

Defining the role of miRNAs regulated by the pluripotency gene NANOG in the development and progression of Testicular Germ Cell Tumours.

By Ryan Patrick Cardenas, BSc (Hons), MRes

Faculty of Medicine and Health Sciences
SVMS, University of Nottingham
Sutton Bonington Campus

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

September 2017

Abstract

Testicular germ cell tumours (TGCTs) are the most common form of cancer diagnosed in young men in the developed world (Ferlay, 2010). The underlying mechanisms responsible for the development and progression of these tumours are not well understood. However, it is believed that *in utero* risk factors may contribute to the disease. TGCTs are believed to be derived from early stage primordial germ cells (PGCs) that failed to mature during embryonic development and remain in an immature state into adulthood (Almstrup *et al.*, 2010; Novotny *et al.*, 2012). Prior to the maturation process, PGCs express high levels of pluripotency factors, including the homeobox transcription factor NANOG (Yamaguchi *et al.*, 2005; Irie *et al.*, 2015). In addition to this, *NANOG* is highly expressed in TGCTs and amplifications of the *NANOG* locus is common in invasive TGCTs, suggesting a role in tumour malignancy (Clark *et al.*, 2004). However, the function of *NANOG* in TGCTs remains still unknown. *NANOG* knockdown experiments in embryonic stem cells (ESCs) identified a key role for NANOG in promoting cell proliferation and pluripotency (Mitsui *et al.*, 2003; Torres and Watt, 2008; Zhang *et al.*, 2009). NANOG has also been shown to regulate microRNAs (miRNAs) that are key for the regulation of the cell cycle and differentiation of ESCs (Barroso-delJesus *et al.*, 2008; Card *et al.*, 2008; Marson *et al.*, 2008).

This study identified miRNAs regulated by NANOG that may contribute to the pathogenesis of TGCTs. Integrated bioinformatic analyses of NANOG ChIP-seq

(hESCs), miRNA microarray (TGCT vs testes) and RNA-seq (TGCT patient samples) identified 4 miRNAs (*miR-512, miR-183, miR-887* and *miR-9*) that were differentially expressed in TGCTs that may be regulated by NANOG. RT-QPCR analysis of TGCT cell lines validated *miR-9* to be expressed at low levels in TGCTs and to be transcriptionally repressed by NANOG. A luciferase reporter system identified that *miR-9-5p* repressed the expression of *LIN28A, NR5A2, CCNE1, KDM7A* and *KDM4C* in HEK293T cell. TCam-2 cells over-expressing *miR-9* were generated and the transcriptome analysed using RNA sequencing. RNA-seq analysis of TCam-2 cells overexpressing *miR-9* identified 209 differentially expressed genes, 3 of which were predicted *miR-9* targets (*SIK1, CD34* and *SMN2*).

This study provides novel molecular insights into the role of NANOG in TGCTs. In particular, it shows that *miR-9* is downregulated in TGCTs because of NANOG-mediated repression. In addition, it identified potential *miR-9* targets that may play an important role in TGCT tumorigenesis. Therefore, the project highlights *miR-9* as a potential biomarker for TGCT detection and a target for therapeutic treatment for TGCTs in the future.

Acknowledgements

The work presented in this thesis was funded by the Biotechnology and Biological Sciences Research Council (BBSRC) and the University Of Nottingham which has facilitated this PhD.

I would first like to express the utmost gratitude to both my supervisors Cinzia Allegrucci and Nigel Mongan, for their patience, guidance and expertise and without them I would not have made it through the PhD. Both have been amazing and inspiring and have changed the way I think and act for the better.

I would also like to thank the members of the technical staff in the Vet School, in particular Belinda Wang in addition to my colleagues Leong, Ahmad, Veronika and Lina and the kind advice and support making the PhD a more pleasant experience.

Finally, a special thanks to my friends Caz, Glenn, Tessa, Leah and Rob as well as my family members Jo, Pablo, Kian and David who have been invaluable in supporting me through this entire process.

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 procedure and work described was performed by me.

