in Section 7.3), the inhibition of tumour suppressor pathways, including p53, $p16^{INK4a}$-RB, $p19^{ARF}$ or $p21^{CIP1}$ (Krizhanovsky and Lowe, 2009, Banito and Gil, 2010, Kareta et al., 2015), a phenomena widely observed in human cancer, has been shown to significantly enhance reprogramming efficiency (Hong et al., 2009a, Kawamura et al., 2009). On the other hand, the most stringent test to evaluate the quality of iPSC and ESCs rely on the ability of these cells to induce the formation of teratoma (a benign germ cell tumour) in immunocompromised mice. According to the cancer stem cells hypothesis (discussed in Section 1.2.1.3), the formation and heterogeneity of tumours are resulted from a small population CSCs which harbour self-renewal potential and are multipotent (Reya et al., 2001, Pardal et al., 2003). Several studies suggested the formation of CSCs is a result of alterations arising in terminally differentiated cells via dedifferentiation mimicking the iPSC reprogramming (Reya et al., 2001). In addition, the iPSC technology have been used to generate cancer and CSC models by immature termination of reprogramming factors (Ohnishi et al., 2014), co-culture of iPSC with tumour-derived extracellular vesicles secreted from Lewis lung carcinoma cells (Yan et al., 2014) or partially differentiation of iPSCs (Nishi et al., 2014b).

1.4.1 Cancer Cell Reprogramming using iPSC technology

Studied have reported the successful derivation of iPSCs from a wide range of cancers, including blood, gastrointestinal, lung, liver, breast and brain, Cancer iPSCs have allowed the investigation of questions pertaining cancer research, for example the cell-of-origin of cancer, mechanisms of cancer initiation and progression as well as the discovery and development of

effective anticancer therapies (Kim et al., 2013, Zhang et al., 2013, Iglesias et al., 2017). These studies have demonstrated the capability of reprogramming to restore pluripotency and differentiation potential in cancer-derived iPSCs, and in most cases, accompanied by the suppression or elimination of malignancies (Stricker and Pollard, 2014). Zhang et al. (2013) reported the successful derivation of iPSCs from sarcoma cell lines. Sarcoma-derived iPSCs restored differentiation potential and capable of giving rise to terminally differentiated cells such as connective tissues and red blood cells. Irreversible loss of uncontrolled proliferation and tumorigenicity through epigenetic reconfiguration predominantly at oncogenes and tumour suppressors were observed as revealed by global DNA promoter methylation and gene expression analyses (Kim et al., 2013, Zhang et al., 2013).

Using the same approach, Kim et al. (2013) successfully generated iPSCs from invasive pancreatic ductal adenocarcinoma (PDAC). PDAC-derived iPSCs exhibited skewed differentiation propensity toward endodermal lineage, giving rise to a large proportion of DBA lectin-positive ductal structures in teratoma, indicating somatic epigenetic memory retention (Bar-Nur et al., 2011). The teratoma-derived pancreatic ductal tissues recapitulated pancreatic intraepithelial neoplasia (PanIN), a precursor state of PDAC. Through analysis of proteins secreted by PanIN-like cells, the HNF4α network was found responsible for cancer progression and might serve as an important marker for early stages of the disease. In this regard, the iPSC technology allowed the generation of early stage

cancer model for the identification of biomarkers for early detection and investigation of pancreatic cancer progression (Bar-Nur et al., 2011, Kim et al., 2013).

The concept of cell-type specific carcinogenesis proposes cancer-related mutations initiate malignant transformation in a tissue- or cell- type specific manner. In other words, despite their presence in all cell types, the effects of such mutations are significantly influenced by cell-type specific contexts (such as developmental or differentiation stages) and environments (such as the activity of cell-cell signalling) (Sieber et al., 2005). Reprogramming of cancer cells to iPSCs followed by subsequent differentiation into different cell types can provide insight into the mechanisms and molecular basis governing cell-type specific carcinogenesis. By reprogramming glioblastoma (GBM)-derived neural stem (GNS) cells and subsequent re-differentiation into different cell lineages, Stricker et al. (2013) reported a widespread cancer-specific DNA hypermethylation, most notably at the promoter regions of testis-derived transcript (*TES*, a cell motility regulator), cyclin-dependent kinase inhibitor 1C (*CDKN1C*, a tumour suppressor) and numerous polycomb-repressive complex 2 (PRC2) target genes. Interestingly, widespread epigenetic changes led to the suppression of infiltrative behaviour and elimination of malignant properties in all lineages via reactivation of tumour suppressors, except for the neural lineage. Results indicated that in the context of GBM, the manifestation of malignant properties was lineage-dependent and these properties can

potentially be suppressed by the epigenome of a different developmental lineage (Stricker et al., 2013).

1.5 CRISPR-Cas9 gene editing

The clustered, regularly interspaced, short palindromic repeats (CRISPR)-Cas9 technology is a useful gene editing tool of mammalian cells (Cong et al., 2013). CRISPR are adaptive immune systems used by many bacteria to provide resistance against the invasion of foreign nucleic acids (Barrangou et al., 2007). The CRISPR-Cas9 gene editing technology mimics the Type II CRISPR systems in bacteria, whereby foreign DNA sequences are incorporated between CRISPR repeat sequences within the host genome. This CRISPR repeat arrays are transcribed and processed into multiple CRISPR RNAs (crRNAs), containing short sequence of the foreign material (the protospacer sequence) and part of the CRISPR repeat. Each crRNA hybridizes with the transactivating CRISPR RNA (tracrRNA) before forming complex with the Cas9 enzyme and leading to the recognition and cleavage of foreign DNA sequence complementary to the protospacer sequence and located immediately upstream of a protospacer adjacent motif (PAM) (5'-NGG for *Streptococcus pyogenes* derived Cas9). Notably, PAM sequence is absent in the CRISPR locus in the host genome therefore self-targeting by crRNA-guided Cas9 is not possible (Sander and Joung, 2014, Wang et al., 2016). This defence mechanism of bacteria was adapted to create a gene editing system by including two components, the Cas9 nuclease and a single guide RNA (sgRNA) containing the crRNA and tracrRNA sequences. The first 20 nucleotides of the crRNA serves as the protospacer sequences in bacteria system and manipulated to target different regions in the genome adjacent to the PAM sites (Wang et al., 2016). PAM sequence recognition is bacteria-dependent and different PAM sequences have been reported such as 3'NGAN or NGNG in *Staphylococcus aureus* (Kleinstiver et al., 2015).

To facilitate gene editing, the sgRNA-guided Cas9 nuclease creates double stranded breaks (DSBs) at specific genomic loci triggering the activation of DNA repair pathways nonhomologous end joining (NHEJ) or homology-directed repair (HDR). The NHEJ is efficient but error-prone while HDR has high fidelity but less efficient. DSBs repaired via NHEJ often results in random mutations (predominantly insertion and deletions, indels) leading to gene knockout. On the other hand, HDR is often utilised to introduce targeted mutations or gene correction by providing a donor DNA template, containing the desired sequence changes, in addition to the Cas9 and sgRNA to the system (Wang et al., 2016) (Figure 1.12). In recent years, the application of CRISPR-Cas9 system has been expanded from gene knockout and targeted editing to gene repression and activation as well as chromosomal loci labelling (Sander and Joung, 2014).

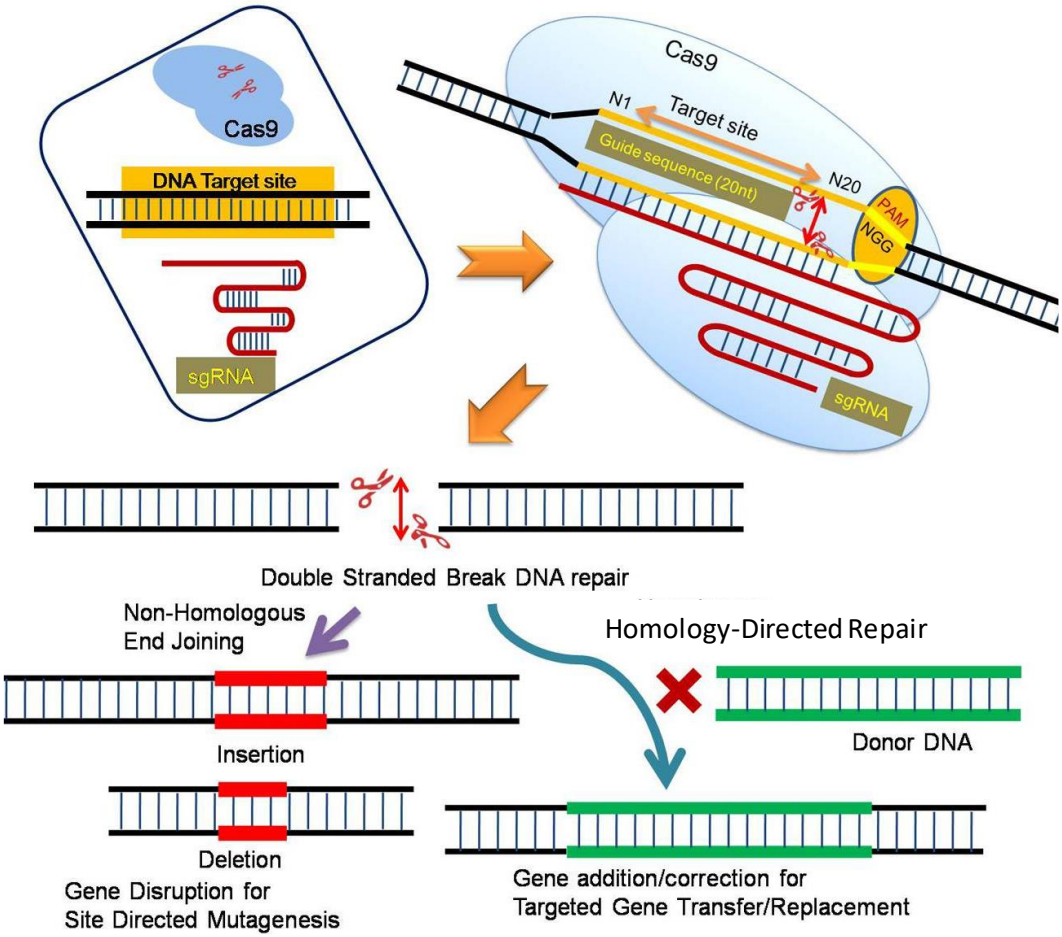


Figure 1.12 The CRISPR-Cas9 gene editing.

(top left) The system consists of Cas9 endonuclease and single guide RNA (sgRNA), which contains a 20 nucleotide (nt) sequence crucial for target site recognition. (top right) Target site should locate immediately 5' upstream of a protospacer adjacent motif (PAM) sequence (NGG is the PAM site for *Streptococcus pyogenes* derived Cas9). (bottom) Cas9 cleavage creates double stranded break. DNA repair occurs through non-homologous end joining (NHEJ) or homology-directed repair (HDR). The former is error prone and results in random indels. The latter is specific and donor DNA harbouring the desired mutation is required to induce HDR-mediated mutagenesis. N1-N20 = the 20 nt upstream of the PAM site in the genome. Figure adapted from Khatodia et al. (2016).

1.6 AIM AND OBJECTIVES

The aim of this project was to identify the contribution of genetic and epigenetic events to breast carcinogenesis and the potential effect of their epigenetic alteration reversal on suppression of malignancy using the iPSC technology. Therefore, the objectives of this study were:

1. To reset epigenetic configuration of breast cancer cells with iPSC technology and generate pluripotent breast cancer-derived iPSCs as cancer models to study the molecular events underlying breast carcinogenesis (Chapter 3, 4, 5).
2. To re-differentiate breast cancer iPSCs into cells of different lineages to study potential elimination of tumorigenicity in differentiated progenies (tumour reversion) (Chapter 6).
3. To evaluate outcomes, identify limitations and discuss the potential significance and future directions of the present study (Chapter 7).

2. METHODOLOGY

2.1 CELL CULTURE

2.1.1 Culture of continuous cell lines

Breast cancer cell lines MCF7 (ATCC, HTB-22), BT-549 (ATCC, HTB-122) and MDA-MB-231 (ATCC, HTB-26) were cultured in Roswell Park Memorial Institute (RPMI) medium 1640 (Invitrogen, 31870-025) supplemented with 10% foetal calf serum (FCS) (Sigma-Aldrich, F7524), 1% penicillin/streptomycin (pen/strep) (Invitrogen, 15070063), 1% L-glutamine (Invitrogen, 21051040), 1% sodium pyruvate (Invitrogen, 11360039) and 1% non-essential amino acids (NEAA) (Invitrogen, 11140035). The immortalised human mammary epithelial cell line MCF10A (ATCC, CRL-10317) was cultured in either human mammary epithelial cell (HuMEC) ready medium (Gibco, 12752-010) (composed of HuMEC basal serum free medium, HuMEC supplement, bovine pituitary extract) supplemented with 1% pen/strep and 100 ng/ml cholera toxin (Sigma-Aldrich, C8052); or DMEM-F12 (Invitrogen, 11330-032) supplemented with 5% horse serum (Invitrogen 16050-122), 1% pen/strep, 20 ng/mL epidermal growth factor (EGF) (PeproTech, AF-100-15), 0.5 mg/mL hydrocortisone, 100 ng/mL cholera toxin and 10 µg/mL insulin. At 70 – 80% confluence, cells were passaged by first washing with sterile phosphate-buffered saline (PBS) and treatment with the dissociating enzyme 0.25%Trypsin-EDTA-phenol red (Gibco, 25200072) for 5-10 mins at 37°C. Culture vessels were gently tapped to facilitate dissociation. The enzyme was deactivated by adding an equal volume of complete RPMI medium or trypsin inhibitor (for MCF10A) (Sigma-Aldrich, 10109886001). Cells were

collected in a 15 mL Falcon tube and centrifuged at 400 xg for 4 min at room temperature. Lastly, cells were resuspended in media and plated at an appropriate density. All cell lines were grown at 37°C and 5% CO_2.

2.1.2 Cell Counting

Cell counting was performed on haemocytometer. Upon dissociation with enzyme and enzyme neutralisation with medium/ enzyme inhibitor, 10 µL of cell suspension was loaded onto haemocytometer. Cells located within the 4 sets of 16 corner square were counted (cell count = x). To estimate the cell density, cell count was divided by four (x ÷ 4 = y) and multiplied by 10^4. Cell density (y) was expressed as number of cells per 1 mL of medium (y = number of cell/mL). Total number of cell was determined by multiplying y with total volume of mixture where 10 µL of cell suspension was derived from. Desired seeding densities for various experiments were achieved by adjusting the volume of cell suspension plated.

2.1.3 Cryopreservation of Cells

Cell were dissociated and spun as described above, cell pellet was resuspended with 1:1 ratio of respective media and 20% sterile dimethyl sulphoxide (DMSO) (Sigma-Aldrich, D2650) in FCS and transferred into sterile cryovials. Cryovials were placed into Mr Frosty freezing container and kept at -80°C overnight before long-term storage in liquid nitrogen (-179°C).

2.1.4 Thawing of Cells

Frozen cells stored in cyrovials from liquid nitrogen were thawed at 37°C in a water bath until a small bit of ice left in the vial. Pre-warm culture medium was added to cryovial dropwise to prevent osmotic shock. Mixture was spun

at 400×g for 4 min and the cell pellet was resuspended in culture medium and plated onto culture vessels.

2.1.5 Stem cells culture

All stem cells, including embryonic stem cells (ESCs) H9 (WiCell, WA09) and induced pluripotent stem cells (iPSCs) were cultured in Essential 8 (E8) medium (Gibco, A1517001) on vitronectin (Gibco, A14700)-coated plate (Chen et al., 2011). Medium was changed every day and cells were passaged at 80 – 90% confluence. To passage, medium was removed, cells were washed with sterile PBS and treated with UltraPure 0.5 mM EDTA pH8.0 solution (Invitrogen, 15575020) for 5 min at 37°C (Beers et al., 2012). EDTA solution was aspirated, cells were dissociated by gently flushing culture vessel surface with E8 medium. Cells were collected in a 15 mL Falcon tube and centrifuge at 400 xg for 4 min. The cell pellet was resuspended with E8 medium and plated at appropriate ratio on culture vessel coated with vitronectin. Rho-associated protein kinase (ROCK) inhibitor (ROCKi) Y-27632 (Selleckchem, S1049) was added to newly plated cells to reduce cell death. The next day, medium was changed and ROCKi treatment stopped.

Vitronectin coating was performed by diluting the vitronectin stock with PBS to 5 µg/mL. An appropriate volume of the diluted vitronectin was distributed (eg. 1 mL for one well of 6-well plate) onto culture vessel and incubate at room temperature for 1 hour (h). Coated plates can be stored at 4°C for as long as two weeks by topping up with excess PBS to prevent drying. Coated plates stored at 4°C were first incubated at 37°C for 1 h

before use. The vitronectin solution was the discarded and fresh medium was added.

2.2 IPSC REPROGRAMMING

2.2.1 X-tremeGEN 9 DNA Transfection Reagent (XG9) (Roche, 06365779001)

Cells were plated 24 h before transfection and medium was changed an hour before transfection. XG9 reagent was left to equilibrate to room temperature for 15 min and briefly vortexed before use. For transfection experiment on one well of a 6-well plate containing 2 mL medium, the reaction mix was prepared as below.

Reaction mix components	Quantity
XG9 reagent	3 µL
Episomal vectors (made up of equal amount of each episomal plasmid)	1 µg
OptiMEM reduced-serum medium (Gibco, 31985062)	100 µL

The mixture was incubated at room temperature for 15 min and then added into medium in a dropwise manner. The plate was swirled and shaken to ensure even distribution.

For iPSC generation, cells were transfected with four episomal vectors encoding *OCT3/4* and shRNA against *TP53* (Addgene #27077), *SOX2/KLF4* (Addgene #27078), *L-MYC/LIN28* (Addgene #27080) and *EBNA1* (Addgene #37624) (Okita et al., 2011) or five vectors with an additional vector encoding the microRNA-302/367 (miR-302/367) cluster (Addgene #31310) (Figure S2-1 – 5). Cells were plated the day before transfection on a 6-well plate pre-coated with vitronectin. On the day of the first transfection (Day

0), culture medium was changed to E8 medium (supplemented with 1 μM hydrocortisone) before switching to Essential 6 medium (Gibco, A1516401) (supplemented with 100 ng/mL human basic fibroblast growth factor (bFGF) (PeproTech, 100-18C) on Day 7. Four consecutive transfections (each time with all episomal vectors) were performed every other day with the last ended on day 6. From Day 20-30, potential iPSC colonies were manually passaged and transferred into a new culture vessel with E8 medium supplemented with 10 μM RCOKi. ROCKi was removed the following day and medium was changed every day. Manual passaging was carried using stem cell cutting tool (H-190-210 S, Vitrolife AB). ESC-like colonies were first cut into smaller pieces before scraped off the culture vessel surface. Floating pieces of ESC-like colonies were then aspirated with the stem cell cutting tool and transferred to a new culture vessel pre-coated with vitronectin.

2.2.2 Amaxa® Cell line Nucleofector kits (Lonza) transfection

Cells were trypsinised and appropriate number of cells was pelleted. For each reaction, cells were resuspended in 100 μL nucleofector solution (made up of 82 μL of nucleofector solution plus 18 μL of supplement). A total of 1 μg of plasmid was added to the cell suspension. The mixture was transferred to cuvette and cuvette was closed with cap. Cells were transfected with the appropriate program in Nucleofector 2b device. A total of 500 μL medium was added to the cuvette and the mixture was transferred immediately into culture vessel.

2.2.3 Fate-reprogramming & fate-maintenance media (FRM & FMM)

Medium composition was as described by (Valamehr et al., 2014) and illustrated in Table 2.1. Traditional feeder-dependent medium composition was also listed for comparison.

Table 2.1 Composition of fate-reprogramming (FRM) and fate-maintenance (FMM) media described by Valamehr et al. (2014) as compared to traditional hESC medium.

Feeder-layer	Valamehr et al., 2014	
	Feeder-free	
Conventional hESC medium	**Fate-reprogramming medium (FRM)**	**Fate-maintenance medium (FMM)**
DMEM/F12	DMEM/F12	DMEM/F12
20% KSR	20% KSR	20% KSR
	1x N2	
	1x B27	
1x Glutamine	1x Glutamine	1x Glutamine
1x Non-EAA	1x Non-EAA	1x Non-EAA
100 µM β-mercaptoethanol	100 µM β-mercaptoethanol	100 µM β-mercaptoethanol
100 ng/mL bFGF	100 ng/mL bFGF	100 ng/mL bFGF
	10 ng/mL LIF	10 ng/mL LIF
	5.0 µM Thiazovivin	5.0 µM Thiazovivin
	0.4 µM PD0325901	0.4 µM PD0325901
	1.0 µM CHIR99021	1.0 µM CHIR99021
	2.0 µM SB431542	

*Abbreviations: KSR = Knockout serum replacement, non-EAA = non-essential amino acids, bFGF = basic fibroblast growth factor, LIF = leukaemia inhibitory factor.
** PD0325901 = non-ATP-competitive MEK inhibitor, CHIR99021 GSK3 inhibitor/WNT activator, SB431542 = Activin/BMP/TGF-β pathway inhibitor.

The protocol for iPSC reprogramming using FRM/FMM system was similar to what described in Section 2.1.1, whereby cells were cultured in FRM medium for the first seven days and transfected four consecutive times every other day (last transfection was on day 6). On day 7, medium was switched to FMM and changed every other day until iPSC colonies emerged. Colonies were isolated and expanded on FMM medium. Cells were passaged using EDTA and ROCKi as described earlier.

2.2.4 Small molecule treatment

When indicated, 5-aza-2'-deoxycytidine (decitabine) and sodium butyrate (NaB) were used at 0.5 mM and 0.5 mM, respectively.

2.3 *IN VITRO* DIFFERENTIATION

2.3.1 Embryoid body (EB) formation

Cells were grown to 60 – 80% confluence before being dissociated with 0.5 mM EDTA solution. EDTA was neutralised by E8 medium and mixture was centrifuged to collect cell pellets. Cell count was performed to enumerate cell density. Cells was resuspended in appropriate volume of E8 medium supplemented with 4 mg/ml poly(vinyl alcohol) (PVA) (Sigma-Aldrich, 341584). A total of 5000 – 8000 cells were plated onto each well of ultra-low attachment 96-well plates (Greiner Bio-One GmbH, 650970). EBs were left to form for 2 days before being transferred to free-floating culture in poly(2-hydroxyethyl methacrylate) (poly-HEMA) (Sigma-Aldrich, P3183)-coated T25 or T75 flask for 4 days. At this stage and onwards, EBs were culture in standard EB medium containing: Dulbecco's Modified Eagle Medium (DMEM) (Gibco, 41965039), 20% FCS, 1% glutamine, 1% NEAA and 0.01 mM 2-Mercaptoethanol (Gibco, 31350-010). EBs were then cultured as free-floating culture or transferred onto gelatin-coated (Sigma, G1890) 24-well plate (one EB per well) for attachment and expansion as monolayer culture for 20 days.

2.3.2 Mammary organoid formation

The mammary differentiation protocol was derived from Qu et al. (2017) with little modification. All media used are listed in Table 2.2.

Table 2.2 Composition of media used for generation of mammary organoid (Qu et al., 2017).

Media	Composition
Mammosphere medium (mEB)	DMEM/ F12 medium (Invitrogen, 11330-032) supplemented with 20 ng/mL epidermal growth factor (EGF, R&D Systems), 20 ng/mL basic fibroblast growth factor (bFGF) (R&D Systems), 5 µg/mL insulin, 1% pen/strep, B27, 4 µg/mL heparin, and 5 µg/mL hydrocortisone
pTHrP	MCF10A medium described in Section 2.1.1 (DMEM/F12-based) plus 100 ng/mL parathyroid hormone (pTHrP)
B+A	MCF10A medium with 1 µg/mL hydrocortisone and 10 µg/mL insulin plus 50 ng/mL FGF10, and 50 ng/mL HGF
Prolactin	MCF10A medium with 1 µg/mL hydrocortisone, and 10 µg/mL insulin plus 10 µg/mL prolactin and 10% FCS

*Abbreviation: FGF10 = fibroblast growth factor 10, HGF = hepatocyte growth factor, FCS = foetal calf serum
**pTHrP (100-09), FGF10 (100-26), HGF (100-39H) and proclactin (100-07) were all purchased from PeproTech.

Embryoid body of pluripotent stem cells (PSCs) were prepared as described earlier but with mEB plus PVA instead of E8 and PVA. EBs formed after two days were cultured in suspension with mEB (without PVA) for 10 days. EBs were embedded in mixed gel (containing 2.5 mg/mL Matrigel (BD Biosciences, 354230) and 1 mg/mL Collagen I (Invitrogen, A1064401)) prepared by mixing 3 portions of 10.1 mg/mL Matrigel with 1 portion of 4 mg/mL Collagen I. Mixed gel containing EBs was cultured first in pTHrP medium for 5 days (to induce mammary commitment), B+A medium for 20 days (to induce branch and alveolar differentiation) and lastly prolactin medium for 5 days (to induce milk protein expression).

To harvest organoids for RNA extraction, the medium was removed and well was washed with PBS. Cells were treated with equal volume of Dispase (1 U/mL) (StemCell Technologies, 07923) and Collagenase IV (1 mg/mL) (Gibco, 17104019) for about 1 h. When all gel dissolved, the

mixture was centrifuged for 4 min at 400 xg. The pellet was collected for RNA extraction.

2.4 SOFT AGAR COLONY FORMATION ASSAY

Cells were seeded in 1 mL of 0.5% noble agar in respective culture media overlaying 1 mL 1% agar in the same medium. After 2 weeks culture, cell colonies were stained with 0.005% crystal violet in 10% ethanol (in PBS) for 1 hour and colonies ≥100 μm were counted under a stereomicroscope. Results were presented as mean ± standard deviation, n = 3. A total of 25000 cells (MCF10A, MCF7 and MCF10A_PIK3CA$^{(+/-)}$) or 5000 cells (BT-549) were seeded per well in 6 well plate.

2.5 CHEMOTHERAPY DRUG SENSITIVITY TEST

Cells were seeded at 60000 cells per well onto a 24-well plate. A day after plating, cells were treated with paclitaxel at concentrations ranging from 0.001 to 10 nM for 24 hours. Cell viability was evaluated by MTT cell proliferation assay. MTT (0.5 mg/ml final concentration) was added to the cells and incubated for 3 hours in the incubator. The medium was then removed and the converted dye was dissolved by incubation with acidic isopropanol (isopropanol containing 0.04N HCl) for 20 minutes at room temperature. Absorbance was measured in a spectrophotometer (Varioskan) using a wavelength of 570nm with background subtraction at 650nm. IC_{50} was calculated by using GraphPad Prism version 7.03.

2.6 IMAGING

Bright-field/phase contrast and some GFP images were taken by inverted microscope (Leica, DFC490). Immunofluorescence and immunohistochemistry staining images were taken by upright microscope (Leica, DFC350).

2.7 IMMUNOLOGICAL TECHNIQUES

2.7.1 Fixation

To prepare cell for immunofluorescence staining, cells were grown as monolayer on coverslips in 24-well plate. At 80 – 90% confluence, cells were washed with PBS and fixed with 4% paraformaldehyde (PFA) for 30 min. PFA was prepared by 4 g of PFA powder in 83 mL of sodium dihydrogen phosphate monohydrate (NaH_2PO_4.H2O) solution (2 g/100 mL in water) and 17 mL of NaOH solution (2.5 g/ 100 mL in water). The mixture was heated at 60°C until dissolved. Wells were washed with PBS twice and stored at 4°C in PBS until the immunostaining was performed.

2.7.2 Immunofluorescence (IF) staining

Immunofluorescence staining was performed on cells fixed on coverslips. Coverslips were washed with PBS and cells blocked by 3% bovine serum albumin (BSA) (Sigma-Aldrich, A9418) (plus 0.1 % Triton x100 (Sigma-Aldrich, T8787) for nucleus-localised proteins) for 30 min at room temperature. BSA was removed and cells were incubated with primary antibodies diluted with 3% BSA solution (with or without Triton x100) for 1 h. The primary antibody was then removed and coverslips washed with 0.1%

Tween 20 solution in PBS thrice and 10 min each time. Cells were incubated with the secondary antibody for 45 min at room temperature followed by 3 washed with PBS-0.1% Tween 20 solution. Cells were stained with 4',6-Diamidino-2-Phenylindole, Dihydrochloride (DAPI) and mounted on microscopic slide and sides were sealed with nail varnish. Cells were visualised as described in Section 2.7. All antibodies used are listed in Table S2-3.

2.7.3 Fluorescence activated cell sorting (FACS) analysis

For green fluorescence protein (GFP)-expressing cells, a total of 1 x 10^5 cells was pelleted and resuspended in 250 µL fresh medium before subjected to FACS analysis. For pluripotency marker staining, a total of 1 x 10^5 cells was pelleted in 1.5 mL tubes. Cells were resuspended with 50 µL primary antibody diluted with culture medium and incubated for 1 h at 4°C. A total of 500 µL medium was added and cells were centrifuged for 4 min at 400 xg. Cell pellet was resuspended with 50 µL secondary antibody diluted with culture medium and incubated for 45 min at 4°C. A total of 500 µL medium was added and cells were centrifuged for 4 min at 400 xg. Cell pellets were resuspended in 250 µL medium and analysed by FACS machine (BD FACS Canto). All FACS data was analysed by Weasel version 3.4.2. All antibodies used are shown in Table S2-4.

2.8 MOLECULAR TECHNIQUE AND ANALYSIS

2.8.1 DNA extraction

Genomic DNA was extracted from frozen cell pellets using the NucleoSpin® Tissue kit (Macherey-Nagel, 740952) according to manufacturer's instruction. First, cells were lysed with proteinase K at 56°C. DNA was bound to the silica membrane in the NucleoSpin Tissue Columns. Subsequently, impurities were removed by washing with two different buffers and pure genomic DNA was eluted with RNase/DNase-free water. Concentration and purity of the DNA samples was determined by using NanoDrop 8000 Spectrophotometer (Thermo Scientific, ND-8000-GL). DNA with a 260/280 ratio between 1.8 and 2.1 was deemed to be of good quality.

2.8.2 RNA extraction

RNA extraction was performed using NucleoSpin RNA kit (Macherey-Nagel, 740955) according to the manufacturer's instruction. Cells were lysed by lysis buffer which immediately inactivated RNases. Lysate viscosity was reduced and cleared from unwanted components by filtration through NucleoSpin Filter. RNA was then adsorbed to the silica membrane in the NucleoSpin® RNA Column and DNA was removed by rDNase treatment. Two different buffers were used to remove remaining salts, metabolites and macromolecular cellular components. Lastly, pure RNA was eluted with RNase-free water before storing at -80°C for long-term storage. Concentration and purity of RNA was determined using NanoDrop 8000 Spectrophotometer as described above.

2.8.3 cDNA synthesis

Reverse-transcription was performed using the Super Scrip IV First-Strand Synthesis System kit (Invitrogen, 18091050) according to manufacturer's instruction. For <5 µg of total RNA, 1 µL of 10 mM dNTP mix (Invitrogen, 18427-013) and 1 µL 50 µM random hexamers (Applied Biosystems, M12897) were added and made up to a volume of 13 µL with nuclease-free water. The mixture was heated to 65°C for 5 minutes incubated on ice for at least 1 minute. Subsequently, 4 µL of 5x SSIV buffer, 1µL of 0.1 M DTT, 1 µL SuperScript IV Reverse Transcriptase and 1µL of RNaseOUT (Invitrogen, 10777-019) were added to the mixture and incubated at 23°C for 10 min, 50-55°C for 10 min and 80°C for 10 min in a thermocycler.

2.8.4 RT-PCR

Reverse transcription polymerase chain reaction (RT-PCR) was used for gene expression detection and amplification of gene of interest for downstream analyses such as sequencing. REDTaq ReadyMix PCR reaction mix (Sigma-Aldrich, R2523) was used. In a volume of 20 µL reaction, 10 µL of REDTaq mix, 0.5µL of each 10 µM forward and reverse primers, 1 to 5µL of template DNA and ddH2O were mixed in sterile PCR tubes. The tubes were transferred into a thermocycler, PCR run with the following conditions and the PCR products were separated on agarose gels before visualised under UV light with ImageQuant Capture 300 (GE Healthcare, 63-0056-52). All primers used in this study are listed in Appendix Table S2-1. Amplicons generated were analysed by Sanger sequencing to confirm amplification of the right target genes (Appendix Figure S2-8).

Steps	Temperature, °C	Time	Cycles
Initial denaturation	95	5 min	1
Denaturation	95	30 s	35
Annealing	45 – 60	30 s	
Extension	72	30 s/kb	
Final extension	72	5 min	1
Hold	4	∞	

2.8.5 RT-qPCR

Reverse transcription real-time polymerase chain reaction (RT-qPCR) was used to quantify gene expression. It was performed using LuminoCt SYBR Green qPCR ReadyMix (Singma-Aldrich, L6544). Each RT-qPCR reaction made up of 15 µL mix: 7.5µL SYBR green, 1 µL Forward primer (5 µM), 1 µl Reverse primer (5 µM), 3.5 µL ddH20 and 2 µl cDNA (0.5 ng/µL). The RT-qPCR running program was as follows; initial 95°C for 5 min, 40 cycles of 95°C for 3 seconds (s), and 72°C – 30 s (end cycle) followed by a melting curve. To determine efficiency of each primer set, cDNA concentration of positive control ranging from 0.01 ng to 10 ng was analysed. Standard curve (Ct against log(quantity of cDNA)) was created on Microsoft Excel and primer efficiency was determined by using online qPCR efficiency calculator (provided by ThermoFisher Scientific, https://www.thermofisher.com/uk/en

/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/tm-calculator.html). Primers with efficiency ranging from 90 – 110% (or 1.8 – 2.1) were considered acceptable. In addition, the $2^{-\Delta\Delta Ct}$ approach was used for normalising and calculating fold change (with primers demonstrating close efficiencies). All primers used in this study for RT-qPCR are shown in Appendix Table S2-2. Primer meting curves and efficiency scatterplots are shown in Appendix Figure S2-6 and S2-7. Amplicons

generated were analysed by Sanger sequencing to confirm amplification of the right target genes (Appendix Figure S2-8).

2.8.6 Gel Electrophoresis

Gel electrophoresis was performed to separate DNA products on agarose gels. Varying concentrations of gel was used to separate DNA products of different sizes (eg. 2% agarose gel was used for 50-150 bp DNA and 1% for DNA between 150 bp – 2 kb). Agarose gel was prepared by dissolving agarose powder (Fisher, BPE1356) in TAE buffer (0.04 M Tris-Acetate, 1 mM EDTA) in a conical flask and heated in a microwave until powder had completely dissolved. Mixture was cooled down to ~50°C before adding Nancy-520 (Sigma-Aldrich, 01494). Mixture was poured into gel cassette with a comb and allowed to solidify. Solidified gel was transferred to gel tank and submerged in fresh TAE. A total of 1-3 µL of DNA ladder (1kB (NEB, N3232S), 100 bp (NEB, N3231S) or Low molecular weight DNA ladder (NEB, N3233S)) was used as size reference. Gel tank was connected to PowerPac basic power supply (Bio-Rad, 1645050) and electrophoresis was performed at 100 – 120 volts for 40 – 70 min. The gel was visualised using ImageQuant Capture 300. PCR products appeared as bands, the sizes of which were determined by comparing with the DNA ladder.

2.8.7 Growing & freezing bacteria

To generate large amount of plasmids DNA for different experiments, bacterial cells containing desired plasmids were grown in Luria-Bertani (LB) nutrient agar (OXOID, CM0003B) containing appropriate antibiotic (e.g. 50 µg/mL ampicillin) overnight. Individual colonies were picked and added to 5 mL of LB nutrient broth containing antibiotic in a universal tube

(Barloworld Scientific LTD, 128A). The mixture was incubated at 37°C and shaken at 150-220 rpm overnight. Bacterial cell pellets were collected by centrifugation for 5 min at 6000 x g. Pellets were subjected to plasmid purification or stored at -20°C. To keep stock for plasmids, 800 µL of the bacteria cells was added to 200 µL glycerol and mixed. This mixture was kept at -80°C for long-term storage.

2.8.8 Plasmids purification

The purification of plasmids from bacterial cultures was performed using a NucleoSpin Plasmid QuickPure (Macherey-Nagel, 740615). Bacterial cell pellet was re-suspended in 250 µL of Buffer A1. The mixture was vortexed before adding 250 µL of Buffer A2. Mixture was inverted 6-8 times and incubated at room temperature for 5 min. A total of 300 µL of Buffer A3 was added and inverted 6-8 times before being centrifuged for 5 min at 11000 xg. The supernatant was loaded on to a NucleoSpin column and centrifuged for 1 min at 11,000 xg to bind the DNA to the column. The membrane was washed and dried before plasmid DNA was eluted with 50 µL ddH20. Concentration and purity of plasmid was determined using Nanodrop machine described earlier. For larger scale plasmid purification, NucleoBond Xtra Midi (Macherey-Nagel, 740410) was used.

2.8.9 Restriction Enzyme Digest

Restriction enzyme (RE) digest was performed in a 20 µL mix containing 1µg of DNA, 2 µL of buffer, 1 µL (10 units) of RE and ddH_2O. Mixture was incubated at 37°C (varied depending on enzymes) for 1 h. Digested products were resolved by gel electrophoresis on agarose gels.

2.8.10 Gel and PCR purification

To purify PCR products from an agarose gel or directly from the PCR mix for downstream analyses or application, the NucleoSpin Gel and PCR clean-up kit (Macherey-Nagel, 740615) was used. For purification from agarose gel, PCR band containing the desired DNA was excised, mixed with buffer NTI (200 µL/100 mg) and incubated at 50°C for 10 – 20 min until the agarose had completely dissolved. For purification from PCR mix, two volumes of buffer NTI was added to one volume of PCR mix. The above mixtures were transferred to a NucleoSpin Gel and PCR clean-up column and spun at 11000 xg for 30 s. The column was washed with buffer NT3 and dried before DNA was eluted with 20 µL of ddH_20. Concentration and purity of plasmid was determined using Nanodrop machine described earlier.

2.9 CRISPR-CAS9 EXPERIMENTAL DESIGN

2.9.1 SgRNA and ssODN design

Short guide RNA (sgRNA) was designed by using the online CRISPR tool (http://crispr.mit.edu). SgRNA (CAAATGAATGATGCACATCA) with the highest quality score and lowest off-target sites was selected. A 141bp long single-stranded oligodeoxynucleotide (ssODN) template: 5'-tgt gtg gaa gat cca atc cat ttt tgt tgt cca gcc acc atg ATG tgc atc att cat ttg ttt cat GAA ata ctc caa agc ctc ttg ctc agt ttt atc taa ggc tag ggt ctt tcg aat gta tgc aat gtc atc aaa aga-3', was designed based on *PIK3CA* genomic DNA sequence (NG_012113.2) obtained from National Centre for Biotechnology Information (NCBI) with SnapGene version 4.1 software.

2.9.2 Cloning

2.9.2.1 Reannealing sgRNA & addition of 5'phosphate group

Short guide RNA (sgRNA) was inserted into a sgRNA expression vector (Addgene #41824) and transfected with Cas9-GFP-containing plasmid (Addgene #44719) during CRISPR-Cas9 experiment. Sense and anti-sense strands of sgRNA containing overhang complementary to overhang created by BsaI restriction digest in sgRNA expression vector were purchased separately from Sigma-Aldrich:

Sense 5' **accg**CAAATGAATGATGCACATCA 3'

Antisense 3' GTTTACTTACTACGTGTAGT**caaa** 5'

*bold and small letter = overhang complementary to BsaI overhang in 41824 plasmid.

Both strands were reannealed through reaction mix as follow**:**

Components	Amount, µL
sgRNA sense (100 µM)	1
sgRNA anti-sense (100 µM)	1
NEB buffer 2 (10x)	5
ddH2O	43
Total	50

The mixture was incubated for 4 min at 95°C, then 10 min at 70°C on heat block and cooled down to room temperature for several hours. Before ligating sgRNA (the insert) to sgRNA expression plasmid, 5'phosphate group was added. The following mixture was prepared:

Components	Amount, µL
Annealed sgRNA (from above)	1
ATP (10Mm)	1
TNK buffer (10x)	1
T4 PNK	1
ddH2O	6
Total	10

The mixture was incubated at 37°C for 30 min and then 65°C 20 min to inactivate enzyme.

2.9.2.2 sgRNA expression vector preparation

The sgRNA expression plasmid vector (Addgene #41824) was treated with the RE BsaI. Excised plasmid was resolved with gel electrophoresis and purified as described earlier. Plasmid was dephosphorylated using calf intestinal alkaline phosphatase (CIP):

Components	Amount, µL
Digested plasmid	1.5 µg
Cutsmart buffer	3
CIP	0.75
ddH2O	X
Total	30

The mixture was incubated at 37°C for 1 h before purification using NucleoSpin Gel and PCR clean-up kit as described earlier.

2.9.2.3 Ligation

To ligate reannealed-sgRNA to excised plasmid, the following mixture was prepared:

Components	Amount, µL
CIP Plasmid	100 ng
Oligo	10
T4 ligase buffer	1.5
T4 ligase	1
ddH2O	X
Total	15

Mixture was incubated at room temperature for 20 min then at 17°C or 4°C overnight.

2.9.2.4 Bacterial transformation

Competent *E. Coli* cells Stbl3 (Invitrogen, C7373-03) were thawed on ice. The ligation reaction was added to the bacteria and incubated on ice for 20 min. The bacteria were then heat shocked at 42°C for 45 seconds and then incubated on ice for 2 minutes. A total of 200 – 250 µL SOC outgrowth medium (Invitrogen, 15544-034) was added to cells and the mixture was incubated at 37°C and shaken at 150 rpm for 45 – 60 min. This mixture was then spread onto LB agar plates containing 50µg/mL kanamycin (Sigma-Aldrich, K1377) and incubated at 37°C overnight. Kanamycin resistant colonies were formed on the plate as a result of successful transformation. Single colonies were picked and expanded in LB broth to generate sufficient plasmid DNA and bacterial stock as described earlier.

2.9.3 SgRNA and Cas9-GFP plasmids transfection

A total of 5 x 10^4 MCF10A cells was plated onto one well of 24-well plate one day before transfection. Used medium was discarded and 400 µL fresh medium was added one hour before transfection. FuGENE 6 transfection reagent (Roche, 11814443001) reaction mix was prepared as follow:

Components	Volume, µL
Cas9-GFP plasmid	0.275 µg
SgRNA plasmid	0.275 µg
FuGENE 6	1.95
OptiMEM	25.5
Total	30

The mixture was incubated at room temperature for 15 min. The reaction mix was added to cells in a dropwise manner. After 24 h, transfection of ssODN was performed with Lipofectamine 3000 transfection reagent

(Invitrogen, L3000001). The reaction mix was prepared by mixing solution A and B:

Components	Volume, µL
Solution A	
ssODN	1.0 µg
P3000 Reagent	2.0
OptiMEM	25
Solution B	
Lipofectamine 3000	3.0
OptiMEM	25

The mixture was incubated for 15 min at room temperature before added to cells. The next day, cells were trypsinised, trypsin was inactivated by equal volume of medium and centrifuged to collect cell pellet. The pellet was resuspended with medium to achieve 1 x 10^5 cells/ mL. GFP^{high} cells were sorted by Beckman Coulter Astrios EQ sorter (BD, B25982). Sorted cells were plated as single cells in 96-well plate filled with fresh medium supplemented with ROCKi to improve cell viability. The next day, EGF was removed from MCF10A medium to facilitate selection for cells harbouring *PIK3CA* mutation. At 70 – 80% confluence, each clone passed into two different 96 well plates, one at 1:4 and the other 3:4. The latter was used for screening with BspH1 enzyme.

2.9.4 DNA extraction

DNA extraction was performed at 90% confluence and in 96-well format with a custom-made lysis buffer:

Components	Amount
Tris pH 7.5-8.0	10 mM
EDTA	10 mM
NaCl	10 mM
Sarcosyl	0.5%
Proteinase K (added fresh)	40 g/mL

Cells were washed twice with 150 µL PBS/ well. A total of 50 µL lysis buffer was added to each well. Plates were covered with parafilm, lid replaced and wrapped in parafilm. Plates were again wrapped in a plastic bag and incubated at 60°C overnight. DNA was precipitated by 100 µL/well 75 mM NaCl dissolved in cold 95% ethanol. Plates were incubated at -80°C for 2 h, spun for 5 min at 3000 rpm. Supernatant was decanted, wells were washed three times with 150 µL/well of 70% ethanol. Residual ethanol was evaporated at room temperature or 60°C. DNA was resuspended with 30 µL/well ddH_2O overnight and stored at -20°C.

2.9.5 Screening

PCR reaction mix (20 µL) and PCR condition were described in Section 2.3.4. A total of 2.0 – 3.0 µL DNA was used per reaction. 3 µL of the PCR reaction mix was digested with BspHI in a 20 µL reaction mix as described in Section 2.3.9. Digested PCR products were resolved by gel electrophoresis.

2.9.1 *TP53* mutation transfection

The Addgene plasmid 85559 carries sgRNA sequence (GGGCAGCTACGGTTTCCGTC) which targets exon 4 of *TP53* and Cas9 protein. A transfection was performed with FuGENE 6 transfection reagent as described earlier. *TP53*-null cells were selected by 10 µM Nutlin-3 (Sigma, N6287) treatment for two weeks before plated as single cells in 96-well plate. To identify mutant clones, the targeted region was amplified by PCR to generate a 418 bp product. The amplicons were subsequently denatured and re-annealed with the following conditions:

Temperature, °C	Time
95	5 min
95 – 85	Ramp rate: -2°C /s
85 – 25	Ramp rate: -0.1°C/s
10	∞

Mutants generated a mixture of homoduplexes (from two complementary single stranded sequences, such as WT/WT or Mut/Mut) and heteroduplexes (products of two non-complementary single stranded sequences, such as WT/Mut) as described in Figure 6.9. These duplexes were resolved by 15% polyacrylamide gel electrophoresis (PAGE) prepared by mixing components as follow for 2 – 3 h at 150 V.

Components	Volume, mL
30% acrylamide (w/v) (Sigma-Aldrich, A3699)	2.5
5x TBE	1.0
H_2O	1.5
10% APS (Sigma-Aldrich, A3678)	83.3 µL
TEMED (Fisher, CAS 110-18-9)	4.17 µL
Total	~5

*APS = Ammonium Persulfate, TEMED = tetramethylethylenediamine, TBE = Tris/Borate/EDTA

The gel was visualised by ImageQuant Capture 300. Potential clones were subjected to Sanger sequencing (service provided by Source BioScience company) to identify mutations.

2.10 WESTERN BLOTTING

2.10.1 Protein extraction

Cells were trypsinised, pelleted in a 1.5 mL tube and dissolved in 150 µL of 1 x RIPA buffer (20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM Na_2EDTA, 1 mM EGTA, 1% NP-40, 1% sodium deoxycholate, 2.5 mM sodium pyrophosphate, 1 mM β-glycerophosphate, 1 mM Na_3VO_4, 1 µg/ml leupeptin; New England Biolabs, 9806S). Protease Inhibitor cocktail (Sigma-Aldrich, P8340) (1:200) was added to each sample for cell lysis. Tubes was vortexed and incubated on ice for 15 min. The solution was centrifuged and proteins were quantified with the Qubit fluorometer (Invitrogen, Q32866). Protein solutions were stored at -80°C for long-term use.

2.10.2 Polyacrylamide Gel Electrophoresis of Protein Samples

Polyacrylamide gel was made up of stacking gel (top) and resolving gel (bottom). Both gels were prepared as follow:

Components	Stacking gel, mL	Resolving gel, mL
H_2O	2.1	1.6
30% acrylamide (w/v)	0.5	2.0
SDS	0.03	0.05
APS	0.03	0.1
TEMED	0.003	0.002
1.0 M Tris pH 6.8	0.38	-
1.5 M Tris pH 8.8	-	1.3
Total	~3	~5

A total of 1-400µg of protein samples were mixed with 4x loading buffer (4x Laemmli Sample Buffer, 1610747) and 50µL/mL β-mercaptoethanol. Samples were denatured at 100°C for 3 min and loaded onto polyacrylamide

gel first resolved at 150 V for ~10 min and then at 180 V for 1 h. Broad Range Protein Marker (NEB, P7702) was used as a loading control.

2.10.3 Western Blotting Transfer

Proteins were transferred to polyvinylidene difluoride (PVDF) membrane (GE Healthcare, 10600023) after being resolved in polyacrylamide gel using Fisherbrand Semi-Dry Blotters (Fisher, 12398556). The PVDF membrane was activated by wetting it in 100% methanol before incubating it in transfer buffer (39mM glycine, 48mM Tris-base, 0.037% SDS and 20% methanol) for 10 min. The membrane was placed on the anode plate, with the gel on top of it. The membrane and gel were sandwiched between 3 Whatman filter papers on each side pre-soaked in transfer buffer to allow the current to pass through. The layers were pressed down by rolling a 5mL pipette over to remove air bubbles. The transfer ran at 80 mA for 30 min.

2.10.4 Membrane Blocking and Antibody Incubation

PVDF membrane was blocked in 5% non-fat milk (Marvel) or 1% albumin from bovine serum (BSA; Sigma-Aldrich, A9418) in Tris-buffered saline (TBS)-tween 20 (TBST) (TBS and 0.5% Tween 20) for 1 hour at room temperature on a shaker. PVDF membrane was stained with primary antibody diluted in 5% milk overnight at 4°C on a roller. The membrane was then washed three times (10 min each) with TBST. Horseradish peroxidase (HRP) secondary antibody diluted 5% milk was added to the membrane and incubated for 1 hour at room temperature on a shaker. The membrane was then washed three times (10 min each) with TBST. To perform chemiluminescence, the membrane was covered with excess

mixture of solution A and solution B at 1:1 ratio (Amersham GE Healthcare, ECL, RPN2235) and incubated for 1 min at room temperature. The excess solutions were drained off on a paper towel and the membrane was wrapped in Saran Wrap film (Dow, SEL-360-500Q). The membrane was visualised using the ChemiDoc™ MP Imaging System (Bio-Rad, 17001402). All antibodies used are shown in Table S2-5.

2.11 STATISTICAL ANALYSIS

All statistical analyses were performed on GraphPad Prism version 7.03. Unless otherwise stated, data were expressed as mean ± standard deviation (SD) for biological replicates. For RT-qPCR and soft agar colony forming assay, data was analysed by one-way analysis of variance (ANOVA) test followed by Bonferroni correction. CD24 and CD44 marker expression was analysed by two-way ANOVA. Chemotherapy drug sensitivity test data was analysed by non-linear regression analysis.

3. REPROGRAMMING OF CONTROL CELL LINE MCF-10A

3.1 INTRODUCTION

The present study aimed to investigate the contribution of genetics and epigenetics to breast carcinogenesis. To achieve this goal, this project aimed to reset the epigenome of breast cancer cell lines via reprogramming to iPSCs. The immortalised mammary epithelial cell line MCF10A was used as a non-malignant control to optimise the reprogramming conditions and subsequent iPSC characterisation protocols.

The use of viruses (lentivirus and retrovirus) in iPSC reprogramming presents high risk of insertional mutagenesis and many other caveats, including incomplete silencing and reactivation of transcription factors thus prohibiting the translational applications of iPSCs (Yamanaka and Takahashi, 2006, Robinton and Daley, 2012, Lu and Zhao, 2013). Various non-integrative and transgene-free methods have been developed to circumvent these drawbacks, striving for safer and more accessible reprogramming methods for iPSC generation (Theunissen and Jaenisch, 2014a) (section 1.3.3.2). The introduction of reprogramming factors via transfection of oriP/EBNA1-based episomal vectors is currently one of the most frequently used non-integrative reprogramming approaches (Yu et al., 2009, Okita et al., 2010). The first episomal-based reprogramming report described the use of seven reprogramming factors (*OCT3/4, SOX2, KLF4, C-MYC, LIN28* and *NANOG, OSKMLN*) in three different vector combinations (Yu et al., 2009). Subsequently, Okita et al. (2011) proposed a modified version of the

reprogramming cocktail encompassing *TP53* short-hairpin RNA (shRNA), four transcription factors (OSKL) and *L-MYC* instead of *C-MYC* in three separate episomal plasmids. Transient downregulation of *TP53* has been linked to enhanced reprogramming efficiency by promoting cell proliferation and survival in previous studies (Zhao et al., 2008, Hong et al., 2009a, Kawamura et al., 2009, Marion et al., 2009, Utikal et al., 2009b, Banito and Gil, 2010). The replacement of *C-MYC* with less-transforming, more efficient and specific *L-MYC* has been shown to reduce the risk of tumorigenicity of iPSC reprogramming (Nakagawa et al., 2010).

The episomal vectors contain two components of the Epstein-Barr virus (EBV): a cis-acting oriP element and a trans-acting *EBNA1* gene. The EBNA1 protein binds vector to the host genome via oriP, enabling extrachromosomal replication once per cell cycle by utilizing the host transcription machinery (Yates et al., 1984, Yates et al., 1985, Okita et al., 2013). The addition of an extra vector encoding the EBNA1 protein to the reprogramming cocktail has been shown to significantly improve reprogramming efficiency (Okita et al., 2013). Upon transfection, episomal vectors reduce by approximately 5% per cell cycle owing to the plasmid synthesis defects and partitioning in the absence of drug selection, permitting the generation of transgene-free iPSCs (Nanbo et al., 2007, Cheng et al., 2012). Although rare incidences of genome integration have been reported, episomal vectors remains one of the most widely used non-integrative reprogramming methods (Schlaeger et al., 2015).

In the present study, iPSC reprogramming was carried out with a combination of four episomal vectors encoding *OCT3/4*, shRNA for *TP53*, *SOX2*, *KLF4*, *LIN28*, *L-MYC* and *EBNA1* (Okita et al., 2011, Okita et al., 2013). The first reprogramming experiment was performed with MCF10A cells. MCF-10A is a non-transformed, spontaneously-immortalised human mammary epithelial cell line that harbours a t(3;9)(p14;p21) reciprocal translocation mutation. Fluorescence *in situ* hybridisation (FISH) and comparative genomic hybridisation (CGH) studies revealed that the t(3;9)(p14;p21) mutation might play a role in the acquisition of the immortalised phenotype by disrupting the expression of the *TAFA1* gene and causing homozygous loss of the *p15/p16* genes (Cowell et al., 2005). MCF10A has been widely used as an *in vitro* model in mammary cell-related studies, most notably breast cancer research where it is used as a surrogate of "normal" non-tumorigenic mammary cell control (Cowell et al., 2005, Konishi et al., 2007, Marella et al., 2009, Qu et al., 2015).

Di Nicolantonio et al. (2008) reported the successful generation of knock-in MCF-10A lines harbouring single or multiple mutated oncogenes and tumour suppressor genes (*BRAF, EGFR, KRAS* and *PIK3CA*). These cell lines serve as potential disease models to study cancer development and enable the discovery and development of anti-cancer therapies (Di Nicolantonio et al., 2008). In the present study, MCF-10A was used as non-malignant control to be reprogrammed alongside its derived isogenic lines, harbouring $PIK3CA^{H1047R(+/-)}$ alone or together with $TP53^{(-/-)}$ mutation, and the breast cancer cell lines MCF7, BT-549 and MDA-MB-231.

Therefore, the aim of this chapter was to optimise the conditions for episomal vector iPSC reprogramming by using the MCF10A cell line. The iPSC line derived was then used to optimise different assays to characterise PSCs, including RT-PCR, RT-qPCR and immunofluorescence staining for the expression of pluripotency markers and served as the normal non-cancerous counterpart for comparative studies performed later in this study.

3.2 RESULTS

3.2.1 Optimisation of transfection conditions

Zhao et al. (2015) reported a highly optimised protocol to generate iPSCs from human lung adenocarcinoma (A549) and mouse melanoma (B16f10) cancer cell lines using episomal vectors and X-tremeGENE 9 DNA transfection reagent (XG9). The same study indicated initial seeding cell density as the most essential factor for successful cancer cell reprogramming (Zhao et al., 2015). Therefore, the transfection efficiency of XG9 was first determined by transfecting MCF10A with the episomal vector pCXLE-EGFP (expressing enhanced green fluorescence protein, EGFP; Addgene #27082) followed by measurement of the GFP signal at different time points with fluorescence-activated cell sorting (FACS) analysis. Three cell densities (D1 = 1 x 10^5, D2 = 0.5 x 10^5 and D3 = 0.25 x 10^5 cells per well of 6-well plate) were used to determine the optimum cell density for iPSC generation.

The highest GFP signal was recorded between 48 – 72 h after transfection for three different densities and slowly decreased at later time points. D1 resulted in confluent culture at 96 h and thus was not selected, as high confluency could impede the emergence and expansion of iPSC colonies. Owing to their transient presence, episomal vectors are required to be transfected for four consecutive times every other day to maintain ectopic expression of transcription factors for successful iPSC generation. The transfection of MCF10A with XG9 resulted in pronounced cell death owing to the toxicity inherent to polymer reagents as previously reported

(Hunter, 2006, Park et al., 2006). Approximately 10-15% of cells remained 48 h after transfection for cells seeded at D3 (lowest cell density) (Figure 3.2). It was deemed that D3 might be too low to provide a suitable number of cells for iPSC generation. Therefore, D2 was considered as the optimum density for MCF10A reprogramming.

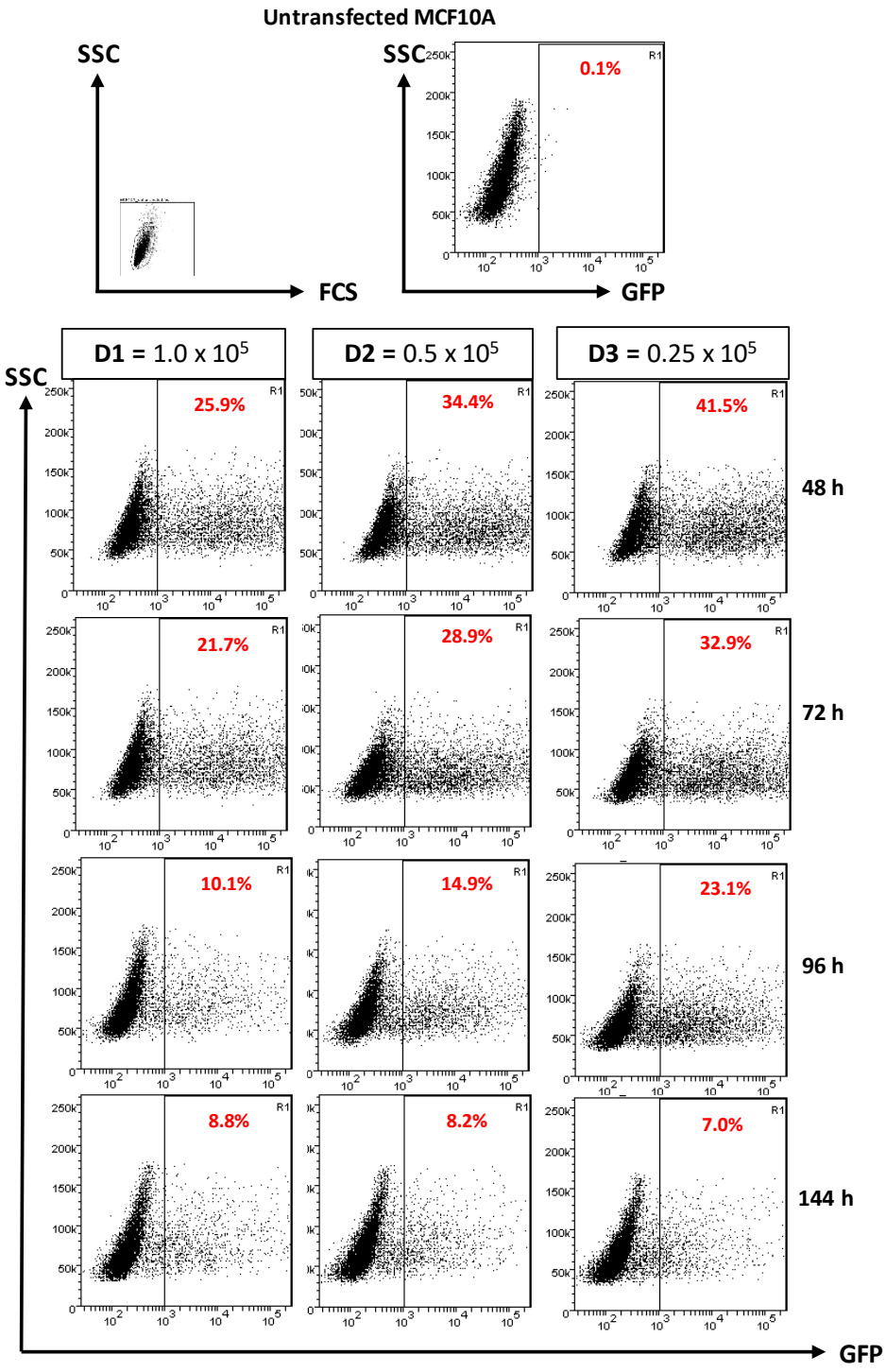


Figure 3.1 Optimization of MCF-10A cell density with X-tremeGENE™ 9 DNA transfection reagent for reprogramming.

Untransfected MCF10A was used as negative control. Transfection efficiencies of three cell densities were assessed based on GFP signal. The highest GFP signal was detected in all three densities at between 48 – 72 h. Signal diminished overtime indicating loss of episomal vector. n = 1. SSC = side scatter.

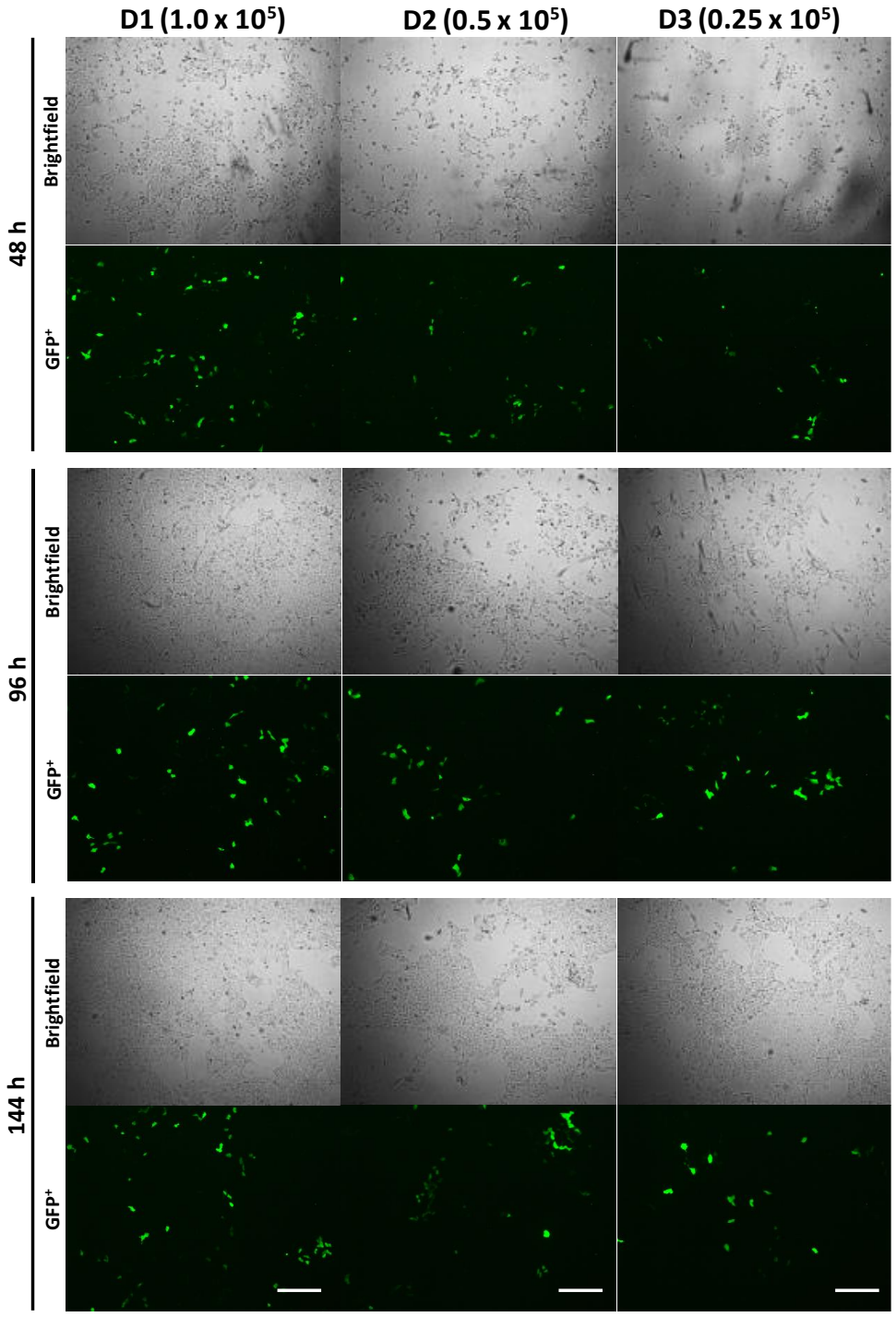


Figure 3.2 Representative images showing GFP$^+$ cells at different time points after transfection under light (top) and fluorescence microscopes (bottom).

Cell death was observed in all three densities. The effect was more pronounced in D3, as only approximately 10% confluence was observed at 48 h after transfection. Cells seeded at D1 were close to 100% confluence at 96 h. D2 was therefore deemed the optimum starting cell density for MCF10A reprogramming. Scale bar = 200 μm.

3.2.2 Episomal vector-mediated reprogramming of MCF10A

After optimising the seeding density, iPSC reprogramming experiment was performed. MCF10A cells were transfected with four episomal vectors encoding *OCT3/4* and shRNA against *TP53* (Addgene #27077), *SOX2/KLF4* (Addgene #27078), *L-MYC/LIN28* (Addgene #27080) and *EBNA1* (Addgene #37624) (Okita et al., 2011) using the X-TremeGENE™ 9 DNA (XG9) transfection reagent. Cells were plated the day before transfection on 100 mm dishes pre-coated with vitronectin. On the day of the first transfection (Day 0), culture medium was changed to Essential 8 (E8) medium (supplemented with 1 μM hydrocortisone) before switching to Essential 6 medium (supplemented with 100 ng/mL human basic fibroblast growth factor, bFGF) on Day 7. Four consecutive transfections were performed every other day. From Day 20-30, potential iPSC colonies were manually passaged and transferred into a new culture vessel with E8 medium supplemented with 10 μM Y-27632 (Rho-associated protein kinase, ROCK inhibitor). The Y27632 was removed the following day and medium was changed every day (Figure 3.3).

A total of five sizable ESC-like colonies were successfully harvested from three 100 mm dishes seeded at 3.7 x 10^5 cells per dish (or equivalent to 0.6 x 10^5 cells per well of 6-well plate), resulting in an efficiency of ~0.0005% (without taking into consideration small and unhealthy colonies that could potentially give rise to iPSC lines). The five iPSC lines were similar to each other in term of growth rate and morphological characteristics, including a high nucleus-to-cytoplasm ratio and preferentially expanded as

compact colonies which exhibited well-defined edges, full-fuilling all morphological characteristics of ESCs. Out of the five iPSC lines isolated, one was randomly selected for further characterisation. Early passages of the selected MCF10-derived iPSC line (named 10AR) were heterogeneous, containing a mixture of reprogrammed (ESC-like) and non-reprogrammed or partially reprogrammed (MCF10A-like) cells. After about 4-6 passages in E8 medium with EDTA-based passaging (which preferentially dissociate pluripotent stem cells while leaving more adherent differentiated cells behind) (Beers et al., 2012), ESC-like cells became predominant in culture. Sizable colonies were again manually passaged and eventually a homogenous culture of 10AR was obtained. A control was set up and cultured with the same E8/E6 media system without being transfected with episomal vectors. The control continued to expand after medium was switched to E8 at D0 of reprogramming and attained confluency without signs of growth attenuation. However, cell death was observed when cells were passaged, proliferation was halted and cells eventually perished, indicating E8 did not support or favour the growth of MCF10A.

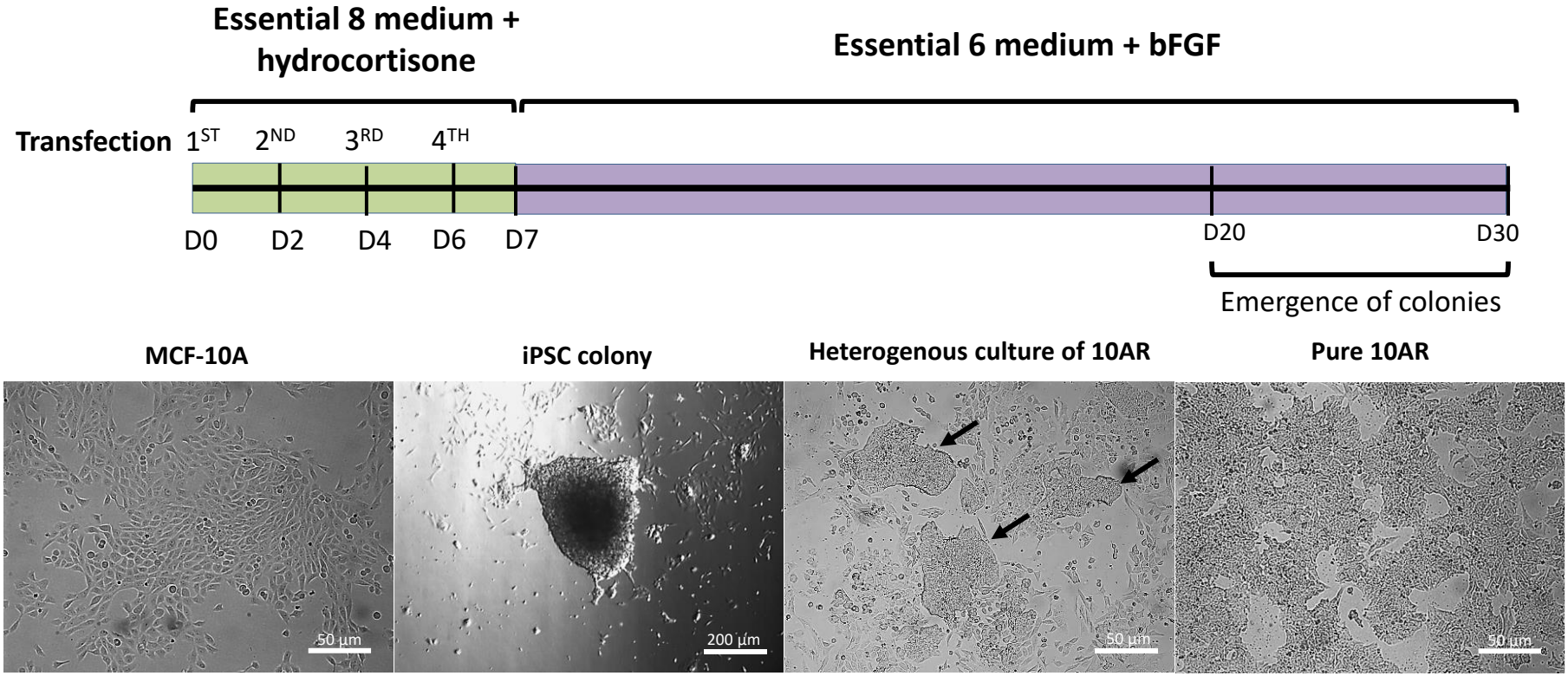


Figure 3.3 Generation of iPSCs from MCF10A.

(top) Reprogramming timeline of MCF10A. On the day of first transfection (Day 0), culture medium was changed to E8 medium (supplemented with 1 µM hydrocortisone) before switching to E6 medium (supplemented with 100 ng/mL bFGF) on Day 7. On day 20 – 30, potential iPSC colonies were manually passaged and expanded in E8 medium. (bottom) Morphologies of MCF-10A at different stages of reprogramming. Early passage iPSCs were heterogeneous, consisting of both non-reprogrammed and ESC-like (black arrows) cells. Through EDTA and manual passaging of ESC-like colonies, pure 10AR cultures were established.

3.2.3 Characterization of the reprogrammed MCF-10A line (10AR)

3.2.3.1 Expression of pluripotency markers

RT-PCR analysis was carried out to detect if episomal traces remained in 10AR using three PCR primers described by Okita et al. (2011). These primers, specific for the plasmids 27077 (pCXLE-hOCT3/4-shP53-F), 27078 (pCXLE-hSOX/KLF4) and 27080 (pCXLE-hLIN28A/L-MYC), spanned part of the respective genes they carry and the episomal vector sequences, thus capable of identifying possible episomal vector leftover in the cells or exogenous gene integration into the host genome. Cells of three different passages (12, 18 and 21) were subjected to analysis. Result indicated loss of all episomal plasmids in 10AR at around passage 18 (Figure 3.4).

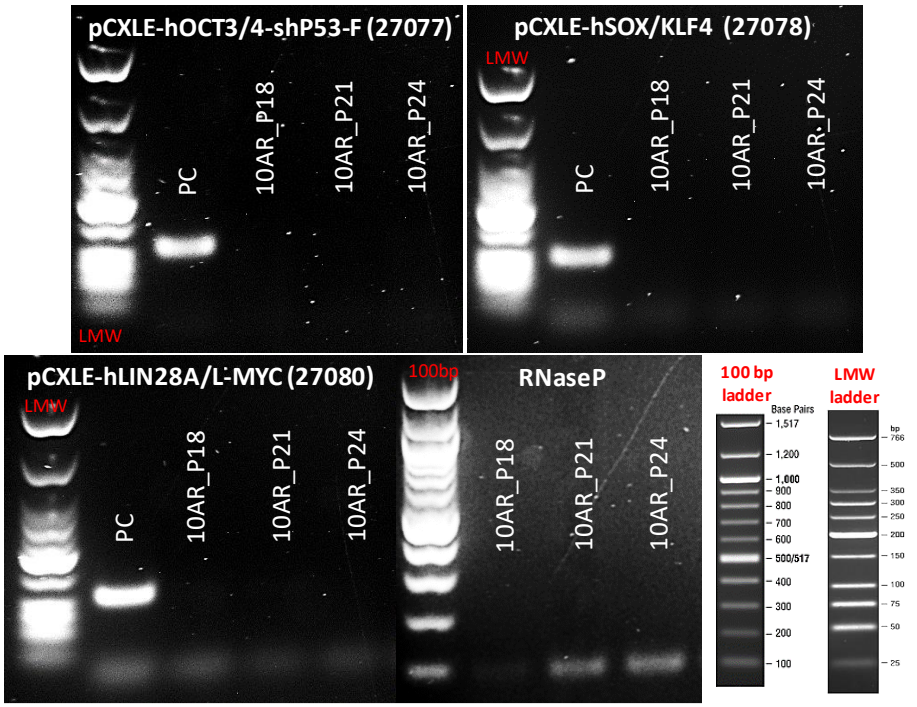


Figure 3.4 Episomal traces in reprogrammed MCF10A (10AR). RT-PCR indicated the absence of all episomal plasmids at as early as passage 18 in 10AR. *RNaseP* was used as a copy number reference. PC = positive control.

Subsequently, RT-PCR was performed to determine the expression of endogenous pluripotency markers, including *REX1, NANOG*, *OCT3/4*, *DPPA4*, and *DNMT3B*. Detection of *NANOG* expression at either transcriptional or translational levels can be problematic owing to the presence of eleven pseudogenes, ten processed (*NANOG P2 – 11*) and one duplicated (*NANOG P1*), in the human genome (Booth and Holland, 2004). Out of the eleven pseudogenes, *NANOG P8*, despite having an amino acid mismatch (Figure S3-1), encodes an active protein functionally similar to *NANOG* (Booth and Holland, 2004, Fairbanks and Maughan, 2006) The expression of *NANOG P8* has been reported in many tumours such as breast (Ezeh et al., 2005, Alldridge et al., 2008), brain (Zbinden et al., 2010), kidney (Bussolati et al., 2008), ovary (Zhang et al., 2008b) and cervix (Ye et al., 2008). Despite the fact that few tumours such as glioma stem cells express both *NANOG* and *NANOG P8*, only *NANOG* has been detected in hESCs, indicating *NANOG* but not *NANOG P8* is an important marker for ESC-like pluripotency (Ambady et al., 2010). In this study, three primer sets capable of discriminating *NANOG* from *NANOG P8* were used. These primers take advantage of the 5'UTR (N444 primer) and 3'UTR (NF2R2 and NF2R3 primers) mismatches between *NANOG* and *NANOG P8* respectively (Zhang et al., 2006, Jeter et al., 2009) (Appendix Figure S3-2). The N444 and NF2R2 primers were specific for *NANOG* while NF2R3 gave rise to two PCR products of different sizes, indicating expression of *NANOG* (467 bp) or *NANOG P8* (446 bp). Only *NANOG* expression, but not *NANOG P8*, was detected in 10AR and H9 in this study. In addition, two forward *OCT3/4* primer, namely the *OCT3/4* F1 and *OCT3/4* F2 that share the same reverse primer was derived from Liedtke et al. (2007) to specifically target *OCT3/4*

responsible for pluripotency by excluding all other pseudogenes and transcript variants.

All endogenous pluripotency markers spanned two or more exons and thus were capable of discerning cDNA template from possible genomic DNA (gDNA) contamination. The embryonic stem cell line H9 was used as a positive control. Negative control (NC) consisted of complete PCR mix without cDNA template. The housekeeping gene, *B-ACTIN* (*ACTB*) was also included as an endogenous control. Result indicated that H9 and 10AR, but not MCF10A, expressed all pluripotency markers tested (Figure 3.5).

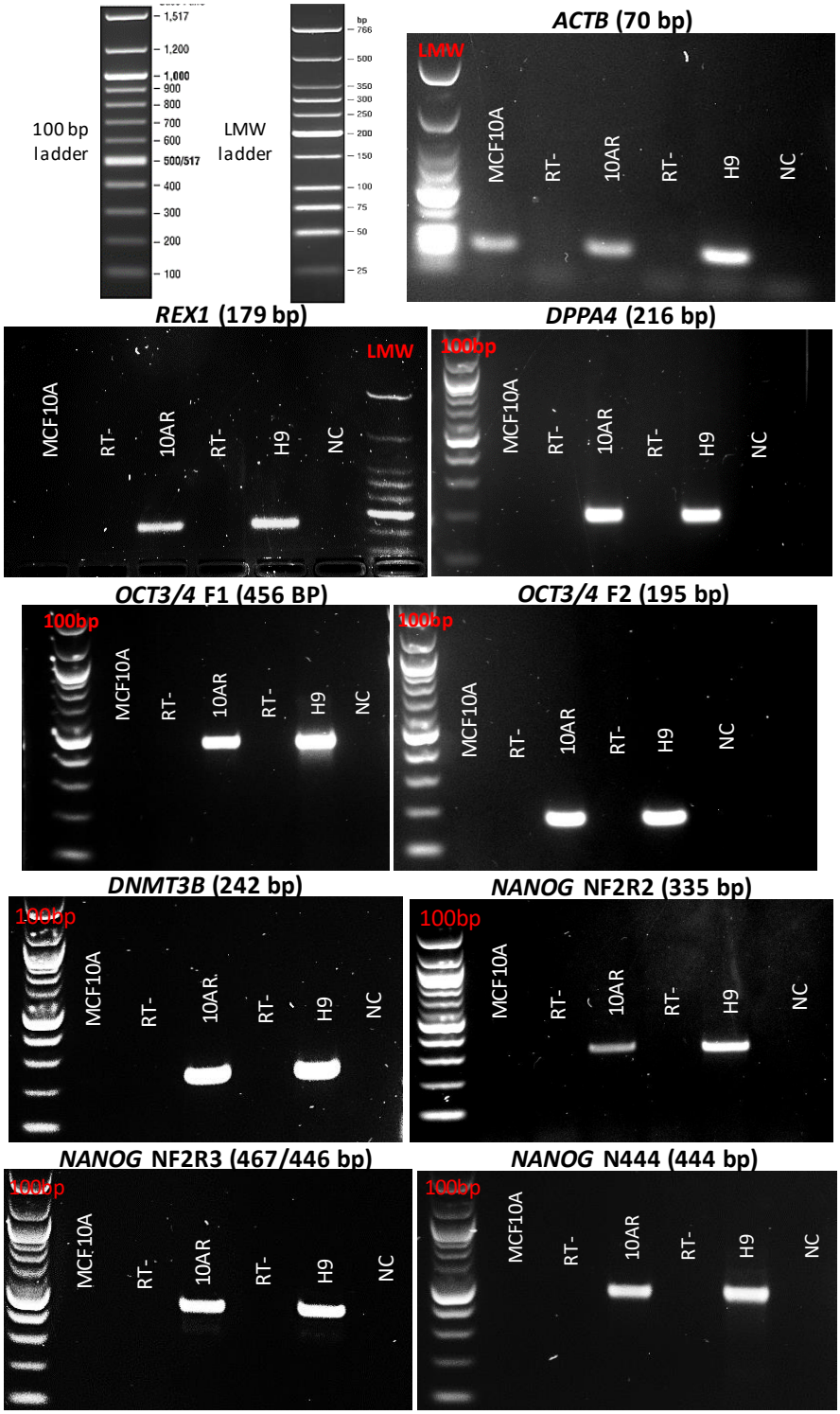


Figure 3.5 Representative RT-PCR of pluripotency markers in 10AR.

10AR expressed all pluripotency markers tested. RT-PCR using NF2R2, N444 and NF2R3 primers indicated 10AR expressed only *NANOG* but not *NANOG P8*. *ACTB* and reverse transcriptase negative (RT-) were included as controls. MCF10A = passage 110 – 117, 10AR = passage 17 – 33, H9 = passage 24 – 30. (n > 3).

Next, real-time PCR (RT-qPCR) was performed to quantify the expression levels of *REX1*, *NANOG*, *DPPA4* and *DNMT3B* in 10AR compared to MCF10A and H9 cells. The expression of *REX1* ($p<0.0005$), *NANOG* ($p<0.05$) (using NF2R2 primer), *DPPA4* ($p<0.005$) and *DNMT3B* ($p<0.005$) was significantly higher in 10AR as compared to MCF10A. However, higher expression of *REX1* ($p<0.0001$) and *NANOG* ($p<0.05$) was observed in H9 cells as compared to 10AR, suggesting a potential discrepancy in term of pluripotency between the two. Nevertheless, results indicated that the expression of pluripotency genes in 10AR was similar to ESCs and significantly different from its non-reprogrammed counterpart MCF10A.

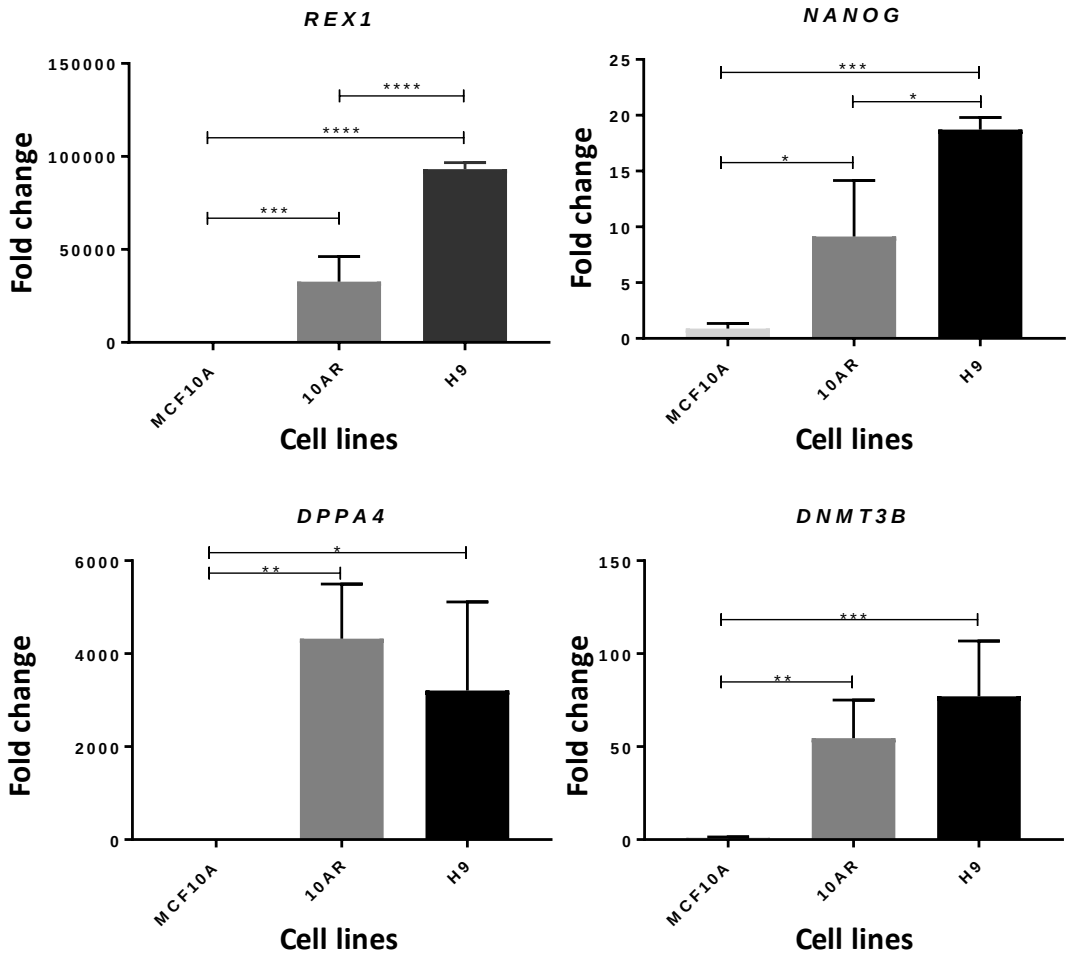


Figure 3.6 Quantification of pluripotency markers expression in 10AR by real-time-PCR (RT-qPCR).

Result were normalized to *ACTB*. Data was analysed by one-way ANOVA followed by Bonferroni correction. * = <0.05, ** = <0.005, *** = <0.0005, **** = < 0.0001. MCF10A = passage 110 – 117, 10AR = passage 17 – 33, H9 = passage 24 – 30. MCF10A, n = 5; 10AR, n = 5; H9, n = 4.

Next, five pluripotency protein markers (TRA-1-81, TRA-1-60, SSEA4, NANOG and OCT3/4) commonly used for PSCs characterisation were also analysed by immunofluorescence staining (Andrews et al., 1984, Marti et al., 2013) (Figure 3.7). Negative controls consisting of 10AR stained by secondary antibodies only were also set up in parallel (Appendix Figure S3-3). In addition, FACS analysis for TRA-1-81, TRA-1-60 and SSEA4 was also performed (Figure 3.8). Results indicated that 10AR, but not MCF10A, was positively stained for all pluripotency markers. Interestingly, consistent results were obtained from immunofluorescence staining and FACS analysis that indicated a relatively lower percentage of TRA-1-81^{+} cells in 10AR as compared to other pluripotency surface markers, such as TRA-1-60 and SSEA-4.

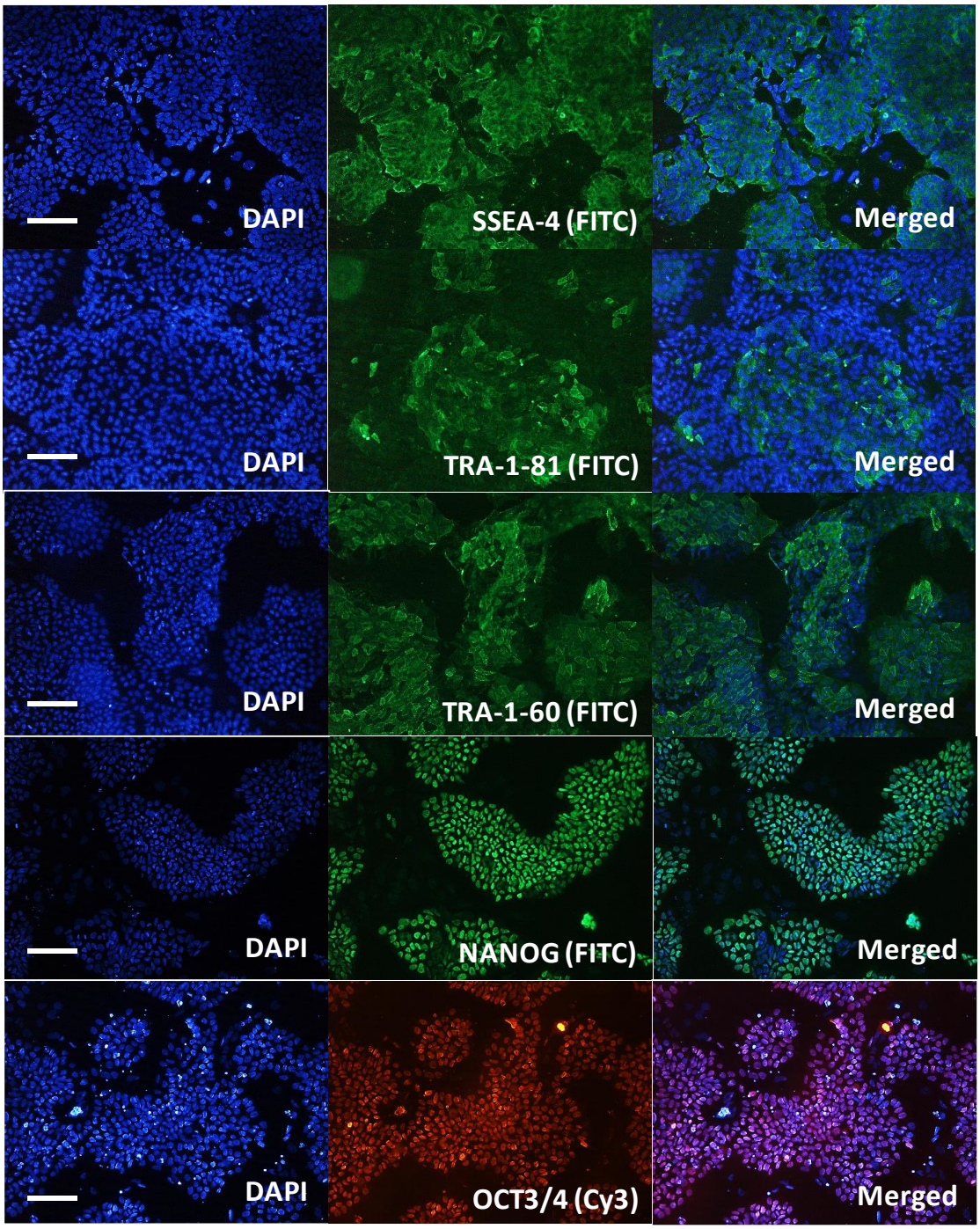


Figure 3.7 Immunofluorescence staining of pluripotency markers in 10AR.

10AR cells were positively stained for all pluripotency markers tested (SSEA-4, TRA-1-81, TRA-1-60, NANOG and OCT3/4). Scale bar = 100 μm. Cell analysed was passage 22. Secondary antibody-only negative control is shown in Figure S3-3. n = 1.

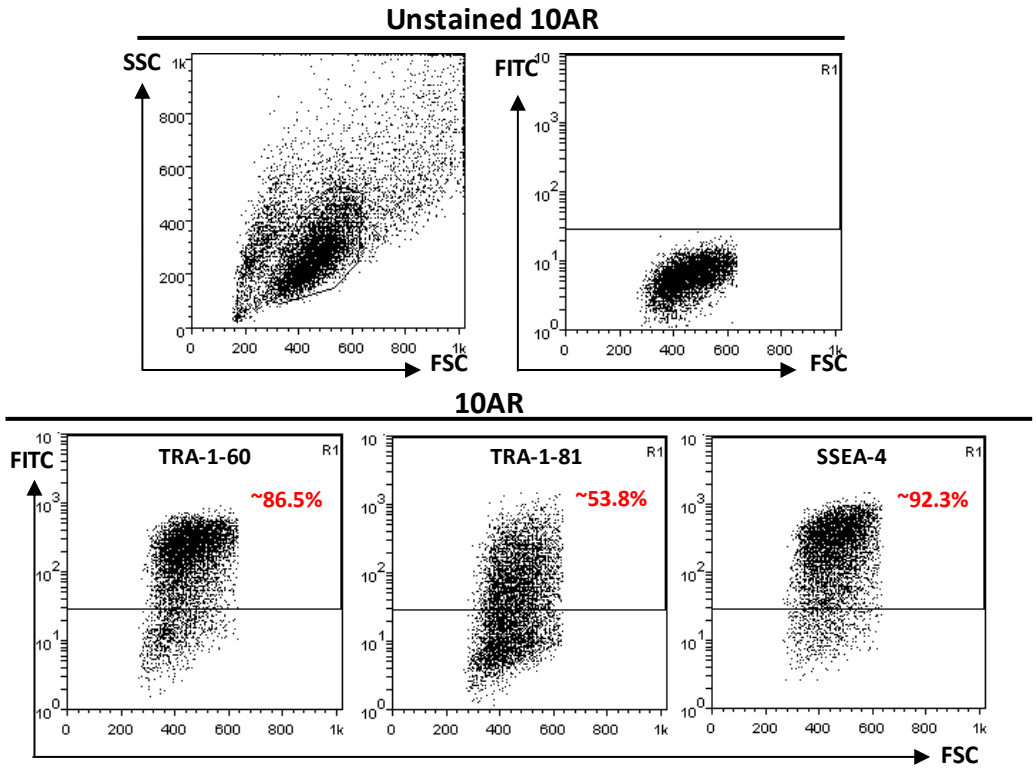


Figure 3.8 FACS analysis of pluripotency markers in 10AR.

The 10AR cells were positively stained for TRA-1-60, TRA-1-81 and SSEA-4. A relatively lower percentage of cells were TRA-1-81 positive as compared to other markers. n = 1. SSC = side scatter, FSC = forward scatter. Cell was passage 25.

3.3 DISCUSSION

Results in this chapter demonstrated the successful derivation of iPSCs from a non-malignant immortalised mammary cell line MCF10A through transfection of a combination of four episomal vectors encoding *OCT3/4*, shRNA for *TP53*, *SOX2, KLF4, LIN28, L-MYC* and *EBNA1* by using the XG9 transfection reagent. A seeding density optimisation protocol as well as a series of analyses for PSC characterisation, including RT-PCR, RT-qPCR and immunofluorescence staining for pluripotency markers expression had also been established by using the MCF10A-derived iPSC line (10AR). 10AR was shown to have acquired ESC-like characteristics morphologically and

molecularly, suggesting mammary cells are amendable to episomal-mediated iPSC reprogramming.

The present study was not the first attempt to reprogram MCF10A cell line. Nishi et al. (2014b) has succeeded in deriving iPSC lines from MCF10A using the classical retrovirus approach. A total of eleven AP^+ colonies (out of 6 x 10^5 cells per 100 mm dish) were harvested (reprogramming efficiency was ~0.002%) and eventually four clones were randomly selected for further analyses. Partial differentiation of MCF10A-iPSC generated cancer stem cell (CSC)-like (iCSCL-10A) cells which exhibited malignant properties of tumour-derived CSCs. In addition, the iCSCL-10A cells acquired a $CD44^+/CD24^{low}$ CSC marker profile, expressed *ALDH1A1* and *SOX2* genes (previously reported in breast CSCs) (Ginestier et al., 2007, Khoury et al., 2012), became resistant to anticancer chemotherapeutic compounds (Taxol and Actinomycin D) and sensitive to CSC-specific drugs Salinomycin and Juglone (Nishi et al., 2014b).

It was suggested that the transformation of MCF10A-iPSC to iCSCL-10A facilitated by the pre-exiting loss of $p16^{INK4a}$ gene caused by the t(3;9)(p14;p21) reciprocal translocation, a mutation allegedly conferred MCF10A cell line its immortality. The $p16^{INK4a}$ is a major component of the tumour suppressor pathway $p16^{INK4a}$/Rb. Its downregulation has been shown to enhance reprogramming efficiency (Li et al., 2009). $P16^{INK4a}$ is also responsible for the stability of p53, whose downregulation is beneficial for iPSC reprogramming (see Section 3.3.3 for detail). Nishi et al. (2014b) described iCSCL-10A as a precious model for CSC study and could facilitate

the development of specific treatment targeting CSCs. Its amenability to reprogramming indicated MCF10A is the pertinent cell line to deploy to optimise the episomal vector reprogramming condition as well as an appropriate control for this study. However, the inclination of malignant transformation of MCF10A-derived iPSCs and its differentiated derivatives should be closely monitored.

3.3.1 Optimum cell density for reprogramming

The gradual loss of GFP^+ in MCF10A cells overtime after initial transfection of the episomal vector pCXLE-EGFP indicated transient presence of episomal vector (Yu et al., 2009, Okita et al., 2011). Approximately 48 h (where GFP signal was the highest) was required for the episomal vector to generate GFP protein. The overall low transfection efficiency at D1 can be explained by the high confluence of cells which formed compact colonies and potentially impeded the accessibility of the transfection reagent-plasmid complexes to cells located within the colonies. In addition, it is well-known that high cell density causes growth inhibition and slowly dividing cells are shown to exhibit poor uptake of foreign materials during transfection (Kunath et al., 2003, Dalby et al., 2004). However, the persistence of GFP signal in cells at D1 at 120 h after transfection was likely due to slow cell division caused by highly confluent culture, which in turn resulted in a lower rate of episomal vector loss. In addition, seeding density D1 resulted in high cell confluency (~90% at 96 h after transfection), which could potentially restrict the emergence and expansion of iPSCs impeding the isolation and clonal expansion of iPSCs (Zhao et al., 2015).

Zhao et al. (2015) suggested transfection can potentially attenuate cancer cells proliferation and expansion rates. Considering the lengthiness of culturing time during iPSC reprogramming (30 days generally), it is highly possible that a subset of non-transfected population can dominate the culture over time, obstructing the emergence and expansion of iPSC colonies and potentially initiating cell layer lifting. The same study concluded that seeding cell density was the most crucial factor for successful iPSC derivation from cancer cells among other criteria, including time (number and timing) of transfection and the transfection reagent concentration ratio. In addition, the authors also emphasised proliferation rate discrepancies, warranting the need for optimisation to identify the optimum seeding density for different cancer cell lines (Zhao et al., 2015).

Cell death inflicted by toxicity of XG9 had dramatic effect on cells seeded at D3, leaving the culture with sparsely attached cells which could compromise the reprogramming process as well as viability and healthy growth of cells due to loss of cell-cell interaction (Kunath et al., 2003). As four consecutive transfections and transfection of multiple plasmids are required for sustained ectopic expression of transcription factors (TFs), it was decided that D3 was unable to generate a suitable cell number for iPSC derivation. Therefore, D2 was eventually selected to proceed with the reprogramming of MCF10A. Overall, the significance of this optimised experimental workflow as well as selection criteria for the identification of optimum seeding density has been translated into successful iPSC derivation from MCF10A.

3.3.2 Loss of episomal vectors

At present, the episomal vector reprogramming method is the second (to Sendai virus) most widely used non-integrating reprogramming approach (De Sousa et al., 2017). It has been widely demonstrated that random integration and subsequent ectopic expression of the reprogramming TFs can result in insertional mutagenesis and carcinogenesis (Okita et al., 2007b, Yu et al., 2007a). A mutation-teeming genome harbouring various genetic mutations ranging from chromosomal to single base aberrations is one characteristic of cancer (Stratton et al., 2009, Vogelstein et al., 2013, Sima and Gilbert, 2014). Therefore, it is conceivable that, in the context of a cancer genome, the integration of TFs could induce further genome alternations and result in unpredictable genomic and phenotypic effects, complicating downstream analyses. Therefore, a non-integrating or possibly an excisable approach (Kotini et al., 2015, Kotini et al., 2017) was the best option for cancer cell reprogramming (De Sousa et al., 2017).

In this study, the loss of episomal vectors with time was demonstrated by the gradual decrease of GFP signal as indicated by FACS analysis (Figure 3.1). Besides, all episomal vectors were lost in 10AR as early as passage 18, consistent with Okita et al. (2011) and Yu et al. (2009), both reported the complete loss of episomal traces at passage ranging from 8 – 20 and 11 – 26, respectively. However, it is worth mentioning that rare incidences of genome integration have been reported in many studies. Three of the pioneer reports have identified incidences of episomal integration: (1) Okita et al. (2010) reported 1/3 of the ESC-like colonies with possible transgene integration; (2) Okita et al. (2011) identified 1/7

iPSC clones having plasmid integration into the genome; (3) and Yu et al. (2009) reported 2/3 of the subclones of one iPSC line with possible genome integration. In addition, Schlaeger et al. (2015) also reported the persistence of episomal vectors at passage 9 – 11 in approximately 40% of iPSC lines. In the present study, only one out of five iPSC lines was subjected to characterization. Therefore, it was not possible to calculate the integration rate. Nevertheless, it is advisable that in future experiment, more iPSC colonies should be analysed to increase the possibility of isolating transgene-free iPSC lines.

3.3.3 Reprogramming efficiency

Despite harbouring mutations (most notably in tumour suppressor genes such as *TP53*) that have been proven beneficial for iPSC reprogramming and expressing pluripotency genes that might or might not be part of the reprogramming cocktail (thought to enhance reprogramming efficiency), cancer cells are found, in majority of the studies, more refractory to iPSC reprogramming as compared to their normal counterparts (Stadtfeld and Hochedlinger, 2010, Gonzalez et al., 2011, Ramos-Mejia et al., 2012).

Low efficiency and slow emergence of iPSC colonies are major drawbacks of the episomal vector reprogramming method (Schlaeger et al., 2015). Modifications have been made since its first use almost a decade ago to circumvent these limitations. Modifications include reducing the number of plasmids (to less than two) carrying all essential TFs to increase the number of cells taking up and expressing all the transgenes, the addition of extra plasmids that encode extra factors known to enhance reprogramming

efficiency (such as *EBNA1*, *SV40TL* and shRNA for *TP53*) and the addition of epigenetic modifiers (such as sodium butyrate) or small molecules (such as GSK-3 inhibitor CHIR99021) at different time points of reprogramming. For comparison, Table 3.1 summarises the efficiency in different studies that used eipsomal vector method for iPSC reprogramming. However, it is worth to bear in mind that all studies in Table 3.1 used different methods to calculate reprogramming efficiency and at different time points during the process.

Table 3.1 The reprogramming efficiency of episomal vector approach.

Factors	Efficiency	Cell type	Reference
OSKML on three plasmids	0.013%	HDF-f fibroblasts, NuFF and BJ fibroblasts	Schlaeger et al. (2015)
OSKMLN and SV40LT on three plasmids	0.0006%	Human new-born foreskin fibroblasts (CRL-2097)	Yu et al. (2009)
OSKM on one plasmid, BM stem cell co-culture system and supplementation of hematopoietic growth factors	1. 13.0 – 20.4% 2. 50%	1. Mature Lineage-committed Myeloid Progenitors (CD34+CD38+) 2. Episome-transfected CD45+CD34- lineage-committed CB cells	Park et al. (2012)
OSKMLN and *SV40LT* on 3 plasmids	0.035%	BM mononuclear cells (MNCs)	Hu et al. (2011)
OSKMLN and *SV40LT* on 3 plasmids	0.4 – 3.6%	CD34+ CB cells	Burridge et al. (2011)
ON on two plasmids	0.1 – 1%	Human fetal neural stem cells	Marchetto et al. (2009)
OSKML on one plasmid with NaB	0.05%	Newborn CB MNCs	Chou et al. (2011)
OSKML shRNA for *TP53* and EBNA1	0.06%	CD34+ CB cells and from PB	Okita et al. (2013)
OSN-fused with VP16 (synthetic factor) and K on one plasmid	0.13%	Human fetal fibroblasts	Wang et al. (2011)
OSKMLN and *SV40LT* with PD0325901, CHIR99021, A-83-01, ROCK inhibitor HA-100 and human LIF	0.1%	Human skin fibroblasts and post-expansion CB cells	Yu et al. (2011b)
OSNL *SV40LT* PD0325901, CHIR99021,	0.12%	Human adult adipose tissue-derived cells	Yu et al. (2011b)

A-83-01, ROCK inhibitor HA-100 and human LIF			
OSKMLN and SV40LT on three plasmids, mRNA for *EBNA1* and NaB	0.02%	Primary fibroblast cells	Chen et al. (2011a)
OS. Improved oriP/EBNA1-based episomal vectors using a SFFV LTR promoter	1-2%	CB CD34+ cells	Meng et al. (2012)
Improved oriP/EBNA1-based episomal vectors using a SFFV LTR promoter. 1. OS + BCL-XL (in lentiviral vector) 2. OSKM + BCL-XL 3. OSKM + BCL-XL 4. OSKM + BCL-XL. Cultured for 4–6 days under conditions that favour non-lymphoid cell expansion	1. 3-6% 2. 10 colonies in 1 mL PB 3. 10-20 colonies in 1 mL PB 4. 18-35 colonies in 1 mL PB	1. CB CD34+ cells 2. PB MNCs 3. T/B-depleted cells PBMNCs (CD3−/19− cells) 4. T/B cell depleted PBMNCs	Su et al. (2013)
OSKMLN on 3 plasmids	1. 0.04 – 0.1% 2. 0.0017%	1. BM MNCs 2. CB MNCs	Hu et al. (2011)
1. OSKMLN and SV40LT on three plasmids 2. OS *SV40LT* or OSN *SV40LT*	1. 1.8 – 41.2% 2. 0.5% - 4.0%	1. Fibroblast cells 2. Fibroblast cells	Valamehr et al. (2014)

* Abbreviations: O = *OCT3/4*, S = *SOX2*, K = *KLF4*, M = *C-MYC*, L = *LIN28*, N = *NANOG*, BM = bone marrow, MNCs = mononuclear cells, CB = cord blood, NaB = sodium butyrate, PB = peripheral blood, LIF = leukaemia inhibitory factor, SFFV = strong spleen focus forming virus, LTR = long terminal repeat.

The inhibition of tumour suppressors such as p53, $p16^{INK4a}$/Rb, $p19^{ARF}$ or $p21^{CIP1}$ (phenomena widely observed in various human cancers) has been shown to significantly enhance reprogramming efficiency (Hong et al., 2009a, Kawamura et al., 2009, Krizhanovsky and Lowe, 2009, Banito and Gil, 2010, Kareta et al., 2015). *TP53* is a tumour suppressor gene responsible for the maintenance of tissue homeostasis (Fridman and Lowe, 2003). In normal circumstances, p53 protein is activated via post-translational modifications (Kruse and Gu, 2008, Krizhanovsky and Lowe, 2009) by intrinsic and extrinsic stress signals that could potentially disrupt genome stability/integrity, cell division and DNA replication. The activation of p53 protein and its network of genes initiate various positive- and

negative-autoregulatory feedback loops in concert with other signal transduction pathways eventually leading to cell cycle arrest, cellular senescence or apoptosis (Krizhanovsky and Lowe, 2009).

In cancer, the inactivation of p53 by mutation compromises this regulatory mechanism and prevents the elimination of cells that sustain DNA damage and cellular stresses, eventually leading to uncontrolled growth and carcinogenesis (Brooks and Gu, 2010). Studies indicated that the p53 pathway is activated during reprogramming potentially by DNA damage signal induced by the OSKM transgenes and this represents a barrier by restricting cell proliferation (via the induction of p21) and promoting apoptosis (Kawamura et al., 2009). The downregulation of *TP53*, its target gene *p21* and genes responsible for its stability ($p19^{Arf}$ and $p16^{Ink4a}$) as well as the prevention of the degradation of its negative regulators, Mdm2 and Mdmx (Wade et al., 2013) enhance reprogramming efficiency. Furthermore, Kawamura et al. (2009) demonstrated that p53 knockdown can replace *KLF4* and *C-MYC* in the OSKM cocktail to generate germline-competent iPSCs.