Ryan Cardenas

Date:

Table of Contents

Abstract ---------- **i**
Acknowledgements ---------- **iii**
Declaration ---------- **iv**
List of Figures ---------- **ix**
List of Tables ---------- **xvi**
Abbreviations ---------- **xviii**
Chapter 1: Introduction ---------- **1**
1.1 Thesis Rationale, Hypothesis, Aims and Objectives ------ 1
1.2 Testicular Germ Cell Tumours (TGCTs) ---------- 4
1.2.1 TGCT Epidemiology ---------- 4
1.2.2 TGCT Aetiology ---------- 7
1.3 Diagnosis and Staging of TGCTs ---------- 12
1.3.1 Biomarkers ---------- 13
1.3.2 Staging ---------- 15
1.4 Treatment of TGCTs ---------- 19
1.5 Developmental origins of TGCTs ---------- 23
1.5.1 Carcinoma *in situ* (CIS) cells ---------- 24
1.5.2 Seminomas ---------- 25
1.5.3 Embryonal Carcinoma Cells ---------- 26
1.6 PGC Development ---------- 28
1.6.1 PGC Specification ---------- 28
1.6.2 PGC Reprogramming ---------- 33
1.6.3 PGC Maturation ---------- 36
1.6.4 Germ Cell Differentiation ---------- 37
1.7 The role of NANOG in PGCs ---------- 39
1.8 MicroRNA Regulation ---------- 43
1.8.1 MicroRNAs Overview ---------- 43
1.8.2 MicroRNA regulation in PGCs ---------- 45
1.8.3 MicroRNA regulation in ESCs ---------- 46
1.9 MicroRNAs and TGCTs ---------- 48
Chapter 2: Methods ---------- **51**
2.1 Bioinformatics Analyses ---------- 51
2.2.1 ChIP-seq Analysis ---------- 51

2.2.2 TGCT Microarray Analysis for miRNAs ---------- 53
2.2.3 miRNA Target Prediction ---------- 54
2.2.4 TGCT TCGA mRNA-seq and miRNA-seq Analysis --- 55
2.2.5 Clustering Analysis ---------- 56
2.2.6 RNA-seq Analysis ---------- 57
2.2.7 Gene ontology and Pathway Analysis ---------- 58
2.3 Testes RNA Samples ---------- 59
2.4 Cell Culture ---------- 59
2.4.1 TGCT Cell Culture ---------- 59
2.4.2 hESC Cell Culture ---------- 61
2.4.3 Counting of cells ---------- 62
2.4.4 Cryopreservation of cells ---------- 62
2.4.5 Thawing of cells ---------- 62
2.5 Generation of Transgenic Cell lines ---------- 63
2.5.1 HEK293T Culture ---------- 63
2.5.2 Production of Lentiviral particles ---------- 64
2.5.3 Antibiotic toxicity testing ---------- 65
2.5.3 Flow Cytometry (FACS) analysis ---------- 66
2.6 Luciferase Assay ---------- 67
2.6.1 Transfection of cells with pGL3 Plasmid ---------- 67
2.6.2 Transfection of cells with psiCHECK-2 Plasmid and miRNA mimics ---------- 67
2.6.3 Luciferase Analysis ---------- 68
2.7 Molecular Techniques and Analyses ---------- 69
2.7.1 RNA Extraction ---------- 69
2.7.2 cDNA Synthesis ---------- 70
2.7.3 RT-QPCR ---------- 71
2.7.4 Cloning ---------- 72
2.7.5 Restriction Enzyme Digest ---------- 73
2.7.6 Bacterial Transformation ---------- 74
2.7.7 Growing Bacteria ---------- 74
2.7.8 Freezing Bacterial Stocks ---------- 74
2.7.9 Plasmid purification by Miniprep ---------- 75
2.7.10 Gel and PCR Purification ---------- 76
2.7.11 Genomic DNA Extraction ---------- 77
2.7.12 Gel Electrophoresis ---------- 78