Different approaches have been used to determine the iPSC reprogramming efficiency. For example, efficiency was identified as the percentage of SSEA-4, TRA-1-81 and CD30 triple positive cells (analysed by FACS) during reprogramming (from day 16 – 22) by Valamehr et al. (2014). On the other hand, Schlaeger et al. (2015) determined reprogramming efficiency through immunological staining with an anti-TRA-1-60 kit of cells at approximately day 16. In addition, Okita et al. (2007b) calculated the

efficiency by dividing the number of GFP^+ colonies (by *Nanog* reporter) by the cells initially seeded. None of the above techniques were used in this study. Instead, the efficiency of MCF10A reprogramming (~0.0005%) were expressed as the number of iPSC lines isolated and expanded divided by the number of starting cells. Therefore, this figure had definitely underestimated the real efficiency as numerous small and unhealthy iPSC-like colonies that could potentially give rise to iPSC lines were observed but not expanded during reprogramming. Nevertheless, the reprogramming efficiency of MCF10A was similar to figure reported by Yu et al. (2009) but was ~4 times and ~120 times less than that of Nishi et al. (2014b) (the MCF10A reprogramming study) and (Okita et al., 2011) (the study where the reprogramming cocktail used in this study was derived) respectively.

3.3.3.1 The E8 and E6 media system

A xeno-free and chemically-defined medium system (Essential 8 and Essential 6 media) was used in this study for iPSC reprogramming and the propagation of stem cells (Chen et al., 2011a). Culture medium plays a vital role in mimicking the *in vivo* growth conditions to maintain the identity and support the growth of specific cell types (Kim and Hayward, 2012, Kim et al., 2014). However, owing to the complexity of *in vivo* biochemical (such as endocrine and paracrine signalling) and biophysical conditions (cell-cell interaction, spatial organization and cell morphology), the development of suitable culture media can be extremely challenging (Kim and Hayward, 2012, Kim et al., 2014).

Despite being derived from the inner cell mass (ICM) of preimplantation embryos, the human embryonic stem cells (hESCs), when cultured in the traditional hESCs medium, exhibit a primed PSC state instead of a naïve state that resembles the ICM. The primed state hESCs shows limited pluripotency *in vivo* and fail to give rise to germline chimeras (Huang et al., 2012). It has been suggested that the hESCs continue to differentiate after being derived from the ICM and acquire stability at the primed state due to the dissimilarity of the *in vitro* to *in vivo* environment (O'Leary et al., 2012, Wu and Izpisua Belmonte, 2015). Recent efforts have contributed to the development of various culture conditions that promote and maintain the naïve pluripotency in hESCs, which include the introduction of various inhibitors and growth factors (Huang et al., 2014, Theunissen et al., 2014).

The traditional hESC medium requires the use of feeder layer and knockout serum replacement (KOSR). The KOSR is a chemically-defined substitute for serum developed by Price et al., (1998), however, the exact formulation has not been disclosed. It has been demonstrated that KOSR exhibit significant lot-to-lot variability (Chaudhry et al., 2008, Tamm et al., 2013) and more differentiated cells has been observed around hESC colonies cultured with KOSR, indicating its limitation in maintaining consistency and pluripotency (Li et al., 2005). Feeder-layer is a monolayer of growth-arrested but viable cells (Xu et al., 2001b, Llames et al., 2015) used to provide substratum for the attachment of hESCs in culture and generate a variety of growth factors essential for the maintenance of pluripotency and stem cell survival. However, like KOSR, feeder cells exhibit remarkable batch variability (Tamm et al., 2013). Moreover, the presence

of feeder cells, generally a xenograft (derived from other animals) or allograft (from the same organism) in culture might compromise the experimental outcomes due to the production of undefined biological products (Lu et al., 2006). In addition, it also limits the clinical applications of hESCs, such as transplantation, by potentially eliciting immunological response or contaminating hESCs with pathogens carried by the feeder cells (Xu et al., 2001a, Lu et al., 2006, Chin et al., 2010, Zou et al., 2016).

Chan et al. (2009) suggested the use of a feeder-free and chemically-defined medium system favoured the generation of *bona fide* iPSCs, an observation also supported by subsequent studies (Chen et al., 2011a, Beers et al., 2012). In addition, it has also been widely proven that E8 medium specifically improves the efficiency of episomal vector reprogramming (Chen et al., 2010a, Chen et al., 2014). Therefore, E6/E8 system used in this study provides various benefits for reprogramming and stem cell propagation as well as circumvent caveats inherent to the traditional feeder layer- and KOSR-dependent culture system (Badenes et al., 2016).

3.3.4 Pluripotency of 10AR

3.3.4.1 The expression of pluripotency markers

Human pluripotent stem cells (hPSCs) are characterized by indefinite proliferation, self-renewal and pluripotency and differentiation potential (Thomson et al., 1998, Yu and Thomson, 2008, De Los Angeles et al., 2016). Like all different cell types in the human body, hPSCs have a distinct transcriptional profile that maintains its identity (Au and Sebastiano, 2014).

Molecularly, numerous pluripotency markers exclusively expressed in PSCs have been reported and widely used to isolate, characterize and distinguish hPSCs from differentiated cells. (Kim et al., 2008a, Orkin et al., 2008, Theunissen et al., 2011). The *REX1* gene, also known as zinc finger protein-42 (*ZFP42*) was first found in mouse F9 teratocarcinoma cells and subsequently in other mouse and human pluripotent and multipotent stem cells (Hosler et al., 1989, Rogers et al., 1991, Ramalho-Santos et al., 2002, Brivanlou et al., 2003, Abeyta et al., 2004, Baal et al., 2004). Recent studies demonstrated that *REX1* marks a specific subpopulation of the most pluripotent hESCs and can be used as a marker to distinguish fully reprogrammed from partially reprogrammed iPSCs (Chan et al., 2009).

The human *NANOG* gene encodes a DNA binding ANTP class homeobox transcription factor (Booth and Holland, 2004). *NANOG* plays a centre role in self-renewal and pluripotency maintenance of hESCs. However, the regulation of *NANOG* expression is not fully understood. *OCT4* and *SOX2* promote *NANOG* expression through binding to *NANOG* promoter (Rodda et al., 2005, Pan and Thomson, 2007). In addition, other transcription factors such as FoxD3 and p53 and signalling pathways (such as BMP and Wnt) also contribute to the maintenance of pluripotency and self-renewal through regulation of pluripotency genes including *NANOG* (Pan and Thomson, 2007). Studies have been focusing on the cooperation of *OCT4, SOX2* and *NANOG,* which activates a core transcriptional regulatory circuitry made up of autoregulatory and feedforward loops to regulate pluripotency, self-renewal and inhibition of cellular differentiation (Boyer et al., 2005). It has been shown that *OCT3/4, SOX2* and *NANOG* share a significant number of target

genes. For instance, Player et al. (2006) reported the co-occupancy of the three transcription factors on *SOX2, OCT3/4, TDGF1, GJA1, SET* and *DPPA4* genes. Besides, Boyer et al. (2005) discovered more than 350 target genes shared between *OCT4, SOX2* and *NANOG*. However, while expression of *NANOG* facilitates pluripotency, high level of *OCT3/4* initiates differentiation (Rodda et al., 2005, Jung et al., 2010, Zeineddine et al., 2014). Therefore, *NANOG* is deemed to be a more stringent marker for pluripotency than *OCT3/4* and *SOX2*.

The detection of *NANOG* expression at either transcriptional or translational levels can be problematic owing to the presence of eleven pseudogenes, ten processed (*NANOG P2 – 11*) and one duplication (*NANOG P1*), in the human genome (Booth and Holland, 2004). Out of the eleven pseudogenes, only *NANOG P8* encodes an active protein functionally similar to *NANOG* (Booth and Holland, 2004, Fairbanks and Maughan, 2006). *NANOG P8* contains a complete open reading frame with one amino acid mismatch (in the C-terminal transactivation domain), and an *Alu* element in the 3'UTR resembling *NANOG* (Booth and Holland, 2004, Zhang et al., 2006, Zbinden et al., 2010) (Appendix Figure S3-1). The expression of *NANOG P8* has been reported in different tumours such as the breast (Ezeh et al., 2005, Alldridge et al., 2008), brain (Zbinden et al., 2010), kidney (Bussolati et al., 2008), ovary (Zhang et al., 2008b) and cervix (Ye et al., 2008). Despite the fact that few cancer stem cells, such as glioma stem cells, express both *NANOG* and *NANOG P8*, only *NANOG* has been detected in hESCs. This indicates that *NANOG* but not *NANOG P8* is an important marker for ESC-like pluripotency (Ambady et al., 2010).

Kohler et al. (2011), Ling et al. (2012) and Han et al. (2012) reported the detection of *NANOG* expression by RT-PCR and/or immuno-chemistry/fluorescence staining in MCF7. However, the PCR primers used in the three studies were not able to differentiate *NANOG* from *NANOG P8* based on the University of California Santa Cruz Genome Browser in-silico PCR result (Appendix Figure S3-2) (Karolchik et al., 2012). Moreover, despite the availability of anti-NANOG and anti-NANOG P8 antibodies, it is unknown if the former (most widely used for NANOG detection) can differentiate NANOG from NANOG P8. As opposed to the above findings, Leis et al. (2012) demonstrated the lack of NANOG expression in MCF7 through immunohistochemistry while Lu et al. (2014) reported a minimal *NANOG* expression in MCF7 by comparing its mRNA expression level with Hues9 (ESC line). Furthermore, by using a universal indiscriminating *NANOG* primer, Jeter et al. (2009) amplified and subsequently analysed the PCR amplicons generated from MCF7 by Sanger sequencing. Results indicated the presence of only *NANOG P8* (the presence of CAG759 → CAC759 mismatch) but not *NANOG* in MCF7 cells. subsequently, Jeter et al. (2009) proposed the use of two PCR primers, the NF2/R2 and NF2/R3 used in this study, to differentiate *NANOG* from *NANOG P8* expression (however, the primers were not tested on MCF7 in that particular study) (Jeter et al., 2009). Similarly, IHC staining indicated that the expression of NANOG protein in MCF7 was rarely nucleus-localised, suggesting possible NANOG P8 but not NANOG expression (Jeter et al., 2009). However, contrasting results in the same study indicated knockdown of *NANOG* with shRNA inhibited MCF7-induced tumour growth *in vivo*. Nevertheless, based on sequence alignment

analysis done in this project (Appendix Figure S3-2), the NF2R2 and NF2R3 were selected for PSC characterisation. Sanger sequencing was performed on amplicons generated using these primers to verify the credibility of the PCR results (Appendix Figure S2-8). In total, three *NANOG* PCR primers were used in this study and they take advantage of the 5'UTR (N444 primer) and 3'UTR (NF2R2 and NF2R3 primers) mismatches between *NANOG* and *NANOG* P8, respectively, to discriminate both genes (Zhang et al., 2006, Jeter et al., 2009). The N444 and NF2R2 primers are specific for *NANOG* while NF2R3 can give rise to two PCR products of different sizes indicating *NANOG* (467 bp) or *NANOG P8* (446 bp) (Appendix Figure S3-2). Consistent with other reprogramming studies, 10AR expressed only *NANOG* but not *NANOG P8,* as shown in Section 3.2.3.1.

In addition, the iPSCs derived from MCF10A were similar to ESCs in various aspects, including morphology, expression level of key pluripotency markers at molecular and protein levels. However, quantitative data as determined by RT-qPCR indicated that the expression levels of *REX1* and *NANOG* in 10AR were significantly lower ($p < 0.0001$ and $p < 0.05$ respectively) than H9. *REX1* and *NANOG* have been shown to be the essential markers of authentic iPSCs and partially reprogrammed cells do not or rarely express these markers (Chan et al., 2009, Plath and Lowry, 2011, Theunissen et al., 2011). Several studies in the mouse demonstrate the vital role of *Nanog* in the reprogramming process. Okita et al. (2007b) demonstrated that *Nanog*-expressing mouse iPSCs exhibited gene expression, DNA methylation profile and germline potential distinct from iPSCs selected for *Fbx15* (despite both capable of forming teratoma) but

similar to mouse ESCs (mESCs). In a different study, *Nanog* has been shown to play a vital role at the final stage of reprogramming. Its overexpression synergized with DNA methylation inhibition promoted the transition of partially reprogrammed iPSCs to naïve pluripotent stem cells in 2i/LIF culture medium (Theunissen et al., 2011). In addition, the overexpression of *Nanog* also enabled the reprogramming of epiblast-derived stem cells to naive pluripotency even under pluripotency unfavourable culture conditions (Theunissen et al., 2011). Lastly, Plath and Lowry (2011) described *Nanog* overexpression as the defined requirement for the generation of *bona fide* iPSCs at the late phase of a proposed reprogramming model. This model is also supported by other similar studies (Polo et al., 2012a, Apostolou et al., 2013, Buganim et al., 2013).

However, in a human context, Chan et al. (2009) identified three types (type I, II and III) of iPSCs based on colony morphology, live immunofluorescence staining (for CD13, SSEA-4, TRA-1-60 and NANOG) and degree of transgene silencing (indicated by the strength of GFP signal). Gene expression and epigenetic profiling revealed type III as *bona fide* iPSCs characterised by having ESC-like morphology (small, compact cells and well-defined colony edges), a $CD13^{-}$/ $SSEA\text{-}4^{+}$/ $TRA\text{-}1\text{-}60^{+}$/ $GFP^{dim/-}$/ $NANOG^{+}$ profile (TRA-1-60 was exclusive to type III, absence of GFP indicated transgene silencing), and selective expression of *ABCG2*, *DNMT3B* and *REX1.* The fact that alkaline phosphatase, SSEA-4, *GDF3*, *hTERT* and NANOG were also detected in type II and/or type I iPSCs suggested that they may not be stringent markers for real iPSCs. Interestingly, type III iPSCs expressed the same level of *NANOG* and *REX1* as compared to the

ESC line H1OGN in the same study (Chan et al., 2009). Considering the discrepancy observed in *NANOG* and *REX1* expressions between 10AR and H9, it is suggested that despite having acquired pluripotent state of true iPSCs (indicated by high *REX1* and *DNMT3B* expressions and TRA-1-60 expression), the 10AR did not achieve an ESC-like pluripotency. However, differentiation potential (an important hallmark of pluripotency) of 10AR should also be taken into consideration and the exact closeness or remoteness of 10AR from ESC-like pluripotent state can only be determined by genome wide analyses such as global gene expression and DNA methylation profiling.

On the other hand, in the present study, immunofluorescence staining indicated 10AR expressed all markers tested but had a relatively lower level of TRA-1-81 as shown in (Figure 3.8) and result was supported by FACS analysis. Lowry et al. (2008) suggested that TRA-1-81^{+} cells identified during reprogramming mark a population that exhibited ESC-like features morphologically and molecularly and will normally result in either partially or faithfully reprogrammed iPSCs at the end of the process. Despite the fact, the authors emphasised there was no sufficient evidence to conclude that TRA-1-81^{-} cells will not result in meaningful iPSC clones (Lowry et al., 2008). This finding was supported by D'Antonio et al. (2017a), who suggested TRA-1-81 is a crucial marker to differentiate high-quality iPSC from inferior ones. Therefore, low level of TRA-1-81 expression in 10AR compared to H9 could also indicate a degree of inferiority in terms of pluripotency as previously discussed. These findings also warranted the

expansion and characterisation of more iPSC clones for identification of the best to proceed with downstream analyses.

3.3.4.2 The retention of MCF10A epigenetic memory and exogenous OCT3/4 expression in 10AR

Early culture of 10AR was heterogeneous, containing a combination of ESC-like and MCF10A-like cells. Homogenous culture only achieved after several passages with EDTA and repeated manual colony passaging. This incidence was likely caused by several factors: 1) contamination by MCF10A or partially-reprogrammed MCF10A cells during manual passaging and; 2) retention of MCF10A epigenetic memory in 10AR and extensive spontaneous differentiation of iPSCs resulted from stress of passaging. During the reprogramming of MCF10A, a control was set up and cultured with the same E8/E6 media system without being transfected with episomal vectors. The control continued to expand after medium was switched to E8 at D0 of reprogramming and attained confluency without signs of growth attenuation. However, cell death was observed when cells were passaged, proliferation halted and cells eventually perished. This observation indicated that the E8 medium, despite not being implicated as selective medium, does not support or favour the growth of non-reprogrammed MCF10A. Therefore, it was very unlikely that the heterogeneous culture of early passage 10AR was due to the contamination of MCF10A, but possibly partially-reprogrammed MCF10A which might be able to survive and propagate in E8 medium during manual passaging.

It has been shown that iPSCs retain epigenetic memory of its cellular origin after the attainment of pluripotency, making newly-derived iPSCs prone to spontaneous differentiation into its original cell type (Nashun et al., 2015, Hore et al., 2016, Ivanov et al., 2016, Zambidis et al., 2016). However, recent findings suggested insignificant differences in DNA methylation (known to be responsible for epigenetic memory) and gene expression level between iPSC lines generated from different cell types of the same individual (Burrows et al., 2016, Kyttala et al., 2016). On the contrary, DNA methylation and gene expression levels as well as differentiation potential and functionality differences were only detected across iPSCs derived from different individuals, potentially resulted from pre-existing genetic variations (Nawy, 2016). Nevertheless, according to previous reports, epigenetic memory of iPSCs can be gradually eliminated by cell passaging (Kim et al., 2010, Polo et al., 2010). In this project, the retention of epigenetic memory coupled with the stress of passaging might have promoted the spontaneous differentiation of 10AR into MCF10A-like cells, thus contributing to the formation of the heterogeneous culture. Differentiation was less profound at higher passage potentially due to the elimination of such memory and establishment of an ESC-like epigenetic and gene expression profile.

In conclusion, in this chapter, an experimental workflow for the determination of optimum seeding cell density for iPSC reprogramming was established. In addition, it has been shown that cells of the mammary lineage can be reprogrammed by using the episomal vector reprogramming method. In addition, iPSC characterisation analyses, including RT-PCR,

immunofluorescence staining and RT-qPCR for pluripotency marker expression, have also been optimised. However, characterisation of 10AR suggested that further optimisation of the reprogramming condition might be required or perhaps more iPSC clones should be screened in order to identify the best clone more closely resemble ESCs for downstream analysis.

4. REPROGRAMMING OF BREAST CANCER CELL LINES

4.1 INTRODUCTION

In this chapter, seeding density of different breast cancer cell lines was optimised by using the experimental workflow established in Chapter 3. The optimum seeding density was used for reprograming under the same condition as previously described for MCF10A. IPSC lines isolated were subjected to characterisation, including RT-PCR, RT-qPCR and immunofluorescence staining for pluripotency marker expression.

As mentioned in Section 1.2.4, epigenetic alterations play an important role in breast cancer and epigenetic studies have provided important insights into the classification, detection, prevention and prognosis of the disease (Rodriguez-Paredes and Esteller, 2011). The reversibility of epigenetic modifications represents interesting targets for breast cancer treatment (Hochedlinger and Jaenisch, 2006, Brower, 2011, Tompkins et al., 2012). As described in Section 1.3, both chemical agents and biological approaches have been employed to suppress and reverse tumorigenicity through the induction of widespread epigenetic reconfiguration in cancer cells. Of all the biological approaches, the induced pluripotent stem cells (iPSC) reprogramming technology is most widely used (Lee et al., 2015). Many studies have reported the successful derivation of iPSCs from a wide range of cancers, including blood, gastrointestinal, lung, liver, breast and brain, to address important questions pertaining cancer research, for example the cell-of-origin of cancers, mechanisms of cancer initiation and progression as well as the discovery and development of

effective anticancer therapies (Kim et al., 2013, Zhang et al., 2013, Iglesias et al., 2017).

The present study aimed to investigate the contribution of epigenetic and genetic factors to breast carcinogenesis. To achieve this goal, the epigenome of breast cancer cells was reset to an ESC-like state through iPSC reprogramming (results presented in this Chapter) followed by the re-differentiation of cancer-derived iPSCs into various lineages and the assessment of tumorigenicity in these differentiated derivatives (results presented in Chapter 6). Three breast cancer cell lines of two different subtypes, namely triple negative/basal (BT-549 and MDA-MB-231) and oestrogen receptor positive/luminal A (ER^+) (MCF7), were reprogrammed. As described in Section 1.2.1.1, the former represents the most aggressive breast cancer subtype and the latter the most prevalent. Therefore, understanding the cause of both breast cancer subtypes can provide significant clinical contribution.

4.2 RESULTS

4.2.1 Optimisation of X-tremeGENE 9 DNA Transfection Reagent (XG9) transfection conditions

To determine the optimum condition for breast cancer cell reprogramming, the transfection efficiency of XG9 was determined by transfecting three breast cancer cell lines MCF7, BT-549 and MDA-MB-231 with the episomal vector pCXLE-EGFP (expressing green fluorescence protein, GFP; Addgene #27082) followed by measurement of the GFP signal at different time points by FACS analysis. Three cell densities (D1 = 1 x 10^5, D2 = 0.5 x 10^5 and D3 = 0.25 x 10^5 cells per well of 6-well plate) were used to determine the optimum cell density for iPSC generation.

The highest GFP signal was recorded at 96 h after transfection in all three breast cancer cell lines (Figure 4.1 - Figure 4.6). Both BT-549 and MCF7 had similar efficiency (the highest at ~60.6% and ~59.5% respectively) and were higher than MDA-MB-231 (~14.8%). The GFP signal diminished after the highest point in all three cell lines. Five time points were analysed for MCF7, namely 24 h, 48 h, 72 h, 96 h and 120 h. The highest GFP signal was observed at 96 h for all three densities (Figure 4.1). After initial transfection, cells at D1 and D2, but not D3, recovered over the course of time and resulted in suitable number of cells at 120 h (Figure 4.2). Therefore, any cell densities within the range of 0.5 – 1.0 x 10^5 would be optimum for reprogramming.

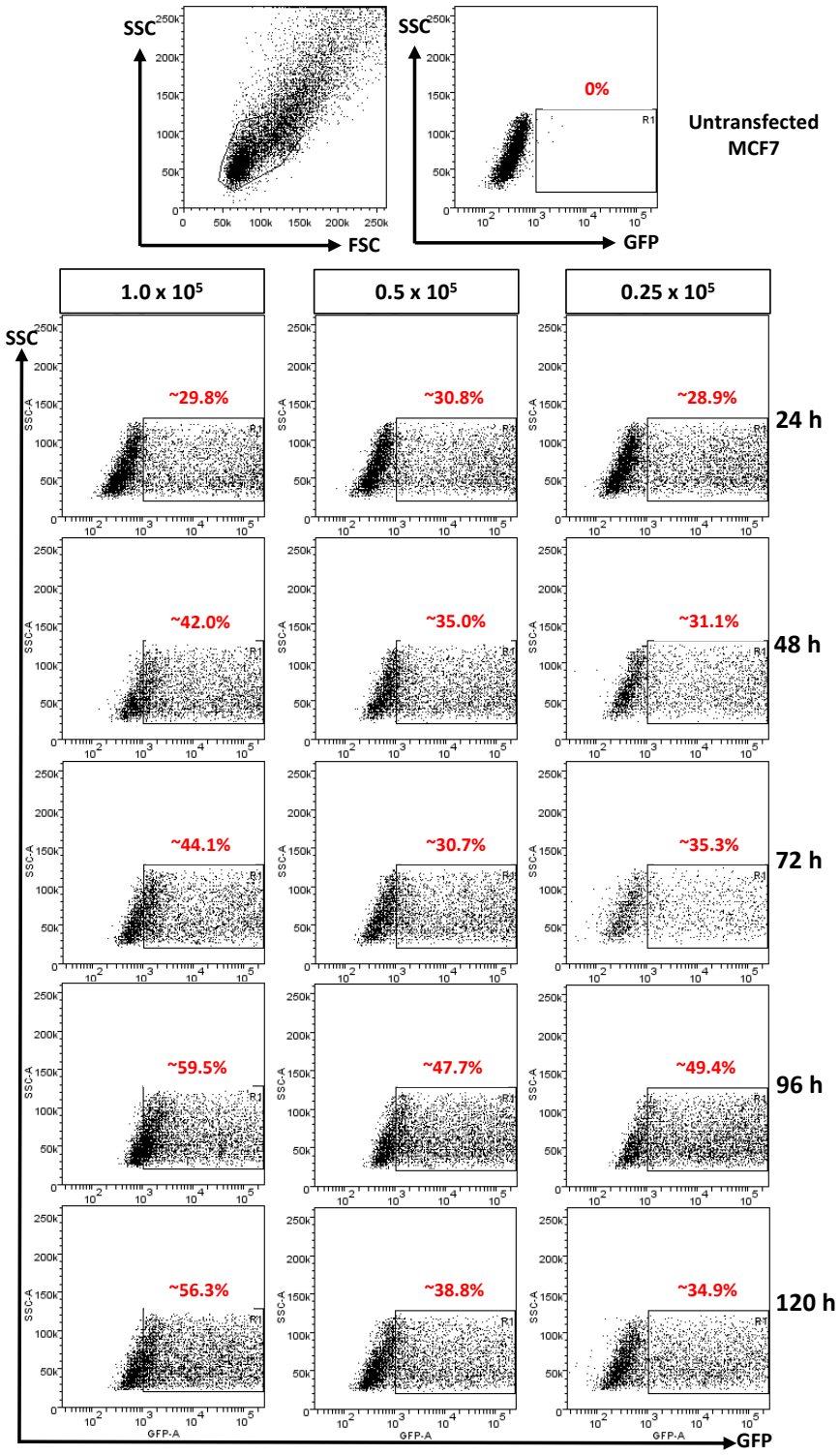


Figure 4.1 Optimization of MCF7 cell density with X-tremeGENE™ 9 DNA transfection reagent for reprogramming.

Untransfected MCF7 was used as negative control. Transfection efficiencies of three cell densities were assessed based on GFP signal. Highest GFP signal was detected in all three groups at 96 h. Signal diminished afterwards indicating loss of episomal vector. n = 1. SSC = side scatter.

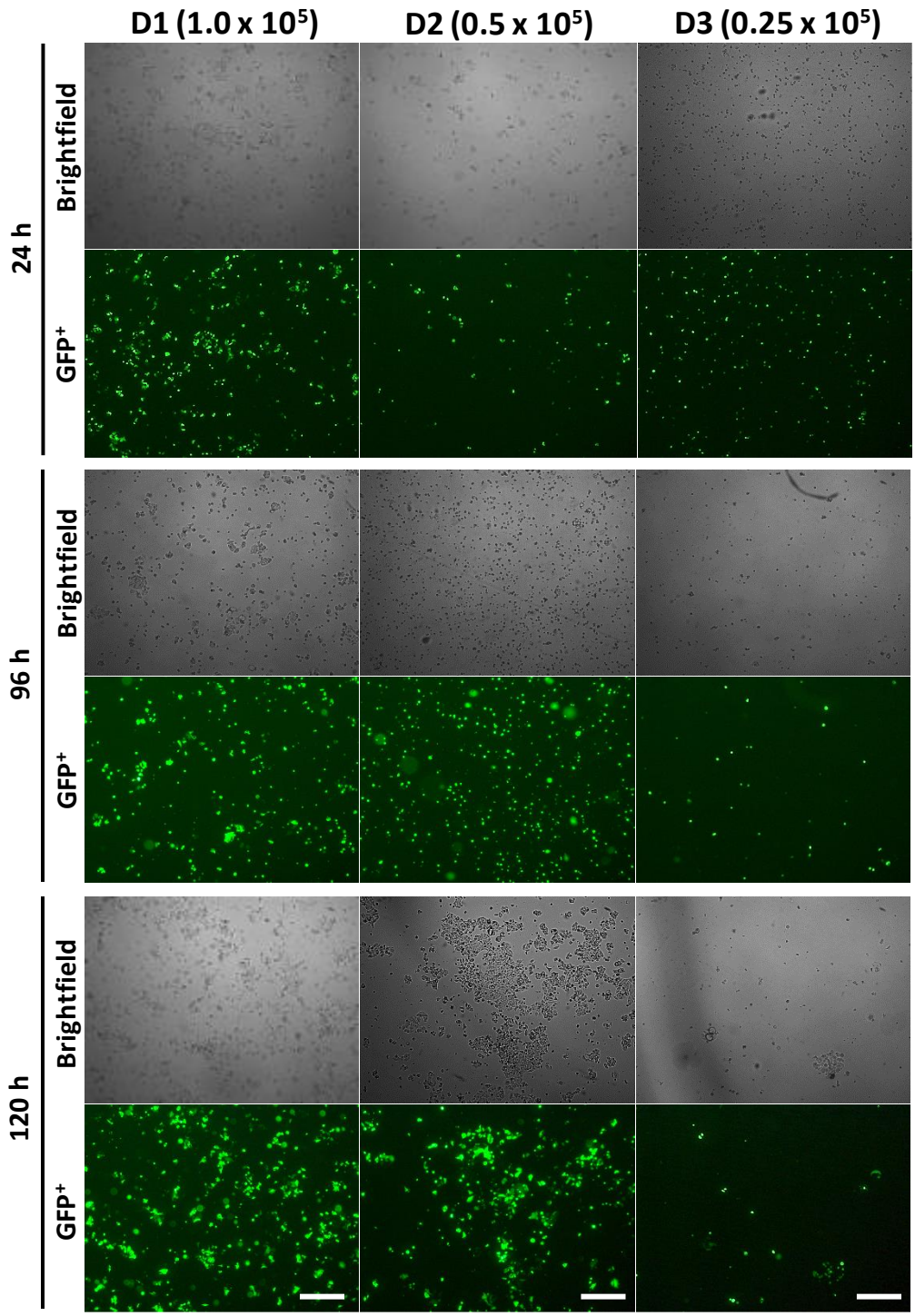


Figure 4.2 Representative images of GFP⁺ MCF7 cells seeded at different densities and different time points after transfection. Cells at D1 and D2, but not D3, recovered well over the course of time and resulted in suitable number of cells at 120 h. n = 1. Scale bar = 200 μm.

Only three time points were analysed for BT-549 and MDA-MB-231. The BT-549 cell line achieved its highest 96 h in all three densities (Figure 4.3). D1 and D2 resulted in highly confluent culture after 96 h, a condition unfavourable for iPSC emergence and isolation (Figure 4.4). Therefore, only 0.25 x 10^5 or a slightly lower seeding density was suitable for BT-549 reprogramming. Lastly, MDA-MB-231 was the least amendable breast cancer cell line to XG9 transfection. The highest GFP signal was recorded at 96 h for D1 and D2 (Figure 4.5). A density between D2 and D3 was deemed optimum for reprogramming based on culture confluence shown in Figure 4.6.

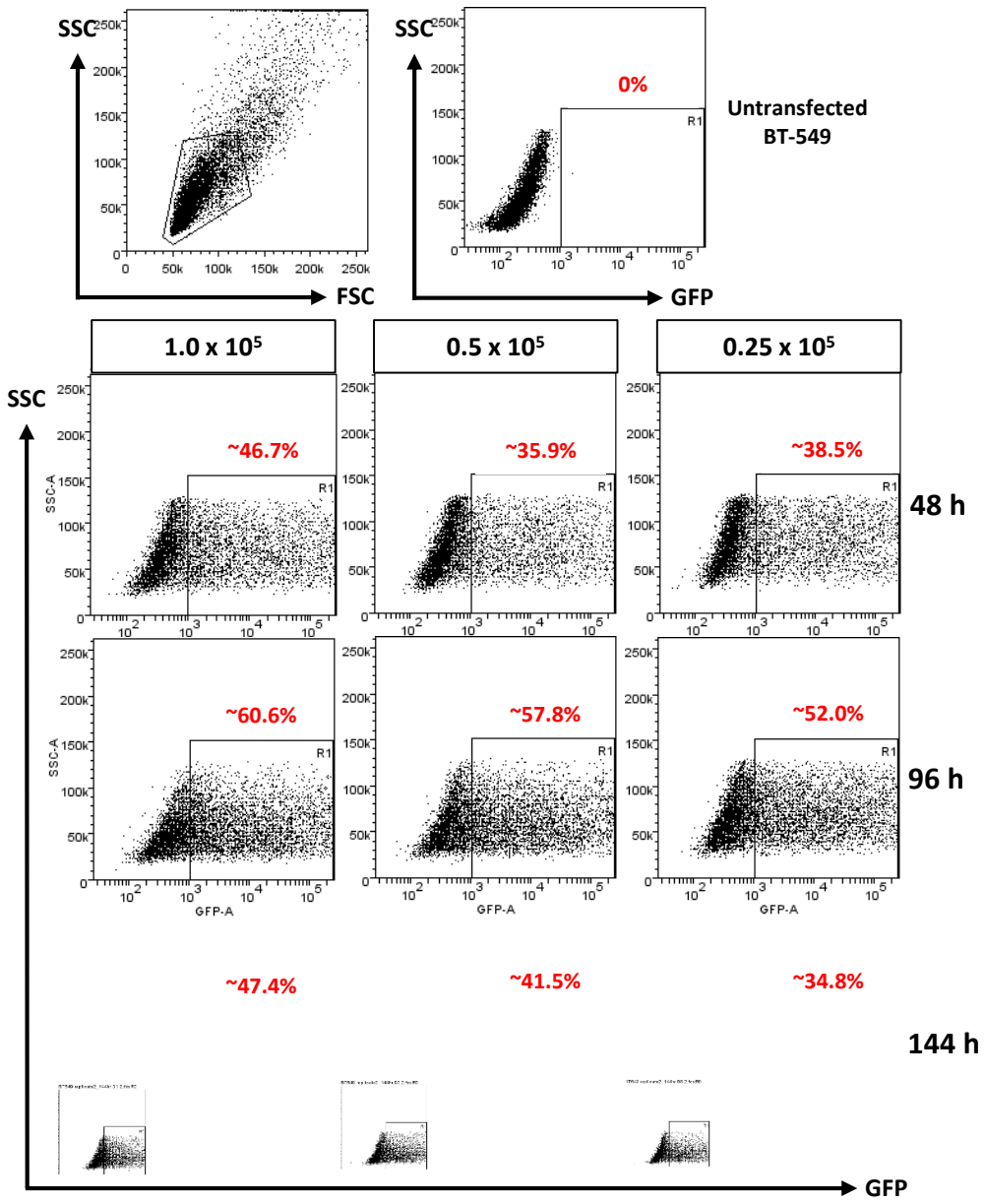


Figure 4.3 Optimization of BT-549 cell density with X-tremeGENE™ 9 DNA transfection reagent for reprogramming.

Untransfected BT-549 was used as negative control. Transfection efficiencies of three cell densities were assessed based on GFP signal. Highest GFP signal recorded at 96 h. Signal diminished over time indicating loss of episomal vector. n = 1. SSC = side scatter.

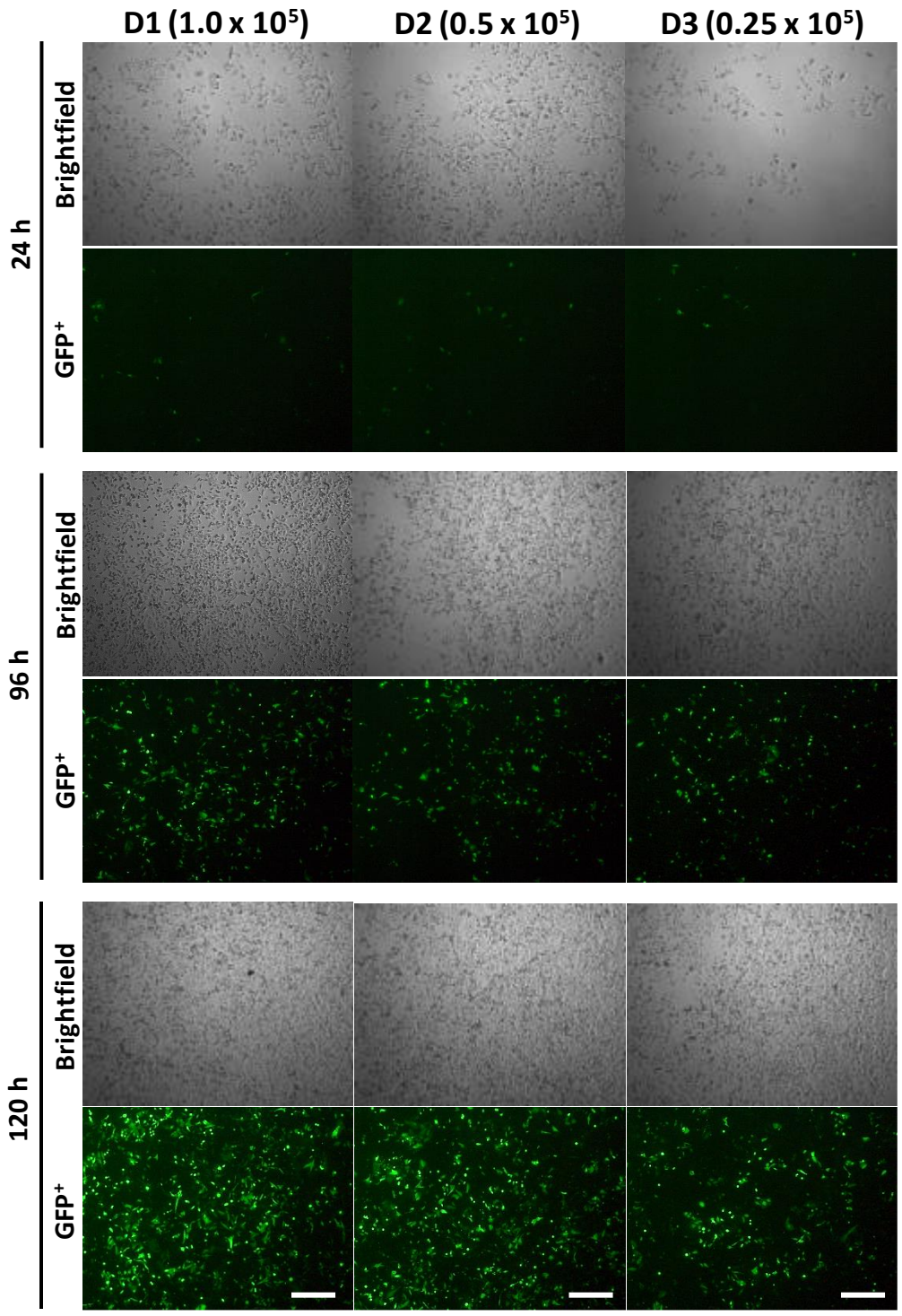


Figure 4.4 Representative images of GFP$^+$ BT-549 cells seeded at different densities at different time points after transfection.

Cells seeded at D1 and D2 became over confluent after 96 h. The D3 was deemed the most suitable seeding density for reprogramming. n = 1. Scale bar = 200 µm.

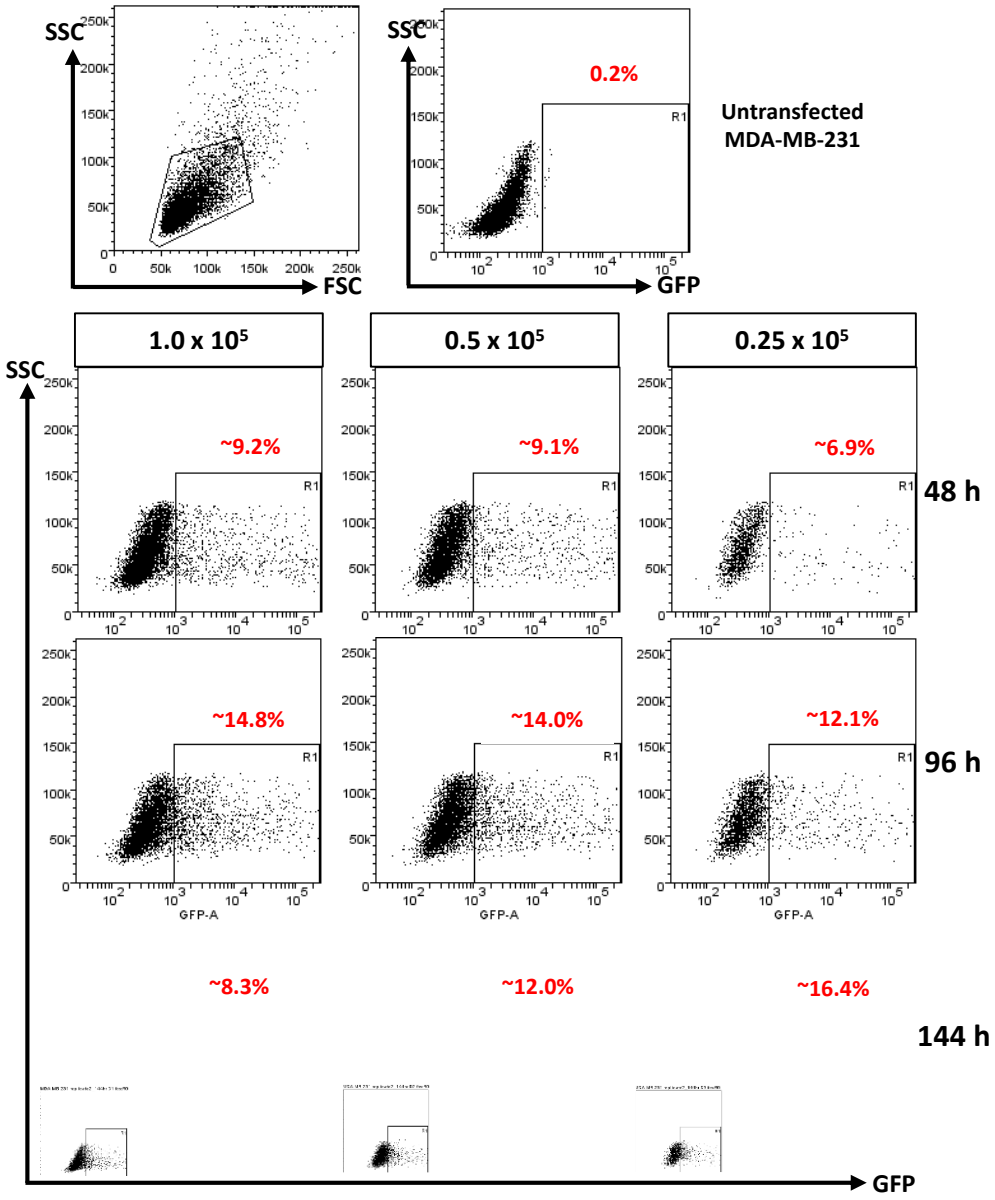


Figure 4.5 Optimization of MDA-MB-231 cell density with X-tremeGENE™ 9 DNA transfection reagent for reprogramming.

Untransfected MDA-MB-231 was used as negative control. Transfection efficiencies of three cell densities were assessed based on GFP signal. Highest GFP signal recorded at 96 h in D1 and D2. Signal diminished over time, except for D3, indicating loss of episomal vector. n = 1. SSC = side scatter.

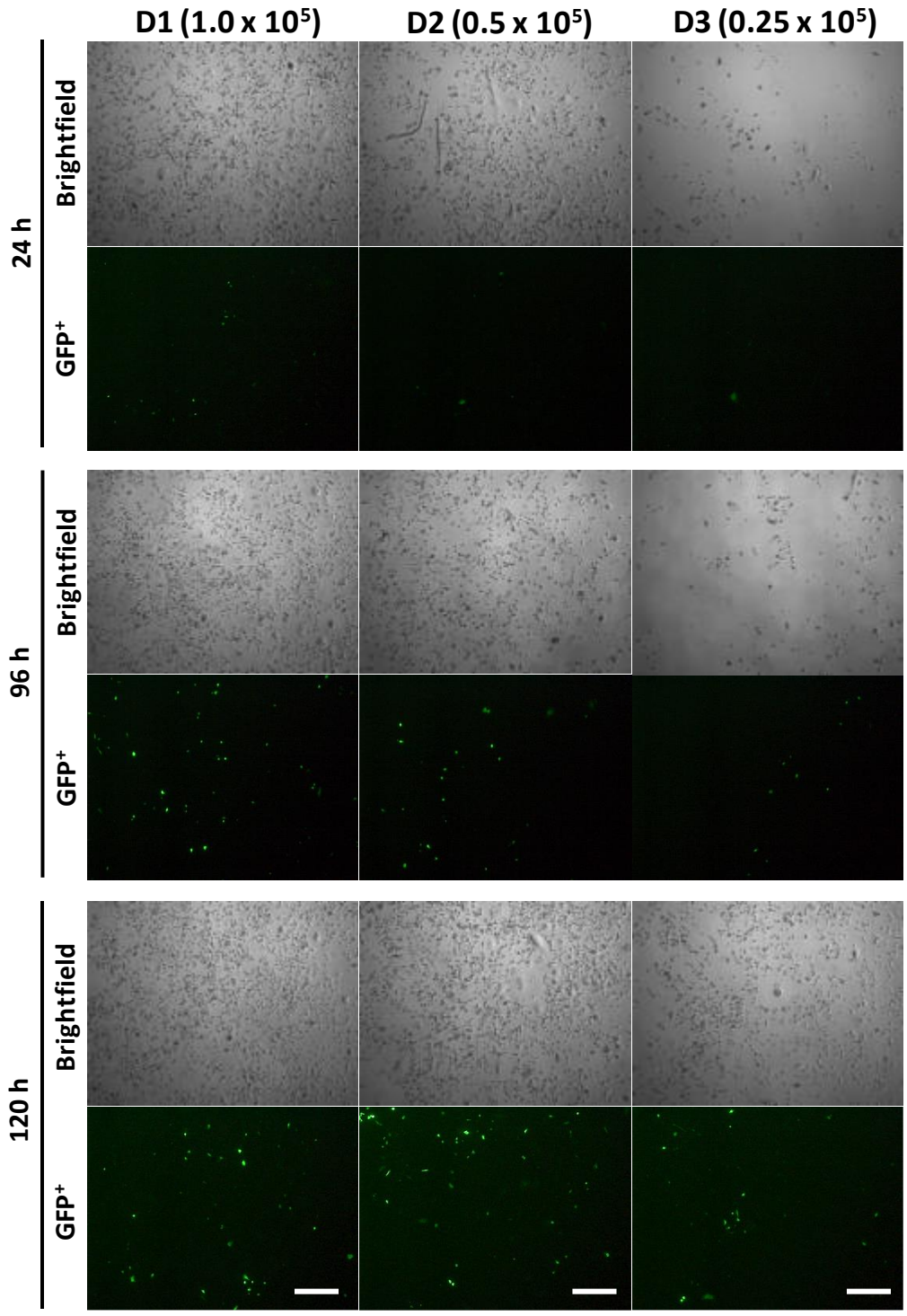


Figure 4.6 Representative images of GFP$^+$ MDA-MB-231 cells seeded at different densities at different time points after transfection.

Cells seeded at D2 resulted in moderate confluent culture and deemed to be most suitable as seeding density for reprogramming. n = 1. Scale bar = 200 µm.

Considering the low reprogramming efficiency of MDA-MB-231, another transfection platform, namely the Amaxa Cell line Nucleofector™ kits (Lonza), was used to see if the overall efficiency for all three breast cancer cell lines could be improved. Two manufacturer-optimised protocols were followed, the Nucleofector™ kit V (for MCF7 and MDA-MB-231) and Nucleofector™ kit T (for BT-549). All nucleofection experiments were carried out using the Nucleofector™ 2b device according to manufacturer instruction, including the use of cell density. Results indicated that the overall efficiency for all cell lines were lower as compared to the XG9 (Appendix Figure 4S-1 & 4S-2). Therefore, the XG9 method was chosen to reprogram the breast cancer cell lines.

4.2.2 Reprogramming of BT-549, MCF7 and MDA-MB-231

The overexpression of reprogramming genes in the three breast cancer cell lines was induced by XG9 transfection reagent-mediated transfection of episomal plasmids encoding *OCT3/4* and shRNA against *TP53* (Addgene #27077), *SOX2/KLF4* (Addgene #27078), *L-MYC/LIN28* (Addgene #27080) and *EBNA1* (Addgene #37624) (Okita et al., 2011). Cells were seeded at density previously optimised: MCF7, BT-549 and MDA-MB-231 at 0.8 x 10^5, 0.25 x 10^5 and 0.5 x 10^5, respectively. Cells were plated the day before transfection on a 6-well plate pre-coated with vitronectin. On the day of the first transfection (Day 0), culture medium was changed to Essential 8 (E8) medium (supplemented with 1 μM hydrocortisone) before switching to Essential 6 medium (supplemented with 100 ng/mL human basic fibroblast growth factor, bFGF) at Day 7. From Day 20-30, potential iPSC colonies

were manually passaged and transferred into a new culture vessel with E8 medium supplemented with 10 μM Y-27632 (ROCK inhibitor). The Y27632 was removed the following day and medium was changed every day (Figure 4.7). Defined iPSC colonies were observed and isolated from BT-549 but not from MCF7 and MDA-MB-231. Nevertheless, the reprogrammed cells were passaged to establish lines that could be stably expanded in the E8 medium (Figure 4.7). One reprogrammed cell line was obtained from MCF7 and named MCF7R. Cells underwent dramatic morphological changes from epithelial-like to irregularly-shaped cells exhibiting distinct cell-cell junction and less-shiny and -smooth appearance. On the other hand, no notable phenotypic changes were observed in two reprogrammed MDA-MB-231 cell lines named MDA R1 and MDA R2.

A total of eight iPSC colonies were manually passaged and four were successfully expanded from BT-549 as individual lines named BT-1, BT-2, BT-4 and BT-8 (Figure 4.8). The other four colonies were discarded as cells reacquired BT-549-like morphology after 2-4 passages in E8 medium. Significant morphological changes were observed in the four reprogrammed cell lines. Cells became more epithelial-like, organized and grown as compact colonies.

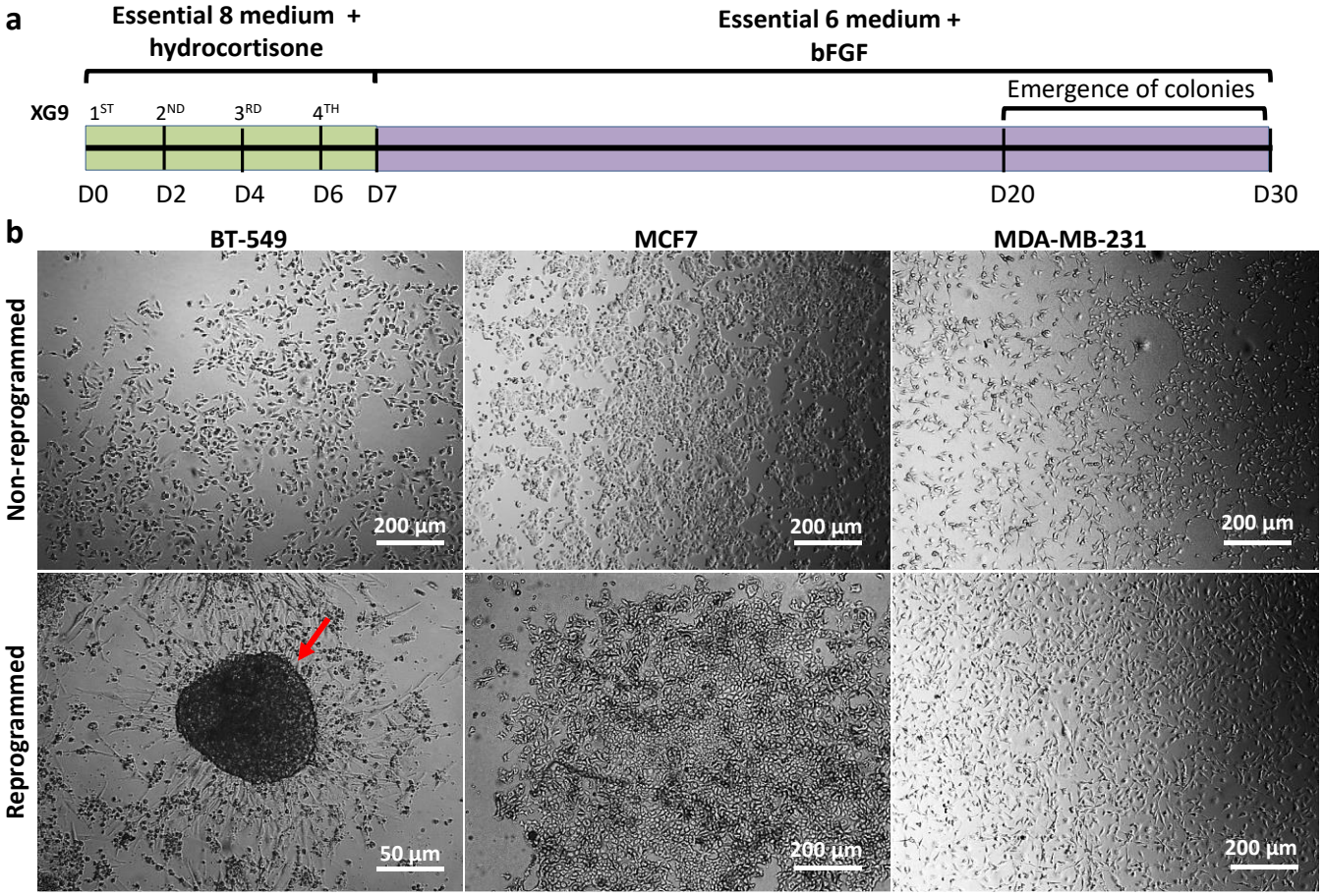


Figure 4.7 Reprogramming of breast cancer cell lines

(a) The reprogramming timeline of BT-549, MCF7 and MDA-MB-231. (b) Representative images of MCF7, MDA-MB-213 and BT-549 before and after reprogramming. Only BT-549 gave rise to iPSC-like colonies (red arrow).

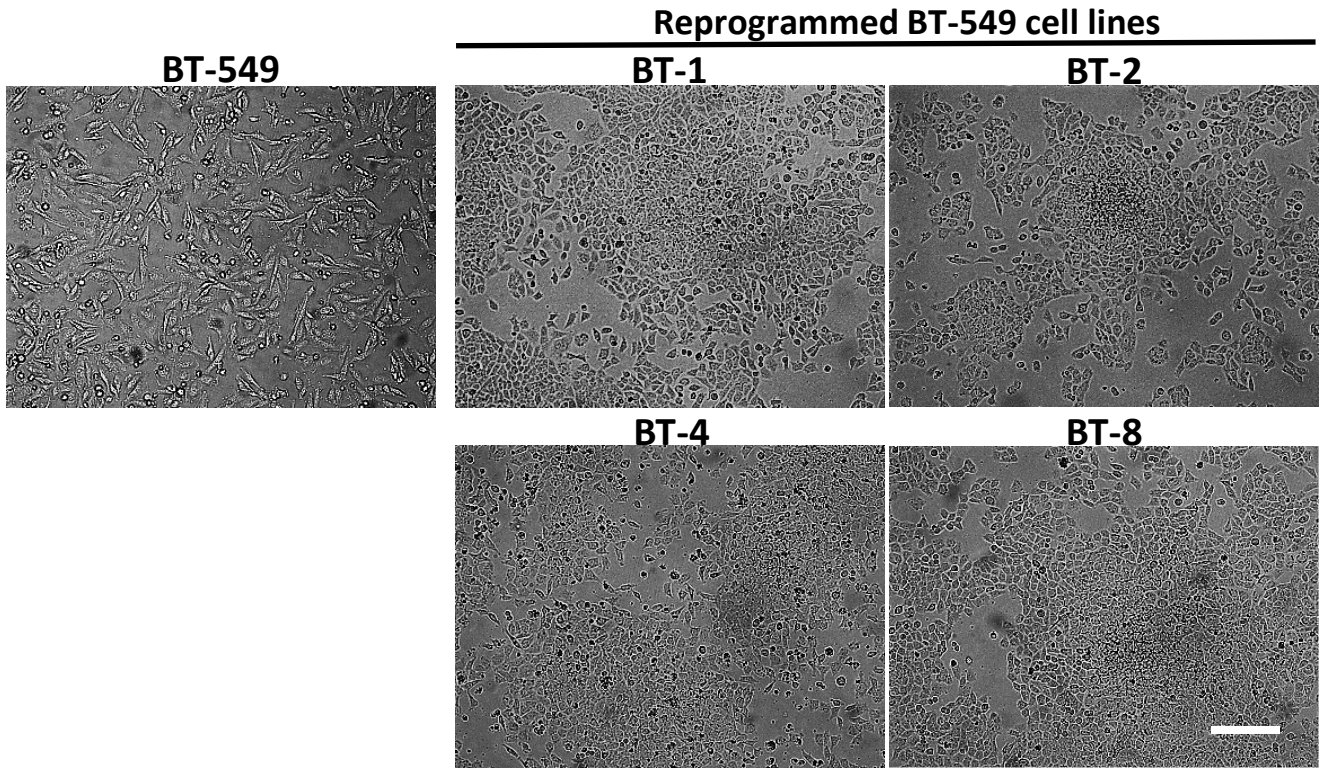


Figure 4.8 Morphology of reprogrammed BT-549 cell lines cultured in Essential 8 medium.

Pronounced morphological differences were observed between BT-549 and the reprogrammed cell lines BT-1 (passage 12), BT-2 (passage 10), BT-4 (passage 8) and BT-8 (passage 12). Scale bar = 50 µm.

4.2.2.1 Expression of pluripotency markers

RT-PCR was performed to determine the expression of endogenous pluripotency markers, including *REX1, NANOG, OCT3/4, DPPA4*, and *DNMT3B*. All endogenous pluripotency markers primers used spanned two or more exons and thus capable of discerning cDNA template from possible genomic DNA (gDNA) contamination. The embryonic stem cell line H9 was used as a positive control. Negative control (NC) consisted of complete PCR mix without cDNA template. The housekeeping gene, *B-ACTIN* (*ACTB*) was also included as an endogenous control. Weak *NANOG* expression was only detected in BT-8 and not BT-549 or other reprogrammed cell lines. On the other hand, both MCF7 and MCF7R expressed *DNMT3B* and the expression of *DPPA4* and *OCT3/4* were only detected after reprogramming. Likewise, MDA-MB-231, MDA R1 and R2 weakly expressed *DNMT3B*. The expression of *DPPA4* and *OCT3/4* was induced after reprogramming in MDA R1 and only *DPPA4* in MDA R2 (Figure 4.9). These results suggested that despite the morphological changes, the propagated cell line failed to express important pluripotency markers. Therefore, a different reprogramming approach was tested.

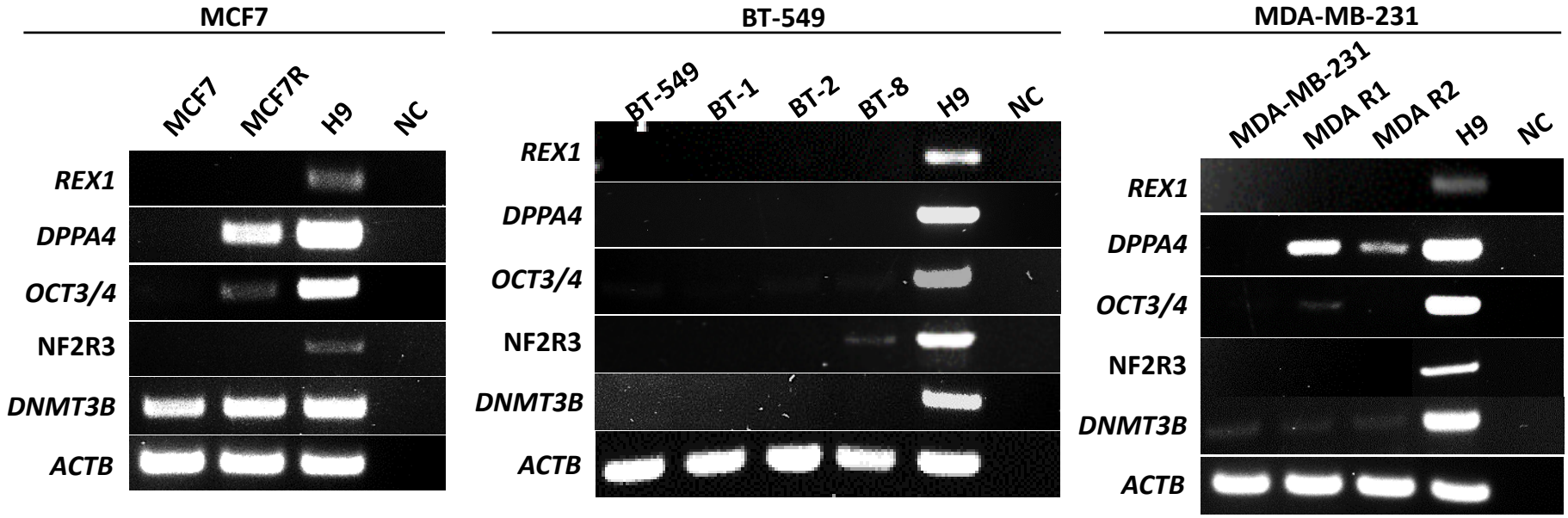


Figure 4.9 Representative RT-PCR of pluripotency markers in reprogrammed MCF-7, BT-549 and MDA-MB-231.

(left) Pluripotency markers expression by MCF7 and its reprogrammed cell line MCF7R. (middle) Pluripotency markers expression by BT-549 and its reprogrammed cell lines (BT-1, BT-2, BT-4 and BT-8). (right) Pluripotency markers expression by MDA-MB-213 and its reprogrammed cell lines MDA R1 and R2. MCF7 = passage 10 – 20, BT-549 = 20 – 30, MDA-MB-231 = passage 20 – 33, MCF7R = passage<10, BT-1, BT-2 and BT-8 = passage 4 – 15, MDA R1 and MDA R2 = passage <10, H9 = passage 24 – 30. (n≥3).

4.2.3 Reprogramming of BT-549 using the FRM/FMM media system

The last section suggested that the reprogramming of the breast cancer cell lines using the E8/E6 media system failed to generate *bona fide* iPSCs but partially reprogrammed cells which expressed only some pluripotency markers were isolated. Thus, a new media system consisting of the fate-reprogramming and fate-maintenance medium (FRM/FMM), as described by Valamehr et al. (2014), was used. FMM is composed of conventional hESC medium supplemented with a combination of small molecules proven to be beneficial for reprogramming and the survival of PSCs. FRM consists of FMM plus the cell culture supplements B-27 and N2, in addition to SB431542 (an Activin/BMP/TGF-β pathway inhibitor) (Section 2.2.3). The FRM/FMM system has been shown to enhance the efficiency of episomal vector-based reprogramming and support iPSC reprogramming with minimal gene combination (*OCT4, SOX2* and *SV40LT*) (Valamehr et al., 2014). Therefore, this highly efficient system was initially used to reprogram BT-549 as it demonstrated higher reprogramming efficiency with the E6/E8 media as compared to MCF7 and MDA-MB-231. Two iPSC colonies were isolated and expanded as individual cell lines named FMM_BT1 and FMM_BT2. In contrast to the fibroblast-like and sparsely distributed morphology observed in parental BT-549, FMM_BT1 and FMM_BT2 were epithelial-like, organized and formed compact colonies. However, only *DPPA4* expression was identified by RT-PCR analysis in both reprogrammed cell lines, but not other pluripotency genes (Figure 4.10).

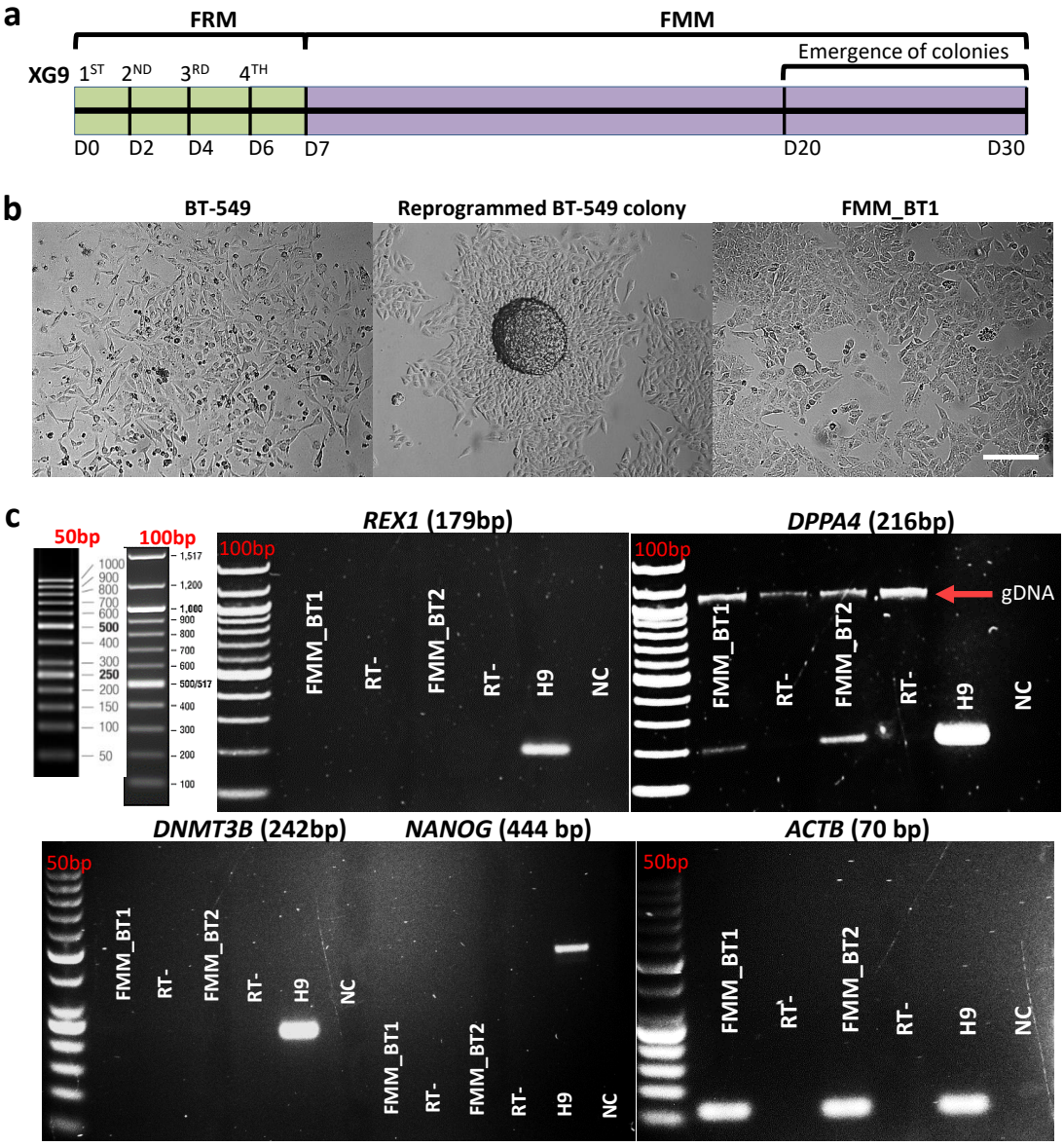


Figure 4.10 Reprogramming of BT-549 using the FRM/FMM media system.

(a) The reprogramming timeline. (b) Morphology of BT-549, reprogrammed BT-549 colony observed during reprogramming and the reprogrammed BT-549 cell line (FMM_BT1). BT-549 and FMM_BT1 exhibited distinct morphologies. (c) Representative RT-PCR for pluripotency genes expression in FMM_BT1 and FMM_BT2. Both cell lines expressed *DPPA4* but not other pluripotency markers. The presence of 1081 bp bands indicated primer used spanned exons and capable of differentiating gDNA contamination (red arrow) from cDNA. FRM = fate-reprogramming medium, FMM = fate-maintenance medium. BT-549 = 20 – 30, FMM_BT1 and FMM_BT2 = passage <10. Scale bar = 50 μm, n≥3.

4.2.4 Conversion of partially-reprogrammed iPSCs (pre-iPSCs) to *bona fide* iPSCs.

Partially-reprogrammed iPSCs (pre-iPSCs) are cells generated during iPSC reprogramming whose identities are incompletely reverted to resemble ESCs and only exhibit a few but not all ESCs hallmarks (Mikkelsen et al., 2008a, Theunissen et al., 2011). Pre-iPSCs can be isolated and expanded as stable cell lines in various stem cell media and characterized by the reactivation of specific subset of pluripotency genes, incomplete elimination of DNA methylation in pluripotency loci and the partial repression of lineage-specific genes (Mikkelsen et al., 2008a). Studies reported that treatment of potent DNA demethylating agents, such as 5-azacytidine (5-aza) and vitamin C, facilitates the conversion of pre-iPSCs to *bona fide* iPSCs. FMM_BTR1 and FMM_BTR2 cell lines exhibited several features of pre-iPSCs, including the expression of some pluripotency markers (*DPPA4*) but not genes of the core pluripotency network (*REX1* and *NANOG*) and the acquisition of ESC-like morphology. Therefore, attempt was made to convert FMM_BTR1 to *bona fide* iPSCs by treating cells with 5-aza, as described in Theunissen et al. (2011). FMM_BTR1 cells were cultured to 80-90% confluence before the treatment of 0.5 mM 5-aza for 48 hours. Treated cells were passaged at least five times, which lasted for about 16 days, before being harvested for analysis. After 5-aza treatment, no significant morphological changes were observed. RT-PCR indicated the activation of *REX1* and *DNMT3B* expression, but not *NANOG* and *OCT3/4*.

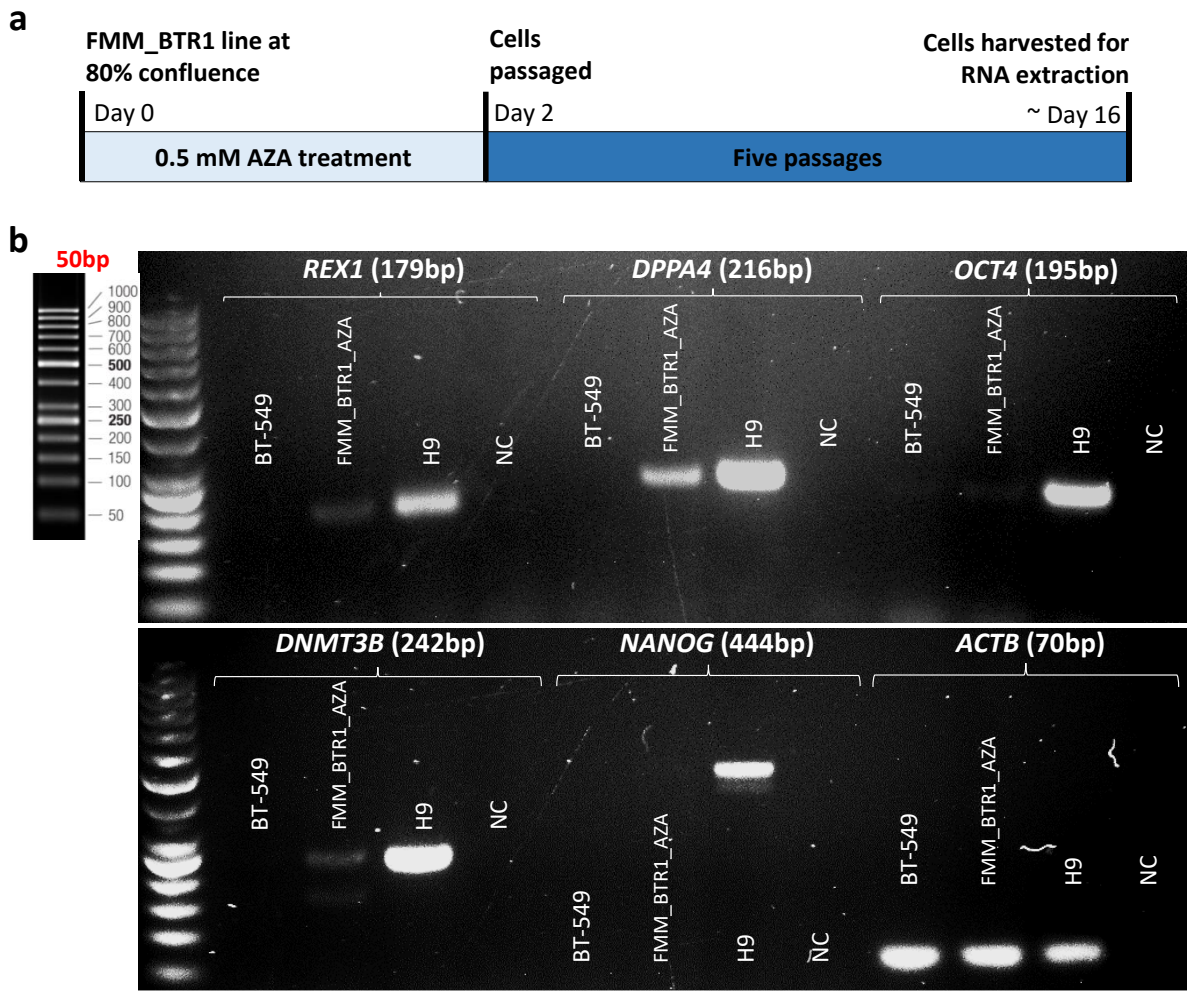


Figure 4.11 Conversion of FMM_BTR1 to *bona fide* iPSCs by 5-Azacytidine (AZA) treatment.

(a) Timeline for 5-aza treatment of FMM_BTR1. (b) Representative RT-PCR analysis of pluripotency markers expression by 5-aza-treated FMM_BTR1 (FMM_BTR1_AZA). NC = negative control, AZA = 5-aza. BT-549 = 20 – 30, FMM_BTR1_AZA = passage <10. (n≥3).

4.2.5 Reprogramming of BT-549 with five plasmids and sodium butyrate

Findings from the previous sections in this chapter indicated that breast cancer cell lines tested were not entirely unamendable to iPSC reprogramming. Although complete refractoriness has been observed in certain cancer types such as B cell-acute lymphoid leukaemia (B-ALL) (Munoz-Lopez et al., 2016). It was hypothesised that "extra forces" were required to push cells up the Waddington's developmental landscape to acquire ESC-like pluripotency (Goldberg et al., 2007, Noble, 2015, Moris et al., 2016). Therefore, the reprogramming cocktail was modified by adding an extra plasmid encoding the *microRNA-302/367* (*miR-302/367*) cluster (Addgene #31310), whose expression, by itself, can reprogram human fibroblasts to iPSCs (Anokye-Danso et al., 2011). In addition, sodium butyrate (NaB), a histone deacetylase (HDAC) inhibitor (HDACi), was also supplemented to increase reprogramming efficiency (Zhang et al., 2011, Wen et al., 2016a).

Again, BT-549 was selected to experiment with the new reprogramming condition involving five plasmids and NaB treatment. Taking into consideration the toxicity of NaB, the seeding density was increased from 0.25 x 10^5 to 0.5 x10^5 per well of 6-well plate. BT-549 cells were transfected four consecutive times every other day with five plasmids using the XG9 transfection reagent. Cells were plated the day before transfection on a 6-well plate pre-coated with vitronectin. On the day of the first transfection (Day 0), culture medium was changed to E8 medium

(supplemented with 1 μM hydrocortisone) before changing to E6 medium supplemented with 100 ng/mL human bFGF and 0.5 mM NaB at Day 8. Medium was also supplemented with 10 μM Y-27632 (ROCKi) to minimise cell death inflicted by the toxicity of NaB. The NaB treatment as well as ROCKi was stopped at day 12. From day 20-30, potential iPSC colonies were manually passaged and transferred into a new culture vessel with E8 medium supplemented with 10 μM ROCKi. The ROCKi was removed the following day and medium was changed every day (Figure 4.12a). A total of 9 iPSC colonies were isolated and five were expanded as individual lines named 5PLA_BTR1, 5PLA_BTR3, 5PLA_BTR4, 5PLA_BTR5 and 5PLA_BTR6. Some iPSC lines were more heterogeneous (5PLA_BTR1 and 3) than the others at early passages. 5PLA_BTR5 was passaged over passage 10 and subjected to characterisation. The 5PLA_BTR5 cells have high nucleus-to-cytoplasm ratio and preferentially expanded as compact colonies which exhibited well-defined edges, closely resembling ESCs and iPSCs isolated from MCF10A in the previous chapter (Figure 4.12).

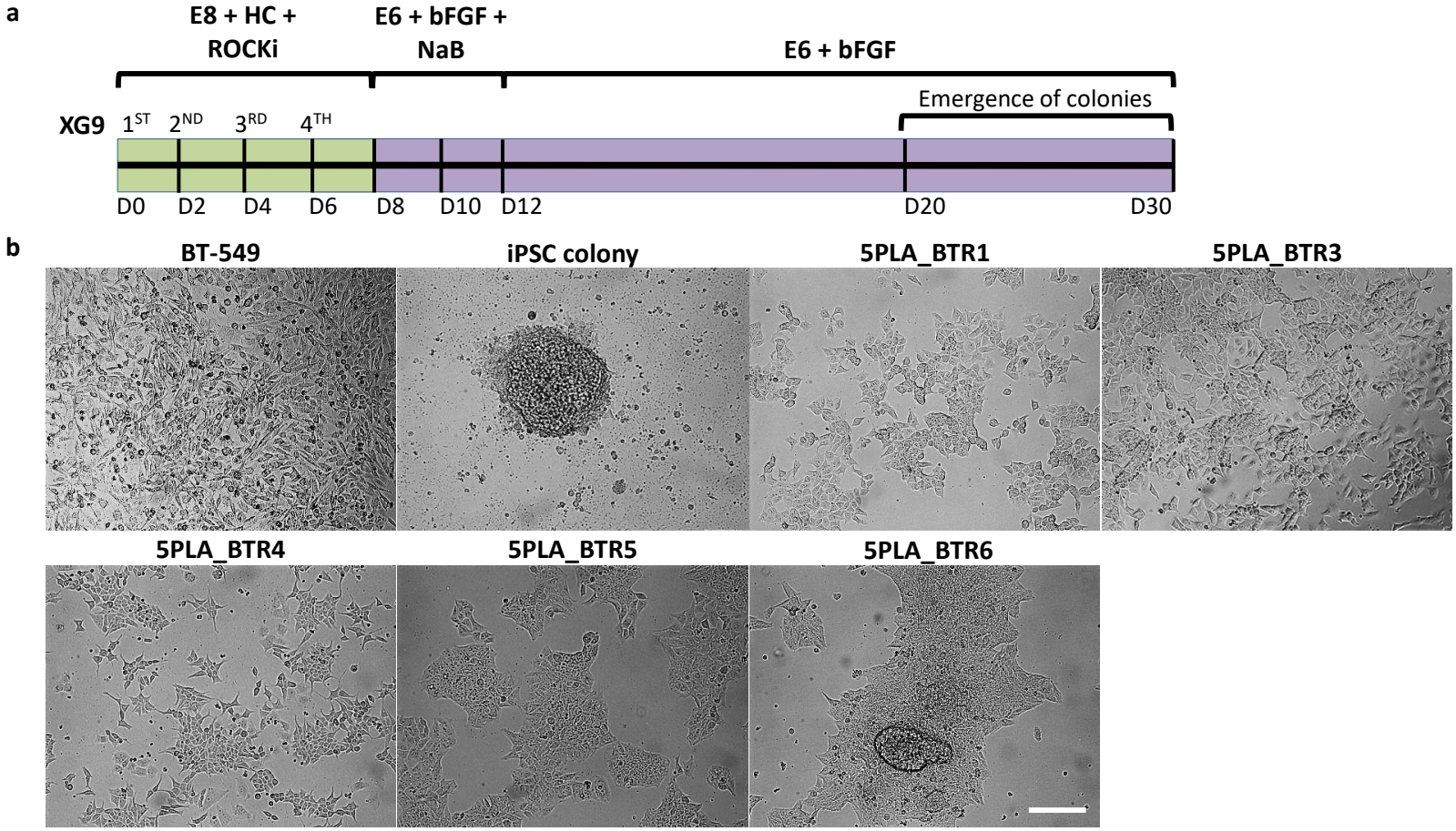


Figure 4.12 Reprogramming of BT-549 with five episomal plasmids and sodium butyrate (NaB).

(a) The reprogramming timeline. (b) Morphology of BT-549, iPSC colony and iPSC lines derived from BT-549. All iPSC lines were at early passage (5 – 8) thus appeared heterogeneous. The 5PLA_BTR5 was expanded and subjected to characterization. E8 = Essential 8 medium, E6 = Essential 6 medium, HC = hydrocortisone, NaB = sodium butyrate, bFGF = basic fibroblast growth factor. Scale bar = 50 μm.

4.2.5.1 Characterisation of 5PLA_BT5

5PLA_BTR5 was characterised by RT-PCR, RT-qPCR and immunofluorescence. RT-PCR indicated that 5PLA_BTR5 lost all episomal vectors at around passage 19 (Figure 4.13).

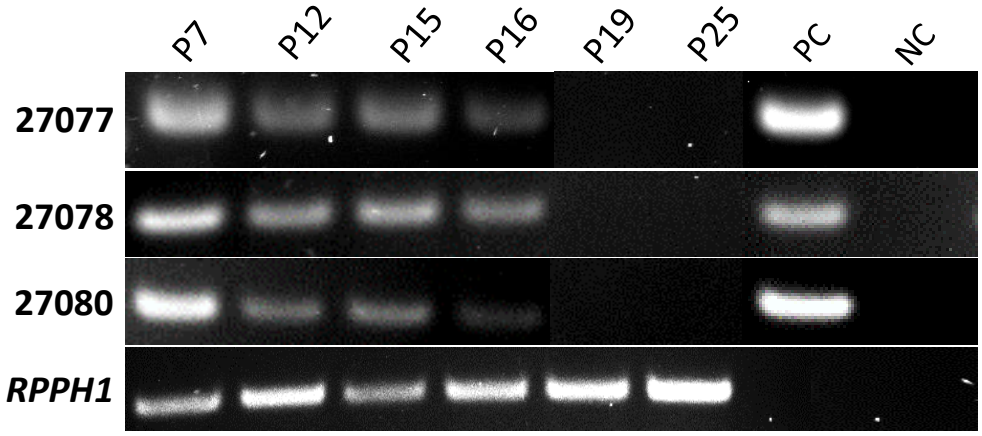


Figure 4.13 Traces of episomal vectors in 5PLA_BTR5.
Traces of episomal vectors 27077, 27078 and 27080 were lost at around passage 19. *RPPH1* = ribonuclease P RNA component H1, P = passage, Addgene plasmids: 27077 = pCXLE-hOCT3/4-shP53-F, 27078 = pCXLE-hSOX/KLF4*,* 27080 = pCXLE-hLIN28A/L-MYC., PC = positive control, NC = negative control. n ≥ 3.

5PLA_BTR5 expressed all pluripotency markers, including *REX1, DPPA4, OCT3/4, NANOG* and *DNMT3B* and at levels significantly higher than BT-549 (*REX1,* $p<0.0005$; *DPPA4,* $p<0.0005$; *OCT3/4*, $p<0.005$; *NANOG,* $p<0.0005$ and *DNMT3B,* $p<<0.0001$) and comparable to ESCs (Figure 4.14 & Figure 4.15). However, lower expression of *REX1* ($p<0.0005$) and higher expression of *NANOG* ($p<0.0005$) and *DNMT3B* ($p<<0.0001$) were observed in 5PLA_BTR5 in comparison to H9. Notably, 5PLA_BTR5 was negative when NF2R2 primer (differentiates *NANOG* from all pseudogenes)

was used and generated a 446 bp PCR band (for *NANOG* P8) instead of 467 bp (for *NANOG*) with NF2R3 primer. The result suggested that iPSC reprogramming activated *NANOG P8* instead of *NANOG* in 5PLA_BTR5. Nevertheless, immunofluorescence staining indicated that 5PLA_BTR5 was positive for TRA-1-81, SSEA-4 and NANOG (Figure 4.16).

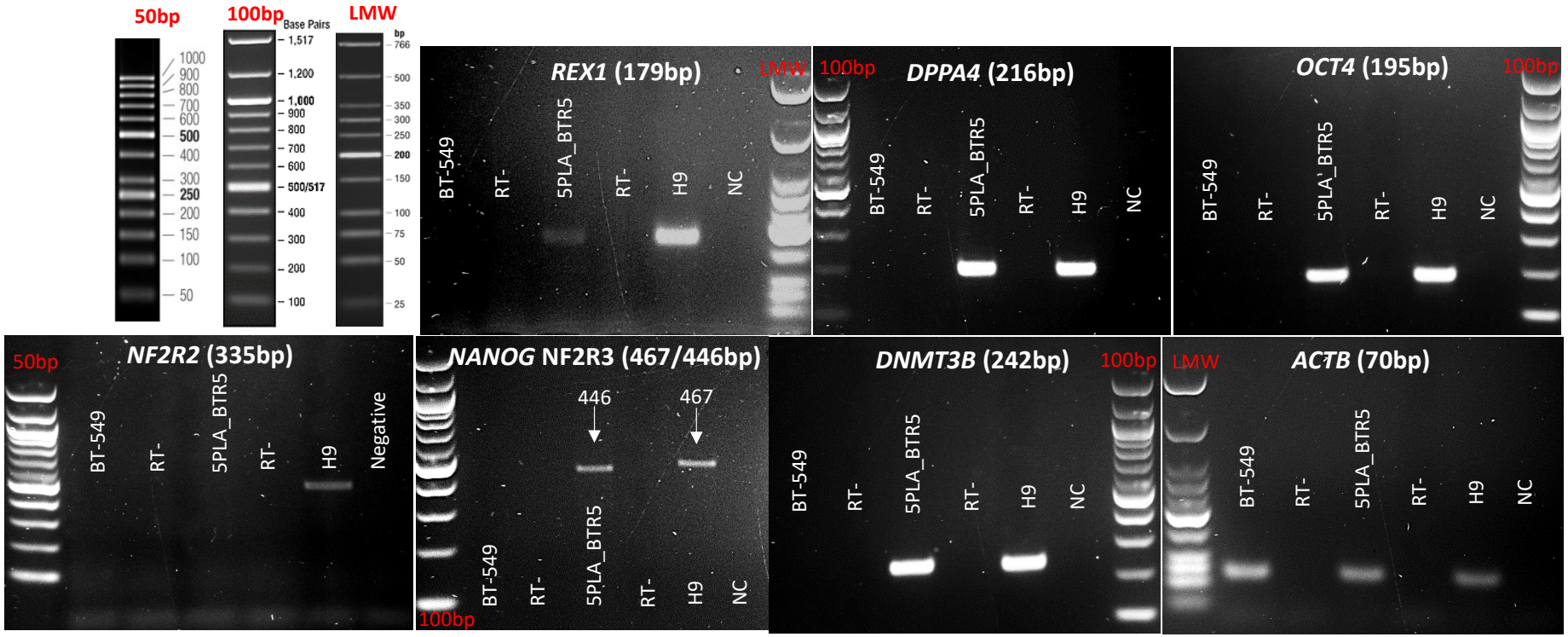


Figure 4.14 Representative RT-PCR of pluripotency markers in 5PLA_BTR5.

5PLA_BTR5 expressed all pluripotency markers. *NANOG* expression as determined by the NF2R2 and NF2R3 primer indicated that 5PLA_BTR5 expressed *NANOG P8* (446 bp) instead of *NANOG* (467 bp). RT- = reverse transcriptase negative, NC = negative control, 50 bp = 50 bp ladder, 100 bp = 100 bp ladder, LMW = low molecular weight ladder. BT-549 = 20 – 30, 5PLA_BTR5 = passage 19 – 30, H9 = passage 24 – 30. n≥3.

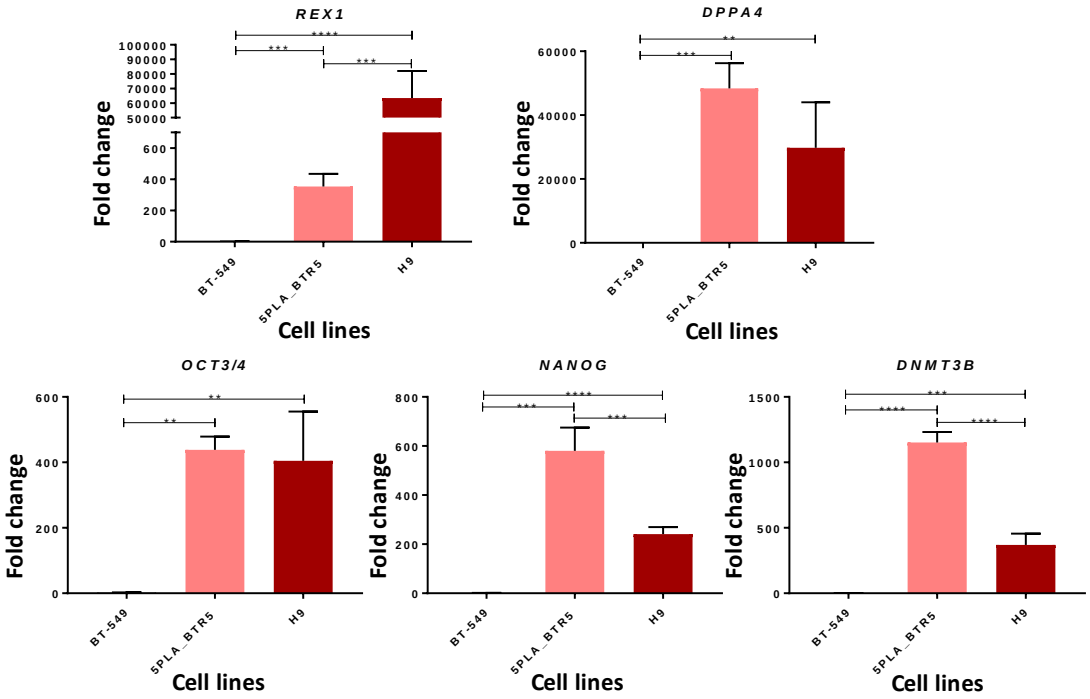


Figure 4.15 Quantification of pluripotency markers expression in 5PLA_BTR5 by real-time-PCR (RT-qPCR).

RT-qPCR for endogenous expression of pluripotent markers (*REX1*, *NANOG*, *OCT3/4, DNMT3B* and *DPPA4*). Results were normalized to *ACTB*. Data was analysed by one-way ANOVA followed by Bonferroni correction. BT-549 = 20 – 30, 5PLA_BTR5 = passage 19 – 30, H9 = passage 24 – 30. * = <0.05, ** = <0.005, *** = <0.0005, **** = < 0.0001. BT-549, n ≥ 3; 5PLA_BTR5, n ≥ 3; H9, n ≥ 3.

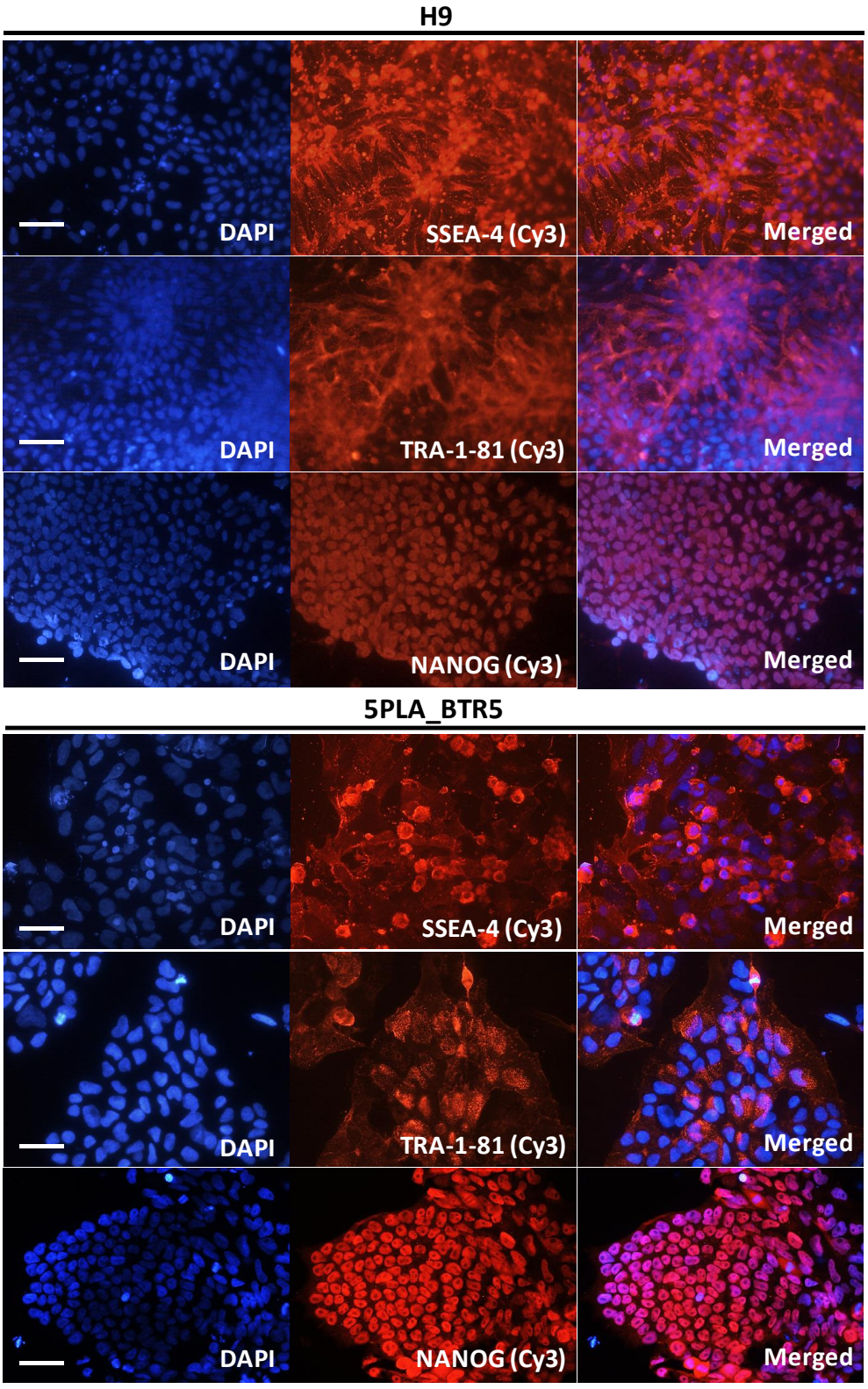


Figure 4.16 Immunofluorescence staining of SSEA-4, TRA-1-81 and NANOG in 5PLA_BTR5.

Like the ESC line H9, 5PLA_BTR5 expressed all three pluripotency markers tested. Scale bar = 50 μm. Secondary antibody-only negative control is shown in Figure S3-3. 5PLA_BTR5 = passage 25. n = 1.

4.2.6 Reprogramming of MCF7 with five plasmids and sodium butyrate treatment

Following the success of BT-549, the same reprogramming condition involving five plasmids and NaB was used to reprogram the ER^+ breast cancer cell line MCF7. Cell seeding density was increased to 0.8 x 10^5 to account for the toxicity of NaB. Reprogramming was performed as previously described. No iPSC colonies were observed or isolated during reprogramming up to day 30. Cells were passaged with EDTA to generate a stable reprogrammed cell line in E8 medium. Cells of the reprogrammed cell lines produced in this section and section 4.2.2 (reprogramming with four plasmids without NaB) exhibited similar morphology, including irregular in shape with distinct cell-cell junction and less-shiny and -smooth appearance. Interestingly, after about three passages, a population of ESC-like cells (flat shape cells which preferentially expanded as compact colonies with well-defined edges) emerged and gradually forming heterogenous culture as passage number increased. At passage five, a homogenous culture dominated by the ESC-like cells was achieved. Only one reprogrammed cell line was established from MCF7 named 5PLA_MCF7R (Figure 4.17).

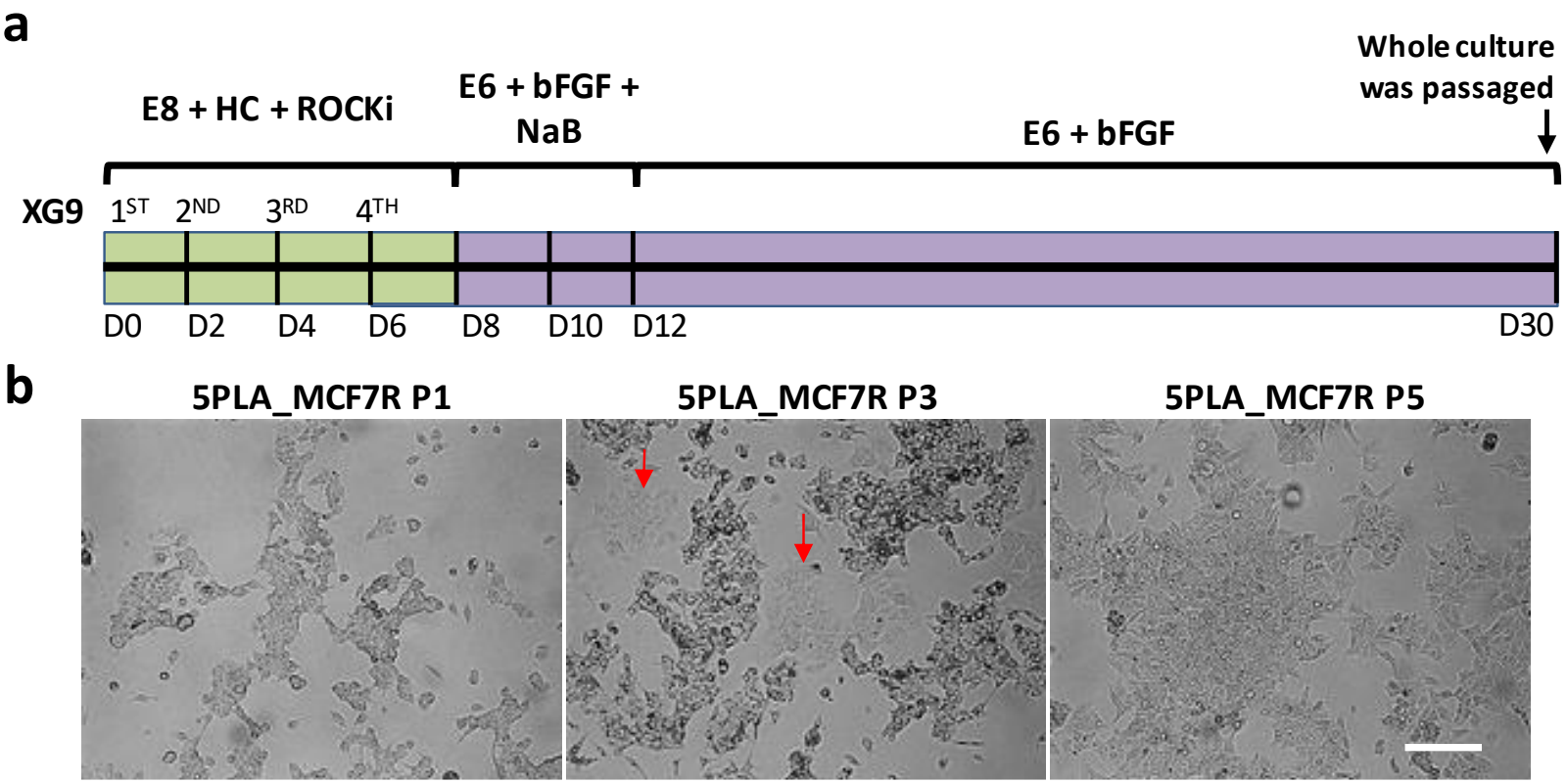


Figure 4.17 Reprogramming of MCF7 with five episomal plasmids in combination with sodium butyrate (NaB) treatment.

(a) The reprogramming timeline. (b) Morphology of reprogrammed MCF7 at different passages. At passage one, 5PLA_MCF7R resembled reprogrammed cell line generated by four-plasmids reprogramming. ESC-like cells emerged gradually (red arrows) as 5PLA_MCF7R cells were passaged and became homogenous at passage 5. E8 = Essential 8 medium, E6 = Essential 6 medium, HC = hydrocortisone, NaB = sodium butyrate, bFGF = basic fibroblast growth factor. Scale bar = 50 μm.

4.2.6.1 Characterisation of 5PLA_MCF7R

5PLA_MCF7R was subjected to characterisation by RT-PCR, qPCR, immunofluorescence and *in vitro* differentiation. 5PLA_MCF7R lost all episomal vectors at around passage 5, in contrast to 5PLA_BTR5 (passage 19), indicating independence from transgenes (Figure 4.18).

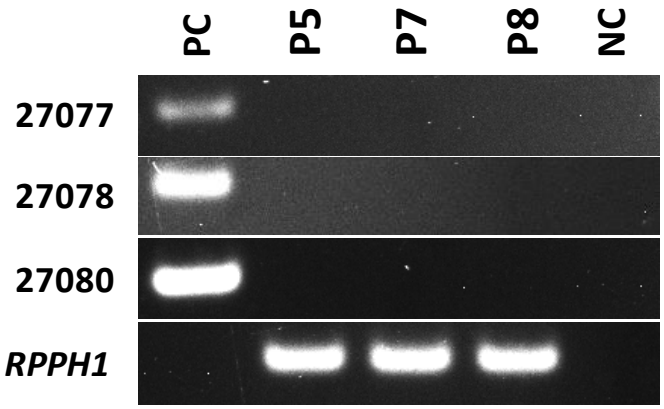


Figure 4.18 Traces of episomal vectors in 5PLA_MCF7R.
Traces of episomes 27077, 27078 and 27080 were lost at around passage 5. NC = negative control, *RPPH1* = ribonuclease P RNA component H1, P = passage, Addgene plasmids: 27077 = pCXLE-hOCT3/4-shP53-F, 27078 = pCXLE-hSOX/KLF4, 27080 = pCXLE-hLIN28A/L-MYC, (n ≥3).

5PLA_MCF7R expressed pluripotency markers *DPPA4, OCT3/4, NANOG* and *DNMT3B* but not *REX1* as indicated by RT-PCR (Figure 4.19). Consistently, all pluripotency markers (*DPPA4,* $p<0.0005$; *OCT3/4*, $p<0.05$; *NANOG,* $p<0.005$ and *DNMT3B,* $p<0.005$) but not *REX1* were detected in 5PLA_MCF7R as compared to MCF7 (Figure 4.20). However, low expression *OCT3/4* ($p< 0.0001$) were observed in 5PLA_MCF7R in comparison to H9.

Similar to 5PLA_BTR5, 5PLA_MCF7R was negative for *NANOG* when using the NF2R2 primer and produced a 446 bp PCR band with NF2R3 primer, indicating *NANOG P8* expression. Lastly, immunofluorescence staining indicated that 5PLA_MCF7R was positive for TRA-1-60, SSEA-4 and NANOG (Figure 4.21). Interestingly a relatively weak TRA-1-60 expression was also observed.

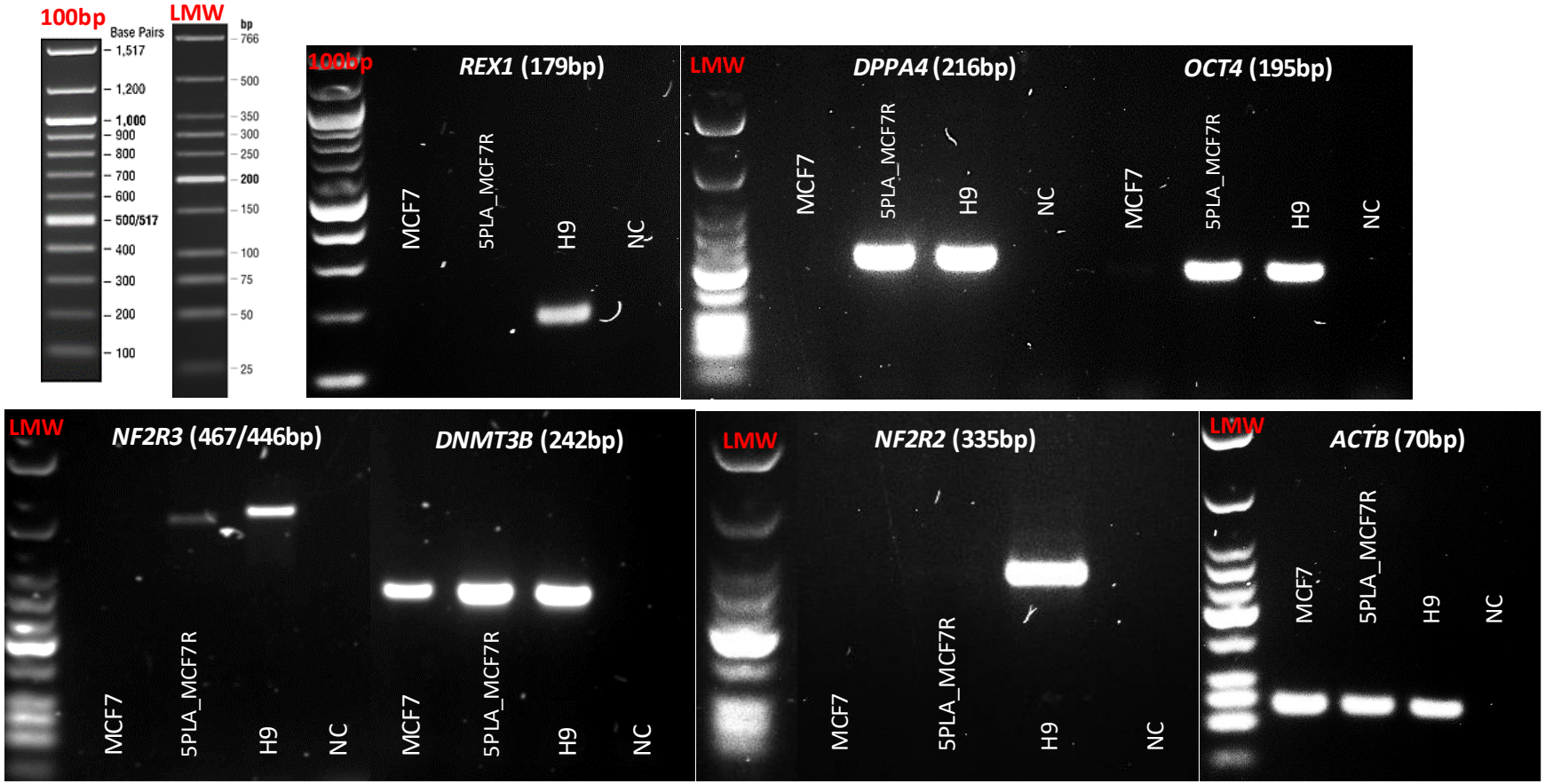


Figure 4.19 Representative RT-PCR of pluripotency markers expression in 5PLA_MCF7R.

The 5PLA_MCF7R expressed all pluripotency markers except for *REX1*. It was negative for NF2R2 primer. A 446 bp PCR band produced by NF2R3 primer indicated *NANOG P8* expression. NC =negative control, 100 bp = 100 bp ladder, LMW = low molecular weight ladder. MCF7 = passage 10 – 20, 5PLA_MCF7R = passage 9 – 20, H9 = passage 24 – 30. n≥3.