2.8 Statistics ---- 79
2.9 PCR Primers ---- 79
Chapter 3: Bioinformatics analysis of TGCT transcriptome and ChIP-Seq datasets ---- 82
3.1 Introduction ---- 82
3.2 Results ---- 90
3.2.1 Analysis of ChIP-Seq data ---- 90
3.2.2 Analysis of Microarray Data ---- 92
3.2.3 Analysis of The Cancer Genome Atlas TGCT dataset ---- 96
3.2.4 Overlapping of Bioinformatics Analyses ---- 102
3.2.5 Analysis of TCGA clinical data ---- 109
3.3 Discussion ---- 113
3.3.1 Identification of miRNAs potentially regulated by NANOG ---- 113
3.3.2 Conclusion ---- 118
3.4 Supplementary ---- 120
Chapter 4: Validation of selected TGCT miRNAs ---- 122
4.1 Introduction ---- 122
4.2 Results ---- 128
4.1.1 Characterisation of TGCT cell lines ---- 128
4.2.2 Expression analysis of miRNAs in TGCT cell lines 138
4.2.3 Characterising NANOG binding to the *miR-9* promoter ---- 142
4.2.4 Overexpression of *miR-9* in TGCTs ---- 155
4.3 Discussion ---- 163
4.3.1 Conclusion ---- 171
4.4 Supplementary ---- 172
Chapter 5: Analysis of chosen miRNA ---- 175
5.1 Introduction ---- 175
5.2 Results ---- 179
5.2.1 Identifying miRNA Predicted Targets ---- 179
5.2.2 Validation of *miR-9* Targets (293T) ---- 182
5.2.3 Validation of *miR-9* Targets in TGCTs ---- 188
5.2.4 Transcriptomic Analysis ---- 191
5.3 Discussion ---- 201
5.4 Supplementary ---- 213

Chapter 6: General Discussion ---------- 216
6.1 Importance of the study: role of miRNA in TGCTs ---- 217
6.2 The role of *miR-9* in TGCTs ---------- 221
6.3 Limitations of this Study ---------- 226
6.4 Future work ---------- 229
Conclusions ---------- 230
Bibliography ---------- 232
Rights and permission to use Figures ---------- 271
Professional Internships for PhD Students ---------- 273
Reflection Form ---------- 273

List of Figures

Figures	Description	Page
Figure 1.1	A map depicting the distribution of TGCTs worldwide by age standardised incidence rate.	4
Figure 1.2	A line plot for the male mortality rates in the UK.	5
Figure 1.3	A depiction of normal PGC development, in comparison to PGCs that fail to mature and give rise to TGCTs.	24
Figure 1.4	Development of PGCs from specification to differentiation.	28
Figure 1.5	An overview for the induction of hPGCLCs.	29
Figure 1.6	A depiction for the specification of hPGCLCs.	30
Figure 1.7	A depiction for epigenetic histone modifications.	33
Figure 1.8	A model for NANOG regulation in ESCs and TGCTs.	37
Figure 1.9	The biogenesis of miRNAs.	40
Figure 1.10	Dicer knockout and its effect on differentiation in ESCs.	43
Figure 2.1	Pflafl equation.	67
Figure 3.1	A flow chart depicting the bioinformatics analysis conducted to produce a list of miRNAs that may be regulated by NANOG and their predicted targets.	86
Figure 3.2	Analysis of ChIP-seq data (NANOG, OCT4 and SOX2).	88
Figure 3.3	Overlap of differentially expressed miRNAs in the microarray data (TGCT vs testes).	89

Figure 3.4	Pearson heatmap and PCA plot for TCGA TGCT mRNA gene expression.	93
Figure 3.5	Pearson heatmap and PCA plot for TCGA TGCT miRNA gene expression.	94
Figure 3.6	Expression of *NANOG, OCT4, SOX2* and *SOX17* in TCGA TGCT patient samples (RNA-seq).	95
Figure 3.7	Heatmap of differentially expressed miRNAs in *NANOG* high, medium and low groups (TCGA TGCTs).	97
Figure 3.8	Overlap of miRNAs from bioinformatics analyses.	99
Figure 3.9	Expression of chosen miRNAs in TCGA TGCT *NANOG* high, medium and low groups.	100
Figure 3.10	ChIP-seq binding profile for NANOG, OCT4, SOX2, H3K4me3 and H3K27me3 at *miR-9-2* locus.	104
Figure 3.11	ChIP-seq binding profile for NANOG, OCT4, SOX2, H3K4me3 and H3K27me3 at *miR-887* locus.	105
Figure 3.12	ChIP-seq binding profile for NANOG, OCT4, SOX2, H3K4me3 and H3K27me3 at *miR-512-1* locus.	106
Figure 3.13	ChIP-seq binding profile for NANOG, OCT4, SOX2, H3K4me3 and H3K27me3 at *miR-183* locus.	107
Figure 3.14	Expression of chosen miRNAs in TCGA TGCT subtype groups.	108
Figure 3.15	Expression of chosen miRNAs in TCGA TGCT CIS absence or presence groups.	109