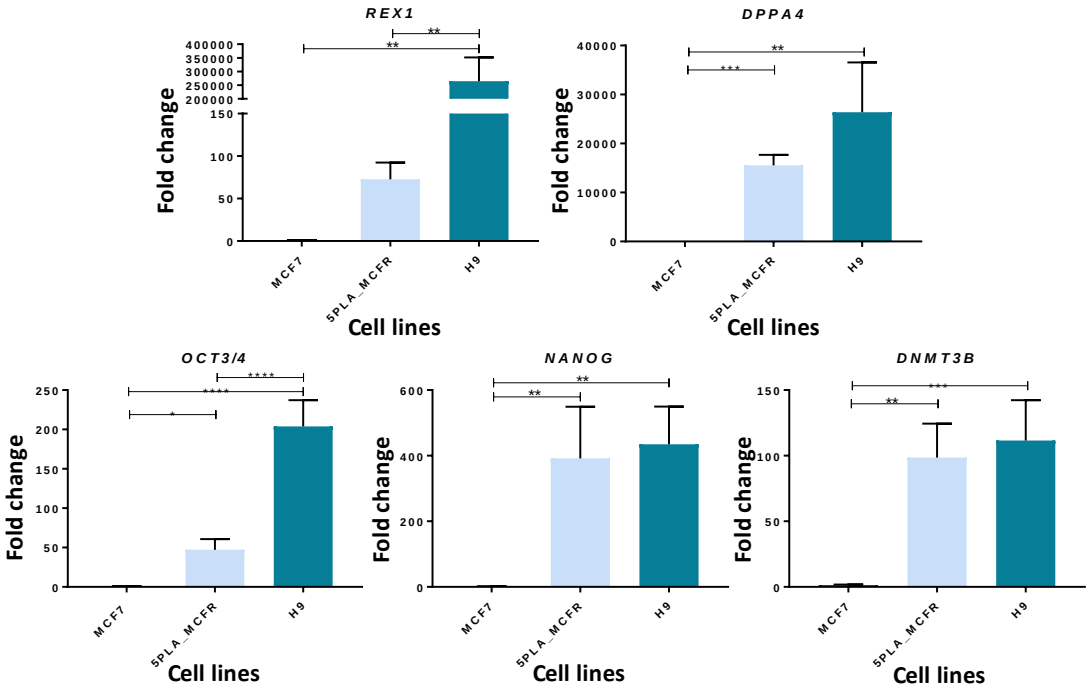


Figure 4.20 Quantification of pluripotency markers expression in 5PLA_MCF7R by real-time PCR (RT-qPCR).

RT-qPCR indicated expression of pluripotent markers (*NANOG*, *OCT3/4, DNMT3B* and *DPPA4* but not *REX1*) by 5PLA_MCF7R at levels significantly higher than its non-reprogrammed counterpart and comparable to H9. Results were normalized to *ACTB*. Data was analysed by one-way ANOVA followed by Bonferroni correction. MCF7 = passage 10 – 20, 5PLA_MCF7R = passage 9 – 20, H9 = passage 24 – 30. * = <0.05, ** = <0.005, *** = <0.0005, **** = < 0.0001. MCF7, n ≥ 3; 5PLA_MCF7R, n ≥ 3; H9, n ≥ 3.

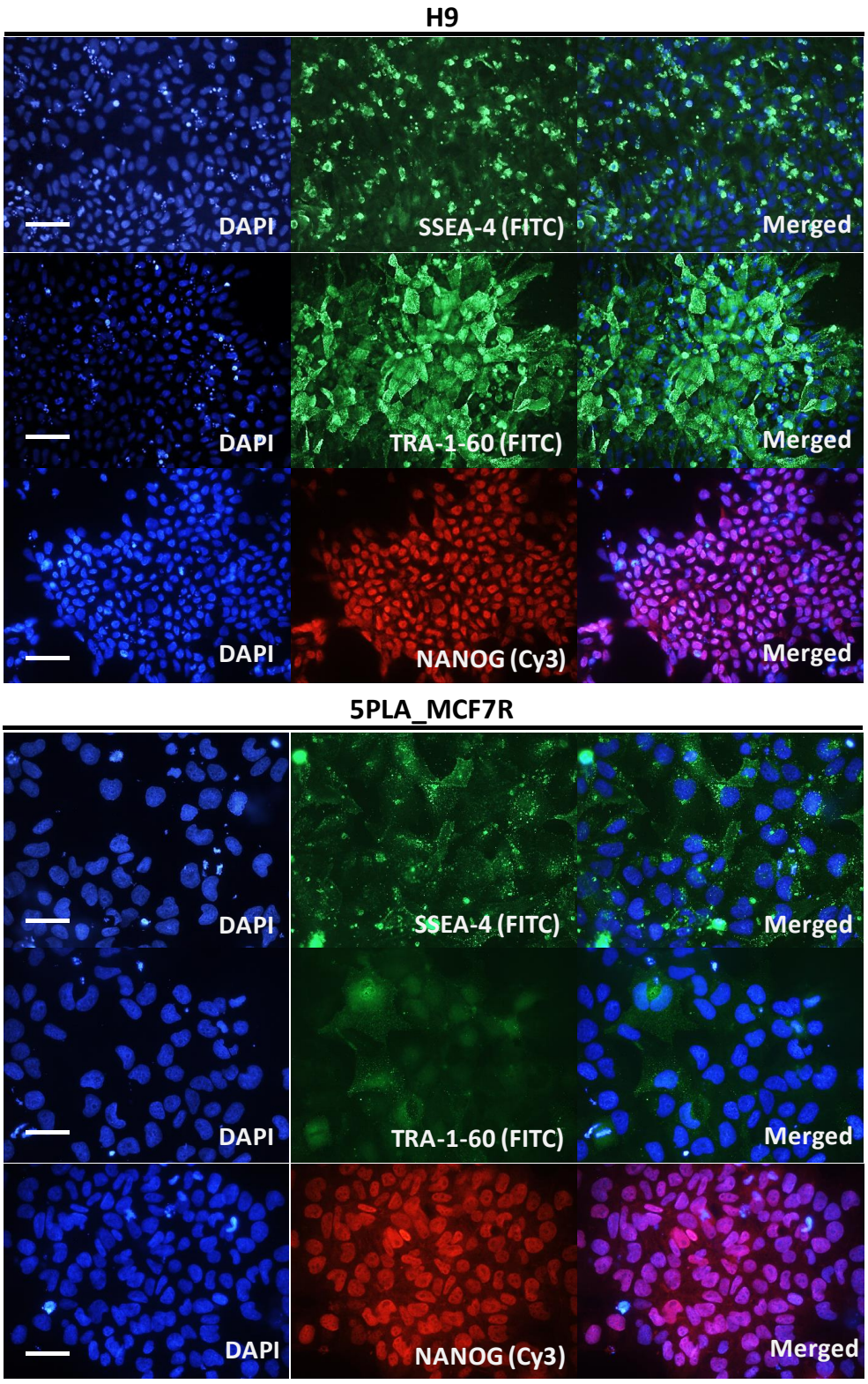


Figure 4.21 Immunofluorescence staining of SSEA-4, TRA-1-60 and NANOG in 5PLA_MCF7R.

5PLA_MCF7R expressed all three pluripotency surface markers. H9 was used as positive control. Scale bar = 50 μm. Secondary antibody-only negative control is shown in Figure S3-3. 5PLA_MCF7R = passage 10. n = 1.

4.2.7 Reprogramming of MDA-MB-231 with five plasmids and sodium butyrate

Another triple-negative breast cancer cell line MDA-MB-231 was also reprogrammed using the improved reprogramming protocol. Experimental set up was as previously described. No iPSC colonies were observed or isolated during the reprogramming. Cells were passaged with EDTA to generate two stable reprogrammed cell lines in E8 medium named 5PLA_MDAR1 and 5PLA_MDAR2 (Figure 4.22). No morphological changes were observed in reprogrammed cells as compared to the non-reprogrammed counterpart. RT-PCR indicated expression of *DPPA4* and *DNMT3B* (also expressed in MDA-MB-231) in reprogrammed cell lines. Weak NANOG expression was also detected in 5PLA_MDAR2.

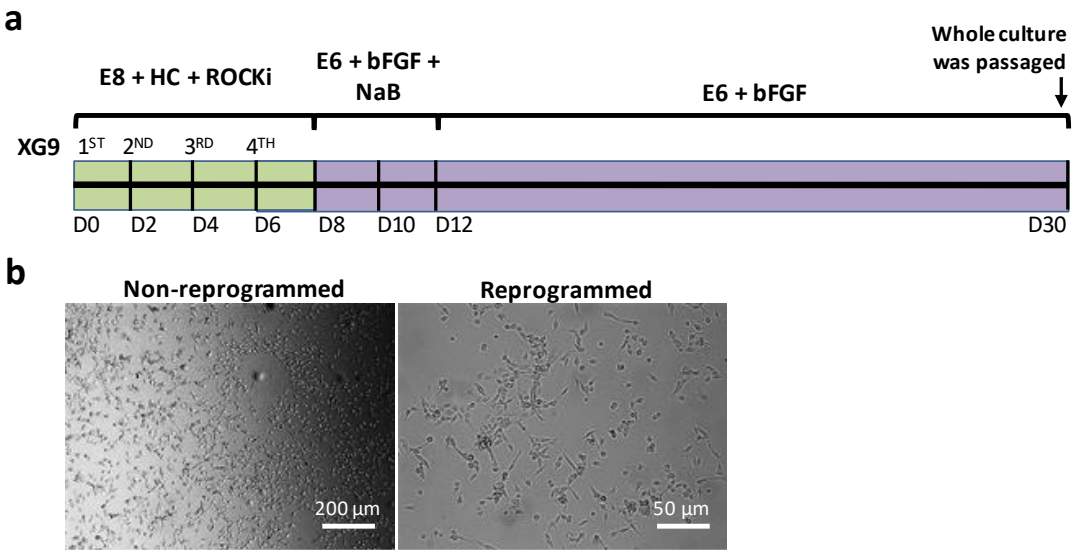


Figure 4.22 Reprogramming of MDA-MB-231 with five episomal plasmids in combination with sodium butyrate (NaB) treatment.

(a) Reprogramming timeline. (b) No obvious change of morphology has been observed in reprogrammed MDA-MB-231. E8 = Essential 8 medium, E6 = Essential 6 medium, HC = hydrocortisone, NaB = sodium butyrate, bFGF = basic fibroblast growth factor.

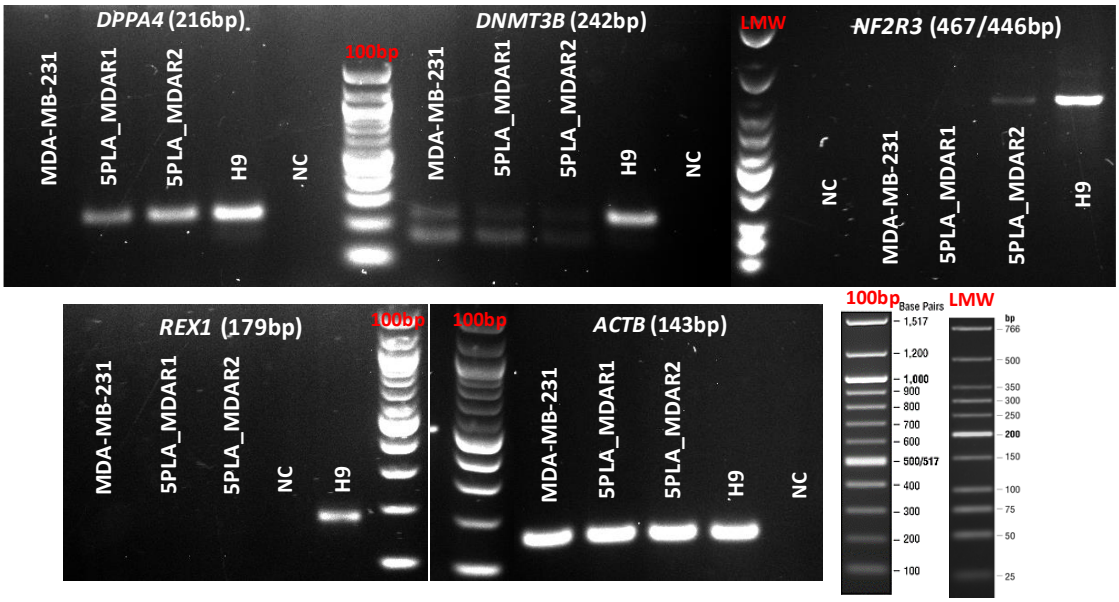


Figure 4.23 Representative RT-PCR of pluripotency markers expression in 5PLA_MDAR1 and 5PLA_MDAR2.

Both 5PLA_MDAR1 and 5PLA_MDAR2 expressed *DPPA4* and DNMT3B (also expressed in MDA-MB-231). Weak expression of NANOG can be detected in 5PLA_MDAR2 using the NF2R3 primer. Double bands observed in *DNMT3B* indicated the amplification of two different isoforms (Duymich et al., 2016). NC = negative control, 100 bp = 100 bp ladder, LMW = low molecular weight ladder. MDA-MB-231 = passage 20 – 33, 5PLA_MDAR1 and 5PLA_MDAR2 = passage <10. n≥3.

4.3 DISCUSSION

Results in this chapter demonstrated that breast cancer cells can be reprogrammed to acquire an ESC-like pluripotency state. However, additional factors, namely *miR-302/367* and NaB treatment were required for successful derivation of cancer-derived iPSCs as compared to the non-malignant control MCF10A. This suggested that breast cancer cells are less susceptible to the reprogramming process. Six iPSC lines were isolated from the triple negative breast cancer cell line BT-549 and one from ER^+ MCF7 cell line. In general, these cancer-derived iPSCs expressed key pluripotency markers and at levels comparable to the ESC line H9. During reprogramming with the improved protocol, seeding cell densities for all three breast cancer cells were slightly increased to account for the toxicity of NaB: 1.0×10^5 0.5×10^5 and 0.8×10^5 for MCF7, BT549 and MDA-MB-231, respectively. Each reprogramming experiment was set up as 3 x 6-well plates. The reprogramming efficiency for BT-549 (six colonies isolated and expanded) and MCF7 (no colony and only one line expanded) worked out as ~0.0007% (again without taking into consideration small or less healthy colonies) and 7.0×10^{-7}%. However, Corominas-Faja et al. (2013) reported similar reprogramming efficiency in MCF7 by comparing, in general, to other studies which used the retrovirus approach. On the contrary, a much lower reprogramming efficiency has been observed in the present study as compared to the control cell line MCF10A.

This study was not the first attempt to reprogram breast cancer cells using the iPSC technology. Corominas-Faja et al. (2013) reported the

successful derivation of partially reprogrammed iPSCs or pre-iPSCs from the MCF7 cell line (MCF7/Rep) by using the classical retrovirus approach. The iPSC-like colonies were observed at as early as day 10 after viral transduction, consistent with the high efficiency and kinetic of retrovirus-mediated reprogramming (Stadtfeld and Hochedlinger, 2010, Gonzalez et al., 2011). The MCF7/Rep cells overexpressed SOX2 and SSEA4 while OCT4, NANOG, SSEA-1, TRA-1–60, and TRA-1–81 were detected at low or moderate levels as indicated by immunofluorescence staining. RT-PCR analysis with Human OSKM factors Expression qBiomarker iPSC PCR array identified endogenous overexpression of only *SOX2* but not the other three factors (OKM) in MCF7/Rep cells, indicating clones isolated were not *bona fide* iPSCs. However, MCF7/Rep cells acquired CSC-like properties, including high levels of aldehyde dehydrogenase 1/3 (ALDH1/3) and CD44, markers of breast cancer stem cells (CSCs) (Al-Hajj et al., 2003a, Ginestier et al., 2007, Badve and Nakshatri, 2012, Liu et al., 2014d, Horimoto et al., 2016). In this study, the reprogrammed MCF7 (5PLA_MCF7R) expressed all pluripotency markers, including *NANOG, OCT3/4, DPPA4* and *DNMT3B*, except for *REX1* as indicated by RT-PCR analysis and high SSEA4, NANOG but low TRA-1-60 by immunofluorescence staining.

4.3.1 Reprogramming of cancer cells

4.3.1.1 Reprogramming efficiency and cancer cell refractoriness to reprogramming

The reprogramming methods, the combination of transcription factors, the somatic cell types and the developmental stage of cells are among the most important factors affecting reprogramming efficiency (Stadtfeld and

Hochedlinger, 2010). In addition, the natural expression of one or more genes making up the Yamanaka cocktail in the parental cells, which often correlates with their developmental stages could also have pronounced effect on successful iPSC derivation. For example, the high expression of *c-MYC* and *KLF4* in keratinocytes alongside transcripts commonly expressed in its iPSC counterpart and hESCs (as revealed by genome-wide transcriptional analysis) contributed to the high amenability of keratinocytes to OSKM reprogramming (with efficiency ~100-fold and kinetic ~2-fold higher than fibroblasts) (Aasen et al., 2008, Aasen and Izpisua Belmonte, 2010). In addition, mouse neural stem cells (mNSCs) express *Sox2, Klf4* and *c-Myc*. The reprogramming of mNSCs was found 7 – 45 times higher than fibroblasts and Sox2 and c-Myc were dispensable in mNSCs reprogramming (Kim et al., 2008b, Kim et al., 2009b). On the other hand, despite having a global transcriptional profile less similar to ESCs as compared to fibroblasts, the *KLF4*- and *c-MYC*-expressing $CD133^{+}$ cord blood cells were successfully reprogrammed by only two factors (*OCT3/4 and SOX2*), while fibroblasts failed to generate any iPSCs under the same condition (Giorgetti et al., 2009). Lastly, the *Sox2*-expressing malignant mouse melanoma cell line R545 was successfully reprogramed by using only the OKM factors (Utikal et al., 2009a).

In breast cancer cells, the expression of pluripotency markers has been reported and correlated with CSC properties, high grade breast cancers and poor prognosis for patients (Jin et al., 1999, Ezeh et al., 2005, Tai et al., 2005, Zhang et al., 2006, Yu et al., 2011a, Liu et al., 2014a). However, the expression of these markers in particular *KLF4, c-MYC*, *NANOG* and *OCT3/4* in breast cancer cell lines remains controversial (Wang

et al., 2005, Yu et al., 2011a). To date, three major isoforms of *OCT3/4* has been described, namely *OCT4A, OCT4B* and *OCT4B1*. Of these isoforms, only *OCT4A* has been detected in pluripotent cells and localised predominantly in the nucleus (Zhao et al., 2011, Hayashi et al., 2015, Villodre et al., 2016). Wang et al. (2009) reported the detection of only a basal level of *OCT3/4B* (similar to normal cell lines HEK293 and fibroblast) localised in the cell cytoplasm and no *OCT3/4A* in MCF7. Cantz et al. (2008) was also unable to detect OCT3/4 in MCF7 with immunofluorescence staining, Western blot, and RT-PCR analyses (primer used spanned exons and specific for *OCT3/4A* but could also target *CLEC4A* intron sequence) (Appendix Figure 4S-3 & 4S-4).

On the contrary, Kim and Nam (2011) suggested that the expression of *OCT3/4* in multiple breast cancer cell lines including MCF7 and MDA-MB-231 was correlated with CSC properties. However, the PCR primer used in that particular study was specific for mouse *Oct3/4* and did not span exons based on the University of California Santa Cruz Genome Browser in-silico PCR analysis and supported by primer-BLAST analysis (Appendix Figure 4S-5) (Karolchik et al., 2007). Moreover, Ezeh et al. (2005) reported the expression of *OCT3/4* in MCF7 by using only RT-PCR with a primer set specific for *OCT3/4A* and spanned exons but could also amplify several other genomic genes and no reverse transcriptase negative (RT-) control was shown in the analysis (Appendix Figure 4S-6). In addition, Hu et al. (2008) demonstrated that the small interference RNA (siRNA) knockdown of *OCT3/4* induced apoptosis and inhibited tumour growth in murine Lewis lung carcinoma 3LL and MCF7 in *in vitro* and *in vivo* models. However, despite

claiming all MCF7 cells expressed *OCT3/4 in vitro*, no data was shown in that particular study and only a primer set specific for mouse *Oct3/4* was described in the methodology (Appendix Figure 4S-7). In this study, no expression of *OCT3/4* was detected in all breast cancer cell lines using a selective primer set derived from Liedtke et al. (2007), which targets specifically the "parental" *OCT3/4A* gene responsible for pluripotency (Appendix Figure 4S-3).

On the other hand, no *NANOG P8* expression was detected in MCF7 in this study using the NF2R3 primer reported by Jeter et al. (2009). As discussed in the previous chapter, Jeter et al. (2009) reported the expression of *NANOG P8* in MCF7 by analysing the sequence of amplicons generated by a universal indiscriminating primer (LDF1/LDR1). There was no result showing NF2R3 primer was used for the detection of *NANOG P8* in that particular study. Nevertheless, Sanger sequencing of PCR product generated using NF2R3 primer from 5PLA_MCF7R indicated that *NANOG P8* instead of *NANOG* was expressed (Appendix Figure 4S-8). At the beginning of the present study, another PCR primer specific for *NANOG* called the N444 was also tested to detect *NANOG* expression (Appendix Figure S3-2). Consistent with the RT-PCR result reported by Zhang et al. (2006), no *NANOG* was detected in MCF7 using this primer. However, in that same study, when RT-PCR product was amplified by a universal *NANOG* primer, cloned and sequenced, of the fifteen clones two were identified as *NANOG* and two *NANOG P8*, suggesting result inconsistency. In addition, RT-PCR for *NANOG* expression in reprogrammed MCF7 (5PLA_MCF7R) with the N444 showed that 5PLA_MCF7R was positive for *NANOG* (Appendix Figure 4S-9),

which contradicted the result obtained from NF2R2 and NF2R3 primers (both indicated only *NANOG P8* expression). Owing to its inconsistency, the N444 primer was not further used in this study for PSC characterisation. It is also worth mentioning that, owing to the low RT-qPCR efficiency of NF2R3 primer and incapability to generate a reliable melting curve (Appendix Figure S2-6 & S2-7), an universal *NANOG* primer (called NANOG_all) reported by (Irie et al., 2015) was used for RT-qPCR analysis of breast cancer iPSC.

Despite the controversies, it was apparent that the alleged expression of various pluripotency markers did not enhance the reprogramming efficiency in all three breast cancer cell lines in this study as compared to their normal counterpart, the MCF10A. No iPSCs were isolated from MDA-MB-231 and iPSCs were only derived from BT-549 and MCF7 when the overall reprogramming efficiency was improved by the addition of extra factors (the miR-302/367 cluster and sodium butyrate). The low efficiency of breast cancer cell reprogramming is consistent with the majority of the cancer cell reprogramming studies using iPSC technology (Mathieu et al., 2011, Hu et al., 2011, Kim et al., 2013, Liu et al., 2014b). However, few studies reported similar (Corominas-Faja et al., 2013, Gandre-Babbe et al., 2013, Stricker et al., 2013) or higher efficiency of cancer cell reprogramming (Zhang et al., 2013, Choong et al., 2014) in comparison to their normal counterparts. Zhang et al. (2013) attributed the enhanced efficiency to the expression of pluripotency markers, including *c-MYC, KLF4* and *SOX2*, in the parental sarcoma cells as well as the absence of tumour suppressors. In this regard, as previously discussed, downregulation of *TP53*, whose activation acts as a barrier for iPSC reprogramming by

restricting cell proliferation and promoting apoptosis, enhances reprogramming efficiency (Banito and Gil, 2010). MDA-MB-231 and BT549 (but not MCF7) are both known to harbour *TP53* mutation at R280K and R249S respectively, resulting in the production of mutant and non-functional TP53 proteins. This could only explain the efficiency discrepancy between MCF7 and BT-549 reprogramming but not between MDA-MB-231 and MCF7. On the other hand, Choong et al. (2014) failed to derive iPSC from an osteoblast cell line hFOB, claiming it was more terminally differentiated than its cancerous counterparts, including four osteosarcoma cell lines (Saos-2, MG-63, G-292 and U-2 OS), and thus less amendable to the induction of pluripotency.

In a review of cancer cell reprogramming, Papapetrou (2016) highlighted the difficulty to fully elucidate the mechanisms underlying the refractoriness of cancer cells to iPSC reprogramming. However, several factors have been speculated to have contributed to these findings. First, the cancer type may play a role. Indeed, a wide range of blood malignancies have demonstrated high amenability to iPSC reprogramming, osteosarcoma consistently showed high reprogramming efficiency as compared to its normal counterpart (eg. osteoblasts). However, exception did exist for a specific subtype of blood malignancy, the B-cell acute lymphoid leukaemia (B-ALL) harbouring mixed-lineage leukaemia 1 (MLL1) gene rearrangement (MLL-rearranged B-ALL) as reported by Munoz-Lopez et al. (2016). The authors failed to derive iPSCs from MLL-rearranged B-ALL despite having performed a wide-range of reprogramming approaches and added various reprogramming "boosters". Global transcriptome and DNA methylome

analyses suggested that the intrinsic differentiation blockage of MLL-rearranged B-ALL resulted in its refractoriness to acquire pluripotency. In addition, it has also been suggested the pre-existing rearrangement mutation was incompatible with pluripotency based on preliminary data (CRISPR-Cas9-edited iPSCs harbouring t(4;11)/MLL-AF4^{+} was lost after few passages). Second, the refractoriness may be cancer mutation-specific. In agreement, patients with distinct myelodysplastic syndrome (MDS)-associated genetic lesions were found having different degrees of reprogramming barriers. In some cases, the reprogramming potential was completely compromised (Kotini et al., 2017).

In addition, the inactivating mutations observed in Fanconi anemia (FA) were found detrimental to reprogramming (Muller et al., 2012). These mutations have significant impact on a signalling pathway responsible for the maintenance of genomic integrity and DNA repair. It has been suggested that the overall instability of the FA cells genome adversely affects the reprogramming efficiency by augmenting cellular and genomic damages during reprogramming (Muller et al., 2012, Navarro et al., 2014, Rio et al., 2014). However, this speculation did not seem to apply to the reprogramming of breast cancer cells in this study. Despite exhibiting overall refractoriness to reprogramming, the intrinsically similar triple negative/ basal breast cancer cell lines, BT-549 and MDA-MB-231 responded differently to iPSC reprogramming.

On the other hand, it has also been suggested that the compromised integrity of cancer genomes could result in their refractoriness to iPSC reprogramming. Kim and Zaret (2015) highlighted the cellular stress responses elicited by aneuploidy (thought to drive genomic instability) (Sheltzer et al., 2011) during reprogramming and the senescence in patient-derived cancer cells induced by previous chemotherapy and irradiation treatment might have potential effect on cancer cell refractoriness. Lai et al. (2013) also suggested that cells at late stages of cancer are less favourable for reprogramming partially due to the presence of more damaging driver mutations, such as *TP53* and *CDKN2A*, which compromise genome integrity and impeded the regulation of pluripotency circuitry. Importantly, Kim et al., 2013 reported the isolation of only one out of eight iPSC lines harbouring the $KRAS^{G12D}$ driver mutation despite having 78% of the starting pancreatic cancer cells positive for the mutation. In this study, *KRAS* mutation has only been reported in MDA-MB-231 (at G13D) (Hollestelle et al., 2007). This might explain the refractoriness of MDA-MB-231 to reprogramming as compared BT-549 and MCF7.

4.3.1.2 Improved reprogramming outcome by additional factors: microRNA-302/367 (miR-302/367) and sodium butyrate

Reprogramming to iPSC is a multistep process and not all cells can be fully reprogrammed to an ESC-like state. Partial reprogramming can give rise to pre-iPSC cells with some degree of pluripotency but incomparable to ESC (Mikkelsen et al., 2008a, Theunissen et al., 2011). Theunissen et al. (2011) reported the isolation of pre-iPSC lines from MEFs and neural stem cells which were exogenous Oct4-dependent. They expressed pluripotency genes

Fgf4 and *Nr0b1*, but not genes of the core pluripotency network (eg. *Nanog* and *Rex1*). Treatment with potent DNA demethylating agent, 5-Azacytidine (5-aza) successfully converted these cells into fully-reprogrammed iPSCs, expressing *Nanog* and *Rex1* and became independent of exogenous *Oct4* (Theunissen et al., 2011). On the other hand, Mikkelsen et al. (2008) converted three pre-iPSCs with different gene expression profiles and morphological similarities to ESCs isolated from MEFs to fully-reprogrammed iPSCs using the same approach plus downregulation of DNA methyltransferase 1 (Dnmt1) by siRNA or shRNA (Mikkelsen et al., 2008).

MicroRNAs (miRs or mirs) are major component of the epigenetic mechanisms alongside DNA methylation and histone modification (Vasilatou et al., 2013), which regulates gene expression at the posttranscriptional level. MiRs have been implicated in the maintenance of self-renewal and pluripotency of stem cells. Melton et al. (2010) reported the antagonistic roles of *let-7* and ESC cell cycle regulating (ESCC) mirs (*mir-291-3p*, *mir-294* and *mir-295* in mouse) in stabilizing the differentiated and self-renewing cell fates through a common pathway respectively. In addition, c-Myc-induced miRs, p53-targeting miRs, and early embryonic miR cluster (EEmiRC) have also been found responsible for the proliferation, inhibition of differentiation and the maintenance of pluripotency in stem cells (Lin et al., 2009, Neveu et al., 2010, Tata et al., 2011).

The human *miR-302/367* cluster, encompassing *miR302a, miR302a*, miR302b, miR302b*, miR302c, miR302c*, miR302d* and *miR367*, was identified as the most abundantly expressed miRNA cluster in hESCs (Suh

et al., 2004, Bar et al., 2008). The human orthologues of the ESCC miRNAs includes members of the human miR-302 cluster (*hsa-miR-302a–d*) and *hsa-miR-372* and *373* (Bar et al., 2008). It has been suggested that during embryonic development, the *miR-302s* target more than 445 genes involved in lineage-specific cell differentiation to prevent premature cell differentiation (Lin et al., 2008). Subsequent studies indicated that the *miR-302/367* cluster maintains hESC self-renewal and pluripotency through 1) the promotion of G1/S phase transition and cell proliferation (Wang et al., 2008); 2) the regulation of cell cycle and apoptosis in a dose dependent manner (Zhang et al., 2015); 3) and the targeting of inhibitors and thus activation of the BMP and TGF-β/Activin/Nodal signalling pathways, which in turn inhibit neural induction and promotes pluripotency as well as other additional pathways such as vesicular transfer, chromatin regulators and cytoskeletal modulators (Lipchina et al., 2011).

Considering its vital role in regulating stem cell pluripotency, it is not surprising that the *miR-302/367* cluster could significantly affect iPSC reprogramming. Barroso-delJesus et al. (2008) identified Nanog, Oct3/4, Sox2, and Rex1 as upstream regulators of *miR302-367* promoter (Card et al., 2008). The overexpression of *hsa-miR-302b* and/or *372* greatly enhanced the reprogramming efficiency of human foreskin (BJ) and lung fibroblasts (MRC-5) through the promotion and inhibition of mesenchymal-epithelial transition (MET) and epithelial-mesenchymal transition (EMT) respectively, by repressing TGFβR2 (Subramanyam et al., 2011). This finding was also supported by Liao et al. (2011) by using the *miR-302/367* cluster. Judson et al. (2009) demonstrated that the member of mouse ESCC

mirs, *mir-294* (to which the human *hsa-miR-302b* and *372* share the same seed sequence with) could substitute c-Myc in the OSKM reprogramming cocktail, despite with 75% of the original efficiency. Banito et al. (2009) showed that the *miR-302* cluster diminished the effect of reprogramming-induced senescence (a reprogramming barrier) and enhanced reprogramming efficiency by targeting the negative cell cycle regulators $p21^{CIP1}$ and p130.

More strikingly, several reports demonstrated the successful derivation of iPSCs from mouse and human somatic and cancer cells by solely overexpressing miR clusters. The overexpression of *miR-302* in human hair follicle cells facilitated iPSCs formation potentially through simultaneous suppression of epigenetic regulators, namely AOF2, AOF1, MECP1-p66 and MECP2 (Lin et al., 2011). Anokye-Danso et al. (2011) reprogrammed mouse and human fibroblast with *miR-302/367* cluster at efficiency approximately two orders of magnitudes higher than the OSKM-mediated reprogramming. In addition, Miyoshi et al. (2011) successfully reprogrammed mouse and human somatic cells by using a combination of *mir-200c* plus *mir-302s* and *mir-369s* family miRs. Moreover, Lin et al. (2008) also derived iPSCs from the human melanoma cell line through overexpression of the *miR-302* family. In this study, the addition of the *miR302/367* cluster (derived from Anokye-Danso et al. (2011)) to the OSKML cocktail permitted the derivation of iPSCs from two breast cancer cell lines (the BT-549 and MCF7) of two different subtypes (TNBC and ER^+).

In this project, sodium butyrate (NaB) was used alongside the *miR-302/367* cluster to facilitate the reprogramming of breast cancer cells. NaB (the sodium salt of short-chain fatty acid butyrate) as well as other histone deacetylase inhibitors (HDACi), namely Trichostatin A and valproic acid, have been shown to enhance iPSC reprogramming efficiency and facilitate the transition of partially reprogrammed cells to *bona fide* iPSCs (Boffa et al., 1978, Ware et al., 2009, Liang et al., 2010, Wen et al., 2016a). NaB is known to have many biological effects on mammalian cells including the inhibition of proliferation and induction of differentiation through gene expression activation and repression at low concentration (Kruh, 1982, Barnard and Warwick, 1993). Treatment with NaB improved the reprogramming efficiency of human adult or foetal fibroblasts by 15- to 51-fold or higher in the OSK or OSM reprogramming conditions, potentially through the promotion of histone H3 acetylation, promoter DNA demethylation, and the expression of pluripotency genes, (Mali et al., 2010). By using mouse fibroblasts, Liang et al. (2010) also reported the potential of NaB to upregulate a number of ESC-associated genes at the early stage of reprogramming mediated by c-Myc to bring about efficient iPSC derivation (improved by as high as 7-fold). Interestingly, Zhang and Wu (2013) demonstrated that NaB acted individually or in concert with two other small molecules (SB431542 and PD0325901, which are inhibitors of TGF-β and MAPK/ERK) to stabilise and upregulate the expression of the miR302/367 cluster. In support to this finding, the authors later demonstrated that NaB induced the expression of *miR-302/367* cluster by interfering the interaction between HDCAs and Oct3/4 bound to its promoter (Zhang et al., 2014b). In a separate study, the authors showed that NaB

was able to rescue the *miR-302/367* cluster-knockdown hESCs from apoptosis by upregulating *miR-302/367* cluster expression (Zhang et al., 2015), confirming the positive relationship between the two factors. In addition, NaB has also been shown to facilitate the transition of pre-iPSC to *bona fide* iPSCs (Kang et al., 2014, Wen et al., 2016b). This proven effect of NaB could have been instrumented in this study considering reprogrammed cell lines (in particular the FMM_BTR1 and FMM_BTR2) generated from the breast cancer cell lines were shown to harbour characteristics of pre-iPSCs.

Considering the fact that *miR302/367* cluster by itself could initiate the reprogramming process, it might have driven the reprogramming of breast cancer cells individually. However, taking into account of the relationship between *miR302/367* cluster and NaB, both factors might have acted synergistically to bring about the formation of iPSCs. Besides, it is also possible that the two factors "loosen" the rigid and mutation-teeming breast cancer cell epigenome and genome individually (*miR302/367* at early stage and NaB at middle stage of the reprogramming) or cooperatively, making them more accessible and amendable to the OSKML factors. No definite conclusion can be drawn at this stage and more reprogramming experiments should be performed in the future by using different combinations of the OSKML factors, *miR302/367* cluster and NaB treatment to identify the true driver/s of the breast cancer cell reprogramming process. Nevertheless, the present study showed that the reprogramming of breast cancer cells is possible under favourable conditions that combine the overexpression of pluripotency factors and chromatin modifiers. Importantly,

however, is that the study showed that reprogramming of breast cancer cells is possible under favourable condition that combine the overexpression of pluripotency factors and chromatin modifiers.

5. REPROGRAMMING MCF10A MUTANTS HARBOURING PIK3CA$^{H1047R+/-}$ AND PIK3CA$^{H1047R+/-}$_TP53$^{-/-}$ MUTATIONS

5.1 INTRODUCTION

The aim of this chapter was to generate two mutant cell lines from MCF10A, namely MCF10A_PIK3CA$^{H1047R+/-}$ and MCF10A_PIK3CA$^{H1047R+/-}$_TP53$^{-/-}$, harbouring the two most commonly mutated genes in breast cancer, *PIK3CA* and *TP53*. The reprogramming of these isogenic cell lines into iPSCs and differentiation into different lineages could help elucidate the potential effects of the two driver mutations on epigenetic regulation/dysregulation in breast cancer. In addition, by rewriting the epigenome of the mutant cell lines, it was speculated that possible epigenetic interventions can be identified to suppress the oncogenic phenotypes conferred by the mutations as previously reported (Croessmann et al., 2017). Last but not least, the cell lines could serve as good cancer models to study the mechanism of action of *PIK3CA* and *TP53* mutations in breast carcinogenesis as well as for screening, discovery and development of effective anti-cancer therapies.

5.1.1 Breast cancer driver mutations

The cancer genome harbours a wide variety of genetic mutations including substitutions, insertions, deletions, copy number variation and chromosomal rearrangement, affecting many genes as well as non-coding regions. In addition, a myriad of epigenetic mutations has also been reported. Together with the influence of tumour microenvironment and cell/tissue differentiation states and other factors affecting the manifestation of the aforementioned mutations, these factors collectively

contribute to the complexity of the cancer genome (Chin et al., 2011, Ajani, 2013).

Driver mutations are genetic alterations that directly confer cancerous properties to the affected cells. On the contrary, mutations that do not directly contribute to cell abnormality and often occur coincidently or as a result of the driver mutations are termed passenger mutations (Bozic et al., 2010). Large scale screenings at the genomic, transcriptomic, proteomic and epigenomic levels have provided valuable insights into the contribution of mutations of different genes to the initiation and progression of breast cancer and successfully identified crucial driver mutations (Nik-Zainal et al., 2016). According to the Cancer Genome Atlas (2012), only mutations in *TP53, PIK3CA* and *GATA3* have >10% occurrence across all breast cancer subtypes, whereas the rest are subtypes-associated, supporting the notion that breast cancer is a heterogeneous disease (Polyak, 2011). The most frequently mutated genes in luminal A tumour were identified as *PIK3CA* (45%) followed by *GATA3 (14%), MAP3K1* (13%)*, TP53 (12%), CDH1* (9%), while *TP53* and *PIK3CA* (29% respectively) were reported in luminal B tumours followed by *GATA3* (15%). *TP53* mutation was detected in 80% and *PIK3CA* (the second most commonly mutated gene) in 9% of basal-like tumours. In HER2-enriched subtype, *HER2* amplification (80%) was the most frequent mutation and the frequency of *TP53* and *PIK3CA* mutations were 72% (second) and 39% (third SGM). In support to the above findings, sequencing of 173 genes in 2433 primary breast tumours identified *PIK3CA* (40.1%), *TP53* (35.4%) and *GATA3* (11.1%) along with four other genes, namely *MUC16* (16.8%), *AHNAK2*

(16.2%), *SYNE1* (12.0%) and *MLL3* (11.4%) as genes having a mutation frequency of >10% across all samples (Pereira et al., 2016a). In addition, Nik-Zainal et al. (2016) also identified *TP53* and *PIK3CA* as the two most frequently mutated genes in breast cancers, having frequencies of 20% and 40% and 80% and 20% in ER^+ and ER^- tumours respectively through whole-genome sequencing of 560 breast cancers.

As previously described in Section 3.3.3, *TP53* is a tumour suppressor gene responsible for the maintenance of tissue homeostasis (Fridman and Lowe, 2003). The inactivation of p53 by mutation in cancer compromises regulatory mechanisms and prevents the elimination of cells that sustain DNA damage and cellular stresses, eventually leading to uncontrolled growth and carcinogenesis (Brooks and Gu, 2010). On the other hand, the phosphatidylinositol 3-kinases (PI3Ks) are a family of lipid kinases known to phosphorylate the 3'-OH of phosphoinositides. It plays a vital role in the PI3K/PTEN/AKT signalling pathway, which involves in numerous essential cellular processes, including cell survival, proliferation, metabolism, cytoskeletal rearrangement and membrane trafficking (Vivanco and Sawyers, 2002). Activating mutation of PIK3CA gene (encoding p110α isoform of catalytic subunit) has been associated extensively reported in various human cancers including breast cancer (Croessmann et al., 2017, Karakas et al., 2006).

5.1.2 Generation of mutants harbouring driver mutations by CRISPR-Cas9 technology

The identification of driver mutations in cancers via genome-wide analyses and the availability of genome-editing tools with great precision, efficiency and simplicity have provided great opportunities to generate models to investigate the mechanisms underlying carcinogenesis. For instance, a wide range of cancer models have been established through the introduction of cancer-specific driver mutation/s into different normal cell types using the CRISPR-Cas9 technology (Drost et al., 2015, Matano et al., 2015, Luo, 2016). Matano et al. (2015) generated a human colorectal cancer model by introducing *APC, SMAD4, TP53, KRAS* and *PIK3CA* mutations into organoids derived from human intestinal epithelium using the CRISPR-Cas9 system. Study of this cancer model suggested that driver mutations contribute to stem cell maintenance but are insufficient to initiate malignant progression, implying a need for the elucidation of functional contribution of other molecular lesions such as epigenetic alterations (Matano et al., 2015).

To achieve the objectives of this chapter, the CRISPR-Cas9 system was used to create two cell lines harbouring $PIK3CA^{H1047(+/-)}$ and $TP53^{(-/-)}$ mutations from MCF10A. These mutant lines were subjected to five episomal plasmids reprogramming into iPSCs under the same experimental conditions reported in the previous chapter. Subsequently, the derived iPSCs were characterised by RT-PCR and immunofluorescence analyses.

5.2 RESULTS

5.2.1 Generation of PIK3CA$^{H1047(+/-)}$ mutant cell line with CRISPR-Cas9

A *PIK3CA* mutation was introduced into a mutation hot spot in exon 21, where a point mutation (CAT → CGT or c.3140A>G) results in one amino acid change from histidine to arginine (p.H1047R). The H1047R mutation is a gain-of-function mutation which has been shown to be an important driver in a wide range of cancers, including breast cancer (Croessmann et al., 2017). One 20bp long short guide RNA (sgRNA) (CAAATGAATGATGCACATCA) was designed by using an online CRISPR tool (http://crispr.mit.edu). The sgRNA sequence was selected based on high quality score and low off-target effect (Figure 5S-1). A 141 bp long single-stranded oligodeoxynucleotide (ssODN) template was designed spanning the mutation site and the sgRNA sequence to induce homology-directed repair (HDR). The ssODN contained the mutated base (from A>G) and a silent mutation that abolished the BspH1 restriction cutting site (5'-TCATGA-3' → 5'-TCATAA-3') present in the sequence to enable restriction enzyme screening for mutant clones (Figure 5.1).

XP_011511196.1 PREDICTED: phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha isoform isoform X1 [Homo sapiens]

MPPRPSSGELWGIHLMPPRILVECLLPNGMIVTLECLREATLITIKHELFKEARKYPLHQLLQDESSYIF
VSVTQEAEREEFFDETRRLCDLRLFQPFLKVIEPVGNREEKILNREIGFAIGMPVCEFDMVKDPEVQDFR
RNILNVCKEAVDLRDLNSPHSRAMYVYPPNVESSPELPKHIYNKLDKGQIIVVIWVIVSPNNDKQKYTLK
INHDCVPEQVIAEAIRKKTRSMLLSSEQLKLCVLEYQGKYILKVCGCDEYFLEKYPLSQYKYIRSCIMLG
RMPNLMLMAKESLYSQLPMDCFTMPSYSRRISTATPYMNGETSTKSLWVINSALRIKILCATYVNVNIRD
IDKIYVRTGIYHGGEPLCDNVNTQRVPCSNPRWNEWLNYDIYIPDLPRAARLCLSICSVKGRKGAKEEHC
PLAWGNINLFDYTDTLVSGKMALNLWPVPHGLEDLLNPIGVTGSNPNKETPCLELEFDWFSSVVKFPDMS
VIEEHANWSVSREAGFSYSHAGLSNRLARDNELRENDKEQLKAISTRDPLSEITEQEKDFLWSHRHYCVT
IPEILPKLLLSVKWNSRDEVAQMYCLVKDWPPIKPEQAMELLDCNYPDPMVRGFAVRCLEKYLTDDKLSQ
YLIQLVQVLKYEQYLDNLLVRFLLKKALTNQRIGHFFFWHLKSEMHNKTVSQRFGLLLESYCRACGMYLK
HLNRQVEAMEKLINLTDILKQEKKDETQKVQMKFLVEQMRRPDFMDALQGFLSPLNPAHQLGNLRLEECR
IMSSAKRPLWLNWENPDIMSELLFQNNEIIFKNGDDLRQDMLTLQIIRIMENIWQNQGLDLRMLPYGCLS
IGDCVGLIEVVRNSHTIMQIQCKGGLKGALQFNSHTLHQWLKDKNKGEIYDAAIDLFTRSCAGYCVATFI
LGIGDRHNSNIMVKDDGQLFHIDFGHFLDHKKKKFGYKRERVPFVLTQDFLIVISKGAQECTKTREFERF
QEMCYKAYLAIRQHANLFINLFSMMLGSGMPELQSFDDIAYIRKTLALDKTEQEALEYFMKQMNDA**H**HGG
WTTKMDWIFHTIKQHALN

p.H1047R

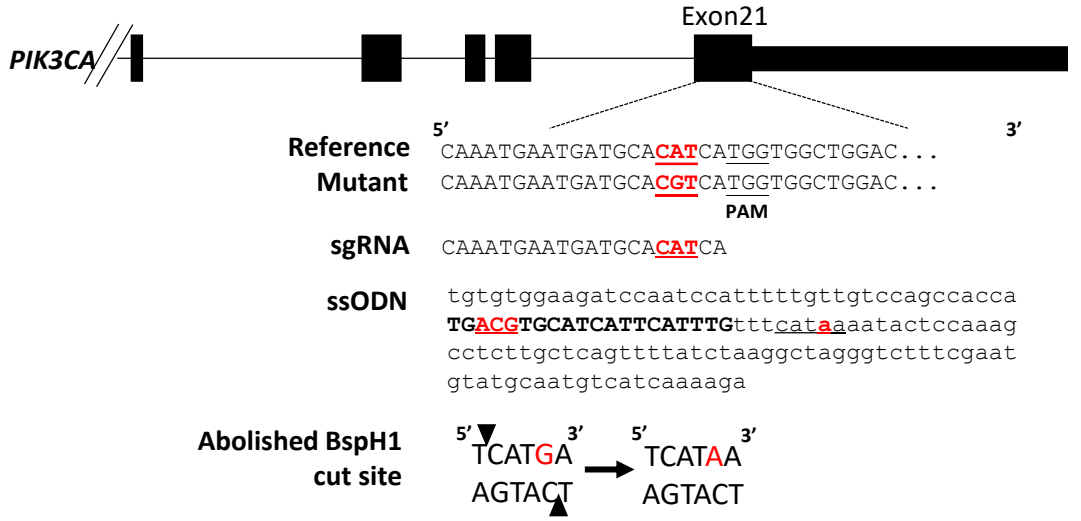


Figure 5.1 sgRNA and single-stranded oligodeoxynucleotide (ssODN) template design for *PIK3CA*$^{H1047R(+/-)}$ mutation.

The *PIK3CA* gene (XP_011511196.1) locates in chromosome 3q26.3 and the p.H1047R mutation occurs in exon 21 of the gene. A point mutation, from CAT (reference or WT) to CGT (mutant) (c.3140A>G) results in one amino acid change from histidine to arginine (p.H1047R) (black arrow). A sgRNA located immediately upstream of a "TGG" protospacer adjacent motif (PAM) site spanning the mutation base was designed. A ssODN harbouring the mutation of interest (underlined, capital letter, in red) and a silent mutation (small letter and underlined) abolishing a BspH1 site was used for HDR and streamline screening of mutant clones. sgRNA = short guide RNA, ssODN = single-stranded oligodeoxynucleotide.

5.2.2 CRISPR-Cas9 gene editing: *PIK3CA$^{H1047R(+/-)}$*

The sgRNA was first cloned into a sgRNA vector (Addgene #41824) carrying the sgRNA scaffold necessary for Cas9 enzyme binding during gene editing. The sgRNA was cloned into a BsaI site immediately upstream of the sgRNA scaffold sequence in the plasmid and a human U6 promoter was located upstream of the sgRNA. Successful insertion was confirmed by Sanger sequencing (Figure 5.3). To perform the gene editing experiment, sgRNA vector was co-transfected with plasmid (Addgene #44719) co-expressing Cas9 enzyme and GFP into MCF10A cells by using the FuGENE 6 transfection reagent. After 24 h, ssODN was transfected using the Lipofectamine 3000 Reagent. GFP^{high} cells were sorted by FACS and plated the next day as single cells on 96-well plates. Cells were allowed to expand to 80-90% confluence in epidermal growth factor negative (EGF^{-}) medium (selection against non-mutants). Each 96-well plate was passaged into two 96-well plates, one at 1:4 ratio (to be cultured continuously until screening was completed) and one at 3:4 ratio (for DNA extraction and screening). DNA extraction was performed in a 96-well format using a custom-made extraction buffer (Chapter 2.9.4) (Figure 5.4).

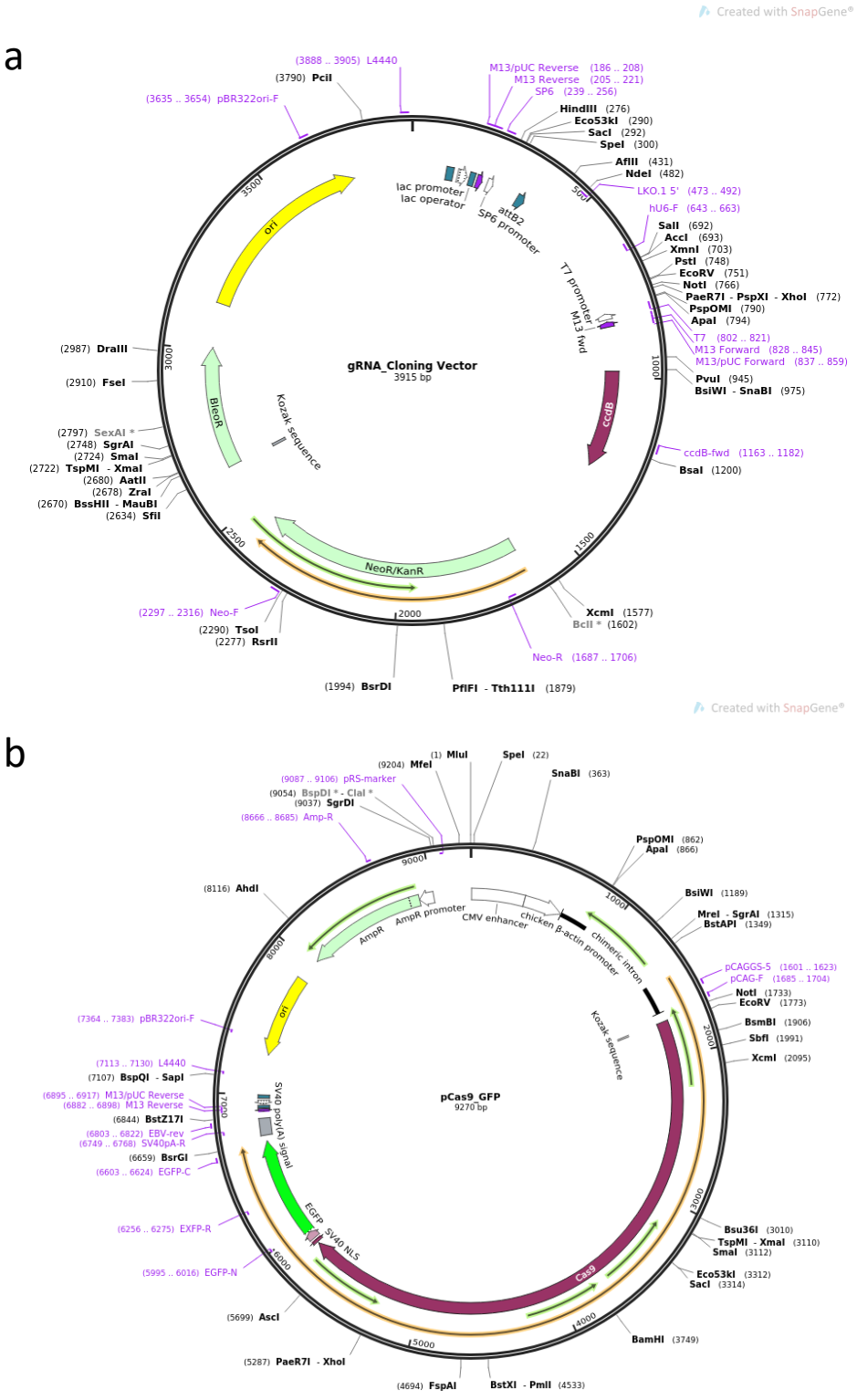


Figure 5.2 Plasmid maps of Addgene plasmid #41824 (a) and #44719 (b).

(a) Plasmid #41824 contains BsaI site for sgRNA insertion. (b) Plasmid #44719 encodes Cas9 linked to enhanced green fluorescence protein (EGFP).

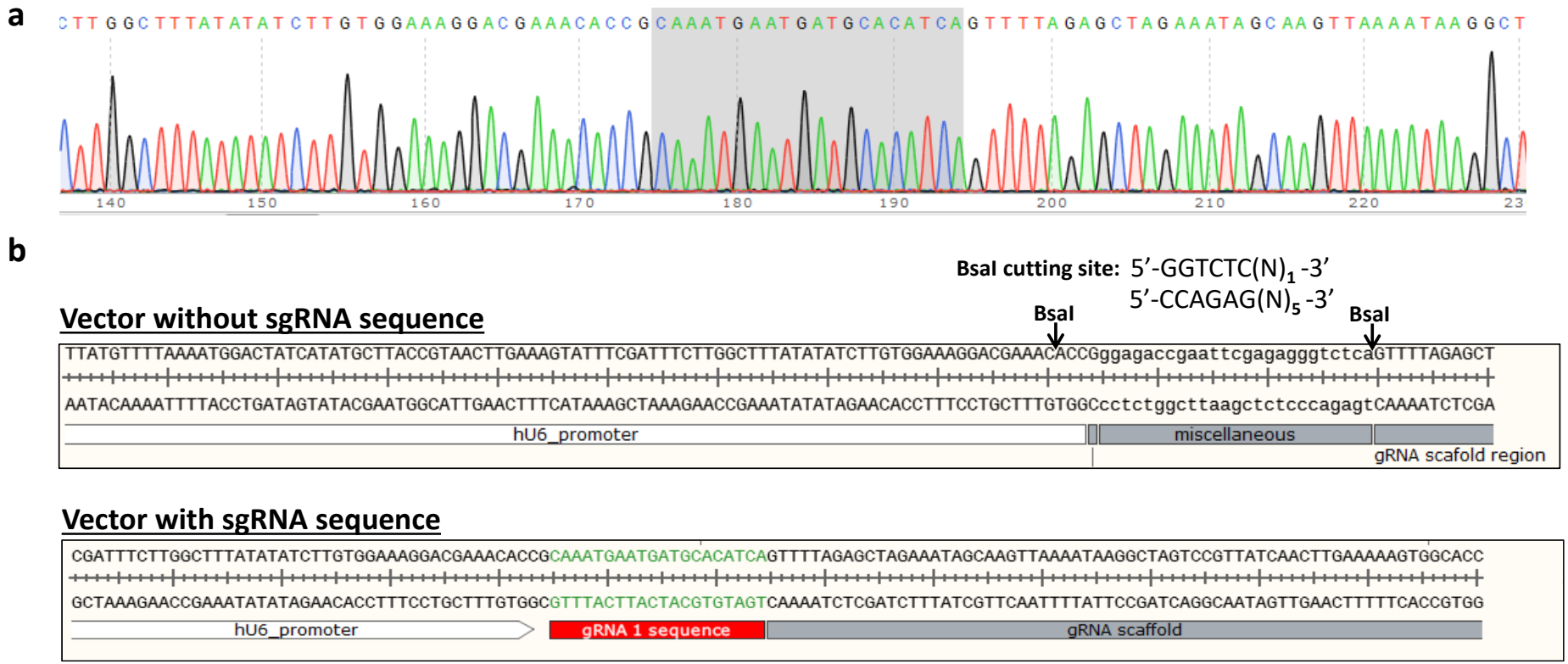


Figure 5.3 Cloning of sgRNA into sgRNA vector (Addgene #41824).

(a) DNA sequencing confirmed the insertion of sgRNA (shaded) into sgRNA vector. (b) sgRNA vector sequences without (top) and with (bottom) sgRNA sequence (bases in green) viewed on SnapGene Viewer software. The sgRNA vector carries sgRNA scaffold, a component required for Cas9 enzyme binding, and a human U6 promoter immediately downstream and upstream of sgRNA. BsaI cutting site is shown.

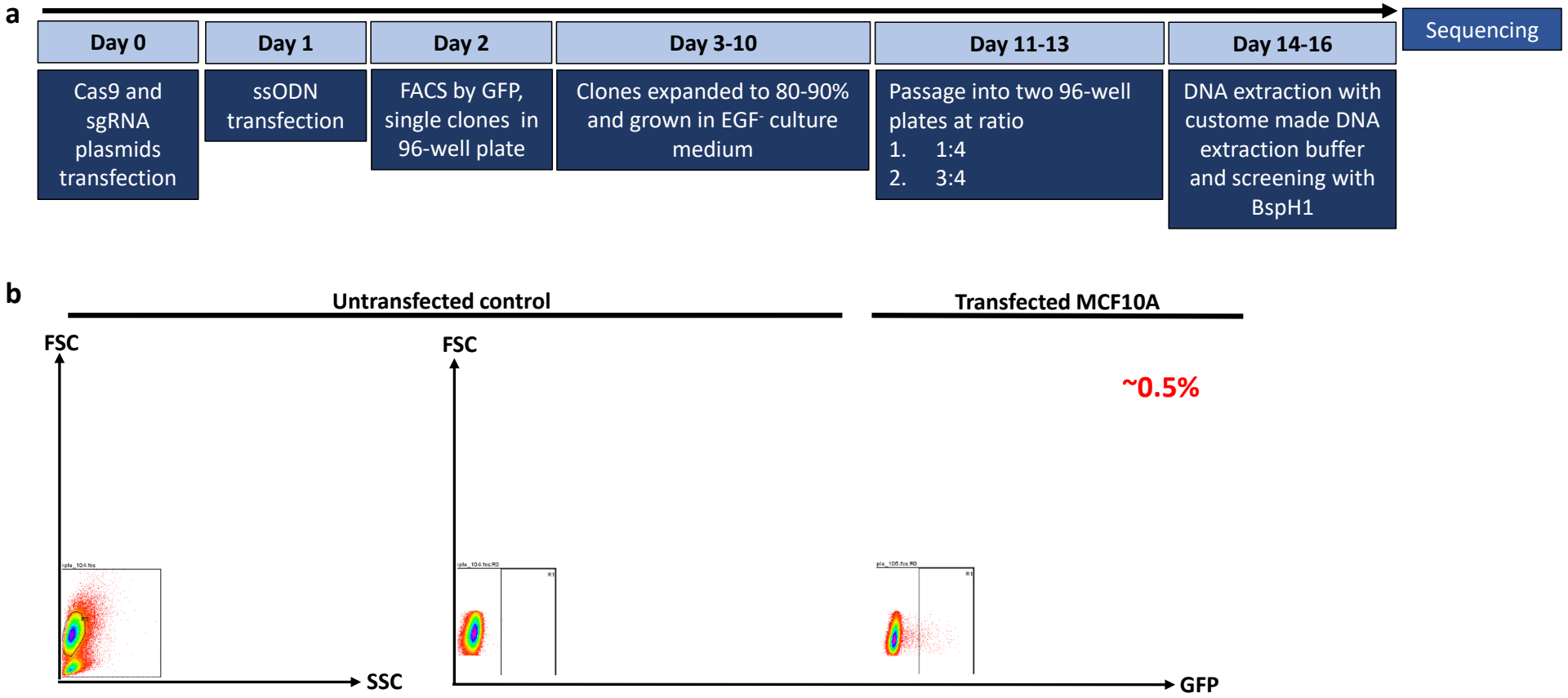


Figure 5.4 CRISPR-Cas9-mediated introduction of PIK3CA$^{H1047(+/-)}$ mutation in MCF10A.

(a) Experimental workflow of CRISPR-Cas9 gene editing. (b) FACS sorting of GFP^{high} cells at day 2. Only a small fraction of cells (~0.5%) with high GFP expression was sorted.

To screen for mutant clones, genomic DNA from each clone was amplified by PCR with primer spanning the mutation site to generate a 597 bp product. PCR products were subsequently digested with BspHI restriction enzyme into products of 421 bp and 176 bp. Before screening for individual clones, the pool GFP^{high} MCF10A cells was screened. Result indicated the presence of an uncut 597 bp band suggesting the presence of cells harbouring the desired *PIK3CA*$^{H1047R(+/-)}$ mutation (Figure 5.5). However, no positive clones were identified after screening of over 200 individual clones (Appendix Figure 5S-2), suggesting the efficiency of HDR was extremely low and more clones needed to be examined to identify the desired *PIK3CA*$^{H1047R(+/-)}$ mutant. A representative sample of clones screened is illustrated in (Figure 5.6).

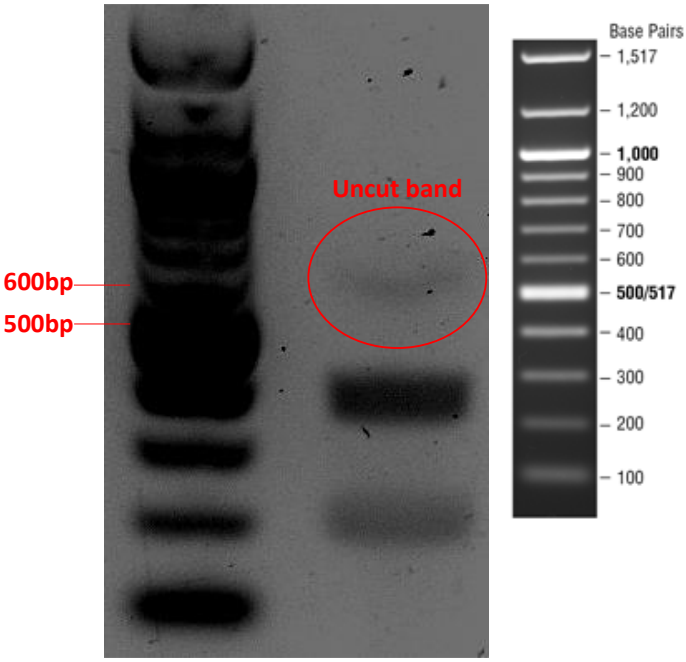


Figure 5.5 BspH1 screening of pool GFP^{high} MCF10A cells.

The presence of uncut 597 bp band indicated the presence of ssODN-directed modified MCF10A cells harbouring the abolished BspH1 cutting site and the *PIK3CA*$^{H1047R(+/-)}$ mutation. n>3 and enzyme incubation time was increased by 60 mins to prevent false positive from incomplete digestion.

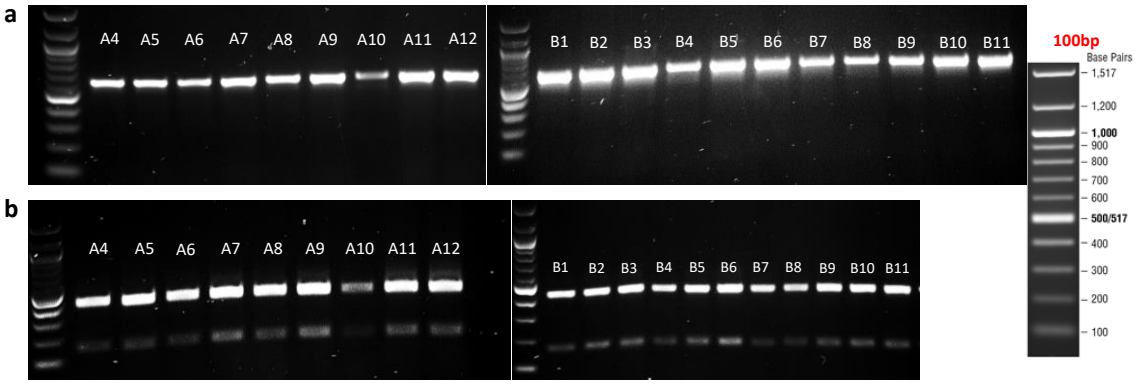


Figure 5.6 Screening of *PIK3CA* mutants.

Only a representative part of the clones was shown. (a) Genomic DNA was extracted and amplified by PCR to generate a 597 bp product. (b) BspHI was used to digest the PCR product (421 bp and 176 bp) to identify the potential abolishment of restriction site. *100 bp ladder (on the right) was used in all experiments shown.

Four clones, namely B11, E6, E11 and H3 were selected randomly for Sanger sequencing analysis. Interestingly, mixed signals were consistently detected from the Cas9 cutting site onwards (in between 4 and 5 bp upstream of the 5'-TGG-3' PAM site). Results confirmed that the BspHI cutting site has not been changed in all four clones (Figure 5.7).

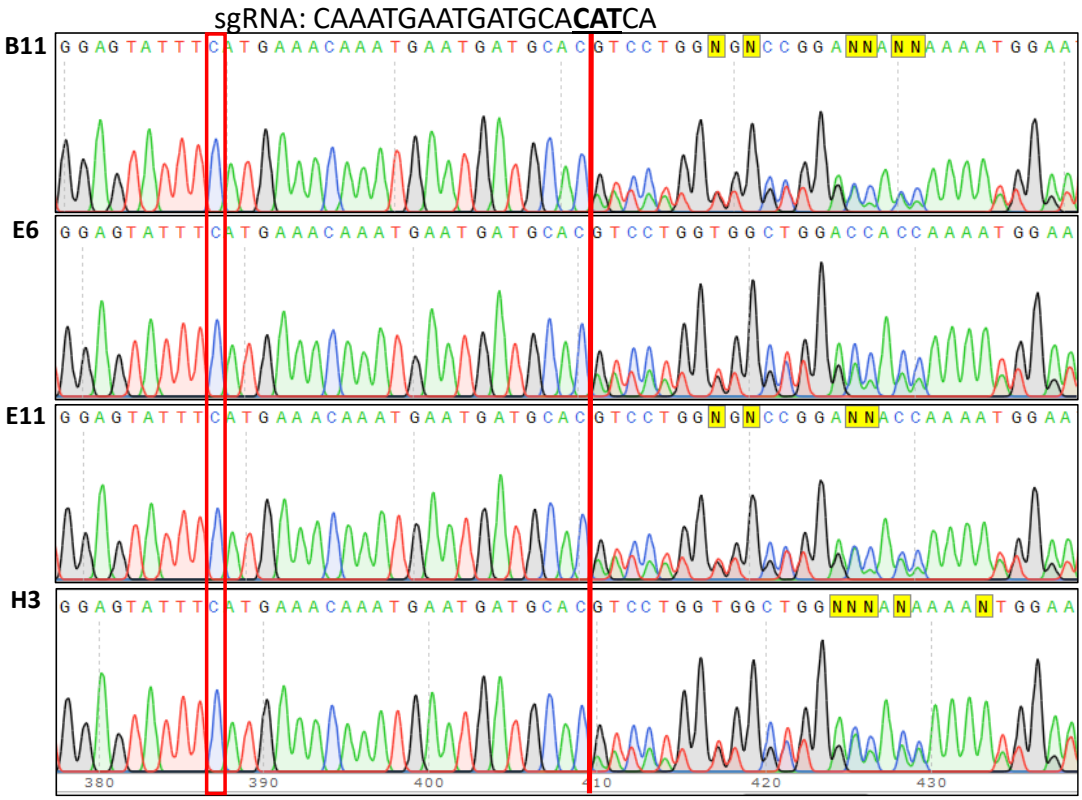


Figure 5.7 Sanger sequencing of four randomly selected GFP^{high} clones

Mixed signal was consistently observed in all four clones, B11, E6, E11 and H3 from a specific point onwards (marked by red line). This point located between 4 and 5 bp upstream of the TGG PAM site and spanned by sgRNA. The BspHI site has not been modified in all four clones (red column). Bold and underlined letters of the sgRNA were the codon of interest.

Subsequently, sequencing results were decoded by using an online tool called CRISPR-ID to identify possible CRISPR-induced mutations (Dehairs et al., 2016). An addition of G nucleotide was identified in all four clones and only B11, E6 and H3 harboured a G>A substitution. In all cases, mutation/s only happened to one allele and all clones carried one unmodified/WT allele. When the mutated allele sequences were translated into protein sequences and aligned with WT protein sequence by using an online tool (McWilliam et al., 2013), a series of protein mutations were detected (Figure 5.8). Results indicated that ssODN-mediated HDR occurred at an extremely low frequency and the non-homologous end joining (NHEJ), a more efficient but mutation-prone pathway, was employed for DNA repair, leaving the target region with random mutations. Unfortunately, time constraint did not allow the screening of more GFP^{high} clones. Instead, a commercially available MCF10A cell line harbouring the $PIK3CA^{H1047R(+/-)}$ mutation was purchased from the Horizon Discovery company. The presence of $PIK3CA^{H1047R(+/-)}$ mutation was confirmed by Sanger sequencing (Figure 5.9). The c cell line was used to create the $PIK3CA^{H1047R(+/-)}$ and $TP53^{(-/-)}$ double mutant.



Figure 5.8 CRISPR-induced mutations found in four GFP^{high} clones.

(a & c) CRISPR-ID analysis of Sanger sequencing results of (a) B3, E6, H3 and (c) E11 clones. An addition of G nucleotide and (G>A) substitution were identified in the first three clones and E11 clone only harboured the former addition mutation. In all cases, only one allele was unaffected. (b & d) When translated into protein sequences, the mutated alleles gave rise to mutated proteins as compared to the WT (in red squares). (Ref = reference sequence or wild type).

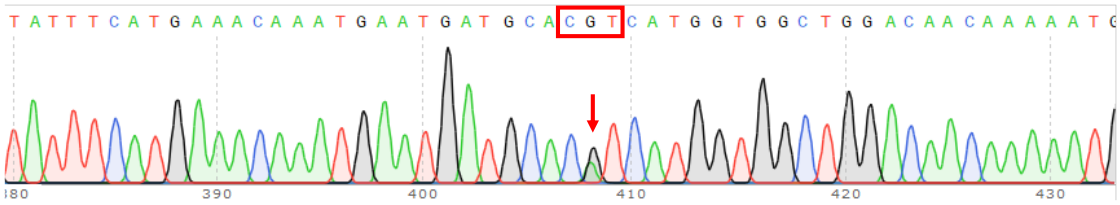


Figure 5.9 Sanger sequencing confirming the presence of CAT>CGT point mutation in MCF10A_PIK3CA$^{H1047R(+/-)}$ cell line from the Horizon Discovery.

(red square) Heterogenous CAT>CGT point mutation in Exon 21 of *PIK3CA* in MCF10A_PIK3CA$^{H1047R(+/-)}$ cell line.

5.2.3 CRISPR-Cas9 gene editing: introduction of *TP53*$^{(-/-)}$ mutation to MCF10A_PIK3CA$^{H1047R(+/-)}$ cell line

In addition to *PIK3CA* mutation, loss-of-function mutation of *TP53* is one of the major drivers of breast cancer. In order to create an *in vitro* model harbouring both mutations, this section aimed to knock-out *TP53* using the CRISPR-Cas9 system in MCF10A_PIK3CA$^{H1047R(+/-)}$ cell line described in the last section. To achieve this aim, a plasmid (Addgene #85559) carrying a sgRNA that targets exon 4 of *TP53* and the Cas9 protein was used (Figure 5.10).

NM_000546.5 Homo sapiens tumor protein p53 (TP53), transcript variant 1, mRNA

GATGGGATTGGGGTTTTCCCCTCCCATGTGCTCAAGACTGGCGCTAAAAGTTTTGAGCTTCTCAAAAGTC
TAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACTTTGCGTTCGGGCTGGGAGCGTG
CTTTCCACGACGGTGACACGCTTCCCTGGATTGGCAGCCAGACTGCCTTCCGGGTCACTGCCATGGAGGA
GCCGCAGTCAGATCCTAGCGTCGAGCCCCCTCTGAGTCAGGAAACATTTTCAGACCTATGGAAACTACTT
CCTGAAAACAACGTTCTGTCCCCCTTGCCGTCCCAAGCAATGGATGATTTGATGCTGTCCCCGGACGATA
TTGAACAATGGTTCACTGAAGACCCAGGTCCAGATGAAGCTCCCAGAATGCCAGAGGCTGCTCCCCCCGT
GGCCCCTGCACCAGCAGCTCCTACACCGGCGGCCCCTGCACCAGCCCCCTCCTGGCCCCTGTCATCTTCT
GTCCCTTCCCAGAAAACCTACCA**GGGCAGCTACGGTTTCCGTC**TGGGCTTCTTGCATTCTGGGACAGCCA
AGTCTGTGACTTGCACGTACTCCCCTGCCCTCAACAAGATGTTTTGCCAACTGGCCAAGACCTGCCCTGT
GCAGCTGTGGGTTGATTCCACACCCCCGCCCGGCACCCGCGTCCGCGCCATGGCCATCTACAAGCAGTCA
CAGCACATGACGGAGGTTGTGAGGCGCTGCCCCCACCATGAGCGCTGCTCAGATAGCGATGGTCTGGCCC
CTCCTCAGCATCTTATCCGAGTGGAAGGAAATTTGCGTGTGGAGTATTTGGATGACAGAAACACTTTTCG
ACATAGTGTGGTGGTGCCCTATGAGCCGCCTGAGGTTGGCTCTGACTGTACCACCATCCACTACAACTAC
…

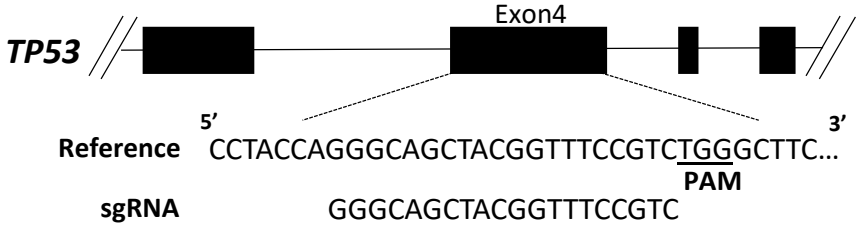


Figure 5.10 sgRNA targeting exon 4 of *TP53* gene.

(top) A partial mRNA sequence of the *TP53* gene (NM_000546.5) derived from NCBI. (bottom) The Cas9-encoding plasmid 85559 carries a sgRNA which targets the exon 4 of *TP53* gene. PAM = protospacer adjacent motif, reference = reference sequence or wild type.

All experiments were performed as described in the last section except for the absence of ssODN. In addition, cells were cultured in medium supplemented with nutlin-3 for the selection of *TP53*$^{(-/-)}$ clones for two weeks before plating as single cells in 96-well plates. Screening of mutant clones was performed by using a protocol described in Zhu et al. (2014) (Figure 5.11). To identify mutant clones, the targeted region was amplified by PCR to generate a 418 bp product. The amplicons were subsequently denatured and re-annealed to generate homoduplexes (when two complementary single stranded sequences reannealed, such as WT/WT or Mut/Mut) and heteroduplexes (when two non-complementary single stranded sequences reannealed, such as WT/Mut or Mut1/Mut2). When resolved by polyacrylamide gel electrophoresis (PAGE, heteroduplexes move relatively slower owing to the presence of an open angle between matched and mismatched bases and homoduplexes of Mut/Mut migrate at a slightly different rate from the WT/WT depending on the mutations. Of the sixteen clones analysed, twelve potential mutant clones harbouring distinct PAGE profiles from the control were identified (Figure 5.12).

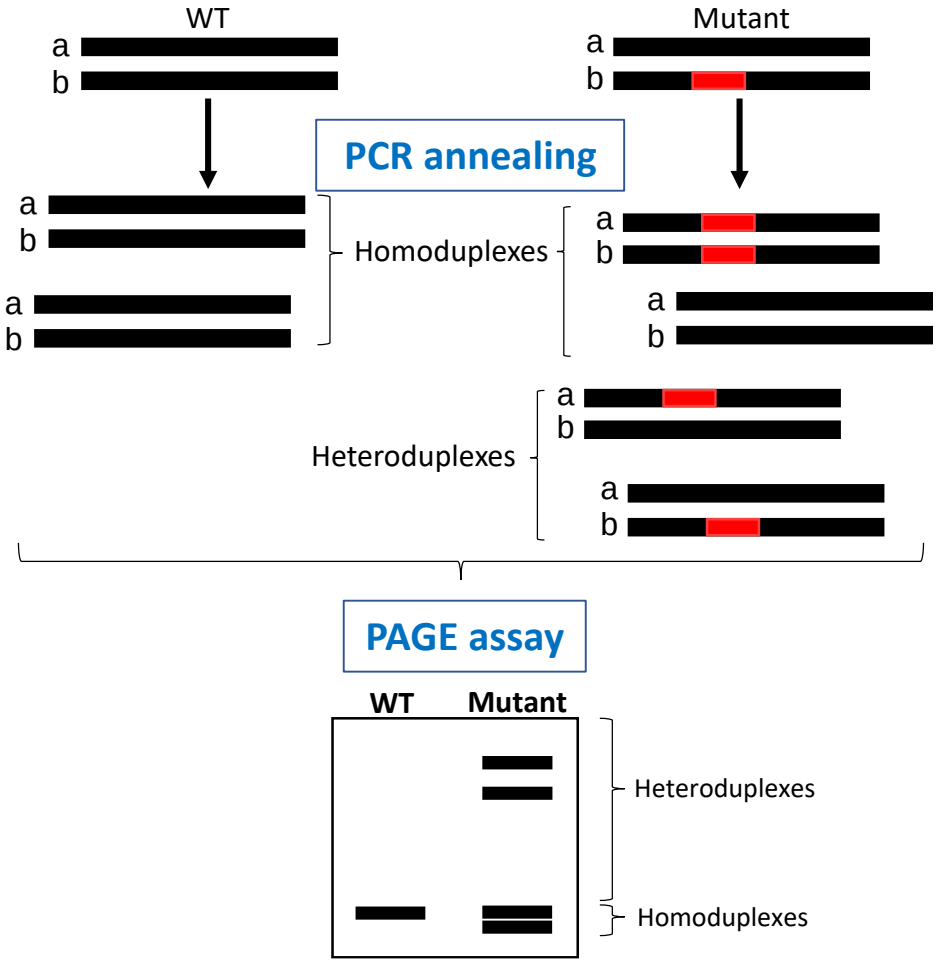


Figure 5.11 Schematic diagram of strategy for the identification of clones harbouring CRISPR-Cas9-induced mutations using PAGE assay.

Double-stranded amplicons spanning the targeted region generated from mutant clone were denatured and re-annealed to generate a mixture of homoduplexes and heteroduplexes. When resolved by PAGE, heteroduplexes migrate slower than homoduplexes due to the presence of mismatched bases. Homoduplexes formed by WT/WT and Mut/Mut also migrate at different rates depending on the mutations harboured by the mutant. WT = wild-type, Mut = mutant, PCR = polymerase chain reaction, PAGE = polyacrylamide gel electrophoresis.

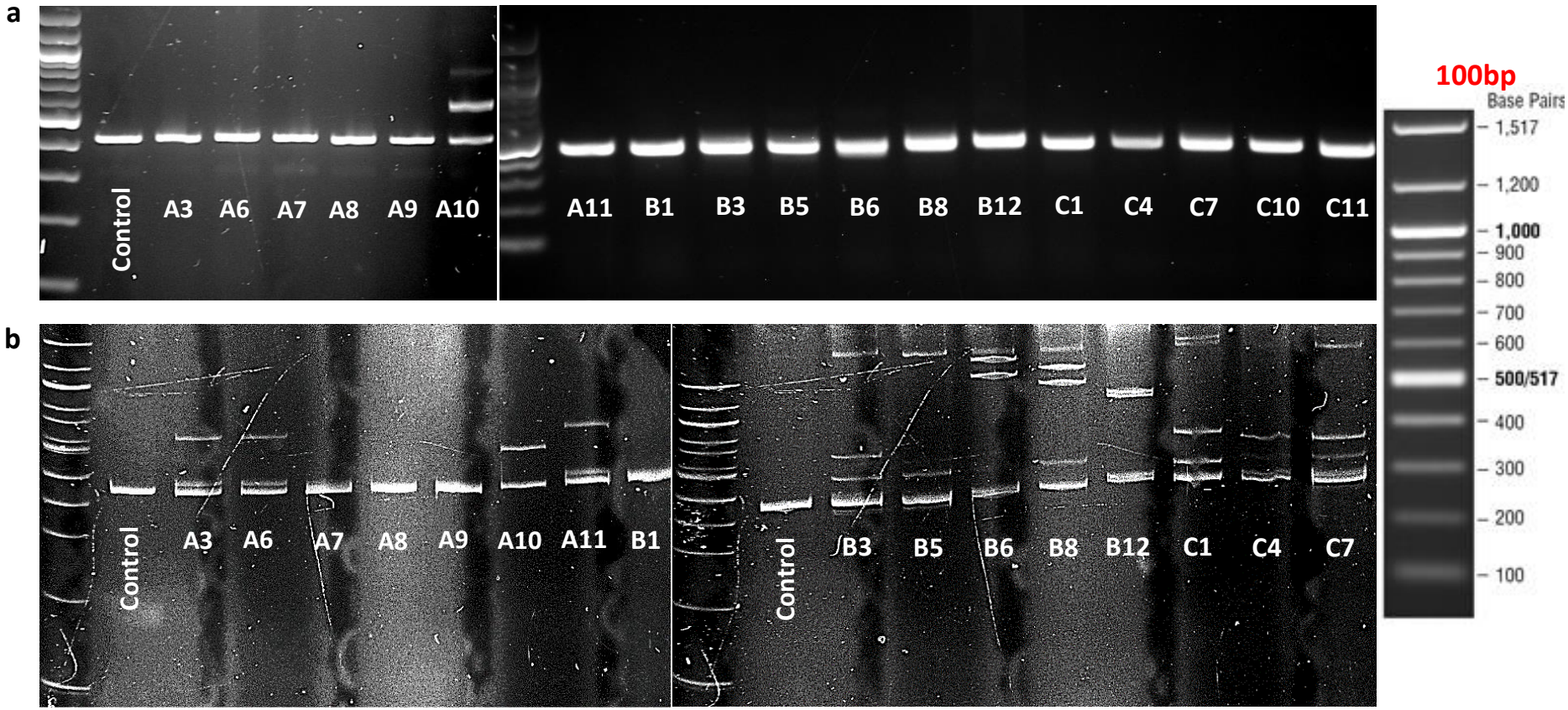


Figure 5.12 Screening for *TP53* mutated clones.

(a) Amplicons (418 bp) spanning the targeted region were generated by PCR. Amplicons were denatured and reannealed by PCR before (b) being resolved in PAGE assay. MCF10A_PIK3CA$^{H1047R(+/-)}$ was used as negative control. (right) 100 bp DNA ladder was used in all the above experiments.

A total of twelve clones, including two seemingly negative clones A8 and A9 and positive clones A3, A10, B3, B5, B6, B8, B12, C1, C4 and C7 were chosen for Western blotting analysis. All clones were treated with nutlin-3 for at least 48 hours before being harvested for the analysis to induce the expression of TP53 protein. The MCF10A_PIK3CA$^{H1047R(+/-)}$ cells, treated and untreated with nutlin-3 were used as controls. Results indicated that all clones, including the clones A8 and A9 did not express TP53. ACTB was used as endogenous control (Figure 5.13).

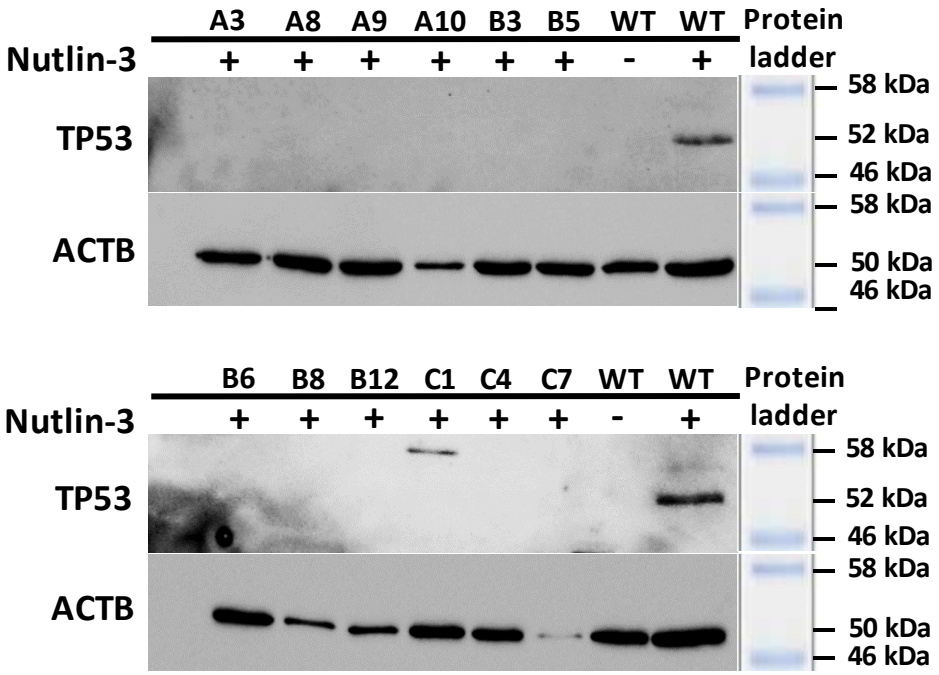


Figure 5.13 Western blotting analysis of TP53 expression.
All clones were pre-treated with nutlin-3 for at least 48 hours before the analysis. Nutlin-3 treated and untreated MCF10A_PIK3CA$^{H1047R(+/-)}$ cells were used as controls. ACTB was used as endogenous control. kDa = kilodalton, + = Nutlin-3 treated, - = Nutlin-3 untreated. Protein ladder = Colour prestained protein standard, broad range (11 – 245 kDa, NEB #P7712S).

To identify the nature of CRISPR-Cas9 induced-mutations that led to the knockout of TP53, PCR amplicons from A3 and A8 were cloned and sequenced. Three and four alleles harbouring different mutations were detected in A3 and A8 respectively. Mutations detected were commonly substitution and INDELs involving one nucleotide but interestingly, an insertion of a ~158 bp sequence was detected in one allele of A8 (Figure 5.14 & Figure 5.15). By analysing the mutations with online available evaluation tool MutationTaster (Schwarz et al., 2010), the deletion and addition of a single "G" nucleotide in allele 3 and allele 1 and 2 of clone A3, respectively (Figure 5.14, red arrow), resulted in frameshift and predicted to be disease causing (Appendix Figure S5-3 & S5-4). In clone A8, same prediction was obtained (frameshift and disease causing) for deletion and substitution observed in allele 1 – 3 (Figure 5.15, red arrow). The ~158 bp insertion in allele 4 was thought to alter poly(A) signal and splice site when analysed with MutationTaster (also disease causing) but predicted to cause frameshift when analysed with Scale Invariant Feature Transform (SIFT) (another online available evaluation tool) (Ng and Henikoff, 2003) (Appendix Figure S5-5, S5-6 & Table S5-1).



Figure 5.14 Alignment of cloned amplicon sequences from A3.

From five sequenced clones three different mutated sequences were identified. Three clones were found having the same mutated sequence as amplicon 1. Mutated regions are in red squares. (red arrow) Deletion and addition of "G" nucleotide was predicted to result in frameshift and possibly deleterious. WT = wild type, 1 – 3 = alelle 1 – 3.

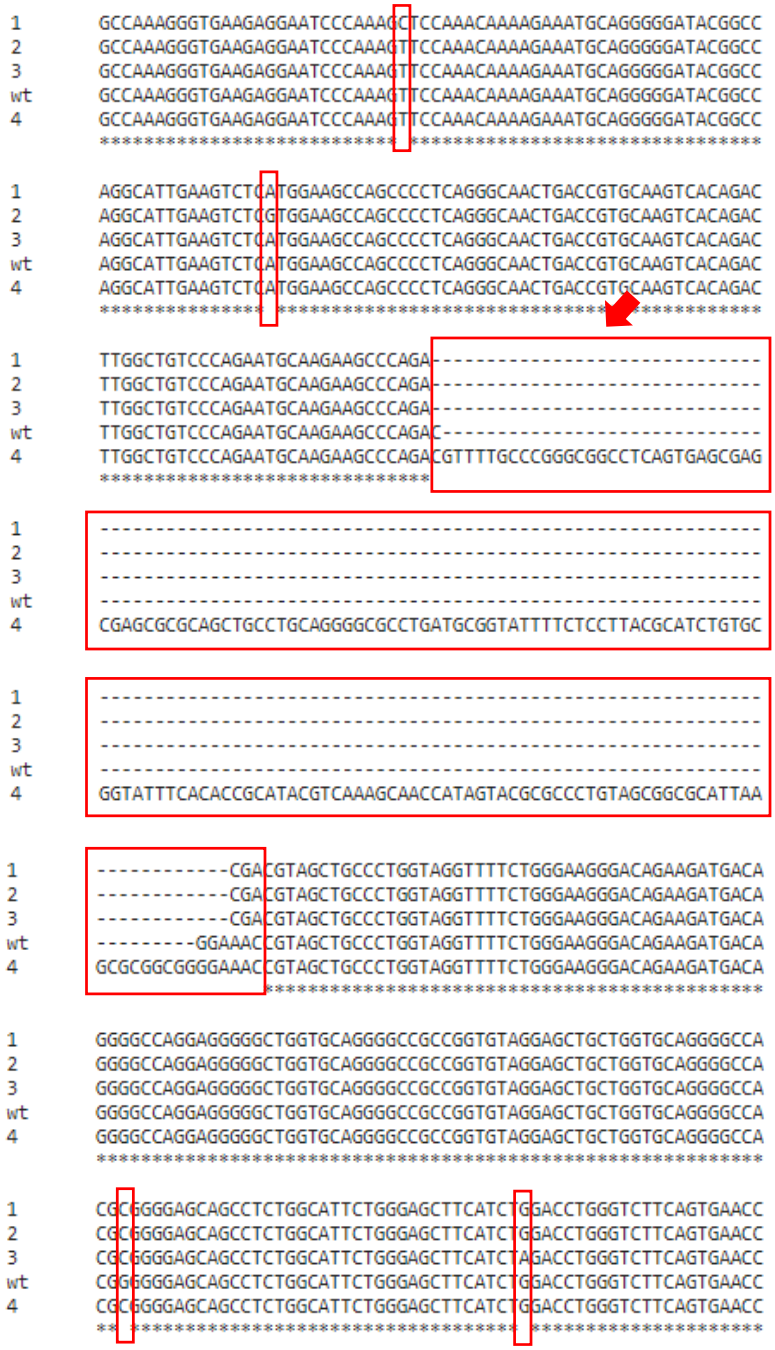


Figure 5.15 Alignment of cloned amplicon sequences from A8.

From five sequenced clones four different mutated sequences were identified. Amplicon 1 and 5 (not shown) had the same sequence. An insertion of ~158 bp DNA sequence was detected in amplicon 4. Mutated regions are in red squares. (red arrow) Mutations in allele 1 – 3 were predicted to result in frameshift and possibly deleterious. Insertion in allele 4 was thought to alter splice site and poly(A) signal when analysed with MutationTaster but predicted to cause frameshift when analysed with SIFT. WT = wild type, 1 – 4 = allele 1 – 4.

5.2.4 Reprogramming of mutant cell lines

The MCF10A_PIK3CA$^{H1047R(+/-)}$ and MCF10A_PIK3CA$^{H1047R(+/-)}$/TP53$^{(-/-)}$ clone A3 and A8 cells were next subjected to iPSC reprogramming. Cells were plated in 6-well plate pre-coated with vitronectin 24 h before the transfection. For reprogramming, cells were transfected four consecutive times every other day with five plasmids using the X-TremeGENE™ 9 DNA (XG9) transfection reagent. On the day of the first transfection (Day 0), culture medium was changed to E8 medium (supplemented with 1 μM hydrocortisone) before changing to E6 medium supplemented with 100 ng/mL human bFGF and 0.5 mM NaB at Day 8. Medium was also supplemented with 10 μM ROCKi to minimise cell death inflicted by the toxicity of NaB. The NaB treatment as well as ROCKi was stopped at day 12. From day 20-30, potential iPSC colonies were manually passaged and transferred into a new culture vessel with E8 medium supplemented with 10 μM ROCKi. The ROCKi was removed the following day and medium was changed every day. From MCF10A_PIK3CA$^{H1047R(+/-)}$, a total of five iPSC colonies were isolated and four were expanded as individual lines named 10A_PIK3CA$^{(-/+)}$ R1, R1.1, R4 and R4.1. Despite showing important morphological features of ESCs, the colonies were less compact and 10A_PIK3CA$^{(-/+)}$ R4.1 was morphologically more similar to cancer-iPSCs (5PLA_BTR1) than to 10AR (iPSC line of wild-type MCF10A). In addition, extensive spontaneous differentiation was constantly observed in all 10A_PIK3CA$^{(-/+)}$ iPSC lines as indicated by the abundant presence of fibroblast-like cells surrounding ESC-like colonies. The 10A_PIK3CA$^{(-/+)}$ R4 was expanded and characterised. For MCF10A_PIK3CA$^{H1047R(+/-)}$/TP53$^{(-/-)}$ clone A3, seven iPSC colonies were harvested and six expanded as

individual lines. Two iPSC colonies were manually passaged and expanded from the clone A8. All iPSC lines closely resembled ESCs morphologically. One iPSC line from both clones was characterised, namely the 10A_PIK3CA$^{(-/+)}$/TP53$^{(-/-)}$ R1.0 (from clone A3) and 10A_PIK3CA$^{(-/+)}$/TP53$^{(-/-)}$ R2.1.3 (from clone A8) (Figure 5.16).

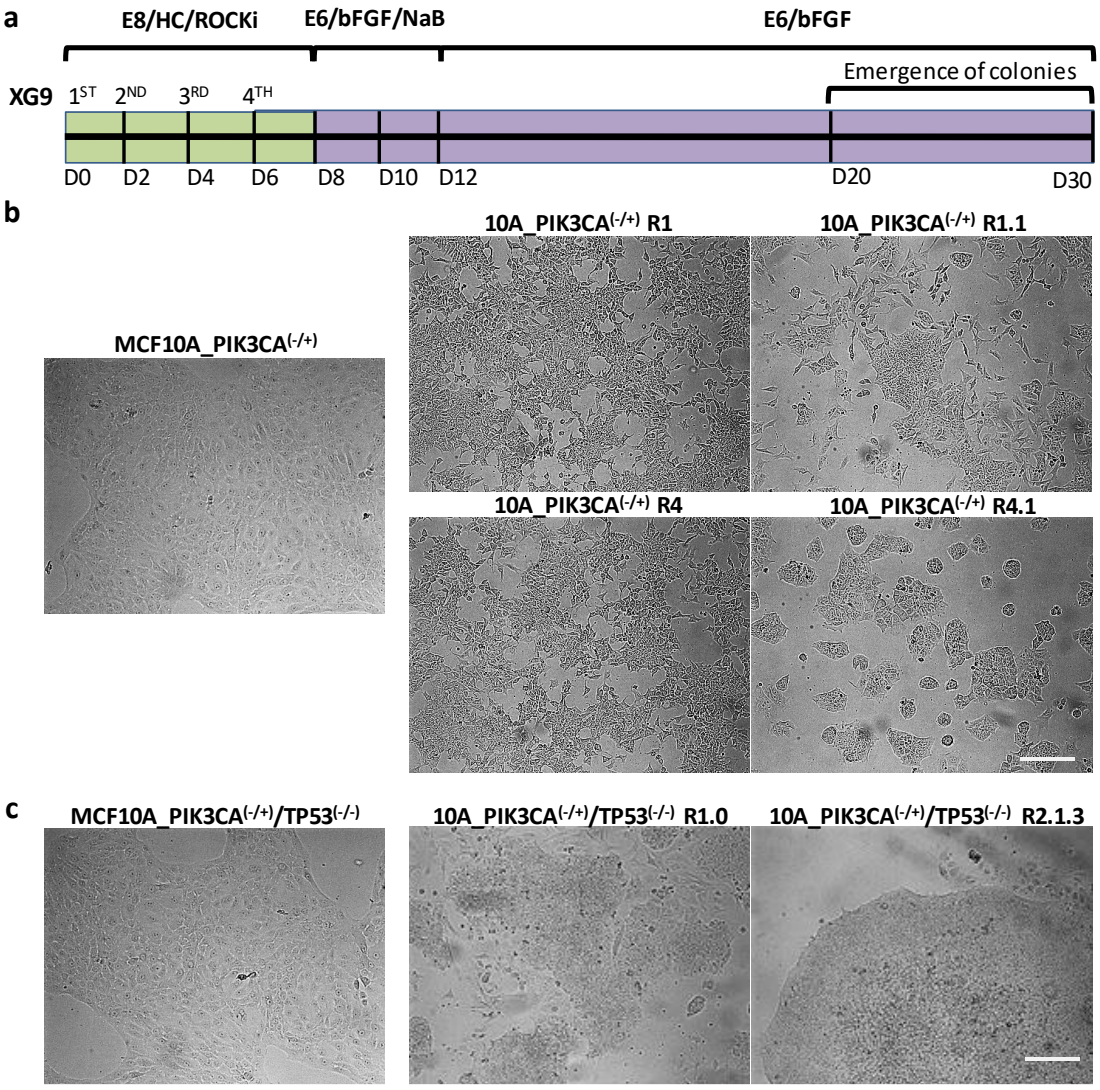


Figure 5.16 Reprogramming of MCF10A mutant cell lines MCF10A_PIK3CA$^{H1047R(+/-)}$ and MCF10A_PIK3CA$^{H1047R(+/-)}$/TP53$^{(-/-)}$.

(a) Reprogramming timeline. (b-c) Morphology of MCF10A_PIK3CA$^{H1047R(+/-)}$ and MCF10A_PIK3CA$^{H1047R(+/-)}$/TP53$^{(-/-)}$ and their respective iPSCs. E8 = Essential 8 medium, E6 = Essential 6 medium, HC = hydrocortisone, NaB = sodium butyrate, bFGF = basic fibroblast growth factor. Scale bar = 50 μm.

5.2.4.1 Characterisation of iPSC lines

All iPSC lines were characterised by RT-PCR and immunofluorescence. Results indicated all iPSCs but not their non-reprogrammed counterparts expressed pluripotency markers tested (Figure 5.17 - Figure 5.18). However, weak SSEA4 expression was consistently detected in 10A_PIK3CA$^{(-/+)}$ R4.

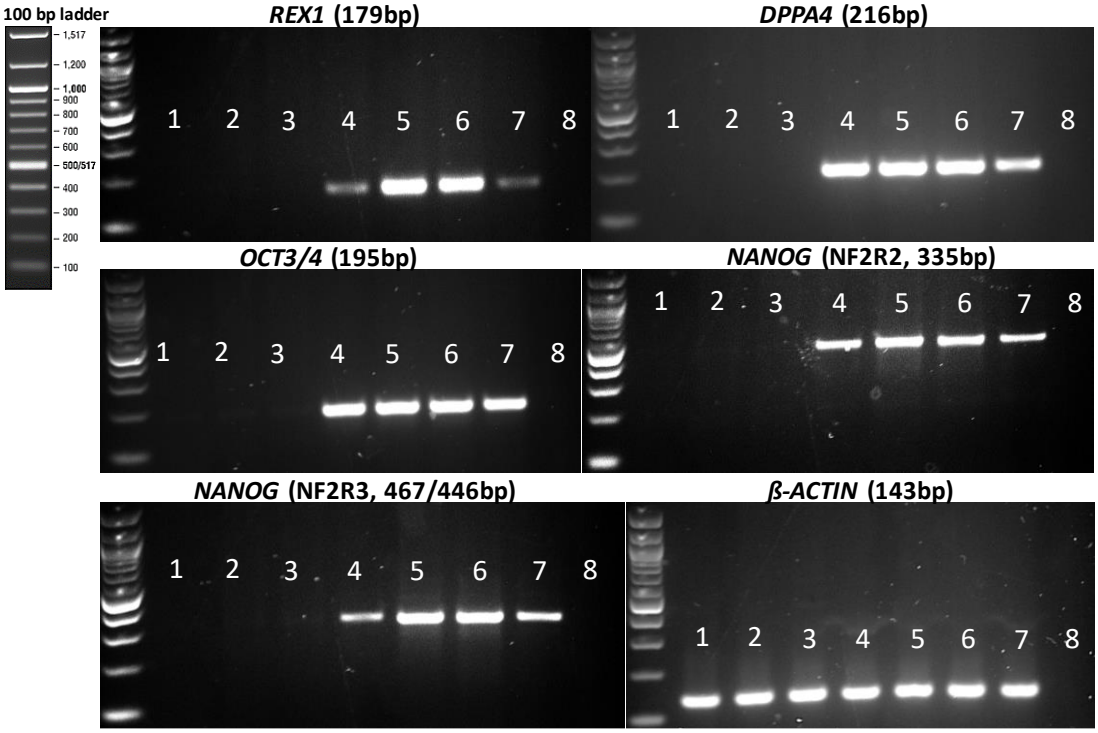
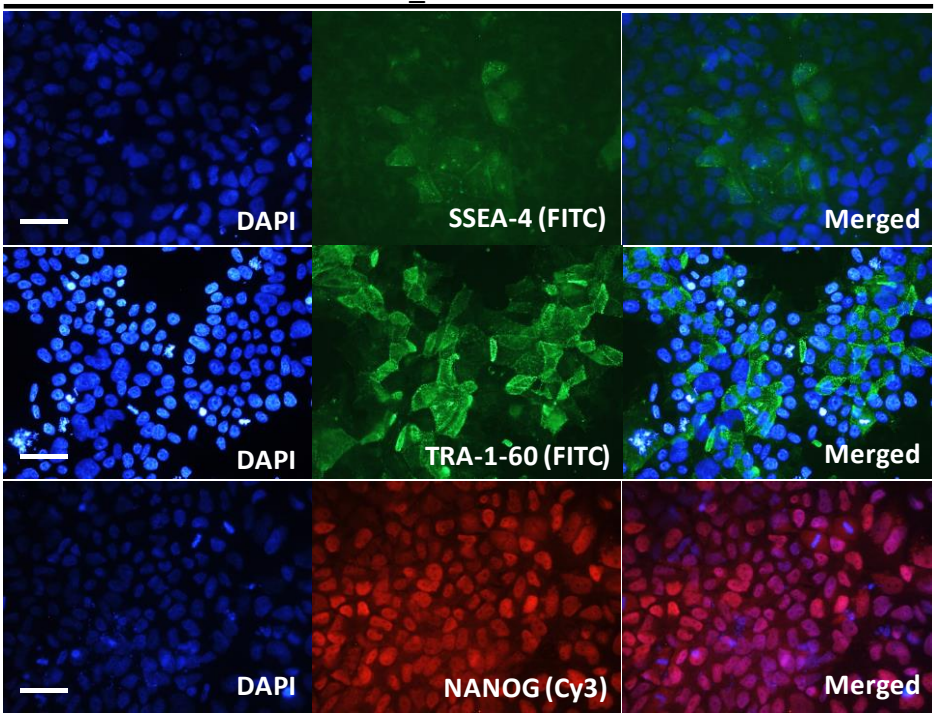


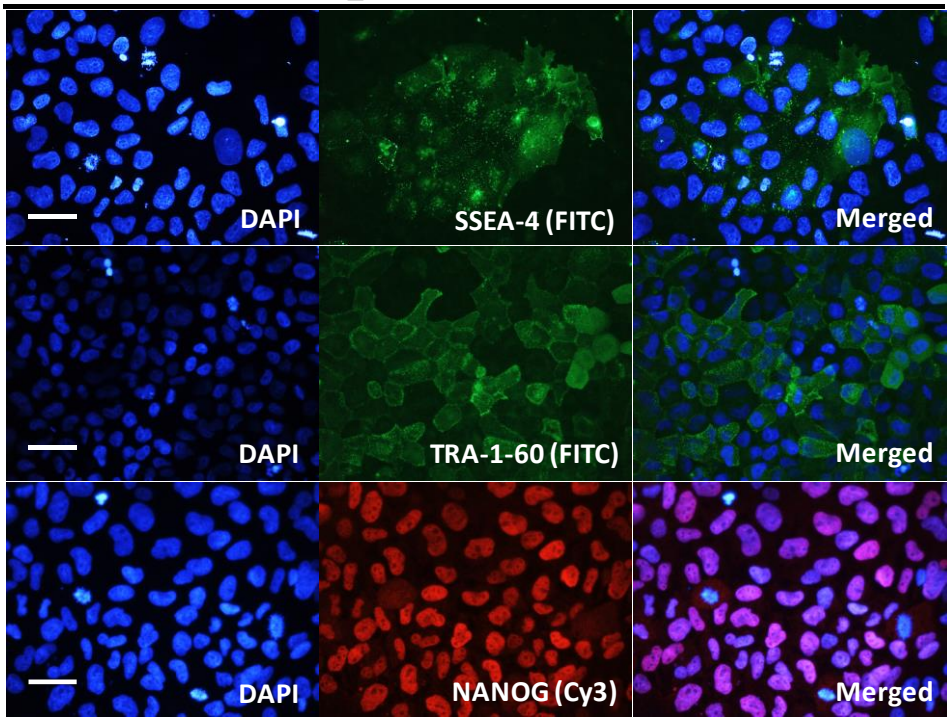
Figure 5.17 Representative RT-PCR of pluripotency markers in MCF10A_PIK3CA$^{H1047R(+/-)}$ and MCF10A_PIK3CA$^{H1047R(+/-)}$/TP53$^{(-/-)}$ and their respective iPSCs

All iPSCs but not their non-reprogrammed counterparts expressed pluripotency markers. ESC line H9 was used as a positive control. 100 bp ladder was used in all experiments shown. 1 = MCF10A_PIK3CA$^{H1047R(+/-)}$, 2 = MCF10A_PIK3CA$^{H1047R(+/-)}$/TP53$^{(-/-)}$ A3, 3 = MCF10A_PIK3CA$^{H1047R(+/-)}$/TP53$^{(-/-)}$ A8, 4 = 10A_PIK3CA$^{(-/+)}$ R4, 5 = 10A_PIK3CA$^{(-/+)}$/TP53$^{(-/-)}$ R1.0, 6 = 10A_PIK3CA$^{(-/+)}$/TP53$^{(-/-)}$ R2.1.3, 7 = H9, 8 = negative control. All cell lines except for H9 were <10 passage, H9 = passage 24 – 30. n ≥ 3.

10A_PIK3CA$^{(-/+)}$ R4



10A_PIK3CA$^{(-/+)}$/TP53$^{(-/-)}$ R1.0



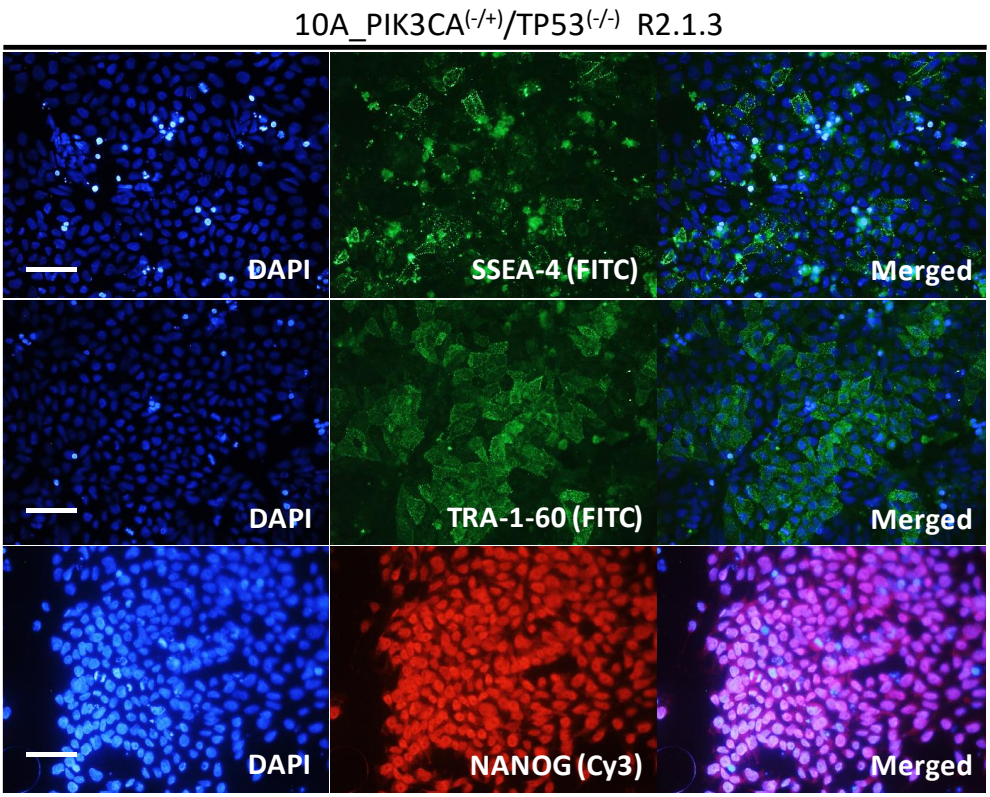


Figure 5.18 Immunofluorescence staining of SSEA-4, TRA-1-60 and NANOG in 10A_PIK3CA$^{(-/+)}$ R4, 10A_PIK3CA$^{(-/+)}$/TP53$^{(-/-)}$ R1.0 and 10A_PIK3CA$^{(-/+)}$/TP53$^{(-/-)}$ R2.1.3.

All iPSCs expressed the three pluripotency markers analysed. Only weak SSEA-4 expression was detected in 10A_PIK3CA$^{(-/+)}$ R4. Scale bar = 50 μm. Secondary antibody-only negative control is shown in Figure S3-3. All cell lines were <10 passage. n = 1.

5.3 DISCUSSION

The aim of this chapter was to introduce two mutations into the MCF10A line that affect the two most commonly mutated genes in breast cancer, namely *PIK3CA* and *TP53*. The creation of the mutant cell line harbouring *PIK3CA*$^{H1047R(+/-)}$ mutation from MCF10A cell line using the CRISPR-Cas9 gene editing system was not successful. Despite the promising result from BspH1 screening of GFP^{high} cell pool, no desired mutants were identified when screening was performed on individual clones. Results suggested that ssODN-mediated HDR pathway DNA repair occurred at an extremely low frequency leading to the scares presence of mutants harbouring the abolished BspH1 site and *PIK3CA*$^{H1047R(+/-)}$ mutation. On the contrary, the error-prone NHEJ pathway was predominantly employed to repair the Cas9-induced DSBs leaving random mutations at the target region. Therefore, the generation of *PIK3CA* and *TP5*3 double mutant was performed on a commercially available MCF10A cell line harbouring the *PIK3CA*$^{H1047R(+/-)}$ mutation (MCF10A_PIK3CA$^{H1047R(+/-)}$). Several potential double mutants were successfully generated by using the CRISPR-Cas9 system. Sequencing of the target loci of two mutants (A3 and A8) revealed the presence of various Cas9-induced mutations including insertions, deletions and substitutions. TP53 knockout was confirmed by Western blotting analysis. Both mutants were subjected to iPSC reprogramming in order to dissect the interaction of the driver mutations with epigenetic factors during reprogramming.

5.3.1 Introduction of *PIK3CA*$^{H1047R(+/-)}$ mutation via CRISPR-Cas9 system

Several hurdles were met during the CRISPR-Cas9 editing of *PIK3CA* gene: First, the inefficiency of Cas9 cleavage. WT Cas9 enzyme is highly efficient and bi-allelic modification is the common outcome in most CRISPR-Cas9 gene editing experiments (Platt et al., 2014). In this study, Sanger sequencing results as resolved by the CRISPR-ID software indicated only mono-allelic modification was obtained in all four clones investigated, indicating the inefficiency of Cas9 cleavage in this study. Despite the fact that heterogeneous modification was the intended outcome for *PIK3CA*$^{H1047R(+/-)}$ mutation, the inefficiency of Cas9 could have potentially reduced the overall number of mutants and further decreased the appearance of clones which have undergone the ssODN-mediated HDR (which itself is a less efficient pathway as compared to the NHEJ) (Platt et al., 2014, Paquet et al., 2016). The sgRNA design has dramatic impact on Cas9 cleavage efficiency (Ren et al., 2014, Paquet et al., 2016, Xu et al., 2017). For example, the distance between then intended mutation site and the Cas9 cleavage site affects the zygosity of mutation and the HDR efficiency (Ran et al., 2013). Therefore, it is advisable that several sgRNAs should be tested to identify one that could contribute to desirable outcome in future experiment.

Second, it was the low efficiency of HDR. In this study, the frequency of ssODN-mediated HDR was low, an observation consistent with other reports (projected to be <1% in the absence of drug selection) (Hisano et

al., 2015, Zhang et al., 2017). Studies demonstrated that site-specific mutagenesis by using ssODN often results in the prevalence of cells harbouring mono-allelic ssODN-mediated repair and having the other allele randomly modified by indels due to the high efficiency of Cas9 and NHEJ pathway (Mali et al., 2013, Wang et al., 2013, Bialk et al., 2016). It is speculated that the low efficiency of ssODN-mediated HDR in this study was due to the ssODN design and the timing of ssODN introduction. The design of ssODN is challenging and one of the factors affecting its efficiency is its length, which has been demonstrated to have significant impact on the HDR efficiency. Yang et al. (2013) tested a range of ssODNs from 50 – 170 nucleotide (nt) and identified the ssODN with the highest HDR efficiency (~1.8%) as 90 nt. Further increase in length reduced the HDR efficiency to as low as ~0.3% for a 170 nt ssODN. In agreement, Ran et al. (2013) suggested that for point mutation, HDR rate reduces ~4-fold at 100 bp away from the Cas9 cutting site. However, an ssODN with high efficiency should contain at least 40 bp homology arms (meaning on both sides or in total should be around 80 bp long) from the Cas9 cleavage site. Nevertheless, ssODNs out of these recommended range have also been used in different studies. For example, a 72 nt ssODN was used by Bialk et al. (2016), 75 nt by Davis and Maizels (2014), 126 nt by Wang et al. (2013) and 180 nt by (Ran et al., 2013). The ssODN used in this study was designed as 141 bp long for improved specificity. Failure to elicit highly efficient HDR in the MCF10A cells suggested that a range of ssODN with different lengths should be tested or at least a shorter ssODN within the reported optimum range (80-100 nt) should be used. In addition, in this study, ssODN was transfected 24 h after Cas9 and sgRNA to avoid sgRNA-guided Cas9

targeting and destruction of ssODN harbouring PAM sites (5'-NGG-3'). It is speculated that the DNA repair mechanism might have completed before the introduction of ssODN and thus ssODN-mediated HDR was not elicited. Therefore, experiment should be performed in the future to co-transfect ssODN along with Cas9 and sgRNA (Ran et al., 2013, Wang et al., 2013, Bialk et al., 2016, Zhang et al., 2017).

5.3.2 Creation of cancer models using CRISPR-CAS9 system

Several studies have reported the generation of breast cancer models by introducing *PIK3CA* mutation (primarily at the E545K and H1047R mutation hot spots) alone or in concert with *TP53* knockout in mammary epithelial cell line such as MCF10A (Isakoff et al., 2005, Zhang et al., 2008a, Gustin et al., 2009, Wu et al., 2014, Croessmann et al., 2017) and HMEC (Zhao et al., 2005) and in mice (Van Keymeulen et al., 2015) (Table 5.1). The Phosphatidylinositol 3-kinases (PI3Ks) are a family of lipid kinases known to phosphorylate the 3'-OH of phosphoinositides. They are part of the PI3K/PTEN/AKT signalling pathway, which involves in numerous essential cellular processes, including cell survival, proliferation, metabolism, cytoskeletal rearrangement and membrane trafficking (Vivanco and Sawyers, 2002). Activating mutation of *PIK3CA* gene (encoding p110α isoform of catalytic subunit) has been associated with PI3K activation and extensively reported in various human cancers including breast cancer (Karakas et al., 2006, Croessmann et al., 2017). The H1047R and E545K, affecting the kinase domain and helical domain of p110α catalytic subunit respectively, are the two most common mutations in *PIK3CA*. The p110α isoform contains a RAS-binding domain (RBD) which binds GTP-bound RAS,

leading to its enhanced activation synergistically with Receptor tyrosine kinases (RTK) activation. The H1047R mutation increases membrane affinity and kinase domain activity of p110α catalytic subunit without the requirement of RAS binding (Burke and Williams, 2015).



Figure 5.19 The domain structure of Class IA p110α catalytic subunit.

Blue arrows = the approximate locations of two mutation hot spots. p85D = p85-binding domain, RBD = RAS-binding domain, C2 = C2 domain.

TP53 instead encodes a crucial tumour suppressor P53, which functions to control cell cycle, proliferation and apoptosis (Sherr, 2006). *TP53* mutations lead to aberrant cell proliferation and survival in a wide range of cancer (Vousden and Prives, 2009, Abraham and O'Neill, 2014).

Table 5.1 Studies that generated PIK3CA mutation models using gene editing technology.

References	Mutation	Gene editing system	Major findings
Isakoff et al. (2005)	E545K or H1047R in MCF10A	Retrovirus	• Cells have high PI3K activity and sensitive to pharmacologic PI3K inhibition • Exhibited feature of cancer cells such as anchorage-independent proliferation in soft agar, EGF-independent proliferation, and resistant to anoikis cell death, abnormal mammary acinar morphogenesis. • Increased resistance to paclitaxel
Zhao et al. (2005)	TP53 plus PIK3CA E545K, H1047R (both affect p110α isoform) or E552K (affect p110ß isoform) in HMEC cell line	Retrovirus	• High soft agar colony formation ability and form orthotopic tumours in nude mice • p110α isoform was more potently activated by missense mutation than p110ß although both possessed tumorigenic potential.

Zhang et al. (2008a)	Nine different somatic mutations in PIK3CA (Q60K, K111N, N345K, C420R, P539R, E542K, E545K, H701P and H1047R) in MCF10A	Lentivirus	• PIK3CA mutants harbour different abilities to promote cell proliferation, colony formation and cell invasion • Parental PIK3CA, the Q60K and the K111N mutants were unable to change morphogenesis in 3D Matrigel assay • p-AKT and p-p70-S6K observed in most mutants in the absence of EGF stimulation and responded effectively to LY294002 (PI3K inhibitor) • Different PIK3CA mutations have different effects on breast carcinogenesis
Gustin et al. (2009)	PIK3CA (E545K or H1047R) in MCF10A	Cre-loxP recombination	• AKT, ERK, and GSK3β phosphorylation and EGF and mTOR-independent cell proliferation observed • Selective targeting of human breast and colorectal cancer cell lines harbouring PIK3CA mutations by lithium chloride (GSK3β inhibitors) and SB216763 and downregulation of GSK3β target gene *CYCLIN D1* observed.
Wu et al. (2014)	PIK3CA (E545K or H1047R) in MCF10A derived from Gustin et al. (2009)	Cre-loxP recombination	• Stable isotope labelling by amino acids in cell culture (SILAC) –based quantitative phosphoproteomic analysis revealed 1. Significant increase phospho-AKT and phospho-MAPK levels 2. Downstream signalling effects are similar between two mutant, slight difference existed such as highly Phosphorylation of PAK2, PAK4 and SLK in Ex9-KI cells 3. Global Protein phosphorylation increased 47% and 33% in E545K- and H1047R-mutants respectively 4. Phosphorylation of cortactin by AKT1 enhanced cell migration and invasion in PI3K mutants.
Van Keymeulen et al. (2015)	$PIK3CA^{H1047R}$ alone or with TP53 knockout in mice	Cre-loxP recombination	• Conditional *pik3ca*^{H1047R} mutation in basal cells (marked by keratin 5) induced the formation of luminal ER^+/PR^+ tumours • In luminal cells (keratin 8), luminal ER^+/PR^+ tumours or basal-like ER^-/PR^- tumours were identified • Double mutation of *pik3ca* and *p53* enhanced development and aggressiveness of mammary tumours • Models provided insights into the cell of origin and tumour heterogeneity of PIK3CA-mutated breast cancer cells
Croessmann et al. (2017)	PIK3CA (E545K or H1047R) plus TP53 alterations (R248W or knockout) in MCF10A	Cre-loxP recombination	• All mutants were non-tumorigenic • Cancer-like cell proliferation, soft agar colony formation, abnormal mammary acinar morphogenesis and genomic heterogeneity were more prominent in double mutant as compared to single mutant and parental MCF10A • Increased sensitivity to paclitaxel in all PIK3CA mutants

*abbreviation: EGF =epidermal growth factor, ER oestrogen receptor, PR = progesterone receptor.

Different gene-editing approaches have been utilized to introduce *PIK3CA* mutation, including retroviral- and lentiviral-mediated site mutagenesis and Cre-loxP recombination but not the CRISPR-Cas9 system. Consistent across all studies, the introduction of *PIK3CA* mutation alone promoted oncogenic properties such as enhanced cell proliferation, colony formation ability and aberrant mammary acinar morphogenesis, but was insufficient to induce tumorigenicity. Tumour formation or tumour-like activity was only detected in the presence of other driver mutations or oncogenic factors (such as ER overexpression). In this study, no pronounced morphological differences were observed between parental and mutant cell lines. However, despite the lack of quantitative data, the mutant cell lines MCF10A_PIK3CA$^{H1047(+/-)}$ and MCF10A_PIK3CA$^{H1047(+/-)}$/TP53$^{(-/-)}$ resulted in confluent culture before day 20 after the initial transcription factors transfection during iPSC reprogramming (an incidence not seen in the parental MCF10A cells), suggesting both cell lines might have acquired high cell proliferation rate in comparison to MCF10A.

While the complexity of cancer genome has rendered the investigation into the contribution of driver mutations to carcinogenesis and their interactions with other driver/passenger mutations challenging, the availability of gene editing technology has permitted the creation of cancer models to circumvent this caveat (Chin et al., 2011). The aforementioned *PIK3CA* mutation studies have laid the ground work and provided important insights into the phenotypic effects to the mechanisms of action of *PIK3CA* mutation as well as its interaction with *TP53* mutation. Isakoff et al. (2005) and Zhao et al. (2005) were among the pioneer studies that demonstrated

the oncogenic effects of *PIK3CA* mutation in genetically modified mammary epithelial cell lines. The latter provided crucial evidence to reinforce the significance of p110α isoform (encoded by the *PIK3CA* gene) mutation among other isoforms such as p110ß (by *PIK3CB*) and p110γ (*PIK3CG*). In addition, Zhang et al. (2008a) further highlighted the importance of the two most common hot spots of *PIK3CA* mutation, E545K and H1047R, in comparison to a panel of 8 other mutation locations. Subsequent studies by Gustin et al. (2009) and Wu et al. (2014) attempted to dissect the mechanisms underlying the promotion of oncogenic properties in cells harbouring E545K and H1047R mutations. Gustin et al. (2009) demonstrated increased GSK3ß phosphorylation (caused by increased AKT and ERK activation) contributed, at least partially, to the EGF- and mTOR-independent cell proliferation in *PIK3CA* mutant. Treatment of breast and colorectal cancer lines with GSK3ß inhibitors downregulated *CYCLIN D1* (A GSK3ß target gene) and led to the selective targeting of *PIK3CA* mutant cells. On the other hand, through stable isotope labelling by amino acids in cell culture (SILAC) quantitative phosphoproteomic analysis and human protein microarray-based AKT1 kinase assays, Wu et al. (2014) reported a wide range of kinases and downstream signalling networks affected by *PIK3CA* mutation and demonstrated that the phosphorylation of cortactin by AKT1 was responsible for the enhanced migration and invasion of cells harbouring *PIK3CA* mutations.

Moreover, by using an *in vivo* system in mice Van Keymeulen et al. (2015) demonstrated that *Pik3ca*^{H1047R} mutation initiated an "oncogene-

induced reprogramming" program to confer multipotency to unipotent basal and luminal cells, leading to the formation of luminal B tumours (ER^+/PR^+) by basal cells and luminal (ER^+/PR^+) or basal tumour (ER^+/PR^+) by luminal cells, suggesting cell fate switching during tumorigenesis. This finding has provided important insights into the cell of origin of breast cancer as well as the mechanisms underlying tumour heterogeneity. In addition, double mutation of *Pik3ca* and *Tp53* facilitated the formation of more aggressive metaplastic tumour in mice (Van Keymeulen et al., 2015). In regards to the collective effect of *TP53* and *PIK3CA* mutation in breast cancer, Croessmann et al. (2017) identified aggressive oncogenic features in double mutant in comparison to single mutant and their parental cells. Notably a greater degree of genomic instability was also observed, which in turn contributed to pronounced cellular heterogeneity.

The present study has taken a different approach to investigate the interaction between the two most common driver mutations (or *PIK3CA* alone) as well as their interactions with epigenetic factors. This aimed to determine the potential contribution of epigenetic regulation/dysregulation to oncogenic phenotypes previously reported and the probable intervention by epigenetic modifiers to suppress these oncogenic effects in breast cancer. Cancers with positive CpG island methylator phenotype (CIMP) are characterised by high degrees of CpG-rich promoters hypermethylation silencing of tumour suppressor or tumour suppressor-related genes and epigenetic instability (Issa, 2004). *PIK3CA* mutation has been associated with high CIMP in several colorectal cancer studies (Nosho et al., 2008, Ahmed et al., 2013). A recent review by Wong et al. (2017) focusing on

cancer metabolism highlighted that promoter hypermethylation silencing of tumour suppressors such as PTEN, LKB1 and VHL, de-repress the PI3K/AKT/mTOR and HIF-1 signalling pathways resulting in metabolic switch to glycolysis in various cancers. In addition, a panel of miRNAs have been shown to regulate the PI3K/AKT/mTOR, LKB1/ AMPK, MYC, and HIF1 signalling cascades and contribute to the cancer metabolic phenotype. On the other hand, Zuo et al. (2011) demonstrated that the activation PI3K/AKT signalling pathway induce the silencing of multiple loci via 1) accumulation of H3K27me3, 2) both H3K27me3 and promoter hypermethylation or 3) predominantly promoter hypermethylation. The collective effect of epigenetic and PI3K/AKT-targeted therapies removed these silencing epigenetic marks and led to the suppression of tumour growth in breast cancer cell lines. Moreover, a recent study by Toska et al. (2017) indicated that the upregulation of AKT in *PIK3CA*-mutated breast cancer cells represses methyltransferase activity of KMT2D via phosphorylation, which in turn disrupts the assembly of a FOXA1-PBX1-ER regulatory network and thus transcriptional activity of ER. This observation explains the resistance of a subset of ER-positive breast cancers to PI3K inhibitor treatment, whereby the upregulation of ER-regulated genes are frequently observed, suggesting the combination of PI3K and KMT2D inhibitors may be more effective against ER^+ breast tumour than PI3K inhibitor alone.

On a different note, off-target mutation represents a major concern of CRISPR-Cas9 gene editing as the endonuclease activity of Cas9 is solely guided by a short 20 nt sequence carried by the sgRNA (Schaefer et al.,

2017). To facilitate the selection of highly efficient sgRNA with low off-target effect, various online tools, including the http://crispr.mit.edu used in this study, are available to predict possible off-target sites of sgRNAs by using different algorithms (Bae et al., 2014, Hodgkins et al., 2015, Koo et al., 2015b). In this study, two online tools reported by Hodgkins et al. (2015) and Bae et al. (2014) were used to identify the off-target sites of sgRNA used for *TP53* knockout (Appendix Figure 5S-7 & 5S-8). All off-target sites predicted have at least three nucleotide mismatches (only five with three mismatches) from the sgRNA, indicating high sgRNA specificity. Nevertheless, analyses should be performed in the future to amplify and sequence the most potent off-target sites to identify possible off target effects.

In this study, several iPSC lines were successfully derived from the two mutant cell lines. All iPSC lines exhibited important ESC-like features, as indicated by RT-PCR and immunofluorescence staining. Results suggested that the presence of *PIK3CA* and *TP53* mutations did not impede the reprogramming process, despite the fact that iPSC lines isolated from the MCF10A_PIK3CA$^{H1047(+/-)}$ cells were less ESC-like. However, more reprogramming experiments should be performed to confirm the results. If it is proven that MCF10A_PIK3CA$^{H1047(+/-)}$ indeed has a lower reprogramming efficiency, it will also be interesting to investigate the molecular nature of the differential response to reprogramming between the two mutants as well as the parental cell line. This could help identify if *PIK3CA* mutation could contribute to reprogramming refractoriness as reported in the other cancer driver mutations such as, *KRAS* (Kim et al., 2013). Considering the

intertwining relationship of PI3K and p53 pathways and the vital roles of various epigenetic mechanisms in the signalling cascades, the reprogramming of MCF10A_PIK3CA$^{H1047(+/-)}$ and MCF10A_PIK3CA$^{H1047(+/-)}$/TP53$^{(-/-)}$ presented a great opportunity to dissect the epigenetic mechanisms underlying oncogenic phenotypes conferred by the *PIK3CA* and *TP53* mutations in breast cancer. In addition, the successful generation of the *PIK3CA*$^{H1047(+/-)}$ and *TP53*$^{(-/-)}$ double mutant provides a stepping stone for the creation of models harbouring multiple subtype-specific somatic mutations to mimic the complexity of the breast cancer genome (Mou et al., 2015, Walton et al., 2017).

In conclusion, in this chapter it had been demonstrated that MCF10A was amendable to CRISPR-Cas9 gene editing. Mutants created can also be reprogrammed to generate iPSCs. The combination of both gene editing and iPSC reprogramming systems allow the investigation of interaction between breast cancer driver mutations and epigenetic regulation/dysregulation.

6. DIFFERENTIATION POTENTIAL AND TUMORIGENICITY OF BREAST CANCER AND MUTANT MCF10A -DERIVED IPSC LINES

6.1 INTRODUCTION

The previous chapters reported the successful derivation of iPSC lines from a non-malignant mammary epithelial cell lines MCF10A, breast cancer cell lines of triple negative and ER^+ subtypes and MCF10A mutants harbouring *PIK3CA* and *TP53* mutations (Table 6.1). The objective of this chapter was to determine the differentiation potential of all iPSC lines through *in vitro* differentiation, their *in vitro* tumorigenicity and chemotherapy drug sensitivity. The scope of these tests was to provide important insights into the capability of iPSC reprogramming-mediated reconfiguration of the breast cancer cell epigenome to re-establish the differentiated phenotype of normal mammary tissue and to suppress or eliminate tumorigenicity. In addition, the impact of genetic alterations could also be determined through assessment of oncogenic phenotypes in isogenic cell lines harbouring the most common driver mutations in breast cancer.

Table 6.1 Summary of iPSC lines isolated and characterisation performed

Cell line	Reprogramming factors used	Number of iPSC /potential iPSC lines derived	iPSC/potential iPSC lines subjected to characterisation	Characterisation		
				RT-PCR	RT-qPCR	IF
Non-malignant control						
MCF10A	OSKML	5	10AR	√	√	√
Cancer cell lines						
MCF7	OSKML	-*	MCF7R	√		
	OSKML + miR-302/367 + NaB	1	5PLA_MCF7R	√	√	√
BT-549	OSKML	-*	BT-1, BT-2, BT8	√		
	OSKML (FRM/FMM)	-**	FMM_BTR1	√		
	OSKML + miR-302/367 + NaB	6	5PLA_BTR5	√	√	√
MDA-MB-231	OSKML	-	MDA R1, MDA R2	√		
	OSKML + miR-302/367 + NaB	-	5PLA_MDAR	√		
CRISPR-Cas9 mutant cell lines						
MCF10A_PIK3CA$^{(+/-)}$	OSKML + miR-302/367 + NaB	5	10A_PIK3CA$^{(+/-)}$ R4	√	√	√
MCF10A_PIK3CA$^{(+/-)}$ /TP53$^{(-/-)}$ clone A3	OSKML + miR-302/367 + NaB	2	10A_PIK3CA$^{(-/+)}$/TP53$^{(-/-)}$ R1.0	√		√
MCF10A_PIK3CA$^{(+/-)}$ /TP53$^{(-/-)}$ clone A8	OSKML + miR-302/367 + NaB	7	10A_PIK3CA$^{(-/+)}$/TP53$^{(-/-)}$ R2.1.3	√		√

Abbreviation: OSKML = *OCT3/4, SOX2, KLF4, L-MYC* and *LIN28*, NaB = sodium butyrate, IF = immunofluorescence staining.
*Morphological changes observed
**Morphological changes and pre-iPSC-like characteristics detected.

6.1.1 Significance of cancer iPSC differentiation

Failure to undergo a normal differentiation process is a characteristic of many cancers (Hanahan and Weinberg, 2011a). Disruption caused by oncogenic mutations generally results in aberrant or blocked differentiation programs, which often occurs during early stages of differentiation. For example, the presence of the oncogenic fusion protein BCR-ABL and upregulated C/EBPγ have been shown to induce differentiation arrest in chronic myeloid leukaemia (CML) (Jacquel et al., 2003) and acute myeloid leukaemia (AML) (Alberich-Jorda et al., 2012) respectively. Treatment with Imatinib, a BCR-ABL inhibitor, induces apoptosis of BCR-ABL positive cells and their differentiation into cells of erythroid lineage. In addition, retinoic acid (RA) and its derivatives have been widely used for differentiation therapy (treatment that induces cancer cell differentiation and stops further proliferation) of acute promyelocytic leukaemia (APL). Furthermore, RA has also been employed for chemoprevention in various cancers including breast, liver and lung (di Masi et al., 2015). Early evidences indicated that epigenetic reprogramming could confer to cancer cells normal differentiation potential (McKinnell et al., 1969). For example, reprogrammed frog renal tumour cells through SCNT could give rise to swimming tadpole with no signs of malignancy. On the other hand, SCNT-reprogrammed mouse medulloblastoma was shown capable of generating embryos which underwent early stages of organogenesis (Li et al., 2003). Recently, (Zhang et al., 2013) demonstrated that sarcoma-derived iPSCs were capable of terminal differentiation to generate bone and fat cells.

The combination of iPSC technology with well-established differentiation induction protocols has provided an unprecedented approach to model cancer. Several studies have reported the generation of cancer models, including myeloproliferative disorder (Ye et al., 2009), CML (Hu et al., 2011, Kumano et al., 2012), AML (Kotini et al., 2017) through re-differentiation of cancer iPSCs. In addition, Kim et al. (2013) generated an early stage model for invasive pancreatic ductal adenocarcinoma (PDAC) through teratoma formation from PDAC-derived iPSC which could facilitate the study of cancer progression and identification of early biomarkers. Therefore, in this chapter, the differentiation potential of breast cancer iPSC lines was determined. Subsequently, the re-differentiated cell lines were characterised and their potential to be used as cancer models were evaluated by performing soft agar colony forming assay and chemo drug sensitivity tests.

6.1.2 Lineage-specificity of cancer mutations

One objective of the present study was to investigate lineage specificity of mutations harboured by breast cancer cells through re-differentiation of reprogrammed breast cancer-derived iPSCs into cells of different lineages followed by assessment of tumorigenicity. Lineage-specific variation in mutations suggests that cells of different lineages possess distinct prevalence to different mutations and suffer from different phenotypic effects caused by the same mutated gene (Chang et al., 2016a). For instance, *PIK3CA* activating mutations are more profoundly detected in breast than lung cancers, while *EGFR* mutations are more common in the latter. In addition, *BRAF* mutations are found predominantly in melanoma

and colon as compared to other tumour types (Garraway and Sellers, 2006). On the other hand, frequency of amino acid changes at codon 12 and 13 caused by substitution of guanine with adenine, cytosine and thymine in *KRAS2* differ remarkably between cancer types (Sieber et al., 2005). Furthermore, the activation of Wnt signalling pathway resulted from mutations of its regulators has been tightly associated with cancer development (Nusse, 2005, Yang, 2012). The pathway regulators affected by mutations vary across cancer types, for example adenomatous polyposis coli (*APC*) in colorectal tumours (Ilyas et al., 1997, Iwao et al., 1998, Sparks et al., 1998), axis inhibitor 1 (*AXIN1*) in sporadic medulloblastoma (Zurawel et al., 1998, Huang et al., 2000, Dahmen et al., 2001) and β-catenin in sporadic endometrial cancers and hepatoblastoma (Fukuchi et al., 1998, Kobayashi et al., 1999, Koch et al., 1999). Therefore, re-differentiation of cancer-iPSCs carrying driver mutations through embryoid body formation and lineage-specific differentiation provides an opportunity to investigate the lineage specificity of pre-existing mutations in breast cancer cells. This could potentially provide general understanding of breast carcinogenesis as well as important insight into specific breast cancer therapies.

6.2 RESULTS

6.2.1 Embryoid body (EB) formation

To determine the differentiation potential of all iPSCs, *in vitro* differentiation was performed. The ESC line H9 was used as a positive control (Figure 6.1). For each ESC and iPSC line, 5000 - 10000 cells were plated onto round bottom ultra-low attachment 96-well plate. Embryoid bodies (EBs) were left to form for two days before being cultured as suspension in polyhema-coated T25 culture flask 4 days. Subsequently, EBs were transferred to culture vessel coated with gelatin to facilitate EB attachment and expansion. After 20 days, cells were trypsinised and harvested for downstream analyses by RT-PCR and immunofluorescence staining for markers of three germ layers. All characterised iPSC lines were tested with the exception of 10A_PIK3CA$^{(-/+)}$/TP53$^{(-/-)}$ cell lines as these were derived near the end of the project and time constraints did not allow the establishment of EBs/mammary organoids and other functional tests that will be described below.

All iPSC lines tested demonstrated differentiation potential to the three germ layers as indicated by RT-PCR results (Figure 6.1 to Figure 6.5, b panels), although some difference in the degree of expression of the markers tested was found. H9-derived EBs (H9_EB) expressed only two of the ectoderm markers (*ALDH1A1* and *COL1A1*). In comparison, the breast cancer iPSC EBs expressed more markers, possibly due to their ectoderm lineage origin. The ectoderm markers expression was more pronounced in the EBs compared to the undifferentiated iPSCs, with the exception of

10A_PIK3CA$^{(+/-)}$ R4 cells as they did not express either *PAX6* or *DCX*. For the mesoderm, H9_EB expressed all the markers tested. *DCN* and *GATA2* had a similar expression profile in EBs derived from breast cancer iPSCs compared to H9, whereas variation was observed for *IGF2* and *ACTA2* expression. For instance, both were not expressed in 5PLA_BTR5 EBs and *ACTA2* was absent in 5PLA_MCF7R EBs. Among the endoderm markers, the expression of *AFP* and *SOX7* in cancer iPSC EBs were consistent to H9. However, both markers were not expressed in 5PLA_MCF7R and 10A_PIK3CA$^{(+/-)}$ R4 EBs. *GATA6* was only detected in H9_EB and 10AR_EB, while *GATA1* and *CDH1* expression did not change consistently between undifferentiated cells and differentiated EBs.

Table 6.2 Expression of three germ layer markers by all pluripotent stem cells and their embryoid bodies (at day 26) determined by RT-PCR. n = 1.

Markers	H9	H9_EB	10AR	10AR_EB	5PLA_BTR5	5PLA_BTR5_EB	5PLA_MCF7R	5PLA_MCF7R_EB	10A_PIK3CA(+/-) R4	10A_PIK3CA(+/-) R4_EB
Ectoderm										
ALDH1A1	-/+	+	+	+	-/+	+	-/+	+	+	+
COL1A1	-/+	+	+	+	+	+	+	+	+	+
PAX6	-	-	-	+	-/+	+	-/+	+	-	-
DCX	-	-	-/+	+	-/+	+	-	+	+	-
Mesoderm										
DCN	-	+	-	+	-	+	-	+	-	+
GATA2	-	+	-	+	-	+	-/+	+	-/+	+
BMP4	-/+	+	+	+	+	+	+	+	+	+
IGF2	-/+	+	-	+	-/+	-	+	+	+	+
ACTA2	-	+	-/+	+	-	-	-	-	+	+
Endoderm										
GATA1	-	-	+	+	-	-	-	-	-	-
SOX7	-/+	+	-	-	-/+	+	-	+	+	-/+
CDH1	+	+	+	+	+	+	-/+	+	+	+
AFP	-	+	-	+	-	+	-	-	-	-
GATA6	-	+	-	+	-	-	-	-	-	-

"-" = negative, "-/+" = weak positive, "+" = strong positive

The RT-PCR results were complemented by immunofluorescence staining for Nestin (NES) for ectoderm, N-cadherin (CDH2) for mesoderm and α-fetoprotein (AFP) for endoderm (Figure 6.1 to Figure 6.5, c panels) (Table 6.3). The expression of NES was induced during EB differentiation for H9 and 10AR cell lines, but only weakly in 5PLA_BTR5_EB and 5PLA_MCF7R _EB or not expressed in 10A_PIK3CA$^{(+/-)}$ R4_EB. CDH2 expression was also present in some of the undifferentiated iPSC lines, but generally induced across all EBs. The expression of AFP however was consistent with the RT-PCR data whereby it was induced in all EBs with the exception of 5PLA_MCF7R_EB and 10A_PIK3CA$^{(+/-)}$ R4_EB. Therefore, based on the markers tested, it could be concluded that the differentiation potential of the 10AR cells was similar to that of the H9 ESC line (with the exception of the expression of SOX7). Among the cancer iPSC lines, 5PLA_MCF7R and 10A_PIK3CA$^{(+/-)}$ R4 demonstrated the lower differentiation potential, based mainly on the negative/low expression of ectoderm marker NES and the endoderm markers AFP and GATA6.

Table 6.3 Immunofluorescence staining for markers of three germ layers. n = 1.

Markers	NES	CDH2	AFP
H9	-/+	-/+	-
H9_EB	+	+	+
10AR	-	-/+	-
10AR_EB	+	+	+
5PLA_BTR5	-	-/+	-
5PLA_BTR5_EB	-/+	+	+
5PLA_MCF7R	-	-	-
5PLA_MCF7R_EB	-/+	+	-
10A_PIK3CA$^{(+/-)}$ R4	-	-	-
10A_PIK3CA$^{(+/-)}$ R4_EB	-	+	-/+

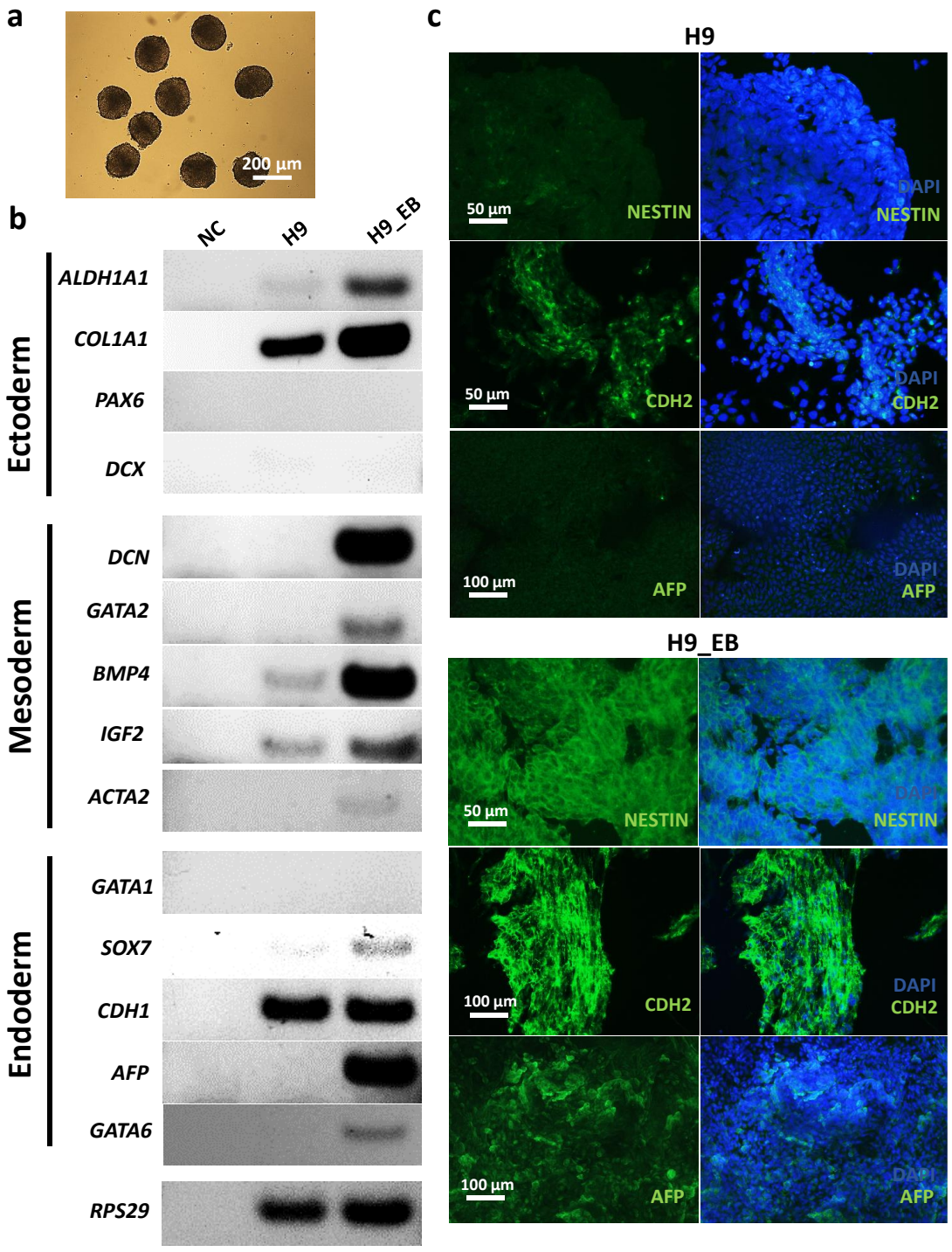


Figure 6.1 H9 *in vitro* differentiation.

(a) Embryoid bodies (EBs). (b-c) RT-PCR analysis and immunofluorescence staining for markers of three germ layers in H9 and H9-derived EBs (H9_EB). NESTIN, CDH2 and AFP are markers for ectoderm, mesoderm and endoderm. Ability to generate cells of three germ layers indicates H9 is pluripotent. n = 1. Secondary antibody-only negative control is shown in Figure S3-3.

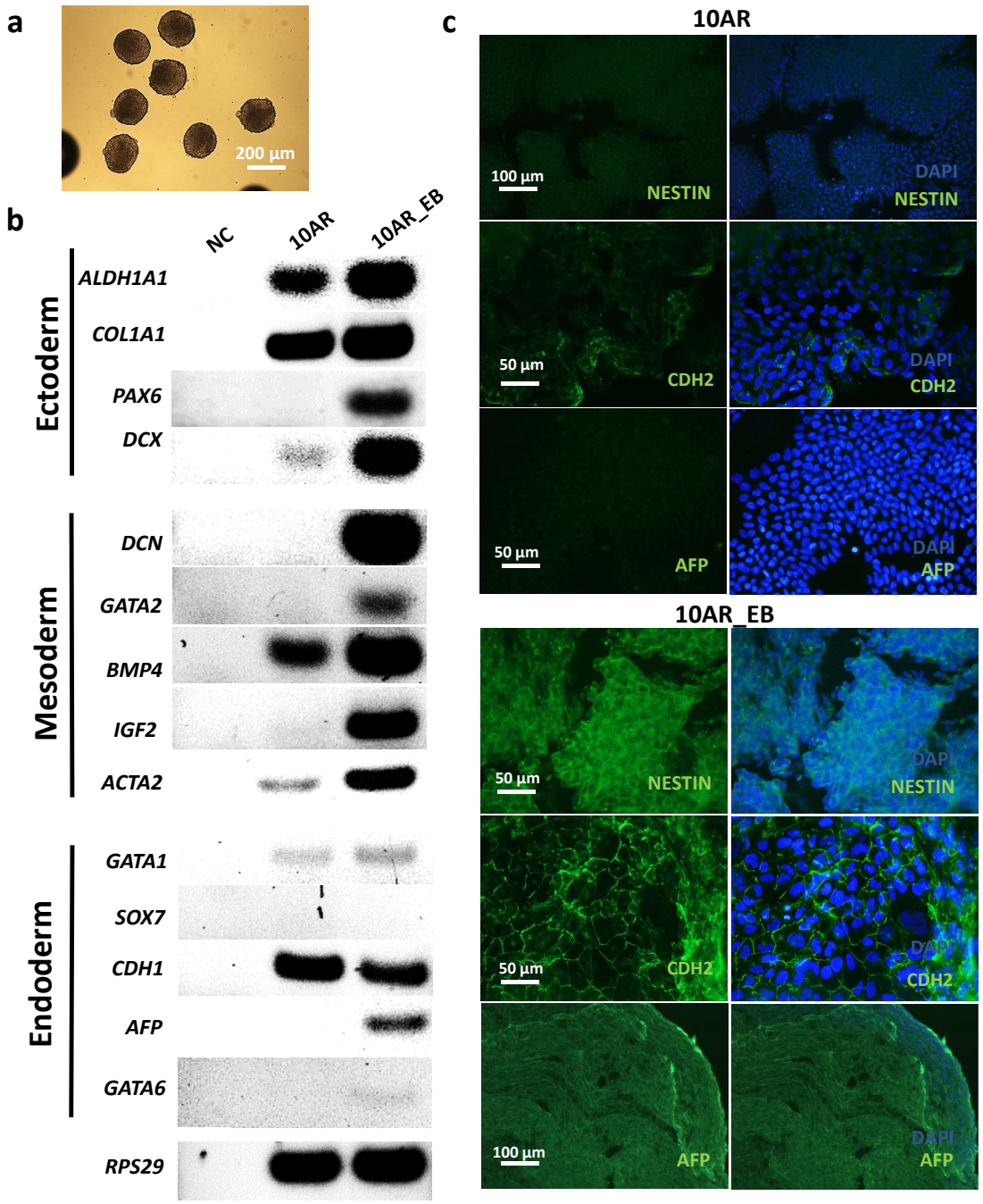


Figure 6.2 10AR *in vitro* differentiation.

(a) Embryoid bodies (EBs). (b-c) RT-PCR analysis and immunofluorescence staining for markers of three germ layers in 10AR and 10AR-derived EBs (10AR_EB). NESTIN, CDH2 and AFP are markers for ectoderm, mesoderm and endoderm. Ability to generate cells of three germ layers indicates 10AR is pluripotent. n = 1. Secondary antibody-only negative control is shown in Figure S3-3.

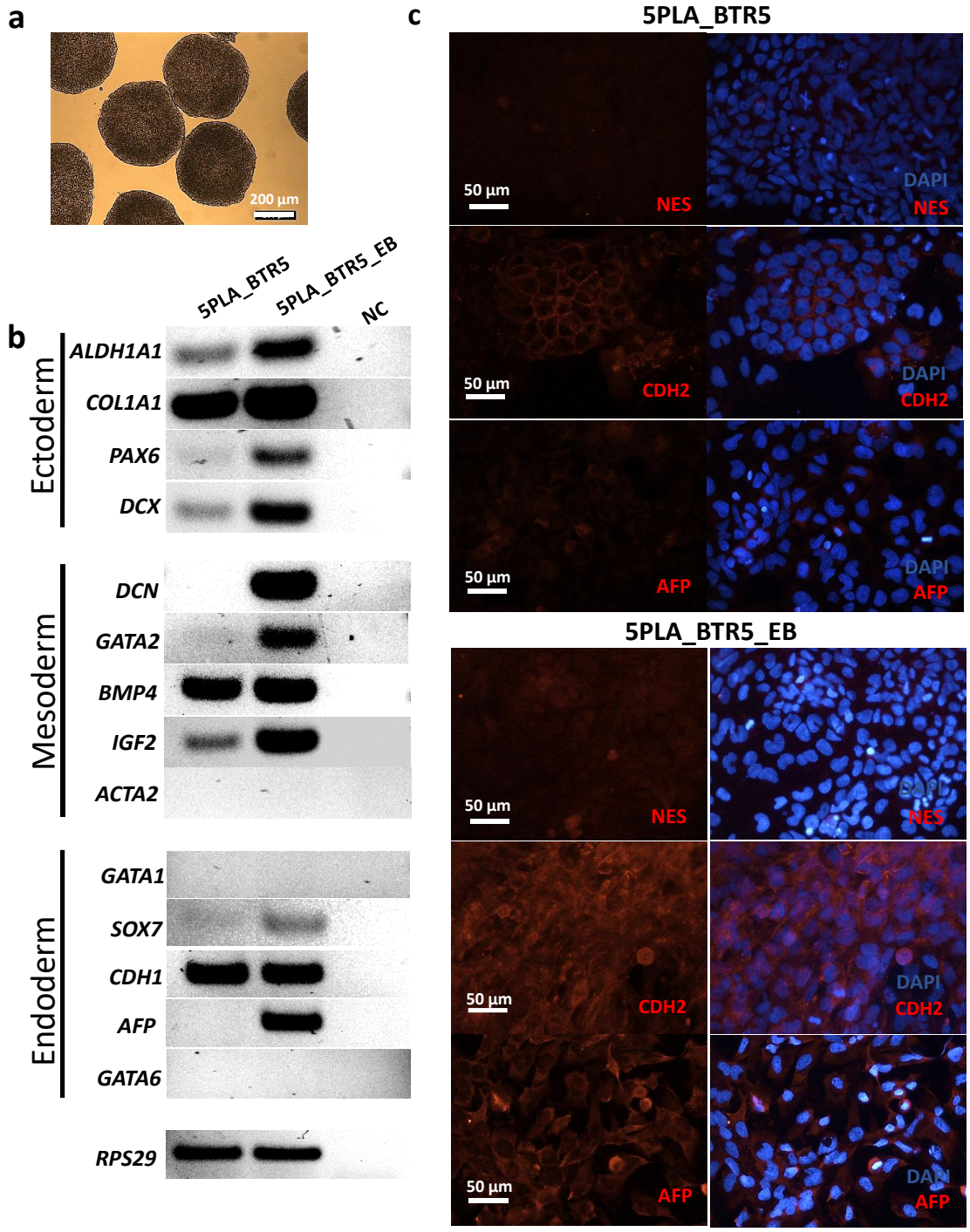


Figure 6.3 5PLA_BTR5 *in vitro* differentiation.

(a) Embryoid bodies (EBs). (b-c) RT-PCR analysis and immunofluorescence staining for markers of three germ layers in 5PLA_BTR5 and 5PLA_BTR5-derived EBs (5PLA_BTR5 _EB). NESTIN, CDH2 and AFP are markers for ectoderm, mesoderm and endoderm. Ability to generate cells of three germ layers indicates 5PLA_BTR5 is pluripotent. n = 1. Secondary antibody-only negative control is shown in Figure S3-3.

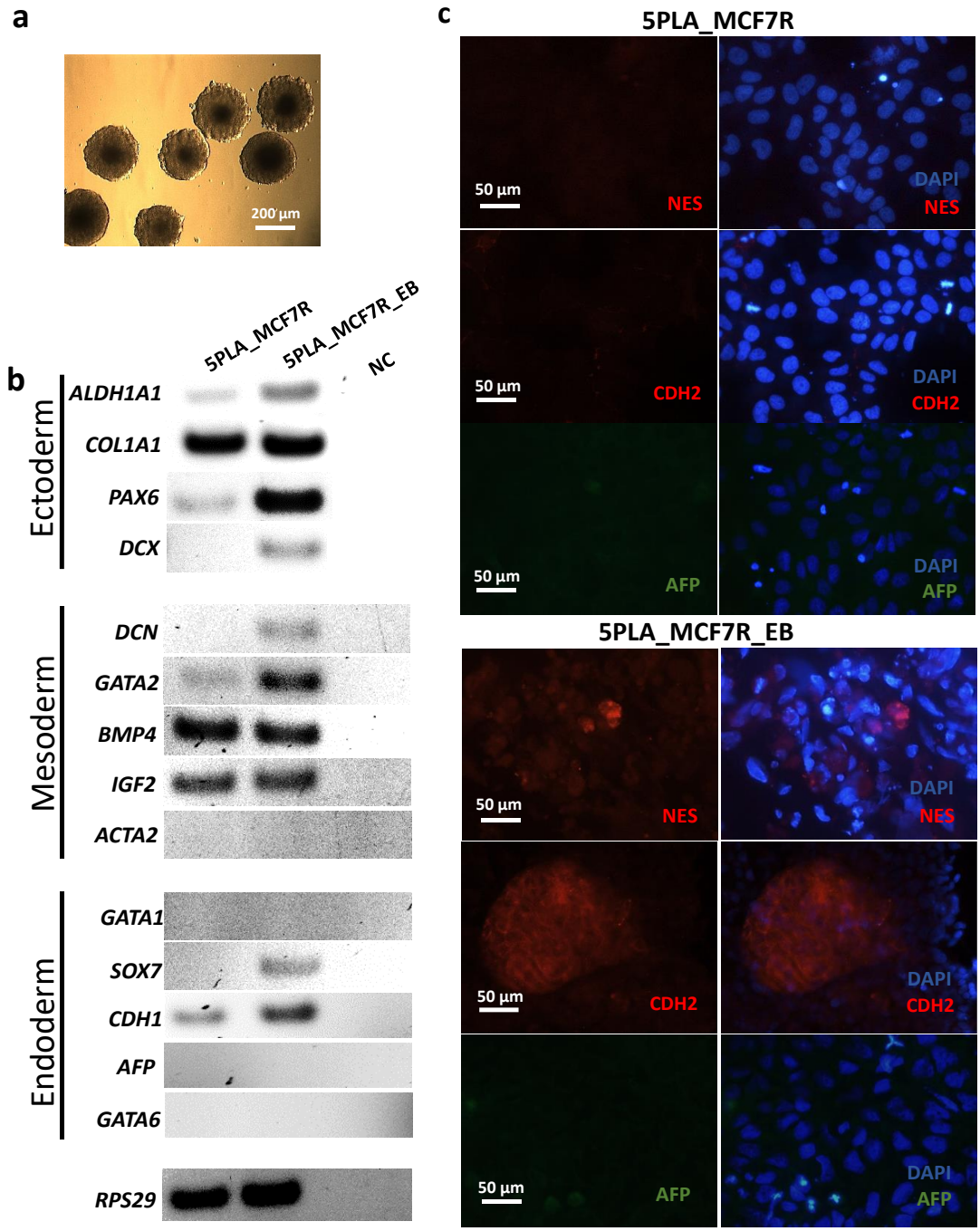


Figure 6.4 5PLA_MCF7R *in vitro* differentiation.

(a) Embryoid bodies (EBs). (b-c) RT-PCR analysis and immunofluorescence staining for markers of three germ layers in 5PLA_MCF7R and 5PLA_MCF7R-derived EBs (5PLA_MCF7R_EB). NESTIN, CDH2 and AFP are markers for ectoderm, mesoderm and endoderm. Absence of AFP^+ cells and poor overall endoderm markers expression indicates that 5PLA_MCF7R does not achieve ESC-like pluripotency. n = 1. Secondary antibody-only negative control is shown in Figure S3-3.

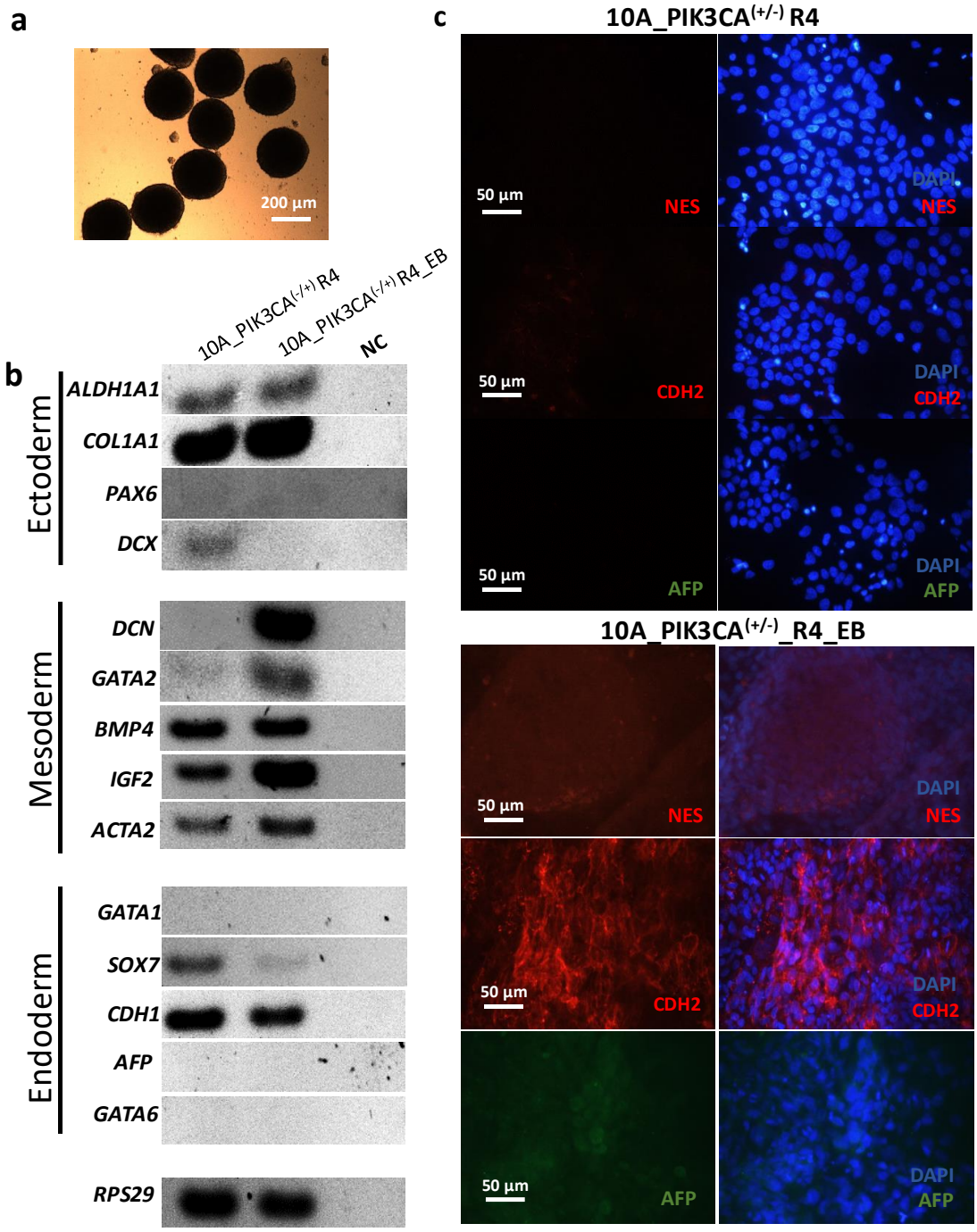


Figure 6.5 10A_PIK3CA$^{(+/-)}$ R4 *in vitro* differentiation.

(a) Embryoid bodies (EBs). (b-c) RT-PCR analysis and immunofluorescence staining for markers of three germ layers in 10A_PIK3CA$^{(+/-)}$ R4 and 10A_PIK3CA$^{(+/-)}$ R4_EB. Absence of NES^{+} cells and poor overall three germ layers markers expression indicates that 10A_PIK3CA$^{(+/-)}$ does not achieve ESC-like pluripotency. n = 1. Secondary antibody-only negative control is shown in Figure S3-3.

6.2.2 Directed differentiation to generate mammary organoids

To test the terminal lineage differentiation ability of iPSCs, a protocol for mammary lineage-directed differentiation described by Qu et al. (2017) was used to generate mammary organoid capable of milk protein secretion. The first step of the two stages protocol involved culturing iPSCs with mammoshpere medium (an ectoderm-favourable medium) to generate and promote ectodermal differentiation in mammary EBs (mEBs). Next, mEBs were trapped in floating mixed gel (made up of Matrigel and Collagen I) and cultured consecutively in three different media composed of horse serum-based MCF10A medium supplemented with three different combinations of growth factors and/or hormones for 5 days (parathyroid hormone), 20 days (fibroblast growth factor 10, FGF10, and hepatocyte growth factor, HGF) and 5 days (prolactin and foetal bovine serum) (Figure 6.6).

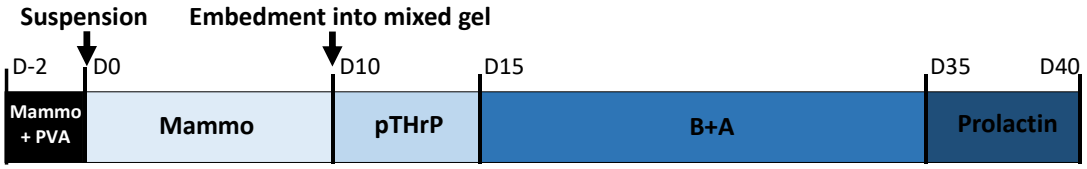


Figure 6.6 Timeline of mammary organoid generation.
Mammo = mammosphere medium, PVA = poly(vinyl alcohol), pTHrP = parathyroid hormone-supplemented medium, B+A = fibroblast growth factor 10- (FGF10) and hepatocyte growth factor- (HGF) supplemented medium for branch and alveolar formation, Prolactin = prolactin supplemented medium.

Mammary organoid formation from the ESC line H9 is illustrated in Figure 6.7. Protrusions were observed before day 15 around the mEBs and eventually developed into branch and alveolar structures mimicking

mammary glands. At day 40, the round-shaped mEBs turned into a complex mesh of branches. A similar developmental process was observed in all iPSC lines (Figure 6.8 to 6.11).

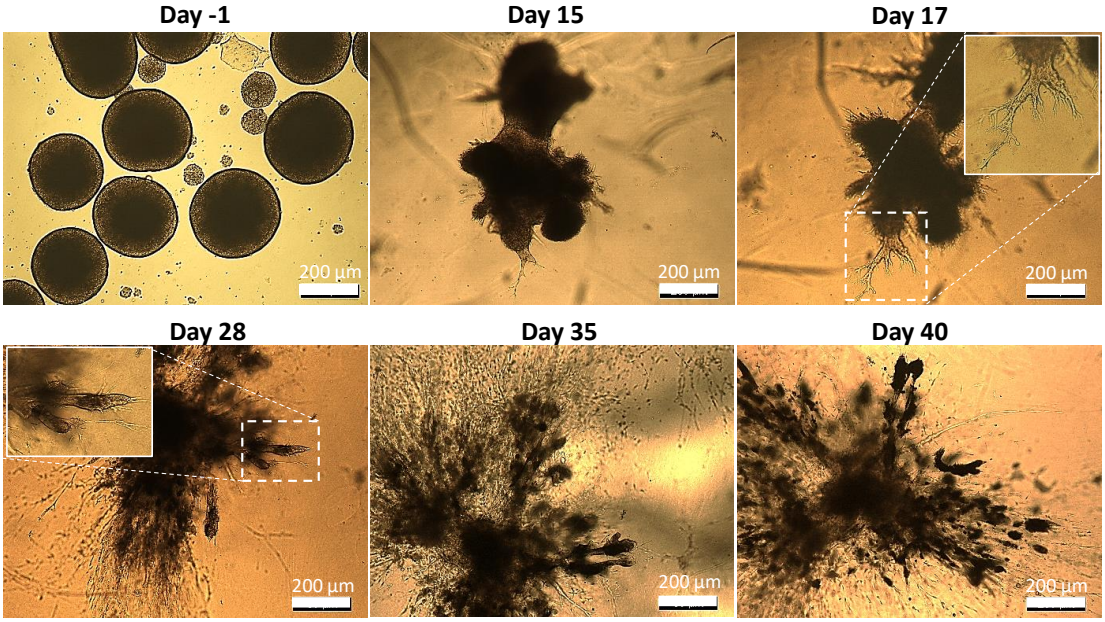


Figure 6.7 Mammary organoid formation from H9.
Process was captured at different time points illustrating formation of branch and alveolar structures mimicking mammary gland. n = 1.

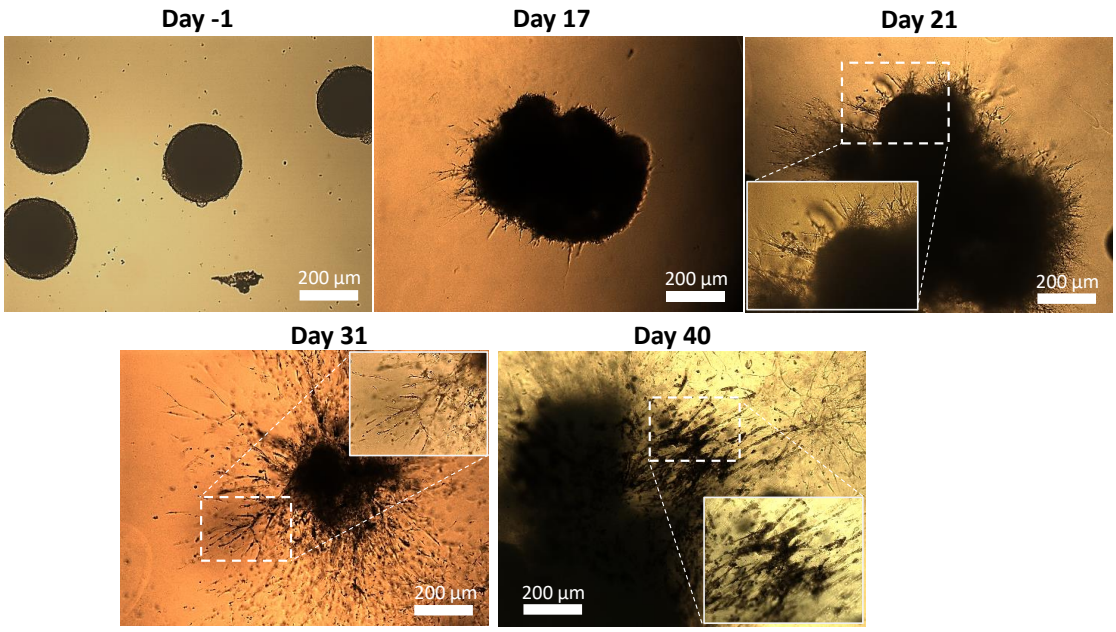


Figure 6.8 Mammary organoid formation from 10AR.

Process was captured at different time points illustrating formation of branch and alveolar structures mimicking mammary gland. n = 1.

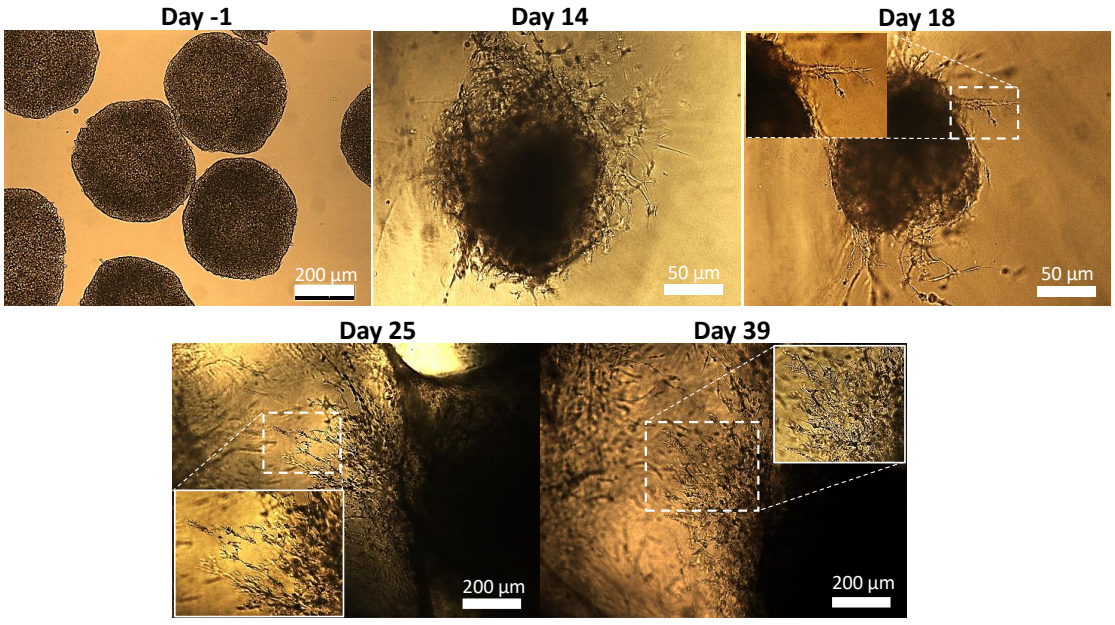


Figure 6.9 Mammary organoid formation from 5PLA_BTR5.

Process was captured at different time points illustrating formation of branch and alveolar structures mimicking mammary gland. n = 1.

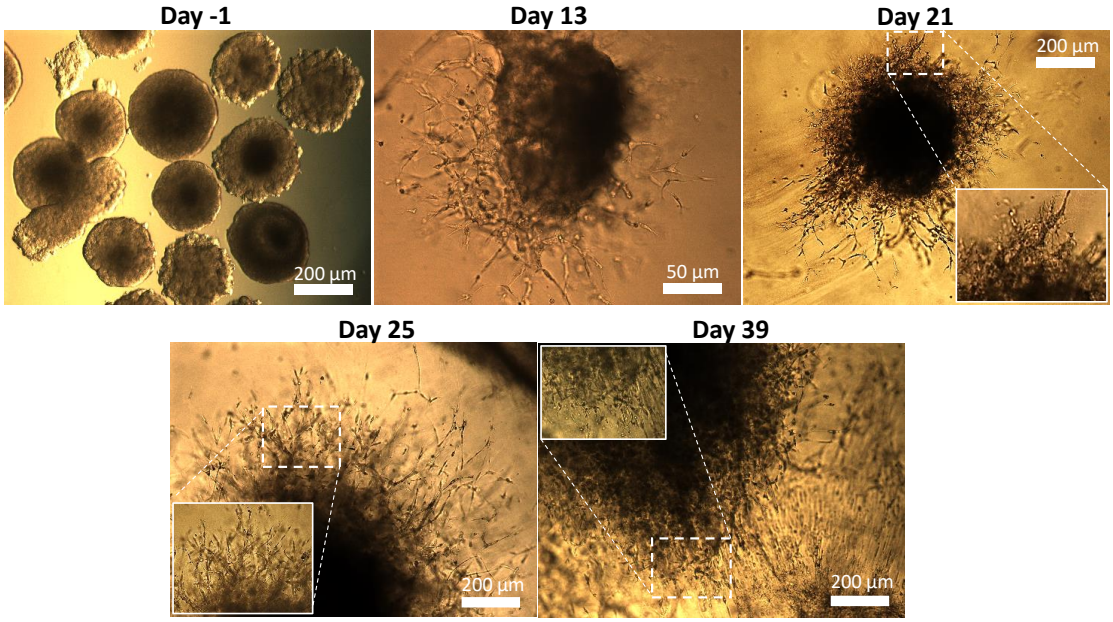


Figure 6.10 Mammary organoid formation from 5PLA_MCF7R.

Process was captured at different time points illustrating formation of branch and alveolar structures mimicking mammary gland. n = 1.

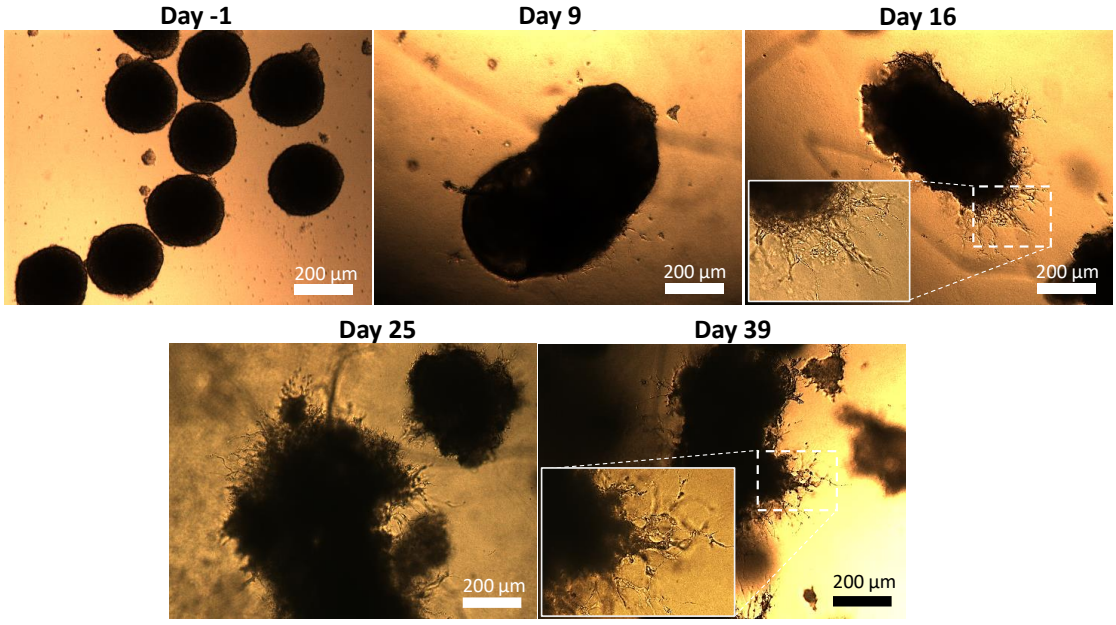


Figure 6.11 Mammary organoid formation from 10A_PIK3CA$^{(+/-)}$ R4.

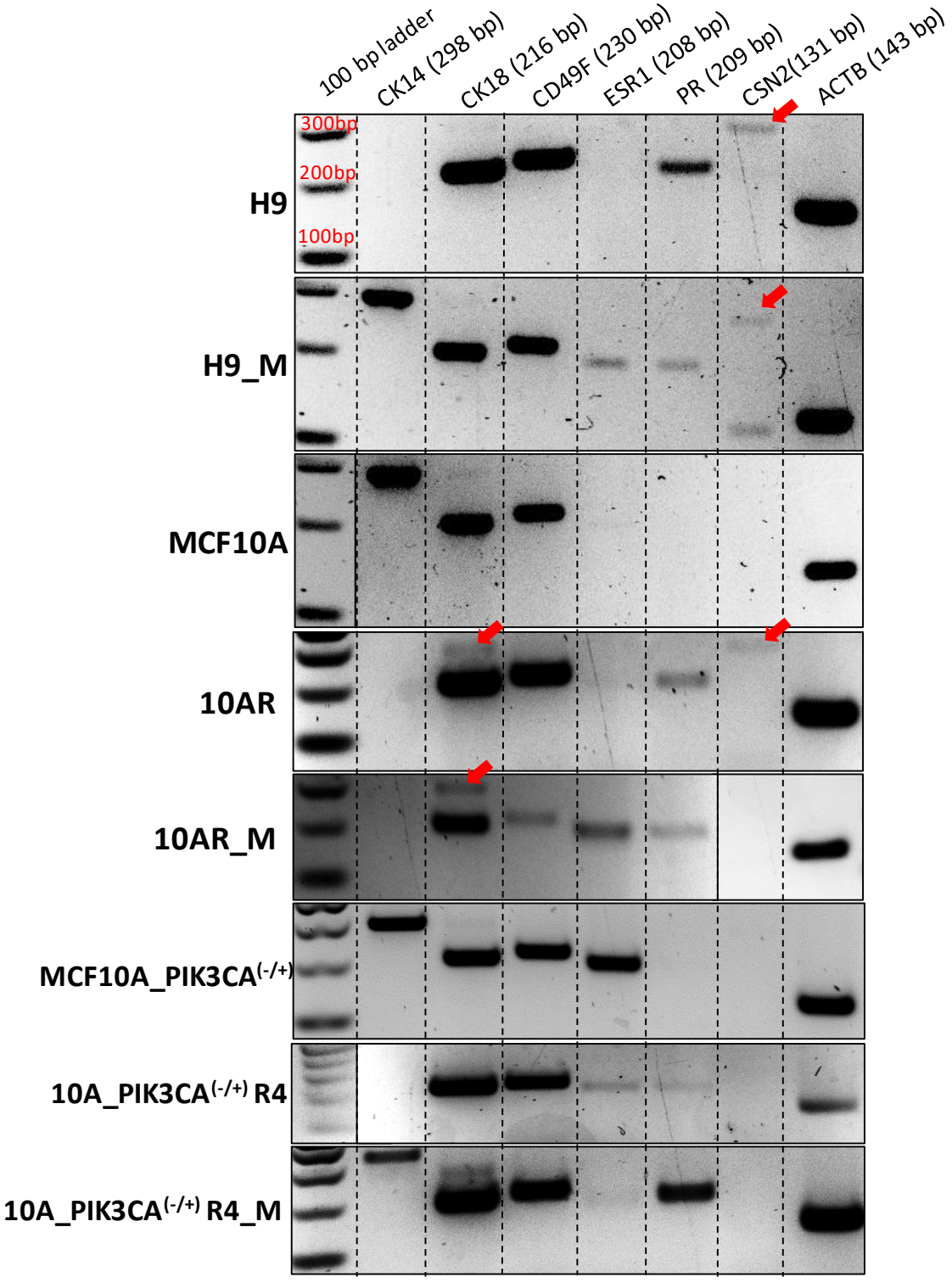
Process was captured at different time points illustrating formation of branch and alveolar structures mimicking mammary gland. n = 1.

Subsequently, organoids were harvested and subjected to RT-PCR analysis to determine the expression of mammary markers, including cytokeratin 14 (*CK14*, a basal marker), *CK18* (a luminal marker), intergrin-α 6 (*CD49f*, human basal breast stem cells marker), oestrogen and progesterone receptors (*ESR1* and *PR*, respectively; both hormones are master regulators of breast development) and *β-casein* (*CSN2*, encodes for milk protein) (Figure 6.12 and Table 6.4). All samples expressed *CK18* and *CD49f* and all pluripotent stem cells (PSCs), including H9, were found expressing *PR*. In H9 cells, the expression of *CK14* (basal marker), *ESR1* and *CSN2* were induced after organoid formation, indicating successful derivation of mammary-like tissues. In MCF10A, *CK14* was lost after reprogramming and *PR* acquired. After organoid formation, 10AR_M did not acquire *CK14* or *CSN2*, however *ESR1* was induced. The expression of *CSN2* was also negative in MCF10A_PIK3CA$^{(+/-)}$ R4. For the cancer iPSC derived organoids, *ESR1* and *CSN2* were detected but *CK14* was not expressed in 5PLA_BTR5. Therefore, all the cell lines tested demonstrated a potential for mammary differentiation, based on the expression of *ESR1* and *PR*, although terminal differentiation indicated by the expression of *CSN2* was only achieved for normal ESC cells (H9), 5PLA_MCF7R and 5PLA_BTR5.

Table 6.4 Expression of mammary markers by mammary organoids generated from different PSCs lines determined by RT-PCR. n = 1.

Markers	*CK14*	*CK18*	*CD49F*	*ESR1*	*PR*	*CSN2*
H9	-	+	+	-	+	-
H9_M	+	+	+	+	+/-	+
MCF10A	+	+	+	-	-	-
10AR	-	+	+	-	+	-
10AR_M	-	+	+	+	+	-
BT-549	-	+	+	-	+	-
5PLA_BTR5	-	+	+	+/-	+	-
5PLA_BTR5_M	-	+	+	+	+	+
MCF7	-	+	+	+	+	-
5PLA_MCF7R	-	+	+	+/-	+	-
5PLA_MCF7R_M	+	+	+	+	+	+
MCF10A_PIK3CA$^{(+/-)}$	+	+	+	+	-	-
10A_PIK3CA$^{(+/-)}$ R4	-	+	+	+	+/-	-
10A_PIK3CA$^{(+/-)}$ R4_M	+	+	+	+/-	+	-

"-" = negative, "-/+" = weak positive, "+" = strong positive


100 bp ladder
CK14 (298 bp)
CK18 (216 bp)
CD49F (230 bp)
ESR1 (208 bp)
PR (209 bp)
CSN2(131 bp)
ACTB (143 bp)
300bp
200bp
100bp
H9
H9_M
MCF10A
10AR
10AR_M
MCF10A_PIK3CA(-/+)
10A_PIK3CA(-/+) R4
10A_PIK3CA(-/+) R4_M

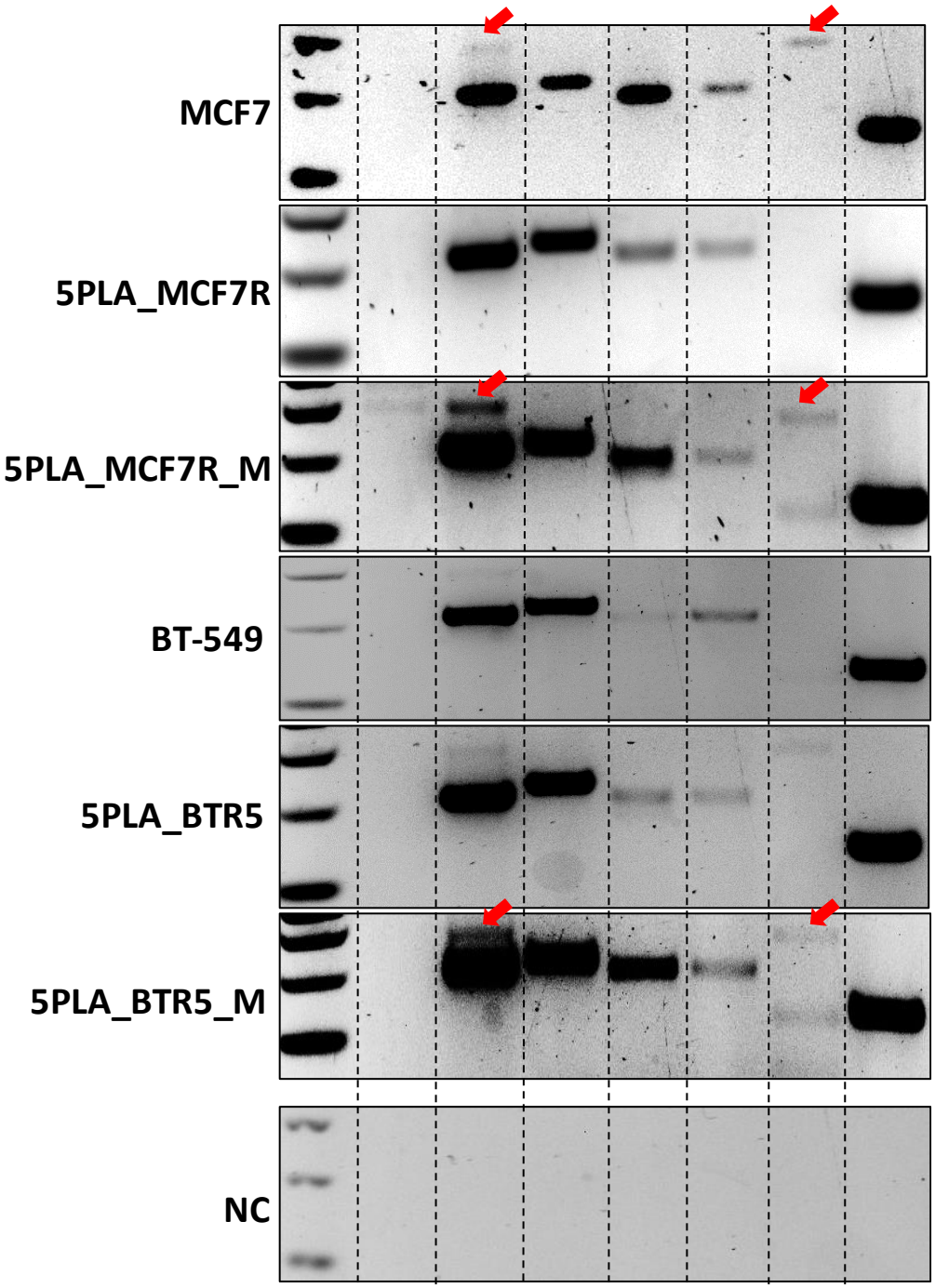


Figure 6.12 RT-PCR analysis for mammary markers expression in ESC- and iPSC-derived mammary organoids as compared to their parental and pluripotent counterparts.

Letter "M" at the end of label indicates mammary organoid generated from respective PSC lines. Red arrow = non-specific PCR products, NC = negative control. n ≥ 3.

6.2.3 CD24 and CD44 surface markers expression

Within the breast, CD24 and CD44 mark epithelial and mesenchymal cells, respectively (Shirasawa et al., 1993, Ye et al., 2005). A profile of $CD44^{+}/CD24^{-/low}$ mark breast CSCs and both markers have been used in breast CSCs and iPSC reprogramming studies (Lee et al., 2012, Hussein et al., 2014, Tonge et al., 2014, Morath et al., 2016). Therefore, the CD24/CD44 profile of all iPSCs, their parental and re-differentiated cells were determined to assess lineage differentiation and potential transformation of iPSCs into CSCs. Two populations of re-differentiated cell were prepared for each iPSC line. They were both generated through EB formation and then cultured either in horse serum-based MCF10A medium (referred to as medium 1, m1) or in EB medium used in the last section (referred to as medium 2, m2). MCF10A medium was used to skew EB differentiation potential towards the mammary lineage, whereas medium 2 was used to evaluate the spontaneous differentiation into the three germ layers. Data was expressed in percentage and as mean ± standard deviation (sd). Representative FACS analysis data for individual series of cell lines was shown in Figure 6.13 to Figure 6.16. Results are summarised and presented as a histogram in Figure 6.17.

In the iPSC lines 10AR and 5PLA_BTR5, there was a significant reduction in the $CD44^{+}/CD24^{-}$ and $CD44^{+}/CD24^{+}$ population with a concomitant increase in the $CD44^{-}/CD24^{+}$ cells compared to the parental cell line. This may indicate a change from a mesenchymal to epithelial phenotype and a possible reduction in CSCs. No significant difference in

$CD44^+/CD24^-$ cells was found for 5PLA_MCF7R and 10A_PIK3CA$^{(+/-)}$ R4 cells but there was a significant increase of $CD44^-/CD24^+$ cells in both cases.

Re-differentiation performed in m2 resulted in cells with profiles more similar to parental cells than the pluripotent counterpart in 10AR and 5PLA_BTR5, suggesting loss of the pluripotent phenotype and recapitulation of the parental one. For example, MCF10A and 10AR_EB_m2 had similar populations of $CD44^+/CD24^-$ and $CD44^-/CD24^-$, while only the $CD44^-/CD24^+$ was similar between 10AR and 10AR_EB_m2. In addition, none of the four populations were significantly different between BT-549 and 5PLA_BTR5_EB_m2, while only $CD44^-/CD24^-$ population was similar between 5PLA_BTR5 and 5PLA_BTR5_EB_m2. 5PLA_MCF7 cells could only be tested in m2 medium as cells failed to grow in m1 medium. Even in this case, recapitulation of parental profile was also observed (all four populations were not significantly different) in 5PLA_MCF7R_EB_m2.

In contrast, re-differentiation of 10AR cells in m1 was highly similar to the undifferentiated counterpart (all comparisons were not statistically significant) but remarkably different (all comparisons were statistically significant) from the parental cells. 5PLA_BTR5 cells responded the same to m1 as in m2 medium (highly similar to parental and different from undifferentiated cells) but with the $CD44^+/CD24^-$ ($p<0.03$) different from parental cell line, suggesting possible loss of CSCs property.

In 10A_PIK3CA$^{(+/-)}$ R4, the profile was similar after differentiation in either m1 or m2 media, with a decrease in $CD44^{-}/CD24^{+}$ and increase in $CD44^{-}/CD24^{-}$ cells ($p<0.0001$) as compared to their pluripotent counterpart. This profile was substantially different from the parental cell line where the predominant population of cells was $CD44^{+}/CD24^{+}$. (Figure 6.13 – 6.16). Therefore, in general, re-differentiation performed with the medium m2 resulted in cells highly similar to the parental cells. However, 10A_PIK3CA$^{(+/-)}$ R4 cells demonstrated a predominant $CD24^{-}CD44^{-}$ profile rather than $CD24^{+}CD44^{+}$ observed in the parental cells, suggesting that the presence of the mutation may affect their differentiation potential.

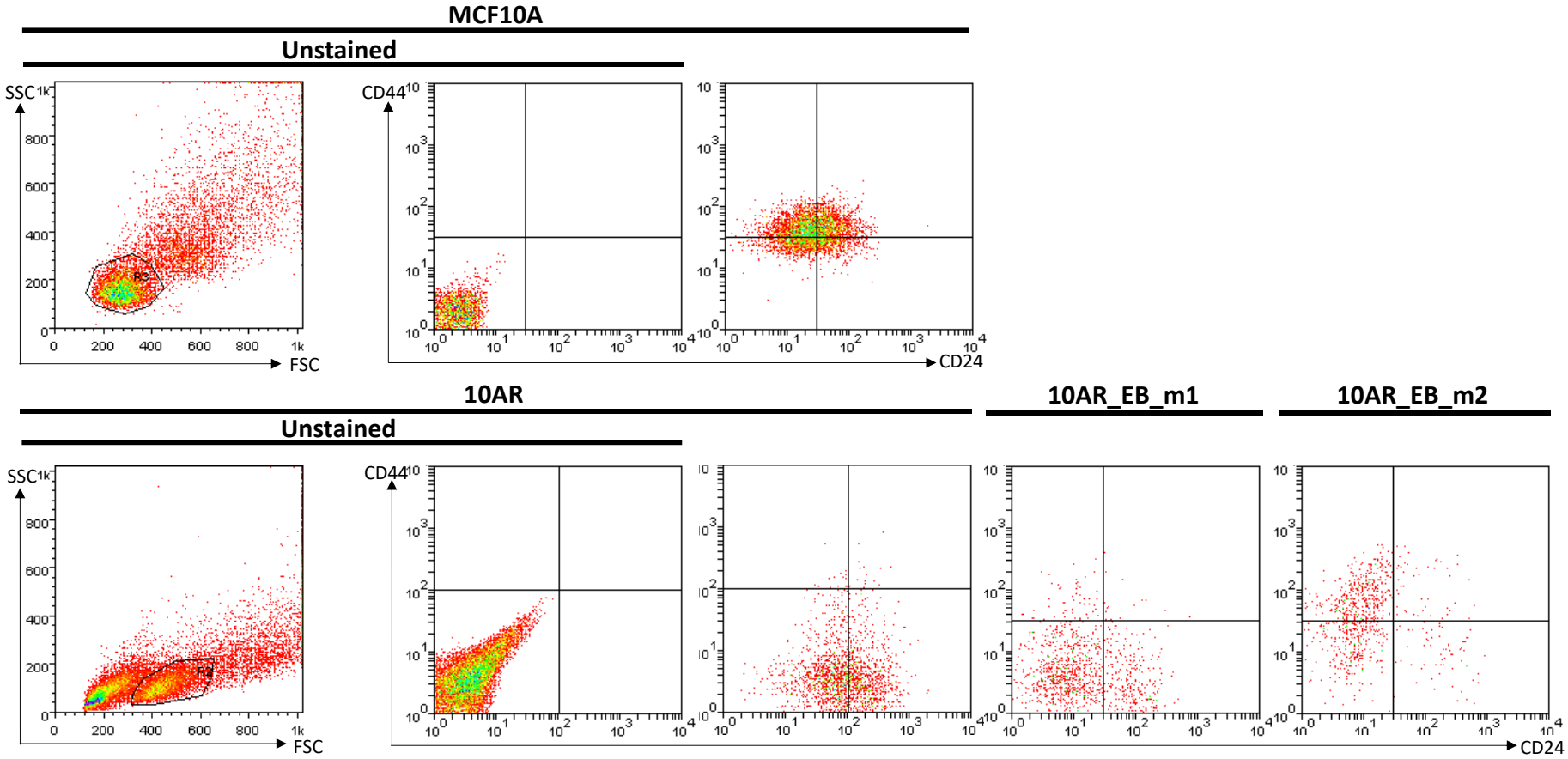


Figure 6.13 Representative FACS analysis of CD24 and CD44 surface markers expression in MCF10A, iPSCs and differentiated derivatives.

Differentiated cells were gated according to parental unstained control. n = 2. Dr Cinzia Allegrucci contributed to the generation of these data.

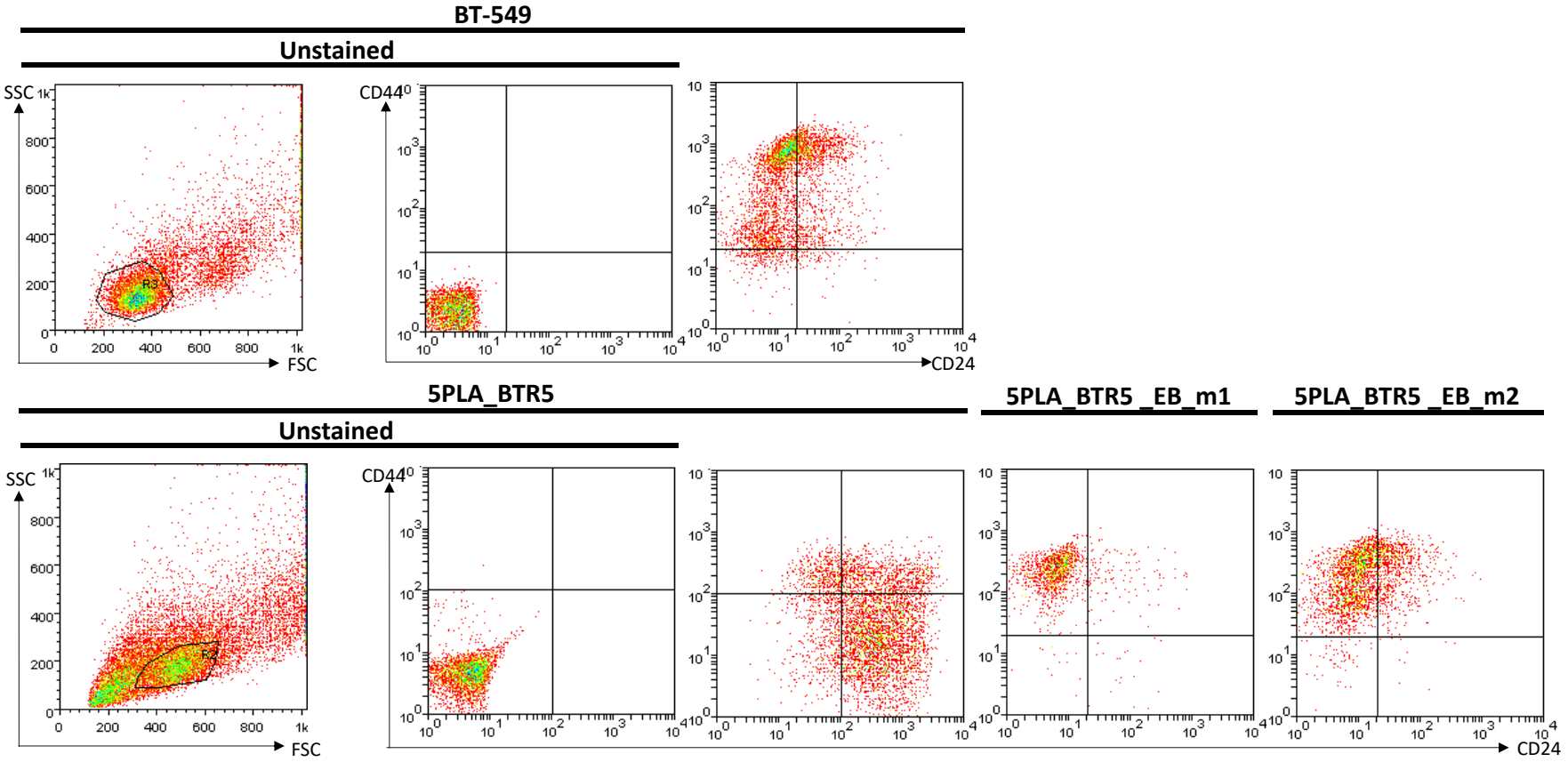


Figure 6.14 Representative FACS analysis of CD24 and CD44 surface markers expression in BT-549, iPSCs and differentiated derivatives.

Differentiated cells were gated according to parental unstained control. n = 2. Dr Cinzia Allegrucci contributed to the generation of these data.

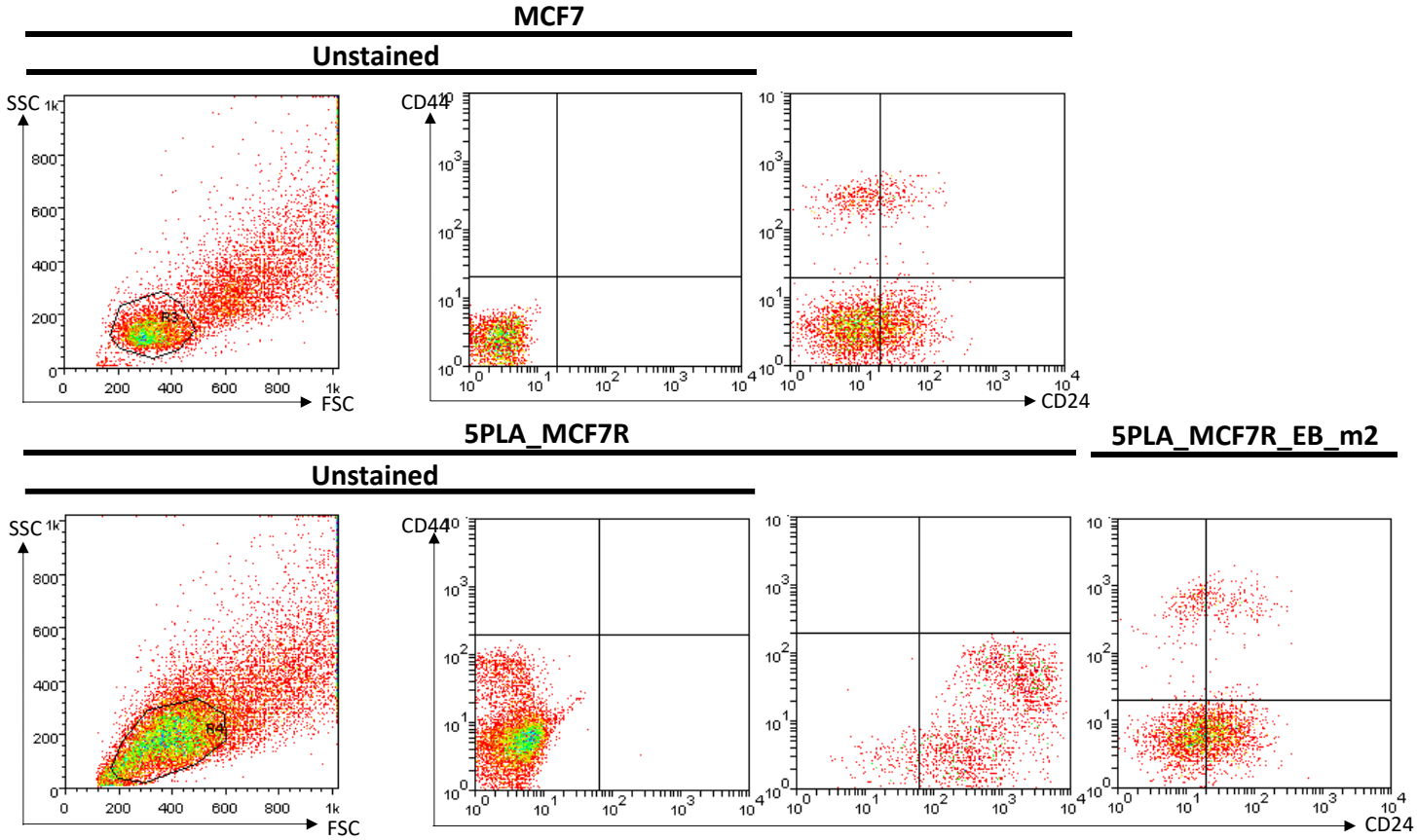


Figure 6.15 Representative FACS analysis of CD24 and CD44 surface markers expression in MCF7, iPSCs and differentiated derivatives.

Differentiated cells were gated according to parental unstained control. n = 2. Dr Cinzia Allegrucci contributed to the generation of these data.

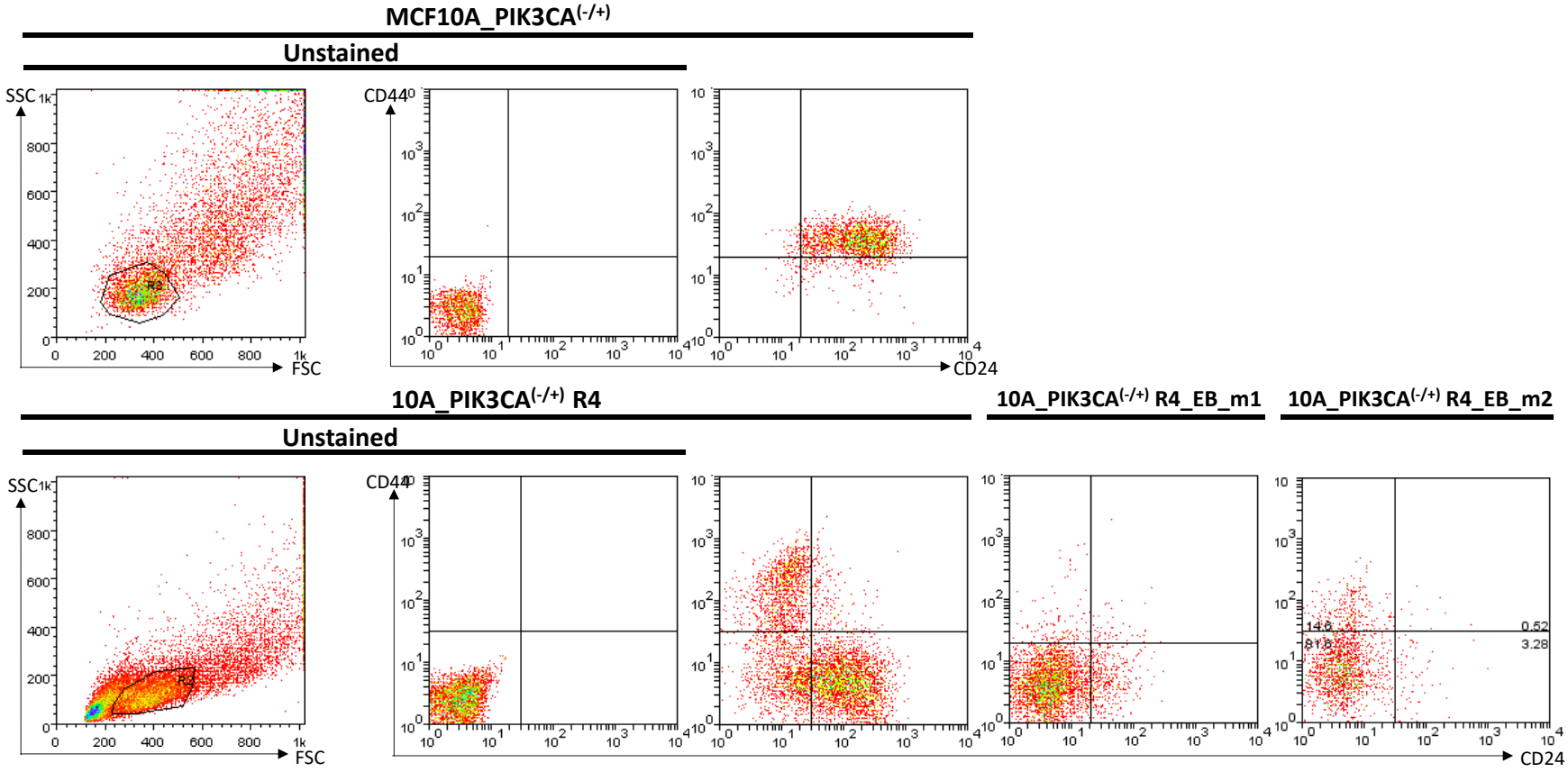


Figure 6.16 Representative FACS analysis of CD24 and CD44 surface markers expression in MCF10A_PIK3CA$^{(+/-)}$, iPSCs and differentiated derivatives.

Differentiated cells were gated according to parental unstained control. n = 2. Dr Cinzia Allegrucci contributed to the generation of these data.

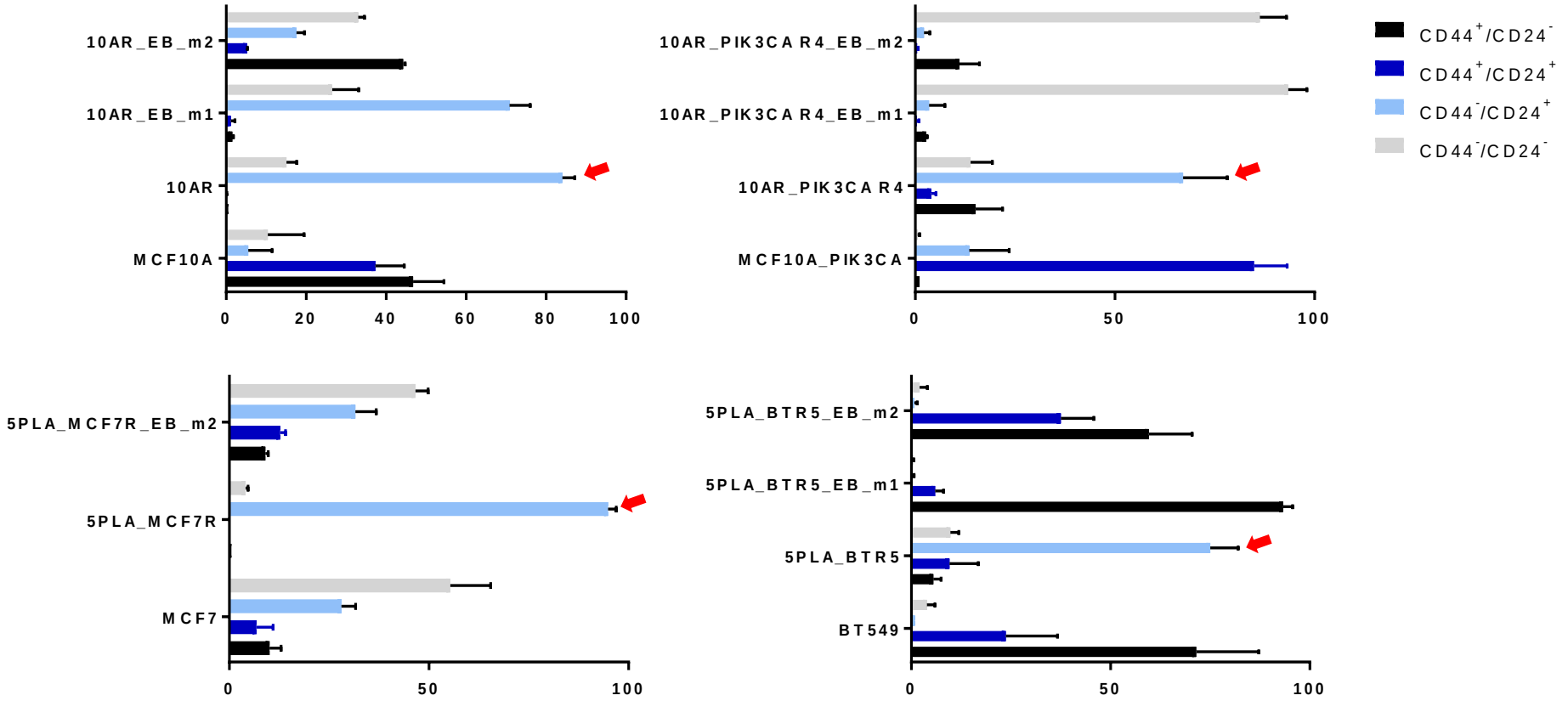


Figure 6.17 CD44/CD24 profiles of parental, iPSCs and re-differentiated cell lines.

An increase of $CD44^{-}/CD24^{+}$ can be observed in all iPSC lines, suggesting acquisition of ESC-like phenotype. (n = 2). Dr Cinzia Allegrucci contributed to the generation of these data.

6.2.4 Soft agar colony formation assay

Soft agar colony formation assay was next performed to assess the impact of iPSC reprogramming on breast cancer *in vitro* tumorigenicity. Colony forming ability of all parental cell lines, their iPSCs as well as their differentiated derivatives (cultured in m1 or m2) was studied. Cells were seeded in 1 ml of 0.5% noble agar in respective culture media overlaying 1 ml 1% agar in the same medium. After 2 weeks culture, cell colonies were stained with 0.05% crystal violet in 10% ethanol/PBS for 1 hour and colonies ≥100 μm counted under stereomicroscope. Results were presented as mean ± sd, n = 3. A total of 25000 cells (MCF10A, MCF7 and MCF10A_PIK3CA$^{(+/-)}$) or 5000 cells (BT-549) were seeded per 6 well. Results indicated a gain of growth in anchorage dependent conditions for MCF10A and MCF10A_PIK3CA$^{(-/+)}$ cells and a loss for MCF7 and BT-549. For MCF10A, differentiated derivatives (but not the pluripotent counterpart) cultured in m1 and m2 produced significantly more colonies ($p<0.0001$ in both cases) as compared to the parental cell (Figure 6.18). For MCF7, growth was remarkably suppressed in iPSCs ($p<0.0001$ in both E8 and m2 systems) as well as differentiated counterpart cultured in m1 ($p<0.0001$) and to a lesser extent in m2 ($p<0.0001$) (Figure 6.19). For BT-549, colony forming ability was suppressed in all pluripotent and differentiated derivatives in all medium systems ($p<0.001$ in all cases) (Figure 6.20). For MCF10A_PIK3CA$^{(+/-)}$, reprogramming remarkably enhanced cell colony forming ability. The highest growth was observed when iPSCs was cultured in E8 medium ($p<0.0001$) followed by culture in m2 ($p<0.0001$) and m1 (not significant) (Figure 6.21). Further analysis of MCF10A_PIK3CA$^{(+/-)}$-derived iPSCs suggested culture media used significantly influenced

tumorigenicity (Figure 6.22a). Enhanced colonies growth was also observed in differentiated derivatives and again to a lesser extent in m1 (Figure 6.22a). To identify if the differentiation process was positively associated with tumorigenicity in MCF10A_PIK3CA$^{(+/-)}$, colony forming ability of parental, pluripotent and differentiated cells were compared according to medium system m1 or m2. Differentiated cells were found more tumorigenic than the pluripotent counterpart in both m1 ($p<0.03$ and $p<0.0005$ in iPSC and differentiated cells respectively) and m2 systems ($p<0.0002$ and $p<0.0001$ in iPSC and differentiated cells, respectively (Figure 6.22b).

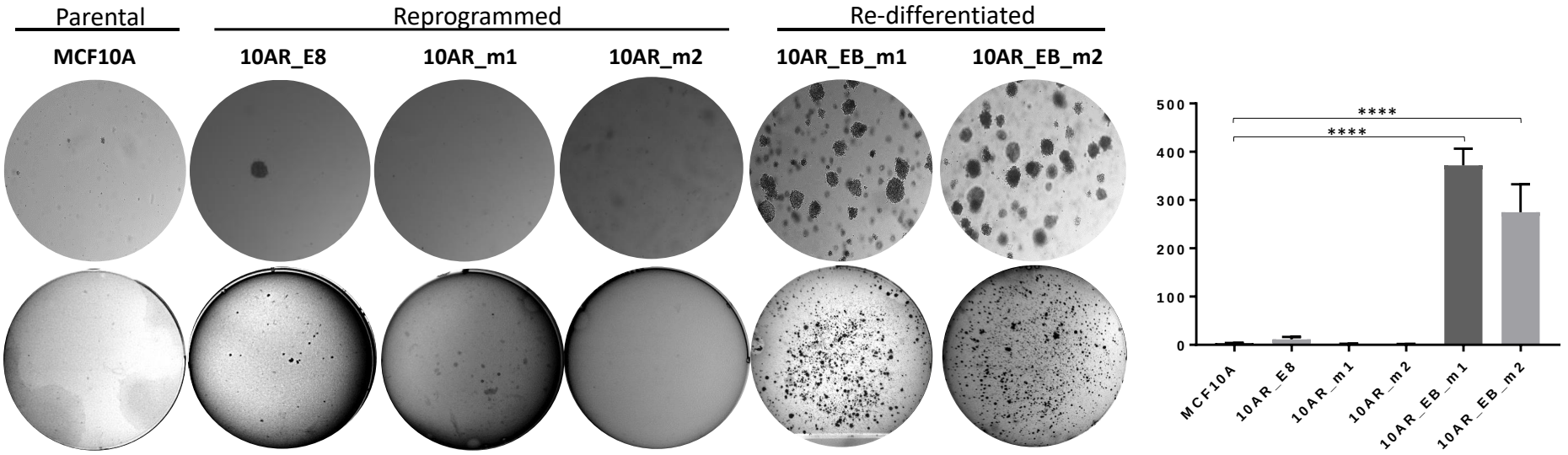


Figure 6.18 Colony forming assay of MCF10A, iPSCs and differentiated derivatives.