Figure 3.16	Expression of chosen miRNAs in TCGA TGCT groups by TGCT stage.	110
Figure 3.17	Expression of chosen miRNAs in TCGA TGCT in patients with a familial history of TGCTs.	111
Figure S3.1	The expression of *NANOG* for TCGA samples for the groups; High, Medium or Low *NANOG* expression.	120
Figure 4.1	Bright field images of the TGCT cell lines.	128
Figure 4.2	Melting curves for the following gene amplicons; *NANOG, OCT4, SOX2, SOX17* and *β-ACTIN*.	130
Figure 4.3	Melting curves for the following gene amplicons; *miR-9-5p*, *miR-9-3p*, *miR-512-3p*, *miR-183-3p* and SNORD75 (U75).	131
Figure 4.4	Calculating primer efficiencies for *NANOG, OCT4, SOX2, SOX17* and *β-ACTIN*.	132
Figure 4.5	Calculating primer efficiencies for *miR-9-5p, miR-9-3p, miR-512-3p, miR-183-3p* and *SNORD75* (U75).	133
Figure 4.6	A pirate plot showing the Ct values across all TGCT samples for *18S, β-ACTIN* and *GAPDH*.	134
Figure 4.7	The expression profiles of *NANOG*, *OCT4* and *SOX2* in TGCT cell lines.	136
Figure 4.8	The expression profile of *SOX17* in TGCT cell lines.	137
Figure 4.9	A pirate plot showing the Ct values across all TGCT samples for *miR-103b, miR-26-5p, RNU6B, miR-26b-5p* and *SNORD75*.	138

Figure 4.10	The expression profiles of *miR-512-3p, miR-183-3p, miR-9-5p and miR-9-3p* in TGCT cell lines	140
Figure 4.11	A schematic showing the procedure by which lentiviral particles were created.	142
Figure 4.12	Puromycin selection experiment for HEK293T cells.	143
Figure 4.13	RT-PCR in HEK293T-WT and HEK293T-NANOG-OE cells.	143
Figure 4.14	A schematic illustrating the genomic region with NANOG enrichment that was amplified by RT-PCR and inserted into pGL3 plasmids.	144
Figure 4.15	The map for pGL3-Basic and pGL3-Promoter.	146
Figure 4.16	Cut pGL3 plasmids indicating correct *miR-9-2* insert.	147
Figure 4.17	Luciferase assay for NANOG binding (pGL3).	148
Figure 4.18	A schematic for pLKO-Tet-On plasmid and shRNA inserts.	149
Figure 4.19	Cut pLKO plasmid to indicate correct *NANOG* shRNA sequence insert.	151
Figure 4.20	Puromycin selection experiment for sh*NANOG* TCam-2 cells.	152
Figure 4.21	RT-QPCR for *NANOG* in *NANOG* KD TCam-2 cells.	153
Figure 4.22	RT-QPCR for *miR-9* in *NANOG* KD TCam-2 cells.	153
Figure 4.23	Cut pLKO plasmid to indicate *miR-9-5p* sequence insert.	155