Tumorigenicity was significantly increased in both differentiated cell lines. Data was analysed by one-way ANOVA followed by Bonferroni correction. **** = $p<0.0001$. n = 3. Dr Cinzia Allegrucci contributed to the generation of these data.

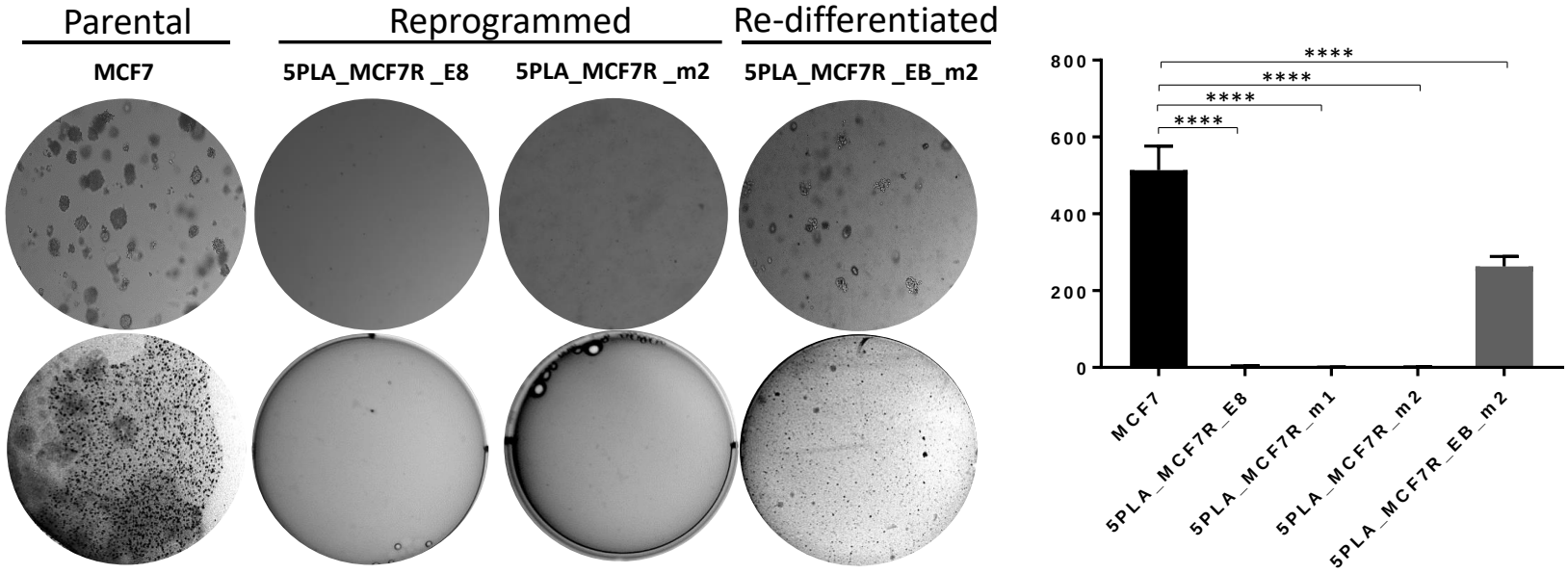


Figure 6.19 Colony forming assay of MCF7, iPSCs and differentiated derivatives.

Tumorigenicity was significantly reduced in iPSCs and both differentiated cell lines, although to a lesser extent in m2. Data was analysed by one-way ANOVA followed by Bonferroni correction. **** = $p<0.0001$. n = 3. Dr Cinzia Allegrucci contributed to the generation of these data.

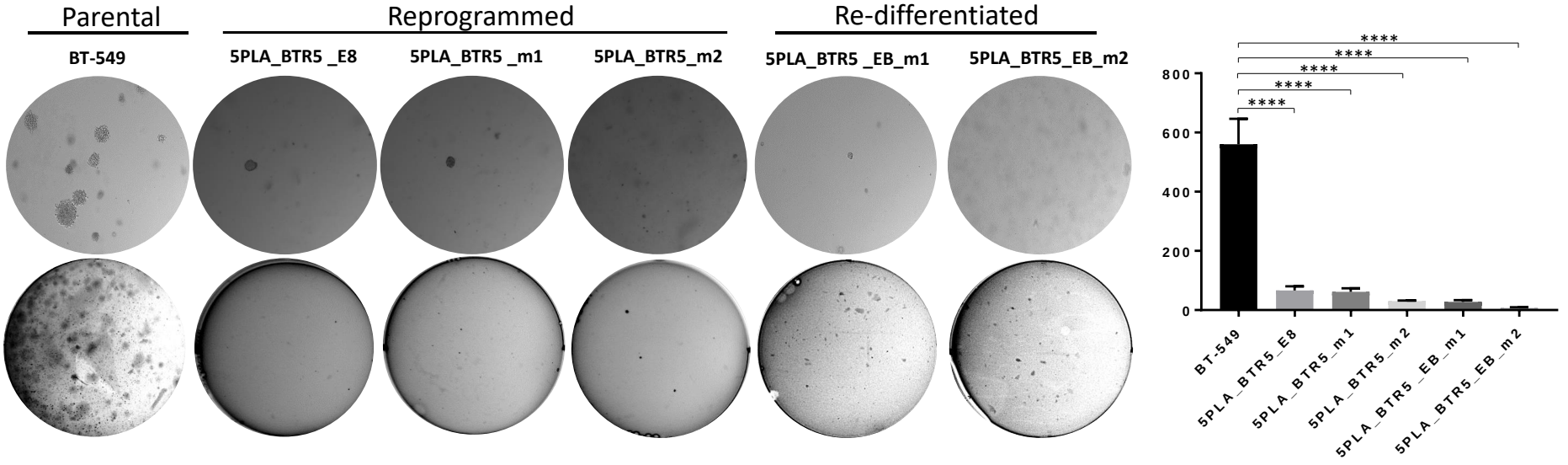


Figure 6.20 Colony forming assay of BT-549, iPSCs and differentiated derivatives.

Tumorigenicity was significantly reduced in all pluripotent and differentiated counterparts. Data was analysed by one-way ANOVA followed by Bonferroni correction. **** = $p<0.0001$. n = 3. Dr Cinzia Allegrucci contributed to the generation of these data.

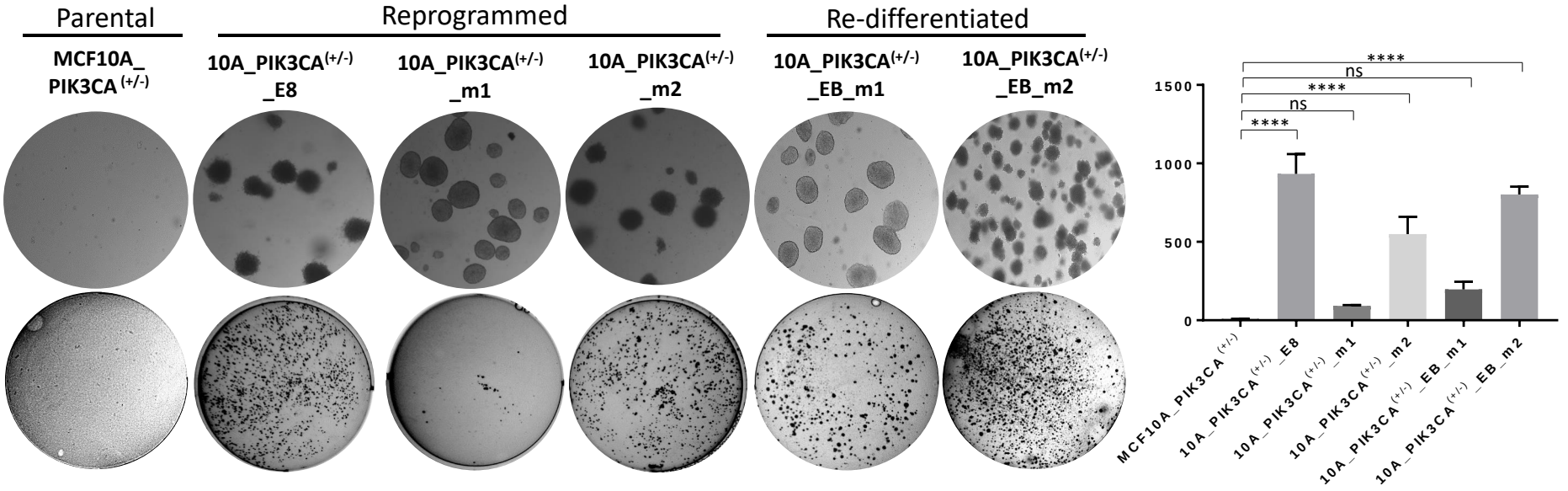


Figure 6.21 Colony forming assay of MCF10A_PIK3CA(+/-), iPSCs and differentiated derivatives.

Tumorigenicity was significantly increased in all pluripotent and differentiated counterparts, except for iPSC and differentiated cells cultured in m1. Data was analysed by one-way ANOVA followed by Bonferroni correction. **** =$p<0.0001$. n = 3. Dr Cinzia Allegrucci contributed to the generation of these data.

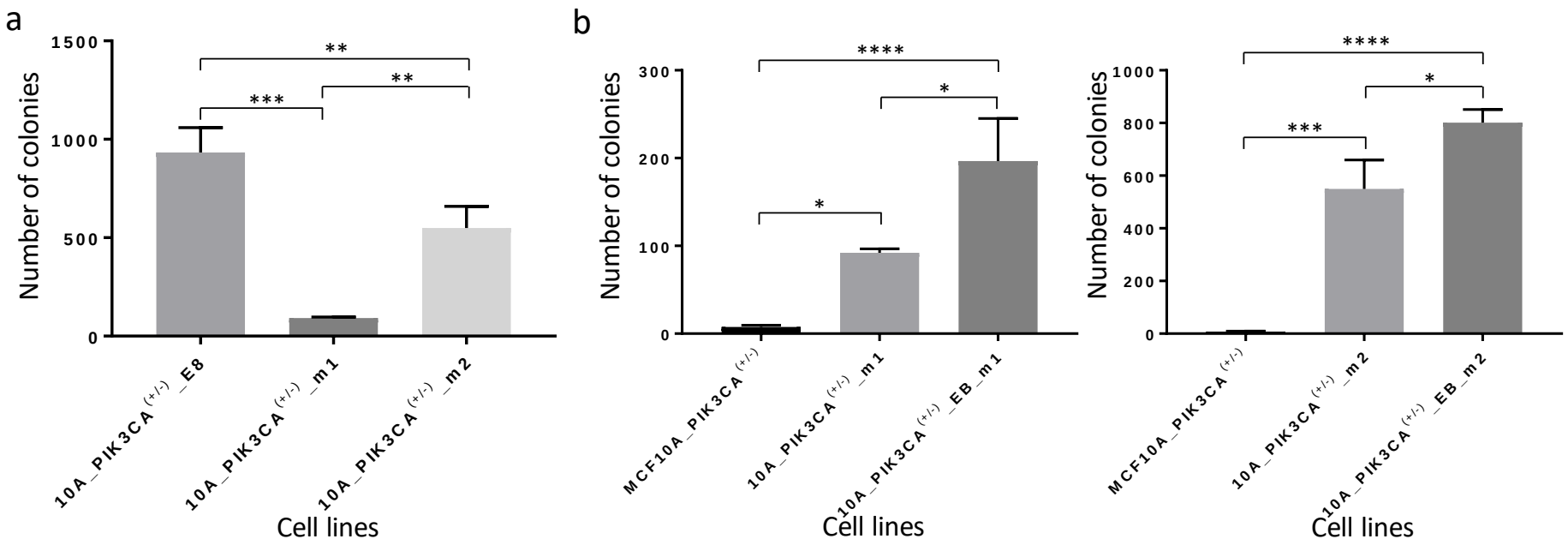


Figure 6.22 Potential effect of medium system, differentiation process and PIK3CA mutation on tumorigenicity.

(a) Effects of different media on colony forming ability of 10AR_PIK3CA$^{(+/-)}$ R4. (b) Differentiation process (in m1 and in m2 medium conditions) enhanced tumorigenicity in 10AR_PIK3CA$^{(+/-)}$ R4. Data was analysed by one-way ANOVA followed by Bonferroni correction. * = $p<0.05$, ** = $p<0.001$, *** = $p<0.0005$, **** = $p<0.0001$. n = 3.

6.2.5 Chemotherapy sensitivity and resistance assays

Next, susceptibility of cells after re-differentiation (in m1 and m2) to the chemotherapy agent paclitaxel was tested. Cells were prepared at 60000 cells per well onto a 24-well plate. A day after plating, cells were treated with paclitaxel at concentrations ranging from 0.001 to 10 nM for 24 hours. Cell viability was evaluated by MTT cell proliferation assay. Results were presented as mean ± standard deviation, n = 3. Treatment of differentiated cells in m1 and m2 media was compared with that of the parental cell lines and the IC_{50} for each cell line and p-values are summarised in Table 6.5. No significant changes were observed in all cell lines except for MCF7, whereby a significant increase in IC_{50} ($p<0.0004$) from 0.07 to 0.11 was observed (Figure 6.23).

Table 6.5 IC_{50} value of paclitaxel for all cell lines. n = 3.

Cell lines	IC_{50}	p-value
MCF10A	0.061	-
10AR_EB_m1	0.051	0.148
10AR_EB_m2	0.053	0.28
MCF7	0.07	-
5PLA_MCF7R_EB_m2	0.11	0.0004*
BT-549	0.19	-
5PLA_BTR5_EB_m1	0.11	0.077
5PLA_BTR5_EB_m2	0.10	0.078
MCF10A_PIK3CA$^{(+/-)}$	0.07	-
10A_PIK3CA$^{(+/-)}$ R4_EB_m1	0.05	0.467
10A_PIK3CA$^{(+/-)}$ R4_EB_m2	0.11	ND

*=statistically significant, ND = not determined

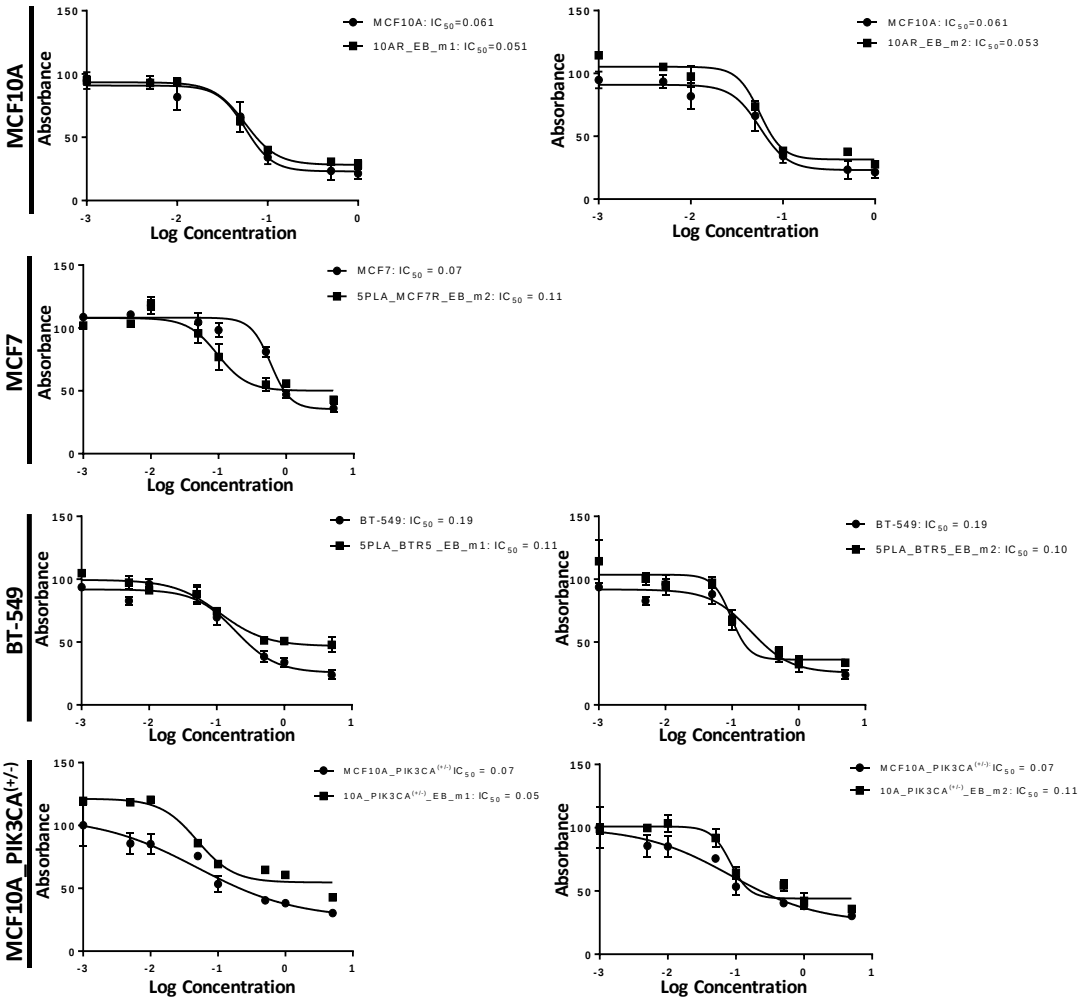


Figure 6.23 IC_{50} value of paclitaxel for parental lines and differentiated derivatives.

Significant change was only observed between MCF7 and 5PLA_MCF7_EB_m2 ($p = 0.0004$). Data was analysed by non-linear regression analysis. n = 3. Dr Cinzia Allegrucci contributed to the generation of these data.

6.3 DISCUSSION

In this chapter, the differentiation potential of all iPSC lines was determined by *in vitro* spontaneous differentiation and directed differentiation to generate mammary organoids. All iPSC lines could give rise to cells of three germ layers despite difference in the degree of marker expression was found. On the other hand, induction of ESR1 and PR expression in mammary organoids indicated all iPSC lines tested were capable of terminal differentiation, however, only normal ESC cells (H9), 5PLA_MCF7R and 5PLA_BTR5 resulted in milk-secreting cells. Analysis of cells CD44/CD24 profile demonstrated that, in general, all iPSC lines acquired a $CD44^{-/low}/CD24^{+}$ profile characteristic of PSCs. Re-differentiation of iPSCs resulted in progenies harbouring similar (BT-549 and MCF7) or distinct (MCF10A and MCF10A_PIK3CA$^{(+/-)}$) profiles from the parental cells, suggesting recapitulation of the parental cellular phenotype or acquisition of different ones. In addition, in the case of 10AR, different differentiation media (m1 or m2) resulted in populations with varying similarities to the parental cells, potentially suggesting the influence of culture medium on cell differentiation. Lastly, results from the soft agar colony forming assay suggested a gain of growth in anchorage dependent conditions for MCF10A and MCF10A_PIK3CA$^{(-/+)}$ cells and a loss for breast cancer cell lines MCF7 and BT-549 after reprogramming, which might correlate with the loss of sensitivity towards paclitaxel in re-differentiated MCF7 cells. A summary of all results in this chapter is summarised in Table 6.6.

Table 6.6 A summary of results obtained in Chapter 6.

Cell lines	Assays				
	EB formation	**Mammary differentiation**	**CD24/CD44 expression**	**Colony formation**	**Drug sensitivity**
10AR	+++	+	+++	Higher in redifferentiated cells	Unchanged after reprogramming
5PLA_BTR5	+++	++	+++	Lower in redifferentiated cells	Unchanged after reprogramming
5PLA_MCF7R	++	+++	+++	Lower in redifferentiated cells	Less susceptible after reprogramming
10A_PIK3CA$^{(+/-)}$ R4	++	++	+++	Higher in redifferentiated cells	Unchanged after reprogramming
Abbreviations	+++ = comparable to H9 ++ = less comparable to H9	+++ = comparable to H9 ++ = less comparable to H9 + = weak	+++ = profile highly similar to H9		

6.3.1 Differentiation potential of iPSC lines

Differentiation potential represents one important hallmark of pluripotent stem cells (PSCs) and can be assessed *in vitro* by embryoid body formation to generate cells of three germ layer (De Los Angeles et al., 2015). Human PSCs differentiation potential is defined as the capability of cells to give rise to progenies of the three primordial germ layers (ectoderm, mesoderm, and endoderm) of the embryos (Marti et al., 2013). The teratoma assay is regarded as the gold standard for pluripotency evaluation (Przyborski, 2005). It involves the implantation of PSCs into an immunocompromised mouse to generate a benign tumour, the teratoma. Pluripotent cells give rise to teratoma consisting of tissues or cells representing three embryonic germ layers and the expression of germ layer markers are normally analysed by immunocytochemistry and gene expression profiling (Wesselschmidt, 2011). However, the necessity of teratoma assay in pluripotency assessment remains controversial. Several reports suggested that the combination of several or all of the following pluripotency tests such as the expression of pluripotency markers, the epigenome status, directed or spontaneous *in vitro* differentiation and the bioinformatics approach (such as the PluriTest) based on genome-wide data sets would be sufficient to robustly predict the pluripotency of PSCs (Muller et al., 2011, Buta et al., 2013, Marti et al., 2013). In addition, the lack of standardisation of teratoma assay methodology (such as the number of cells injected, the degree of host immunodeficiency and site of injection) and the ethical concerns also challenge the use of teratoma assay in routine pluripotency evaluation (Muller et al., 2010, Dressel, 2011, Buta et al., 2013). Interestingly, the incapability of sarcomas-derived (Zhang et al., 2013) and breast cancer cell-

derived (Corominas-Faja et al., 2013) iPSCs to generate benign teratomas, but successful derivation of terminally differentiated cells (bone and fat cells) from sarcoma-iPSCs suggest that the teratoma assay might not be the best approach for differentiation potential evaluation of cancer iPSCs. Therefore, in this study, *in vitro* spontaneous (embryoid body formation into three germ layers) and directed differentiation (into milk-secreting mammary organoids) were performed (Qu et al., 2017).

All differentiation markers, except *ACTA2* and *GATA6*, used in this study were described by D'Antonio et al. (2017a). In that study, the authors identified twelve defining markers (four for each germ layer) and used them for RT-qPCR analysis to generate a heat map from which "germ layer scores" (and thus differentiation potential) of PSC lines were quantified. Notably, the collective effect of the expression of four genes for each germ layer was taken into consideration rather than individual expression. In this study, the levels of marker expression were not quantified and the differentiation potential of all iPSC lines were estimated based on the number of germ layer markers expressed by different EBs-derived from iPSC lines as shown by RT-PCR analysis and immunofluorescence staining and. Therefore, qRT-PCR analysis, as reported by Qu et al. (2017), should be performed to specifically quantify differentiation potential of iPSCs.

Nevertheless, results indicated differences in differentiation potential to generate cells of three germ layers in different iPSC lines which closely related to their respective pluripotent states as indicated by analyses of pluripotency markers expression. For example, 5PLA_MCF7R did not

expressed *REX1* and only expressed TRA-1-60 at low level and also showed poor differentiation potential in this chapter (Figure 6.4). On the other hand, one interesting finding analysing EBs using the 12 markers was the consistently high *CDH1* (endoderm) expression in iPSCs. In fact, upregulation of *CDH1* has been detected at early stages of iPSC reprogramming during mesenchymal to epithelial transition (MET) (Liang and Zhang, 2013a), and it is associated with successful reprogramming. (Chen et al., 2010b) reported that *CDH1* can replace *OCT3/4* in the OSKM cocktail (Redmer et al., 2011) and is a hESC specific marker capable of differentiating hESC from EB and other differentiated derivatives. Therefore, *CDH1* might not be a stringent marker for endodermal lineage as suggested by Qu et al. (2017).

Results of mammary organoid formation indicated that only the ESC line H9 and one iPSC line, the 5PLA_MCF7R, generated organoids expressing all mammary markers tested. Interestingly, Qu et al. (2017), the same study the mammary organoid differentiation protocol was derived from, reported the absence of a discernible basal layer (marked by p63, a basal marker) and ESR1 protein expression in organoids cultured under lactogenic condition (with the supplement of prolactin). The authors explained that the basal layer is normally disrupted (Yallowitz et al., 2014) and ESR1 protein expression suppressed (Hatsumi and Yamamuro, 2006) in lactating human breast. However, these organoids still contained a group of $CK14^{+}$ (a basal marker) cells which thought to be keratinocytes based on morphology and localisation, suggesting that the proposed culture system could potentially

give rise to non-mammary epithelial cells. In contrast to these findings, all organoids generated in the present study expressed *ESR1* at the mRNA level with or without the concomitant expression of *CSN2*. However, it is possible the mRNA expression could not translate to protein expression as reported in the previous study. To confirm this, immunohistochemistry (IHC) staining such that performed by Qu et al. (2017) should be performed on the organoids generated in the present study. The same IHC experiment should also be performed with CK14 antibody and other basal markers, such as p63 which specifically target basal population of the mammary epithelial cells, should be employed to verify the identity of the *CK14*-expressing cells obtained in this study. Nevertheless, *CSN2* expression detected in H9-, 5PLA_BTR5- and 5PLA_MCF7R-derived organoids indicated that mammary identity has been successful acquired. In the case of 10AR and 10A_PIK3CA$^{(+/-)}$, however, the absence of *CSN2* expression could be explained by the acquisition of a tumorigenic phenotype (discussed in detailed in the next section), which might have interfered with the normal differentiation process to generate milk-secreting cells.

6.3.2 CD24 and CD44 marker expression

CD24 and CD44 have been widely used to isolate breast cancer stem cells (CSCs) along with other surface markers as discussed in Section 1.2.1.3. CD24 was first identified in murine hematopoietic and neural cells as heat-stable antigen linked to the cells membrane through a glycosylphosphatidylinositol (GPI) anchor and thought to be involved in signalling pathways leading to the activation of cell aggregation (Sammar et al., 1997). In normal cells, CD24 is known to mark epithelial cell type (Ye

et al., 2005). In breast cancer, low expression of CD24 has been associated with CSCs and PSCs features (such as expression of pluripotency markers and differentiation potential) and thus tumour aggressiveness and poor prognosis (Baumann et al., 2012, Lee et al., 2012). In addition, CD24 has also been implicated in iPSC reprogramming. A recent study by Tonge et al. (2014) demonstrated the gradual increase of CD24 expression (accompanied by accumulation of H3K4me3 and H3K36me3 activation marks at the CD24 locus) shortly after OSKM induction of MEF reprogramming and reduced expression (with the appearance of H3K27me3 repressive mark) when ESC-like state was attained. A subsequent study suggested $CD24^{high}$ as a unique marker that can distinguish the F-class mouse iPSC (miPSC) population (a less superior pluripotent state) from ESC-like miPSCs. The same study also identified high CD24 expression in primed hESCs and hiPSCs (both pluripotent states are less superior than the naïve state) and low expression in human naïve-ESCs (a pluripotent state that mimics hESC *in vivo*) (Shakiba et al., 2015), indicating regulation of CD24 expression is largely conserved in both mouse and human systems.

The CD44 protein family belongs to the class of cell adhesion molecules (CAMs) and comprises multifunctional proteins. It has been reported that CD44 interacts with extracellular matrix components, growth factors and cytokines present in the tumour microenvironment and contribute to CSC maintenance in various cancers such as breast, cervical, colorectal and pancreatic (Morath et al., 2016). In normal cells, CD44 marks cells with mesenchymal cellular phenotype (Ramos et al., 2016). On the other hand, loss of CD44 was observed during iPSC reprogramming of

mouse embryonic fibroblasts (MEFs) (in both F-class and ESC-like iPSCs) (Hussein et al., 2014). O'Malley et al. (2013) made use of the dynamic changes of two surface markers, CD44 and ICAM1 in combination with *Nanog* reporter to isolate intermediates (such that progress from $CD44^{+}/ICAM1^{-}$ starting cells, $CD44^{-}/ICAM1^{-}$ intermediates to $CD44^{-}/ICAM1^{+}$ reprogrammed cells) during MEF reprogramming to construct a molecular route map of iPSC generation.

CD24 and CD44 markers are associated with breast stem cell properties (Al-Hajj et al., 2003b). The induction of epithelial to mesenchymal transition (EMT) in non-tumorigenic immortalised cell line HMLE generates $CD44^{high}/CD24^{low/-}$ mesenchymal cells which show mammary epithelial stem cell-like characteristic such as mammoshpere and mammary duct-like structure-forming ability (Mani et al., 2008). These findings are consistent with Prat et al. (2013) who identified a subpopulation of $CD44^{+}/CD24^{-}$ and $CD44^{-}/CD24^{-}$ cells having mammary stem cells/ bipotent progenitor phenotype and thought to top the mammary differentiation hierarchy. Opposing evidence suggested CSCs markers including CD24 and CD44 identifies distinct subpopulations in different breast cancer patient cells and cell lines instead of a common stem cell-like subpopulation, suggesting variability across samples and inconsistency of the markers (Liu et al., 2014c). Many studies also suggested high expression of CD24 in progenitor and metabolically active cells as compared to terminally differentiated cells (Fang et al., 2010, Shakiba et al., 2015). In agreement, in two studies using a mouse model, a subpopulation of $Lin^{-}CD29^{hi}CD24^{+}$ and $CD45^{-}Ter119^{-}CD31^{-}CD140a^{-}CD24^{med}CD49f^{high}$ cells were

shown capable of reconstituting a mammary gland with single cell *in vivo*. Therefore, in this study the use of CD24 and CD44 was used to identify cellular phenotypes (epithelial and mesenchymal) in the context of iPSCs and differentiated progenies. The use of CD24/CD44 as mammary stem cell markers was indicative of re-acquisition of mammary stem cell properties similar to the parental cell lines.

IPSC lines derived from MCF10A, BT-549, MCF7 and MCF10A_PIK3CA$^{(-/+)}$ harboured an expression profile of CD24^{+}/CD44^{-}, suggesting the acquisition of ESC-like profile. Results were consistent with gastrointestinal cancer-derived iPSCs reported by Miyoshi et al. (2010b). When cultured in m2 (EB medium) all iPSCs-derived EBs gave rise to differentiated progenies with highly similar CD24/CD44 expression profile as their parental cells, except for 10AR and 10A_PIK3CA$^{(+/-)}$-derived re-differentiated cells, potentially due to the acquisition of tumorigenicity (discussed later in this chapter). Less similarity was observed in differentiated cell lines cultured in m1 potentially due to incomplete differentiation (caused by low content of serum, which contains a variety of growth factors important for ESCs *in vitro* differentiation) (Schuldiner et al., 2000, Khoo et al., 2005, Nakano et al., 2012)) or skewed differentiation propensity. However, the latter is unlikely when comparing re-differentiated MCF10A cultured in m1 and m2. Cells in m2 had a CD24/CD44 profile more similar to parental MCF10A despite m1 being the medium normally used to culture the parental cell line. Therefore, m1 might not be a suitable medium to skew EB differentiation propensity to the mammary lineage. A different

medium system such as the mammoshpere medium mentioned in Section 6.2.2 (Qu et al., 2017) should be tested in future experiments.

6.3.3 Impact of iPSC reprogramming on breast cancer tumorigenicity

In vitro colony forming assay was performed to assess the impact of iPSC reprogramming on breast cancer tumorigenicity. Consistent with the study by Nishi et al. (2014b), iPSC reprogramming and re-differentiation of MCF10A in EB medium (m2 medium in this study) resulted in increased growth in anchorage-independent conditions. As discussed in Section 3.3, these differentiated cells reported by Nishi et al. (2014) acquired a CSC identity characterised by $CD44^{+}/CD24^{low}$ profile, expression of *ALDH1A1*, *SOX2* (both are known breast CSC markers) and epithelial-to-mesenchymal transition (EMT)-related genes such as *Snail* and *Slug*. In addition, resistance to paclitaxel treatment and susceptibility to CSC-specific drug have also been reported. In this study, 10AR_EB_m2 was found harbouring a $CD44^{+}/CD24^{low}$ profile. However, the possible role of pre-exiting loss of $p16^{INK4a}$ gene in tumorigenicity acquisition requires supports from further evidence such that reported by Nishi et al. (2014b), including a series of $p16^{INK4a}$ rescue experiments. For MCF7, significant loss of colony forming ability was observed in 5PLA_MCF7R and 5PLA_MCF7R_EB_m2 as compared to parental MCF7, although low degree of malignancy was restored in the latter. Likewise, similar results were observed in BT-549 whereby colony forming ability was significantly reduced in both iPSCs and re-differentiated cells. In contrast to 5PLA_MCF7R_EB_m2, no restoration of tumorigenicity was observed in all 5PLA_BTR5-derived differentiated progenies (both from m1 and m2), suggesting high degree of malignancy suppression. Consistent

with some cancer cell reprogramming studies, results suggested epigenetic reconfiguration induced by iPSC reprogramming is capable of suppressing or eliminating tumorigenicity of cancer cells (Zhang et al., 2013). However, the degree of suppression seemed to be subtype dependent in the present study.

In the case of MCF10A_PIK3CA$^{(+/-)}$, results obtained were consistent with the report by Gustin et al. (2009), the study describing the first derivation of MCF10A_PIK3CA$^{(+/-)}$ cells. The sole introduction of *PIK3CA* mutation did not result in increased colony forming ability as compared to WT MCF10A cell line, despite exhibiting epidermal growth factor (EGF)-independent growth. This is in contrast to other studies listed in Table 6.7 (Isakoff et al., 2005, Zhao et al., 2005, Zhang et al., 2008a, Van Keymeulen et al., 2015).

Table 6.7 Colony forming assay performed on different *PIK3CA* mutation studies

References	Number of colonies	Cells seeded
The present study	7.7±2.1	2.5 x 10^4
Isakoff et al. (2005)*	100	5 × 10^4
Zhao et al. (2005)*	800 (HMEC)	5 × 10^4
Zhang et al. (2008a)*	45	5× 10^3
Gustin et al. (2009)	Did not form colonies in soft agar or tumours in nude mouse xenograft assays	3 x 10^4
Van Keymeulen et al. (2015)*	40 in mutant versus 20 in wild type	1000 cells
Croessmann et al. (2017)	Mutant derived from Gustin et al. (2009). No significant increase in colonies number as compared to wild type	2 x 10^4

*Difference was statistically significant

Besides, in contrast to MCF10A-derived iPSCs (10AR), significant increase of growth in soft agar was observed in reprogrammed MCF10A_PIK3CA$^{(+/-)}$ (the 10A_PIK3CA$^{(+/-)}$ R4), suggesting a possible contribution of the *PIK3CA* mutation individually or in concert with loss of

p16^{INK4a} gene to tumorigenicity in the pluripotent state. Acquisition of tumorigenicity in the iPSC state could help explaining the low quality of pluripotency of 10A_PIK3CA$^{(+/-)}$ R4 (as indicated by pluripotency marker expression and morphology shown in Section 5.2.4). As discussed in Section 1.4, iPSC reprogramming and malignant transformation share numerous similarities (Ohnishi et al., 2014). Therefore, MCF10A_PIK3CA$^{(+/-)}$ cells may have taken an alternative "transforming route" instead of one that leads to pluripotency upon induction of OSKM overexpression, resulting in tumorigenicity acquisition. The potential effect of *PIK3CA* mutation in directing the selection of transformation route represents an interesting topic of research.

Another interesting observation in this experiment was the varying degrees of tumorigenicity observed between cells cultured in different media at both pluripotent and differentiated states. E8 is a stem cell medium while m1 and m2 are culture media specific for MCF10A (mammary epithelial cells) and induction of spontaneous *in vitro* differentiation. Two theories are proposed to explain this tumorigenicity discrepancy. First, lineage-specificity of *PIK3CA* mutation. *PIK3CA* mutation has been implicated in a wide range of cancer types, including breast, colon, ovarian, liver, pancreatic as well as germ cell (Karakas et al., 2006, Zehir et al., 2017b). Despite the availability of prevalence estimates, no association between *PIK3CA* mutation and the degree of tumorigenicity has been reported (Zehir et al., 2017b). An iPSC pluripotent state, closely resembling germ cell tumour, could potentially enhance the oncogenic effect of *PIK3CA* mutation resulting in high colony forming ability. On the other hand, spontaneous

differentiation induced by m2 generated a variety of cells of three germ layers. Each cell type might manifest varying degrees of aggressiveness as a result of *PIK3CA* mutation contributing to the overall high colony forming ability as compared to cells cultured in m1 (which is thought to generate a relatively homogenous group of cells representing the mammary lineage). Second, the medium composition could too play a role. E8 medium contains various growth factors that promote cell proliferation and stem cell maintenance (Chen et al., 2011b), some of which have been associated with an aggressive cancer phenotype. For example, NODAL has been shown to enhance cancer aggressiveness by promoting epithelial–mesenchymal transition (EMT) (Shen, 2007, Fang et al., 2013).

6.3.4 Chemotherapy sensitivity assay

Paclitaxel (or Taxol) is an antimitotic drug which binds and stabilises tubulin polymerisation into microtubules, interfering mitosis and causing cytotoxicity in proliferative cells. It is widely used to treat breast, ovarian and non-small cell lung cancers (Bollag et al., 1995, Dumontet and Jordan, 2010). However, the development of drug resistance remains one major problems of paclitaxel usage in addition to toxicity. In this study, paclitaxel was used to study the impact of iPSC reprogramming to drug sensitivity or resistance in cancer cells and their re-differentiated derivatives. No significant changes were observed in all cell lines except for MCF7, whereby the IC_{50} was significantly increased in re-differentiated cells as compared to the parental cells. As mentioned earlier, colony forming assay indicated loss of tumorigenicity in re-differentiated MCF7 cells. Therefore, the increased IC_{50} could be a result of malignant properties lose, in particular reduced

proliferation rate, which rendered re-differentiated MCF7 cells less susceptible to paclitaxel. However, although slower growth in culture was observed in re-differentiated cell line, no firm conclusion can be drawn at this stage and proliferation assays should be performed as future work to verify this speculation.

Despite the reduced tumorigenicity and large IC_{50} difference between parental BT-549 and re-differentiated cells, the result was not statistically significant. This could be due to the natural low susceptibility of the parental BT-549 to paclitaxel. Indeed, Hiser et al. (2006) reported approximately 5-times higher IC_{50} (less susceptible) in BT-549 as compared to MCF7. Consistently, a 3-fold difference was observed in the present study. Mechanisms underlying varying susceptibilities to paclitaxel between cancer types and subtypes remain elusive. Studies suggested it might be due to the potential effects of expression level of different tubulin isotypes (Hiser et al., 2006), tau protein (Rouzier et al., 2005) and mesenchymal (BT-549) or epithelial (MCF7) cellular phenotype (Byers et al., 2013, Mezencev et al., 2016). Nevertheless, results in the present study suggested that paclitaxel might not be a suitable drug for testing on BT-549. Drugs with proven efficacy on triple negative breast cancer such as Poly (ADP ribose)-polymerase (PARP) inhibitors (Schoonen et al., 2017) should be used in future experiment to address the association of reprogramming with drug sensitivity in BT-549.

Overall, results in this chapter indicated the importance of epigenetic and genetic alterations in breast cancer. Global epigenetic reconfiguration induced by iPSC reprogramming reverted tumorigenicity of breast cancer cells. The extent of reversion was subtype dependent. However, genome-wide studies such as global DNA methylation and expression profiling are required to elucidate the mechanisms underlying tumour reversion. Likewise, the same approaches should be employed to dissect the contribution of *PIK3CA* to breast cancer initiation and progression using models created in this chapter. Lastly, the developmental state or identity of cells cultured in different media which exhibited varying degree of tumorigenic potential should be assessed to identify possible lineage-specificity of *PIK3CA* mutation, which in turn could provide clues into the cellular origin of breast cancer.

7. GENERAL DISCUSSION

The aim of this study was to identify the potential of iPSC reprogramming to suppress or eliminate tumorigenicity by resetting the global epigenetic configuration of breast cancer cells. To achieve this, the episomal vector approach was used to derive iPSCs from breast cancer cell lines representing basal-like and ER^+ subtypes. In this chapter, the results obtained from previous experimental chapters are discussed in the context of current literature together with their strengths, possible applications, limitations and future directions.

Results from Chapter 3 suggested cells of the mammary lineage (MCF10A cell line) were amendable to episomal vector (OSKML + shRNA against *TP53* in four plasmids) iPSC reprogramming in addition to the classical retroviral approach (OSKM) reported by (Nishi et al., 2014b). In Chapter 4, the optimised experimental condition was applied to reprogram three breast cancer cell lines, namely MCF7 (ER^+), BT-549 (basal-like) and MDA-MB-231 (basal-like), but without much success, suggesting possible refractoriness in breast cancer cells to iPSC reprogramming (Papapetrou, 2016). Successful derivation of cancer-iPSCs was only achieved after the incorporation of an additional vector encoding the *microRNA-302/367* (*miR-302/367*) cluster (Anokye-Danso et al., 2011) together with sodium butyrate (NaB) treatment (Zhang and Wu, 2013). This could be due to enhanced overall reprogramming efficiency or the pursuit of an alternative route to pluripotency (Tonge et al., 2014, Jackson et al., 2016). Following the successful reprogramming of MCF10A and the establishment of the more

efficient five plasmids plus NaB reprogramming cocktail, two MCF10A models harbouring the two most frequently mutated genes in breast cancer (*PIK3CA* and *TP53* driver mutations) (Nik-Zainal et al., 2016) were developed through CRISPR-Cas9 gene editing and the mutants were reprogrammed with the five plasmids plus NaB reprogramming method.

In addition, pluripotent stem cells (PSCs) characterisation protocols, namely RT-PCR, RT-qPCR analyses and immunofluorescence staining of pluripotency marker expression, were also optimised in Chapter 3 and subsequently applied to characterise iPSCs generated in Chapter 4 and 5. These protocols were complemented by *in vitro* differentiation experiments in Chapter 6, including embryoid body formation and mammary organoid generation to assess the state of pluripotency in all iPSC lines isolated. In Chapter 6, the overall impact of iPSC reprogramming on two phenotypic hallmarks of breast cancer, tumorigenicity and drug sensitivity, were evaluated by soft agar colony formation assay and paclitaxel sensitivity test (Hanahan and Weinberg, 2011a). Results indicated loss or suppressed tumorigenicity in BT-549 and MCF7 breast cancer cell lines. On the contrary, colony forming ability were induced and enhanced in WT MCF10A and MCF10A_PIK3CA$^{(+/-)}$ isogenic line. Furthermore, reduced paclitaxel susceptibility had also been observed in MCF7 after reprogramming. However, more evidence is required to determine if reduced susceptibility was due to the acquisition of drug resistance (an undesirable outcome for cancer treatment) or decreased proliferation rate (recapitulating that of normal non-tumorigenic cells and thus not selectively targeted by paclitaxel). Nevertheless, all evidences collectively implicated the vital role

of epigenetics in breast cancer initiation and progression and in line with the hypothesis of this study, which stated tumour reversion can potentially be achieved via reprogramming of the breast cancer epigenome (Allegrucci et al., 2011, Powers and Pollack, 2016).

7.1 TUMOUR REVERSION THROUGH IPSC REPROGRAMMING

Cancer cell iPSC reprogramming provides an unprecedented platform to investigate the contribution of epigenetics to cancer independent of genetic alterations. As discussed in Section 1.3.2, many studies have demonstrated the suppression or elimination of malignancies through iPSC reprogramming of cancer cells (Stricker et al., 2013). Some of these studies have provided important insights into the mechanisms underlying reprogramming-induced tumour suppression. Through global DNA promoter methylation and gene expression analyses, Zhang et al. (2013) demonstrated epigenetic reconfiguration predominantly at oncogenes and tumour suppressors were responsible for malignancy suppression in sarcoma-derived iPSCs and their re-differentiated progenies. In addition, Stricker et al. (2013) demonstrated widespread changes in a panel of 691 differentially methylated genes between normal and cancer-derived neural stem cells in glioblastoma (GBM)-derived iPSCs, including demethylation of *TES* and *CDKN1C* promoter regions, both have been previously associated with GBM. However, re-acquisition of malignancy in GBM-iPSC-derived neural lineage but not cells of other lineages suggested developmental lineage-specific epigenome maybe crucial for tumour suppression in addition to reprogramming-induced DNA methylation changes. Furthermore, Miyoshi et al. (2010a) suggested

DNA demethylation and reactivation of tumour suppressor gene *p16^{INK4A}* in re-differentiated gastrointestinal cancer-iPSCs contributed to reduced oncogenic phenotypes. Through genome-wide methylation analysis, Mahalingam et al. (2012) also reported the reversal of aberrantly methylated promoters in non-small cell lung cancer (NSCLC) in NSCLC-derived iPSCs. Changes of oncogene and tumour suppressor methylation status have also been observed. Together with findings in this study, the above evidences suggested suppression of tumorigenicity and potentially tumour reversion can be achieved by resetting the cancer epigenome. The combination of iPSC reprogramming and genome-wide analyses could potentially allow the elucidation of underlying mechanisms of the reversion process.

7.2 CANCER CELL REPROGAMMING AND RE-DIFFERENTIATION – *IN VITRO* CANCER MODELLING

Self-renewal and differentiation potential are two hallmarks of pluripotency (De Los Angeles et al., 2015). The reprogramming of patient-derived cancer cells allows derivation of cancer iPSCs harbouring all pre-existing genetic mutations. The availability of well-established differentiation protocols permits the induction of PSC differentiation into virtually all cell types making up the human body (Keller, 2005). Therefore, cancer cell reprogramming followed by re-differentiation has provided an alternative approach to generate cancer models closely recapitulate cancer initiation and progression *in vivo* (Gillet et al., 2013, Papapetrou, 2016). Furthermore, self-renewal of cancer iPSCs allows the generation of, theoretically, unlimited supply of samples for various research studies. Ye et al. (2009)

reported the successful derivation of iPSCs from patients with myeloproliferative disorders. Re-differentiation of iPSCs generated hematopoietic progenitor cells harbouring gene expression characteristic of parental cells and exhibited disease phenotypes. In addition, Hu et al. (2011) demonstrated the recapitulation of chronic myeloid leukaemia (CML) signature, the BCR-ABL fusion gene and a 4-way translocation between chromosomes 1, 9, 22, and 11 characteristic of patient cells in CML-derived iPSCs. Through re-differentiation of invasive pancreatic ductal adenocarcinoma (PDAC)-derived iPSCs, Kim et al. (2013) isolated pancreatic ductal tissues from teratoma which recapitulated pancreatic intraepithelial neoplasia, a precursor state of PDAC. This allowed the establishment of early stage cancer model for the identification of early stage biomarkers and investigation of cancer progression. Lastly, through generation of iPSC lines from patient samples representing different stages of leukaemia, such that from preleukaemia, low risk myelodysplastic syndrome (MDS), high risk MDS to MDS/acute myeloid leukaemia (AML), Kotini et al. (2017) constructed a phenotypic roadmap of the disease (progressive loss of hematopoietic progenitor differentiation potential, a phenotypic hallmark of the disease), and identified genetic factors responsible for disease progression as well as evaluated responses of candidate drugs at a stage-specific manner. In the present study, successful derivation of breast cancer iPSCs from two different breast cancer subtypes and from cells carrying defined breast cancer driver mutations have laid the groundwork for primary patient-derived breast cancer cells reprogramming. This approach would allow the investigation of mechanisms underlying breast carcinogenesis and the evaluation of drug responses in a

patient/subtype-specific manner. Schematic diagram in Figure 8.1 illustrates the potential application of the iPSC technology in cancer research.

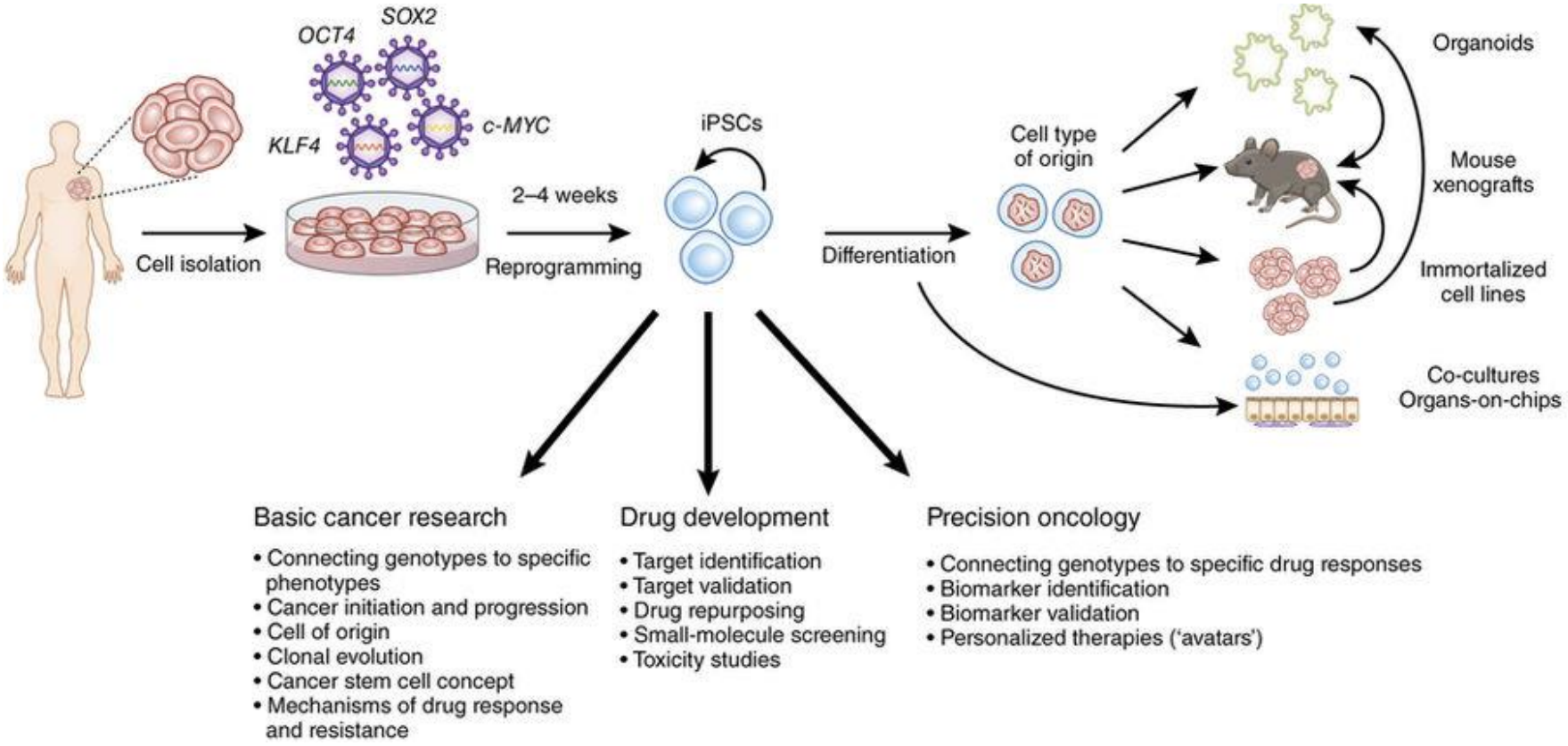


Figure 7.1 The potential applications of iPSC technology in cancer research.

The derivation of iPSCs from patient sample allow the unlimited supply of cell source, which can be integrated into other model systems to facilitate cancer research. Cancer iPSCs could also be used to address problems pertaining cancer research striving for better understanding of cancer, development of effective treatment as well as precision oncology. Figure derived from Papapetrou, E. P. (2016) (Licence number: 4292990131089).

7.3 LIMITATIONS OF THE STUDY

Despite its high reproducibility, iPSC generation remains greatly inefficient, having typically <1% of transfected cells successfully progress into *bona fide* iPSCs. The ease and efficiency of reprogramming are significantly dependent on the somatic origin as well as the reprogramming methods (Yamanaka, 2012b). The initial reprogramming procedure relied on the induction of OSKM overexpression mediated by retrovirus or lentivirus integration into the host genome. This approach possesses high risk of insertional mutagenesis affecting host genome and epigenome, which in turn hampers translational applications of iPSCs and could potentially lead to tumorigenicity (Howe et al., 2008). In addition, incomplete silencing and transgene reactivation have also been reported in iPSCs and their differentiated derivatives (Okita et al., 2007b). This incidence represents a major caveat as the components of the original Yamanaka OSKM cocktails, most notably *c-MYC*, have been associated with cancer (Hayden, 2011, Yamanaka, 2012b, Ma et al., 2013, Semi et al., 2013, Rouhani et al., 2016). The *c-MYC* is a widely studied proto-oncogene involved in cell cycle, proliferation and metabolic pathways. The overexpression of *c-MYC* upon mutation promotes genetic instability, angiogenesis, cell survival, proliferation, differentiation blockade and has been associated with poor outcomes in a wide range of human cancers (Vennstrom et al., 1982, Chang et al., 2008, Dang, 2012, Farrell and Sears, 2014, Hsieh et al., 2015, Xue et al., 2015). Indeed, the replacement of *c-Myc* with another *Myc* family member, *L-Myc* or omitting *c-Myc* from the OSKM cocktail has resulted in less transformation activity (Aoi et al. 2008; Nakagawa et al., 2008).

In addition, *SOX2* has been shown as a lineage-survival oncogene (whose malignant effect will only be manifested in cell lineage that rely on its expression for normal development) in lung and oesophageal squamous cell carcinomas (Bass et al., 2009). *SOX2* was also responsible for the CSC properties in ovarian and colorectal cancers and skin squamous-cell carcinoma (Boumahdi et al., 2014, Chebouti et al., 2016, Lundberg et al., 2016). On the other hand, *KLF4* acts as a tumour suppressor gene and oncogene in a context-dependent manner (Rowland and Peeper, 2006). For instance, the overexpression of *KLF4* was responsible for pancreatic ductal adenocarcinoma initiation (Maddipati and Katz, 2016, Wei et al., 2016) but suppressed breast cancer cells (SK-BR-3 cell line) tumorigenicity *in vivo* (Wang et al., 2015). Last but not least, *OCT3/4* has been implicated in lung cancer (Chen et al., 2008, Chiou et al., 2010), glioblastoma (Ikushima et al., 2011, Seymour et al., 2015), oral (Baillie et al., 2017), head and neck (Koo et al., 2015a) as well as breast cancers (Liu et al., 2014a). It was too reported as a marker for germ cells tumours and its overexpression promoted the malignant potential of ESC-derived tumours (Gidekel et al., 2003, Jones et al., 2004, Rijlaarsdam et al., 2011). To circumvents the abovementioned limitations, a wide range of non-integrative and transgene-free reprogramming methods have been developed since the discovery of iPSCs, striving for safer and more accessible reprogramming approaches for iPSC generation (Theunissen and Jaenisch, 2014b). In the present study, all iPSC lines have been shown devoid of exogenous gene expression. Nevertheless, together with the downregulation of *TP53*, it should be taken into consideration that the reprograming cocktail used in this study could

represent a confounding factor for tumour initiation observed in MCF10A and *PIK3CA* mutation isogenic line.

Another limitation of the study was the use of cancer cell lines. Cultured cancer cell lines are the most frequently used *in vitro* models in cancer research (Gillet et al., 2013). However, how closely the cancer cell lines reflect cancer cells within the *in vivo* environment remains highly debatable. For example, long term *in vitro* passage has been shown to promote expansion and selection of fast growing population reducing heterogeneity observed in primary tumour. This incidence was allegedly observed in MCF7, whereby a small $CD24^{-}/CD44^{+}$ (cancer stem cell phenotype) population was isolated distinct from the major population which harboured a $CD24^{+/}CD44^{-}$ profile. In this study, all reprogramming experiments were performed on established cell lines and only one iPSC line per each cancer cell line was fully characterised due to time constraints. The same reprogramming conditions optimised in Chapter 3 and Chapter 4 should be applied to reprogram primary samples derived from cancer patients or patient derived xenografts (PDX) to validate findings obtained by using cancer cell lines. To date, only reprogramming of primary samples from various blood cancer cells and pancreatic adenocarcinoma have been reported (Papapetrou, 2016). Lastly, the current study focused only on breast cancer of two different intrinsic subtypes, the basal-like and ER^{+} subtypes, and only two and one cell line was used for both subtypes respectively. The number of cell lines should be increased for each subtype or breast cancer cells of other subtypes should be included to provide

outcomes that are more credible in a subtype-specific manner or a general view on breast cancer.

7.4 FUTURE WORK

Future work should include different aspects of this project, including the reprogramming process, the characterisation of the cell lines, the use of complementary *in vitro* and *in vivo* functional tests and the study of the gene networks involved in the reprogramming process.

As illustrated in Table 3.1, different strategies were utilised to evaluate efficiency of episomal vector-mediated iPSC reprogramming. These include the generation of NANOG or OCT3/4 green fluorescence protein (GFP) reporter cell lines and FACS analysis of pluripotency markers (such as SSEA4, TRA-1-81 or NANOG) at different time points of reprogramming. In this study, the reprogramming efficiency was not quantitatively determined. It has been reported that certain cancer cells were more refractory to iPSC reprogramming in comparison to non-tumorigenic counterparts. Consistent results were also obtained in the present study whereby breast cancer cell lines MCF7, BT-549 and MDA-MB-231 seemed to be less amendable to iPSC reprogramming than the non-tumorigenic mammary epithelial cell line MCF10A. However, quantitative analyses should be employed and appropriate experimental replicates should be performed before statistically relevant conclusions are drawn.

Subsequently, further analyses should be carried out to identify mechanisms underlying refractoriness (eg. the presence of KRAS mutation as suggested by Kim et al. (2013)), which could help provide understandings into iPSC reprogramming and potentially its crosstalk with breast tumorigenesis. As shown in Chapter 4, successful reprogramming of breast cancer cell lines required the addition of *miR-302/367* cluster expression and NaB treatment, as both have been demonstrated to increase reprogramming efficiency (Zhang et al., 2015) and have been used individually to efficiently reprogramming different cell types (Anokye-Danso et al., 2011). In an effort to dissect the molecular mechanisms underlying iPSC reprogramming, (Buganim et al., 2012) suggested that the pluripotency circuitry can be turned on through multiple entry points and the combination of different factors (*Lin28, Sall4, Esrrb* and *Dppa2*) has been demonstrated with high reprogramming efficiency in the same study. Therefore, it will be interesting to investigate the potential effects of the additional factors by performing reprogramming experiments using different combinations of the OSKML + shRNA_TP53 factors, miR302/367 cluster and NaB treatment. Comparative studies should then be performed to identify the differentially activated genes/pathways which could be useful for efficient breast cancer cell reprogramming or iPSC reprogramming in general.

Additional work should be done to fully characterise the existing iPSC lines and to derive iPSCs from patient primary tumour cells. During the development of the project, several iPSC clonal cell lines were derived for each cell type, but they were not fully characterised due to time constraints.

The generation of iPSC lines and the characterisation involving differentiation experiments are lengthy processes that take several months of work. Therefore, only the best iPSC clones (based on growth characteristics and morphology) were selected for further study. In the future, more clonal lines (at least two) should be characterised to account for any variation in iPSC reprogramming to confirm the general validity of the results. In addition, the MCF10A cell line was selected as "normal control" given its non-tumorigenic characteristics. Although it is widely accepted by the scientist community that this cell line closely represents normal mammary epithelial cells, the results obtained in this project demonstrate that the process on immortalisation may have an impact on their tumour initiation. Therefore, reprogramming of normal primary cells of the breast, such as HMEC, should be attempted in the future. As previously mentioned, the use of primary tumour cells would also be essential to fully evaluate the application of the iPSC technology in the breast cancer setting and to advance translational applications.

The assessment of tumorigenicity represented an important end point of the current study. In the future, more *in vitro* functional tests such as proliferation, oncosphere and invasion assays should be performed. In addition, *in vivo* assay such as xenotransplantation into immuno-compromised nude mice should also be performed to confirm the malignancy status of all cell lines and their differentiated progenies. As discussed in Chapter 6, sensitivity of different cell lines towards different chemotherapy drugs, in particular drugs specific for triple negative breast

cancer (such as PARP inhibitors), should be tested to evaluate cancer cell susceptibility or resistance after reprogramming.

Finally, the gene networks involved in the cancer reprogramming should be studied in the future. Genome-wide studies such as global DNA methylation and gene expression profiling could provide highly informative data to answers important questions pertaining this study. First, the level of pluripotency of all iPSCs generated. Using analyses such as principle component analysis (PCA) of transcriptomic data the closeness or remoteness of iPSC lines to ESCs and the parental cell lines can be identified, which could offer explanation of different levels of pluripotency marker expression and differentiation potential observed. Second, genomic and epigenomic studies could help identify differentially expressed genes or methylated regions responsible for pluripotency acquisition, which could help identify possible pathways or regulation mechanisms responsible for efficient reprogramming and refractoriness in cancer cells. Subsequently, pathways inhibitors or agonists can be introduced to overcome reprogramming barriers and facilitate the reprogramming breast cancer cells. Most importantly, as demonstrated in different studies discussed earlier in this chapter, genome-wide analyses will be crucial to provide important insights into molecular mechanisms underlying tumour reversion (in the cases of MCF7 and BT-549) and initiation (MCF10A and MCF10A_PIK3CA$^{(+/-)}$) induced by iPSC reprogramming. Last but not least, the transcriptome and epigenome of reprogrammed cancer cells, differentiated derivatives and parental cells should be studied to identify novel biomarkers for breast cancer detection and the discovery of novel druggable targets.

7.5 CONCLUSION

In conclusion, this study has successfully optimised iPSC reprogramming conditions involving the transfection of five episomal plasmids encoding OSKML, *EBNA1*, shRNA against *TP53* and miR-302/367 cluster together with NaB treatment for breast cancer cells. IPSC lines were successfully generated from one basal-like (BT-549) and ER^- subtype (MCF7) breast cancer cell line as well as two isogenic MCF10A cell lines harbouring *PIK3CA* mutation and both *PIK3CA* and *TP53* mutations (the two most commonly mutated genes in breast cancer). These iPSC lines (iPSCs from *PIK3CA* and *TP53* double mutant was not fully evaluated) showed different levels of pluripotency markers expression and differentiation potential, which could be due to intrinsic factors, such as the presence of *PIK3CA* mutation in isogenic line that could had potentially directed cells toward a "transforming route". The reprogramming process suppressed *in vitro* tumorigenicity in both breast cancer cell lines but had opposite effect on the MCF10A_PIK3CA$^{(+/-)}$ isogenic and the control WT MCF10A cell lines potentially due to the presence of pre-exiting loss of $p16^{INK4a}$ as demonstrated by previous study (Nishi et al., 2014b). In both cases, results suggested that epigenetics plays a vital role in breast carcinogenesis and that cancer iPSC reprogramming could be used to model tumour progression. Further analyses into the molecular mechanisms underlying the reprogramming process could provide important insights into tumour initiation and reversion, which in turn could facilitate the identification of biomarkers as well as the discovery of novel strategies or targets for breast cancer treatment.

APPENDIX

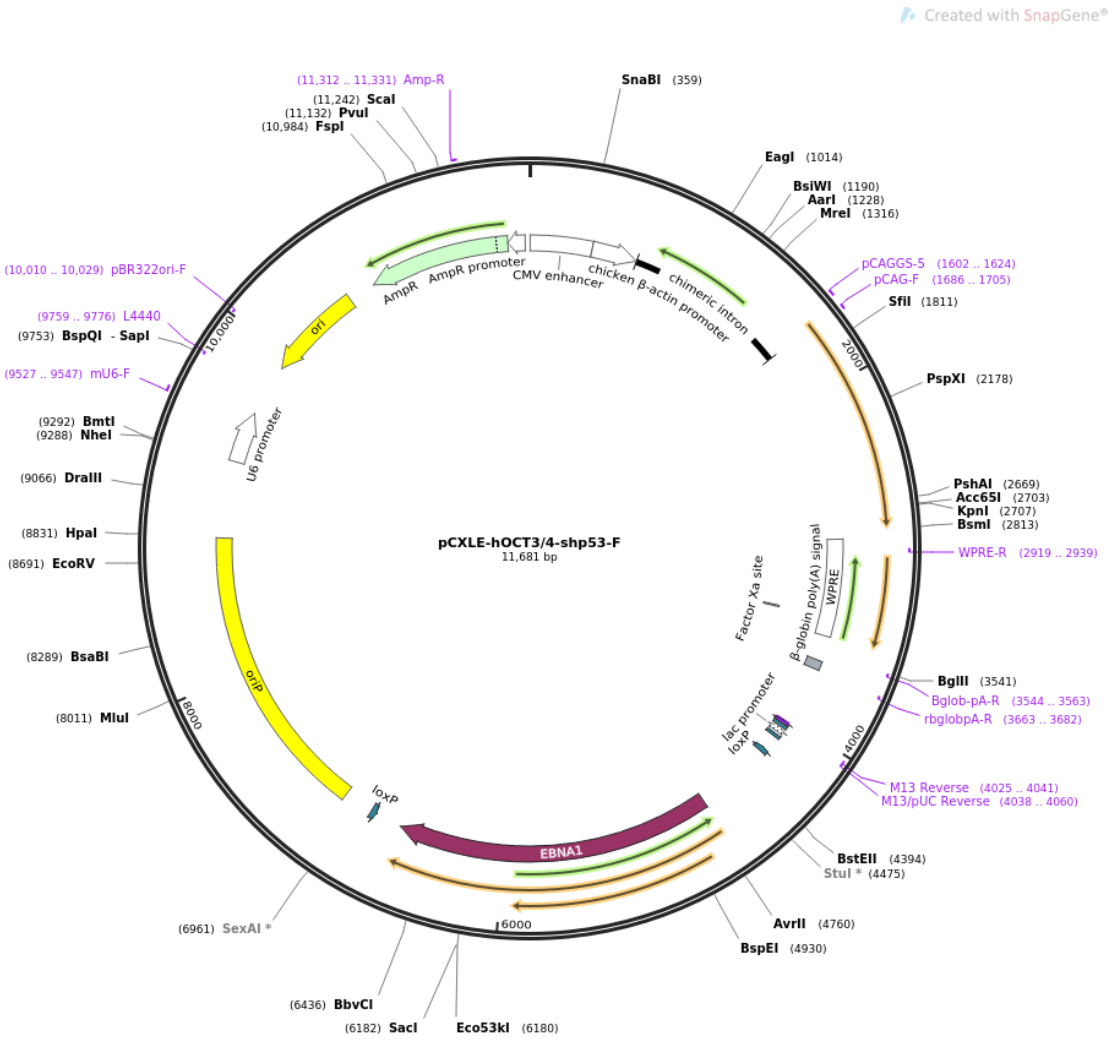


Figure S2-1 Plasmid map of Addgene plasmid #27077.

Plasmid #27077 encodes OCT3/4 and short-hairpin RNA against *TP53.*

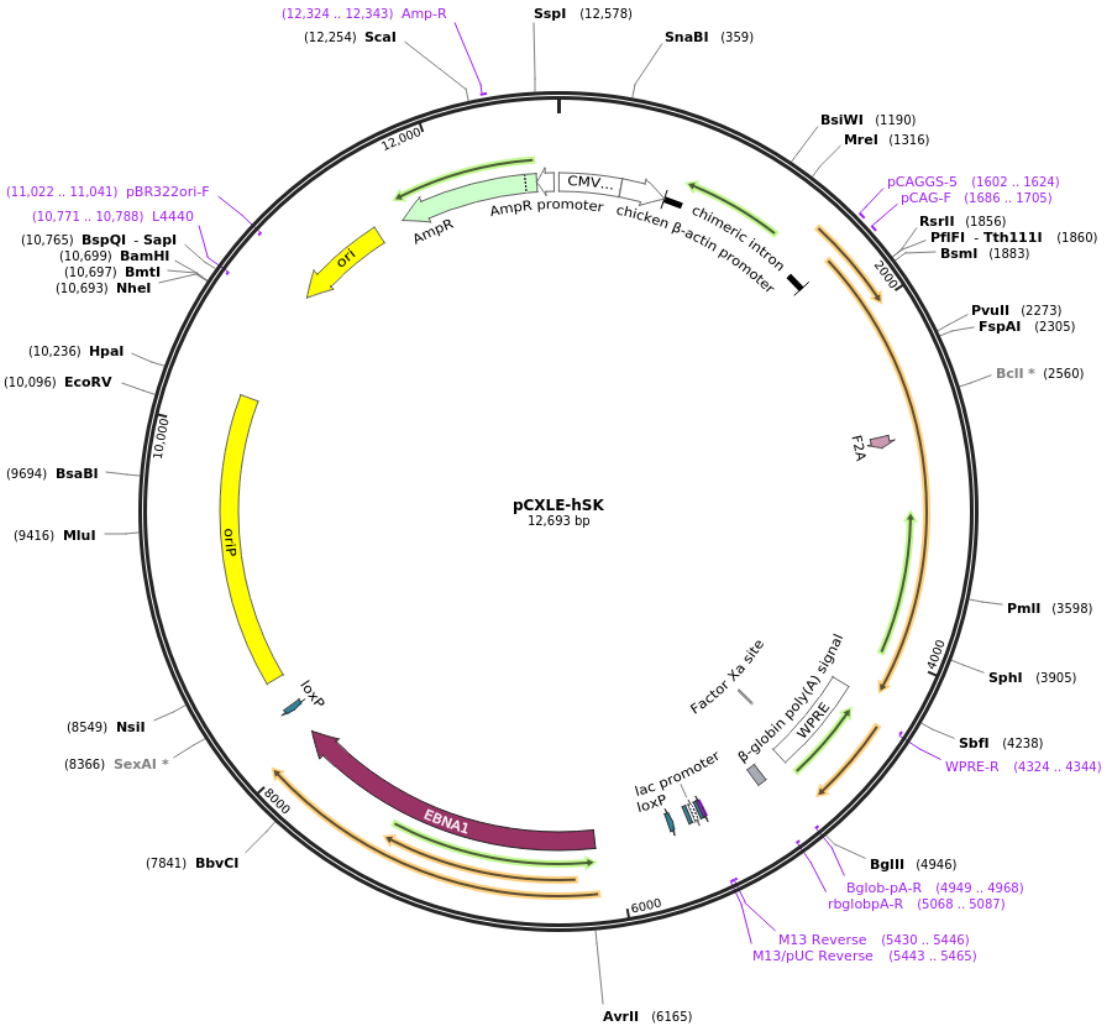


Figure S2-2 Plasmid map of Addgene plasmid #27078.

Plasmid #27078 encodes human *SOX2* and *KLF4.*

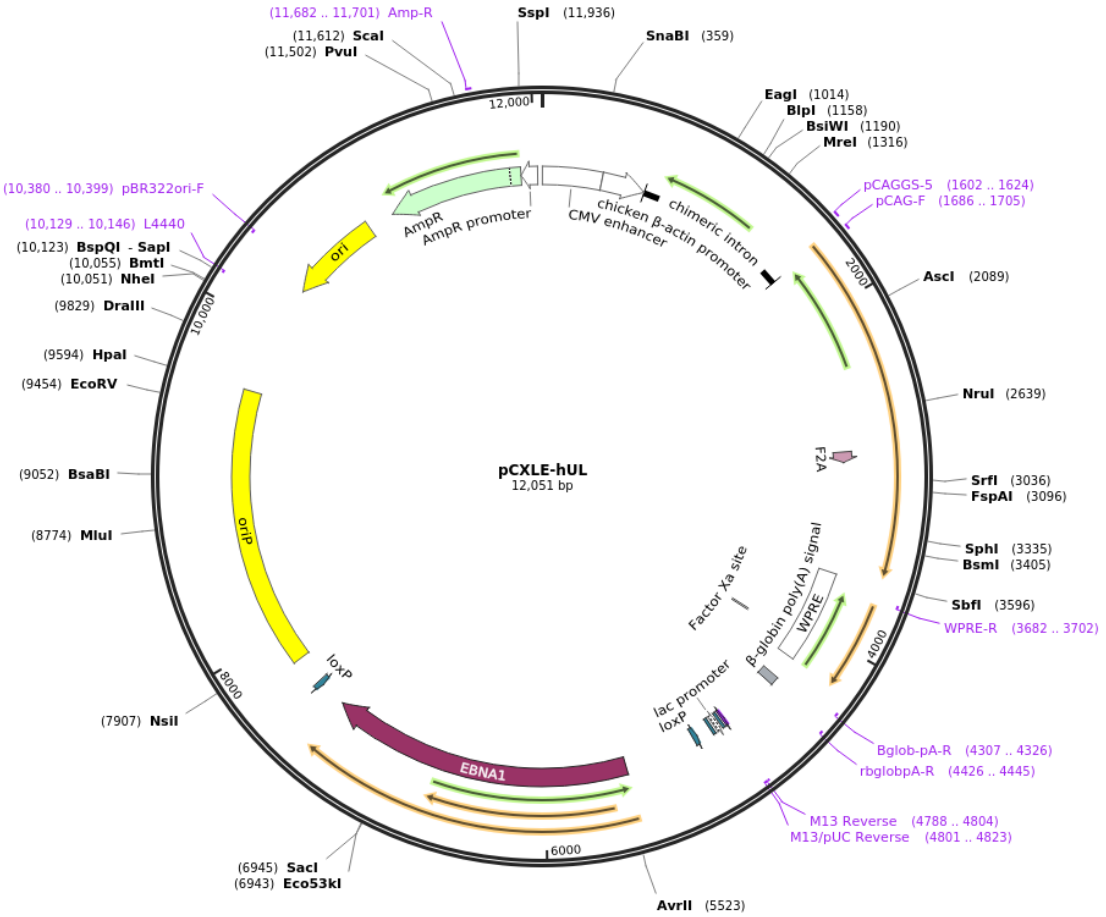


Figure S2-3 Plasmid map of Addgene plasmid #27080.

Plasmid #27080 encodes human *L-MYC* and *LIN28.*

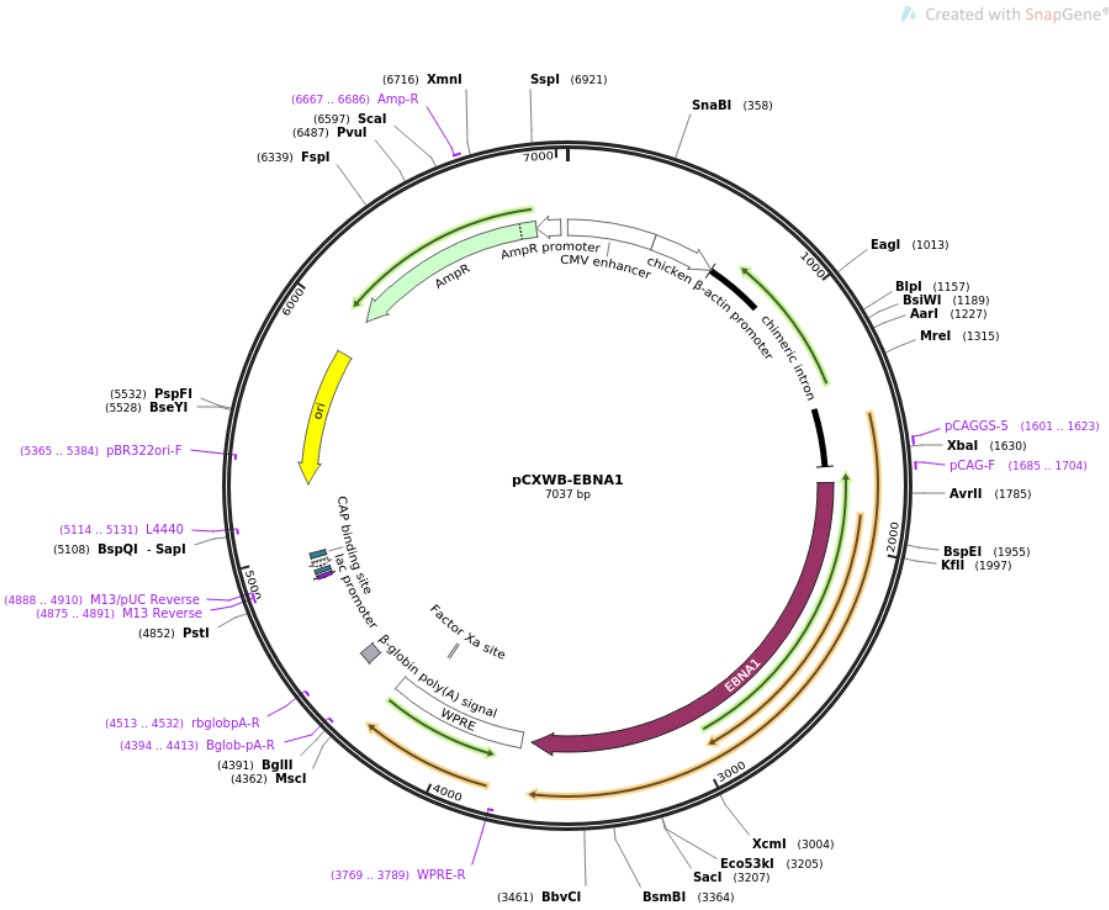


Figure S2-4 Plasmid map of Addgene plasmid #37624.

Plasmid #37624 encodes *EBNA1.*

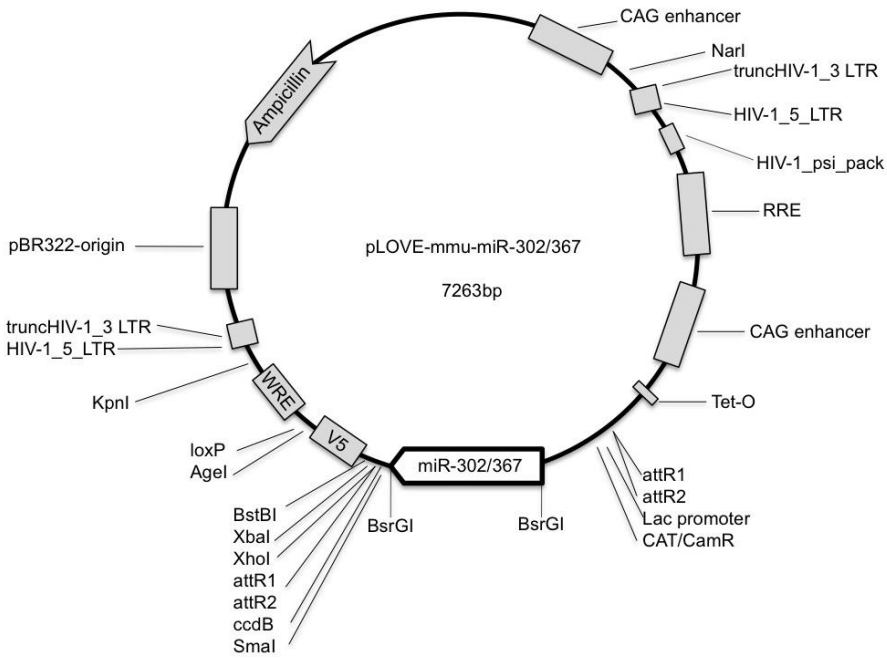


Figure S2-5 Plasmid map of Addgene plasmid #31310.

Plasmid #31310 encodes *miR302/367.*

Table S2-1 Primers used for RT-PCR.

Genes		5′ – 3′	cDNA	gDNA	Ta
Housekeeping genes					
RPLP0	F	ATAGACAGGAGCGCTATCCGC	129	320	60
	R	GACACCCATCCCGCGGTC			
ACTB	F	CCTCGCCTTTGCCGATCC	70	930	58
	R	CGCGGCGATATCATCATCC			
*ACTB*_143	F	ATAGCAACGTACATGGCTGG	143	-	60
	R	CACCTCTACAATGAGCTGC			
Endogenous pluripotency markers					
REX1	F	GAAGAGGCCTTCACTCTAGTAGTG	179	3206	56
	R	TTTCTGGTGTCTTGTCTTTGCCCG			
All *NANOG* genes (NANOG_all)	F	TGCTGAGATGCCTCACACGGA	155	3339	
	R	TGACCGGGACCTTGTCTTCCTT			
NANOG (NF2/R2)	F	TGAGACGGAGTCTTGCTCTG	335	-	55
	R	GGCTTCTATCAATGTTGTCCTTA			
NANOG/ NANOGP8 (NF2/R3)	F	TGAGACGGAGTCTTGCTCTG	467/ 446	-	56
	R	CCTCCCAATCCCAAACAATA			
NANOG (N444)	F	ATTATAAATCTAGAGACTCC	444	-	46
	R	TTGTTTGCCTTTGGGACTGGT			
DPPA4	F	TGCCTCCAATTTTCCACCCC	216	1081	58
	R	CATGGCCCATAAACAGGTGC			
OCT3/4	F1	AGCCCTCATTTCACCAGGCC	456	-	58
	F2	GATGGCGTACTGTGGGCCC	195	-	58
	R	TGGGACTCCTCCGGGTTTTG			
DNMT3B	F	TGCTGCTCACAGGGCCCGATACTT C	242	3910	58
	R	TCCTTTCGAGCTCAGTGCACCACAAAAC			
Episomal vectors					
OCT3/4-shp53	F	GCGCCGGCAGGAAGGAAATG	124	#27077	
	R	TCGGCCTGTGTATATCCCAGGGTG			
KLF4	F	CCACCTCGCCTTACACATGAAGA	156	#27078	
	R	TAGCGTAAAAGGAGCAACATAG			
SOX2	F	TTCACATGTCCCAGCACTACCAGA	111	#27078	
	R	TTTGTTTGACAGGAGCGACAAT			
L-MYC	F	GGCTGAGAAGAGGATGGCTAC	122	#27080	
	R	TTTGTTTGACAGGAGCGACAAT			
LIN28	F	AGCCATATGGTAGCCTCATGTCCGC	251	#27080	
	R	TAGCGTAAAAGGAGCAACATAG			
CRISPR-Cas9					
RNaseP	F	AAGGTCTGAGACTAGGGCCAGA	109	109	55
	R	TCTCCTGCCCAGTCTGACCT			
PIK3CA	F	CAGGCATTGTTGTAGGTGCT	-	597	55
	R	CTATGCAATCGGTCTTTGC			
TP53	F	ATGATTTGATGCTGTCCCCGG	-	418	68
	R	GGGACAGGAGTCAGAGATCAC			Q5
Three Germ Layers markers					
Mesoderm					
DCN	F	AGAGTACCTGGTGGGCTGG	97	-	60
	R	GTGGGCAGAAGTCACTTGAT			
GATA2	F	CTGTCTGCAACGCCTGTG	89	589	60
	R	GTTCCGAGTCTGGATCCCTT			
BMP4	F	GCACTGGTCTTGAGTATCCTG	135	1099	60
	R	TGCTGAGGTTAAAGAGGAAACG			
IGF2	F	AGACGTACTGTGCTACCCC	121	414	60
	R	TGCTTCCAGGTGTCATATTGG			
aSMA	F	AAGTACCCGATAGAACATGGCA	196	-	60
	R	ACATGGCTGGGACATTGAAA			

Endoderm					
GATA1	F	TGCGGCCTCTATCACAAGATG	82	-	60
	R	CTGCCCGTTTACTGACAATCA			
SOX7	F	TCGACGCCCTGGATCAACT	143	143	60
	R	CTGGGAGACCGGAACATGC			
CDH1	F	CGAGAGCTACACGTTCACGG	119	-	60
	R	GGGTGTCGAGGGAAAAATAGG			
AFP	F	CTTTGGGCTGCTCGCTATGA	176	-	60
	R	ATGGCTTGGAAAGTTCGGGTC			
GATA6	F	TCTACAGCAAGATGAACGGCCT	207	-	65
	R	TTTTCATAGCAAGTGGTCTGGG			
Ectoderm					
ALDH1A1	F	CTGCTGGCGACAATGGAGT	89	-	60
	R	CGCAATGTTTTGATGCAGCCT			
COL1A1	F	GAGGGCCAAGACGAAGACATC	140	-	60
	R	CAGATCACGTCATCGCACAAC			
PAX6	F	AACGATAACATACCAAGCGTGT	120	-	60
	R	GGTCTGCCCGTTCAACATC			
DCX	F	TGCCTCAGGGAGTGCGTTA	94	-	60
	R	GAACAGACATAGCTTTCCCCTTC			
Mammary Differentiation Markers					
CD24	F	AACTAATGCCACCACCAAGG	188	188	56
	R	CCTGTTTTTCCTTGCCACAT			
CD44	F	GGCTTTCAATAGCACCTTGC	152	3685	56
	R	ACACCCCTGTGTTGTTTGCT			
CK18	F	AAGTAAAAGGCCTACAAGCCCA	216	301	60
	R	CAGCTCCAACCTCAGCAGA			
CK14	F	GTCCCTACTTCAAGACCATTG	298	2446	60
	R	CCTCTCAGGGCATTCATCTC			
CD49f	F	GGGAGATTGGAGCTTTTGTG	230	1832	56
	R	CATGCTCAGTCTCTCCACCA			
ESR1	F	TGGAGATCTTCGACATGCTG	208	-	60
	R	GCCATCAGGTGGATCAAAGT			
PR	F	AAATCATTGCCAGGTTTTCG	209	-	58
	R	TGCCACATGGTAAGGCATAA			
CSN2	F	AAGGTCCTCATCCTCGCC	131	2074	58
	R	TGCTGGTCCTCATGTTTAAC			

Highlighted in orange are primers used for real-time PCR

Table S2-2 Primers used for RT-qPCR.

Genes		5′ – 3′	cDNA	gDNA	Ta
Housekeeping genes					
ACTB	F	CCTCGCCTTTGCCGATCC	70	930	58
	R	CGCGGCGATATCATCATCC			
Endogenous pluripotency markers					
REX1	F	GAAGAGGCCTTCACTCTAGTAGTG	179	3206	56
	R	TTTCTGGTGTCTTGTCTTTGCCCG			
All *NANOG* genes (NANOG_all)	F	TGCTGAGATGCCTCACACGGA	155	3339	
	R	TGACCGGGACCTTGTCTTCCTT			
NANOG (NF2/R2)	F	TGAGACGGAGTCTTGCTCTG	335	-	55
	R	GGCTTCTATCAATGTTGTCCTTA			
NANOG/ NANOGP8 (NF2/R3)	F	TGAGACGGAGTCTTGCTCTG	467/ 446	-	56
	R	CCTCCCAATCCCAAACAATA			
DPPA4	F	TGCCTCCAATTTTCCACCCC	216	1081	58
	R	CATGGCCCATAAACAGGTGC			
OCT3/4	F2	GATGGCGTACTGTGGGCCC	195	-	58
	R	TGGGACTCCTCCGGGTTTTG			
DNMT3B	F	TGCTGCTCACAGGGCCCGATACTT C	242	3910	58
	R	TCCTTTCGAGCTCAGTGCACCACAAAAC			

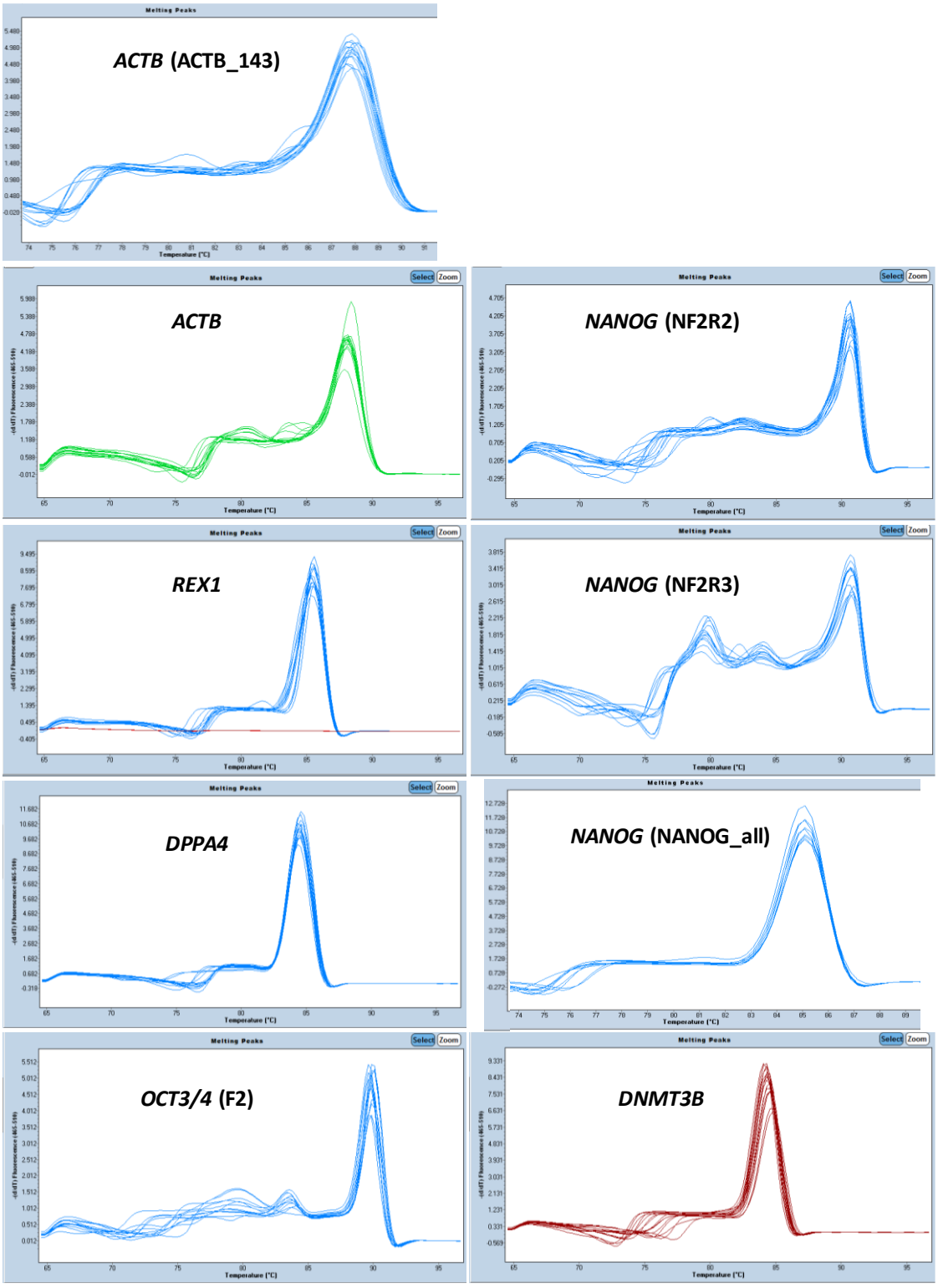


Figure S2-6 Melting curves for primers used for RT-qPCR.

Melting curves for amplicons generated from different primers used in RT-qPCR. A single peak indicates a single product amplified.

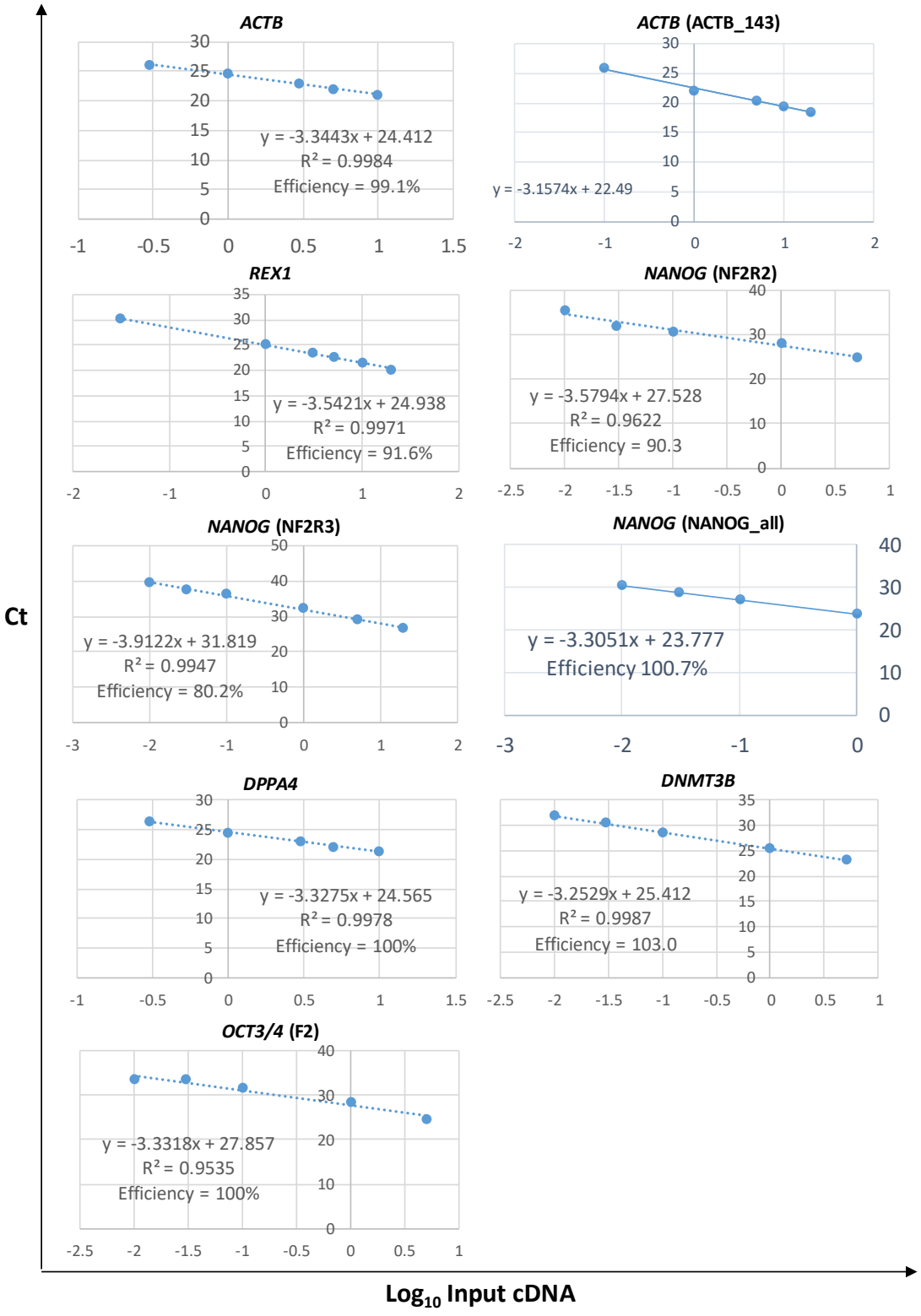


Figure S2-7 Efficiency of primers used for RT-qPCR.

RT-qPCR results were presented in a scatterplot of C_t value against Log_{10} input cDNA created on Microsoft Excel. The Thermo-Fisher qPCR efficiency calculator was used to calculate efficiencies.

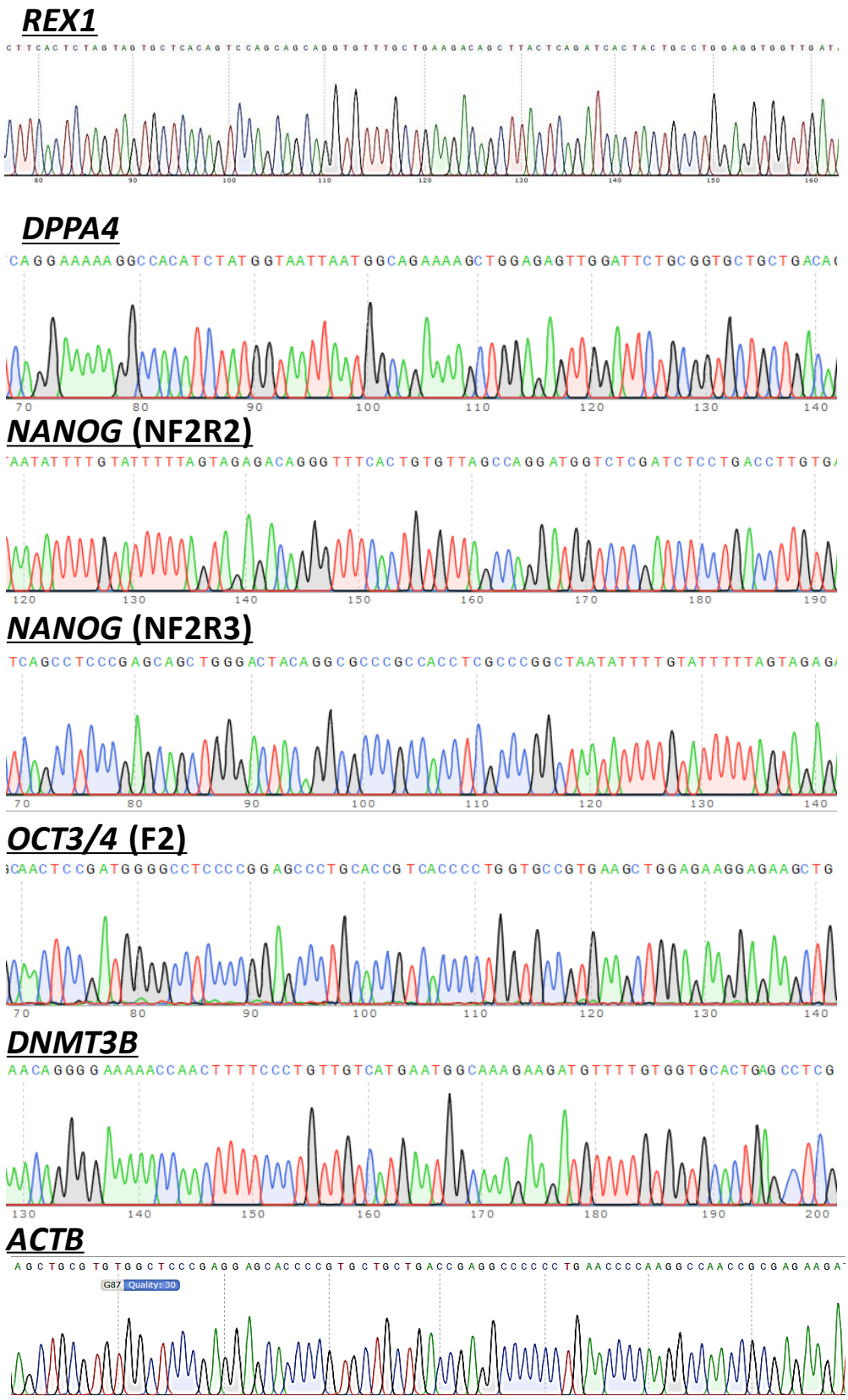


Figure S2-8 Sanger sequencing of amplicons generated by different PCR primer.

Results suggested all primers were specific for their respective targets.

Table S2-3 Antibodies used for immunofluorescence staining.

Antibodies	Company (Cat #)	Animal	Dilution
Primary antibody			
TRA-1-60	Hybridoma Bank (MAB-4381)	Mouse	1:10
TRA-1-81	Hybridoma Bank (MAB-4360)	Mouse	1:10
SSEA-4	Hybridoma Bank (MC-813-70)	Mouse	1:10
OCT3/4	Santa Cruz (sc-8628)	Goat	1:300
NANOG	Peprotech (500-P236)	Rabbit	1:300
AFP	Chemicon (2004189)	Mouse	1:100
AFP	R&D (189502)	Mouse	1:100
NES	Abcam (ab5968)	Rabbit	1:250
CDH2	Cell signalling tech (13116)	Rabbit	1:200
Secondary Antibody			
Cy3	Millipore AP180C	Donkey anti Goat	1:500
Alexa 488	ThermoFisher A-21206	Donkey anti rabbit	1:500
FITC	Jackson Laboratories 115-095-068	Goat anti Mouse	1:100

Highlighted in orange are antibodies also used for FACS

Table S2-4 Antibodies used for FACS analysis.

Antibodies	Fluorophores	Company (Cat #)	Animal	Dilution
TRA-1-81	PE	BD (560885)	Mouse	1:5
TRA-1-60	-	Hybridoma Bank (MAB-4360)	Mouse	1:10
SSEA-4	-	Hybridoma Bank (MC-813-70)	Mouse	1:10
CD24	FITC	eBioscience (11024742)	Mouse	1:20
CD24	PE	eBioscience (120241)	Mouse	1:150
CD44	FITC	eBioscience (11044182)	Mouse	1:100
CD44	PE	eBioscience (120441)	Mouse	1:150
CD44	APC	eBioscience (170441)	Mouse	1:300

Table S2-5 Antibodies used for Western blotting.

Antibodies	Company (Cat #)	Animal	Dilution
Primary antibody			
P53	eBioscience (17-8883)	Mouse	1:500
ACTB	BD (560885)	Mouse	1:500
Secondary antibody			
HRP	GE Healthcare NA931	Sheep anti-mouse	1:10000

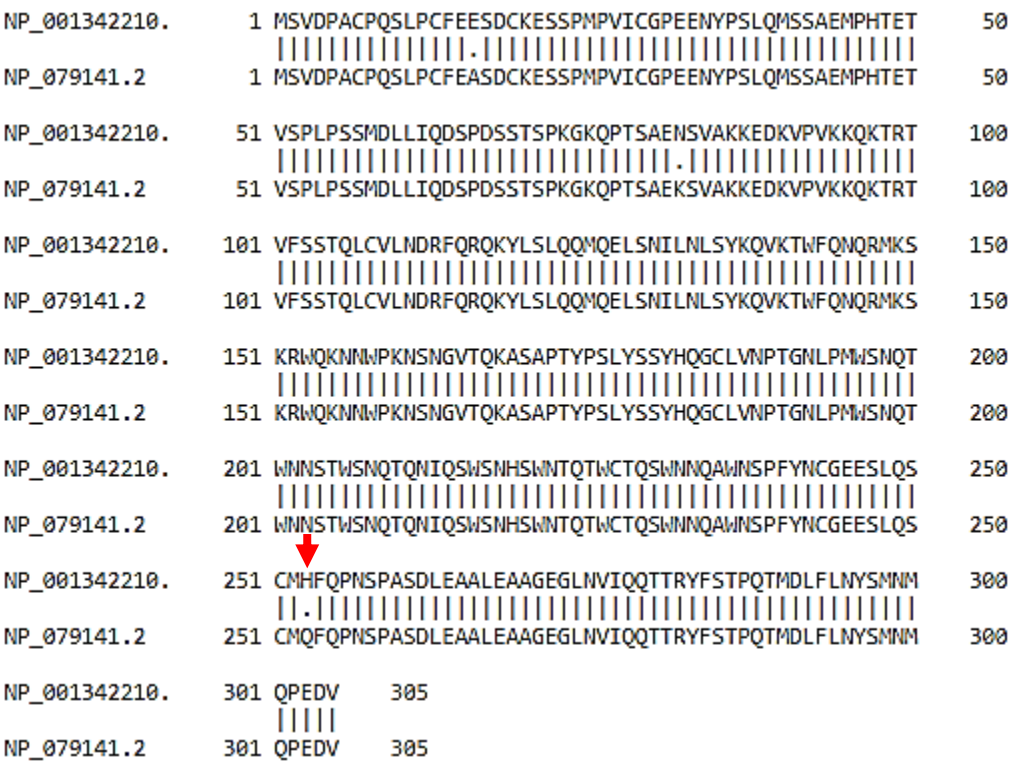


Figure S3-1 Protein sequences of NANOG (NP_079141.2) and NANOG P8 (NP_001342210.1). A nucleotide mismatch (CAG759 → CAC759) causes an amino acid change from Gln to His in NANOG and NANOGP8 respectively.

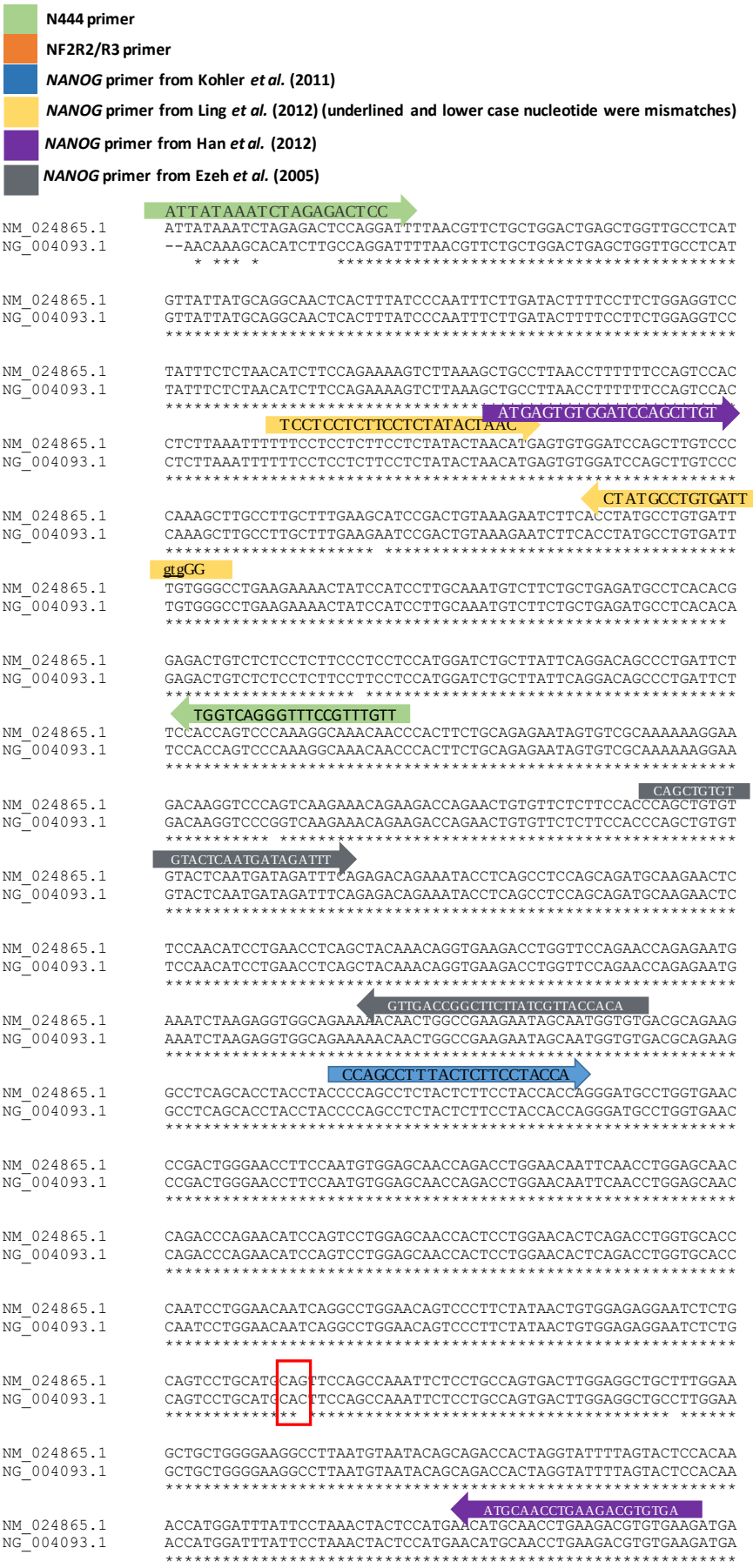

N444 primer
NF2R2/R3 primer
NANOG primer from Kohler et al. (2011)
NANOG primer from Ling et al. (2012) (underlined and lower case nucleotide were mismatches)
NANOG primer from Han et al. (2012)
NANOG primer from Ezeh et al. (2005)
ATTATAAATCTAGAGACTCC
ATGAGTGTGGATCCAGCTTGT
TCCTCCTCTTCCTCTATACTAAC
CTATGCCTGTGATT
gtgGG
TGGTCAGGGTTTCCGTTTGTT
CAGCTGTGT
GTACTCAATGATAGATTT
GTTGACCGGCTTCTTATCGTTACCACA
CCAGCCTTTACTCTTCCTACCA
ATGCAACCTGAAGACGTGTGA



Figure S3-2 Binding sites of different primers on *NANOG* (NM_024865.1) and *NANOG P8* (NG_004093.1).