Figure 4.24	Puromycin selection experiment for pLKO *miR-9-5p*Tcam-2 cells	156
Figure 4.25	RT-QPCR for *miR-9* in TCam-2 *miR-9-5p* (pLKO) overexpressing cells.	157
Figure 4.26	Enzyme digestions for the MDH1-PGK-GFP plasmids.	158
Figure 4.27	Fluorescent images for TCam-2 MDH1-PGK-GFP-miR9 and MDH1-PGK-GFP-Scrambled infected TCam-2 cells.	159
Figure 4.28	FACS analysis for TCam-2 MDH1-PGK-GFP cells.	160
Figure 4.29	RT-QPCR for *miR-9* in TCam-2 *miR-9* (MDH1-GFP) overexpressing cells.	161
Figure 4.30	A depiction of the *pre-miR-9* and the pLKO shRNA for *miR-9-5p.*	168
Figure S4.1	Sequencing results for the RT-QPCR amplicon for *miR-26p-5p, SNORD75, miR-9-5p, miR-9-3p* and *miR-512-3p.*	171
Figure S4.2	Sequencing results for the RT-QPCR amplicon for *NANOG, SOX2, LIN28A, KDM4C* and *GAPDH.*	172
Figure S4.3	Sequencing results for the amplicon for the pGL3-Basic-*miR-9*-2	173
Figure 5.1	Overlap of miRNA targets using five different prediction algorithms for *miR-9-5p* and *miR-9-3p.*	179
Figure 5.2	Overlap of *miR-9* predicted targets with genes that were highly expressed in TGCT patient samples with high NANOG expression	180

Figure 5.3	A depiction of amplifying a *miR-9* target region using PCR and inserting this downstream of a luciferase gene.	182
Figure 5.4	Cut psiCHECK-2 plasmids and run on gel.	183
Figure 5.5	PsiCHECK-2 luciferase assay for *NR5A2, KDM4C* and *ESRRB.*	185
Figure 5.6	PsiCHECK-2 luciferase assay for *Lin28A, KDM7A* and *CCNE1*.	186
Figure 5.7	Expression analysis for *LIN28A* and *KDM4C* in TCam-2 cells that overexpressed *miR-9*	187
Figure 5.8	A heatmap of 209 differentially expressed genes identified in TCam-2 Scrambled and TCam-2 *miR-9* groups	188
Figure 5.9	An overlap of *miR-9-5p* and *miR-9-3p* predicted targets with genes that were lowly expressed TCam-2 *miR-9* cells.	191
Figure 5.10	STRING analysis for genes differentially expressed between TCam-2 Scrambled and TCam-2 *miR-9* groups.	197
Figure 5.11	A plot for gene expression using RNA-seq samples for TCam-2 Scrambled and TCam-2 *miR-9*.	199
Figure 5.12	A depiction of NANOGs regulation of miR-9 in TGCTs.	211
Figure S5.1	Sequencing results for the amplicon for psiCHECK-2 plasmids for *miR-9* binding sites for the genes *KDM4C, NR5A2_1, NR5A2_2, ESRRB, KDM7A* and *CCNE1*.	212

Figure 6.1 An overview for NANOG regulation of *miR-9* and the downstream effects in different cell types. 224

List of Tables

Tables	Description	Page
Table 1.1	Tests recommended to be performed at diagnosis of TGCTs for staging according to EAU guidelines.	12
Table 1.2	Guidelines to classify the TGCT using tumour size and distance of metastases.	16
Table 1.3	Staging of TGCTs.	18
Table 1.4	Prognosis based staging system.	20
Table 2.1	ChIP-seq datasets used for analysis and their accessions.	50
Table 2.2	Descriptions for the stratification of the TCGA RNA-seq patient samples.	54
Table 2.3	Transfer plasmids used for viral transduction for over-expression or knockdown of genes.	63
Table 2.4	PCR cycling program for Q5 polymerase PCR.	71
Table 2.5	Viral plasmids used in this study and the enzymes used to cut these plasmids.	72
Table 2.6	Primers for miRNA RT-QPCR	78
Table 2.7	Primers for NANOG enriched genomic loci upstream of *miR-9-2* TSS (pGL3).	79
Table 2.8	Primers for psiCHECK-2 inserts for *miR-9* predicted targets.	79
Table 2.9	RT-QPCR primers for predicted *miR-9* targets.	80
Table 3.1	Heatmap of differentially expressed miRNAs from microarray data (TGCT vs testes).	92
Table 3.2	The stratification of TCGA TGCT samples based on NANOG expression.	99
Table 3.3	MicroRNAs identified after bioinformatics overlaps.	103
Table S3.1	MiRNAs with upstream binding for NANOG, OCT4 or SOX2 (10kb upstream of TSSs).	119
Table 4.1	Description of the TGCT cell lines.	127
Table 5.1	The predicted *miR-9* target gene list and the expected fragment size cloned into psiCHECK-2.	183
Table 5.2	Top 20 differentially expressed genes when comparing TCam-2 Scrambled and TCam-2 *miR-9*.	189