(red box) A nucleotide mismatch (CAG759 → CAC759) between *NANOG* and *NANOGP8*. Position of arrows indicates sequence where different primers bind. N444 and NF2R2 primers take advantage of the 5'UTR and 3'UTR mismatches respectively to distinguish genes. forward arrow: forward primer, backward arrow: reverse primer.

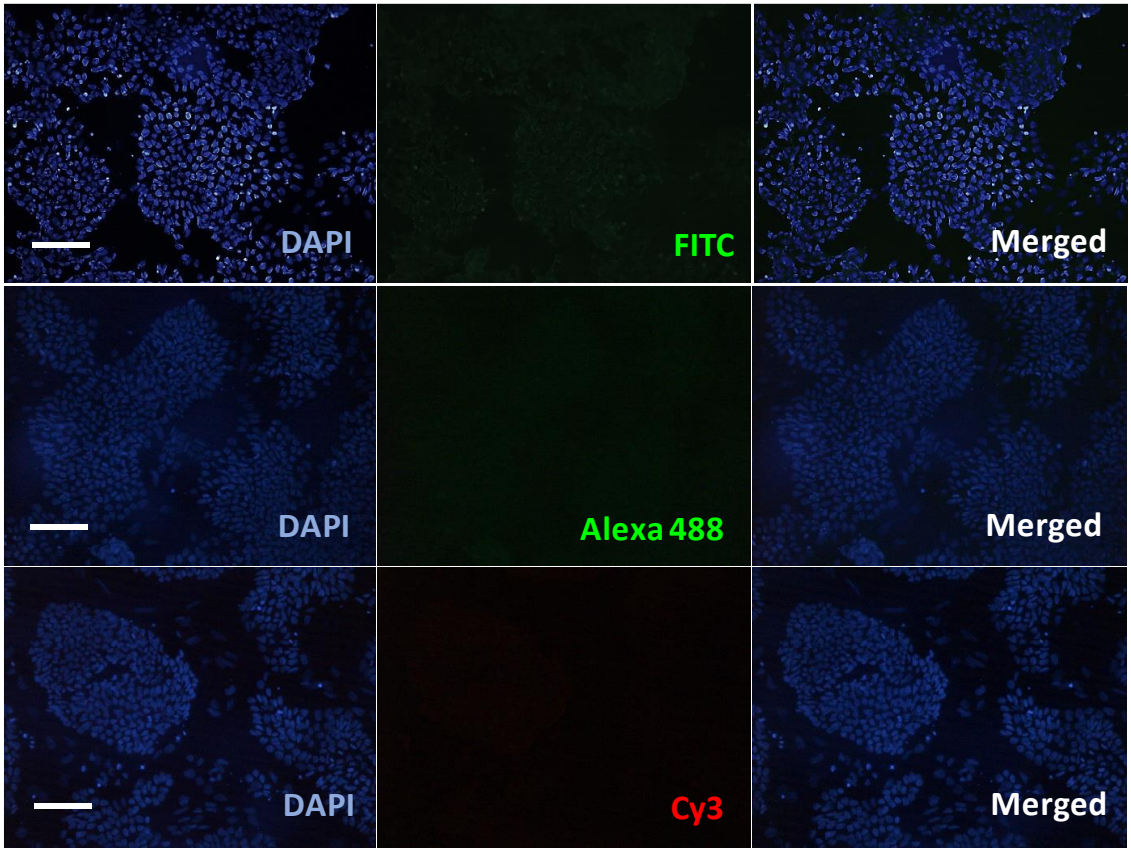


Figure S3-3 Negative controls for immunofluorescence staining.

Negative controls consisted of 10AR stained by secondary antibodies only. Scale bar = 100 μm.

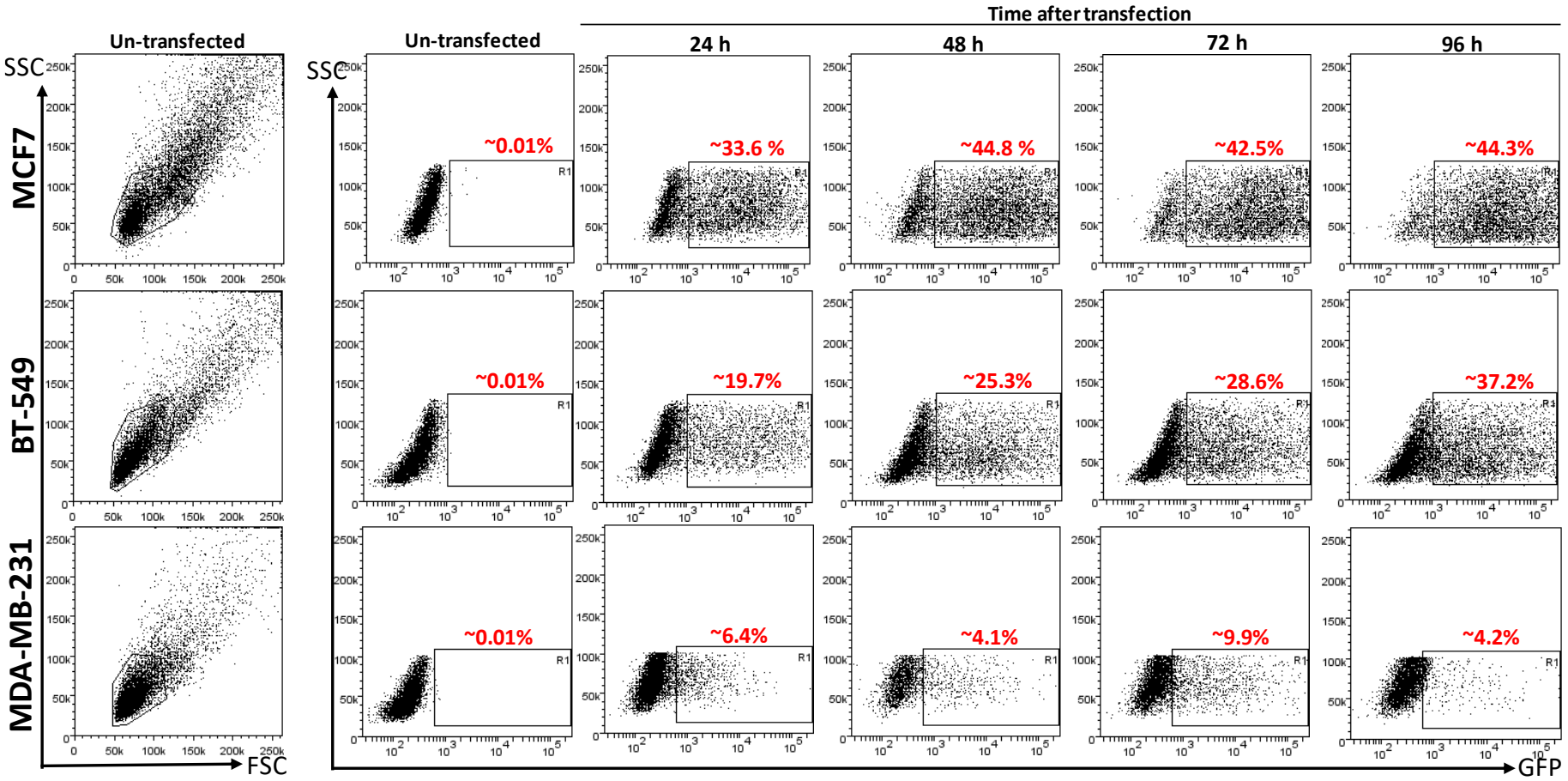


Figure 4S-1 Amaxa® Cell line Nucleofector™ kits (Lonza) transfection.

Breast cancer cells were transfected with pCXLE-EGFP (Addgene #27082) plasmid and GFP signal was monitored at different time points using FACS analysis. Overall transfection efficiency was low as compared to XG9 in all three cell lines. n = 1. SSC = side scatter.

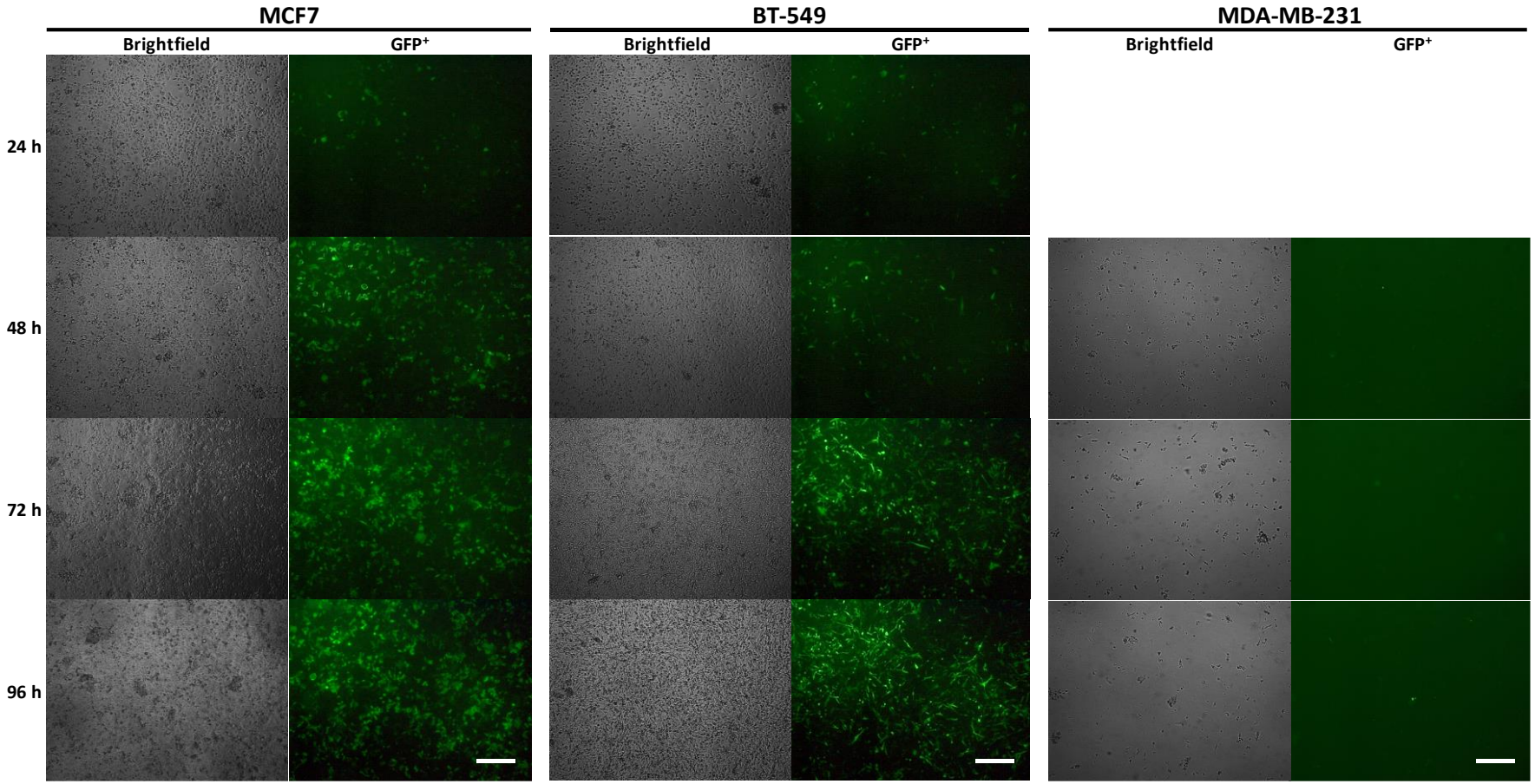


Figure 4S-2 GFP+ cells in three different breast cancer cell lines at different time points after Nucleofector™ kit transfection.

Cells were approximately 80-90% confluent at 24 h after initial transfection in MCF7 and BT-549. On the contrary, heavy cell death was observed in MDA-MB-231. Sustained GFP expression was apparent in MCF7 and BT-549 to as long as 96 h. GFP signal remained low in MDA-MB-231 and barely detectable under such magnification. Scale bar = 200 μm.

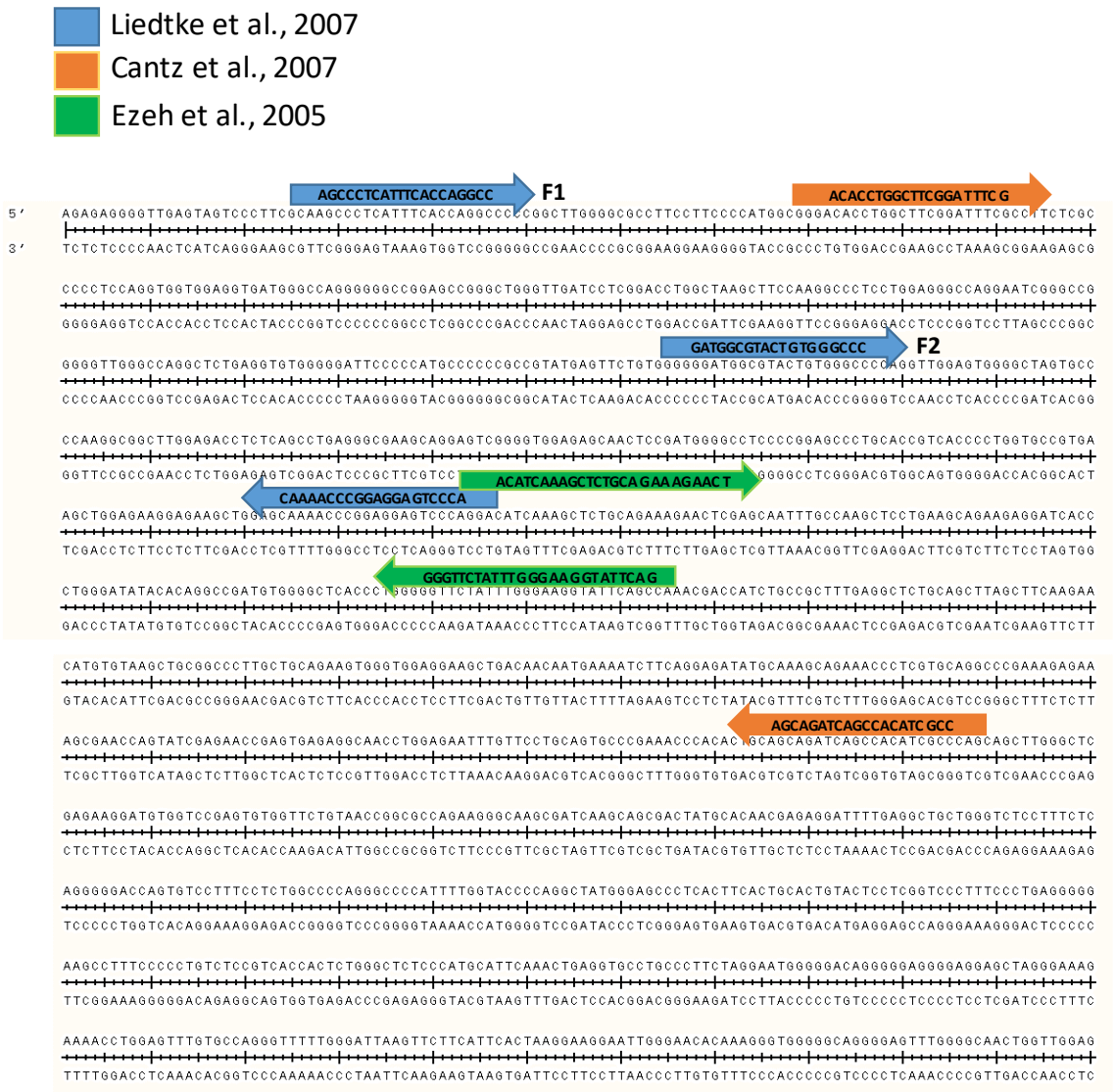


Figure 4S-3 Binding sites of primers from different studies on the original OCT3/4 gene sequence (NM_002701.5).

Primers from studies above all recognise and amplify the original *OCT3/4* sequence. Primers from Liedtke et al. (2007) were used in the present study as the other two primers also recognise other non-*OCT3/4* sequences in the genome (see following figures).

Forward primer Reverse primer

>chr12:8134062-8134843 782bp ACACCTGGCTTCGGATTTCG GGCGATGTGGCTGATCTGCT
ACACCTGGCTTCGGATTTCGccttctcacccccctccaggcggtggaggtg
atgggccaggggggccggagccgggctgggttgatcctcggacctggcta
agcttccaaggccctcctggagggccaggaatcgggccggggtttgggcc
aggctctgaggagtgggggattcccccatgtcccccgccgtatgagttct
gcgggggggatggcgtactgtgggcctcagactggagtggggctagtgccc
caagacggcttggagacctctcagcctgagggcgaagcaggagtcggggt
ggagagcaactccgatggggcctccccggagccctgcaccgtcccctctg
gtgccgtgaagctggagaaggagaagctggagcaaaacccggaggagtcc
caggacatcaaagctctgcagaaagaactcgagcaatttgccaagctcct
gaagcagaagaggatcaccctgggatatacacaggccgatgtggctcacc
ctgggggttctatttgggaaggtgttcagccaaacgaccatctgccgctt
tgaggctctgcagcttagcttcaagaacatgtgtgagctgcggcccttgc
tgcagaagtgggtggaggaagctgacaacaatgaaaatcttcaggagata
tgcaaagcagaaaccctcgtgcaggcccgaaagagaaagcgaaccagtat
cgagaaccaagtgagaggcaacctggagaatttgttcctgcggtgcccga
aacccacactgcAGCAGATCAGCCACATCGCC

Figure 4S-4 *OCT3/4* primers reported by Cantz et al. (2008).

University of California Santa Cruz (UCSC) Genome Browser in-silico PCR analysis indicated possible amplification of the intron sequence (chr12: 8134062 – 8134843) of *CELC4A* resulting in a 782 bp product of the same size as *OCT3/4* gene.

	Sequence (5'->3')	Length	Tm	GC%
Forward primer	GCATTCAAACTGAGGCACCA	20	59.04	50.00
Reverse primer	AGCTTCTTTCCCCATCCCA	19	58.19	52.63

Products on target templates

>XM_021149626.1 PREDICTED: Mus caroli POU class 5 homeobox 1 (Pou5f1), transcript variant X2, mRNA

```
product length = 153
Forward primer  1    GCATTCAAACTGAGGCACCA  20
Template        753  ....................  772

Reverse primer  1    AGCTTCTTTCCCCATCCCA  19
Template        905  ....................  887
```

>XM_021149625.1 PREDICTED: Mus caroli POU class 5 homeobox 1 (Pou5f1), transcript variant X1, mRNA

```
product length = 153
Forward primer  1     GCATTCAAACTGAGGCACCA  20
Template        1115  ....................  1134

Reverse primer  1     AGCTTCTTTCCCCATCCCA  19
Template        1267  ....................  1249
```

>NM_001252452.1 Mus musculus POU domain, class 5, transcription factor 1 (Pou5f1), transcript variant 2, mRNA

```
product length = 153
Forward primer  1    GCATTCAAACTGAGGCACCA  20
Template        753  ....................  772

Reverse primer  1    AGCTTCTTTCCCCATCCCA  19
Template        905  ....................  887
```

Figure 4S-5 *OCT3/4* primers reported by Kim and Nam (2011).

Primer-BLAST analysis indicated primer is specific for *Mus musculus* and not human *OCT3/4*.

Forward primer Reverse primer

>chr1:155433593+155433719 127bp ACATCAAAGCTCTGCAGAAAGAACT CTGAATACCTTCCCAAATAGAACCC
ACATCAAAGCTCTGCAGAAAGAACTcgaacaatttgccaagctcctgaag
cagaagaggatcaccctgggatatacacaggccgatgtggggctcaccct
ggGGGTTCTATTTGGGAAGGTgTTCAG
>chr8:127416279+127416405 127bp ACATCAAAGCTCTGCAGAAAGAACT CTGAATACCTTCCCAAATAGAACCC
ACATCAAAGCTCTGCAGAAAGAACTcgagcaatttgccaagctcctgaag
cagaagaggatcaccctgggatatacacaggccgatgtggggctcatcct
ggGGGTTCTATTTGGGAAGGTgTTCAG
>chr10:68010742-68010868 127bp ACATCAAAGCTCTGCAGAAAGAACT CTGAATACCTTCCCAAATAGAACCC
AtcTCAAAGCTCTGCAGAAAGAACTgagcaatttgccaagctcctgaag
cagaagaggatcaccctgggatatgcacaggctgatgtggggcttaccct
ggGGGTTCTATTTGGGAAtaTgTTCAG
>chr12:8134315-8134439 125bp ACATCAAAGCTCTGCAGAAAGAACT CTGAATACCTTCCCAAATAGAACCC
ACATCAAAGCTCTGCAGAAAGAACTcgagcaatttgccaagctcctgaag
cagaagaggatcaccctgggatatacacaggccgatgtggctcaccctgg
GGGTTCTATTTGGGAAGGTgTTCAG

Figure 4S-6 OCT3/4 primer reported by Ezeh et al. (2005).

UCSC in silico PCR analysis indicated primers can amplify different genomic regions in addition to *OCT3/4* gene, including *ASH1L, CASC8, HERC4, CLEC4A*.

	Sequence (5'->3')	Length	Tm	GC%
Forward primer	GCTGTATCCTTTCCTCTGCC	20	58.04	55.00
Reverse primer	TCTTGTCTACCTCCCTTGCC	20	58.72	55.00

Products on target templates

>XM_021149626.1 PREDICTED: Mus caroli POU class 5 homeobox 1 (Pou5f1), transcript variant X2, mRNA

```
product length = 210
Forward primer  1    GCTGTATCCTTTCCTCTGCC  20
Template        610  ....................  629

Reverse primer  1    TCTTGTCTACCTCCCTTGCC  20
Template        819  ....................  800
```

>XM_021149625.1 PREDICTED: Mus caroli POU class 5 homeobox 1 (Pou5f1), transcript variant X1, mRNA

```
product length = 210
Forward primer  1    GCTGTATCCTTTCCTCTGCC  20
Template        972  ....................  991

Reverse primer  1     TCTTGTCTACCTCCCTTGCC  20
Template        1181  ....................  1162
```

>NM_001252452.1 Mus musculus POU domain, class 5, transcription factor 1 (Pou5f1), transcript variant 2, mRNA

```
product length = 210
Forward primer  1    GCTGTATCCTTTCCTCTGCC  20
Template        610  ....................  629

Reverse primer  1    TCTTGTCTACCTCCCTTGCC  20
Template        819  ....................  800
```

Figure 4S-7 *OCT3/4* primer reported by Hu et al. (2008).

Primer-BLAST analysis indicated primer is specific for *Mus musculus* and not human *OCT3/4*.

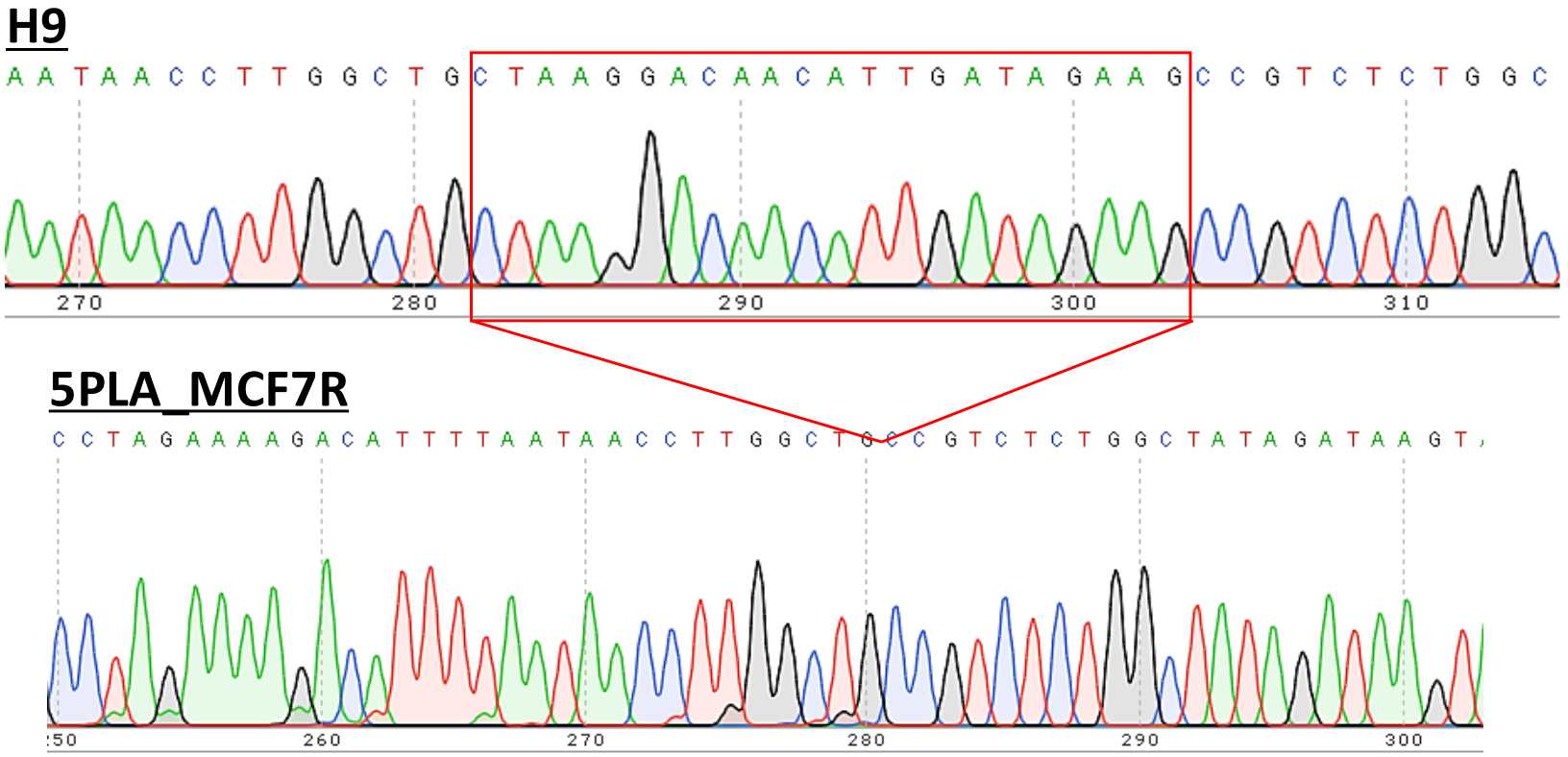


Figure 4S-8 Sanger sequencing of PCR products from H9 and 5PLA_MCF7R amplified by NF2R3 primer. *NANOG P8* does not contain a 22 bp sequence which is present in the 5' UTR of *NANOG* (Appendix Figure 3S-2). Result indicated *NANOG* and *NANOG P8* expression in H9 and 5PLA_MCF7R respectively.

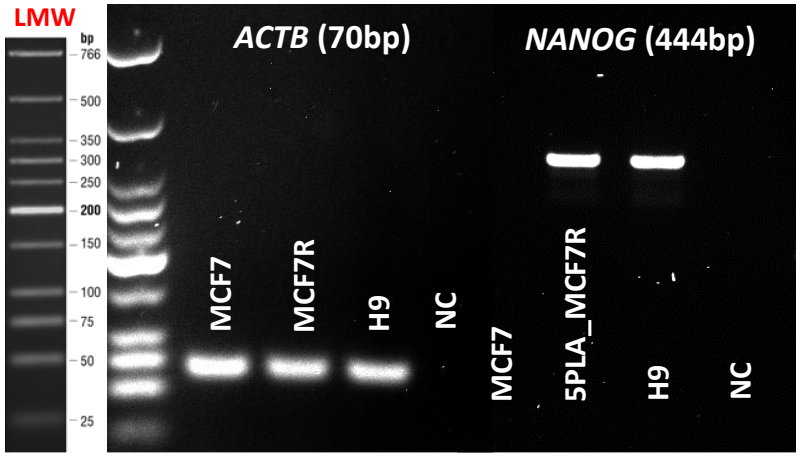


Figure 4S-9 RT-PCR for *NANOG* expression in MCF7 and reprogrammed MCF7 (5PLA_MCF7R).

Positivity of 5PLA_MCF7R to N444 indicated the presence of *NANOG*, an observation contradicted results obtained by NF2R2 and NF2R3 primers.

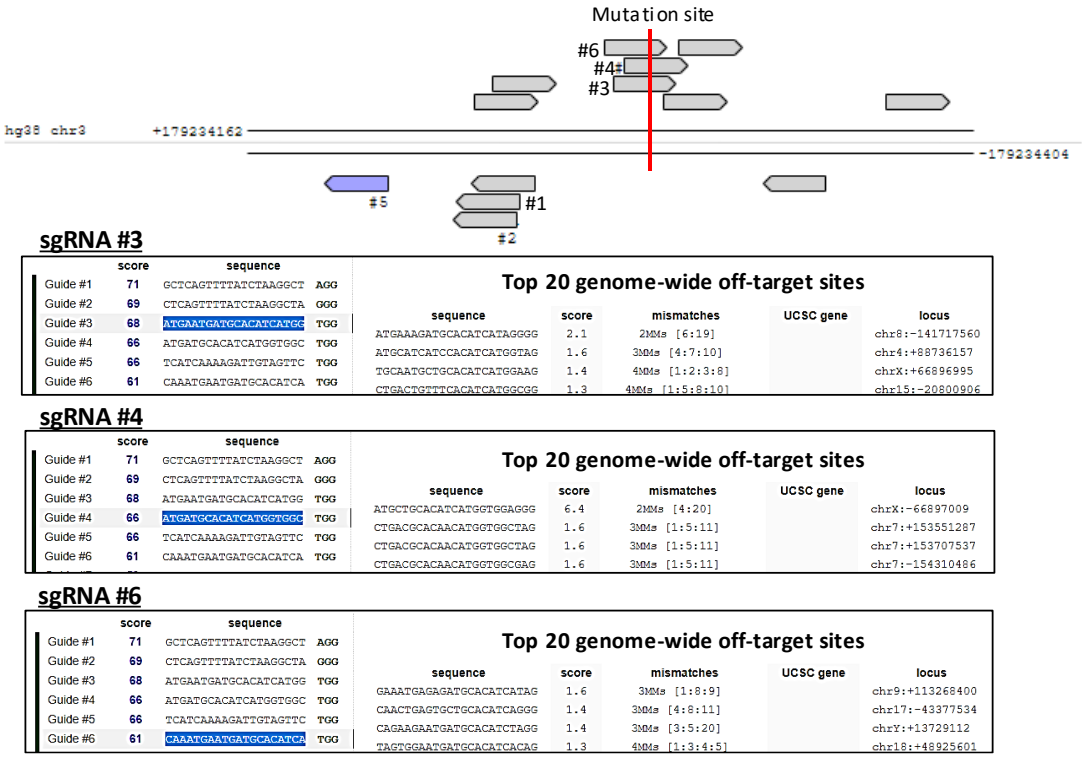
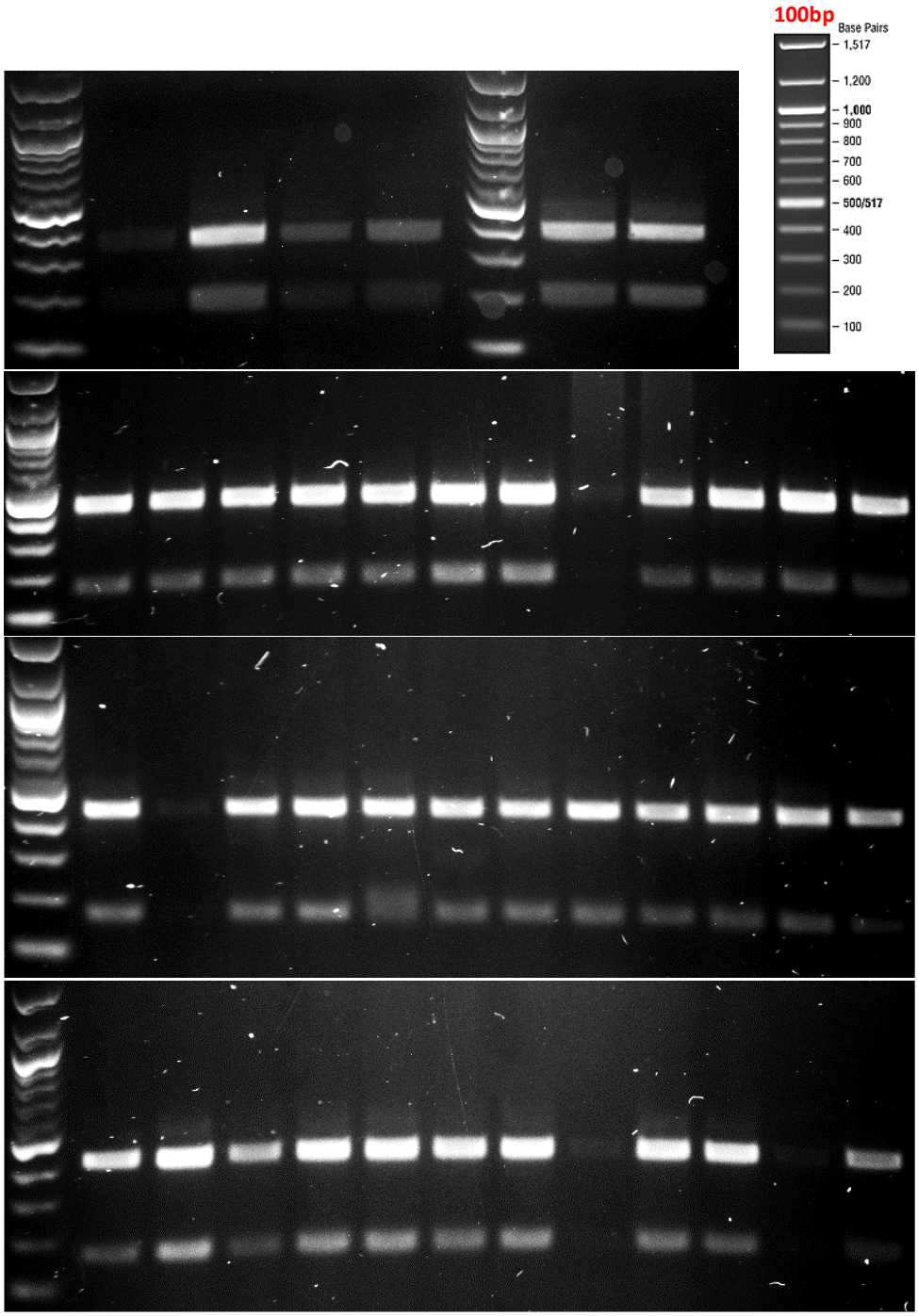
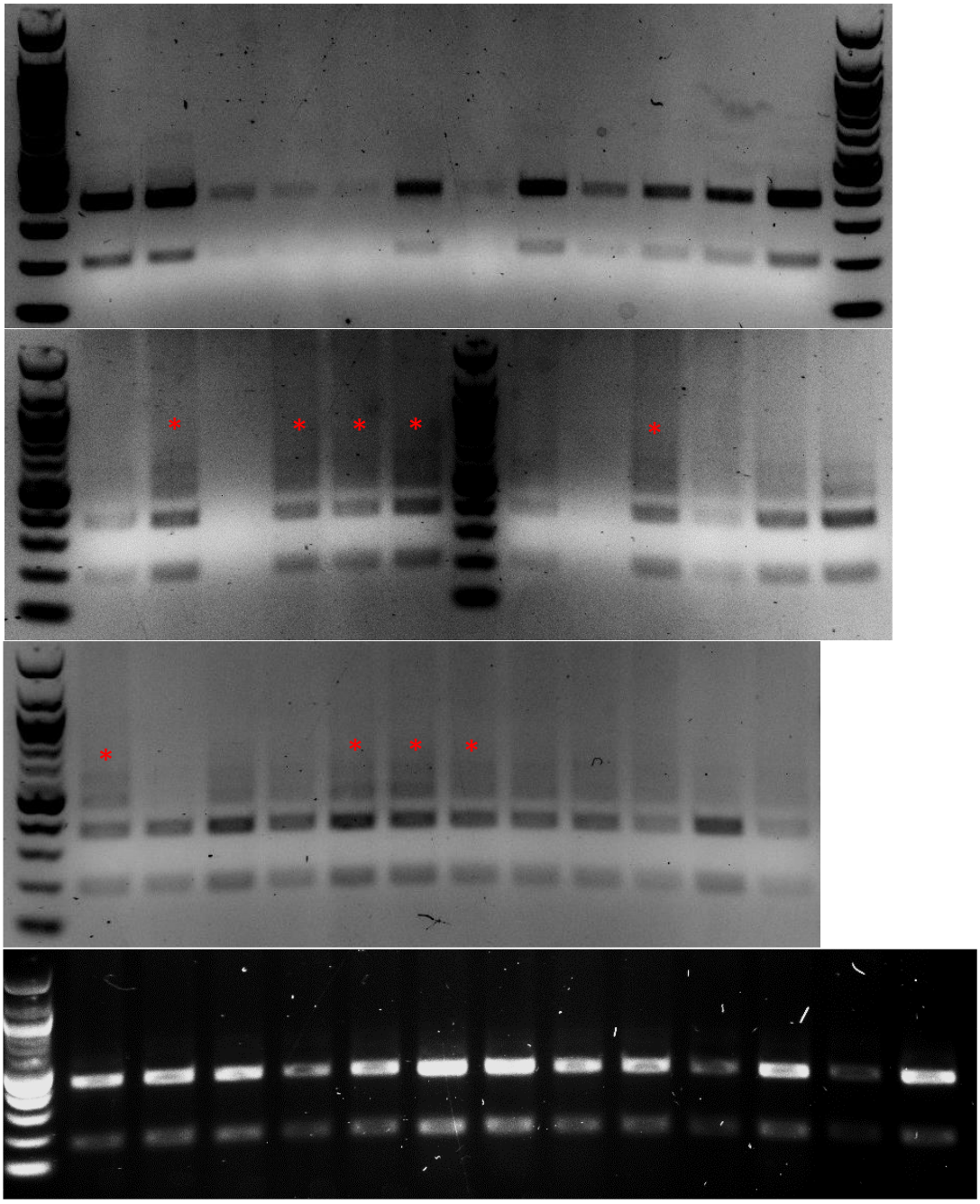


Figure 5S-1 sgRNA design with online CRISPR tool (crispr.mit.edu).

Despite having higher quality scores, sgRNA #1, #2 and #5 locate further from the mutation site (red line), which could potentially compromise efficiency of gene editing. sgRNA #3 and #4 have one off-target site with 2 nucleotide mismatches, which could potentially augment off-target effect of both sgRNAs. sgRNA #6 is deemed to be the optimum sgRNA.


100bp
Base Pairs
1,517
1,200
1,000
900
800
700
600
500/517
400
300
200
100



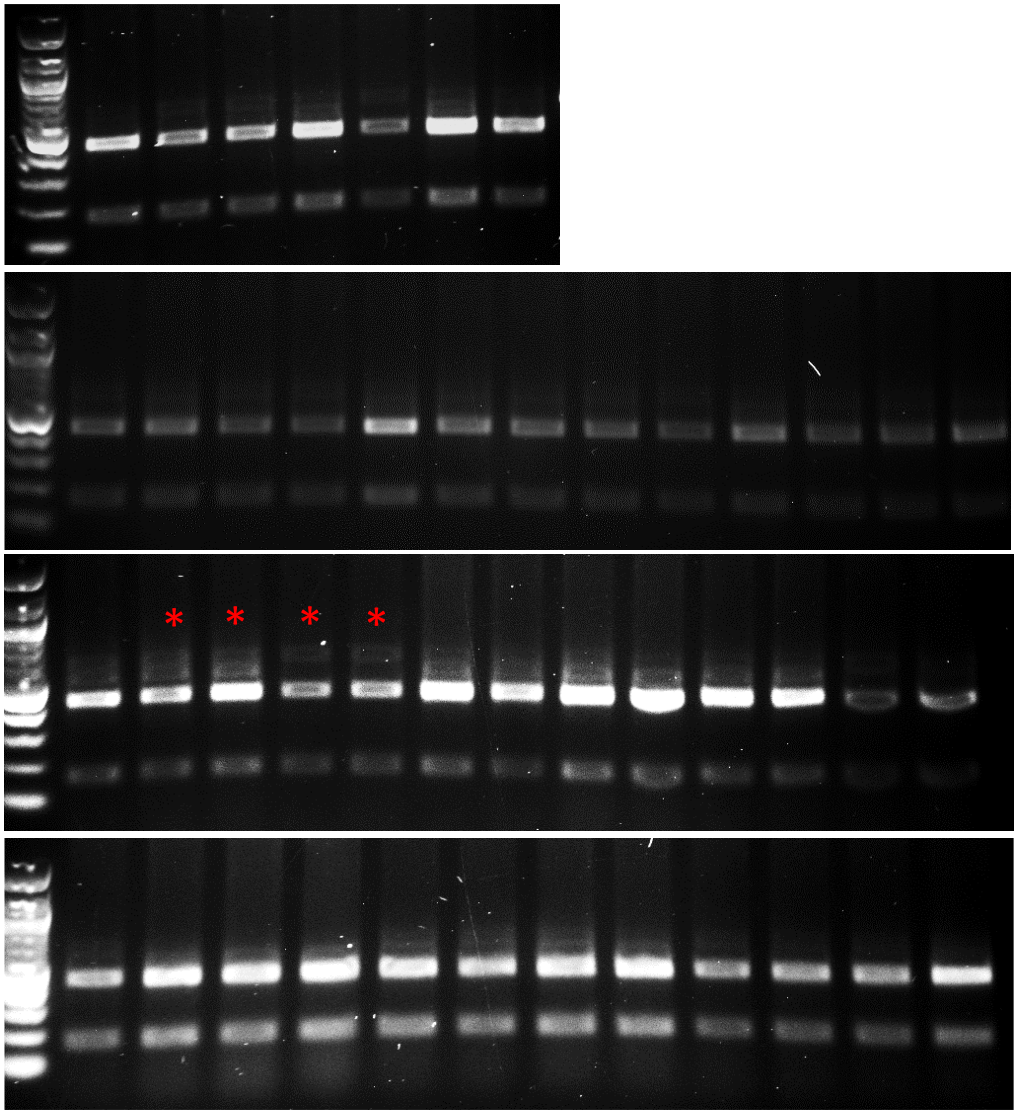
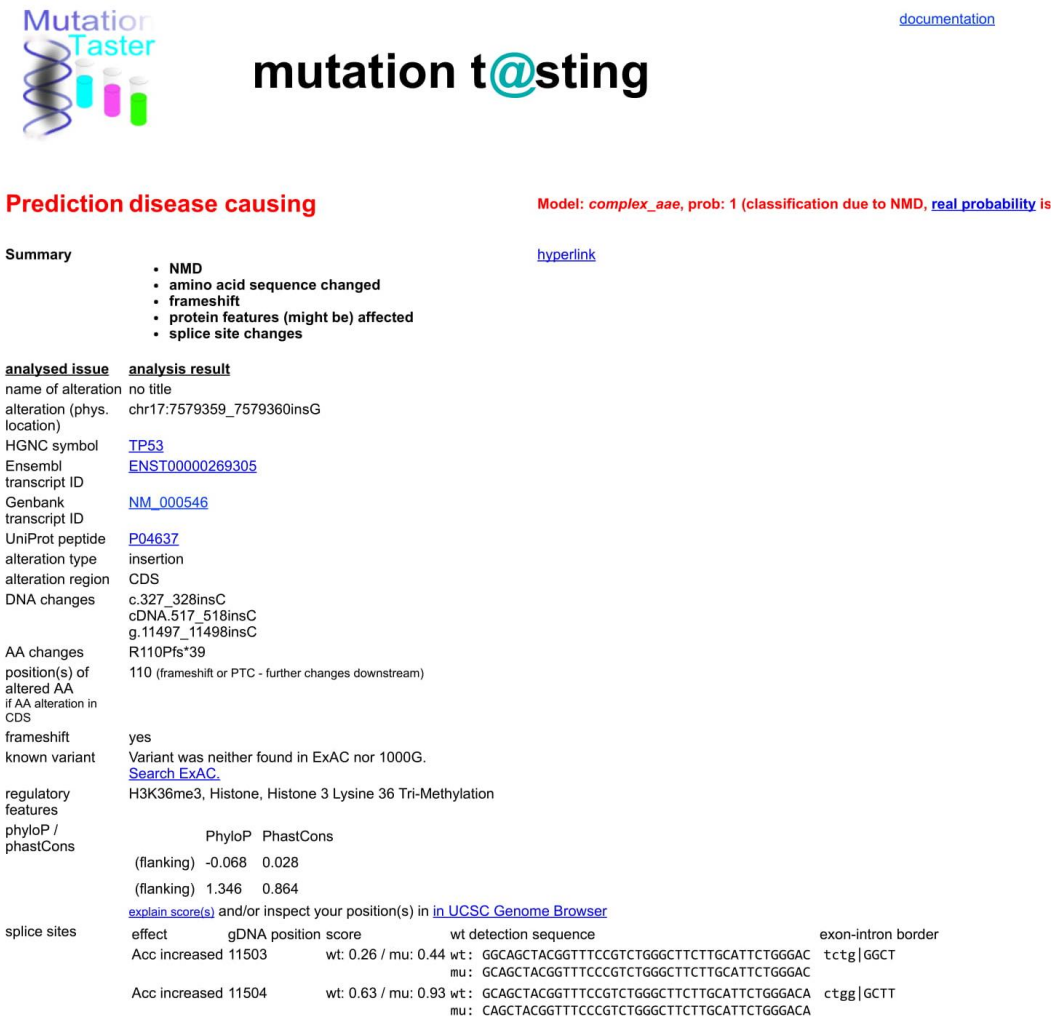


Figure 5S-2 Screening of PIK3CA mutants with BspH1 restriction enzyme.

Genomic DNA was extracted and amplified by PCR to generate a 597 bp product. BspHI was used to digest the PCR product (421 bp and 176 bp) to identify the potential abolishment of restriction site by HDR. *100 bp ladder was used in all experiments shown. * = clones re-screened and found negative.



documentation

mutation t@sting

Prediction disease causing

Model: *complex_aae*, prob: 1 (classification due to NMD, real probability is hyperlink

Summary

- **NMD**
- **amino acid sequence changed**
- **frameshift**
- **protein features (might be) affected**
- **splice site changes**

analysed issue	analysis result
name of alteration	no title
alteration (phys. location)	chr17:7579359_7579360insG
HGNC symbol	TP53
Ensembl transcript ID	ENST00000269305
Genbank transcript ID	NM_000546
UniProt peptide	P04637
alteration type	insertion
alteration region	CDS
DNA changes	c.327_328insC cDNA.517_518insC g.11497_11498insC
AA changes	R110Pfs*39
position(s) of altered AA if AA alteration in CDS	110 (frameshift or PTC - further changes downstream)
frameshift	yes
known variant	Variant was neither found in ExAC nor 1000G. Search ExAC.
regulatory features	H3K36me3, Histone, Histone 3 Lysine 36 Tri-Methylation
phyloP / phastCons	PhyloP PhastCons (flanking) -0.068 0.028 (flanking) 1.346 0.864 explain score(s) and/or inspect your position(s) in in UCSC Genome Browser

splice sites

effect	gDNA position	score	wt detection sequence	exon-intron border
Acc increased	11503	wt: 0.26 / mu: 0.44	wt: GGCAGCTACGGTTTCCGTCTGGGCTTCTTGCATTCTGGGAC mu: GCAGCTACGGTTTCCCGTCTGGGCTTCTTGCATTCTGGGAC	tctg|GGCT
Acc increased	11504	wt: 0.63 / mu: 0.93	wt: GCAGCTACGGTTTCCGTCTGGGCTTCTTGCATTCTGGGACA mu: CAGCTACGGTTTCCCGTCTGGGCTTCTTGCATTCTGGGACA	ctgg|GCTT

Figure S5-3 Evaluation of mutation in clone A3 allele 1 and 2 using MutationTaster.

Insertion was predicted to cause frameshift and was possibly disease causing.



documentation

mutation t@sting

Alteration TP53allele1

Prediction disease causing

Model: *complex_aae*, prob: 1 (classification due to ClinVar, real probability is s

Summary

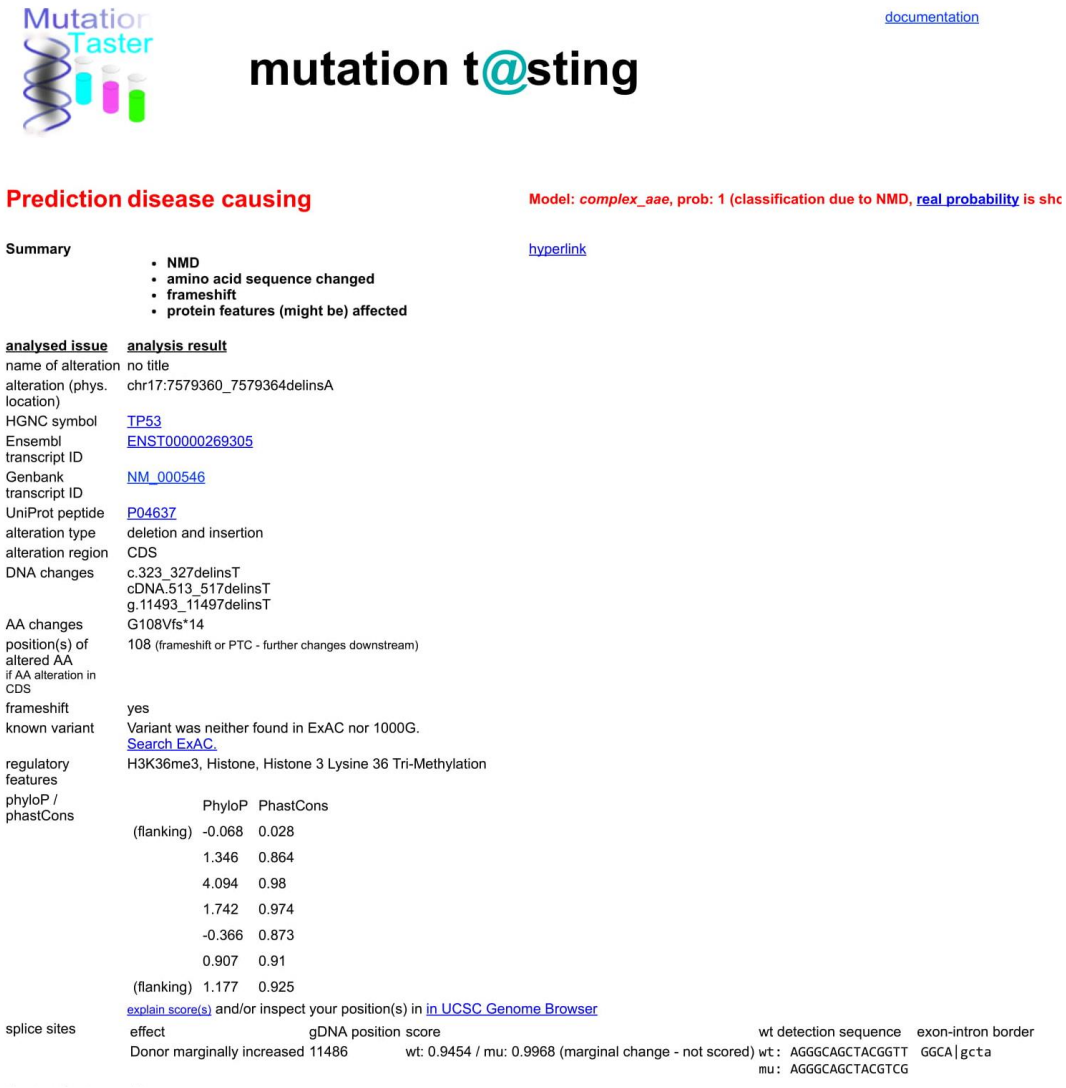
hyperlink

- **NMD**
- **amino acid sequence changed**
- **frameshift**
- **known disease mutation: rs587780066 (pathogenic)**
- **protein features (might be) affected**
- **splice site changes**

analysed issue	analysis result
name of alteration	TP53allele1
alteration (phys. location)	chr17:7579359_7579359delG
HGNC symbol	TP53
Ensembl transcript ID	ENST00000269305
Genbank transcript ID	NM_000546
UniProt peptide	P04637
alteration type	deletion
alteration region	CDS
DNA changes	c.328_328delC cDNA.518_518delC g.11498_11498delC
AA changes	R110Vfs*13
position(s) of altered AA if AA alteration in CDS	110 (frameshift or PTC - further changes downstream)
frameshift	yes
known variant	Reference ID: rs587780066 Allele '-' was neither found in ExAC nor 1000G. **known disease mutation: rs587780066 (pathogenic for *Hereditary_cancer-predisposing_syndrome*)** dbSNP NCBI variation viewer
regulatory features	H3K36me3, Histone, Histone 3 Lysine 36 Tri-Methylation
phyloP / phastCons	PhyloP PhastCons (flanking) 0.199 0.004 -0.068 0.028 (flanking) 1.346 0.864 explain score(s) and/or inspect your position(s) in in UCSC Genome Browser
splice sites	effect gDNA position score wt detection sequence exon-intron border Donor gained 11492 0.49 mu: GCTACGGTTTCGTCT TACG|gttt

Figure S5-4 Evaluation of mutation in clone A3 allele 3 using MutationTaster.

Deletion was predicted to cause frameshift shown to be known disease mutation reported in dbSNP (rs587780066).



documentation

mutation t@sting

Prediction disease causing

Model: *complex_aae*, prob: 1 (classification due to NMD, real probability is shc

Summary

hyperlink

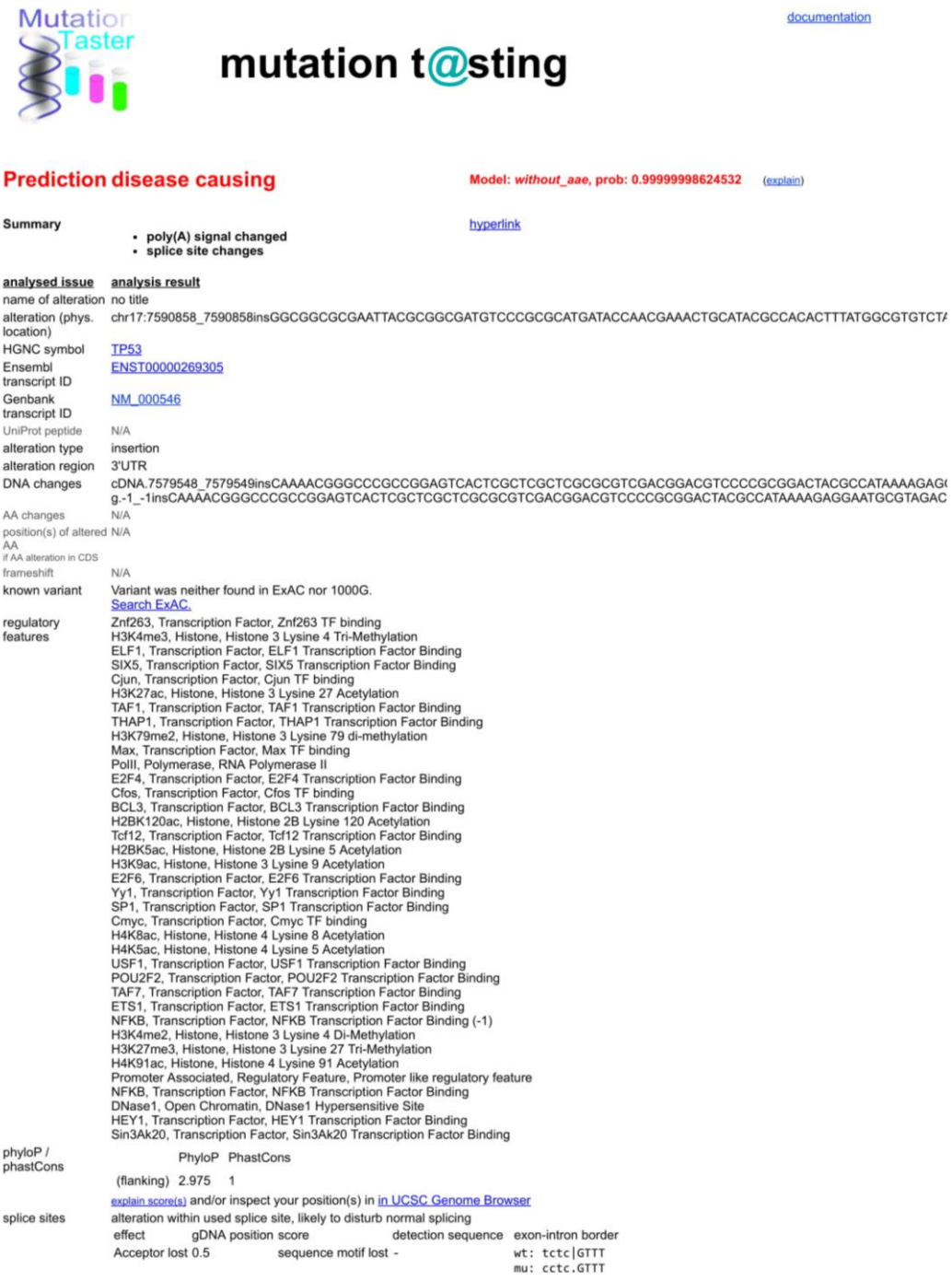
- **NMD**
- **amino acid sequence changed**
- **frameshift**
- **protein features (might be) affected**

analysed issue	analysis result
name of alteration	no title
alteration (phys. location)	chr17:7579360_7579364delinsA
HGNC symbol	TP53
Ensembl transcript ID	ENST00000269305
Genbank transcript ID	NM_000546
UniProt peptide	P04637
alteration type	deletion and insertion
alteration region	CDS
DNA changes	c.323_327delinsT cDNA.513_517delinsT g.11493_11497delinsT
AA changes	G108Vfs*14
position(s) of altered AA if AA alteration in CDS	108 (frameshift or PTC - further changes downstream)
frameshift	yes
known variant	Variant was neither found in ExAC nor 1000G. Search ExAC.
regulatory features	H3K36me3, Histone, Histone 3 Lysine 36 Tri-Methylation
phyloP / phastCons	PhyloP PhastCons (flanking) -0.068 0.028 1.346 0.864 4.094 0.98 1.742 0.974 -0.366 0.873 0.907 0.91 (flanking) 1.177 0.925 explain score(s) and/or inspect your position(s) in in UCSC Genome Browser

splice sites	effect	gDNA position	score	wt detection sequence	exon-intron border
	Donor marginally increased	11486	wt: 0.9454 / mu: 0.9968 (marginal change - not scored)	wt: AGGGCAGCTACGGTT mu: AGGGCAGCTACGTCG	GGCA|gcta

Figure S5-5 Evaluation of mutation in clone A8 allele 1 – 3 using MutationTaster.

Deletion and substitution were predicted to cause frameshift and were possibly disease causing.



MutationTaster

documentation

mutation t@sting

Prediction disease causing

Model: *without_aae*, prob: 0.99999998624532 (explain)

Summary hyperlink

- **poly(A) signal changed**
- **splice site changes**

analysed issue	analysis result
name of alteration	no title
alteration (phys. location)	chr17:7590858_7590858insGGCGGCGCGAATTACGCGGCGATGTCCCGCGCATGATACCAACGAAACTGCATACGCCACACTTTATGGCGTGTCTA
HGNC symbol	TP53
Ensembl transcript ID	ENST00000269305
Genbank transcript ID	NM_000546
UniProt peptide	N/A
alteration type	insertion
alteration region	3'UTR
DNA changes	cDNA.7579548_7579549insCAAAACGGGCCCGCCGGAGTCACTCGCTCGCTCGCGCGTCGACGGACGTCCCCGCGGACTACGCCATAAAAGAG(g.-1_-1insCAAAACGGGCCCGCCGGAGTCACTCGCTCGCTCGCGCGTCGACGGACGTCCCCGCGGACTACGCCATAAAAGAGGAATGCGTAGAC
AA changes	N/A
position(s) of altered AA if AA alteration in CDS	N/A
frameshift	N/A
known variant	Variant was neither found in ExAC nor 1000G. Search ExAC.
regulatory features	Znf263, Transcription Factor, Znf263 TF binding H3K4me3, Histone, Histone 3 Lysine 4 Tri-Methylation ELF1, Transcription Factor, ELF1 Transcription Factor Binding SIX5, Transcription Factor, SIX5 Transcription Factor Binding Cjun, Transcription Factor, Cjun TF binding H3K27ac, Histone, Histone 3 Lysine 27 Acetylation TAF1, Transcription Factor, TAF1 Transcription Factor Binding THAP1, Transcription Factor, THAP1 Transcription Factor Binding H3K79me2, Histone, Histone 3 Lysine 79 di-methylation Max, Transcription Factor, Max TF binding PolII, Polymerase, RNA Polymerase II E2F4, Transcription Factor, E2F4 Transcription Factor Binding Cfos, Transcription Factor, Cfos TF binding BCL3, Transcription Factor, BCL3 Transcription Factor Binding H2BK120ac, Histone, Histone 2B Lysine 120 Acetylation Tcf12, Transcription Factor, Tcf12 Transcription Factor Binding H2BK5ac, Histone, Histone 2B Lysine 5 Acetylation H3K9ac, Histone, Histone 3 Lysine 9 Acetylation E2F6, Transcription Factor, E2F6 Transcription Factor Binding Yy1, Transcription Factor, Yy1 Transcription Factor Binding SP1, Transcription Factor, SP1 Transcription Factor Binding Cmyc, Transcription Factor, Cmyc TF binding H4K8ac, Histone, Histone 4 Lysine 8 Acetylation H4K5ac, Histone, Histone 4 Lysine 5 Acetylation USF1, Transcription Factor, USF1 Transcription Factor Binding POU2F2, Transcription Factor, POU2F2 Transcription Factor Binding TAF7, Transcription Factor, TAF7 Transcription Factor Binding ETS1, Transcription Factor, ETS1 Transcription Factor Binding NFKB, Transcription Factor, NFKB Transcription Factor Binding (-1) H3K4me2, Histone, Histone 3 Lysine 4 Di-Methylation H3K27me3, Histone, Histone 3 Lysine 27 Tri-Methylation H4K91ac, Histone, Histone 4 Lysine 91 Acetylation Promoter Associated, Regulatory Feature, Promoter like regulatory feature NFKB, Transcription Factor, NFKB Transcription Factor Binding DNase1, Open Chromatin, DNase1 Hypersensitive Site HEY1, Transcription Factor, HEY1 Transcription Factor Binding Sin3Ak20, Transcription Factor, Sin3Ak20 Transcription Factor Binding
phyloP / phastCons	PhyloP PhastCons (flanking) 2.975 1 explain score(s) and/or inspect your position(s) in in UCSC Genome Browser
splice sites	alteration within used splice site, likely to disturb normal splicing effect gDNA position score detection sequence exon-intron border Acceptor lost 0.5 sequence motif lost - wt: tctc|GTTT mu: cctc.GTTT

Figure S5-6 Evaluation of mutation in clone A8 allele 4 using MutationTaster.

Insertion was predicted to alter poly(A) signal and splice site and was possibly disease causing.

Table S5-1 Evaluation of mutation in clone A8 allele 4 using SIFT. Insertion was predicted to cause frameshift but possibly not damaging or deleterious.

	Found in dbSNP?			PROVEAN Prediction			SIFT Prediction*		
	Total	Yes	No (novel)	Neutral	Deleterious	Not available	Tolerated	Damaging	Not available
1. Protein-coding	**1**	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**1**
(1) Single AA Change	0	0	0	0	0	0	0	0	0
(2) Synonymous	0	0	0	0	0	0	0	0	0
(3) Deletion	0	0	0	0	0	0	0	0	0
(4) Insertion	0	0	0	0	0	0	0	0	0
(5) Multiple AA Change	0	0	0	0	0	0	0	0	0
(6) Frameshift	1	0	1	0	0	1	0	0	1
(7) Nonsense	0	0	0	0	0	0	0	0	0
(8) Unknown	0	0	0	0	0	0	0	0	0
(9) Input error	0	0	0	0	0	0	0	0	0
2. Non protein-coding	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
3. Input format error	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**

a

Target Sequence	Mismatch	Number of Found Targets
GGGCAGCTACGGTTTCCGTCNGG	0	1
GGGCAGCTACGGTTTCCGTCNGG	3	5
GGGCAGCTACGGTTTCCGTCNGG	4	23
GGGCAGCTACGGTTTCCGTCNGG	5	356
GGGCAGCTACGGTTTCCGTCNGG	6	3332

b

Bulge Type	Target	Chromosome	Position	Direction	Mismatches	Bulge Size
X	crRNA: GGGCAGCTACGGTTTCCGTCNGG DNA: GcGCAGCTACGGTcTCCGcCCGG	chr3	9793024	-	3	0
X	crRNA: GGGCAGCTACGGTTTCCGTCNGG DNA: GGGCAGCaACtGTTTCCtTCGGG	chr3	10666288	+	3	0
X	crRNA: GGGCAGCTACGGTTTCCGTCNGG DNA: GGGCAGCTACGGTTTCttTtTGG	chr5	179595509	+	3	0
X	crRNA: GGGCAGCTACGGTTTCCGTCNGG DNA: GGGCAGCTACGGTTTCCGTCTGG	chr17	7676034	-	0	0
X	crRNA: GGGCAGCTACGGTTTCCGTCNGG DNA: GGGaAGCTACtGTaTCCGTCAGG	chr10	24516719	+	3	0
X	crRNA: GGGCAGCTACGGTTTCCGTCNGG DNA: GGGCAGaTACaGTTTCCtTCTGG	chr11	77137838	-	3	0

Figure S5-7 Off target sites identification with online tool Cas-OFFinder (Bae et al., 2014).

(a) No off-target sites with less than three nucleotide mismatches were identified and only five off-target sites were detected, indicating the high specificity of sgRNA. (b) Sequences and the chromosomal locations of five off-target sites with three nucleotide mismatches.

a

Off-Target Counts	All	Exonic	Intronic	Intergenic
Total	29	6	16	7
Summary	{0: 1, 1: 0, 2: 0, 3: 5, 4: 23}	{0:1, 1:0, 2:0, 3:0, 4:5}	{0:0, 1:0, 2:0, 3:5, 4:11}	{0:0, 1:0, 2:0, 3:0, 4:7}

b

WGE ID	Location	Sequence	Mismatches	Strand	Type
1143558245	Original CRISPR	GGGCAGCTACGGTTTCCGTC TGG (reversed)			Exonic
1143558245	17:7676035-7676057	GGGCAGCTACGGTTTCCGTC TGG	0	-	Exonic
1065321774	10:24516720-24516742	GGGAAGCTACTGTATCCGTC AGG	3	+	Intronic
1085520772	11:77137839-77137861	GGGCAGATACAGTTTCCTTC TGG	3	-	Intronic
950148130	3:10666289-10666311	GGGCAGCAACTGTTTCCTTC GGG	3	+	Intronic
950021853	3:9793025-9793047	GCGCAGCTACGGTCTCCGCC CGG	3	-	Intronic
1002493364	5:179595510-179595532	GGGCAGCTACGGTTTCTTTT TGG	3	+	Intronic

Figure S5-8 Off-target sites identification with WTSI Genome Editing tool.

(a) A total of 29 off-target sites were identified, 6, 16 and 7 were exonic, intronic and intergenic respectively. All sites have at least three nucleotide mismatches (b) Sequences and the chromosomal locations of five off-target sites with three nucleotide mismatches.

REQUEST PERMISSIONS

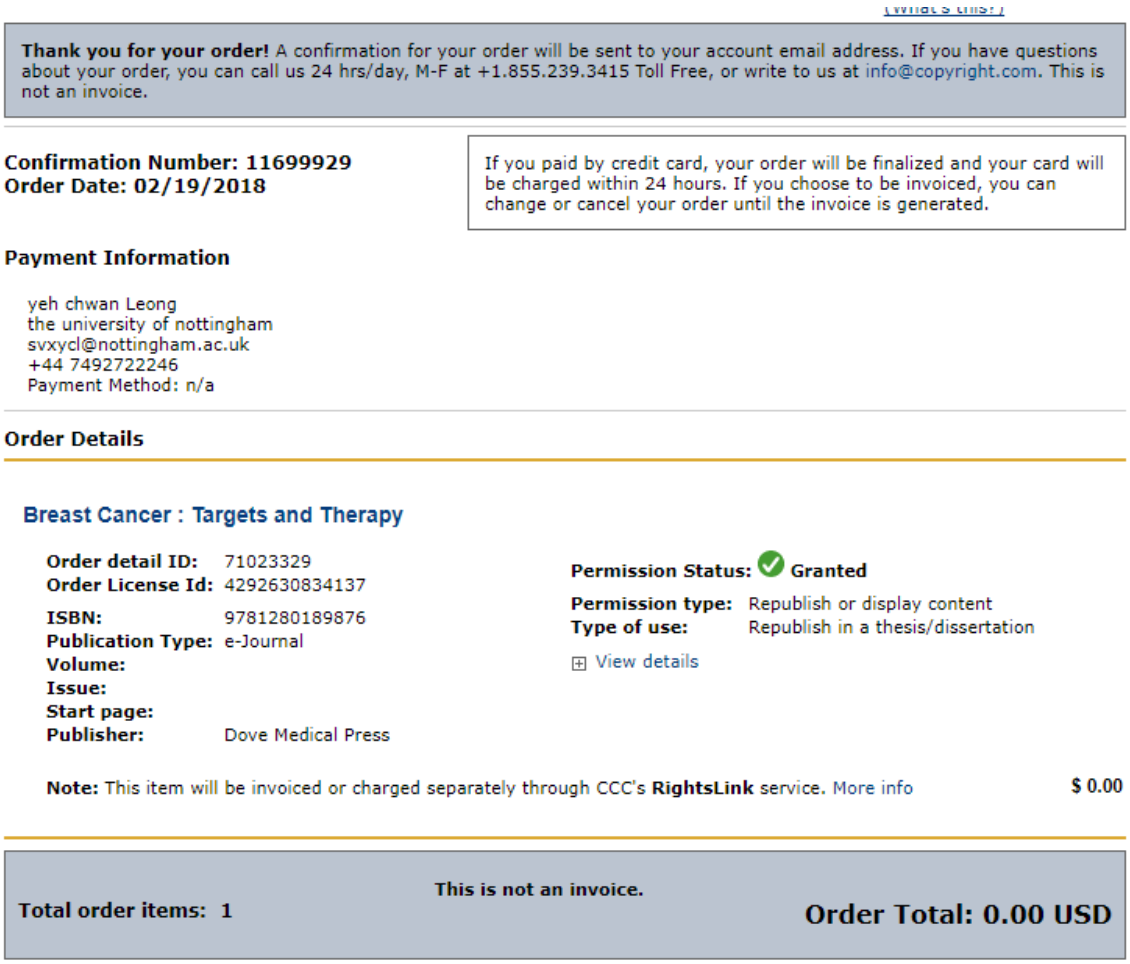
Figure 1.1

License Number	4292611399315
License date	Feb 19, 2018
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Cancer
Licensed Content Title	The molecular programme of tumour reversion: the steps beyond malignant transformation
Licensed Content Author	Adam Telerman, Robert Amson
Licensed Content Date	Jan 30, 2009
Licensed Content Volume	9
Licensed Content Issue	3
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Will you be translating?	no
Circulation/distribution	<501
Author of this Springer Nature content	no
Title	REPROGRAMMING TO CANCER INDUCED PLURIPOTENT STEM CELLS ELUCIDATES THE CONTRIBUTION OF GENETIC AND EPIGENETIC ALTERATIONS TO BREAST CARCINOGENESIS
Instructor name	Dr. Cinzia Allegrucci
Institution name	the university of nottingham
Expected presentation date	Apr 2018
Portions	Figure 1: Theoretical models of tumour reversion
Attachment	
Requestor Location	the university of nottingham Sutton Bonington Campus Sutton Bonington, Leicestershire LE12 5RD United Kingdom Attn: the university of nottingham
Billing Type	Invoice
Billing address	the university of nottingham Sutton Bonington Campus Sutton Bonington, United Kingdom LE12 5RD Attn: the university of nottingham
Total	0.00 GBP

Figure 1.2

License Number	4292621042501
License date	Feb 19, 2018
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature
Licensed Content Title	The cancer genome
Licensed Content Author	Michael R. Stratton, Peter J. Campbell, P. Andrew Futreal
Licensed Content Date	Apr 9, 2009
Licensed Content Volume	458
Licensed Content Issue	7239
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Will you be translating?	no
Circulation/distribution	<501
Author of this Springer Nature content	no
Title	REPROGRAMMING TO CANCER INDUCED PLURIPOTENT STEM CELLS ELUCIDATES THE CONTRIBUTION OF GENETIC AND EPIGENETIC ALTERATIONS TO BREAST CARCINOGENESIS
Instructor name	Dr. Cinzia Allegrucci
Institution name	the university of nottingham
Expected presentation date	Apr 2018
Portions	Figure 1. The lineage of mitotic cell divisions from the fertilized egg to a single cell within a cancer showing the timing of the somatic mutations acquired by the cancer cell and the processes that contribute to them
Attachment	
Requestor Location	the university of nottingham Sutton Bonington Campus Sutton Bonington, Leicestershire LE12 5RD United Kingdom Attn: the university of nottingham
Billing Type	Invoice
Billing address	the university of nottingham Sutton Bonington Campus Sutton Bonington, United Kingdom LE12 5RD Attn: the university of nottingham
Total	0.00 USD

Figure 1.7



(What's this?)

Thank you for your order! A confirmation for your order will be sent to your account email address. If you have questions about your order, you can call us 24 hrs/day, M-F at +1.855.239.3415 Toll Free, or write to us at info@copyright.com. This is not an invoice.

Confirmation Number: 11699929
Order Date: 02/19/2018

If you paid by credit card, your order will be finalized and your card will be charged within 24 hours. If you choose to be invoiced, you can change or cancel your order until the invoice is generated.

Payment Information

yeh chwan Leong
the university of nottingham
svxycl@nottingham.ac.uk
+44 7492722246
Payment Method: n/a

Order Details

Breast Cancer : Targets and Therapy

Order detail ID: 71023329
Order License Id: 4292630834137
ISBN: 9781280189876
Publication Type: e-Journal
Volume:
Issue:
Start page:
Publisher: Dove Medical Press

Permission Status: Granted
Permission type: Republish or display content
Type of use: Republish in a thesis/dissertation
View details

Note: This item will be invoiced or charged separately through CCC's **RightsLink** service. More info $ 0.00

Total order items: 1

This is not an invoice.

Order Total: 0.00 USD

Figure 1.8

License Number	4292631332964
License date	Feb 19, 2018
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Cell Research
Licensed Content Title	Embryonic stem cell and induced pluripotent stem cell: an epigenetic perspective
Licensed Content Author	Gaoyang Liang, Yi Zhang
Licensed Content Date	Dec 18, 2012
Licensed Content Volume	23
Licensed Content Issue	1
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Will you be translating?	no
Circulation/distribution	<501
Author of this Springer Nature content	no
Title	REPROGRAMMING TO CANCER INDUCED PLURIPOTENT STEM CELLS ELUCIDATES THE CONTRIBUTION OF GENETIC AND EPIGENETIC ALTERATIONS TO BREAST CARCINOGENESIS
Instructor name	Dr. Cinzia Allegrucci
Institution name	the university of nottingham
Expected presentation date	Apr 2018
Portions	Figure 1 The path from a somatic cell to a refined iPSC and the putative epigenetic barriers during the process
Attachment	
Requestor Location	the university of nottingham Sutton Bonington Campus Sutton Bonington, Leicestershire LE12 5RD United Kingdom Attn: the university of nottingham
Billing Type	Invoice
Billing address	the university of nottingham Sutton Bonington Campus Sutton Bonington, United Kingdom LE12 5RD Attn: the university of nottingham
Total	0.00 USD

Figure 1.9

License Number	4292631509455
License date	Feb 19, 2018
Licensed Content Publisher	Elsevier
Licensed Content Publication	Cell Stem Cell
Licensed Content Title	How Tets and Cytoskeleton Dynamics MET in Reprogramming
Licensed Content Author	Laurent David,Calley L. Hirsch
Licensed Content Date	Apr 3, 2014
Licensed Content Volume	14
Licensed Content Issue	4
Licensed Content Pages	2
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Original figure numbers	Figure 1 Regulation of the Mesenchymal-to-Epithelial Transition during the Initiation Phase of Reprogramming
Title of your thesis/dissertation	REPROGRAMMING TO CANCER INDUCED PLURIPOTENT STEM CELLS ELUCIDATES THE CONTRIBUTION OF GENETIC AND EPIGENETIC ALTERATIONS TO BREAST CARCINOGENESIS
Publisher of new work	the university of nottingham
Author of new work	Dr. Cinzia Allegrucci
Expected completion date	Apr 2018
Estimated size (number of pages)	1
Attachment	
Requestor Location	the university of nottingham Sutton Bonington Campus Sutton Bonington, Leicestershire LE12 5RD United Kingdom Attn: the university of nottingham
Publisher Tax ID	GB 494 6272 12
Total	0.00 USD

Figure 1.10

License Number	4292640171745
License date	Feb 19, 2018
Licensed Content Publisher	Elsevier
Licensed Content Publication	Cell
Licensed Content Title	Single-Cell Expression Analyses during Cellular Reprogramming Reveal an Early Stochastic and a Late Hierarchic Phase
Licensed Content Author	Yosef Buganim,Dina A. Faddah,Albert W. Cheng,Elena Itskovich,Styliani Markoulaki,Kibibi Ganz,Sandy L. Klemm,Alexander van Oudenaarden,Rudolf Jaenisch
Licensed Content Date	Sep 14, 2012
Licensed Content Volume	150
Licensed Content Issue	6
Licensed Content Pages	14
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Original figure numbers	Figure 7 Two Phases in Reprogramming
Title of your thesis/dissertation	REPROGRAMMING TO CANCER INDUCED PLURIPOTENT STEM CELLS ELUCIDATES THE CONTRIBUTION OF GENETIC AND EPIGENETIC ALTERATIONS TO BREAST CARCINOGENESIS
Publisher of new work	the university of nottingham
Author of new work	Dr. Cinzia Allegrucci
Expected completion date	Apr 2018
Estimated size (number of pages)	1
Attachment	
Requestor Location	the university of nottingham Sutton Bonington Campus Sutton Bonington, Leicestershire LE12 5RD United Kingdom Attn: the university of nottingham
Publisher Tax ID	GB 494 6272 12
Total	0.00 USD

Figure 1.11

License Number	4292640348079
License date	Feb 19, 2018
Licensed Content Publisher	Elsevier
Licensed Content Publication	Cell Stem Cell
Licensed Content Title	Molecular Control of Induced Pluripotency
Licensed Content Author	Thorold W. Theunissen,Rudolf Jaenisch
Licensed Content Date	Jun 5, 2014
Licensed Content Volume	14
Licensed Content Issue	6
Licensed Content Pages	15
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Original figure numbers	Figure 2 Models of Epigenetic Reprogramming
Title of your thesis/dissertation	REPROGRAMMING TO CANCER INDUCED PLURIPOTENT STEM CELLS ELUCIDATES THE CONTRIBUTION OF GENETIC AND EPIGENETIC ALTERATIONS TO BREAST CARCINOGENESIS
Publisher of new work	the university of nottingham
Author of new work	Dr. Cinzia Allegrucci
Expected completion date	Apr 2018
Estimated size (number of pages)	1
Attachment	
Requestor Location	the university of nottingham Sutton Bonington Campus Sutton Bonington, Leicestershire LE12 5RD United Kingdom Attn: the university of nottingham
Publisher Tax ID	GB 494 6272 12
Total	0.00 USD

Figure 7.1

License Number	4292990131089
License date	Feb 20, 2018
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Medicine
Licensed Content Title	Patient-derived induced pluripotent stem cells in cancer research and precision oncology
Licensed Content Author	Eirini P Papapetrou
Licensed Content Date	Dec 6, 2016
Licensed Content Volume	22
Licensed Content Issue	12
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Will you be translating?	no
Circulation/distribution	<501
Author of this Springer Nature content	no
Title	REPROGRAMMING TO CANCER INDUCED PLURIPOTENT STEM CELLS ELUCIDATES THE CONTRIBUTION OF GENETIC AND EPIGENETIC ALTERATIONS TO BREAST CARCINOGENESIS
Instructor name	Dr. Cinzia Allegrucci
Institution name	the university of nottingham
Expected presentation date	Apr 2018
Portions	Figure 1: An overview of iPSCs and cancer modeling.
Attachment	
Requestor Location	the university of nottingham Sutton Bonington Campus Sutton Bonington, Leicestershire LE12 5RD United Kingdom Attn: the university of nottingham
Billing Type	Invoice
Billing address	the university of nottingham Sutton Bonington Campus Sutton Bonington, United Kingdom LE12 5RD Attn: the university of nottingham
Total	0.00 USD

BIBLIOGRAPHY

AASEN, T. & IZPISUA BELMONTE, J. C. 2010. Isolation and cultivation of human keratinocytes from skin or plucked hair for the generation of induced pluripotent stem cells. *Nat Protoc,* 5**,** 371-82.

AASEN, T., RAYA, A., BARRERO, M. J., GARRETA, E., CONSIGLIO, A., GONZALEZ, F., VASSENA, R., BILIC, J., PEKARIK, V., TISCORNIA, G., EDEL, M., BOUE, S. & IZPISUA BELMONTE, J. C. 2008. Efficient and rapid generation of induced pluripotent stem cells from human keratinocytes. *Nat Biotechnol,* 26**,** 1276-84.

ABEYTA, M. J., CLARK, A. T., RODRIGUEZ, R. T., BODNAR, M. S., PERA, R. A. & FIRPO, M. T. 2004. Unique gene expression signatures of independently-derived human embryonic stem cell lines. *Hum Mol Genet,* 13**,** 601-8.

ABRAHAM, A. G. & O'NEILL, E. 2014. PI3K/Akt-mediated regulation of p53 in cancer. *Biochem Soc Trans,* 42**,** 798-803.

AHMED, D., EIDE, P. W., EILERTSEN, I. A., DANIELSEN, S. A., EKNAES, M., HEKTOEN, M., LIND, G. E. & LOTHE, R. A. 2013. Epigenetic and genetic features of 24 colon cancer cell lines. *Oncogenesis,* 2**,** e71.

AJANI, J. A. 2013. Challenges imposed by the complexity of cancer genome. *Lancet Oncol,* 14**,** e291-2.

AL-HAJJ, M. & CLARKE, M. F. 2004. Self-renewal and solid tumor stem cells. *Oncogene,* 23**,** 7274-82.

AL-HAJJ, M., WICHA, M. S., BENITO-HERNANDEZ, A., MORRISON, S. J. & CLARKE, M. F. 2003a. Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A,* 100**,** 3983-8.

AL-HAJJ, M., WICHA, M. S., BENITO-HERNANDEZ, A., MORRISON, S. J. & CLARKE, M. F. 2003b. Prospective identification of tumorigenic breast cancer cells (vol 100, pg 3983, 2003). *Proceedings of the National Academy of Sciences of the United States of America,* 100**,** 6890-6890.

AL-RAYYAN, N., LITCHFIELD, L. M., IVANOVA, M. M., RADDE, B. N., CHENG, A., ELBEDEWY, A. & KLINGE, C. M. 2014. 5-Aza-2-deoxycytidine and trichostatin A increase COUP-TFII expression in antiestrogen-resistant breast cancer cell lines. *Cancer Letters,* 347**,** 139-150.

ALBERICH-JORDA, M., WOUTERS, B., BALASTIK, M., SHAPIRO-KOSS, C., ZHANG, H., DI RUSCIO, A., RADOMSKA, H. S., EBRALIDZE, A. K., AMABILE, G., YE, M., ZHANG, J., LOWERS, I., AVELLINO, R., MELNICK, A., FIGUEROA, M. E., VALK, P. J., DELWEL, R. & TENEN, D. G. 2012. C/EBPgamma deregulation results in differentiation arrest in acute myeloid leukemia. *J Clin Invest,* 122**,** 4490-504.

ALEXANDROV, L. B., NIK-ZAINAL, S., WEDGE, D. C., APARICIO, S. A. J. R., BEHJATI, S., BIANKIN, A. V., BIGNELL, G. R., BOLLI, N., BORG, A., BORRESEN-DALE, A. L., BOYAULT, S., BURKHARDT, B., BUTLER, A. P., CALDAS, C., DAVIES, H. R., DESMEDT, C., EILS, R., EYFJORD, J. E., FOEKENS, J. A., GREAVES, M., HOSODA, F., HUTTER, B., ILICIC, T., IMBEAUD, S., IMIELINSK, M., JAGER, N., JONES, D. T. W., JONES, D., KNAPPSKOG, S., KOOL, M., LAKHANI, S. R., LOPEZ-OTIN, C., MARTIN, S., MUNSHI, N. C., NAKAMURA, H., NORTHCOTT, P. A., PAJIC, M., PAPAEMMANUIL, E., PARADISO, A., PEARSON, J. V., PUENTE, X. S., RAINE, K., RAMAKRISHNA, M., RICHARDSON, A. L., RICHTER, J., ROSENSTIEL, P., SCHLESNER, M., SCHUMACHER, T. N., SPAN, P. N., TEAGUE, J. W., TOTOKI, Y., TUTT, A. N. J., VALDES-MAS, R., VAN BUUREN, M. M., VAN 'T VEER, L., VINCENT-SALOMON, A., WADDELL, N., YATES, L. R., ZUCMAN-ROSSI, J., FUTREAL, P. A., MCDERMOTT, U., LICHTER, P., MEYERSON, M., GRIMMOND, S. M., SIEBERT, R., CAMPO, E., SHIBATA, T., PFISTER, S. M., CAMPBELL, P. J., STRATTON, M. R., GENOME,

A. P. C., CONSORTIUM, I. B. C., CONSORTIUM, I. M.-S. & PEDBRAIN, I. 2013. Signatures of mutational processes in human cancer. *Nature,* 500**,** 415-+.

ALI, H. R., RUEDA, O. M., CHIN, S. F., CURTIS, C., DUNNING, M. J., APARICIO, S. A. & CALDAS, C. 2014. Genome-driven integrated classification of breast cancer validated in over 7,500 samples. *Genome Biol,* 15**,** 431.

ALLDRIDGE, L., METODIEVA, G., GREENWOOD, C., AL-JANABI, K., THWAITES, L., SAUVEN, P. & METODIEV, M. 2008. Proteome profiling of breast tumors by gel electrophoresis and nanoscale electrospray ionization mass spectrometry. *J Proteome Res,* 7**,** 1458-69.

ALLEGRUCCI, C., RUSHTON, M. D., DIXON, J. E., SOTTILE, V., SHAH, M., KUMARI, R., WATSON, S., ALBERIO, R. & JOHNSON, A. D. 2011. Epigenetic reprogramming of breast cancer cells with oocyte extracts. *Mol Cancer,* 10**,** 7.

AMABILE, G., DI RUSCIO, A., MULLER, F., WELNER, R. S., YANG, H., EBRALIDZE, A. K., ZHANG, H., LEVANTINI, E., QI, L., MARTINELLI, G., BRUMMELKAMP, T., LE BEAU, M. M., FIGUEROA, M. E., BOCK, C. & TENEN, D. G. 2015. Dissecting the role of aberrant DNA methylation in human leukaemia. *Nat Commun,* 6**,** 7091.

AMBADY, S., MALCUIT, C., KASHPUR, O., KOLE, D., HOLMES, W. F., HEDBLOM, E., PAGE, R. L. & DOMINKO, T. 2010. Expression of NANOG and NANOGP8 in a variety of undifferentiated and differentiated human cells. *Int J Dev Biol,* 54**,** 1743-54.

ANDREWS, P. W., BANTING, G., DAMJANOV, I., ARNAUD, D. & AVNER, P. 1984. Three monoclonal antibodies defining distinct differentiation antigens associated with different high molecular weight polypeptides on the surface of human embryonal carcinoma cells. *Hybridoma,* 3**,** 347-61.

ANOKYE-DANSO, F., TRIVEDI, C. M., JUHR, D., GUPTA, M., CUI, Z., TIAN, Y., ZHANG, Y., YANG, W., GRUBER, P. J., EPSTEIN, J. A. & MORRISEY, E. E. 2011. Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. *Cell Stem Cell,* 8**,** 376-88.

APOSTOLOU, E., FERRARI, F., WALSH, R. M., BAR-NUR, O., STADTFELD, M., CHELOUFI, S., STUART, H. T., POLO, J. M., OHSUMI, T. K., BOROWSKY, M. L., KHARCHENKO, P. V., PARK, P. J. & HOCHEDLINGER, K. 2013. Genome-wide chromatin interactions of the Nanog locus in pluripotency, differentiation, and reprogramming. *Cell Stem Cell,* 12**,** 699-712.

APOSTOLOU, E. & HOCHEDLINGER, K. 2013. Chromatin dynamics during cellular reprogramming. *Nature,* 502**,** 462-71.

ASKANAZY, M. 1907. Die Teratome nach ihrem Bau, ihrem Verlauf, ihrer Genese und im Vergleich zum experimentellen Teratoid. *Verhandlungen der Deutschen Pathologischen Gesellschaft,* 11**,** 39-82.

ATCHLEY, D. P., ALBARRACIN, C. T., LOPEZ, A., VALERO, V., AMOS, C. I., GONZALEZ-ANGULO, A. M., HORTOBAGYI, G. N. & ARUN, B. K. 2008. Clinical and pathologic characteristics of patients with BRCA-positive and BRCA-negative breast cancer. *J Clin Oncol,* 26**,** 4282-8.

AU, K. F. & SEBASTIANO, V. 2014. The transcriptome of human pluripotent stem cells. *Curr Opin Genet Dev,* 28**,** 71-7.

BAAL, N., REISINGER, K., JAHR, H., BOHLE, R. M., LIANG, O., MUNSTEDT, K., RAO, C. V., PREISSNER, K. T. & ZYGMUNT, M. T. 2004. Expression of transcription factor Oct-4 and other embryonic genes in CD133 positive cells from human umbilical cord blood. *Thromb Haemost,* 92**,** 767-75.

BADENES, S. M., FERNANDES, T. G., CORDEIRO, C. S., BOUCHER, S., KUNINGER, D., VEMURI, M. C., DIOGO, M. M. & CABRAL, J. M. 2016. Defined Essential 8 Medium and Vitronectin Efficiently Support Scalable Xeno-Free Expansion of Human Induced

Pluripotent Stem Cells in Stirred Microcarrier Culture Systems. *PLoS One,* 11**,** e0151264.

BADVE, S. & NAKSHATRI, H. 2012. Breast-cancer stem cells-beyond semantics. *Lancet Oncol,* 13**,** e43-8.

BAE, S., PARK, J. & KIM, J. S. 2014. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics,* 30**,** 1473-5.

BAILLIE, R., TAN, S. T. & ITINTEANG, T. 2017. Cancer Stem Cells in Oral Cavity Squamous Cell Carcinoma: A Review. *Front Oncol,* 7**,** 112.

BANERJI, S., CIBULSKIS, K., RANGEL-ESCARENO, C., BROWN, K. K., CARTER, S. L., FREDERICK, A. M., LAWRENCE, M. S., SIVACHENKO, A. Y., SOUGNEZ, C., ZOU, L., CORTES, M. L., FERNANDEZ-LOPEZ, J. C., PENG, S., ARDLIE, K. G., AUCLAIR, D., BAUTISTA-PINA, V., DUKE, F., FRANCIS, J., JUNG, J., MAFFUZ-AZIZ, A., ONOFRIO, R. C., PARKIN, M., PHO, N. H., QUINTANAR-JURADO, V., RAMOS, A. H., REBOLLAR-VEGA, R., RODRIGUEZ-CUEVAS, S., ROMERO-CORDOBA, S. L., SCHUMACHER, S. E., STRANSKY, N., THOMPSON, K. M., URIBE-FIGUEROA, L., BASELGA, J., BEROUKHIM, R., POLYAK, K., SGROI, D. C., RICHARDSON, A. L., JIMENEZ-SANCHEZ, G., LANDER, E. S., GABRIEL, S. B., GARRAWAY, L. A., GOLUB, T. R., MELENDEZ-ZAJGLA, J., TOKER, A., GETZ, G., HIDALGO-MIRANDA, A. & MEYERSON, M. 2012. Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature,* 486**,** 405-9.

BANITO, A. & GIL, J. 2010. Induced pluripotent stem cells and senescence: learning the biology to improve the technology. *EMBO Rep,* 11**,** 353-9.

BANITO, A., RASHID, S. T., ACOSTA, J. C., LI, S., PEREIRA, C. F., GETI, I., PINHO, S., SILVA, J. C., AZUARA, V., WALSH, M., VALLIER, L. & GIL, J. 2009. Senescence impairs successful reprogramming to pluripotent stem cells. *Genes Dev,* 23**,** 2134-9.

BANNISTER, A. J. & KOUZARIDES, T. 2011. Regulation of chromatin by histone modifications. *Cell Res,* 21**,** 381-95.

BAO, S., WU, Q., MCLENDON, R. E., HAO, Y., SHI, Q., HJELMELAND, A. B., DEWHIRST, M. W., BIGNER, D. D. & RICH, J. N. 2006. Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature,* 444**,** 756-60.

BAR-NUR, O., RUSS, H. A., EFRAT, S. & BENVENISTY, N. 2011. Epigenetic memory and preferential lineage-specific differentiation in induced pluripotent stem cells derived from human pancreatic islet beta cells. *Cell Stem Cell,* 9**,** 17-23.

BAR, M., WYMAN, S. K., FRITZ, B. R., QI, J., GARG, K. S., PARKIN, R. K., KROH, E. M., BENDORAITE, A., MITCHELL, P. S., NELSON, A. M., RUZZO, W. L., WARE, C., RADICH, J. P., GENTLEMAN, R., RUOHOLA-BAKER, H. & TEWARI, M. 2008. MicroRNA discovery and profiling in human embryonic stem cells by deep sequencing of small RNA libraries. *Stem Cells,* 26**,** 2496-505.

BARNARD, J. A. & WARWICK, G. 1993. Butyrate rapidly induces growth inhibition and differentiation in HT-29 cells. *Cell Growth Differ,* 4**,** 495-501.

BARRANGOU, R., FREMAUX, C., DEVEAU, H., RICHARDS, M., BOYAVAL, P., MOINEAU, S., ROMERO, D. A. & HORVATH, P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science,* 315**,** 1709-12.

BARROSO-DELJESUS, A., ROMERO-LOPEZ, C., LUCENA-AGUILAR, G., MELEN, G. J., SANCHEZ, L., LIGERO, G., BERZAL-HERRANZ, A. & MENENDEZ, P. 2008. Embryonic stem cell-specific miR302-367 cluster: human gene structure and functional characterization of its core promoter. *Mol Cell Biol,* 28**,** 6609-19.

BASS, A. J., WATANABE, H., MERMEL, C. H., YU, S., PERNER, S., VERHAAK, R. G., KIM, S. Y., WARDWELL, L., TAMAYO, P., GAT-VIKS, I., RAMOS, A. H., WOO, M. S., WEIR, B. A., GETZ, G., BEROUKHIM, R., O'KELLY, M., DUTT, A., ROZENBLATT-ROSEN, O., DZIUNYCZ, P., KOMISAROF, J., CHIRIEAC, L. R., LAFARGUE, C. J., SCHEBLE, V., WILBERTZ, T., MA,

C., RAO, S., NAKAGAWA, H., STAIRS, D. B., LIN, L., GIORDANO, T. J., WAGNER, P., MINNA, J. D., GAZDAR, A. F., ZHU, C. Q., BROSE, M. S., CECCONELLO, I., RIBEIRO, U., JR., MARIE, S. K., DAHL, O., SHIVDASANI, R. A., TSAO, M. S., RUBIN, M. A., WONG, K. K., REGEV, A., HAHN, W. C., BEER, D. G., RUSTGI, A. K. & MEYERSON, M. 2009. SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. *Nat Genet,* 41**,** 1238-42.

BATISTA, P. J. & CHANG, H. Y. 2013. Long noncoding RNAs: cellular address codes in development and disease. *Cell,* 152**,** 1298-307.

BATLLE, E. & CLEVERS, H. 2017. Cancer stem cells revisited. *Nat Med,* 23**,** 1124-1134.

BAUMANN, P., THIELE, W., CREMERS, N., MUPPALA, S., KRACHULEC, J., DIEFENBACHER, M., KASSEL, O., MUDDULURU, G., ALLGAYER, H., FRAME, M. & SLEEMAN, J. P. 2012. CD24 interacts with and promotes the activity of c-src within lipid rafts in breast cancer cells, thereby increasing integrin-dependent adhesion. *Cell Mol Life Sci,* 69**,** 435-48.

BAYANI, J., YAO, C. Q., QUINTAYO, M. A., YAN, F., HAIDER, S., D'COSTA, A., BROOKES, C. L., VAN DE VELDE, C. J. H., HASENBURG, A., KIEBACK, D. G., MARKOPOULOS, C., DIRIX, L., SEYNAEVE, C., REA, D., BOUTROS, P. C. & BARTLETT, J. M. S. 2017. Molecular stratification of early breast cancer identifies drug targets to drive stratified medicine. *NPJ Breast Cancer,* 3**,** 3.

BAYART, E. & COHEN-HAGUENAUER, O. 2013. Technological overview of iPS induction from human adult somatic cells. *Curr Gene Ther,* 13**,** 73-92.

BEERS, J., GULBRANSON, D. R., GEORGE, N., SINISCALCHI, L. I., JONES, J., THOMSON, J. A. & CHEN, G. 2012. Passaging and colony expansion of human pluripotent stem cells by enzyme-free dissociation in chemically defined culture conditions. *Nat Protoc,* 7**,** 2029-40.

BEN-PORATH, I., THOMSON, M. W., CAREY, V. J., GE, R., BELL, G. W., REGEV, A. & WEINBERG, R. A. 2008. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat Genet,* 40**,** 499-507.

BERDASCO, M. & ESTELLER, M. 2010. Aberrant epigenetic landscape in cancer: how cellular identity goes awry. *Dev Cell,* 19**,** 698-711.

BIALK, P., SANSBURY, B., RIVERA-TORRES, N., BLOH, K., MAN, D. & KMIEC, E. B. 2016. Analyses of point mutation repair and allelic heterogeneity generated by CRISPR/Cas9 and single-stranded DNA oligonucleotides. *Sci Rep,* 6**,** 32681.

BILOUSOVA, G. & ROOP, D. R. 2014. Induced pluripotent stem cells in dermatology: potentials, advances, and limitations. *Cold Spring Harb Perspect Med,* 4**,** a015164.

BJERKVIG, R., TYSNES, B. B., ABOODY, K. S., NAJBAUER, J. & TERZIS, A. J. 2005. Opinion: the origin of the cancer stem cell: current controversies and new insights. *Nat Rev Cancer,* 5**,** 899-904.

BLANCAFORT, P., JIN, J. & FRYE, S. 2013. Writing and rewriting the epigenetic code of cancer cells: from engineered proteins to small molecules. *Mol Pharmacol,* 83**,** 563-76.

BLANPAIN, C. 2013. Tracing the cellular origin of cancer. *Nat Cell Biol,* 15**,** 126-34.

BLELLOCH, R., VENERE, M., YEN, J. & RAMALHO-SANTOS, M. 2007. Generation of induced pluripotent stem cells in the absence of drug selection. *Cell Stem Cell,* 1**,** 245-7.

BLELLOCH, R. H., HOCHEDLINGER, K., YAMADA, Y., BRENNAN, C., KIM, M., MINTZ, B., CHIN, L. & JAENISCH, R. 2004. Nuclear cloning of embryonal carcinoma cells. *Proc Natl Acad Sci U S A,* 101**,** 13985-90.

BLOWS, F. M., DRIVER, K. E., SCHMIDT, M. K., BROEKS, A., VAN LEEUWEN, F. E., WESSELING, J., CHEANG, M. C., GELMON, K., NIELSEN, T. O., BLOMQVIST, C., HEIKKILA, P., HEIKKINEN, T., NEVANLINNA, H., AKSLEN, L. A., BEGIN, L. R., FOULKES, W. D., COUCH, F. J., WANG, X., CAFOUREK, V., OLSON, J. E., BAGLIETTO, L., GILES, G. G., SEVERI, G., MCLEAN, C. A., SOUTHEY, M. C., RAKHA, E., GREEN, A. R., ELLIS, I. O., SHERMAN, M.

E., LISSOWSKA, J., ANDERSON, W. F., COX, A., CROSS, S. S., REED, M. W., PROVENZANO, E., DAWSON, S. J., DUNNING, A. M., HUMPHREYS, M., EASTON, D. F., GARCIA-CLOSAS, M., CALDAS, C., PHAROAH, P. D. & HUNTSMAN, D. 2010. Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. *PLoS Med,* 7**,** e1000279.

BOCK, C., KISKINIS, E., VERSTAPPEN, G., GU, H., BOULTING, G., SMITH, Z. D., ZILLER, M., CROFT, G. F., AMOROSO, M. W., OAKLEY, D. H., GNIRKE, A., EGGAN, K. & MEISSNER, A. 2011. Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell,* 144**,** 439-52.

BOFFA, L. C., VIDALI, G., MANN, R. S. & ALLFREY, V. G. 1978. Suppression of histone deacetylation in vivo and in vitro by sodium butyrate. *J Biol Chem,* 253**,** 3364-6.

BOLLAG, D. M., MCQUENEY, P. A., ZHU, J., HENSENS, O., KOUPAL, L., LIESCH, J., GOETZ, M., LAZARIDES, E. & WOODS, C. M. 1995. Epothilones, a new class of microtubule-stabilizing agents with a taxol-like mechanism of action. *Cancer Res,* 55**,** 2325-33.

BOOTH, H. A. & HOLLAND, P. W. 2004. Eleven daughters of NANOG. *Genomics,* 84**,** 229-38.

BOULTING, G. L., KISKINIS, E., CROFT, G. F., AMOROSO, M. W., OAKLEY, D. H., WAINGER, B. J., WILLIAMS, D. J., KAHLER, D. J., YAMAKI, M., DAVIDOW, L., RODOLFA, C. T., DIMOS, J. T., MIKKILINENI, S., MACDERMOTT, A. B., WOOLF, C. J., HENDERSON, C. E., WICHTERLE, H. & EGGAN, K. 2011. A functionally characterized test set of human induced pluripotent stem cells. *Nat Biotechnol,* 29**,** 279-86.

BOUMAHDI, S., DRIESSENS, G., LAPOUGE, G., RORIVE, S., NASSAR, D., LE MERCIER, M., DELATTE, B., CAAUWE, A., LENGLEZ, S., NKUSI, E., BROHEE, S., SALMON, I., DUBOIS, C., DEL MARMOL, V., FUKS, F., BECK, B. & BLANPAIN, C. 2014. SOX2 controls tumour initiation and cancer stem-cell functions in squamous-cell carcinoma. *Nature,* 511**,** 246-50.

BOYER, L. A., LEE, T. I., COLE, M. F., JOHNSTONE, S. E., LEVINE, S. S., ZUCKER, J. P., GUENTHER, M. G., KUMAR, R. M., MURRAY, H. L., JENNER, R. G., GIFFORD, D. K., MELTON, D. A., JAENISCH, R. & YOUNG, R. A. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell,* 122**,** 947-56.

BOZIC, I., ANTAL, T., OHTSUKI, H., CARTER, H., KIM, D., CHEN, S., KARCHIN, R., KINZLER, K. W., VOGELSTEIN, B. & NOWAK, M. A. 2010. Accumulation of driver and passenger mutations during tumor progression. *Proc Natl Acad Sci U S A,* 107**,** 18545-50.

BRAMBRINK, T., FOREMAN, R., WELSTEAD, G. G., LENGNER, C. J., WERNIG, M., SUH, H. & JAENISCH, R. 2008. Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell Stem Cell,* 2**,** 151-9.

BRENA, R. M. & COSTELLO, J. F. 2007. Genome-epigenome interactions in cancer. *Hum Mol Genet,* 16 Spec No 1**,** R96-105.

BRIVANLOU, A. H., GAGE, F. H., JAENISCH, R., JESSELL, T., MELTON, D. & ROSSANT, J. 2003. Stem cells. Setting standards for human embryonic stem cells. *Science,* 300**,** 913-6.

BROOKS, C. L. & GU, W. 2010. New insights into p53 activation. *Cell Res,* 20**,** 614-21.

BROWER, V. 2011. Epigenetics: Unravelling the cancer code. *Nature,* 471**,** S12-3.

BUGANIM, Y., FADDAH, D. A., CHENG, A. W., ITSKOVICH, E., MARKOULAKI, S., GANZ, K., KLEMM, S. L., VAN OUDENAARDEN, A. & JAENISCH, R. 2012. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell,* 150**,** 1209-22.

BUGANIM, Y., FADDAH, D. A. & JAENISCH, R. 2013. Mechanisms and models of somatic cell reprogramming. *Nat Rev Genet,* 14**,** 427-39.

BURKE, J. E. & WILLIAMS, R. L. 2015. Synergy in activating class I PI3Ks. *Trends Biochem Sci,* 40**,** 88-100.

BURNS, K. H. 2017. Transposable elements in cancer. *Nat Rev Cancer,* 17**,** 415-424.

BURRELL, R. A., MCGRANAHAN, N., BARTEK, J. & SWANTON, C. 2013. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature,* 501**,** 338-45.

BURRIDGE, P. W., THOMPSON, S., MILLROD, M. A., WEINBERG, S., YUAN, X., PETERS, A., MAHAIRAKI, V., KOLIATSOS, V. E., TUNG, L. & ZAMBIDIS, E. T. 2011. A universal system for highly efficient cardiac differentiation of human induced pluripotent stem cells that eliminates interline variability. *PLoS One,* 6**,** e18293.

BURROWS, C. K., BANOVICH, N. E., PAVLOVIC, B. J., PATTERSON, K., GALLEGO ROMERO, I., PRITCHARD, J. K. & GILAD, Y. 2016. Genetic Variation, Not Cell Type of Origin, Underlies the Majority of Identifiable Regulatory Differences in iPSCs. *PLoS Genet,* 12**,** e1005793.

BUSSOLATI, B., BRUNO, S., GRANGE, C., FERRANDO, U. & CAMUSSI, G. 2008. Identification of a tumor-initiating stem cell population in human renal carcinomas. *FASEB J,* 22**,** 3696-705.

BUTA, C., DAVID, R., DRESSEL, R., EMGARD, M., FUCHS, C., GROSS, U., HEALY, L., HESCHELER, J., KOLAR, R., MARTIN, U., MIKKERS, H., MULLER, F. J., SCHNEIDER, R. K., SEILER, A. E., SPIELMANN, H. & WEITZER, G. 2013. Reconsidering pluripotency tests: do we still need teratoma assays? *Stem Cell Res,* 11**,** 552-62.

BYERS, L. A., DIAO, L., WANG, J., SAINTIGNY, P., GIRARD, L., PEYTON, M., SHEN, L., FAN, Y., GIRI, U., TUMULA, P. K., NILSSON, M. B., GUDIKOTE, J., TRAN, H., CARDNELL, R. J., BEARSS, D. J., WARNER, S. L., FOULKS, J. M., KANNER, S. B., GANDHI, V., KRETT, N., ROSEN, S. T., KIM, E. S., HERBST, R. S., BLUMENSCHEIN, G. R., LEE, J. J., LIPPMAN, S. M., ANG, K. K., MILLS, G. B., HONG, W. K., WEINSTEIN, J. N., WISTUBA, II, COOMBES, K. R., MINNA, J. D. & HEYMACH, J. V. 2013. An epithelial-mesenchymal transition gene signature predicts resistance to EGFR and PI3K inhibitors and identifies Axl as a therapeutic target for overcoming EGFR inhibitor resistance. *Clin Cancer Res,* 19**,** 279-90.

BYLER, S., GOLDGAR, S., HEERBOTH, S., LEARY, M., HOUSMAN, G., MOULTON, K. & SARKAR, S. 2014. Genetic and epigenetic aspects of breast cancer progression and therapy. *Anticancer Res,* 34**,** 1071-7.

CALVET, C. Y., ANDRE, F. M. & MIR, L. M. 2014. The culture of cancer cell lines as tumorspheres does not systematically result in cancer stem cell enrichment. *PLoS One,* 9**,** e89644.

CANCER GENOME ATLAS, N. 2012. Comprehensive molecular portraits of human breast tumours. *Nature,* 490**,** 61-70.

CANCER GENOME ATLAS RESEARCH, N., WEINSTEIN, J. N., COLLISSON, E. A., MILLS, G. B., SHAW, K. R., OZENBERGER, B. A., ELLROTT, K., SHMULEVICH, I., SANDER, C. & STUART, J. M. 2013. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet,* 45**,** 1113-20.

CANTZ, T., KEY, G., BLEIDISSEL, M., GENTILE, L., HAN, D. W., BRENNE, A. & SCHOLER, H. R. 2008. Absence of OCT4 expression in somatic tumor cell lines. *Stem Cells,* 26**,** 692-7.

CAO, Z., LI, Y., CHEN, Z., WANG, H., ZHANG, M., ZHOU, N., WU, R., LING, Y., FANG, F., LI, N. & ZHANG, Y. 2015. Genome-Wide Dynamic Profiling of Histone Methylation during Nuclear Transfer-Mediated Porcine Somatic Cell Reprogramming. *PLoS One,* 10**,** e0144897.

CARD, D. A., HEBBAR, P. B., LI, L., TROTTER, K. W., KOMATSU, Y., MISHINA, Y. & ARCHER, T. K. 2008. Oct4/Sox2-regulated miR-302 targets cyclin D1 in human embryonic stem cells. *Mol Cell Biol,* 28**,** 6426-38.

CARETTE, J. E., PRUSZAK, J., VARADARAJAN, M., BLOMEN, V. A., GOKHALE, S., CAMARGO, F. D., WERNIG, M., JAENISCH, R. & BRUMMELKAMP, T. R. 2010. Generation of iPSCs from cultured human malignant cells. *Blood,* 115**,** 4039-42.

CAREY, B. W., MARKOULAKI, S., HANNA, J., SAHA, K., GAO, Q., MITALIPOVA, M. & JAENISCH, R. 2009. Reprogramming of murine and human somatic cells using a single polycistronic vector. *Proc Natl Acad Sci U S A,* 106**,** 157-62.

CAREY, B. W., MARKOULAKI, S., HANNA, J. H., FADDAH, D. A., BUGANIM, Y., KIM, J., GANZ, K., STEINE, E. J., CASSADY, J. P., CREYGHTON, M. P., WELSTEAD, G. G., GAO, Q. & JAENISCH, R. 2011. Reprogramming factor stoichiometry influences the epigenetic state and biological properties of induced pluripotent stem cells. *Cell Stem Cell,* 9**,** 588-98.

CAREY, L. A., DEES, E. C., SAWYER, L., GATTI, L., MOORE, D. T., COLLICHIO, F., OLLILA, D. W., SARTOR, C. I., GRAHAM, M. L. & PEROU, C. M. 2007. The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. *Clin Cancer Res,* 13**,** 2329-34.

CECH, T. R. & STEITZ, J. A. 2014. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell,* 157**,** 77-94.

CHAN, E. M., RATANASIRINTRAWOOT, S., PARK, I. H., MANOS, P. D., LOH, Y. H., HUO, H., MILLER, J. D., HARTUNG, O., RHO, J., INCE, T. A., DALEY, G. Q. & SCHLAEGER, T. M. 2009. Live cell imaging distinguishes bona fide human iPS cells from partially reprogrammed cells. *Nat Biotechnol,* 27**,** 1033-7.

CHANG, M. T., ASTHANA, S., GAO, S. P., LEE, B. H., CHAPMAN, J. S., KANDOTH, C., GAO, J., SOCCI, N. D., SOLIT, D. B., OLSHEN, A. B., SCHULTZ, N. & TAYLOR, B. S. 2016a. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nat Biotechnol,* 34**,** 155-63.

CHANG, M. T., ASTHANA, S., GAO, S. P., LEE, B. H., CHAPMAN, J. S., KANDOTH, C., GAO, J. J., SOCCI, N. D., SOLIT, D. B., OLSHEN, A. B., SCHULTZ, N. & TAYLOR, B. S. 2016b. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nature Biotechnology,* 34**,** 155-+.

CHANG, T. C., YU, D., LEE, Y. S., WENTZEL, E. A., ARKING, D. E., WEST, K. M., DANG, C. V., THOMAS-TIKHONENKO, A. & MENDELL, J. T. 2008. Widespread microRNA repression by Myc contributes to tumorigenesis. *Nat Genet,* 40**,** 43-50.

CHARAFE-JAUFFRET, E., GINESTIER, C., IOVINO, F., WICINSKI, J., CERVERA, N., FINETTI, P., HUR, M. H., DIEBEL, M. E., MONVILLE, F., DUTCHER, J., BROWN, M., VIENS, P., XERRI, L., BERTUCCI, F., STASSI, G., DONTU, G., BIRNBAUM, D. & WICHA, M. S. 2009. Breast cancer cell lines contain functional cancer stem cells with metastatic capacity and a distinct molecular signature. *Cancer Res,* 69**,** 1302-13.

CHAUDHRY, M. A., VITALIS, T. Z., BOWEN, B. D. & PIRET, J. M. 2008. Basal medium composition and serum or serum replacement concentration influences on the maintenance of murine embryonic stem cells. *Cytotechnology,* 58**,** 173-9.

CHEANG, M. C., CHIA, S. K., VODUC, D., GAO, D., LEUNG, S., SNIDER, J., WATSON, M., DAVIES, S., BERNARD, P. S., PARKER, J. S., PEROU, C. M., ELLIS, M. J. & NIELSEN, T. O. 2009. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. *J Natl Cancer Inst,* 101**,** 736-50.

CHEBOUTI, I., BLASSL, C., WIMBERGER, P., NEUBAUER, H., FEHM, T., KIMMIG, R. & KASIMIR-BAUER, S. 2016. Analysis of disseminated tumor cells before and after platinum based chemotherapy in primary ovarian cancer. Do stem cell like cells predict prognosis? *Oncotarget,* 7**,** 26454-64.

CHEN, D., SUN, Y., WEI, Y., ZHANG, P., REZAEIAN, A. H., TERUYA-FELDSTEIN, J., GUPTA, S., LIANG, H., LIN, H. K., HUNG, M. C. & MA, L. 2012. LIFR is a breast cancer metastasis suppressor upstream of the Hippo-YAP pathway and a prognostic marker. *Nat Med,* 18**,** 1511-7.

CHEN, G., GULBRANSON, D. R., HOU, Z., BOLIN, J. M., RUOTTI, V., PROBASCO, M. D., SMUGA-OTTO, K., HOWDEN, S. E., DIOL, N. R., PROPSON, N. E., WAGNER, R., LEE, G. O.,

ANTOSIEWICZ-BOURGET, J., TENG, J. M. & THOMSON, J. A. 2011a. Chemically defined conditions for human iPSC derivation and culture. *Nat Methods,* 8**,** 424-9.

CHEN, G., HOU, Z., GULBRANSON, D. R. & THOMSON, J. A. 2010a. Actin-myosin contractility is responsible for the reduced viability of dissociated human embryonic stem cells. *Cell Stem Cell,* 7**,** 240-8.

CHEN, G. K., GULBRANSON, D. R., HOU, Z. G., BOLIN, J. M., RUOTTI, V., PROBASCO, M. D., SMUGA-OTTO, K., HOWDEN, S. E., DIOL, N. R., PROPSON, N. E., WAGNER, R., LEE, G. O., ANTOSIEWICZ-BOURGET, J., TENG, J. M. C. & THOMSON, J. A. 2011b. Chemically defined conditions for human iPSC derivation and culture. *Nature Methods,* 8**,** 424-U76.

CHEN, J., CHEN, X., LI, M., LIU, X., GAO, Y., KOU, X., ZHAO, Y., ZHENG, W., ZHANG, X., HUO, Y., CHEN, C., WU, Y., WANG, H., JIANG, C. & GAO, S. 2016. Hierarchical Oct4 Binding in Concert with Primed Epigenetic Rearrangements during Somatic Cell Reprogramming. *Cell Rep,* 14**,** 1540-1554.

CHEN, K. G., MALLON, B. S., MCKAY, R. D. & ROBEY, P. G. 2014. Human pluripotent stem cell culture: considerations for maintenance, expansion, and therapeutics. *Cell Stem Cell,* 14**,** 13-26.

CHEN, T., YUAN, D., WEI, B., JIANG, J., KANG, J., LING, K., GU, Y., LI, J., XIAO, L. & PEI, G. 2010b. E-cadherin-mediated cell-cell contact is critical for induced pluripotent stem cell generation. *Stem Cells,* 28**,** 1315-25.

CHEN, Y. C., HSU, H. S., CHEN, Y. W., TSAI, T. H., HOW, C. K., WANG, C. Y., HUNG, S. C., CHANG, Y. L., TSAI, M. L., LEE, Y. Y., KU, H. H. & CHIOU, S. H. 2008. Oct-4 expression maintained cancer stem-like properties in lung cancer-derived CD133-positive cells. *PLoS One,* 3**,** e2637.

CHENG, L., HANSEN, N. F., ZHAO, L., DU, Y., ZOU, C., DONOVAN, F. X., CHOU, B. K., ZHOU, G., LI, S., DOWEY, S. N., YE, Z., PROGRAM, N. C. S., CHANDRASEKHARAPPA, S. C., YANG, H., MULLIKIN, J. C. & LIU, P. P. 2012. Low incidence of DNA sequence variation in human induced pluripotent stem cells generated by nonintegrating plasmid expression. *Cell Stem Cell,* 10**,** 337-44.

CHIK, F. & SZYF, M. 2011. Effects of specific DNMT gene depletion on cancer cell transformation and breast cancer cell invasion; toward selective DNMT inhibitors. *Carcinogenesis,* 32**,** 224-32.

CHIN, A. C., PADMANABHAN, J., OH, S. K. & CHOO, A. B. 2010. Defined and serum-free media support undifferentiated human embryonic stem cell growth. *Stem Cells Dev,* 19**,** 753-61.

CHIN, L. & GRAY, J. W. 2008. Translating insights from the cancer genome into clinical practice. *Nature,* 452**,** 553-63.

CHIN, L., HAHN, W. C., GETZ, G. & MEYERSON, M. 2011. Making sense of cancer genomic data. *Genes Dev,* 25**,** 534-55.

CHIN, M. H., MASON, M. J., XIE, W., VOLINIA, S., SINGER, M., PETERSON, C., AMBARTSUMYAN, G., AIMIUWU, O., RICHTER, L., ZHANG, J., KHVOROSTOV, I., OTT, V., GRUNSTEIN, M., LAVON, N., BENVENISTY, N., CROCE, C. M., CLARK, A. T., BAXTER, T., PYLE, A. D., TEITELL, M. A., PELEGRINI, M., PLATH, K. & LOWRY, W. E. 2009. Induced Pluripotent Stem Cells and Embryonic Stem Cells Are Distinguished by Gene Expression Signatures. *Cell Stem Cell,* 5**,** 111-123.

CHIOU, S. H., WANG, M. L., CHOU, Y. T., CHEN, C. J., HONG, C. F., HSIEH, W. J., CHANG, H. T., CHEN, Y. S., LIN, T. W., HSU, H. S. & WU, C. W. 2010. Coexpression of Oct4 and Nanog enhances malignancy in lung adenocarcinoma by inducing cancer stem cell-like properties and epithelial-mesenchymal transdifferentiation. *Cancer Res,* 70**,** 10433-44.

CHOONG, P. F., TEH, H. X., TEOH, H. K., ONG, H. K., CHOO, K. B., SUGII, S., CHEONG, S. K. & KAMARUL, T. 2014. Heterogeneity of osteosarcoma cell lines led to variable responses in reprogramming. *Int J Med Sci,* 11**,** 1154-60.

CHOU, B. K., MALI, P., HUANG, X., YE, Z., DOWEY, S. N., RESAR, L. M., ZOU, C., ZHANG, Y. A., TONG, J. & CHENG, L. 2011. Efficient human iPS cell derivation by a non-integrating plasmid from blood cells with unique epigenetic and gene expression signatures. *Cell Res,* 21**,** 518-29.

CHRONIS, C., FIZIEV, P., PAPP, B., BUTZ, S., BONORA, G., SABRI, S., ERNST, J. & PLATH, K. 2017. Cooperative Binding of Transcription Factors Orchestrates Reprogramming. *Cell,* 168**,** 442-459 e20.

CHUANG, J. C. & JONES, P. A. 2007. Epigenetics and microRNAs. *Pediatr Res,* 61**,** 24R-29R.

CIRIELLO, G., MILLER, M. L., AKSOY, B. A., SENBABAOGLU, Y., SCHULTZ, N. & SANDER, C. 2013. Emerging landscape of oncogenic signatures across human cancers. *Nat Genet,* 45**,** 1127-33.

CONG, L., RAN, F. A., COX, D., LIN, S., BARRETTO, R., HABIB, N., HSU, P. D., WU, X., JIANG, W., MARRAFFINI, L. A. & ZHANG, F. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science,* 339**,** 819-23.

CONNOLLY, R. M., JANKOWITZ, R. C., ANDREOPOULOU, E., ALLRED, J. B., JETER, S. C., ADAM, B. M., ESPINOZA-DELGADO, I., BAYLIN, S. B., ZAHNOW, C. A., AHUJA, N., DAVIDSON, N. E. & STEARNS, V. 2011. A Phase 2 Study Investigating the Safety, Efficacy and Surrogate Biomarkers of Response of 5-Azacitidine (5-AZA) and Entinostat (MS-275) in Patients with Advanced Breast Cancer. *Cancer Research,* 71.

CONNOLLY, R. M., LI, H., JANKOWITZ, R. C., ZHANG, Z., RUDEK, M. A., JETER, S. C., SLATER, S. A., POWERS, P., WOLFF, A. C., FETTING, J. H., BRUFSKY, A., PIEKARZ, R., AHUJA, N., LAIRD, P. W., SHEN, H., WEISENBERGER, D. J., COPE, L., HERMAN, J. G., SOMLO, G., GARCIA, A. A., JONES, P. A., BAYLIN, S. B., DAVIDSON, N. E., ZAHNOW, C. A. & STEARNS, V. 2017. Combination Epigenetic Therapy in Advanced Breast Cancer with 5-Azacitidine and Entinostat: A Phase II National Cancer Institute/Stand Up to Cancer Study. *Clin Cancer Res,* 23**,** 2691-2701.

CONSORTIUM, E. P. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature,* 489**,** 57-74.

CORDENONSI, M., ZANCONATO, F., AZZOLIN, L., FORCATO, M., ROSATO, A., FRASSON, C., INUI, M., MONTAGNER, M., PARENTI, A. R., POLETTI, A., DAIDONE, M. G., DUPONT, S., BASSO, G., BICCIATO, S. & PICCOLO, S. 2011. The Hippo transducer TAZ confers cancer stem cell-related traits on breast cancer cells. *Cell,* 147**,** 759-72.

COROMINAS-FAJA, B., CUFI, S., OLIVERAS-FERRAROS, C., CUYAS, E., LOPEZ-BONET, E., LUPU, R., ALARCON, T., VELLON, L., IGLESIAS, J. M., LEIS, O., MARTIN, A. G., VAZQUEZ-MARTIN, A. & MENENDEZ, J. A. 2013. Nuclear reprogramming of luminal-like breast cancer cells generates Sox2-overexpressing cancer stem-like cellular states harboring transcriptional activation of the mTOR pathway. *Cell Cycle,* 12**,** 3109-24.

COUCH, F. J., KUCHENBAECKER, K. B., MICHAILIDOU, K., MENDOZA-FANDINO, G. A., NORD, S., LILYQUIST, J., OLSWOLD, C., HALLBERG, E., AGATA, S., AHSAN, H., AITTOMAKI, K., AMBROSONE, C., ANDRULIS, I. L., ANTON-CULVER, H., ARNDT, V., ARUN, B. K., ARVER, B., BARILE, M., BARKARDOTTIR, R. B., BARROWDALE, D., BECKMANN, L., BECKMANN, M. W., BENITEZ, J., BLANK, S. V., BLOMQVIST, C., BOGDANOVA, N. V., BOJESEN, S. E., BOLLA, M. K., BONANNI, B., BRAUCH, H., BRENNER, H., BURWINKEL, B., BUYS, S. S., CALDES, T., CALIGO, M. A., CANZIAN, F., CARPENTER, J., CHANG-CLAUDE, J., CHANOCK, S. J., CHUNG, W. K., CLAES, K. B., COX, A., CROSS, S. S., CUNNINGHAM, J. M., CZENE, K., DALY, M. B., DAMIOLA, F., DARABI, H., DE LA HOYA, M., DEVILEE, P., DIEZ, O., DING, Y. C., DOLCETTI, R., DOMCHEK, S. M., DORFLING, C. M., DOS-SANTOS-SILVA, I., DUMONT, M., DUNNING, A. M., ECCLES, D. M., EHRENCRONA, H., EKICI, A.

B., ELIASSEN, H., ELLIS, S., FASCHING, P. A., FIGUEROA, J., FLESCH-JANYS, D., FORSTI, A., FOSTIRA, F., FOULKES, W. D., FRIEBEL, T., FRIEDMAN, E., FROST, D., GABRIELSON, M., GAMMON, M. D., GANZ, P. A., GAPSTUR, S. M., GARBER, J., GAUDET, M. M., GAYTHER, S. A., GERDES, A. M., GHOUSSAINI, M., GILES, G. G., GLENDON, G., GODWIN, A. K., GOLDBERG, M. S., GOLDGAR, D. E., GONZALEZ-NEIRA, A., GREENE, M. H., GRONWALD, J., GUENEL, P., GUNTER, M., HAEBERLE, L., HAIMAN, C. A., HAMANN, U., HANSEN, T. V., HART, S., HEALEY, S., HEIKKINEN, T., HENDERSON, B. E., HERZOG, J., et al. 2016. Identification of four novel susceptibility loci for oestrogen receptor negative breast cancer. *Nat Commun,* 7**,** 11375.

COWELL, J. K., LADUCA, J., ROSSI, M. R., BURKHARDT, T., NOWAK, N. J. & MATSUI, S. 2005. Molecular characterization of the t(3;9) associated with immortalization in the MCF10A cell line. *Cancer Genet Cytogenet,* 163**,** 23-9.

CRISCIONE, S. W., ZHANG, Y., THOMPSON, W., SEDIVY, J. M. & NERETTI, N. 2014. Transcriptional landscape of repetitive elements in normal and cancer human cells. *BMC Genomics,* 15**,** 583.

CROESSMANN, S., WONG, H. Y., ZABRANSKY, D. J., CHU, D., ROSEN, D. M., CIDADO, J., COCHRAN, R. L., DALTON, W. B., ERLANGER, B., CRAVERO, K., BUTTON, B., KYKER-SNOWMAN, K., HURLEY, P. J., LAURING, J. & PARK, B. H. 2017. PIK3CA mutations and TP53 alterations cooperate to increase cancerous phenotypes and tumor heterogeneity. *Breast Cancer Res Treat,* 162**,** 451-464.

CURTIS, C., SHAH, S. P., CHIN, S. F., TURASHVILI, G., RUEDA, O. M., DUNNING, M. J., SPEED, D., LYNCH, A. G., SAMARAJIWA, S., YUAN, Y., GRAF, S., HA, G., HAFFARI, G., BASHASHATI, A., RUSSELL, R., MCKINNEY, S., GROUP, M., LANGEROD, A., GREEN, A., PROVENZANO, E., WISHART, G., PINDER, S., WATSON, P., MARKOWETZ, F., MURPHY, L., ELLIS, I., PURUSHOTHAM, A., BORRESEN-DALE, A. L., BRENTON, J. D., TAVARE, S., CALDAS, C. & APARICIO, S. 2012. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature,* 486**,** 346-52.

D'ANTONIO, M., WOODRUFF, G., NATHANSON, J. L., D'ANTONIO-CHRONOWSKA, A., ARIAS, A., MATSUI, H., WILLIAMS, R., HERRERA, C., REYNA, S. M., YEO, G. W., GOLDSTEIN, L. S., PANOPOULOS, A. D. & FRAZER, K. A. 2017a. High-Throughput and Cost-Effective Characterization of Induced Pluripotent Stem Cells. *Stem Cell Reports*.

D'ANTONIO, M., WOODRUFF, G., NATHANSON, J. L., D'ANTONIO-CHRONOWSKA, A., ARIAS, A., MATSUI, H., WILLIAMS, R., HERRERA, C., REYNA, S. M., YEO, G. W., GOLDSTEIN, L. S. B., PANOPOULOS, A. D. & FRAZER, K. A. 2017b. High-Throughput and Cost-Effective Characterization of Induced Pluripotent Stem Cells. *Stem Cell Reports,* 8**,** 1101-1111.

DAHMEN, R. P., KOCH, A., DENKHAUS, D., TONN, J. C., SORENSEN, N., BERTHOLD, F., BEHRENS, J., BIRCHMEIER, W., WIESTLER, O. D. & PIETSCH, T. 2001. Deletions of AXIN1, a component of the WNT/wingless pathway, in sporadic medulloblastomas. *Cancer Res,* 61**,** 7039-43.

DALBY, B., CATES, S., HARRIS, A., OHKI, E. C., TILKINS, M. L., PRICE, P. J. & CICCARONE, V. C. 2004. Advanced transfection with Lipofectamine 2000 reagent: primary neurons, siRNA, and high-throughput applications. *Methods,* 33**,** 95-103.

DANG, C. V. 2012. MYC on the path to cancer. *Cell,* 149**,** 22-35.

DASKALOS, A., NIKOLAIDIS, G., XINARIANOS, G., SAVVARI, P., CASSIDY, A., ZAKOPOULOU, R., KOTSINAS, A., GORGOULIS, V., FIELD, J. K. & LILOGLOU, T. 2009. Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer. *Int J Cancer,* 124**,** 81-7.

DAVID, L. & HIRSCH, C. L. 2014. How Tets and Cytoskeleton Dynamics MET in Reprogramming. *Cell Stem Cell,* 14**,** 417-418.

DAVIS, L. & MAIZELS, N. 2014. Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. *Proc Natl Acad Sci U S A,* 111**,** E924-32.

DAWSON, M. A. & KOUZARIDES, T. 2012. Cancer epigenetics: from mechanism to therapy. *Cell,* 150**,** 12-27.

DE LOS ANGELES, A., FERRARI, F., XI, R., FUJIWARA, Y., BENVENISTY, N., DENG, H., HOCHEDLINGER, K., JAENISCH, R., LEE, S., LEITCH, H. G., LENSCH, M. W., LUJAN, E., PEI, D., ROSSANT, J., WERNIG, M., PARK, P. J. & DALEY, G. Q. 2015. Hallmarks of pluripotency. *Nature,* 525**,** 469-78.

DE LOS ANGELES, A., FERRARI, F., XI, R., FUJIWARA, Y., BENVENISTY, N., DENG, H., HOCHEDLINGER, K., JAENISCH, R., LEE, S., LEITCH, H. G., LENSCH, M. W., LUJAN, E., PEI, D., ROSSANT, J., WERNIG, M., PARK, P. J. & DALEY, G. Q. 2016. Corrigendum: Hallmarks of pluripotency. *Nature,* 531**,** 400.

DE SOUSA, P. A., STEEG, R., WACHTER, E., BRUCE, K., KING, J., HOEVE, M., KHADUN, S., MCCONNACHIE, G., HOLDER, J., KURTZ, A., SELTMANN, S., DEWENDER, J., REIMANN, S., STACEY, G., O'SHEA, O., CHAPMAN, C., HEALY, L., ZIMMERMANN, H., BOLTON, B., RAWAT, T., ATKIN, I., VEIGA, A., KUEBLER, B., SERANO, B. M., SARIC, T., HESCHELER, J., BRUSTLE, O., PEITZ, M., THIELE, C., GEIJSEN, N., HOLST, B., CLAUSEN, C., LAKO, M., ARMSTRONG, L., GUPTA, S. K., KVIST, A. J., HICKS, R., JONEBRING, A., BROLEN, G., EBNETH, A., CABRERA-SOCORRO, A., FOERCH, P., GERAERTS, M., STUMMANN, T. C., HARMON, S., GEORGE, C., STREETER, I., CLARKE, L., PARKINSON, H., HARRISON, P. W., FAULCONBRIDGE, A., CHERUBIN, L., BURDETT, T., TRIGUEROS, C., PATEL, M. J., LUCAS, C., HARDY, B., PREDAN, R., DOKLER, J., BRAJNIK, M., KEMINER, O., PLESS, O., GRIBBON, P., CLAUSSEN, C., RINGWALD, A., KREISEL, B., COURTNEY, A. & ALLSOPP, T. E. 2017. Rapid establishment of the European Bank for induced Pluripotent Stem Cells (EBiSC) - the Hot Start experience. *Stem Cell Res,* 20**,** 105-114.

DEAN, M., FOJO, T. & BATES, S. 2005. Tumour stem cells and drug resistance. *Nat Rev Cancer,* 5**,** 275-84.

DEATON, A. M. & BIRD, A. 2011. CpG islands and the regulation of transcription. *Genes Dev,* 25**,** 1010-22.

DEHAIRS, J., TALEBI, A., CHERIFI, Y. & SWINNEN, J. V. 2016. CRISP-ID: decoding CRISPR mediated indels by Sanger sequencing. *Sci Rep,* 6**,** 28973.

DENG, J., SHOEMAKER, R., XIE, B., GORE, A., LEPROUST, E. M., ANTOSIEWICZ-BOURGET, J., EGLI, D., MAHERALI, N., PARK, I. H., YU, J. Y., DALEY, G. Q., EGGAN, K., HOCHEDLINGER, K., THOMSON, J., WANG, W., GAO, Y. & ZHANG, K. 2009. Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nature Biotechnology,* 27**,** 353-360.

DEWI, D., ISHII, H., HARAGUCHI, N., NISHIKAWA, S., KANO, Y., FUKUSUMI, T., OHTA, K., MIYAZAKI, S., OZAKI, M., SAKAI, D., SATOH, T., NAGANO, H., DOKI, Y. & MORI, M. 2012. Reprogramming of gastrointestinal cancer cells. *Cancer Sci,* 103**,** 393-9.

DI MASI, A., LEBOFFE, L., DE MARINIS, E., PAGANO, F., CICCONI, L., ROCHETTE-EGLY, C., LO-COCO, F., ASCENZI, P. & NERVI, C. 2015. Retinoic acid receptors: from molecular mechanisms to cancer therapy. *Mol Aspects Med,* 41**,** 1-115.

DI NICOLANTONIO, F., ARENA, S., GALLICCHIO, M., ZECCHIN, D., MARTINI, M., FLONTA, S. E., STELLA, G. M., LAMBA, S., CANCELLIERE, C., RUSSO, M., GEUNA, M., APPENDINO, G., FANTOZZI, R., MEDICO, E. & BARDELLI, A. 2008. Replacement of normal with mutant alleles in the genome of normal human cells unveils mutation-specific drug responses. *Proc Natl Acad Sci U S A,* 105**,** 20864-9.

DI STEFANO, B., SARDINA, J. L., VAN OEVELEN, C., COLLOMBET, S., KALLIN, E. M., VICENT, G. P., LU, J., THIEFFRY, D., BEATO, M. & GRAF, T. 2014. C/EBPalpha poises B cells for rapid reprogramming into induced pluripotent stem cells. *Nature,* 506**,** 235-9.

DOI, A., PARK, I. H., WEN, B., MURAKAMI, P., ARYEE, M. J., IRIZARRY, R., HERB, B., LADD-ACOSTA, C., RHO, J. S., LOEWER, S., MILLER, J., SCHLAEGER, T., DALEY, G. Q. & FEINBERG, A. P. 2009. Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nature Genetics,* 41**,** 1350-U123.

DOWNING, T. L., SOTO, J., MOREZ, C., HOUSSIN, T., FRITZ, A., YUAN, F., CHU, J., PATEL, S., SCHAFFER, D. V. & LI, S. 2013. Biophysical regulation of epigenetic state and cell reprogramming. *Nat Mater,* 12**,** 1154-62.

DOWSETT, M., SESTAK, I., LOPEZ-KNOWLES, E., SIDHU, K., DUNBIER, A. K., COWENS, J. W., FERREE, S., STORHOFF, J., SCHAPER, C. & CUZICK, J. 2013. Comparison of PAM50 risk of recurrence score with oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy. *J Clin Oncol,* 31**,** 2783-90.

DRESSEL, R. 2011. Effects of histocompatibility and host immune responses on the tumorigenicity of pluripotent stem cells. *Seminars in Immunopathology,* 33**,** 573-591.

DROST, J., VAN JAARSVELD, R. H., PONSIOEN, B., ZIMBERLIN, C., VAN BOXTEL, R., BUIJS, A., SACHS, N., OVERMEER, R. M., OFFERHAUS, G. J., BEGTHEL, H., KORVING, J., VAN DE WETERING, M., SCHWANK, G., LOGTENBERG, M., CUPPEN, E., SNIPPERT, H. J., MEDEMA, J. P., KOPS, G. J. & CLEVERS, H. 2015. Sequential cancer mutations in cultured human intestinal stem cells. *Nature,* 521**,** 43-7.

DRUKER, B. J., TAMURA, S., BUCHDUNGER, E., OHNO, S., SEGAL, G. M., FANNING, S., ZIMMERMANN, J. & LYDON, N. B. 1996. Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. *Nat Med,* 2**,** 561-6.

DRUKKER, C. A., VAN DEN HOUT, H. C., SONKE, G. S., BRAIN, E., BONNEFOI, H., CARDOSO, F., GOLDHIRSCH, A., HARBECK, N., HONKOOP, A. H., KOORNSTRA, R. H., VAN LAARHOVEN, H. W., PORTIELJE, J. E., SCHNEEWEISS, A., SMORENBURG, C. H., STOUTHARD, J., LINN, S. C. & SCHMIDT, M. K. 2014. Risk estimations and treatment decisions in early stage breast cancer: agreement among oncologists and the impact of the 70-gene signature. *Eur J Cancer,* 50**,** 1045-54.

DU, W. & ELEMENTO, O. 2015. Cancer systems biology: embracing complexity to develop better anticancer therapeutic strategies. *Oncogene,* 34**,** 3215-25.

DUMONTET, C. & JORDAN, M. A. 2010. Microtubule-binding agents: a dynamic field of cancer therapeutics. *Nat Rev Drug Discov,* 9**,** 790-803.

DUYMICH, C. E., CHARLET, J., YANG, X. J., JONES, P. A. & LIANG, G. N. 2016. DNMT3B isoforms without catalytic activity stimulate gene body methylation as accessory proteins in somatic cells. *Nature Communications,* 7.

EARLY BREAST CANCER TRIALISTS' COLLABORATIVE, G., DAVIES, C., GODWIN, J., GRAY, R., CLARKE, M., CUTTER, D., DARBY, S., MCGALE, P., PAN, H. C., TAYLOR, C., WANG, Y. C., DOWSETT, M., INGLE, J. & PETO, R. 2011. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. *Lancet,* 378**,** 771-84.

EARLY BREAST CANCER TRIALISTS' COLLABORATIVE, G., DOWSETT, M., FORBES, J. F., BRADLEY, R., INGLE, J., AIHARA, T., BLISS, J., BOCCARDO, F., COATES, A., COOMBES, R. C., CUZICK, J., DUBSKY, P., GNANT, M., KAUFMANN, M., KILBURN, L., PERRONE, F., REA, D., THURLIMANN, B., VAN DE VELDE, C., PAN, H., PETO, R., DAVIES, C. & GRAY, R. 2015. Aromatase inhibitors versus tamoxifen in early breast cancer: patient-level meta-analysis of the randomised trials. *Lancet,* 386**,** 1341-52.

EDGE, S. B. & COMPTON, C. C. 2010. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol,* 17**,** 1471-4.

EROLES, P., BOSCH, A., PEREZ-FIDALGO, J. A. & LLUCH, A. 2012. Molecular biology in breast cancer: intrinsic subtypes and signaling pathways. *Cancer Treat Rev,* 38**,** 698-707.

ESTELLER, M. 2007. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet,* 8**,** 286-98.

ESTELLER, M., CORN, P. G., BAYLIN, S. B. & HERMAN, J. G. 2001. A gene hypermethylation profile of human cancer. *Cancer Res,* 61**,** 3225-9.

EZEH, U. I., TUREK, P. J., REIJO, R. A. & CLARK, A. T. 2005. Human embryonic stem cell genes OCT4, NANOG, STELLAR, and GDF3 are expressed in both seminoma and breast carcinoma. *Cancer,* 104**,** 2255-65.

FACKLER, M. J., MCVEIGH, M., MEHROTRA, J., BLUM, M. A., LANGE, J., LAPIDES, A., GARRETT, E., ARGANI, P. & SUKUMAR, S. 2004. Quantitative multiplex methylation-specific PCR assay for the detection of promoter hypermethylation in multiple genes in breast cancer. *Cancer Res,* 64**,** 4442-52.

FAIRBANKS, D. J. & MAUGHAN, P. J. 2006. Evolution of the NANOG pseudogene family in the human and chimpanzee genomes. *BMC Evol Biol,* 6**,** 12.

FAN, C., OH, D. S., WESSELS, L., WEIGELT, B., NUYTEN, D. S., NOBEL, A. B., VAN'T VEER, L. J. & PEROU, C. M. 2006. Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med,* 355**,** 560-9.

FANG, R., ZHANG, G., GUO, Q., NING, F., WANG, H., CAI, S. H. & DU, J. 2013. Nodal promotes aggressive phenotype via Snail-mediated epithelial-mesenchymal transition in murine melanoma. *Cancer Letters,* 333**,** 66-75.

FANG, X., ZHENG, P., TANG, J. & LIU, Y. 2010. CD24: from A to Z. *Cell Mol Immunol,* 7**,** 100-3.

FARRELL, A. S. & SEARS, R. C. 2014. MYC degradation. *Cold Spring Harb Perspect Med,* 4.

FEINBERG, A. P. & TYCKO, B. 2004. The history of cancer epigenetics. *Nat Rev Cancer,* 4**,** 143-53.

FENG, B., NG, J. H., HENG, J. C. & NG, H. H. 2009. Molecules that promote or enhance reprogramming of somatic cells to induced pluripotent stem cells. *Cell Stem Cell,* 4**,** 301-12.

FENG, W., SHEN, L., WEN, S., ROSEN, D. G., JELINEK, J., HU, X., HUAN, S., HUANG, M., LIU, J., SAHIN, A. A., HUNT, K. K., BAST, R. C., JR., SHEN, Y., ISSA, J. P. & YU, Y. 2007. Correlation between CpG methylation profiles and hormone receptor status in breast cancers. *Breast Cancer Res,* 9**,** R57.

FESTUCCIA, N., OSORNO, R., HALBRITTER, F., KARWACKI-NEISIUS, V., NAVARRO, P., COLBY, D., WONG, F., YATES, A., TOMLINSON, S. R. & CHAMBERS, I. 2012. Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. *Cell Stem Cell,* 11**,** 477-90.

FISCHINGER, P. J., NOMURA, S., PEEBLES, P. T., HAAPALA, D. K. & BASSIN, R. H. 1972. Reversion of murine sarcoma virus transformed mouse cells: variants without a rescuable sarcoma virus. *Science,* 176**,** 1033-5.

FLAVAHAN, W. A., DRIER, Y., LIAU, B. B., GILLESPIE, S. M., VENTEICHER, A. S., STEMMER-RACHAMIMOV, A. O., SUVA, M. L. & BERNSTEIN, B. E. 2016. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature,* 529**,** 110-4.

FLEISCHER, T., TEKPLI, X., MATHELIER, A., WANG, S., NEBDAL, D., DHAKAL, H. P., SAHLBERG, K. K., SCHLICHTING, E., OSLO BREAST CANCER RESEARCH, C., BORRESEN-DALE, A. L., BORGEN, E., NAUME, B., ESKELAND, R., FRIGESSI, A., TOST, J., HURTADO, A. & KRISTENSEN, V. N. 2017. DNA methylation at enhancers identifies distinct breast cancer lineages. *Nat Commun,* 8**,** 1379.

FOLMES, C. D., NELSON, T. J., MARTINEZ-FERNANDEZ, A., ARRELL, D. K., LINDOR, J. Z., DZEJA, P. P., IKEDA, Y., PEREZ-TERZIC, C. & TERZIC, A. 2011. Somatic oxidative bioenergetics transitions into pluripotency-dependent glycolysis to facilitate nuclear reprogramming. *Cell Metab,* 14**,** 264-71.

FONG, P. C., BOSS, D. S., YAP, T. A., TUTT, A., WU, P., MERGUI-ROELVINK, M., MORTIMER, P., SWAISLAND, H., LAU, A., O'CONNOR, M. J., ASHWORTH, A., CARMICHAEL, J., KAYE,

S. B., SCHELLENS, J. H. & DE BONO, J. S. 2009. Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N Engl J Med,* 361**,** 123-34.

FORD, D., EASTON, D. F., BISHOP, D. T., NAROD, S. A. & GOLDGAR, D. E. 1994. Risks of cancer in BRCA1-mutation carriers. Breast Cancer Linkage Consortium. *Lancet,* 343**,** 692-5.

FRAGA, M. F., BALLESTAR, E., VILLAR-GAREA, A., BOIX-CHORNET, M., ESPADA, J., SCHOTTA, G., BONALDI, T., HAYDON, C., ROPERO, S. & PETRIE, K. 2005. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nature genetics,* 37**,** 391-400.

FRIDMAN, J. S. & LOWE, S. W. 2003. Control of apoptosis by p53. *Oncogene,* 22**,** 9030-40.

FRIERSON, H. F., JR., WOLBER, R. A., BEREAN, K. W., FRANQUEMONT, D. W., GAFFEY, M. J., BOYD, J. C. & WILBUR, D. C. 1995. Interobserver reproducibility of the Nottingham modification of the Bloom and Richardson histologic grading scheme for infiltrating ductal carcinoma. *Am J Clin Pathol,* 103**,** 195-8.

FU, T., LIU, Y., LI, K., WAN, W., PAPPOU, E. P., IACOBUZIO-DONAHUE, C. A., KERNER, Z., BAYLIN, S. B., WOLFGANG, C. L. & AHUJA, N. 2016. Tumors with unmethylated MLH1 and the CpG island methylator phenotype are associated with a poor prognosis in stage II colorectal cancer patients. *Oncotarget,* 7**,** 86480-86489.

FUKUCHI, T., SAKAMOTO, M., TSUDA, H., MARUYAMA, K., NOZAWA, S. & HIROHASHI, S. 1998. Beta-catenin mutation in carcinoma of the uterine endometrium. *Cancer Res,* 58**,** 3526-8.

FULLGRABE, J., KAVANAGH, E. & JOSEPH, B. 2011. Histone onco-modifications. *Oncogene,* 30**,** 3391-403.

FUSAKI, N., BAN, H., NISHIYAMA, A., SAEKI, K. & HASEGAWA, M. 2009. Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. *Proc Jpn Acad Ser B Phys Biol Sci,* 85**,** 348-62.

FYODOROV, D. V., ZHOU, B. R., SKOULTCHI, A. I. & BAI, Y. 2017. Emerging roles of linker histones in regulating chromatin structure and function. *Nat Rev Mol Cell Biol*.

GANDRE-BABBE, S., PALURU, P., ARIBEANA, C., CHOU, S. T., BRESOLIN, S., LU, L., SULLIVAN, S. K., TASIAN, S. K., WENG, J., FAVRE, H., CHOI, J. K., FRENCH, D. L., LOH, M. L. & WEISS, M. J. 2013. Patient-derived induced pluripotent stem cells recapitulate hematopoietic abnormalities of juvenile myelomonocytic leukemia. *Blood,* 121**,** 4925-9.

GARRAWAY, L. A. & LANDER, E. S. 2013. Lessons from the cancer genome. *Cell,* 153**,** 17-37.

GARRAWAY, L. A. & SELLERS, W. R. 2006. Lineage dependency and lineage-survival oncogenes in human cancer. *Nat Rev Cancer,* 6**,** 593-602.

GARZON, R., MARCUCCI, G. & CROCE, C. M. 2010. Targeting microRNAs in cancer: rationale, strategies and challenges. *Nat Rev Drug Discov,* 9**,** 775-89.

GAYTHER, S. A., BATLEY, S. J., LINGER, L., BANNISTER, A., THORPE, K., CHIN, S. F., DAIGO, Y., RUSSELL, P., WILSON, A., SOWTER, H. M., DELHANTY, J. D., PONDER, B. A., KOUZARIDES, T. & CALDAS, C. 2000. Mutations truncating the EP300 acetylase in human cancers. *Nat Genet,* 24**,** 300-3.

GHOSH, Z., WILSON, K. D., WU, Y., HU, S. J., QUERTERMOUS, T. & WU, J. C. 2010. Persistent Donor Cell Gene Expression among Human Induced Pluripotent Stem Cells Contributes to Differences with Human Embryonic Stem Cells. *Plos One,* 5.

GIDEKEL, S., PIZOV, G., BERGMAN, Y. & PIKARSKY, E. 2003. Oct-3/4 is a dose-dependent oncogenic fate determinant. *Cancer Cell,* 4**,** 361-70.

GILBERT, N., BOYLE, S., FIEGLER, H., WOODFINE, K., CARTER, N. P. & BICKMORE, W. A. 2004. Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell,* 118**,** 555-66.

GILLET, J. P., VARMA, S. & GOTTESMAN, M. M. 2013. The clinical relevance of cancer cell lines. *J Natl Cancer Inst,* 105**,** 452-8.

GINESTIER, C., HUR, M. H., CHARAFE-JAUFFRET, E., MONVILLE, F., DUTCHER, J., BROWN, M., JACQUEMIER, J., VIENS, P., KLEER, C. G., LIU, S., SCHOTT, A., HAYES, D., BIRNBAUM, D., WICHA, M. S. & DONTU, G. 2007. ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. *Cell Stem Cell,* 1**,** 555-67.

GIORGETTI, A., MONTSERRAT, N., AASEN, T., GONZALEZ, F., RODRIGUEZ-PIZA, I., VASSENA, R., RAYA, A., BOUE, S., BARRERO, M. J., CORBELLA, B. A., TORRABADELLA, M., VEIGA, A. & IZPISUA BELMONTE, J. C. 2009. Generation of induced pluripotent stem cells from human cord blood using OCT4 and SOX2. *Cell Stem Cell,* 5**,** 353-7.

GOBEL, G., AUER, D., GAUGG, I., SCHNEITTER, A., LESCHE, R., MULLER-HOLZNER, E., MARTH, C. & DAXENBICHLER, G. 2011. Prognostic significance of methylated RASSF1A and PITX2 genes in blood- and bone marrow plasma of breast cancer patients. *Breast Cancer Res Treat,* 130**,** 109-17.

GOLDBERG, A. D., ALLIS, C. D. & BERNSTEIN, E. 2007. Epigenetics: a landscape takes shape. *Cell,* 128**,** 635-8.

GOMMANS, W. M. & BEREZIKOV, E. 2012. Controlling miRNA regulation in disease. *Methods Mol Biol,* 822**,** 1-18.

GONCALVES, R., WARNER, W. A., LUO, J. & ELLIS, M. J. 2014. New concepts in breast cancer genomics and genetics. *Breast Cancer Res,* 16**,** 460.

GONZALEZ, F., BOUE, S. & IZPISUA BELMONTE, J. C. 2011. Methods for making induced pluripotent stem cells: reprogramming a la carte. *Nat Rev Genet,* 12**,** 231-42.

GREAVES, M. & MALEY, C. C. 2012. Clonal evolution in cancer. *Nature,* 481**,** 306-13.

GREENMAN, C., STEPHENS, P., SMITH, R., DALGLIESH, G. L., HUNTER, C., BIGNELL, G., DAVIES, H., TEAGUE, J., BUTLER, A., STEVENS, C., EDKINS, S., O'MEARA, S., VASTRIK, I., SCHMIDT, E. E., AVIS, T., BARTHORPE, S., BHAMRA, G., BUCK, G., CHOUDHURY, B., CLEMENTS, J., COLE, J., DICKS, E., FORBES, S., GRAY, K., HALLIDAY, K., HARRISON, R., HILLS, K., HINTON, J., JENKINSON, A., JONES, D., MENZIES, A., MIRONENKO, T., PERRY, J., RAINE, K., RICHARDSON, D., SHEPHERD, R., SMALL, A., TOFTS, C., VARIAN, J., WEBB, T., WEST, S., WIDAA, S., YATES, A., CAHILL, D. P., LOUIS, D. N., GOLDSTRAW, P., NICHOLSON, A. G., BRASSEUR, F., LOOIJENGA, L., WEBER, B. L., CHIEW, Y. E., DEFAZIO, A., GREAVES, M. F., GREEN, A. R., CAMPBELL, P., BIRNEY, E., EASTON, D. F., CHENEVIX-TRENCH, G., TAN, M. H., KHOO, S. K., TEH, B. T., YUEN, S. T., LEUNG, S. Y., WOOSTER, R., FUTREAL, P. A. & STRATTON, M. R. 2007. Patterns of somatic mutation in human cancer genomes. *Nature,* 446**,** 153-8.

GUO, S., ZI, X., SCHULZ, V. P., CHENG, J., ZHONG, M., KOOCHAKI, S. H., MEGYOLA, C. M., PAN, X., HEYDARI, K., WEISSMAN, S. M., GALLAGHER, P. G., KRAUSE, D. S., FAN, R. & LU, J. 2014. Nonstochastic reprogramming from a privileged somatic cell state. *Cell,* 156**,** 649-62.

GUO, W. 2014. Concise review: breast cancer stem cells: regulatory networks, stem cell niches, and disease relevance. *Stem Cells Transl Med,* 3**,** 942-8.

GUO, W., KECKESOVA, Z., DONAHER, J. L., SHIBUE, T., TISCHLER, V., REINHARDT, F., ITZKOVITZ, S., NOSKE, A., ZURRER-HARDI, U., BELL, G., TAM, W. L., MANI, S. A., VAN OUDENAARDEN, A. & WEINBERG, R. A. 2012. Slug and Sox9 cooperatively determine the mammary stem cell state. *Cell,* 148**,** 1015-28.

GURDON, J. B. 1962. The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. *J Embryol Exp Morphol,* 10**,** 622-40.

GUSTIN, J. P., KARAKAS, B., WEISS, M. B., ABUKHDEIR, A. M., LAURING, J., GARAY, J. P., COSGROVE, D., TAMAKI, A., KONISHI, H., KONISHI, Y., MOHSENI, M., WANG, G., ROSEN, D. M., DENMEADE, S. R., HIGGINS, M. J., VITOLO, M. I., BACHMAN, K. E. &

PARK, B. H. 2009. Knockin of mutant PIK3CA activates multiple oncogenic pathways. *Proc Natl Acad Sci U S A,* 106**,** 2835-40.

HAMAM, R., HAMAM, D., ALSALEH, K. A., KASSEM, M., ZAHER, W., ALFAYEZ, M., ALDAHMASH, A. & ALAJEZ, N. M. 2017. Circulating microRNAs in breast cancer: novel diagnostic and prognostic biomarkers. *Cell Death Dis,* 8**,** e3045.

HAN, J., ZHANG, F., YU, M., ZHAO, P., JI, W., ZHANG, H., WU, B., WANG, Y. & NIU, R. 2012. RNA interference-mediated silencing of NANOG reduces cell proliferation and induces G0/G1 cell cycle arrest in breast cancer cells. *Cancer Lett,* 321**,** 80-8.

HANAHAN, D. & WEINBERG, R. A. 2011a. Hallmarks of cancer: the next generation. *Cell,* 144**,** 646-74.

HANAHAN, D. & WEINBERG, R. A. 2011b. Hallmarks of Cancer: The Next Generation. *Cell,* 144**,** 646-674.

HANNA, J., SAHA, K., PANDO, B., VAN ZON, J., LENGNER, C. J., CREYGHTON, M. P., VAN OUDENAARDEN, A. & JAENISCH, R. 2009. Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature,* 462**,** 595-U63.

HAO, D., WANG, L. & DI, L. J. 2016. Distinct mutation accumulation rates among tissues determine the variation in cancer risk. *Sci Rep,* 6**,** 19458.

HARRIS, H., MILLER, O. J., KLEIN, G., WORST, P. & TACHIBANA, T. 1969. Suppression of malignancy by cell fusion. *Nature,* 223**,** 363-8.

HARRIS, H. & WATKINS, J. F. 1965. Hybrid Cells Derived from Mouse and Man: Artificial Heterokaryons of Mammalian Cells from Different Species. *Nature,* 205**,** 640-6.

HARTMAN, A. R., KALDATE, R. R., SAILER, L. M., PAINTER, L., GRIER, C. E., ENDSLEY, R. R., GRIFFIN, M., HAMILTON, S. A., FRYE, C. A., SILBERMAN, M. A., WENSTRUP, R. J. & SANDBACH, J. F. 2012. Prevalence of BRCA mutations in an unselected population of triple-negative breast cancer. *Cancer,* 118**,** 2787-95.

HATSUMI, T. & YAMAMURO, Y. 2006. Downregulation of estrogen receptor gene expression by exogenous 17 beta-estradiol in the mammary glands of lactating mice. *Experimental Biology and Medicine,* 231**,** 311-316.

HATZIMICHAEL, E. & CROOK, T. 2013. Cancer epigenetics: new therapies and new challenges. *J Drug Deliv,* 2013**,** 529312.

HAYASHI, H., ARAO, T., TOGASHI, Y., KATO, H., FUJITA, Y., DE VELASCO, M. A., KIMURA, H., MATSUMOTO, K., TANAKA, K., OKAMOTO, I., ITO, A., YAMADA, Y., NAKAGAWA, K. & NISHIO, K. 2015. The OCT4 pseudogene POU5F1B is amplified and promotes an aggressive phenotype in gastric cancer. *Oncogene,* 34**,** 199-208.

HAYDEN, E. C. 2011. Stem cells: The growing pains of pluripotency. *Nature,* 473**,** 272-4.

HAYES, J., PERUZZI, P. P. & LAWLER, S. 2014. MicroRNAs in cancer: biomarkers, functions and therapy. *Trends Mol Med,* 20**,** 460-9.

HENDRIX, M. J., SEFTOR, E. A., SEFTOR, R. E., KASEMEIER-KULESA, J., KULESA, P. M. & POSTOVIT, L. M. 2007. Reprogramming metastatic tumour cells with embryonic microenvironments. *Nat Rev Cancer,* 7**,** 246-55.

HENG, J. C., FENG, B., HAN, J., JIANG, J., KRAUS, P., NG, J. H., ORLOV, Y. L., HUSS, M., YANG, L., LUFKIN, T., LIM, B. & NG, H. H. 2010. The nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells. *Cell Stem Cell,* 6**,** 167-74.

HERSCHKOWITZ, J. I., SIMIN, K., WEIGMAN, V. J., MIKAELIAN, I., USARY, J., HU, Z., RASMUSSEN, K. E., JONES, L. P., ASSEFNIA, S., CHANDRASEKHARAN, S., BACKLUND, M. G., YIN, Y., KHRAMTSOV, A. I., BASTEIN, R., QUACKENBUSH, J., GLAZER, R. I., BROWN, P. H., GREEN, J. E., KOPELOVICH, L., FURTH, P. A., PALAZZO, J. P., OLOPADE, O. I., BERNARD, P. S., CHURCHILL, G. A., VAN DYKE, T. & PEROU, C. M. 2007. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol,* 8**,** R76.

HISANO, Y., SAKUMA, T., NAKADE, S., OHGA, R., OTA, S., OKAMOTO, H., YAMAMOTO, T. & KAWAHARA, A. 2015. Precise in-frame integration of exogenous DNA mediated by CRISPR/Cas9 system in zebrafish. *Sci Rep,* 5**,** 8841.

HISER, L., AGGARWAL, A., YOUNG, R., FRANKFURTER, A., SPANO, A., CORREIA, J. J. & LOBERT, S. 2006. Comparison of beta-tubulin mRNA and protein levels in 12 human cancer cell lines. *Cell Motil Cytoskeleton,* 63**,** 41-52.

HO, L., MILLER, E. L., RONAN, J. L., HO, W. Q., JOTHI, R. & CRABTREE, G. R. 2011. esBAF facilitates pluripotency by conditioning the genome for LIF/STAT3 signalling and by regulating polycomb function. *Nat Cell Biol,* 13**,** 903-13.

HOCHEDLINGER, K., BLELLOCH, R., BRENNAN, C., YAMADA, Y., KIM, M., CHIN, L. & JAENISCH, R. 2004. Reprogramming of a melanoma genome by nuclear transplantation. *Genes Dev,* 18**,** 1875-85.

HOCHEDLINGER, K. & JAENISCH, R. 2006. Nuclear reprogramming and pluripotency. *Nature,* 441**,** 1061-7.

HODGKINS, A., FARNE, A., PERERA, S., GREGO, T., PARRY-SMITH, D. J., SKARNES, W. C. & IYER, V. 2015. WGE: a CRISPR database for genome engineering. *Bioinformatics,* 31**,** 3078-80.

HOLLESTELLE, A., ELSTRODT, F., NAGEL, J. H., KALLEMEIJN, W. W. & SCHUTTE, M. 2007. Phosphatidylinositol-3-OH kinase or RAS pathway mutations in human breast cancer cell lines. *Mol Cancer Res,* 5**,** 195-201.

HOLM, K., HEGARDT, C., STAAF, J., VALLON-CHRISTERSSON, J., JONSSON, G., OLSSON, H., BORG, A. & RINGNER, M. 2010. Molecular subtypes of breast cancer are associated with characteristic DNA methylation patterns. *Breast Cancer Res,* 12**,** R36.

HOLM, K., STAAF, J., LAUSS, M., AINE, M., LINDGREN, D., BENDAHL, P. O., VALLON-CHRISTERSSON, J., BARKARDOTTIR, R. B., HOGLUND, M., BORG, A., JONSSON, G. & RINGNER, M. 2016. An integrated genomics analysis of epigenetic subtypes in human breast tumors links DNA methylation patterns to chromatin states in normal mammary cells. *Breast Cancer Res,* 18**,** 27.

HOLOCH, D. & MOAZED, D. 2015. RNA-mediated epigenetic regulation of gene expression. *Nat Rev Genet,* 16**,** 71-84.

HONETH, G., BENDAHL, P. O., RINGNER, M., SAAL, L. H., GRUVBERGER-SAAL, S. K., LOVGREN, K., GRABAU, D., FERNO, M., BORG, A. & HEGARDT, C. 2008. The CD44+/CD24- phenotype is enriched in basal-like breast tumors. *Breast Cancer Res,* 10**,** R53.

HONG, H., TAKAHASHI, K., ICHISAKA, T., AOI, T., KANAGAWA, O., NAKAGAWA, M., OKITA, K. & YAMANAKA, S. 2009a. Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. *Nature,* 460**,** 1132-5.

HONG, H., TAKAHASHI, K., ICHISAKA, T., AOI, T., KANAGAWA, O., NAKAGAWA, M., OKITA, K. & YAMANAKA, S. 2009b. Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. *Nature,* 460**,** 1132-U95.

HOQUE, M. O., PRENCIPE, M., POETA, M. L., BARBANO, R., VALORI, V. M., COPETTI, M., GALLO, A. P., BRAIT, M., MAIELLO, E., APICELLA, A., ROSSIELLO, R., ZITO, F., STEFANIA, T., PARADISO, A., CARELLA, M., DALLAPICCOLA, B., MURGO, R., CAROSI, I., BISCEGLIA, M., FAZIO, V. M., SIDRANSKY, D. & PARRELLA, P. 2009. Changes in CpG islands promoter methylation patterns during ductal breast carcinoma progression. *Cancer Epidemiol Biomarkers Prev,* 18**,** 2694-700.

HORE, T. A., VON MEYENN, F., RAVICHANDRAN, M., BACHMAN, M., FICZ, G., OXLEY, D., SANTOS, F., BALASUBRAMANIAN, S., JURKOWSKI, T. P. & REIK, W. 2016. Retinol and ascorbate drive erasure of epigenetic memory and enhance reprogramming to naive pluripotency by complementary mechanisms. *Proceedings of the National Academy of Sciences of the United States of America,* 113**,** 12202-12207.

HORIMOTO, Y., ARAKAWA, A., SASAHARA, N., TANABE, M., SAI, S., HIMURO, T. & SAITO, M. 2016. Combination of Cancer Stem Cell Markers CD44 and CD24 Is Superior to ALDH1 as a Prognostic Indicator in Breast Cancer Patients with Distant Metastases. *PLoS One,* 11**,** e0165253.

HOSHINO, H., NAGANO, H., HARAGUCHI, N., NISHIKAWA, S., TOMOKUNI, A., KANO, Y., FUKUSUMI, T., SAITO, T., OZAKI, M., SAKAI, D., SATOH, T., EGUCHI, H., SEKIMOTO, M., DOKI, Y., MORI, M. & ISHII, H. 2012. Hypoxia and TP53 deficiency for induced pluripotent stem cell-like properties in gastrointestinal cancer. *Int J Oncol,* 40**,** 1423-30.

HOSLER, B. A., LAROSA, G. J., GRIPPO, J. F. & GUDAS, L. J. 1989. Expression of REX-1, a gene containing zinc finger motifs, is rapidly reduced by retinoic acid in F9 teratocarcinoma cells. *Mol Cell Biol,* 9**,** 5623-9.

HOU, P., LI, Y., ZHANG, X., LIU, C., GUAN, J., LI, H., ZHAO, T., YE, J., YANG, W., LIU, K., GE, J., XU, J., ZHANG, Q., ZHAO, Y. & DENG, H. 2013. Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. *Science,* 341**,** 651-4.

HOWE, S. J., MANSOUR, M. R., SCHWARZWAELDER, K., BARTHOLOMAE, C., HUBANK, M., KEMPSKI, H., BRUGMAN, M. H., PIKE-OVERZET, K., CHATTERS, S. J., DE RIDDER, D., GILMOUR, K. C., ADAMS, S., THORNHILL, S. I., PARSLEY, K. L., STAAL, F. J., GALE, R. E., LINCH, D. C., BAYFORD, J., BROWN, L., QUAYE, M., KINNON, C., ANCLIFF, P., WEBB, D. K., SCHMIDT, M., VON KALLE, C., GASPAR, H. B. & THRASHER, A. J. 2008. Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. *J Clin Invest,* 118**,** 3143-50.

HSIEH, A. L., WALTON, Z. E., ALTMAN, B. J., STINE, Z. E. & DANG, C. V. 2015. MYC and metabolism on the path to cancer. *Semin Cell Dev Biol,* 43**,** 11-21.

HU, K., YU, J., SUKNUNTHA, K., TIAN, S., MONTGOMERY, K., CHOI, K. D., STEWART, R., THOMSON, J. A. & SLUKVIN, II 2011. Efficient generation of transgene-free induced pluripotent stem cells from normal and neoplastic bone marrow and cord blood mononuclear cells. *Blood,* 117**,** e109-19.

HU, T., LIU, S., BREITER, D. R., WANG, F., TANG, Y. & SUN, S. 2008. Octamer 4 small interfering RNA results in cancer stem cell-like cell apoptosis. *Cancer Res,* 68**,** 6533-40.

HU, X., ZHANG, L., MAO, S. Q., LI, Z., CHEN, J., ZHANG, R. R., WU, H. P., GAO, J., GUO, F., LIU, W., XU, G. F., DAI, H. Q., SHI, Y. G., LI, X., HU, B., TANG, F., PEI, D. & XU, G. L. 2014. Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. *Cell Stem Cell,* 14**,** 512-22.

HUANG, H., MAHLER-ARAUJO, B. M., SANKILA, A., CHIMELLI, L., YONEKAWA, Y., KLEIHUES, P. & OHGAKI, H. 2000. APC mutations in sporadic medulloblastomas. *Am J Pathol,* 156**,** 433-7.

HUANG, K., MARUYAMA, T. & FAN, G. 2014. The naive state of human pluripotent stem cells: a synthesis of stem cell and preimplantation embryo transcriptome analyses. *Cell Stem Cell,* 15**,** 410-415.

HUANG, Y. & DAVIDSON, N. E. 2016. Targeting tumorigenicity of breast cancer stem-like cells using combination epigenetic therapy: something old and something new. *J Thorac Dis,* 8**,** 2971-2974.

HUANG, Y., OSORNO, R., TSAKIRIDIS, A. & WILSON, V. 2012. In Vivo differentiation potential of epiblast stem cells revealed by chimeric embryo formation. *Cell Rep,* 2**,** 1571-8.

HUNTER, A. C. 2006. Molecular hurdles in polyfectin design and mechanistic background to polycation induced cytotoxicity. *Adv Drug Deliv Rev,* 58**,** 1523-31.

HUSSEIN, S. M., PURI, M. C., TONGE, P. D., BENEVENTO, M., CORSO, A. J., CLANCY, J. L., MOSBERGEN, R., LI, M., LEE, D. S., CLOONAN, N., WOOD, D. L., MUNOZ, J., MIDDLETON, R., KORN, O., PATEL, H. R., WHITE, C. A., SHIN, J. Y., GAUTHIER, M. E., LE CAO, K. A., KIM, J. I., MAR, J. C., SHAKIBA, N., RITCHIE, W., RASKO, J. E.,

GRIMMOND, S. M., ZANDSTRA, P. W., WELLS, C. A., PREISS, T., SEO, J. S., HECK, A. J., ROGERS, I. M. & NAGY, A. 2014. Genome-wide characterization of the routes to pluripotency. *Nature,* 516**,** 198-206.

ICHIDA, J. K., BLANCHARD, J., LAM, K., SON, E. Y., CHUNG, J. E., EGLI, D., LOH, K. M., CARTER, A. C., DI GIORGIO, F. P., KOSZKA, K., HUANGFU, D., AKUTSU, H., LIU, D. R., RUBIN, L. L. & EGGAN, K. 2009. A small-molecule inhibitor of tgf-Beta signaling replaces sox2 in reprogramming by inducing nanog. *Cell Stem Cell,* 5**,** 491-503.

IGLESIAS, J. M., GUMUZIO, J. & MARTIN, A. G. 2017. Linking Pluripotency Reprogramming and Cancer. *Stem Cells Transl Med,* 6**,** 335-339.

IKUSHIMA, H., TODO, T., INO, Y., TAKAHASHI, M., SAITO, N., MIYAZAWA, K. & MIYAZONO, K. 2011. Glioma-initiating cells retain their tumorigenicity through integration of the Sox axis and Oct4 protein. *J Biol Chem,* 286**,** 41434-41.

ILIOPOULOS, D., LINDAHL-ALLEN, M., POLYTARCHOU, C., HIRSCH, H. A., TSICHLIS, P. N. & STRUHL, K. 2010. Loss of miR-200 inhibition of Suz12 leads to polycomb-mediated repression required for the formation and maintenance of cancer stem cells. *Mol Cell,* 39**,** 761-72.

ILLINGWORTH, R. S. & BIRD, A. P. 2009. CpG islands--'a rough guide'. *FEBS Lett,* 583**,** 1713-20.

ILYAS, M., TOMLINSON, I. P., ROWAN, A., PIGNATELLI, M. & BODMER, W. F. 1997. Beta-catenin mutations in cell lines established from human colorectal cancers. *Proc Natl Acad Sci U S A,* 94**,** 10330-4.

IRIE, N., WEINBERGER, L., TANG, W. W., KOBAYASHI, T., VIUKOV, S., MANOR, Y. S., DIETMANN, S., HANNA, J. H. & SURANI, M. A. 2015. SOX17 is a critical specifier of human primordial germ cell fate. *Cell,* 160**,** 253-68.

ISAKOFF, S. J., ENGELMAN, J. A., IRIE, H. Y., LUO, J., BRACHMANN, S. M., PEARLINE, R. V., CANTLEY, L. C. & BRUGGE, J. S. 2005. Breast cancer-associated PIK3CA mutations are oncogenic in mammary epithelial cells. *Cancer Res,* 65**,** 10992-1000.

ISSA, J. P. 2004. CpG island methylator phenotype in cancer. *Nat Rev Cancer,* 4**,** 988-93.

IVANOV, N. A., TAO, R., CHENOWETH, J. G., BRANDTJEN, A., MIGHDOLL, M. I., GENOVA, J. D., MCKAY, R. D., JIA, Y. K., WEINBERGER, D. R., KLEINMAN, J. E., HYDE, T. M. & JAFFE, A. E. 2016. Strong Components of Epigenetic Memory in Cultured Human Fibroblasts Related to Site of Origin and Donor Age. *Plos Genetics,* 12.

IWAO, K., NAKAMORI, S., KAMEYAMA, M., IMAOKA, S., KINOSHITA, M., FUKUI, T., ISHIGURO, S., NAKAMURA, Y. & MIYOSHI, Y. 1998. Activation of the beta-catenin gene by interstitial deletions involving exon 3 in primary colorectal carcinomas without adenomatous polyposis coli mutations. *Cancer Res,* 58**,** 1021-6.

JACKSON, S. A., OLUFS, Z. P., TRAN, K. A., ZAIDAN, N. Z. & SRIDHARAN, R. 2016. Alternative Routes to Induced Pluripotent Stem Cells Revealed by Reprogramming of the Neural Lineage. *Stem Cell Reports,* 6**,** 302-11.

JACQUEL, A., HERRANT, M., LEGROS, L., BELHACENE, N., LUCIANO, F., PAGES, G., HOFMAN, P. & AUBERGER, P. 2003. Imatinib induces mitochondria-dependent apoptosis of the Bcr-Abl-positive K562 cell line and its differentiation toward the erythroid lineage. *FASEB J,* 17**,** 2160-2.

JAENISCH, R. & BIRD, A. 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet,* 33 Suppl**,** 245-54.

JANG, H. & SERRA, C. 2014. Nutrition, epigenetics, and diseases. *Clin Nutr Res,* 3**,** 1-8.

JERONIMO, C., MONTEIRO, P., HENRIQUE, R., DINIS-RIBEIRO, M., COSTA, I., COSTA, V. L., FILIPE, L., CARVALHO, A. L., HOQUE, M. O., PAIS, I., LEAL, C., TEIXEIRA, M. R. & SIDRANSKY, D. 2008. Quantitative hypermethylation of a small panel of genes augments the diagnostic accuracy in fine-needle aspirate washings of breast lesions. *Breast Cancer Res Treat,* 109**,** 27-34.

JETER, C. R., BADEAUX, M., CHOY, G., CHANDRA, D., PATRAWALA, L., LIU, C., CALHOUN-DAVIS, T., ZAEHRES, H., DALEY, G. Q. & TANG, D. G. 2009. Functional evidence that the self-renewal gene NANOG regulates human tumor development. *Stem Cells,* 27**,** 993-1005.

JEZKOVA, E., KAJO, K., ZUBOR, P., GRENDAR, M., MALICHEROVA, B., MENDELOVA, A., DOKUS, K., LASABOVA, Z., PLANK, L. & DANKO, J. 2016. Methylation in promoter regions of PITX2 and RASSF1A genes in association with clinicopathological features in breast cancer patients. *Tumour Biol*.

JIN, B., LI, Y. & ROBERTSON, K. D. 2011. DNA methylation: superior or subordinate in the epigenetic hierarchy? *Genes Cancer,* 2**,** 607-17.

JIN, T., BRANCH, D. R., ZHANG, X., QI, S., YOUNGSON, B. & GOSS, P. E. 1999. Examination of POU homeobox gene expression in human breast cancer cells. *Int J Cancer,* 81**,** 104-12.

JONES, P. A. & BAYLIN, S. B. 2002. The fundamental role of epigenetic events in cancer. *Nat Rev Genet,* 3**,** 415-28.

JONES, P. A., ISSA, J. P. & BAYLIN, S. 2016. Targeting the cancer epigenome for therapy. *Nat Rev Genet,* 17**,** 630-41.

JONES, T. D., ULBRIGHT, T. M., EBLE, J. N., BALDRIDGE, L. A. & CHENG, L. 2004. OCT4 staining in testicular tumors: a sensitive and specific marker for seminoma and embryonal carcinoma. *Am J Surg Pathol,* 28**,** 935-40.

JORDAN, C. T., GUZMAN, M. L. & NOBLE, M. 2006. Cancer stem cells. *N Engl J Med,* 355**,** 1253-61.

JUDES, G., DAGDEMIR, A., KARSLI-CEPPIOGLU, S., LEBERT, A., ECHEGUT, M., NGOLLO, M., BIGNON, Y. J., PENAULT-LLORCA, F. & BERNARD-GALLON, D. 2016. H3K4 acetylation, H3K9 acetylation and H3K27 methylation in breast tumor molecular subtypes. *Epigenomics,* 8**,** 909-24.

JUDSON, R. L., BABIARZ, J. E., VENERE, M. & BLELLOCH, R. 2009. Embryonic stem cell-specific microRNAs promote induced pluripotency. *Nat Biotechnol,* 27**,** 459-61.

JUNG, M., PETERSON, H., CHAVEZ, L., KAHLEM, P., LEHRACH, H., VILO, J. & ADJAYE, J. 2010. A data integration approach to mapping OCT4 gene regulatory networks operative in embryonic stem cells and embryonal carcinoma cells. *PLoS One,* 5**,** e10709.

KAIKKONEN, M. U., LAM, M. T. & GLASS, C. K. 2011. Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovasc Res,* 90**,** 430-40.

KAJABOVA, V., SMOLKOVA, B., ZMETAKOVA, I., SEBOVA, K., KRIVULCIK, T., BELLA, V., KAJO, K., MACHALEKOVA, K. & FRIDRICHOVA, I. 2013. RASSF1A Promoter Methylation Levels Positively Correlate with Estrogen Receptor Expression in Breast Cancer Patients. *Transl Oncol,* 6**,** 297-304.

KAJI, K., NORRBY, K., PACA, A., MILEIKOVSKY, M., MOHSENI, P. & WOLTJEN, K. 2009. Virus-free induction of pluripotency and subsequent excision of reprogramming factors. *Nature,* 458**,** 771-5.

KALA, R. & TOLLEFSBOL, T. O. 2016. A Novel Combinatorial Epigenetic Therapy Using Resveratrol and Pterostilbene for Restoring Estrogen Receptor-alpha (ERalpha) Expression in ERalpha-Negative Breast Cancer Cells. *PLoS One,* 11**,** e0155057.

KANG, S. J., PARK, Y. I., SO, B. & KANG, H. G. 2014. Sodium butyrate efficiently converts fully reprogrammed induced pluripotent stem cells from mouse partially reprogrammed cells. *Cell Reprogram,* 16**,** 345-54.

KARAKAS, B., BACHMAN, K. E. & PARK, B. H. 2006. Mutation of the PIK3CA oncogene in human cancers. *Br J Cancer,* 94**,** 455-9.

KARETA, M. S., GORGES, L. L., HAFEEZ, S., BENAYOUN, B. A., MARRO, S., ZMOOS, A. F., CECCHINI, M. J., SPACEK, D., BATISTA, L. F., O'BRIEN, M., NG, Y. H., ANG, C. E., VAKA, D., ARTANDI, S. E., DICK, F. A., BRUNET, A., SAGE, J. & WERNIG, M. 2015. Inhibition

of pluripotency networks by the Rb tumor suppressor restricts reprogramming and tumorigenesis. *Cell Stem Cell,* 16**,** 39-50.

KAROLCHIK, D., BEJERANO, G., HINRICHS, A. S., KUHN, R. M., MILLER, W., ROSENBLOOM, K. R., ZWEIG, A. S., HAUSSLER, D. & KENT, W. J. 2007. Comparative genomic analysis using the UCSC genome browser. *Methods Mol Biol,* 395**,** 17-34.

KAROLCHIK, D., HINRICHS, A. S. & KENT, W. J. 2012. The UCSC Genome Browser. *Curr Protoc Bioinformatics,* Chapter 1**,** Unit1 4.

KARSLI-CEPPIOGLU, S., DAGDEMIR, A., JUDES, G., LEBERT, A., PENAULT-LLORCA, F., BIGNON, Y. J. & BERNARD-GALLON, D. 2017. The Epigenetic Landscape of Promoter Genome-wide Analysis in Breast Cancer. *Sci Rep,* 7**,** 6597.

KASINSKI, A. L. & SLACK, F. J. 2011. Epigenetics and genetics. MicroRNAs en route to the clinic: progress in validating and targeting microRNAs for cancer therapy. *Nat Rev Cancer,* 11**,** 849-64.

KATZ, T. A., VASILATOS, S. N., HARRINGTON, E., OESTERREICH, S., DAVIDSON, N. E. & HUANG, Y. 2014. Inhibition of histone demethylase, LSD2 (KDM1B), attenuates DNA methylation and increases sensitivity to DNMT inhibitor-induced apoptosis in breast cancer cells. *Breast Cancer Research and Treatment,* 146**,** 99-108.

KAWAMURA, T., SUZUKI, J., WANG, Y. V., MENENDEZ, S., MORERA, L. B., RAYA, A., WAHL, G. M. & IZPISUA BELMONTE, J. C. 2009. Linking the p53 tumour suppressor pathway to somatic cell reprogramming. *Nature,* 460**,** 1140-4.

KELLER, G. 2005. Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes Dev,* 19**,** 1129-55.

KHATODIA, S., BHATOTIA, K., PASSRICHA, N., KHURANA, S. M. & TUTEJA, N. 2016. The CRISPR/Cas Genome-Editing Tool: Application in Improvement of Crops. *Front Plant Sci,* 7**,** 506.

KHOO, M. L., MCQUADE, L. R., SMITH, M. S., LEES, J. G., SIDHU, K. S. & TUCH, B. E. 2005. Growth and differentiation of embryoid bodies derived from human embryonic stem cells: effect of glucose and basic fibroblast growth factor. *Biol Reprod,* 73**,** 1147-56.

KHOURY, T., ADEMUYIWA, F. O., CHANDRASEKHAR, R., JABBOUR, M., DELEO, A., FERRONE, S., WANG, Y. & WANG, X. 2012. Aldehyde dehydrogenase 1A1 expression in breast cancer is associated with stage, triple negativity, and outcome to neoadjuvant chemotherapy. *Mod Pathol,* 25**,** 388-97.

KIDA, Y. S., KAWAMURA, T., WEI, Z., SOGO, T., JACINTO, S., SHIGENO, A., KUSHIGE, H., YOSHIHARA, E., LIDDLE, C., ECKER, J. R., YU, R. T., ATKINS, A. R., DOWNES, M. & EVANS, R. M. 2015. ERRs Mediate a Metabolic Switch Required for Somatic Cell Reprogramming to Pluripotency. *Cell Stem Cell,* 16**,** 547-55.

KIM, D., KIM, C. H., MOON, J. I., CHUNG, Y. G., CHANG, M. Y., HAN, B. S., KO, S., YANG, E., CHA, K. Y., LANZA, R. & KIM, K. S. 2009a. Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins. *Cell Stem Cell,* 4**,** 472-6.

KIM, D., WU, X., YOUNG, A. T. & HAYNES, C. L. 2014. Microfluidics-based in vivo mimetic systems for the study of cellular biology. *Acc Chem Res,* 47**,** 1165-73.

KIM, J., CHU, J., SHEN, X., WANG, J. & ORKIN, S. H. 2008a. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell,* 132**,** 1049-61.

KIM, J. & HAYWARD, R. C. 2012. Mimicking dynamic in vivo environments with stimuli-responsive materials for cell culture. *Trends Biotechnol,* 30**,** 426-39.

KIM, J., HOFFMAN, J. P., ALPAUGH, R. K., RHIM, A. D., REICHERT, M., STANGER, B. Z., FURTH, E. E., SEPULVEDA, A. R., YUAN, C. X., WON, K. J., DONAHUE, G., SANDS, J., GUMBS, A. A. & ZARET, K. S. 2013. An iPSC line from human pancreatic ductal adenocarcinoma undergoes early to invasive stages of pancreatic cancer progression. *Cell Rep,* 3**,** 2088-99.

KIM, J. & ZARET, K. S. 2015. Reprogramming of human cancer cells to pluripotency for models of cancer progression. *EMBO J,* 34**,** 739-47.

KIM, J. B., GREBER, B., ARAUZO-BRAVO, M. J., MEYER, J., PARK, K. I., ZAEHRES, H. & SCHOLER, H. R. 2009b. Direct reprogramming of human neural stem cells by OCT4. *Nature,* 461**,** 649-3.

KIM, J. B., ZAEHRES, H., WU, G., GENTILE, L., KO, K., SEBASTIANO, V., ARAUZO-BRAVO, M. J., RUAU, D., HAN, D. W., ZENKE, M. & SCHOLER, H. R. 2008b. Pluripotent stem cells induced from adult neural stem cells by reprogramming with two factors. *Nature,* 454**,** 646-50.

KIM, K., DOI, A., WEN, B., NG, K., ZHAO, R., CAHAN, P., KIM, J., ARYEE, M. J., JI, H., EHRLICH, L. I., YABUUCHI, A., TAKEUCHI, A., CUNNIFF, K. C., HONGGUANG, H., MCKINNEY-FREEMAN, S., NAVEIRAS, O., YOON, T. J., IRIZARRY, R. A., JUNG, N., SEITA, J., HANNA, J., MURAKAMI, P., JAENISCH, R., WEISSLEDER, R., ORKIN, S. H., WEISSMAN, I. L., FEINBERG, A. P. & DALEY, G. Q. 2010. Epigenetic memory in induced pluripotent stem cells. *Nature,* 467**,** 285-90.

KIM, K., ZHAO, R., DOI, A., NG, K., UNTERNAEHRER, J., CAHAN, P., HUO, H. G., LOH, Y. H., ARYEE, M. J., LENSCH, M. W., LI, H., COLLINS, J. J., FEINBERG, A. P. & DALEY, G. Q. 2011. Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent stem cells. *Nature Biotechnology,* 29**,** 1117-1119.

KIM, M. J., RO, J. Y., AHN, S. H., KIM, H. H., KIM, S. B. & GONG, G. 2006. Clinicopathologic significance of the basal-like subtype of breast cancer: a comparison with hormone receptor and Her2/neu-overexpressing phenotypes. *Hum Pathol,* 37**,** 1217-26.

KIM, R. J. & NAM, J. S. 2011. OCT4 Expression Enhances Features of Cancer Stem Cells in a Mouse Model of Breast Cancer. *Lab Anim Res,* 27**,** 147-52.

KIM, W., KIM, R., PARK, G., PARK, J. W. & KIM, J. E. 2012. Deficiency of H3K79 histone methyltransferase Dot1-like protein (DOT1L) inhibits cell proliferation. *J Biol Chem,* 287**,** 5588-99.

KLEINSMITH, L. J. & PIERCE, G. B., JR. 1964. Multipotentiality of Single Embryonal Carcinoma Cells. *Cancer Res,* 24**,** 1544-51.

KLEINSTIVER, B. P., PREW, M. S., TSAI, S. Q., NGUYEN, N. T., TOPKAR, V. V., ZHENG, Z. & JOUNG, J. K. 2015. Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nat Biotechnol,* 33**,** 1293-1298.

KLOTEN, V., BECKER, B., WINNER, K., SCHRAUDER, M. G., FASCHING, P. A., ANZENEDER, T., VEECK, J., HARTMANN, A., KNUCHEL, R. & DAHL, E. 2013. Promoter hypermethylation of the tumor-suppressor genes ITIH5, DKK3, and RASSF1A as novel biomarkers for blood-based breast cancer screening. *Breast Cancer Res,* 15**,** R4.

KNAUPP, A. S., BUCKBERRY, S., PFLUEGER, J., LIM, S. M., FORD, E., LARCOMBE, M. R., ROSSELLO, F. J., DE MENDOZA, A., ALAEI, S., FIRAS, J., HOLMES, M. L., NAIR, S. S., CLARK, S. J., NEFZGER, C. M., LISTER, R. & POLO, J. M. 2017. Transient and Permanent Reconfiguration of Chromatin and Transcription Factor Occupancy Drive Reprogramming. *Cell Stem Cell,* 21**,** 834-845 e6.

KOBAYASHI, K., SAGAE, S., NISHIOKA, Y., TOKINO, T. & KUDO, R. 1999. Mutations of the beta-catenin gene in endometrial carcinomas. *Jpn J Cancer Res,* 90**,** 55-9.

KOBOLDT, D. C., FULTON, R. S., MCLELLAN, M. D., SCHMIDT, H., KALICKI-VEIZER, J., MCMICHAEL, J. F., FULTON, L. L., DOOLING, D. J., DING, L., MARDIS, E. R., WILSON, R. K., ALLY, A., BALASUNDARAM, M., BUTTERFIELD, Y. S. N., CARLSEN, R., CARTER, C., CHU, A., CHUAH, E., CHUN, H. J. E., COOPE, R. J. N., DHALLA, N., GUIN, R., HIRST, C., HIRST, M., HOLT, R. A., LEE, D., LI, H. Y. I., MAYO, M., MOORE, R. A., MUNGALL, A. J., PLEASANCE, E., ROBERTSON, A. G., SCHEIN, J. E., SHAFIEI, A., SIPAHIMALANI, P., SLOBODAN, J. R., STOLL, D., TAM, A., THIESSEN, N., VARHOL, R. J., WYE, N., ZENG, T., ZHAO, Y. J., BIROL, I., JONES, S. J. M., MARRA, M. A., CHERNIACK, A. D., SAKSENA, G.,

ONOFRIO, R. C., PHO, N. H., CARTER, S. L., SCHUMACHER, S. E., TABAK, B., HERNANDEZ, B., GENTRY, J., NGUYEN, H., CRENSHAW, A., ARDLIE, K., BEROUKHIM, R., WINCKLER, W., GETZ, G., GABRIEL, S. B., MEYERSON, M., CHIN, L., PARK, P. J., KUCHERLAPATI, R., HOADLEY, K. A., AUMAN, J. T., FAN, C., TURMAN, Y. J., SHI, Y., LI, L., TOPAL, M. D., HE, X. P., CHAO, H. H., PRAT, A., SILVA, G. O., IGLESIA, M. D., ZHAO, W., USARY, J., BERG, J. S., ADAMS, M., BOOKER, J., WU, J. Y., GULABANI, A., BODENHEIMER, T., HOYLE, A. P., SIMONS, J. V., SOLOWAY, M. G., MOSE, L. E., JEFFERYS, S. R., BALU, S., PARKER, J. S., HAYES, D. N., PEROU, C. M., MALIK, S., MAHURKAR, S., SHEN, H., WEISENBERGER, D. J., TRICHE, T., et al. 2012. Comprehensive molecular portraits of human breast tumours. *Nature,* 490**,** 61-70.

KOCH, A., DENKHAUS, D., ALBRECHT, S., LEUSCHNER, I., VON SCHWEINITZ, D. & PIETSCH, T. 1999. Childhood hepatoblastomas frequently carry a mutated degradation targeting box of the beta-catenin gene. *Cancer Res,* 59**,** 269-73.

KOCHE, R. P., SMITH, Z. D., ADLI, M., GU, H., KU, M., GNIRKE, A., BERNSTEIN, B. E. & MEISSNER, A. 2011. Reprogramming factor expression initiates widespread targeted chromatin remodeling. *Cell Stem Cell,* 8**,** 96-105.

KOHLER, E. E., COWAN, C. E., CHATTERJEE, I., MALIK, A. B. & WARY, K. K. 2011. NANOG induction of fetal liver kinase-1 (FLK1) transcription regulates endothelial cell proliferation and angiogenesis. *Blood,* 117**,** 1761-9.

KONISHI, H., KARAKAS, B., ABUKHDEIR, A. M., LAURING, J., GUSTIN, J. P., GARAY, J. P., KONISHI, Y., GALLMEIER, E., BACHMAN, K. E. & PARK, B. H. 2007. Knock-in of mutant K-ras in nontumorigenic human epithelial cells as a new model for studying K-ras mediated transformation. *Cancer Res,* 67**,** 8460-7.

KOO, B. S., LEE, S. H., KIM, J. M., HUANG, S., KIM, S. H., RHO, Y. S., BAE, W. J., KANG, H. J., KIM, Y. S., MOON, J. H. & LIM, Y. C. 2015a. Oct4 is a critical regulator of stemness in head and neck squamous carcinoma cells. *Oncogene,* 34**,** 2317-24.

KOO, T., LEE, J. & KIM, J. S. 2015b. Measuring and Reducing Off-Target Activities of Programmable Nucleases Including CRISPR-Cas9. *Mol Cells,* 38**,** 475-81.

KORKAYA, H., LIU, S. & WICHA, M. S. 2011. Breast cancer stem cells, cytokine networks, and the tumor microenvironment. *J Clin Invest,* 121**,** 3804-9.

KOTINI, A. G., CHANG, C. J., BOUSSAAD, I., DELROW, J. J., DOLEZAL, E. K., NAGULAPALLY, A. B., PERNA, F., FISHBEIN, G. A., KLIMEK, V. M., HAWKINS, R. D., HUANGFU, D., MURRY, C. E., GRAUBERT, T., NIMER, S. D. & PAPAPETROU, E. P. 2015. Functional analysis of a chromosomal deletion associated with myelodysplastic syndromes using isogenic human induced pluripotent stem cells. *Nat Biotechnol,* 33**,** 646-55.

KOTINI, A. G., CHANG, C. J., CHOW, A., YUAN, H., HO, T. C., WANG, T., VORA, S., SOLOVYOV, A., HUSSER, C., OLSZEWSKA, M., TERUYA-FELDSTEIN, J., PERUMAL, D., KLIMEK, V. M., SPYRIDONIDIS, A., RAMPAL, R. K., SILVERMAN, L., REDDY, E. P., PAPAEMMANUIL, E., PAREKH, S., GREENBAUM, B. D., LESLIE, C. S., KHARAS, M. G. & PAPAPETROU, E. P. 2017. Stage-Specific Human Induced Pluripotent Stem Cells Map the Progression of Myeloid Transformation to Transplantable Leukemia. *Cell Stem Cell,* 20**,** 315-328 e7.

KOUZARIDES, T. 2007. Chromatin modifications and their function. *Cell,* 128**,** 693-705.

KRELL, J., FRAMPTON, A. E. & STEBBING, J. 2013. MicroRNAs in the cancer clinic. *Front Biosci (Elite Ed),* 5**,** 204-13.

KRESO, A. & DICK, J. E. 2014. Evolution of the cancer stem cell model. *Cell Stem Cell,* 14**,** 275-91.

KRIZHANOVSKY, V. & LOWE, S. W. 2009. Stem cells: The promises and perils of p53. *Nature,* 460**,** 1085-6.

KRUH, J. 1982. Effects of sodium butyrate, a new pharmacological agent, on cells in culture. *Mol Cell Biochem,* 42**,** 65-82.

KRUSE, J. P. & GU, W. 2008. SnapShot: p53 posttranslational modifications. *Cell,* 133**,** 930-30 e1.

KULIS, M., HEATH, S., BIBIKOVA, M., QUEIROS, A. C., NAVARRO, A., CLOT, G., MARTINEZ-TRILLOS, A., CASTELLANO, G., BRUN-HEATH, I., PINYOL, M., BARBERAN-SOLER, S., PAPASAIKAS, P., JARES, P., BEA, S., RICO, D., ECKER, S., RUBIO, M., ROYO, R., HO, V., KLOTZLE, B., HERNANDEZ, L., CONDE, L., LOPEZ-GUERRA, M., COLOMER, D., VILLAMOR, N., AYMERICH, M., ROZMAN, M., BAYES, M., GUT, M., GELPI, J. L., OROZCO, M., FAN, J. B., QUESADA, V., PUENTE, X. S., PISANO, D. G., VALENCIA, A., LOPEZ-GUILLERMO, A., GUT, I., LOPEZ-OTIN, C., CAMPO, E. & MARTIN-SUBERO, J. I. 2012. Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nat Genet,* 44**,** 1236-42.

KUMANO, K., ARAI, S., HOSOI, M., TAOKA, K., TAKAYAMA, N., OTSU, M., NAGAE, G., UEDA, K., NAKAZAKI, K., KAMIKUBO, Y., ETO, K., ABURATANI, H., NAKAUCHI, H. & KUROKAWA, M. 2012. Generation of induced pluripotent stem cells from primary chronic myelogenous leukemia patient samples. *Blood,* 119**,** 6234-42.

KUNATH, K., VON HARPE, A., FISCHER, D., PETERSON, H., BICKEL, U., VOIGT, K. & KISSEL, T. 2003. Low-molecular-weight polyethylenimine as a non-viral vector for DNA delivery: comparison of physicochemical properties, transfection efficiency and in vivo distribution with high-molecular-weight polyethylenimine. *Journal of Controlled Release,* 89**,** 113-125.

KUO, K. K., LEE, K. T., CHEN, K. K., YANG, Y. H., LIN, Y. C., TSAI, M. H., WUPUTRA, K., LEE, Y. L., KU, C. C., MIYOSHI, H., NAKAMURA, Y., SAITO, S., WU, C. C., CHAI, C. Y., ECKNER, R., STEVE LIN, C. L., WANG, S. S., WU, D. C., LIN, C. S. & YOKOYAMA, K. K. 2016. Positive Feedback Loop of OCT4 and c-JUN Expedites Cancer Stemness in Liver Cancer. *Stem Cells,* 34**,** 2613-2624.

KYTTALA, A., MORAGHEBI, R., VALENSISI, C., KETTUNEN, J., ANDRUS, C., PASUMARTHY, K. K., NAKANISHI, M., NISHIMURA, K., OHTAKA, M., WELTNER, J., VAN HANDEL, B., PARKKONEN, O., SINISALO, J., JALANKO, A., HAWKINS, R. D., WOODS, N. B., OTONKOSKI, T. & TROKOVIC, R. 2016. Genetic Variability Overrides the Impact of Parental Cell Type and Determines iPSC Differentiation Potential. *Stem Cell Reports,* 6**,** 200-212.

LAI, J., KONG, C. M., MAHALINGAM, D., XIE, X. & WANG, X. 2013. Elite model for the generation of induced pluripotent cancer cells (iPCs). *PLoS One,* 8**,** e56702.

LAKHANI, S. R., ELLIS. I.O., SCHNITT, S.J., TAN, P.H., VAN DE VIJVER, M.J. 2012. WHO Classification of Tumours of the Breast. Fourth Edition. *WHO.*

LAMOUILLE, S., XU, J. & DERYNCK, R. 2014. Molecular mechanisms of epithelial-mesenchymal transition. *Nature Reviews Molecular Cell Biology,* 15**,** 178-196.

LAW, J. A. & JACOBSEN, S. E. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet,* 11**,** 204-20.

LAWRENCE, M., DAUJAT, S. & SCHNEIDER, R. 2016. Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends Genet,* 32**,** 42-56.

LEE, D. F., SU, J., KIM, H. S., CHANG, B., PAPATSENKO, D., ZHAO, R., YUAN, Y., GINGOLD, J., XIA, W., DARR, H., MIRZAYANS, R., HUNG, M. C., SCHANIEL, C. & LEMISCHKA, I. R. 2015. Modeling familial cancer with induced pluripotent stem cells. *Cell,* 161**,** 240-54.

LEE, K. M., JU, J. H., JANG, K., YANG, W., YI, J. Y., NOH, D. Y. & SHIN, I. 2012. CD24 regulates cell proliferation and transforming growth factor beta-induced epithelial to mesenchymal transition through modulation of integrin beta1 stability. *Cell Signal,* 24**,** 2132-42.

LEHMANN-CHE, J., HAMY, A. S., PORCHER, R., BARRITAULT, M., BOUHIDEL, F., HABUELLELAH, H., LEMAN-DETOURS, S., DE ROQUANCOURT, A., CAHEN-DOIDY, L., BOURSTYN, E.,

DE CREMOUX, P., DE BAZELAIRE, C., ALBITER, M., GIACCHETTI, S., CUVIER, C., JANIN, A., ESPIE, M., DE THE, H. & BERTHEAU, P. 2013. Molecular apocrine breast cancers are aggressive estrogen receptor negative tumors overexpressing either HER2 or GCDFP15. *Breast Cancer Res,* 15**,** R37.

LEIS, O., EGUIARA, A., LOPEZ-ARRIBILLAGA, E., ALBERDI, M. J., HERNANDEZ-GARCIA, S., ELORRIAGA, K., PANDIELLA, A., REZOLA, R. & MARTIN, A. G. 2012. Sox2 expression in breast tumours and activation in breast cancer stem cells. *Oncogene,* 31**,** 1354-65.

LENNARTSSON, A. & EKWALL, K. 2009. Histone modification patterns and epigenetic codes. *Biochim Biophys Acta,* 1790**,** 863-8.

LI, D., LIU, J., YANG, X., ZHOU, C., GUO, J., WU, C., QIN, Y., GUO, L., HE, J., YU, S., LIU, H., WANG, X., WU, F., KUANG, J., HUTCHINS, A. P., CHEN, J. & PEI, D. 2017. Chromatin Accessibility Dynamics during iPSC Reprogramming. *Cell Stem Cell,* 21**,** 819-833 e6.

LI, H., COLLADO, M., VILLASANTE, A., STRATI, K., ORTEGA, S., CANAMERO, M., BLASCO, M. A. & SERRANO, M. 2009. The Ink4/Arf locus is a barrier for iPS cell reprogramming. *Nature,* 460**,** 1136-9.

LI, L., CONNELLY, M. C., WETMORE, C., CURRAN, T. & MORGAN, J. I. 2003. Mouse embryos cloned from brain tumors. *Cancer Res,* 63**,** 2733-6.

LI, M. H., FU, S. B. & XIAO, H. S. 2015a. Genome-wide analysis of microRNA and mRNA expression signatures in cancer. *Acta Pharmacol Sin,* 36**,** 1200-11.

LI, R., LIANG, J., NI, S., ZHOU, T., QING, X., LI, H., HE, W., CHEN, J., LI, F., ZHUANG, Q., QIN, B., XU, J., LI, W., YANG, J., GAN, Y., QIN, D., FENG, S., SONG, H., YANG, D., ZHANG, B., ZENG, L., LAI, L., ESTEBAN, M. A. & PEI, D. 2010. A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell,* 7**,** 51-63.

LI, S. Y., SUN, R., WANG, H. X., SHEN, S., LIU, Y., DU, X. J., ZHU, Y. H. & JUN, W. 2015b. Combination therapy with epigenetic-targeted and chemotherapeutic drugs delivered by nanoparticles to enhance the chemotherapy response and overcome resistance by breast cancer stem cells. *J Control Release,* 205**,** 7-14.

LI, Y., POWELL, S., BRUNETTE, E., LEBKOWSKI, J. & MANDALAM, R. 2005. Expansion of human embryonic stem cells in defined serum-free medium devoid of animal-derived products. *Biotechnol Bioeng,* 91**,** 688-98.

LI, Y., ZHANG, Y., LI, S., LU, J., CHEN, J., WANG, Y., LI, Y., XU, J. & LI, X. 2015c. Genome-wide DNA methylome analysis reveals epigenetically dysregulated non-coding RNAs in human breast cancer. *Sci Rep,* 5**,** 8790.

LIANG, G., TARANOVA, O., XIA, K. & ZHANG, Y. 2010. Butyrate promotes induced pluripotent stem cell generation. *J Biol Chem,* 285**,** 25516-21.

LIANG, G. & ZHANG, Y. 2013a. Embryonic stem cell and induced pluripotent stem cell: an epigenetic perspective. *Cell Res,* 23**,** 49-69.

LIANG, G. & ZHANG, Y. 2013b. Genetic and epigenetic variations in iPSCs: potential causes and implications for application. *Cell Stem Cell,* 13**,** 149-59.

LIAO, B., BAO, X., LIU, L., FENG, S., ZOVOILIS, A., LIU, W., XUE, Y., CAI, J., GUO, X., QIN, B., ZHANG, R., WU, J., LAI, L., TENG, M., NIU, L., ZHANG, B., ESTEBAN, M. A. & PEI, D. 2011. MicroRNA cluster 302-367 enhances somatic cell reprogramming by accelerating a mesenchymal-to-epithelial transition. *J Biol Chem,* 286**,** 17359-64.

LIAO, X. H., LI, Y. Q., WANG, N., ZHENG, L., XING, W. J., ZHAO, D. W., YAN, T. B., WANG, Y., LIU, L. Y., SUN, X. G., HU, P., ZHOU, H. & ZHANG, T. C. 2014. Re-expression and epigenetic modification of maspin induced apoptosis in MCF-7 cells mediated by myocardin. *Cell Signal,* 26**,** 1335-46.

LIEDTKE, S., ENCZMANN, J., WACLAWCZYK, S., WERNET, P. & KOGLER, G. 2007. Oct4 and its pseudogenes confuse stem cell research. *Cell Stem Cell,* 1**,** 364-6.

LIM, E., VAILLANT, F., WU, D., FORREST, N. C., PAL, B., HART, A. H., ASSELIN-LABAT, M. L., GYORKI, D. E., WARD, T., PARTANEN, A., FELEPPA, F., HUSCHTSCHA, L. I., THORNE, H. J., KCONFAB, FOX, S. B., YAN, M., FRENCH, J. D., BROWN, M. A., SMYTH, G. K., VISVADER, J. E. & LINDEMAN, G. J. 2009. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat Med,* 15**,** 907-13.

LIN, C. H., JACKSON, A. L., GUO, J., LINSLEY, P. S. & EISENMAN, R. N. 2009. Myc-regulated microRNAs attenuate embryonic stem cell differentiation. *EMBO J,* 28**,** 3157-70.

LIN, S. L., CHANG, D. C., CHANG-LIN, S., LIN, C. H., WU, D. T., CHEN, D. T. & YING, S. Y. 2008. Mir-302 reprograms human skin cancer cells into a pluripotent ES-cell-like state. *RNA,* 14**,** 2115-24.

LIN, S. L., CHANG, D. C., LIN, C. H., YING, S. Y., LEU, D. & WU, D. T. 2011. Regulation of somatic cell reprogramming through inducible mir-302 expression. *Nucleic Acids Res,* 39**,** 1054-65.

LING, G. Q., CHEN, D. B., WANG, B. Q. & ZHANG, L. S. 2012. Expression of the pluripotency markers Oct3/4, Nanog and Sox2 in human breast cancer cell lines. *Oncol Lett,* 4**,** 1264-1268.

LING, H., GIRNITA, L., BUDA, O. & CALIN, G. A. 2017. Non-coding RNAs: the cancer genome dark matter that matters! *Clin Chem Lab Med,* 55**,** 705-714.

LIPCHINA, I., ELKABETZ, Y., HAFNER, M., SHERIDAN, R., MIHAILOVIC, A., TUSCHL, T., SANDER, C., STUDER, L. & BETEL, D. 2011. Genome-wide identification of microRNA targets in human ES cells reveals a role for miR-302 in modulating BMP response. *Genes Dev,* 25**,** 2173-86.

LISTER, R., PELIZZOLA, M., DOWEN, R. H., HAWKINS, R. D., HON, G., TONTI-FILIPPINI, J., NERY, J. R., LEE, L., YE, Z., NGO, Q. M., EDSALL, L., ANTOSIEWICZ-BOURGET, J., STEWART, R., RUOTTI, V., MILLAR, A. H., THOMSON, J. A., REN, B. & ECKER, J. R. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature,* 462**,** 315-22.

LIU, T., SUN, B., ZHAO, X., LI, Y., GU, Q., DONG, X. & LIU, F. 2014a. OCT4 expression and vasculogenic mimicry formation positively correlate with poor prognosis in human breast cancer. *Int J Mol Sci,* 15**,** 19634-49.

LIU, Y., CHENG, H., GAO, S., LU, X., HE, F., HU, L., HOU, D., ZOU, Z., LI, Y., ZHANG, H., XU, J., KANG, L., WANG, Q., YUAN, W., GAO, S. & CHENG, T. 2014b. Reprogramming of MLL-AF9 leukemia cells into pluripotent stem cells. *Leukemia,* 28**,** 1071-80.

LIU, Y., NENUTIL, R., APPLEYARD, M. V., MURRAY, K., BOYLAN, M., THOMPSON, A. M. & COATES, P. J. 2014c. Lack of correlation of stem cell markers in breast cancer stem cells. *British Journal of Cancer,* 110**,** 2063-2071.

LIU, Y., NENUTIL, R., APPLEYARD, M. V., MURRAY, K., BOYLAN, M., THOMPSON, A. M. & COATES, P. J. 2014d. Lack of correlation of stem cell markers in breast cancer stem cells. *Br J Cancer,* 110**,** 2063-71.

LLAMES, S., GARCIA-PEREZ, E., MEANA, A., LARCHER, F. & DEL RIO, M. 2015. Feeder Layer Cell Actions and Applications. *Tissue Engineering Part B-Reviews,* 21**,** 345-353.

LOBO, N. A., SHIMONO, Y., QIAN, D. & CLARKE, M. F. 2007. The biology of cancer stem cells. *Annu Rev Cell Dev Biol,* 23**,** 675-99.

LOEB, L. A. 2011. Human cancers express mutator phenotypes: origin, consequences and targeting. *Nat Rev Cancer,* 11**,** 450-7.

LOWRY, W. E., RICHTER, L., YACHECHKO, R., PYLE, A. D., TCHIEU, J., SRIDHARAN, R., CLARK, A. T. & PLATH, K. 2008. Generation of human induced pluripotent stem cells from dermal fibroblasts. *Proc Natl Acad Sci U S A,* 105**,** 2883-8.

LU, J., GETZ, G., MISKA, E. A., ALVAREZ-SAAVEDRA, E., LAMB, J., PECK, D., SWEET-CORDERO, A., EBERT, B. L., MAK, R. H., FERRANDO, A. A., DOWNING, J. R., JACKS, T., HORVITZ,

H. R. & GOLUB, T. R. 2005. MicroRNA expression profiles classify human cancers. *Nature,* 435**,** 834-8.

LU, J., HOU, R., BOOTH, C. J., YANG, S. H. & SNYDER, M. 2006. Defined culture conditions of human embryonic stem cells. *Proc Natl Acad Sci U S A,* 103**,** 5688-93.

LU, X., MAZUR, S. J., LIN, T., APPELLA, E. & XU, Y. 2014. The pluripotency factor nanog promotes breast cancer tumorigenesis and metastasis. *Oncogene,* 33**,** 2655-64.

LU, X. & ZHAO, T. 2013. Clinical therapy using iPSCs: hopes and challenges. *Genomics Proteomics Bioinformatics,* 11**,** 294-8.

LUNDBERG, I. V., EDIN, S., EKLOF, V., OBERG, A., PALMQVIST, R. & WIKBERG, M. L. 2016. SOX2 expression is associated with a cancer stem cell state and down-regulation of CDX2 in colorectal cancer. *BMC Cancer,* 16**,** 471.

LUO, J. 2016. CRISPR/Cas9: From Genome Engineering to Cancer Drug Discovery. *Trends Cancer,* 2**,** 313-324.

LUSTBERG, M. B. & RAMASWAMY, B. 2011. Epigenetic Therapy in Breast Cancer. *Curr Breast Cancer Rep,* 3**,** 34-43.

LYKO, F. 2017. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nat Rev Genet*.

MA, H., MOREY, R., O'NEIL, R. C., HE, Y. P., DAUGHTRY, B., SCHULTZ, M. D., HARIHARAN, M., NERY, J. R., CASTANON, R., SABATINI, K., THIAGARAJAN, R. D., TACHIBANA, M., KANG, E. J., TIPPNER-HEDGES, R., AHMED, R., GUTIERREZ, N. M., VAN DYKEN, C., POLAT, A., SUGAWARA, A., SPARMAN, M., GOKHALE, S., AMATO, P., WOLF, D. P., ECKER, J. R., LAURENT, L. C. & MITALIPOV, S. 2014. Abnormalities in human pluripotent cells due to reprogramming mechanisms. *Nature,* 511**,** 177-+.

MA, T., XIE, M., LAURENT, T. & DING, S. 2013. Progress in the reprogramming of somatic cells. *Circ Res,* 112**,** 562-74.

MADDIPATI, R. & KATZ, J. P. 2016. KLF4 Initiates Acinar Cell Reprogramming and Is Essential for the Early Stages of Pancreatic Carcinogenesis. *Cancer Cell,* 29**,** 247-8.

MAHALINGAM, D., KONG, C. M., LAI, J., TAY, L. L., YANG, H. & WANG, X. 2012. Reversal of aberrant cancer methylome and transcriptome upon direct reprogramming of lung cancer cells. *Sci Rep,* 2**,** 592.

MAHERALI, N., SRIDHARAN, R., XIE, W., UTIKAL, J., EMINLI, S., ARNOLD, K., STADTFELD, M., YACHECHKO, R., TCHIEU, J., JAENISCH, R., PLATH, K. & HOCHEDLINGER, K. 2007. Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell,* 1**,** 55-70.

MALHOTRA, G. K., ZHAO, X., BAND, H. & BAND, V. 2010. Histological, molecular and functional subtypes of breast cancers. *Cancer Biol Ther,* 10**,** 955-60.

MALI, P., CHOU, B. K., YEN, J., YE, Z., ZOU, J., DOWEY, S., BRODSKY, R. A., OHM, J. E., YU, W., BAYLIN, S. B., YUSA, K., BRADLEY, A., MEYERS, D. J., MUKHERJEE, C., COLE, P. A. & CHENG, L. 2010. Butyrate greatly enhances derivation of human induced pluripotent stem cells by promoting epigenetic remodeling and the expression of pluripotency-associated genes. *Stem Cells,* 28**,** 713-20.

MALI, P., YANG, L., ESVELT, K. M., AACH, J., GUELL, M., DICARLO, J. E., NORVILLE, J. E. & CHURCH, G. M. 2013. RNA-guided human genome engineering via Cas9. *Science,* 339**,** 823-6.

MALKIN, D., LI, F. P., STRONG, L. C., FRAUMENI, J. F., JR., NELSON, C. E., KIM, D. H., KASSEL, J., GRYKA, M. A., BISCHOFF, F. Z., TAINSKY, M. A. & ET AL. 1990. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science,* 250**,** 1233-8.

MANI, S. A., GUO, W., LIAO, M. J., EATON, E. N., AYYANAN, A., ZHOU, A. Y., BROOKS, M., REINHARD, F., ZHANG, C. C., SHIPITSIN, M., CAMPBELL, L. L., POLYAK, K., BRISKEN, C.,

YANG, J. & WEINBERG, R. A. 2008. The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell,* 133**,** 704-15.

MARCHETTO, M. C., YEO, G. W., KAINOHANA, O., MARSALA, M., GAGE, F. H. & MUOTRI, A. R. 2009. Transcriptional signature and memory retention of human-induced pluripotent stem cells. *PLoS One,* 4**,** e7076.

MARELLA, N. V., MALYAVANTHAM, K. S., WANG, J., MATSUI, S., LIANG, P. & BEREZNEY, R. 2009. Cytogenetic and cDNA microarray expression analysis of MCF10 human breast cancer progression cell lines. *Cancer Res,* 69**,** 5946-53.

MARGUERON, R. & REINBERG, D. 2011. The Polycomb complex PRC2 and its mark in life. *Nature,* 469**,** 343-9.

MARION, R. M., STRATI, K., LI, H., MURGA, M., BLANCO, R., ORTEGA, S., FERNANDEZ-CAPETILLO, O., SERRANO, M. & BLASCO, M. A. 2009. A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. *Nature,* 460**,** 1149-53.

MARTI, M., MULERO, L., PARDO, C., MORERA, C., CARRIO, M., LARICCHIA-ROBBIO, L., ESTEBAN, C. R. & IZPISUA BELMONTE, J. C. 2013. Characterization of pluripotent stem cells. *Nat Protoc,* 8**,** 223-53.