Table 5.3	Genes downregulated in TCam-2 *miR-9* cells that are also predicted targets of *miR-9-5p* or *miR-9-3p*.	192
Table 5.4	Gene ontology (GO) for biological processes for genes differentially expressed in TCam-2 Scrambled and TCam-2 *miR-9*.	193
Table 5.5	Pathway identified using the Reactome Pathway Database for genes differentially expressed in TCam-2 Scrambled and TCam-2 *miR-9*.	194
Table 5.6	Pathway identified using the KEGG Pathway Database for genes differentially expressed in TCam-2 Scrambled and TCam-2 *miR-9*.	195
Table 5.7	Differentially expressed genes (TCam-2 Scrambled vs TCam-2 *miR-9*) that were also enriched in TCGA TGCT patient samples.	196

Abbreviations

Abbreviation	Description
%	Percent
µl	Microlitre
AFP	Alpha-fetoprotein
AP	Alkaline phosphatase
BEP	Bleomycin, etoposide and cisplatin
bp	Base pairs
BPA	Bisphenol A
BSA	Bovine serum albumin
cDNA	Complementary DNA
CDS	Coding sequence region
ChIP	Chromatin immunoprecipitation
CIS	Carcinoma in situ
CO2	Carbon dioxide
Ct	Threshold cycle
CT	Computer topography
DAPI	4',6-diamidino-2-phenylindole
DMEM	Dulbecco's (Modified Eagle's) minimal essential medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
dNTP	Deoxyribonucleotide triphosphate
EB	Embryoid body
EC	Embryonal Carcinoma
ED	Endocrine Disruptor
EGC	Embryonic germ cells
EP	Etoposide and cisplatin
Epi	Epiblast
EpiLC	Epiblast like cells
EpiSC	Epiblast stem cells
ERK	Extracellular-signal-regulated kinase
ESC	Embryonic stem cells
ESCC	ESC specific cell cycle
FACs	Fluorescence Activated Cell Sorter or Flow cytometry
FBS	Fetal bovine serum
GEO	Gene expression omnibus
GFP	Green fluorescence protein
GP-130	Glycoprotein-130

Grb2	Growth factor receptor binding protein 2
GWAS	Genome wide association studies
HEK293	Human embryonic kidneys 293 cells
hESC	Human embryonic stem cells
hPSC	Human pluripotent stem cells
iPSC	Induced pluripotency stem cells
Kb	Kilobase
LDH	Lactate dehydrogenase
M	Molar
MEHP	Mono-2-ethylhexylphthalate
mESC	Mouse embryonic stem cells
miRNAs	MicroRNAs
ml	Millilitre
mM	Milimolar
MMC-MEF	Mitomycin C inactivated mouse embryonic fibroblast
MRI	Magnetic resonance imaging
mRNA	Messenger ribonucleic acid
ng	Nanogram
^{o}C	Degrees Celsius
PBS	Phosphate buffered saline
PC	Pricinciple component
PCA	Principle component analysis
PCR	Polymerase chain reaction
PGC	Primordial germ cell
PGCLCs	PGC-like cells
PLAP	Placental alkaline phosphatase
PSC	Pluripotent stem cells
RISC	RNA-inducing silencing complex
RNA	Ribonucleic acid
RNAi	RNA interference
RNA-seq	RNA sequencing
RPMI	Roswell Park Memorial Institute
RT	Room temperature
RT-PCR	Reverse transcriptase-polymerase chain reaction
SD	Standard Deviation
SDS	Sodium dodecyl sulfate
SEM	Seminoma
S.E.M	Standard error of the mean
Seq	sequencing
shRNA	Short hairpin RNA

SNP	Single nucleotide polymorphism
TCGA	The cancer genome atlas
TDS	Testicular dysgenesis syndrome
Tera	Teratoma
TGCTs	Testicular germ cell tumours
TSS	Transcriptional start site
UTR	Untranslated region
x g	Times gravity
YST	Yolk Sac Tumour
β-HCG	Beta human chorionic gonadotrophin

Chapter 1: Introduction

1.1 Thesis Rationale, Hypothesis, Aims and Objectives

Testicular germ cell tumour (TGCT) malignancy is a relatively rare disease, making up approximately 1% of all new cancers in males (Znaor *et al.*, 2014). However, TGCTs remains the most common form of cancer in males aged 15-44 in many countries, particularly developed countries, and over the past few decades incidence of this disease has been steadily increasing (Ferlay, 2010). There are two main subtypes of TGCTs: seminomas (SEM) which show histological and transcriptomic similarities to embryonic germ cells (Rashid *et al.*, 2004; Irie *et al.*, 2015) and non-seminomas, which possess extra-embryonal and embryonal features (Rapley *et al.*, 2009).

Both seminomas and embryonal carcinomas (ECs) are undifferentiated tumours, expressing high levels of the pluripotency markers NANOG and OCT4 (de Jong *et al.*, 2005; Hart *et al.*, 2005). Indeed, transcriptomic analysis of TGCT subtypes have identified that ECs have an expression profile more similar to embryonic stem cells (ESCs), whereas SEM have an expression profile more related to primordial germ cells (PGCs) (Sperger *et al.*, 2003; Irie *et al.*, 2015). Invasive TGCTs have amplifications of chromosome 12p, where *NANOG* is located, and as such it has been suggested that overexpression may play a role in TGCT malignancy (Clark *et al.*, 2004). NANOG knockdown in human ESCs (hESCs) increased cellular differentiation and decreased proliferation (Zhang *et al.*, 2009). Similar to this, conditional knockdown of Nanog in mouse PGCs contributes to reduced proliferation and, as a result, reduced PGC number

(Yamaguchi *et al.*, 2009). This indicated a clear role for NANOG in the regulation of pluripotency and proliferation of pluripotent cells.

In ESCs, pluripotency is maintained by expression of pluripotency genes and epigenetic silencing of pro-differentiation genes. MicroRNAs (miRNAs), together with other epigenetic modifications, play a key role in this regulation (Wang and Blelloch, 2011; Greve, Judson and Blelloch, 2013). Knockdown of key components of the miRNA biogenesis pathway, Dicer or Dgcr8, results in the downregulation of all canonical miRNA transcripts resulting in abnormal differentiation and reduced proliferation in mESCs (Kanellopoulou *et al.*, 2005; Wang *et al.*, 2007). Similar to this, conditional knockout of *Dicer* in mouse PGCs exhibited reduced proliferation and PGC number (Hayashi *et al.*, 2008). During early stages of mouse PGC development the *miR-290-295* cluster is highly expressed followed by its downregulation once PGCs reach the gonads (Hayashi *et al.*, 2008). This is of particular interest as the human homologue of the mouse *miR-290-295*, *miR-371-373*, is highly expressed in TGCTs (Murray *et al.*, 2011). *MiR-373* has been shown to target tumour suppressors such as *LATS2* in order to promote proliferation in TGCTs (Voorhoeve *et al.*, 2006). Pluripotency transcription factors, including NANOG, bind to the promoter regions of key miRNAs, including *miR-290-295*, in order to regulate miRNA expression in ESCs (Barroso-delJesus *et al.*, 2008; Card *et al.*, 2008; Marson *et al.*, 2008). This suggests that perhaps NANOG is also regulates *miR-371-373*, the human homologue of *miR-290-295*, in TGCTs. To our knowledge, there has not been a comprehensive study investigating NANOG regulation of miRNAs in TGCTs.

Therefore, the hypothesis of this project is that NANOG transcriptionally regulates a subset of miRNAs that are involved in TGCT development and progression.