MARTINEZ-GARCIA, E., POPOVIC, R., MIN, D. J., SWEET, S. M., THOMAS, P. M., ZAMDBORG, L., HEFFNER, A., WILL, C., LAMY, L., STAUDT, L. M., LEVENS, D. L., KELLEHER, N. L. & LICHT, J. D. 2011. The MMSET histone methyl transferase switches global histone methylation and alters gene expression in t(4;14) multiple myeloma cells. *Blood,* 117**,** 211-20.

MARUSYK, A., ALMENDRO, V. & POLYAK, K. 2012. Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer,* 12**,** 323-34.

MATAGA, M. A., ROSENTHAL, S., HEERBOTH, S., DEVALAPALLI, A., KOKOLUS, S., EVANS, L. R., LONGACRE, M., HOUSMAN, G. & SARKAR, S. 2012. Anti-breast cancer effects of histone deacetylase inhibitors and calpain inhibitor. *Anticancer Res,* 32**,** 2523-9.

MATANO, M., DATE, S., SHIMOKAWA, M., TAKANO, A., FUJII, M., OHTA, Y., WATANABE, T., KANAI, T. & SATO, T. 2015. Modeling colorectal cancer using CRISPR-Cas9-mediated engineering of human intestinal organoids. *Nat Med,* 21**,** 256-62.

MATHIEU, J. & RUOHOLA-BAKER, H. 2017. Metabolic remodeling during the loss and acquisition of pluripotency. *Development,* 144**,** 541-551.

MATHIEU, J., ZHANG, Z., ZHOU, W., WANG, A. J., HEDDLESTON, J. M., PINNA, C. M., HUBAUD, A., STADLER, B., CHOI, M., BAR, M., TEWARI, M., LIU, A., VESSELLA, R., ROSTOMILY, R., BORN, D., HORWITZ, M., WARE, C., BLAU, C. A., CLEARY, M. A., RICH, J. N. & RUOHOLA-BAKER, H. 2011. HIF induces human embryonic stem cell markers in cancer cells. *Cancer Res,* 71**,** 4640-52.

MATTICK, J. S. & MAKUNIN, I. V. 2006. Non-coding RNA. *Hum Mol Genet,* 15 Spec No 1**,** R17-29.

MATTOUT, A., BIRAN, A. & MESHORER, E. 2011. Global epigenetic changes during somatic cell reprogramming to iPS cells. *J Mol Cell Biol,* 3**,** 341-50.

MCKINNELL, R. G., DEGGINS, B. A. & LABAT, D. D. 1969. Transplantation of pluripotential nuclei from triploid frog tumors. *Science,* 165**,** 394-6.

MCWILLIAM, H., LI, W., ULUDAG, M., SQUIZZATO, S., PARK, Y. M., BUSO, N., COWLEY, A. P. & LOPEZ, R. 2013. Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res,* 41**,** W597-600.

MEACHAM, C. E. & MORRISON, S. J. 2013. Tumour heterogeneity and cancer cell plasticity. *Nature,* 501**,** 328-37.

MEDEMA, J. P. 2013. Cancer stem cells: the challenges ahead. *Nat Cell Biol,* 15**,** 338-44.

MEINDL, A., HELLEBRAND, H., WIEK, C., ERVEN, V., WAPPENSCHMIDT, B., NIEDERACHER, D., FREUND, M., LICHTNER, P., HARTMANN, L., SCHAAL, H., RAMSER, J., HONISCH, E.,

KUBISCH, C., WICHMANN, H. E., KAST, K., DEISSLER, H., ENGEL, C., MULLER-MYHSOK, B., NEVELING, K., KIECHLE, M., MATHEW, C. G., SCHINDLER, D., SCHMUTZLER, R. K. & HANENBERG, H. 2010. Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. *Nat Genet,* 42**,** 410-4.

MEKALA, J. R., NAUSHAD, S. M., PONNUSAMY, L., ARIVAZHAGAN, G., SAKTHIPRASAD, V. & PAL-BHADRA, M. 2018. Epigenetic regulation of miR-200 as the potential strategy for the therapy against triple-negative breast cancer. *Gene,* 641**,** 248-258.

MELTON, C., JUDSON, R. L. & BLELLOCH, R. 2010. Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature,* 463**,** 621-6.

MENG, X., NEISES, A., SU, R. J., PAYNE, K. J., RITTER, L., GRIDLEY, D. S., WANG, J., SHENG, M., LAU, K. H., BAYLINK, D. J. & ZHANG, X. B. 2012. Efficient reprogramming of human cord blood CD34+ cells into induced pluripotent stem cells with OCT4 and SOX2 alone. *Mol Ther,* 20**,** 408-16.

MENSCHIKOWSKI, M., HAGELGANS, A., NACKE, B., JANDECK, C., SUKOCHEVA, O. & SIEGERT, G. 2015. Epigenetic control of phospholipase A2 receptor expression in mammary cancer cells. *BMC Cancer,* 15**,** 971.

MEYER, M. J., FLEMING, J. M., LIN, A. F., HUSSNAIN, S. A., GINSBURG, E. & VONDERHAAR, B. K. 2010. CD44posCD49fhiCD133/2hi defines xenograft-initiating cells in estrogen receptor-negative breast cancer. *Cancer Res,* 70**,** 4624-33.

MEZENCEV, R., MATYUNINA, L. V., JABBARI, N. & MCDONALD, J. F. 2016. Snail-induced epithelial-to-mesenchymal transition of MCF-7 breast cancer cells: systems analysis of molecular changes and their effect on radiation and drug sensitivity. *BMC Cancer,* 16**,** 236.

MIKKELSEN, T. S., HANNA, J., ZHANG, X., KU, M., WERNIG, M., SCHORDERET, P., BERNSTEIN, B. E., JAENISCH, R., LANDER, E. S. & MEISSNER, A. 2008a. Dissecting direct reprogramming through integrative genomic analysis. *Nature,* 454**,** 49-55.

MIKKELSEN, T. S., HANNA, J., ZHANG, X. L., KU, M. C., WERNIG, M., SCHORDERET, P., BERNSTEIN, B. E., JAENISCH, R., LANDER, E. S. & MEISSNER, A. 2008b. Dissecting direct reprogramming through integrative genomic analysis. *Nature,* 454**,** 49-U1.

MIYOSHI, N., ISHII, H., NAGAI, K., HOSHINO, H., MIMORI, K., TANAKA, F., NAGANO, H., SEKIMOTO, M., DOKI, Y. & MORI, M. 2010a. Defined factors induce reprogramming of gastrointestinal cancer cells. *Proc Natl Acad Sci U S A,* 107**,** 40-5.

MIYOSHI, N., ISHII, H., NAGAI, K., HOSHINO, H., MIMORI, K., TANAKA, F., NAGANO, H., SEKIMOTO, M., DOKI, Y. & MORI, M. 2010b. Defined factors induce reprogramming of gastrointestinal cancer cells. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 40-45.

MIYOSHI, N., ISHII, H., NAGANO, H., HARAGUCHI, N., DEWI, D. L., KANO, Y., NISHIKAWA, S., TANEMURA, M., MIMORI, K., TANAKA, F., SAITO, T., NISHIMURA, J., TAKEMASA, I., MIZUSHIMA, T., IKEDA, M., YAMAMOTO, H., SEKIMOTO, M., DOKI, Y. & MORI, M. 2011. Reprogramming of mouse and human cells to pluripotency using mature microRNAs. *Cell Stem Cell,* 8**,** 633-8.

MOLYNEUX, G., GEYER, F. C., MAGNAY, F. A., MCCARTHY, A., KENDRICK, H., NATRAJAN, R., MACKAY, A., GRIGORIADIS, A., TUTT, A., ASHWORTH, A., REIS-FILHO, J. S. & SMALLEY, M. J. 2010. BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. *Cell Stem Cell,* 7**,** 403-17.

MORATH, I., HARTMANN, T. N. & ORIAN-ROUSSEAU, V. 2016. CD44: More than a mere stem cell marker. *International Journal of Biochemistry & Cell Biology,* 81**,** 166-173.

MORIS, N., PINA, C. & ARIAS, A. M. 2016. Transition states and cell fate decisions in epigenetic landscapes. *Nat Rev Genet,* 17**,** 693-703.

MOU, H., KENNEDY, Z., ANDERSON, D. G., YIN, H. & XUE, W. 2015. Precision cancer mouse models through genome editing with CRISPR-Cas9. *Genome Med,* 7**,** 53.

MOUTINHO, C. & ESTELLER, M. 2017. MicroRNAs and Epigenetics. *Adv Cancer Res,* 135**,** 189-220.

MUELLER, D., BACH, C., ZEISIG, D., GARCIA-CUELLAR, M. P., MONROE, S., SREEKUMAR, A., ZHOU, R., NESVIZHSKII, A., CHINNAIYAN, A., HESS, J. L. & SLANY, R. K. 2007. A role for the MLL fusion partner ENL in transcriptional elongation and chromatin modification. *Blood,* 110**,** 4445-54.

MULLER, F. J., GOLDMANN, J., LOSER, P. & LORING, J. F. 2010. A Call to Standardize Teratoma Assays Used to Define Human Pluripotent Cell Lines. *Cell Stem Cell,* 6**,** 412-414.

MULLER, F. J., SCHULDT, B. M., WILLIAMS, R., MASON, D., ALTUN, G., PAPAPETROU, E. P., DANNER, S., GOLDMANN, J. E., HERBST, A., SCHMIDT, N. O., ALDENHOFF, J. B., LAURENT, L. C. & LORING, J. F. 2011. A bioinformatic assay for pluripotency in human cells. *Nat Methods,* 8**,** 315-7.

MULLER, L. U., MILSOM, M. D., HARRIS, C. E., VYAS, R., BRUMME, K. M., PARMAR, K., MOREAU, L. A., SCHAMBACH, A., PARK, I. H., LONDON, W. B., STRAIT, K., SCHLAEGER, T., DEVINE, A. L., GRASSMAN, E., D'ANDREA, A., DALEY, G. Q. & WILLIAMS, D. A. 2012. Overcoming reprogramming resistance of Fanconi anemia cells. *Blood,* 119**,** 5449-57.

MUNOZ-LOPEZ, A., ROMERO-MOYA, D., PRIETO, C., RAMOS-MEJIA, V., AGRAZ-DOBLAS, A., VARELA, I., BUSCHBECK, M., PALAU, A., CARVAJAL-VERGARA, X., GIORGETTI, A., FORD, A., LAKO, M., GRANADA, I., RUIZ-XIVILLE, N., RODRIGUEZ-PERALES, S., TORRES-RUIZ, R., STAM, R. W., FUSTER, J. L., FRAGA, M. F., NAKANISHI, M., CAZZANIGA, G., BARDINI, M., COBO, I., BAYON, G. F., FERNANDEZ, A. F., BUENO, C. & MENENDEZ, P. 2016. Development Refractoriness of MLL-Rearranged Human B Cell Acute Leukemias to Reprogramming into Pluripotency. *Stem Cell Reports,* 7**,** 602-618.

MUVARAK, N. E., CHOWDHURY, K., XIA, L. M., ROBERT, C., CHOI, E. Y., CAI, Y., BELLANI, M., ZOU, Y., SINGH, Z. N., DUONG, V. H., RUTHERFORD, T., NAGARIA, P., BENTZEN, S. M., SEIDMAN, M. M., BAER, M. R., LAPIDUS, R. G., BAYLIN, S. B. & RASSOO, F. V. 2016. Enhancing the Cytotoxic Effects of PARP Inhibitors with DNA Demethylating Agents - A Potential Therapy for Cancer. *Cancer Cell,* 30**,** 637-650.

NAGAI, K., ISHII, H., MIYOSHI, N., HOSHINO, H., SAITO, T., SATO, T., TOMIMARU, Y., KOBAYASHI, S., NAGANO, H., SEKIMOTO, M., DOKI, Y. & MORI, M. 2010. Long-term culture following ES-like gene-induced reprogramming elicits an aggressive phenotype in mutated cholangiocellular carcinoma cells. *Biochem Biophys Res Commun,* 395**,** 258-63.

NAKAGAWA, M., TAKIZAWA, N., NARITA, M., ICHISAKA, T. & YAMANAKA, S. 2010. Promotion of direct reprogramming by transformation-deficient Myc. *Proc Natl Acad Sci U S A,* 107**,** 14152-7.

NAKANO, T., ANDO, S., TAKATA, N., KAWADA, M., MUGURUMA, K., SEKIGUCHI, K., SAITO, K., YONEMURA, S., EIRAKU, M. & SASAI, Y. 2012. Self-formation of optic cups and storable stratified neural retina from human ESCs. *Cell Stem Cell,* 10**,** 771-85.

NANBO, A., SUGDEN, A. & SUGDEN, B. 2007. The coupling of synthesis and partitioning of EBV's plasmid replicon is revealed in live cells. *EMBO J,* 26**,** 4252-62.

NARLIKAR, G. J., SUNDARAMOORTHY, R. & OWEN-HUGHES, T. 2013. Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell,* 154**,** 490-503.

NARSINH, K. H., JIA, F., ROBBINS, R. C., KAY, M. A., LONGAKER, M. T. & WU, J. C. 2011. Generation of adult human induced pluripotent stem cells using nonviral minicircle DNA vectors. *Nat Protoc,* 6**,** 78-88.

NASHUN, B., HILL, P. W. S. & HAJKOVA, P. 2015. Reprogramming of cell fate: epigenetic memory and the erasure of memories past. *Embo Journal,* 34**,** 1296-1308.

NAVARRO, S., MOLEIRO, V., MOLINA-ESTEVEZ, F. J., LOZANO, M. L., CHINCHON, R., ALMARZA, E., QUINTANA-BUSTAMANTE, O., MOSTOSLAVSKY, G., MAETZIG, T., GALLA, M.,

HEINZ, N., SCHIEDLMEIER, B., TORRES, Y., MODLICH, U., SAMPER, E., RIO, P., SEGOVIA, J. C., RAYA, A., GUENECHEA, G., IZPISUA-BELMONTE, J. C. & BUEREN, J. A. 2014. Generation of iPSCs from genetically corrected Brca2 hypomorphic cells: implications in cell reprogramming and stem cell therapy. *Stem Cells,* 32**,** 436-46.

NAVIN, N., KENDALL, J., TROGE, J., ANDREWS, P., RODGERS, L., MCINDOO, J., COOK, K., STEPANSKY, A., LEVY, D., ESPOSITO, D., MUTHUSWAMY, L., KRASNITZ, A., MCCOMBIE, W. R., HICKS, J. & WIGLER, M. 2011. Tumour evolution inferred by single-cell sequencing. *Nature,* 472**,** 90-4.

NAWY, T. 2016. STEM CELLS. Reprogrammed cells leave their past lives behind. *Nat Methods,* 13**,** 292.

NEVEU, P., KYE, M. J., QI, S., BUCHHOLZ, D. E., CLEGG, D. O., SAHIN, M., PARK, I. H., KIM, K. S., DALEY, G. Q., KORNBLUM, H. I., SHRAIMAN, B. I. & KOSIK, K. S. 2010. MicroRNA profiling reveals two distinct p53-related human pluripotent stem cell states. *Cell Stem Cell,* 7**,** 671-81.

NG, E. K., LI, R., SHIN, V. Y., JIN, H. C., LEUNG, C. P., MA, E. S., PANG, R., CHUA, D., CHU, K. M., LAW, W. L., LAW, S. Y., POON, R. T. & KWONG, A. 2013. Circulating microRNAs as specific biomarkers for breast cancer detection. *PLoS One,* 8**,** e53141.

NG, P. C. & HENIKOFF, S. 2003. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res,* 31**,** 3812-4.

NIK-ZAINAL, S., ALEXANDROV, L. B., WEDGE, D. C., VAN LOO, P., GREENMAN, C. D., RAINE, K., JONES, D., HINTON, J., MARSHALL, J., STEBBINGS, L. A., MENZIES, A., MARTIN, S., LEUNG, K., CHEN, L., LEROY, C., RAMAKRISHNA, M., RANCE, R., LAU, K. W., MUDIE, L. J., VARELA, I., MCBRIDE, D. J., BIGNELL, G. R., COOKE, S. L., SHLIEN, A., GAMBLE, J., WHITMORE, I., MADDISON, M., TARPEY, P. S., DAVIES, H. R., PAPAEMMANUIL, E., STEPHENS, P. J., MCLAREN, S., BUTLER, A. P., TEAGUE, J. W., JONSSON, G., GARBER, J. E., SILVER, D., MIRON, P., FATIMA, A., BOYAULT, S., LANGEROD, A., TUTT, A., MARTENS, J. W., APARICIO, S. A., BORG, A., SALOMON, A. V., THOMAS, G., BORRESEN-DALE, A. L., RICHARDSON, A. L., NEUBERGER, M. S., FUTREAL, P. A., CAMPBELL, P. J., STRATTON, M. R. & BREAST CANCER WORKING GROUP OF THE INTERNATIONAL CANCER GENOME, C. 2012a. Mutational processes molding the genomes of 21 breast cancers. *Cell,* 149**,** 979-93.

NIK-ZAINAL, S., DAVIES, H., STAAF, J., RAMAKRISHNA, M., GLODZIK, D., ZOU, X., MARTINCORENA, I., ALEXANDROV, L. B., MARTIN, S., WEDGE, D. C., VAN LOO, P., JU, Y. S., SMID, M., BRINKMAN, A. B., MORGANELLA, S., AURE, M. R., LINGJAERDE, O. C., LANGEROD, A., RINGNER, M., AHN, S. M., BOYAULT, S., BROCK, J. E., BROEKS, A., BUTLER, A., DESMEDT, C., DIRIX, L., DRONOV, S., FATIMA, A., FOEKENS, J. A., GERSTUNG, M., HOOIJER, G. K., JANG, S. J., JONES, D. R., KIM, H. Y., KING, T. A., KRISHNAMURTHY, S., LEE, H. J., LEE, J. Y., LI, Y., MCLAREN, S., MENZIES, A., MUSTONEN, V., O'MEARA, S., PAUPORTE, I., PIVOT, X., PURDIE, C. A., RAINE, K., RAMAKRISHNAN, K., RODRIGUEZ-GONZALEZ, F. G., ROMIEU, G., SIEUWERTS, A. M., SIMPSON, P. T., SHEPHERD, R., STEBBINGS, L., STEFANSSON, O. A., TEAGUE, J., TOMMASI, S., TREILLEUX, I., VAN DEN EYNDEN, G. G., VERMEULEN, P., VINCENT-SALOMON, A., YATES, L., CALDAS, C., VAN'T VEER, L., TUTT, A., KNAPPSKOG, S., TAN, B. K., JONKERS, J., BORG, A., UENO, N. T., SOTIRIOU, C., VIARI, A., FUTREAL, P. A., CAMPBELL, P. J., SPAN, P. N., VAN LAERE, S., LAKHANI, S. R., EYFJORD, J. E., THOMPSON, A. M., BIRNEY, E., STUNNENBERG, H. G., VAN DE VIJVER, M. J., MARTENS, J. W., BORRESEN-DALE, A. L., RICHARDSON, A. L., KONG, G., THOMAS, G. & STRATTON, M. R. 2016. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature,* 534**,** 47-54.

NIK-ZAINAL, S., VAN LOO, P., WEDGE, D. C., ALEXANDROV, L. B., GREENMAN, C. D., LAU, K. W., RAINE, K., JONES, D., MARSHALL, J., RAMAKRISHNA, M., SHLIEN, A., COOKE, S. L.,

HINTON, J., MENZIES, A., STEBBINGS, L. A., LEROY, C., JIA, M., RANCE, R., MUDIE, L. J., GAMBLE, S. J., STEPHENS, P. J., MCLAREN, S., TARPEY, P. S., PAPAEMMANUIL, E., DAVIES, H. R., VARELA, I., MCBRIDE, D. J., BIGNELL, G. R., LEUNG, K., BUTLER, A. P., TEAGUE, J. W., MARTIN, S., JONSSON, G., MARIANI, O., BOYAULT, S., MIRON, P., FATIMA, A., LANGEROD, A., APARICIO, S. A., TUTT, A., SIEUWERTS, A. M., BORG, A., THOMAS, G., SALOMON, A. V., RICHARDSON, A. L., BORRESEN-DALE, A. L., FUTREAL, P. A., STRATTON, M. R., CAMPBELL, P. J. & BREAST CANCER WORKING GROUP OF THE INTERNATIONAL CANCER GENOME, C. 2012b. The life history of 21 breast cancers. *Cell,* 149**,** 994-1007.

NISHI, M., AKUTSU, H., KUDOH, A., KIMURA, H., YAMAMOTO, N., UMEZAWA, A., LEE, S. W. & RYO, A. 2014a. Induced cancer stem-like cells as a model for biological screening and discovery of agents targeting phenotypic traits of cancer stem cell. *Oncotarget,* 5**,** 8665-80.

NISHI, M., SAKAI, Y., AKUTSU, H., NAGASHIMA, Y., QUINN, G., MASUI, S., KIMURA, H., PERREM, K., UMEZAWA, A., YAMAMOTO, N., LEE, S. W. & RYO, A. 2014b. Induction of cells with cancer stem cell properties from nontumorigenic human mammary epithelial cells by defined reprogramming factors. *Oncogene,* 33**,** 643-52.

NISHIMURA, K., AIZAWA, S., NUGROHO, F. L., SHIOMITSU, E., TRAN, Y. T., BUI, P. L., BORISOVA, E., SAKURAGI, Y., TAKADA, H., KURISAKI, A., HAYASHI, Y., FUKUDA, A., NAKANISHI, M. & HISATAKE, K. 2017. A Role for KLF4 in Promoting the Metabolic Shift via TCL1 during Induced Pluripotent Stem Cell Generation. *Stem Cell Reports,* 8**,** 787-801.

NOBLE, D. 2015. Conrad Waddington and the origin of epigenetics. *J Exp Biol,* 218**,** 816-8.

NODA, M., KITAYAMA, H., MATSUZAKI, T., SUGIMOTO, Y., OKAYAMA, H., BASSIN, R. H. & IKAWA, Y. 1989. Detection of genes with a potential for suppressing the transformed phenotype associated with activated ras genes. *Proc Natl Acad Sci U S A,* 86**,** 162-6.

NOGUCHI, K., EGUCHI, H., KONNO, M., KAWAMOTO, K., NISHIDA, N., KOSEKI, J., WADA, H., MARUBASHI, S., NAGANO, H., DOKI, Y., MORI, M. & ISHII, H. 2015. Susceptibility of pancreatic cancer stem cells to reprogramming. *Cancer Sci,* 106**,** 1182-7.

NOSHO, K., KAWASAKI, T., OHNISHI, M., SUEMOTO, Y., KIRKNER, G. J., ZEPF, D., YAN, L., LONGTINE, J. A., FUCHS, C. S. & OGINO, S. 2008. PIK3CA mutation in colorectal cancer: relationship with genetic and epigenetic alterations. *Neoplasia,* 10**,** 534-41.

NUSSE, R. 2005. Wnt signaling in disease and in development. *Cell Res,* 15**,** 28-32.

O'BRIEN, K. M., COLE, S. R., TSE, C. K., PEROU, C. M., CAREY, L. A., FOULKES, W. D., DRESSLER, L. G., GERADTS, J. & MILLIKAN, R. C. 2010. Intrinsic breast tumor subtypes, race, and long-term survival in the Carolina Breast Cancer Study. *Clin Cancer Res,* 16**,** 6100-10.

O'LEARY, T., HEINDRYCKX, B., LIERMAN, S., VAN BRUGGEN, D., GOEMAN, J. J., VANDEWOESTYNE, M., DEFORCE, D., DE SOUSA LOPES, S. M. & DE SUTTER, P. 2012. Tracking the progression of the human inner cell mass during embryonic stem cell derivation. *Nat Biotechnol,* 30**,** 278-82.

O'MALLEY, J., SKYLAKI, S., IWABUCHI, K. A., CHANTZOURA, E., RUETZ, T., JOHNSSON, A., TOMLINSON, S. R., LINNARSSON, S. & KAJI, K. 2013. High-resolution analysis with novel cell-surface markers identifies routes to iPS cells. *Nature,* 499**,** 88-91.

OHI, Y., QIN, H., HONG, C. B., BLOUIN, L., POLO, J. M., GUO, T. X., QI, Z. X., DOWNEY, S. L., MANOS, P. D., ROSSI, D. J., YU, J. W., HEBROK, M., HOCHEDLINGER, K., COSTELLO, J. F., SONG, J. S. & RAMALHO-SANTOS, M. 2011. Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. *Nature Cell Biology,* 13**,** 541-U328.

OHNISHI, K., SEMI, K., YAMAMOTO, T., SHIMIZU, M., TANAKA, A., MITSUNAGA, K., OKITA, K., OSAFUNE, K., ARIOKA, Y., MAEDA, T., SOEJIMA, H., MORIWAKI, H., YAMANAKA, S., WOLTJEN, K. & YAMADA, Y. 2014. Premature termination of reprogramming in vivo

leads to cancer development through altered epigenetic regulation. *Cell,* 156**,** 663-77.

OKADA, Y., FENG, Q., LIN, Y., JIANG, Q., LI, Y., COFFIELD, V. M., SU, L., XU, G. & ZHANG, Y. 2005. hDOT1L links histone methylation to leukemogenesis. *Cell,* 121**,** 167-78.

OKITA, K., HONG, H., TAKAHASHI, K. & YAMANAKA, S. 2010. Generation of mouse-induced pluripotent stem cells with plasmid vectors. *Nat Protoc,* 5**,** 418-28.

OKITA, K., ICHISAKA, T. & YAMANAKA, S. 2007a. Generation of germline-competent induced pluripotent stem cells. *Nature,* 448**,** 313-U1.

OKITA, K., ICHISAKA, T. & YAMANAKA, S. 2007b. Generation of germline-competent induced pluripotent stem cells. *Nature,* 448**,** 313-7.

OKITA, K., MATSUMURA, Y., SATO, Y., OKADA, A., MORIZANE, A., OKAMOTO, S., HONG, H., NAKAGAWA, M., TANABE, K., TEZUKA, K., SHIBATA, T., KUNISADA, T., TAKAHASHI, M., TAKAHASHI, J., SAJI, H. & YAMANAKA, S. 2011. A more efficient method to generate integration-free human iPS cells. *Nat Methods,* 8**,** 409-12.

OKITA, K., YAMAKAWA, T., MATSUMURA, Y., SATO, Y., AMANO, N., WATANABE, A., GOSHIMA, N. & YAMANAKA, S. 2013. An efficient nonviral method to generate integration-free human-induced pluripotent stem cells from cord blood and peripheral blood cells. *Stem Cells,* 31**,** 458-66.

ONDER, T. T., KARA, N., CHERRY, A., SINHA, A. U., ZHU, N., BERNT, K. M., CAHAN, P., MANCARCI, B. O., UNTERNAEHRER, J., GUPTA, P. B., LANDER, E. S., ARMSTRONG, S. A. & DALEY, G. Q. 2012. Chromatin-modifying enzymes as modulators of reprogramming. *Nature,* 483**,** 598-U119.

ORKIN, S. H., WANG, J., KIM, J., CHU, J., RAO, S., THEUNISSEN, T. W., SHEN, X. & LEVASSEUR, D. N. 2008. The transcriptional network controlling pluripotency in ES cells. *Cold Spring Harb Symp Quant Biol,* 73**,** 195-202.

OSAFUNE, K., CARON, L., BOROWIAK, M., MARTINEZ, R. J., FITZ-GERALD, C. S., SATO, Y., COWAN, C. A., CHIEN, K. R. & MELTON, D. A. 2008. Marked differences in differentiation propensity among human embryonic stem cell lines. *Nature Biotechnology,* 26**,** 313-315.

OZDAG, H., TESCHENDORFF, A. E., AHMED, A. A., HYLAND, S. J., BLENKIRON, C., BOBROW, L., VEERAKUMARASIVAM, A., BURTT, G., SUBKHANKULOVA, T., ARENDS, M. J., COLLINS, V. P., BOWTELL, D., KOUZARIDES, T., BRENTON, J. D. & CALDAS, C. 2006. Differential expression of selected histone modifier genes in human solid cancers. *BMC Genomics,* 7**,** 90.

PAN, G. & THOMSON, J. A. 2007. Nanog and transcriptional networks in embryonic stem cell pluripotency. *Cell Res,* 17**,** 42-9.

PANOPOULOS, A. D., YANES, O., RUIZ, S., KIDA, Y. S., DIEP, D., TAUTENHAHN, R., HERRERIAS, A., BATCHELDER, E. M., PLONGTHONGKUM, N., LUTZ, M., BERGGREN, W. T., ZHANG, K., EVANS, R. M., SIUZDAK, G. & BELMONTE, J. C. I. 2012a. The metabolome of induced pluripotent stem cells reveals metabolic changes occurring in somatic cell reprogramming. *Cell Research,* 22**,** 168-177.

PANOPOULOS, A. D., YANES, O., RUIZ, S., KIDA, Y. S., DIEP, D., TAUTENHAHN, R., HERRERIAS, A., BATCHELDER, E. M., PLONGTHONGKUM, N., LUTZ, M., BERGGREN, W. T., ZHANG, K., EVANS, R. M., SIUZDAK, G. & IZPISUA BELMONTE, J. C. 2012b. The metabolome of induced pluripotent stem cells reveals metabolic changes occurring in somatic cell reprogramming. *Cell Res,* 22**,** 168-77.

PAPAPETROU, E. P. 2016. Patient-derived induced pluripotent stem cells in cancer research and precision oncology. *Nat Med,* 22**,** 1392-1401.

PAQUET, D., KWART, D., CHEN, A., SPROUL, A., JACOB, S., TEO, S., OLSEN, K. M., GREGG, A., NOGGLE, S. & TESSIER-LAVIGNE, M. 2016. Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature,* 533**,** 125-9.

PARDAL, R., CLARKE, M. F. & MORRISON, S. J. 2003. Applying the principles of stem-cell biology to cancer. *Nat Rev Cancer,* 3**,** 895-902.

PARK, T. G., JEONG, J. H. & KIM, S. W. 2006. Current status of polymeric gene delivery systems. *Adv Drug Deliv Rev,* 58**,** 467-86.

PARK, T. S., HUO, J. S., PETERS, A., TALBOT, C. C., JR., VERMA, K., ZIMMERLIN, L., KAPLAN, I. M. & ZAMBIDIS, E. T. 2012. Growth factor-activated stem cell circuits and stromal signals cooperatively accelerate non-integrated iPSC reprogramming of human myeloid progenitors. *PLoS One,* 7**,** e42838.

PATHANIA, R., KOLHE, R. B., RAMACHANDRAN, S., MARIAPPAN, G., THAKUR, P., PRASAD, P. D., GANAPATHY, V. & THANGARAJU, M. 2016. Combination of DNMT and HDAC inhibitors reprogram cancer stem cell signaling to overcome drug resistance. *Cancer Research,* 76.

PATTABIRAMAN, D. R. & WEINBERG, R. A. 2014. Tackling the cancer stem cells - what challenges do they pose? *Nat Rev Drug Discov,* 13**,** 497-512.

PEREIRA, B., CHIN, S. F., RUEDA, O. M., VOLLAN, H. K., PROVENZANO, E., BARDWELL, H. A., PUGH, M., JONES, L., RUSSELL, R., SAMMUT, S. J., TSUI, D. W., LIU, B., DAWSON, S. J., ABRAHAM, J., NORTHEN, H., PEDEN, J. F., MUKHERJEE, A., TURASHVILI, G., GREEN, A. R., MCKINNEY, S., OLOUMI, A., SHAH, S., ROSENFELD, N., MURPHY, L., BENTLEY, D. R., ELLIS, I. O., PURUSHOTHAM, A., PINDER, S. E., BORRESEN-DALE, A. L., EARL, H. M., PHAROAH, P. D., ROSS, M. T., APARICIO, S. & CALDAS, C. 2016a. Erratum: The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes. *Nat Commun,* 7**,** 11908.

PEREIRA, B., CHIN, S. F., RUEDA, O. M., VOLLAN, H. K., PROVENZANO, E., BARDWELL, H. A., PUGH, M., JONES, L., RUSSELL, R., SAMMUT, S. J., TSUI, D. W., LIU, B., DAWSON, S. J., ABRAHAM, J., NORTHEN, H., PEDEN, J. F., MUKHERJEE, A., TURASHVILI, G., GREEN, A. R., MCKINNEY, S., OLOUMI, A., SHAH, S., ROSENFELD, N., MURPHY, L., BENTLEY, D. R., ELLIS, I. O., PURUSHOTHAM, A., PINDER, S. E., BORRESEN-DALE, A. L., EARL, H. M., PHAROAH, P. D., ROSS, M. T., APARICIO, S. & CALDAS, C. 2016b. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat Commun,* 7**,** 11479.

PEROU, C. M. 2011. Molecular stratification of triple-negative breast cancers. *Oncologist,* 16 Suppl 1**,** 61-70.

PEROU, C. M. & BORRESEN-DALE, A. L. 2011. Systems biology and genomics of breast cancer. *Cold Spring Harb Perspect Biol,* 3.

PEROU, C. M., SORLIE, T., EISEN, M. B., VAN DE RIJN, M., JEFFREY, S. S., REES, C. A., POLLACK, J. R., ROSS, D. T., JOHNSEN, H., AKSLEN, L. A., FLUGE, O., PERGAMENSCHIKOV, A., WILLIAMS, C., ZHU, S. X., LONNING, P. E., BORRESEN-DALE, A. L., BROWN, P. O. & BOTSTEIN, D. 2000. Molecular portraits of human breast tumours. *Nature,* 406**,** 747-52.

PFISTER, S., REA, S., TAIPALE, M., MENDRZYK, F., STRAUB, B., ITTRICH, C., THUERIGEN, O., SINN, H. P., AKHTAR, A. & LICHTER, P. 2008. The histone acetyltransferase hMOF is frequently downregulated in primary breast carcinoma and medulloblastoma and constitutes a biomarker for clinical outcome in medulloblastoma. *Int J Cancer,* 122**,** 1207-13.

PFISTER, S. X. & ASHWORTH, A. 2017. Marked for death: targeting epigenetic changes in cancer. *Nat Rev Drug Discov,* 16**,** 241-263.

PIERCE, G. B. & DIXON, F. J., JR. 1959. Testicular teratomas. I. Demonstration of teratogenesis by metamorphosis of multipotential cells. *Cancer,* 12**,** 573-83.

PLATH, K. & LOWRY, W. E. 2011. Progress in understanding reprogramming to the induced pluripotent state. *Nat Rev Genet,* 12**,** 253-65.

PLATT, R. J., CHEN, S., ZHOU, Y., YIM, M. J., SWIECH, L., KEMPTON, H. R., DAHLMAN, J. E., PARNAS, O., EISENHAURE, T. M., JOVANOVIC, M., GRAHAM, D. B., JHUNJHUNWALA, S., HEIDENREICH, M., XAVIER, R. J., LANGER, R., ANDERSON, D. G., HACOHEN, N., REGEV, A., FENG, G., SHARP, P. A. & ZHANG, F. 2014. CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell,* 159**,** 440-55.

PLAYER, A., WANG, Y., BHATTACHARYA, B., RAO, M., PURI, R. K. & KAWASAKI, E. S. 2006. Comparisons between transcriptional regulation and RNA expression in human embryonic stem cell lines. *Stem Cells Dev,* 15**,** 315-23.

POGRIBNY, I. P., ROSS, S. A., TRYNDYAK, V. P., POGRIBNA, M., POIRIER, L. A. & KARPINETS, T. V. 2006. Histone H3 lysine 9 and H4 lysine 20 trimethylation and the expression of Suv4-20h2 and Suv-39h1 histone methyltransferases in hepatocarcinogenesis induced by methyl deficiency in rats. *Carcinogenesis,* 27**,** 1180-6.

POLO, J. M., ANDERSSEN, E., WALSH, R. M., SCHWARZ, B. A., NEFZGER, C. M., LIM, S. M., BORKENT, M., APOSTOLOU, E., ALAEI, S., CLOUTIER, J., BAR-NUR, O., CHELOUFI, S., STADTFELD, M., FIGUEROA, M. E., ROBINTON, D., NATESAN, S., MELNICK, A., ZHU, J., RAMASWAMY, S. & HOCHEDLINGER, K. 2012a. A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell,* 151**,** 1617-32.

POLO, J. M., ANDERSSEN, E., WALSH, R. M., SCHWARZ, B. A., NEFZGER, C. M., LIM, S. M., BORKENT, M., APOSTOLOU, E., ALAEI, S., CLOUTIER, J., BAR-NUR, O., CHELOUFI, S., STADTFELD, M., FIGUEROA, M. E., ROBINTON, D., NATESAN, S., MELNICK, A., ZHU, J. F., RAMASWAMY, S. & HOCHEDLINGER, K. 2012b. A Molecular Roadmap of Reprogramming Somatic Cells into iPS Cells. *Cell,* 151**,** 1617-1632.

POLO, J. M., LIU, S., FIGUEROA, M. E., KULALERT, W., EMINLI, S., TAN, K. Y., APOSTOLOU, E., STADTFELD, M., LI, Y., SHIODA, T., NATESAN, S., WAGERS, A. J., MELNICK, A., EVANS, T. & HOCHEDLINGER, K. 2010. Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. *Nat Biotechnol,* 28**,** 848-55.

POLYAK, K. 2011. Heterogeneity in breast cancer. *J Clin Invest,* 121**,** 3786-8.

PORTELA, A. & ESTELLER, M. 2010. Epigenetic modifications and human disease. *Nat Biotechnol,* 28**,** 1057-68.

POSTOVIT, L. M., MARGARYAN, N. V., SEFTOR, E. A., KIRSCHMANN, D. A., LIPAVSKY, A., WHEATON, W. W., ABBOTT, D. E., SEFTOR, R. E. & HENDRIX, M. J. 2008. Human embryonic stem cell microenvironment suppresses the tumorigenic phenotype of aggressive cancer cells. *Proc Natl Acad Sci U S A,* 105**,** 4329-34.

POSTOVIT, L. M., SEFTOR, E. A., SEFTOR, R. E. & HENDRIX, M. J. 2006. A three-dimensional model to study the epigenetic effects induced by the microenvironment of human embryonic stem cells. *Stem Cells,* 24**,** 501-5.

POWERS, S. & POLLACK, R. E. 2016. Inducing stable reversion to achieve cancer control. *Nat Rev Cancer,* 16**,** 266-70.

PRAT, A., KARGINOVA, O., PARKER, J. S., FAN, C., HE, X., BIXBY, L., HARRELL, J. C., ROMAN, E., ADAMO, B., TROESTER, M. & PEROU, C. M. 2013. Characterization of cell lines derived from breast cancers and normal mammary tissues for the study of the intrinsic molecular subtypes. *Breast Cancer Res Treat,* 142**,** 237-55.

PRZYBORSKI, S. A. 2005. Differentiation of human embryonic stem cells after transplantation in immune-deficient mice. *Stem Cells,* 23**,** 1242-50.

QU, Y., HAN, B., GAO, B., BOSE, S., GONG, Y., WAWROWSKY, K., GIULIANO, A. E., SAREEN, D. & CUI, X. 2017. Differentiation of Human Induced Pluripotent Stem Cells to Mammary-like Organoids. *Stem Cell Reports,* 8**,** 205-215.

QU, Y., HAN, B., YU, Y., YAO, W., BOSE, S., KARLAN, B. Y., GIULIANO, A. E. & CUI, X. 2015. Evaluation of MCF10A as a Reliable Model for Normal Human Mammary Epithelial Cells. *PLoS One,* 10**,** e0131285.

RADPOUR, R., BAREKATI, Z., KOHLER, C., SCHUMACHER, M. M., GRUSSENMEYER, T., JENOE, P., HARTMANN, N., MOES, S., LETZKUS, M., BITZER, J., LEFKOVITS, I., STAEDTLER, F. & ZHONG, X. Y. 2011. Integrated epigenetics of human breast cancer: synoptic investigation of targeted genes, microRNAs and proteins upon demethylation treatment. *PLoS One,* 6**,** e27355.

RADPOUR, R., KOHLER, C., HAGHIGHI, M. M., FAN, A. X., HOLZGREVE, W. & ZHONG, X. Y. 2009. Methylation profiles of 22 candidate genes in breast cancer using high-throughput MALDI-TOF mass array. *Oncogene,* 28**,** 2969-78.

RAIS, Y., ZVIRAN, A., GEULA, S., GAFNI, O., CHOMSKY, E., VIUKOV, S., MANSOUR, A. A., CASPI, I., KRUPALNIK, V., ZERBIB, M., MAZA, I., MOR, N., BARAN, D., WEINBERGER, L., JAITIN, D. A., LARA-ASTIASO, D., BLECHER-GONEN, R., SHIPONY, Z., MUKAMEL, Z., HAGAI, T., GILAD, S., AMANN-ZALCENSTEIN, D., TANAY, A., AMIT, I., NOVERSHTERN, N. & HANNA, J. H. 2015. Corrigendum: Deterministic direct reprogramming of somatic cells to pluripotency. *Nature,* 520**,** 710.

RAMACHANDRAN, S., IENT, J., GOTTGENS, E. L., KRIEG, A. J. & HAMMOND, E. M. 2015. Epigenetic Therapy for Solid Tumors: Highlighting the Impact of Tumor Hypoxia. *Genes (Basel),* 6**,** 935-56.

RAMALHO-SANTOS, M., YOON, S., MATSUZAKI, Y., MULLIGAN, R. C. & MELTON, D. A. 2002. "Stemness": transcriptional profiling of embryonic and adult stem cells. *Science,* 298**,** 597-600.

RAMOS-MEJIA, V., FRAGA, M. F. & MENENDEZ, P. 2012. iPSCs from cancer cells: challenges and opportunities. *Trends Mol Med,* 18**,** 245-7.

RAMOS, T. L., SANCHEZ-ABARCA, L. I., MUNTION, S., PRECIADO, S., PUIG, N., LOPEZ-RUANO, G., HERNANDEZ-HERNANDEZ, A., REDONDO, A., ORTEGA, R., RODRIGUEZ, C., SANCHEZ-GUIJO, F. & DEL CANIZO, C. 2016. MSC surface markers (CD44, CD73, and CD90) can identify human MSC-derived extracellular vesicles by conventional flow cytometry. *Cell Communication and Signaling,* 14.

RAN, F. A., HSU, P. D., WRIGHT, J., AGARWALA, V., SCOTT, D. A. & ZHANG, F. 2013. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc,* 8**,** 2281-2308.

REDMER, T., DIECKE, S., GRIGORYAN, T., QUIROGA-NEGREIRA, A., BIRCHMEIER, W. & BESSER, D. 2011. E-cadherin is crucial for embryonic stem cell pluripotency and can replace OCT4 during somatic cell reprogramming. *EMBO Rep,* 12**,** 720-6.

REN, R. 2005. Mechanisms of BCR-ABL in the pathogenesis of chronic myelogenous leukaemia. *Nat Rev Cancer,* 5**,** 172-83.

REN, X., YANG, Z., XU, J., SUN, J., MAO, D., HU, Y., YANG, S. J., QIAO, H. H., WANG, X., HU, Q., DENG, P., LIU, L. P., JI, J. Y., LI, J. B. & NI, J. Q. 2014. Enhanced specificity and efficiency of the CRISPR/Cas9 system with optimized sgRNA parameters in Drosophila. *Cell Rep,* 9**,** 1151-62.

REYA, T. & CLEVERS, H. 2005. Wnt signalling in stem cells and cancer. *Nature,* 434**,** 843-50.

REYA, T., MORRISON, S. J., CLARKE, M. F. & WEISSMAN, I. L. 2001. Stem cells, cancer, and cancer stem cells. *Nature,* 414**,** 105-11.

RIJLAARSDAM, M. A., VAN HERK, H. A., GILLIS, A. J., STOOP, H., JENSTER, G., MARTENS, J., VAN LEENDERS, G. J., DINJENS, W., HOOGLAND, A. M., TIMMERMANS, M. & LOOIJENGA, L. H. 2011. Specific detection of OCT3/4 isoform A/B/B1 expression in solid (germ cell) tumours and cell lines: confirmation of OCT3/4 specificity for germ cell tumours. *Br J Cancer,* 105**,** 854-63.

RIO, P., BANOS, R., LOMBARDO, A., QUINTANA-BUSTAMANTE, O., ALVAREZ, L., GARATE, Z., GENOVESE, P., ALMARZA, E., VALERI, A., DIEZ, B., NAVARRO, S., TORRES, Y., TRUJILLO, J. P., MURILLAS, R., SEGOVIA, J. C., SAMPER, E., SURRALLES, J., GREGORY, P. D., HOLMES, M. C., NALDINI, L. & BUEREN, J. A. 2014. Targeted gene therapy and cell reprogramming in Fanconi anemia. *EMBO Mol Med,* 6**,** 835-48.

ROBINTON, D. A. & DALEY, G. Q. 2012. The promise of induced pluripotent stem cells in research and therapy. *Nature,* 481**,** 295-305.

RODDA, D. J., CHEW, J. L., LIM, L. H., LOH, Y. H., WANG, B., NG, H. H. & ROBSON, P. 2005. Transcriptional regulation of nanog by OCT4 and SOX2. *J Biol Chem,* 280**,** 24731-7.

RODRIGUEZ-PAREDES, M. & ESTELLER, M. 2011. Cancer epigenetics reaches mainstream oncology. *Nat Med,* 17**,** 330-9.

ROGERS, M. B., HOSLER, B. A. & GUDAS, L. J. 1991. Specific expression of a retinoic acid-regulated, zinc-finger gene, Rex-1, in preimplantation embryos, trophoblast and spermatocytes. *Development,* 113**,** 815-24.

ROMANO, G., VENEZIANO, D., ACUNZO, M. & CROCE, C. M. 2017. Small non-coding RNA and cancer. *Carcinogenesis,* 38**,** 485-491.

ROUHANI, F. J., NIK-ZAINAL, S., WUSTER, A., LI, Y., CONTE, N., KOIKE-YUSA, H., KUMASAKA, N., VALLIER, L., YUSA, K. & BRADLEY, A. 2016. Mutational History of a Human Cell Lineage from Somatic to Induced Pluripotent Stem Cells. *PLoS Genet,* 12**,** e1005932.

ROUZIER, R., RAJAN, R., WAGNER, P., HESS, K. R., GOLD, D. L., STEC, J., AYERS, M., ROSS, J. S., ZHANG, P., BUCHHOLZ, T. A., KUERER, H., GREEN, M., ARUN, B., HORTOBAGYI, G. N., SYMMANS, W. F. & PUSZTAI, L. 2005. Microtubule-associated protein tau: a marker of paclitaxel sensitivity in breast cancer. *Proc Natl Acad Sci U S A,* 102**,** 8315-20.

ROWLAND, B. D. & PEEPER, D. S. 2006. KLF4, p21 and context-dependent opposing forces in cancer. *Nat Rev Cancer,* 6**,** 11-23.

RUSSNES, H. G., LINGJAERDE, O. C., BORRESEN-DALE, A. L. & CALDAS, C. 2017. Breast Cancer Molecular Stratification: From Intrinsic Subtypes to Integrative Clusters. *Am J Pathol,* 187**,** 2152-2162.

RYCAJ, K. & TANG, D. G. 2015. Cell-of-Origin of Cancer versus Cancer Stem Cells: Assays and Interpretations. *Cancer Res,* 75**,** 4003-11.

SAKURAI, K., TALUKDAR, I., PATIL, V. S., DANG, J., LI, Z., CHANG, K. Y., LU, C. C., DELORME-WALKER, V., DERMARDIROSSIAN, C., ANDERSON, K., HANEIN, D., YANG, C. S., WU, D., LIU, Y. & RANA, T. M. 2014. Kinome-wide functional analysis highlights the role of cytoskeletal remodeling in somatic cell reprogramming. *Cell Stem Cell,* 14**,** 523-34.

SAMAVARCHI-TEHRANI, P., GOLIPOUR, A., DAVID, L., SUNG, H. K., BEYER, T. A., DATTI, A., WOLTJEN, K., NAGY, A. & WRANA, J. L. 2010. Functional Genomics Reveals a BMP-Driven Mesenchymal-to-Epithelial Transition in the Initiation of Somatic Cell Reprogramming. *Cell Stem Cell,* 7**,** 64-77.

SAMMAR, M., GULBINS, E., HILBERT, K., LANG, F. & ALTEVOGT, P. 1997. Mouse CD24 as a signaling molecule for integrin-mediated cell binding: functional and physical association with src-kinases. *Biochem Biophys Res Commun,* 234**,** 330-4.

SANCHO-MARTINEZ, I., NIVET, E., XIA, Y., HISHIDA, T., AGUIRRE, A., OCAMPO, A., MA, L., MOREY, R., KRAUSE, M. N., ZEMBRZYCKI, A., ANSORGE, O., VAZQUEZ-FERRER, E., DUBOVA, I., REDDY, P., LAM, D., HISHIDA, Y., WU, M. Z., ESTEBAN, C. R., O'LEARY, D., WAHL, G. M., VERMA, I. M., LAURENT, L. C. & IZPISUA BELMONTE, J. C. 2016. Establishment of human iPSC-based models for the study and targeting of glioma initiating cells. *Nat Commun,* 7**,** 10743.

SANDER, J. D. & JOUNG, J. K. 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol,* 32**,** 347-55.

SANDOVAL, J. & ESTELLER, M. 2012. Cancer epigenomics: beyond genomics. *Curr Opin Genet Dev,* 22**,** 50-5.

SATO, H., OKA, T., SHINNOU, Y., KONDO, T., WASHIO, K., TAKANO, M., TAKATA, K., MORITO, T., HUANG, X., TAMURA, M., KITAMURA, Y., OHARA, N., OUCHIDA, M., OHSHIMA, K., SHIMIZU, K., TANIMOTO, M., TAKAHASHI, K., MATSUOKA, M., UTSUNOMIYA, A. & YOSHINO, T. 2010. Multi-step aberrant CpG island hyper-methylation is associated with the progression of adult T-cell leukemia/lymphoma. *Am J Pathol,* 176**,** 402-15.

SATO, T., VAN ES, J. H., SNIPPERT, H. J., STANGE, D. E., VRIES, R. G., VAN DEN BORN, M., BARKER, N., SHROYER, N. F., VAN DE WETERING, M. & CLEVERS, H. 2011. Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature,* 469**,** 415-8.

SCHAEFER, K. A., WU, W. H., COLGAN, D. F., TSANG, S. H., BASSUK, A. G. & MAHAJAN, V. B. 2017. Unexpected mutations after CRISPR-Cas9 editing in vivo. *Nat Methods,* 14**,** 547-548.

SCHLAEGER, T. M., DAHERON, L., BRICKLER, T. R., ENTWISLE, S., CHAN, K., CIANCI, A., DEVINE, A., ETTENGER, A., FITZGERALD, K., GODFREY, M., GUPTA, D., MCPHERSON, J., MALWADKAR, P., GUPTA, M., BELL, B., DOI, A., JUNG, N., LI, X., LYNES, M. S., BROOKES, E., CHERRY, A. B., DEMIRBAS, D., TSANKOV, A. M., ZON, L. I., RUBIN, L. L., FEINBERG, A. P., MEISSNER, A., COWAN, C. A. & DALEY, G. Q. 2015. A comparison of non-integrating reprogramming methods. *Nat Biotechnol,* 33**,** 58-63.

SCHMITZ, J., SCHWAB, J., SCHWENCK, J., CHEN, Q., QUINTANILLA-MARTINEZ, L., HAHN, M., WIETEK, B., SCHWENZER, N., STAEBLER, A., KOHLHOFER, U., AINA, O. H., HUBBARD, N. E., REISCHL, G., BOROWSKY, A. D., BRUCKER, S., NIKOLAOU, K., LA FOUGERE, C., CARDIFF, R. D., PICHLER, B. J. & SCHMID, A. M. 2016. Decoding Intratumoral Heterogeneity of Breast Cancer by Multiparametric In Vivo Imaging: A Translational Study. *Cancer Res,* 76**,** 5512-22.

SCHNITT, S. J. 2010. Classification and prognosis of invasive breast cancer: from morphology to molecular taxonomy. *Mod Pathol,* 23 Suppl 2**,** S60-4.

SCHOONEN, P. M., TALENS, F., STOK, C., GOGOLA, E., HEIJINK, A. M., BOUWMAN, P., FOIJER, F., TARSOUNAS, M., BLATTER, S., JONKERS, J., ROTTENBERG, S. & VAN VUGT, M. A. T. M. 2017. Progression through mitosis promotes PARP inhibitor-induced cytotoxicity in homologous recombination-deficient cancer cells. *Nature Communications,* 8.

SCHULDINER, M., YANUKA, O., ITSKOVITZ-ELDOR, J., MELTON, D. A. & BENVENISTY, N. 2000. Effects of eight growth factors on the differentiation of cells derived from human embryonic stem cells. *Proc Natl Acad Sci U S A,* 97**,** 11307-12.

SCHWARZ, J. M., RODELSPERGER, C., SCHUELKE, M. & SEELOW, D. 2010. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods,* 7**,** 575-6.

SEMI, K., MATSUDA, Y., OHNISHI, K. & YAMADA, Y. 2013. Cellular reprogramming and cancer development. *Int J Cancer,* 132**,** 1240-8.

SERRANO-GOMEZ, S. J., SANABRIA-SALAS, M. C., HERNANDEZ-SUAREZ, G., GARCIA, O., SILVA, C., ROMERO, A., MEJIA, J. C., MIELE, L., FEJERMAN, L. & ZABALETA, J. 2016. High prevalence of luminal B breast cancer intrinsic subtype in Colombian women. *Carcinogenesis,* 37**,** 669-676.

SEYMOUR, T., NOWAK, A. & KAKULAS, F. 2015. Targeting Aggressive Cancer Stem Cells in Glioblastoma. *Front Oncol,* 5**,** 159.

SHAH, M. & ALLEGRUCCI, C. 2012. Keeping an open mind: highlights and controversies of the breast cancer stem cell theory. *Breast Cancer (Dove Med Press),* 4**,** 155-66.

SHAH, S. P., ROTH, A., GOYA, R., OLOUMI, A., HA, G., ZHAO, Y., TURASHVILI, G., DING, J., TSE, K., HAFFARI, G., BASHASHATI, A., PRENTICE, L. M., KHATTRA, J., BURLEIGH, A., YAP, D., BERNARD, V., MCPHERSON, A., SHUMANSKY, K., CRISAN, A., GIULIANY, R., HERAVI-MOUSSAVI, A., ROSNER, J., LAI, D., BIROL, I., VARHOL, R., TAM, A., DHALLA, N., ZENG, T., MA, K., CHAN, S. K., GRIFFITH, M., MORADIAN, A., CHENG, S. W., MORIN, G. B., WATSON, P., GELMON, K., CHIA, S., CHIN, S. F., CURTIS, C., RUEDA, O. M., PHAROAH, P. D., DAMARAJU, S., MACKEY, J., HOON, K., HARKINS, T., TADIGOTLA, V., SIGAROUDINIA, M., GASCARD, P., TLSTY, T., COSTELLO, J. F., MEYER, I. M., EAVES, C. J., WASSERMAN, W. W., JONES, S., HUNTSMAN, D., HIRST, M., CALDAS, C., MARRA, M. A. & APARICIO, S. 2012. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature,* 486**,** 395-9.

SHAKIBA, N., WHITE, C. A., LIPSITZ, Y. Y., YACHIE-KINOSHITA, A., TONGE, P. D., HUSSEIN, S. M., PURI, M. C., ELBAZ, J., MORRISSEY-SCOOT, J., LI, M., MUNOZ, J., BENEVENTO, M., ROGERS, I. M., HANNA, J. H., HECK, A. J., WOLLSCHEID, B., NAGY, A. & ZANDSTRA, P. W. 2015. CD24 tracks divergent pluripotent states in mouse and human cells. *Nat Commun,* 6**,** 7329.

SHAN, M., YIN, H., LI, J., LI, X., WANG, D., SU, Y., NIU, M., ZHONG, Z., WANG, J., ZHANG, X., KANG, W. & PANG, D. 2016. Detection of aberrant methylation of a six-gene panel in serum DNA for diagnosis of breast cancer. *Oncotarget,* 7**,** 18485-94.

SHELTZER, J. M., BLANK, H. M., PFAU, S. J., TANGE, Y., GEORGE, B. M., HUMPTON, T. J., BRITO, I. L., HIRAOKA, Y., NIWA, O. & AMON, A. 2011. Aneuploidy drives genomic instability in yeast. *Science,* 333**,** 1026-30.

SHEN, H. & LAIRD, P. W. 2013. Interplay between the cancer genome and epigenome. *Cell,* 153**,** 38-55.

SHEN, M. M. 2007. Nodal signaling: developmental roles and regulation. *Development,* 134**,** 1023-1034.

SHERR, C. J. 2006. Divorcing ARF and p53: an unsettled case. *Nat Rev Cancer,* 6**,** 663-73.

SHIRASAWA, T., AKASHI, T., SAKAMOTO, K., TAKAHASHI, H., MARUYAMA, N. & HIROKAWA, K. 1993. Gene expression of CD24 core peptide molecule in developing brain and developing non-neural tissues. *Dev Dyn,* 198**,** 1-13.

SIEBER, O. M., TOMLINSON, S. R. & TOMLINSON, I. P. 2005. Tissue, cell and stage specificity of (epi)mutations in cancers. *Nat Rev Cancer,* 5**,** 649-55.

SIMA, J. & GILBERT, D. M. 2014. Complex correlations: replication timing and mutational landscapes during cancer and genome evolution. *Curr Opin Genet Dev,* 25**,** 93-100.

SIMON, C., CHAGRAOUI, J., KROSL, J., GENDRON, P., WILHELM, B., LEMIEUX, S., BOUCHER, G., CHAGNON, P., DROUIN, S., LAMBERT, R., RONDEAU, C., BILODEAU, A., LAVALLEE, S., SAUVAGEAU, M., HEBERT, J. & SAUVAGEAU, G. 2012. A key role for EZH2 and associated genes in mouse and human adult T-cell acute leukemia. *Genes Dev,* 26**,** 651-6.

SINGOVSKI, G., BERNAL, C., KUCIAK, M., SIEGL-CACHEDENIER, I., CONOD, A. & RUIZ, I. A. A. 2016. In vivo epigenetic reprogramming of primary human colon cancer cells enhances metastases. *J Mol Cell Biol,* 8**,** 157-73.

SINN, H. P. & KREIPE, H. 2013. A Brief Overview of the WHO Classification of Breast Tumors, 4th Edition, Focusing on Issues and Updates from the 3rd Edition. *Breast Care (Basel),* 8**,** 149-54.

SKIBINSKI, A. & KUPERWASSER, C. 2015. The origin of breast tumor heterogeneity. *Oncogene,* 34**,** 5309-16.

SMITH, Z. D., NACHMAN, I., REGEV, A. & MEISSNER, A. 2010. Dynamic single-cell imaging of direct reprogramming reveals an early specifying event. *Nat Biotechnol,* 28**,** 521-6.

SOLDNER, F., HOCKEMEYER, D., BEARD, C., GAO, Q., BELL, G. W., COOK, E. G., HARGUS, G., BLAK, A., COOPER, O., MITALIPOVA, M., ISACSON, O. & JAENISCH, R. 2009. Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors. *Cell,* 136**,** 964-77.

SONG, J., NOH, J. H., LEE, J. H., EUN, J. W., AHN, Y. M., KIM, S. Y., LEE, S. H., PARK, W. S., YOO, N. J., LEE, J. Y. & NAM, S. W. 2005. Increased expression of histone deacetylase 2 is found in human gastric cancer. *APMIS,* 113**,** 264-8.

SORLIE, T., PEROU, C. M., TIBSHIRANI, R., AAS, T., GEISLER, S., JOHNSEN, H., HASTIE, T., EISEN, M. B., VAN DE RIJN, M., JEFFREY, S. S., THORSEN, T., QUIST, H., MATESE, J. C., BROWN, P. O., BOTSTEIN, D., LONNING, P. E. & BORRESEN-DALE, A. L. 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A,* 98**,** 10869-74.

SOTIRIOU, C. & PICCART, M. J. 2007. Taking gene-expression profiling to the clinic: when will molecular signatures become relevant to patient care? *Nat Rev Cancer,* 7**,** 545-53.

SOUFI, A., DONAHUE, G. & ZARET, K. S. 2012. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell,* 151**,** 994-1004.

SOUFI, A., GARCIA, M. F., JAROSZEWICZ, A., OSMAN, N., PELLEGRINI, M. & ZARET, K. S. 2015. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell,* 161**,** 555-568.

SPARKS, A. B., MORIN, P. J., VOGELSTEIN, B. & KINZLER, K. W. 1998. Mutational analysis of the APC/beta-catenin/Tcf pathway in colorectal cancer. *Cancer Res,* 58**,** 1130-4.

SRIDHARAN, R., TCHIEU, J., MASON, M. J., YACHECHKO, R., KUOY, E., HORVATH, S., ZHOU, Q. & PLATH, K. 2009. Role of the Murine Reprogramming Factors in the Induction of Pluripotency. *Cell,* 136**,** 364-377.

STADTFELD, M. & HOCHEDLINGER, K. 2010. Induced pluripotency: history, mechanisms, and applications. *Genes Dev,* 24**,** 2239-63.

STADTFELD, M., MAHERALI, N., BREAULT, D. T. & HOCHEDLINGER, K. 2008a. Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell Stem Cell,* 2**,** 230-40.

STADTFELD, M., NAGAYA, M., UTIKAL, J., WEIR, G. & HOCHEDLINGER, K. 2008b. Induced pluripotent stem cells generated without viral integration. *Science,* 322**,** 945-9.

STAHLHUT, C. & SLACK, F. J. 2013. MicroRNAs and the cancer phenotype: profiling, signatures and clinical implications. *Genome Med,* 5**,** 111.

STEFANSSON, O. A. & ESTELLER, M. 2013. Epigenetic modifications in breast cancer and their role in personalized medicine. *Am J Pathol,* 183**,** 1052-1063.

STEFANSSON, O. A., MORAN, S., GOMEZ, A., SAYOLS, S., ARRIBAS-JORBA, C., SANDOVAL, J., HILMARSDOTTIR, H., OLAFSDOTTIR, E., TRYGGVADOTTIR, L., JONASSON, J. G., EYFJORD, J. & ESTELLER, M. 2015. A DNA methylation-based definition of biologically distinct breast cancer subtypes. *Mol Oncol,* 9**,** 555-68.

STEFANSSON, O. A., VILLANUEVA, A., VIDAL, A., MARTI, L. & ESTELLER, M. 2012. BRCA1 epigenetic inactivation predicts sensitivity to platinum-based chemotherapy in breast and ovarian cancer. *Epigenetics,* 7**,** 1225-9.

STEPHENS, P. J., TARPEY, P. S., DAVIES, H., VAN LOO, P., GREENMAN, C., WEDGE, D. C., NIK-ZAINAL, S., MARTIN, S., VARELA, I., BIGNELL, G. R., YATES, L. R., PAPAEMMANUIL, E., BEARE, D., BUTLER, A., CHEVERTON, A., GAMBLE, J., HINTON, J., JIA, M., JAYAKUMAR, A., JONES, D., LATIMER, C., LAU, K. W., MCLAREN, S., MCBRIDE, D. J., MENZIES, A., MUDIE, L., RAINE, K., RAD, R., CHAPMAN, M. S., TEAGUE, J., EASTON, D., LANGEROD, A., OSLO BREAST CANCER, C., LEE, M. T., SHEN, C. Y., TEE, B. T., HUIMIN, B. W., BROEKS, A., VARGAS, A. C., TURASHVILI, G., MARTENS, J., FATIMA, A., MIRON, P., CHIN, S. F., THOMAS, G., BOYAULT, S., MARIANI, O., LAKHANI, S. R., VAN DE VIJVER, M., VAN 'T VEER, L., FOEKENS, J., DESMEDT, C., SOTIRIOU, C., TUTT, A., CALDAS, C., REIS-FILHO, J. S., APARICIO, S. A., SALOMON, A. V., BORRESEN-DALE, A. L., RICHARDSON, A. L., CAMPBELL, P. J., FUTREAL, P. A. & STRATTON, M. R. 2012. The landscape of cancer genes and mutational processes in breast cancer. *Nature,* 486**,** 400-4.

STINGL, J. & CALDAS, C. 2007. Molecular heterogeneity of breast carcinomas and the cancer stem cell hypothesis. *Nat Rev Cancer,* 7**,** 791-9.

STRAHL, B. D. & ALLIS, C. D. 2000. The language of covalent histone modifications. *Nature,* 403**,** 41-5.

STRATTON, M. R., CAMPBELL, P. J. & FUTREAL, P. A. 2009. The cancer genome. *Nature,* 458**,** 719-24.

STRICKER, S. & POLLARD, S. 2014. Reprogramming cancer cells to pluripotency: an experimental tool for exploring cancer epigenetics. *Epigenetics,* 9**,** 798-802.

STRICKER, S. H., FEBER, A., ENGSTROM, P. G., CAREN, H., KURIAN, K. M., TAKASHIMA, Y., WATTS, C., WAY, M., DIRKS, P., BERTONE, P., SMITH, A., BECK, S. & POLLARD, S. M. 2013. Widespread resetting of DNA methylation in glioblastoma-initiating cells suppresses malignant cellular behavior in a lineage-dependent manner. *Genes Dev,* 27**,** 654-69.

SU, R. J., BAYLINK, D. J., NEISES, A., KIROYAN, J. B., MENG, X., PAYNE, K. J., TSCHUDY-SENEY, B., DUAN, Y., APPLEBY, N., KEARNS-JONKER, M., GRIDLEY, D. S., WANG, J., LAU, K. H. & ZHANG, X. B. 2013. Efficient generation of integration-free ips cells from human adult peripheral blood using BCL-XL together with Yamanaka factors. *PLoS One,* 8**,** e64496.

SUBRAMANYAM, D., LAMOUILLE, S., JUDSON, R. L., LIU, J. Y., BUCAY, N., DERYNCK, R. & BLELLOCH, R. 2011. Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells. *Nat Biotechnol,* 29**,** 443-8.

SUH, M. R., LEE, Y., KIM, J. Y., KIM, S. K., MOON, S. H., LEE, J. Y., CHA, K. Y., CHUNG, H. M., YOON, H. S., MOON, S. Y., KIM, V. N. & KIM, K. S. 2004. Human embryonic stem cells express a unique set of microRNAs. *Dev Biol,* 270**,** 488-98.

SUN, J., MENG, H., YAO, L., LV, M., BAI, J., ZHANG, J., WANG, L., OUYANG, T., LI, J., WANG, T., FAN, Z., FAN, T., LIN, B. & XIE, Y. 2017. Germline Mutations in Cancer Susceptibility Genes in a Large Series of Unselected Breast Cancer Patients. *Clin Cancer Res,* 23**,** 6113-6119.

SUNAMI, E., SHINOZAKI, M., SIM, M. S., NGUYEN, S. L., VU, A. T., GIULIANO, A. E. & HOON, D. S. 2008. Estrogen receptor and HER2/neu status affect epigenetic differences of tumor-related genes in primary breast tumors. *Breast Cancer Res,* 10**,** R46.

SUVA, M. L., RHEINBAY, E., GILLESPIE, S. M., PATEL, A. P., WAKIMOTO, H., RABKIN, S. D., RIGGI, N., CHI, A. S., CAHILL, D. P., NAHED, B. V., CURRY, W. T., MARTUZA, R. L., RIVERA, M. N., ROSSETTI, N., KASIF, S., BEIK, S., KADRI, S., TIROSH, I., WORTMAN, I., SHALEK, A. K., ROZENBLATT-ROSEN, O., REGEV, A., LOUIS, D. N. & BERNSTEIN, B. E. 2014. Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. *Cell,* 157**,** 580-94.

SVEJSTRUP, J. Q. 2010. The interface between transcription and mechanisms maintaining genome integrity. *Trends Biochem Sci,* 35**,** 333-8.

SWEENEY, C., BERNARD, P. S., FACTOR, R. E., KWAN, M. L., HABEL, L. A., QUESENBERRY, C. P., JR., SHAKESPEAR, K., WELTZIEN, E. K., STIJLEMAN, I. J., DAVIS, C. A., EBBERT, M. T., CASTILLO, A., KUSHI, L. H. & CAAN, B. J. 2014. Intrinsic subtypes from PAM50 gene expression assay in a population-based breast cancer cohort: differences by age, race, and tumor characteristics. *Cancer Epidemiol Biomarkers Prev,* 23**,** 714-24.

TAI, M. H., CHANG, C. C., KIUPEL, M., WEBSTER, J. D., OLSON, L. K. & TROSKO, J. E. 2005. Oct4 expression in adult human stem cells: evidence in support of the stem cell theory of carcinogenesis. *Carcinogenesis,* 26**,** 495-502.

TAKAHASHI, K., TANABE, K., OHNUKI, M., NARITA, M., ICHISAKA, T., TOMODA, K. & YAMANAKA, S. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell,* 131**,** 861-72.

TAKAHASHI, K. & YAMANAKA, S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell,* 126**,** 663-76.

TAMM, C., PIJUAN GALITO, S. & ANNEREN, C. 2013. A comparative study of protocols for mouse embryonic stem cell culturing. *PLoS One,* 8**,** e81156.

TATA, P. R., TATA, N. R., KUHL, M. & SIRBU, I. O. 2011. Identification of a novel epigenetic regulatory region within the pluripotency associated microRNA cluster, EEmiRC. *Nucleic Acids Res,* 39**,** 3574-81.

TATE, C. R., RHODES, L. V., SEGAR, H. C., DRIVER, J. L., POUNDER, F. N., BUROW, M. E. & COLLINS-BUROW, B. M. 2012. Targeting triple-negative breast cancer cells with the histone deacetylase inhibitor panobinostat. *Breast Cancer Res,* 14**,** R79.

TELERMAN, A. & AMSON, R. 2009. The molecular programme of tumour reversion: the steps beyond malignant transformation. *Nat Rev Cancer,* 9**,** 206-16.

TETTEH, P. W., BASAK, O., FARIN, H. F., WIEBRANDS, K., KRETZSCHMAR, K., BEGTHEL, H., VAN DEN BORN, M., KORVING, J., DE SAUVAGE, F., VAN ES, J. H., VAN OUDENAARDEN, A. & CLEVERS, H. 2016. Replacement of Lost Lgr5-Positive Stem Cells through Plasticity of Their Enterocyte-Lineage Daughters. *Cell Stem Cell,* 18**,** 203-13.

THEUNISSEN, T. W. & JAENISCH, R. 2014a. Molecular control of induced pluripotency. *Cell Stem Cell,* 14**,** 720-34.

THEUNISSEN, T. W. & JAENISCH, R. 2014b. Molecular Control of Induced Pluripotency. *Cell Stem Cell,* 14**,** 720-734.

THEUNISSEN, T. W., POWELL, B. E., WANG, H., MITALIPOVA, M., FADDAH, D. A., REDDY, J., FAN, Z. P., MAETZEL, D., GANZ, K., SHI, L., LUNGJANGWA, T., IMSOONTHORNRUKSA, S., STELZER, Y., RANGARAJAN, S., D'ALESSIO, A., ZHANG, J., GAO, Q., DAWLATY, M. M., YOUNG, R. A., GRAY, N. S. & JAENISCH, R. 2014. Systematic Identification of Culture Conditions for Induction and Maintenance of Naive Human Pluripotency. *Cell Stem Cell,* 15**,** 524-526.

THEUNISSEN, T. W., VAN OOSTEN, A. L., CASTELO-BRANCO, G., HALL, J., SMITH, A. & SILVA, J. C. 2011. Nanog overcomes reprogramming barriers and induces pluripotency in minimal conditions. *Curr Biol,* 21**,** 65-71.

THOMSON, J. A., ITSKOVITZ-ELDOR, J., SHAPIRO, S. S., WAKNITZ, M. A., SWIERGIEL, J. J., MARSHALL, V. S. & JONES, J. M. 1998. Embryonic stem cell lines derived from human blastocysts. *Science,* 282**,** 1145-7.

TOMITA, Y., LEE, M. J., LEE, S., TOMITA, S., CHUMSRI, S., CRUICKSHANK, S., ORDENTLICH, P. & TREPEL, J. B. 2016. The interplay of epigenetic therapy and immunity in locally recurrent or metastatic estrogen receptor-positive breast cancer: Correlative analysis of ENCORE 301, a randomized, placebo-controlled phase II trial of exemestane with or without entinostat. *Oncoimmunology,* 5**,** e1219008.

TOMPKINS, J. D., HALL, C., CHEN, V. C., LI, A. X., WU, X., HSU, D., COUTURE, L. A. & RIGGS, A. D. 2012. Epigenetic stability, adaptability, and reversibility in human embryonic stem cells. *Proc Natl Acad Sci U S A,* 109**,** 12544-9.

TONGE, P. D., CORSO, A. J., MONETTI, C., HUSSEIN, S. M., PURI, M. C., MICHAEL, I. P., LI, M., LEE, D. S., MAR, J. C., CLOONAN, N., WOOD, D. L., GAUTHIER, M. E., KORN, O., CLANCY, J. L., PREISS, T., GRIMMOND, S. M., SHIN, J. Y., SEO, J. S., WELLS, C. A., ROGERS, I. M. & NAGY, A. 2014. Divergent reprogramming routes lead to alternative stem-cell states. *Nature,* 516**,** 192-7.

TOSKA, E., OSMANBEYOGLU, H. U., CASTEL, P., CHAN, C., HENDRICKSON, R. C., ELKABETS, M., DICKLER, M. N., SCALTRITI, M., LESLIE, C. S., ARMSTRONG, S. A. & BASELGA, J. 2017. PI3K pathway regulates ER-dependent transcription in breast cancer through the epigenetic regulator KMT2D. *Science,* 355**,** 1324-1330.

TOYOOKA, K. O., TOYOOKA, S., VIRMANI, A. K., SATHYANARAYANA, U. G., EUHUS, D. M., GILCREASE, M., MINNA, J. D. & GAZDAR, A. F. 2001. Loss of expression and aberrant methylation of the CDH13 (H-cadherin) gene in breast and lung carcinomas. *Cancer Res,* 61**,** 4556-60.

TRYNDYAK, V. P., KOVALCHUK, O. & POGRIBNY, I. P. 2006. Loss of DNA methylation and histone H4 lysine 20 trimethylation in human breast cancer cells is associated with

aberrant expression of DNA methyltransferase 1, Suv4-20h2 histone methyltransferase and methyl-binding proteins. *Cancer Biol Ther,* 5**,** 65-70.

TSAI, H. C., LI, H. L., VAN NESTE, L., CAI, Y., ROBERT, C., RASSOOL, F. V., SHIN, J. J., HARBOM, K. M., BEATY, R., PAPPOU, E., HARRIS, J., YEN, R. W. C., AHUJA, N., BROCK, M. V., STEARNS, V., FELLER-KOPMAN, D., LIN, Y. C., WELM, A. L., ISSA, J. P., MINN, I., MATSUI, W., JANG, Y. Y., SHARKIS, S. J., BAYLIN, S. B. & ZAHNOW, C. A. 2012. Transient low doses of DNA demethylating agents exert durable antitumor effects on hematological and epithelial tumor cells. *Cancer Research,* 72.

TURNER, N. C. & REIS-FILHO, J. S. 2006. Basal-like breast cancer and the BRCA1 phenotype. *Oncogene,* 25**,** 5846-53.

TUYNDER, M., SUSINI, L., PRIEUR, S., BESSE, S., FIUCCI, G., AMSON, R. & TELERMAN, A. 2002. Biological models and genes of tumor reversion: cellular reprogramming through tpt1/TCTP and SIAH-1. *Proc Natl Acad Sci U S A,* 99**,** 14976-81.

UTIKAL, J., MAHERALI, N., KULALERT, W. & HOCHEDLINGER, K. 2009a. Sox2 is dispensable for the reprogramming of melanocytes and melanoma cells into induced pluripotent stem cells. *J Cell Sci,* 122**,** 3502-10.

UTIKAL, J., POLO, J. M., STADTFELD, M., MAHERALI, N., KULALERT, W., WALSH, R. M., KHALIL, A., RHEINWALD, J. G. & HOCHEDLINGER, K. 2009b. Immortalization eliminates a roadblock during cellular reprogramming into iPS cells. *Nature,* 460**,** 1145-8.

VALAMEHR, B., ROBINSON, M., ABUJAROUR, R., REZNER, B., VRANCEANU, F., LE, T., MEDCALF, A., LEE, T. T., FITCH, M., ROBBINS, D. & FLYNN, P. 2014. Platform for induction and maintenance of transgene-free hiPSCs resembling ground state pluripotent stem cells. *Stem Cell Reports,* 2**,** 366-81.

VAN HOESEL, A. Q., VAN DE VELDE, C. J., KUPPEN, P. J., LIEFERS, G. J., PUTTER, H., SATO, Y., ELASHOFF, D. A., TURNER, R. R., SHAMONKI, J. M., DE KRUIJF, E. M., VAN NES, J. G., GIULIANO, A. E. & HOON, D. S. 2012. Hypomethylation of LINE-1 in primary tumor has poor prognosis in young breast cancer patients: a retrospective cohort study. *Breast Cancer Res Treat,* 134**,** 1103-14.

VAN KEYMEULEN, A., LEE, M. Y., OUSSET, M., BROHEE, S., RORIVE, S., GIRADDI, R. R., WUIDART, A., BOUVENCOURT, G., DUBOIS, C., SALMON, I., SOTIRIOU, C., PHILLIPS, W. A. & BLANPAIN, C. 2015. Reactivation of multipotency by oncogenic PIK3CA induces breast tumour heterogeneity. *Nature,* 525**,** 119-23.

VASILATOU, D., PAPAGEORGIOU, S. G., DIMITRIADIS, G. & PAPPA, V. 2013. Epigenetic alterations and microRNAs: new players in the pathogenesis of myelodysplastic syndromes. *Epigenetics,* 8**,** 561-70.

VEECK, J., ROPERO, S., SETIEN, F., GONZALEZ-SUAREZ, E., OSORIO, A., BENITEZ, J., HERMAN, J. G. & ESTELLER, M. 2010. BRCA1 CpG island hypermethylation predicts sensitivity to poly(adenosine diphosphate)-ribose polymerase inhibitors. *J Clin Oncol,* 28**,** e563-4; author reply e565-6.

VENNSTROM, B., SHEINESS, D., ZABIELSKI, J. & BISHOP, J. M. 1982. Isolation and characterization of c-myc, a cellular homolog of the oncogene (v-myc) of avian myelocytomatosis virus strain 29. *J Virol,* 42**,** 773-9.

VIALE, G. 2012. The current state of breast cancer classification. *Ann Oncol,* 23 Suppl 10**,** x207-10.

VILLODRE, E. S., KIPPER, F. C., PEREIRA, M. B. & LENZ, G. 2016. Roles of OCT4 in tumorigenesis, cancer therapy resistance and prognosis. *Cancer Treat Rev,* 51**,** 1-9.

VISVADER, J. E. 2009. Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. *Genes Dev,* 23**,** 2563-77.

VISVADER, J. E. 2011. Cells of origin in cancer. *Nature,* 469**,** 314-22.

VIVANCO, I. & SAWYERS, C. L. 2002. The phosphatidylinositol 3-Kinase AKT pathway in human cancer. *Nat Rev Cancer,* 2**,** 489-501.

VODUC, K. D., CHEANG, M. C., TYLDESLEY, S., GELMON, K., NIELSEN, T. O. & KENNECKE, H. 2010. Breast cancer subtypes and the risk of local and regional relapse. *J Clin Oncol,* 28**,** 1684-91.

VOGELSTEIN, B., PAPADOPOULOS, N., VELCULESCU, V. E., ZHOU, S., DIAZ, L. A., JR. & KINZLER, K. W. 2013. Cancer genome landscapes. *Science,* 339**,** 1546-58.

VOUSDEN, K. H. & PRIVES, C. 2009. Blinded by the Light: The Growing Complexity of p53. *Cell,* 137**,** 413-31.

WADDINGTON, C. H. 2012. The epigenotype. 1942. *Int J Epidemiol,* 41**,** 10-3.

WADE, M., LI, Y. C. & WAHL, G. M. 2013. MDM2, MDMX and p53 in oncogenesis and cancer therapy. *Nat Rev Cancer,* 13**,** 83-96.

WALTON, J. B., FARQUHARSON, M., MASON, S., PORT, J., KRUSPIG, B., DOWSON, S., STEVENSON, D., MURPHY, D., MATZUK, M., KIM, J., COFFELT, S., BLYTH, K. & MCNEISH, I. A. 2017. CRISPR/Cas9-derived models of ovarian high grade serous carcinoma targeting Brca1, Pten and Nf1, and correlation with platinum sensitivity. *Sci Rep,* 7**,** 16827.

WANG, B., ZHAO, M. Z., CUI, N. P., LIN, D. D., ZHANG, A. Y., QIN, Y., LIU, C. Y., YAN, W. T., SHI, J. H. & CHEN, B. P. 2015. Kruppel-like factor 4 induces apoptosis and inhibits tumorigenic progression in SK-BR-3 breast cancer cells. *FEBS Open Bio,* 5**,** 147-54.

WANG, H., LA RUSSA, M. & QI, L. S. 2016. CRISPR/Cas9 in Genome Editing and Beyond. *Annu Rev Biochem,* 85**,** 227-64.

WANG, H., YANG, H., SHIVALILA, C. S., DAWLATY, M. M., CHENG, A. W., ZHANG, F. & JAENISCH, R. 2013. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell,* 153**,** 910-8.

WANG, X., ZHAO, Y., XIAO, Z., CHEN, B., WEI, Z., WANG, B., ZHANG, J., HAN, J., GAO, Y., LI, L., ZHAO, H., ZHAO, W., LIN, H. & DAI, J. 2009. Alternative translation of OCT4 by an internal ribosome entry site and its novel function in stress response. *Stem Cells,* 27**,** 1265-75.

WANG, Y., BASKERVILLE, S., SHENOY, A., BABIARZ, J. E., BAEHNER, L. & BLELLOCH, R. 2008. Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. *Nat Genet,* 40**,** 1478-83.

WANG, Y., CHEN, J., HU, J. L., WEI, X. X., QIN, D., GAO, J., ZHANG, L., JIANG, J., LI, J. S., LIU, J., LAI, K. Y., KUANG, X., ZHANG, J., PEI, D. & XU, G. L. 2011. Reprogramming of mouse and human somatic cells by high-performance engineered factors. *EMBO Rep,* 12**,** 373-8.

WANG, Y., WATERS, J., LEUNG, M. L., UNRUH, A., ROH, W., SHI, X., CHEN, K., SCHEET, P., VATTATHIL, S., LIANG, H., MULTANI, A., ZHANG, H., ZHAO, R., MICHOR, F., MERIC-BERNSTAM, F. & NAVIN, N. E. 2014. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature,* 512**,** 155-60.

WANG, Y. H., LIU, S., ZHANG, G., ZHOU, C. Q., ZHU, H. X., ZHOU, X. B., QUAN, L. P., BAI, J. F. & XU, N. Z. 2005. Knockdown of c-Myc expression by RNAi inhibits MCF-7 breast tumor cells growth in vitro and in vivo. *Breast Cancer Res,* 7**,** R220-8.

WARE, C. B., WANG, L., MECHAM, B. H., SHEN, L., NELSON, A. M., BAR, M., LAMBA, D. A., DAUPHIN, D. S., BUCKINGHAM, B., ASKARI, B., LIM, R., TEWARI, M., GARTLER, S. M., ISSA, J. P., PAVLIDIS, P., DUAN, Z. & BLAU, C. A. 2009. Histone deacetylase inhibition elicits an evolutionarily conserved self-renewal program in embryonic stem cells. *Cell Stem Cell,* 4**,** 359-69.

WEI, D., WANG, L., YAN, Y., JIA, Z., GAGEA, M., LI, Z., ZUO, X., KONG, X., HUANG, S. & XIE, K. 2016. KLF4 Is Essential for Induction of Cellular Identity Change and Acinar-to-Ductal Reprogramming during Early Pancreatic Carcinogenesis. *Cancer Cell,* 29**,** 324-38.

WEINHOLD, B. 2006. Epigenetics: the science of change. *Environ Health Perspect,* 114**,** A160-7.

WEN, W., ZHANG, J. P., XU, J., SU, R. J., NEISES, A., JI, G. Z., YUAN, W., CHENG, T. & ZHANG, X. B. 2016a. Enhanced Generation of Integration-free iPSCs from Human Adult Peripheral Blood Mononuclear Cells with an Optimal Combination of Episomal Vectors. *Stem Cell Reports,* 6**,** 873-84.

WEN, W., ZHANG, J. P., XU, J., SU, R. J., NEISES, A., JI, G. Z., YUAN, W., CHENG, T. & ZHANG, X. B. 2016b. Enhanced Generation of Integration-free iPSCs from Human Adult Peripheral Blood Mononuclear Cells with an Optimal Combination of Episomal Vectors. *Stem Cell Reports,* 6**,** 873-884.

WERNIG, M., MEISSNER, A., FOREMAN, R., BRAMBRINK, T., KU, M. C., HOCHEDLINGER, K., BERNSTEIN, B. E. & JAENISCH, R. 2007. In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature,* 448**,** 318-U2.

WESSELSCHMIDT, R. L. 2011. The teratoma assay: an in vivo assessment of pluripotency. *Methods Mol Biol,* 767**,** 231-41.

WHITE, A. C. & LOWRY, W. E. 2015. Refining the role for adult stem cells as cancer cells of origin. *Trends Cell Biol,* 25**,** 11-20.

WHO. Feb 2017 2017. *RE: Cancer Fact Sheet.*

WIDSCHWENDTER, M., SIEGMUND, K. D., MULLER, H. M., FIEGL, H., MARTH, C., MULLER-HOLZNER, E., JONES, P. A. & LAIRD, P. W. 2004. Association of breast cancer DNA methylation profiles with hormone receptor status and response to tamoxifen. *Cancer Res,* 64**,** 3807-13.

WILSON, A. S., POWER, B. E. & MOLLOY, P. L. 2007. DNA hypomethylation and human diseases. *Biochim Biophys Acta,* 1775**,** 138-62.

WONG, C. C., QIAN, Y. & YU, J. 2017. Interplay between epigenetics and metabolism in oncogenesis: mechanisms and therapeutic approaches. *Oncogene,* 36**,** 3359-3374.

WRIGHT, M. H., CALCAGNO, A. M., SALCIDO, C. D., CARLSON, M. D., AMBUDKAR, S. V. & VARTICOVSKI, L. 2008. Brca1 breast tumors contain distinct CD44+/CD24- and CD133+ cells with cancer stem cell characteristics. *Breast Cancer Res,* 10**,** R10.

WU, J. & IZPISUA BELMONTE, J. C. 2015. Dynamic Pluripotent Stem Cell States and Their Applications. *Cell Stem Cell,* 17**,** 509-25.

WU, X., RENUSE, S., SAHASRABUDDHE, N. A., ZAHARI, M. S., CHAERKADY, R., KIM, M. S., NIRUJOGI, R. S., MOHSENI, M., KUMAR, P., RAJU, R., ZHONG, J., YANG, J., NEISWINGER, J., JEONG, J. S., NEWMAN, R., POWERS, M. A., SOMANI, B. L., GABRIELSON, E., SUKUMAR, S., STEARNS, V., QIAN, J., ZHU, H., VOGELSTEIN, B., PARK, B. H. & PANDEY, A. 2014. Activation of diverse signalling pathways by oncogenic PIK3CA mutations. *Nat Commun,* 5**,** 4961.

WYKE, J. A., BEAMAND, J. A. & VARMUS, H. E. 1980. Factors affecting phenotypic reversion of rat cells transformed by avian sarcoma virus. *Cold Spring Harb Symp Quant Biol,* 44 Pt 2**,** 1065-75.

XU, C., INOKUMA, M. S., DENHAM, J., GOLDS, K., KUNDU, P., GOLD, J. D. & CARPENTER, M. K. 2001a. Feeder-free growth of undifferentiated human embryonic stem cells. *Nat Biotechnol,* 19**,** 971-4.

XU, C. H., INOKUMA, M. S., DENHAM, J., GOLDS, K., KUNDU, P., GOLD, J. D. & CARPENTER, M. K. 2001b. Feeder-free growth of undifferentiated human embryonic stem cells. *Nature Biotechnology,* 19**,** 971-974.

XU, X., DUAN, D. & CHEN, S. J. 2017. CRISPR-Cas9 cleavage efficiency correlates strongly with target-sgRNA folding stability: from physical mechanism to off-target assessment. *Sci Rep,* 7**,** 143.

XUE, G., YAN, H. L., ZHANG, Y., HAO, L. Q., ZHU, X. T., MEI, Q. & SUN, S. H. 2015. c-Myc-mediated repression of miR-15-16 in hypoxia is induced by increased HIF-2alpha and promotes tumor angiogenesis and metastasis by upregulating FGF2. *Oncogene,* 34**,** 1393-406.

YALLOWITZ, A. R., ALEXANDROVA, E. M., TALOS, F., XU, S., MARCHENKO, N. D. & MOLL, U. M. 2014. p63 is a prosurvival factor in the adult mammary gland during post-lactational involution, affecting PI-MECs and ErbB2 tumorigenesis. *Cell Death Differ,* 21**,** 645-54.

YAMANAKA, S. 2012a. Induced Pluripotent Stem Cells: Past, Present, and Future. *Cell Stem Cell,* 10**,** 678-684.

YAMANAKA, S. 2012b. Induced pluripotent stem cells: past, present, and future. *Cell Stem Cell,* 10**,** 678-84.

YAMANAKA, S. & TAKAHASHI, K. 2006. [Induction of pluripotent stem cells from mouse fibroblast cultures]. *Tanpakushitsu Kakusan Koso,* 51**,** 2346-51.

YAN, T., MIZUTANI, A., CHEN, L., TAKAKI, M., HIRAMOTO, Y., MATSUDA, S., SHIGEHIRO, T., KASAI, T., KUDOH, T., MURAKAMI, H., MASUDA, J., HENDRIX, M. J., STRIZZI, L., SALOMON, D. S., FU, L. & SENO, M. 2014. Characterization of cancer stem-like cells derived from mouse induced pluripotent stem cells transformed by tumor-derived extracellular vesicles. *J Cancer,* 5**,** 572-84.

YANG, L., GUELL, M., BYRNE, S., YANG, J. L., DE LOS ANGELES, A., MALI, P., AACH, J., KIM-KISELAK, C., BRIGGS, A. W., RIOS, X., HUANG, P. Y., DALEY, G. & CHURCH, G. 2013. Optimization of scarless human stem cell genome editing. *Nucleic Acids Res,* 41**,** 9049-61.

YANG, Y. 2012. Wnt signaling in development and disease. *Cell Biosci,* 2**,** 14.

YANG, Z., DONG, D., ZHANG, Z., CRABBE, M. J., WANG, L. & ZHONG, Y. 2012. Preferential regulation of stably expressed genes in the human genome suggests a widespread expression buffering role of microRNAs. *BMC Genomics,* 13 Suppl 7**,** S14.

YAO, K., GOLDSCHMIDT, R., TURK, M., WESSELING, J., STORK-SLOOTS, L., DE SNOO, F. & CRISTOFANILLI, M. 2015. Molecular subtyping improves diagnostic stratification of patients with primary breast cancer into prognostically defined risk groups. *Breast Cancer Res Treat,* 154**,** 81-8.

YATES, J., WARREN, N., REISMAN, D. & SUGDEN, B. 1984. A cis-acting element from the Epstein-Barr viral genome that permits stable replication of recombinant plasmids in latently infected cells. *Proc Natl Acad Sci U S A,* 81**,** 3806-10.

YATES, J. L., WARREN, N. & SUGDEN, B. 1985. Stable replication of plasmids derived from Epstein-Barr virus in various mammalian cells. *Nature,* 313**,** 812-5.

YE, F., ZHOU, C., CHENG, Q., SHEN, J. & CHEN, H. 2008. Stem-cell-abundant proteins Nanog, Nucleostemin and Musashi1 are highly expressed in malignant cervical epithelial cells. *BMC Cancer,* 8**,** 108.

YE, P., SIMONIAN, M., NADKARNI, M. A., DECARLO, A. A., CHAPPLE, C. C. & HUNTER, N. 2005. Identification of epithelial auto-antigens associated with periodontal disease. *Clin Exp Immunol,* 139**,** 328-37.

YE, Z., ZHAN, H., MALI, P., DOWEY, S., WILLIAMS, D. M., JANG, Y. Y., DANG, C. V., SPIVAK, J. L., MOLITERNO, A. R. & CHENG, L. 2009. Human-induced pluripotent stem cells from blood cells of healthy donors and patients with acquired blood disorders. *Blood,* 114**,** 5473-80.

YERSAL, O. & BARUTCA, S. 2014. Biological subtypes of breast cancer: Prognostic and therapeutic implications. *World J Clin Oncol,* 5**,** 412-24.

YOKOTA, J. 2000. Tumor progression and metastasis. *Carcinogenesis,* 21**,** 497-503.

YOSHIOKA, N., GROS, E., LI, H. R., KUMAR, S., DEACON, D. C., MARON, C., MUOTRI, A. R., CHI, N. C., FU, X. D., YU, B. D. & DOWDY, S. F. 2013. Efficient generation of human iPSCs by a synthetic self-replicative RNA. *Cell Stem Cell,* 13**,** 246-54.

YU, F., LI, J., CHEN, H., FU, J., RAY, S., HUANG, S., ZHENG, H. & AI, W. 2011a. Kruppel-like factor 4 (KLF4) is required for maintenance of breast cancer stem cells and for cell migration and invasion. *Oncogene,* 30**,** 2161-72.

YU, J., CHAU, K. F., VODYANIK, M. A., JIANG, J. & JIANG, Y. 2011b. Efficient feeder-free episomal reprogramming with small molecules. *PLoS One,* 6**,** e17557.

YU, J., HU, K., SMUGA-OTTO, K., TIAN, S., STEWART, R., SLUKVIN, II & THOMSON, J. A. 2009. Human induced pluripotent stem cells free of vector and transgene sequences. *Science,* 324**,** 797-801.

YU, J. & THOMSON, J. A. 2008. Pluripotent stem cell lines. *Genes Dev,* 22**,** 1987-97.

YU, J., VODYANIK, M. A., SMUGA-OTTO, K., ANTOSIEWICZ-BOURGET, J., FRANE, J. L., TIAN, S., NIE, J., JONSDOTTIR, G. A., RUOTTI, V., STEWART, R., SLUKVIN, II & THOMSON, J. A. 2007a. Induced pluripotent stem cell lines derived from human somatic cells. *Science,* 318**,** 1917-20.

YU, J. Y., VODYANIK, M. A., SMUGA-OTTO, K., ANTOSIEWICZ-BOURGET, J., FRANE, J. L., TIAN, S., NIE, J., JONSDOTTIR, G. A., RUOTTI, V., STEWART, R., SLUKVIN, I. I. & THOMSON, J. A. 2007b. Induced pluripotent stem cell lines derived from human somatic cells. *Science,* 318**,** 1917-1920.

ZACK, T. I., SCHUMACHER, S. E., CARTER, S. L., CHERNIACK, A. D., SAKSENA, G., TABAK, B., LAWRENCE, M. S., ZHSNG, C. Z., WALA, J., MERMEL, C. H., SOUGNEZ, C., GABRIEL, S. B., HERNANDEZ, B., SHEN, H., LAIRD, P. W., GETZ, G., MEYERSON, M. & BEROUKHIM, R. 2013. Pan-cancer patterns of somatic copy number alteration. *Nat Genet,* 45**,** 1134-40.

ZAKRZEWSKI, F., WEISSHAAR, B., FUCHS, J., BANNACK, E., MINOCHE, A. E., DOHM, J. C., HIMMELBAUER, H. & SCHMIDT, T. 2011. Epigenetic profiling of heterochromatic satellite DNA. *Chromosoma,* 120**,** 409-22.

ZAMBIDIS, E., ZIMMERLIN, L., PARK, T. S., HUO, J. S., VERMA, K., PATHER, S., TALBOT, C., AGARWAL, J., STEPPAN, D., CONSIDINE, M., GUO, H., ZHANG, Y., COPE, L., FRIEDMAN, A. & BAYLIN, S. 2016. Stable Reversion of Conventional Human Pluripotent Stem Cells to a Mouse ESC-like Naive Ground State Erases Somatic Donor Epigenetic Memory and Significantly Improves Their Hemato-Vascular Differentiation Potency. *Biology of Blood and Marrow Transplantation,* 22**,** S426-S426.

ZANG, C., WANG, T., DENG, K., LI, B., HU, S., QIN, Q., XIAO, T., ZHANG, S., MEYER, C. A., HE, H. H., BROWN, M., LIU, J. S., XIE, Y. & LIU, X. S. 2016. High-dimensional genomic data bias correction and data integration using MANCIE. *Nat Commun,* 7**,** 11305.

ZBINDEN, M., DUQUET, A., LORENTE-TRIGOS, A., NGWABYT, S. N., BORGES, I. & RUIZ I ALTABA, A. 2010. NANOG regulates glioma stem cells and is essential in vivo acting in a cross-functional network with GLI1 and p53. *EMBO J,* 29**,** 2659-74.

ZEHIR, A., BENAYED, R., SHAH, R. H., SYED, A., MIDDHA, S., KIM, H. R., SRINIVASAN, P., GAO, J., CHAKRAVARTY, D., DEVLIN, S. M., HELLMANN, M. D., BARRON, D. A., SCHRAM, A. M., HAMEED, M., DOGAN, S., ROSS, D. S., HECHTMAN, J. F., DELAIR, D. F., YAO, J., MANDELKER, D. L., CHENG, D. T., CHANDRAMOHAN, R., MOHANTY, A. S., PTASHKIN, R. N., JAYAKUMARAN, G., PRASAD, M., SYED, M. H., REMA, A. B., LIU, Z. Y., NAFA, K., BORSU, L., SADOWSKA, J., CASANOVA, J., BACARES, R., KIECKA, I. J., RAZUMOVA, A., SON, J. B., STEWART, L., BALDI, T., MULLANEY, K. A., AL-AHMADIE, H., VAKIANI, E., ABESHOUSE, A. A., PENSON, A. V., JONSSON, P., CAMACHO, N., CHANG, M. T., WON, H. H., GROSS, B. E., KUNDRA, R., HEINS, Z. J., CHEN, H. W., PHILLIPS, S., ZHANG, H., WANG, J., OCHOA, A., WILLS, J., EUBANK, M., THOMAS, S. B., GARDOS, S. M., REALES, D. N., GALLE, J., DURANY, R., CAMBRIA, R., ABIDA, W., CERCEK, A., FELDMAN, D. R., GOUNDER, M. M., HAKIMI, A. A., HARDING, J. J., IYER, G., JANJIGIAN, Y. Y., JORDAN, E. J., KELLY, C. M., LOWERY, M. A., MORRIS, L. G. T., OMURO, A. M., RAJ, N., RAZAVI, P., SHOUSHTARI, A. N., SHUKLA, N., SOUMERAI, T. E., VARGHESE, A. M., YAEGER, R., COLEMAN, J., BOCHNER, B., RIELY, G. J., SALTZ, L. B., SCHER, H. I., SABBATINI, P. J., ROBSON, M. E., KLIMSTRA, D. S., TAYLOR, B. S., BASELGA, J., SCHULTZ, N., HYMAN, D. M., ARCILA, M. E., SOLIT, D. B., LADANYI, M. & BERGER, M. F. 2017a. Erratum:

Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med,* 23**,** 1004.

ZEHIR, A., BENAYED, R., SHAH, R. H., SYED, A., MIDDHA, S., KIM, H. R., SRINIVASAN, P., GAO, J., CHAKRAVARTY, D., DEVLIN, S. M., HELLMANN, M. D., BARRON, D. A., SCHRAM, A. M., HAMEED, M., DOGAN, S., ROSS, D. S., HECHTMAN, J. F., DELAIR, D. F., YAO, J., MANDELKER, D. L., CHENG, D. T., CHANDRAMOHAN, R., MOHANTY, A. S., PTASHKIN, R. N., JAYAKUMARAN, G., PRASAD, M., SYED, M. H., REMA, A. B., LIU, Z. Y., NAFA, K., BORSU, L., SADOWSKA, J., CASANOVA, J., BACARES, R., KIECKA, I. J., RAZUMOVA, A., SON, J. B., STEWART, L., BALDI, T., MULLANEY, K. A., AL-AHMADIE, H., VAKIANI, E., ABESHOUSE, A. A., PENSON, A. V., JONSSON, P., CAMACHO, N., CHANG, M. T., WON, H. H., GROSS, B. E., KUNDRA, R., HEINS, Z. J., CHEN, H. W., PHILLIPS, S., ZHANG, H., WANG, J., OCHOA, A., WILLS, J., EUBANK, M., THOMAS, S. B., GARDOS, S. M., REALES, D. N., GALLE, J., DURANY, R., CAMBRIA, R., ABIDA, W., CERCEK, A., FELDMAN, D. R., GOUNDER, M. M., HAKIMI, A. A., HARDING, J. J., IYER, G., JANJIGIAN, Y. Y., JORDAN, E. J., KELLY, C. M., LOWERY, M. A., MORRIS, L. G. T., OMURO, A. M., RAJ, N., RAZAVI, P., SHOUSHTARI, A. N., SHUKLA, N., SOUMERAI, T. E., VARGHESE, A. M., YAEGER, R., COLEMAN, J., BOCHNER, B., RIELY, G. J., SALTZ, L. B., SCHER, H. I., SABBATINI, P. J., ROBSON, M. E., KLIMSTRA, D. S., TAYLOR, B. S., BASELGA, J., SCHULTZ, N., HYMAN, D. M., ARCILA, M. E., SOLIT, D. B., LADANYI, M. & BERGER, M. F. 2017b. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med,* 23**,** 703-713.

ZEINEDDINE, D., HAMMOUD, A. A., MORTADA, M. & BOEUF, H. 2014. The Oct4 protein: more than a magic stemness marker. *Am J Stem Cells,* 3**,** 74-82.

ZHANG, H., LIU, G., DZIUBINSKI, M., YANG, Z., ETHIER, S. P. & WU, G. 2008a. Comprehensive analysis of oncogenic effects of PIK3CA mutations in human mammary epithelial cells. *Breast Cancer Res Treat,* 112**,** 217-27.

ZHANG, J., WANG, X., LI, M., HAN, J., CHEN, B., WANG, B. & DAI, J. 2006. NANOGP8 is a retrogene expressed in cancers. *FEBS J,* 273**,** 1723-30.

ZHANG, J. P., LI, X. L., LI, G. H., CHEN, W., ARAKAKI, C., BOTIMER, G. D., BAYLINK, D., ZHANG, L., WEN, W., FU, Y. W., XU, J., CHUN, N., YUAN, W., CHENG, T. & ZHANG, X. B. 2017. Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. *Genome Biol,* 18**,** 35.

ZHANG, L., DENG, L., CHEN, F., YAO, Y., WU, B., WEI, L., MO, Q. & SONG, Y. 2014a. Inhibition of histone H3K79 methylation selectively inhibits proliferation, self-renewal and metastatic potential of breast cancer. *Oncotarget,* 5**,** 10665-77.

ZHANG, S., BALCH, C., CHAN, M. W., LAI, H. C., MATEI, D., SCHILDER, J. M., YAN, P. S., HUANG, T. H. & NEPHEW, K. P. 2008b. Identification and characterization of ovarian cancer-initiating cells from primary human tumors. *Cancer Res,* 68**,** 4311-20.

ZHANG, X., CRUZ, F. D., TERRY, M., REMOTTI, F. & MATUSHANSKY, I. 2013. Terminal differentiation and loss of tumorigenicity of human cancers via pluripotency-based reprogramming. *Oncogene,* 32**,** 2249-60, 2260 e1-21.

ZHANG, Z., GAO, Y., GORDON, A., WANG, Z. Z., QIAN, Z. & WU, W. S. 2011. Efficient generation of fully reprogrammed human iPS cells via polycistronic retroviral vector and a new cocktail of chemical compounds. *PLoS One,* 6**,** e26592.

ZHANG, Z., HONG, Y., XIANG, D., ZHU, P., WU, E., LI, W., MOSENSON, J. & WU, W. S. 2015. MicroRNA-302/367 cluster governs hESC self-renewal by dually regulating cell cycle and apoptosis pathways. *Stem Cell Reports,* 4**,** 645-57.

ZHANG, Z. & WU, W. S. 2013. Sodium butyrate promotes generation of human induced pluripotent stem cells through induction of the miR302/367 cluster. *Stem Cells Dev,* 22**,** 2268-77.

ZHANG, Z., XIANG, D. & WU, W. S. 2014b. Sodium butyrate facilitates reprogramming by derepressing OCT4 transactivity at the promoter of embryonic stem cell-specific miR-302/367 cluster. *Cell Reprogram,* 16**,** 130-9.

ZHAO, H., DAVIES, T. J., NING, J., CHANG, Y., SACHAMITR, P., SATTLER, S., FAIRCHILD, P. J. & HUANG, F. P. 2015. A highly optimized protocol for reprogramming cancer cells to pluripotency using nonviral plasmid vectors. *Cell Reprogram,* 17**,** 7-18.

ZHAO, J. J., LIU, Z., WANG, L., SHIN, E., LODA, M. F. & ROBERTS, T. M. 2005. The oncogenic properties of mutant p110alpha and p110beta phosphatidylinositol 3-kinases in human mammary epithelial cells. *Proc Natl Acad Sci U S A,* 102**,** 18443-8.

ZHAO, S., YUAN, Q., HAO, H., GUO, Y., LIU, S., ZHANG, Y., WANG, J., LIU, H., WANG, F., LIU, K., LING, E. A. & HAO, A. 2011. Expression of OCT4 pseudogenes in human tumours: lessons from glioma and breast carcinoma. *J Pathol,* 223**,** 672-82.

ZHAO, Y., YIN, X., QIN, H., ZHU, F., LIU, H., YANG, W., ZHANG, Q., XIANG, C., HOU, P., SONG, Z., LIU, Y., YONG, J., ZHANG, P., CAI, J., LIU, M., LI, H., LI, Y., QU, X., CUI, K., ZHANG, W., XIANG, T., WU, Y., ZHAO, Y., LIU, C., YU, C., YUAN, K., LOU, J., DING, M. & DENG, H. 2008. Two supporting factors greatly improve the efficiency of human iPSC generation. *Cell Stem Cell,* 3**,** 475-9.

ZHU, X., XU, Y., YU, S., LU, L., DING, M., CHENG, J., SONG, G., GAO, X., YAO, L., FAN, D., MENG, S., ZHANG, X., HU, S. & TIAN, Y. 2014. An efficient genotyping method for genome-modified animals and human cells generated with CRISPR/Cas9 system. *Sci Rep,* 4**,** 6420.

ZOU, Q., WU, M. J., ZHONG, L. W., FAN, Z. X., ZHANG, B., CHEN, Q. & MA, F. 2016. Development of a Xeno-Free Feeder-Layer System from Human Umbilical Cord Mesenchymal Stem Cells for Prolonged Expansion of Human Induced Pluripotent Stem Cells in Culture. *Plos One,* 11.

ZUO, T., LIU, T. M., LAN, X., WENG, Y. I., SHEN, R., GU, F., HUANG, Y. W., LIYANARACHCHI, S., DEATHERAGE, D. E., HSU, P. Y., TASLIM, C., RAMASWAMY, B., SHAPIRO, C. L., LIN, H. J., CHENG, A. S., JIN, V. X. & HUANG, T. H. 2011. Epigenetic silencing mediated through activated PI3K/AKT signaling in breast cancer. *Cancer Res,* 71**,** 1752-62.

ZURAWEL, R. H., CHIAPPA, S. A., ALLEN, C. & RAFFEL, C. 1998. Sporadic medulloblastomas contain oncogenic beta-catenin mutations. *Cancer Res,* 58**,** 896-9.