Therefore, the aim of this project was to identify NANOG-regulated miRNAs that are important in TGCTs.

The objectives of this project were:

1. To identify a list of candidate miRNAs regulated by NANOG using available bioinformatic datasets of TGCT patient samples (Chapter 3).

2. To validate the expression of candidate miRNAs and their potential NANOG-mediated regulation using TGCT cell lines as models of TGCTs (Chapter 4).

3. To produce a list of candidate mRNA targets for chosen NANOG-regulated miRNA(s) and to validate miRNA-mRNA inhibition of translation in TGCT cell lines (Chapter 5).

4. To discuss the achievements and limitations of the study in order to assess its potential significance and future directions (Chapter 6).

1.2 Testicular Germ Cell Tumours (TGCTs)

1.2.1 TGCT Epidemiology

TGCTs are a relatively rare disease but are the most common neoplasms in males aged between 20 and 40 years old, with highest incidence rates in western and northern Europe (Figure 1.1), effecting 1-12 per 100,000 males per year (Rosen *et al.*, 2011; Shanmugalingam *et al.*, 2013; Znaor *et al.*, 2014). TGCTs can be classified histologically into three main groups; (I) seminomas, (II) non-seminomas and (III) spermatocytic seminomas (Sesterhenn and Davis Jr, 2004).

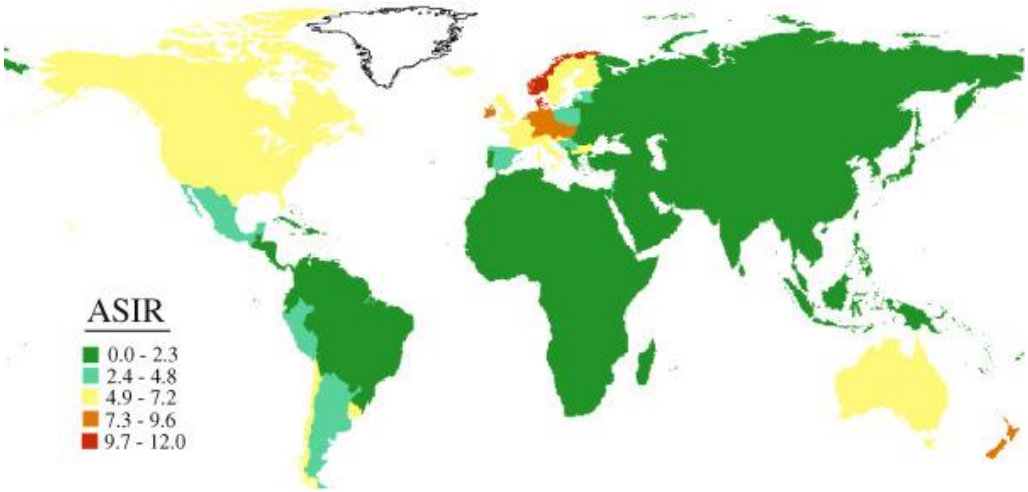


Figure 1.1 A map depicting the distribution of TGCTs worldwide by age standardised incidence rate (ASIR); red indicates a highest incidence of TGCTs followed by orange, yellow and green. Image adapted from (Rosen *et al.*, 2011).

Seminomas and non-seminoma are the most prevalent TGCT subtypes comprising of 56% and 43% of all TGCT tumours, respectively. In contrast, spermatocytic seminomas occur only in 0.6% of patients (Ries *et al.*, 2008). Both non-seminomas and seminomas develop in younger men aged 25-29 and 35-39, respectively, whereas spermatocytic seminomas patients are generally older (50-54yrs). (Ries *et al.*, 2008). Overall, TGCTs are most commonly diagnosed during adolescence and risk peaks between the ages of 30-34 years, this is in contrast to other cancers which typically show an increase in prevalence with age (Cancer Research UK, 2017). The prevalence of TGCTs has been increasing over the past few decades particularly in developed countries (Ferlay, 2010) as shown in Figure 1.2. TGCT prevalence in the UK showed an increase by 28% in males since the 1990s and had increased by 46% in Great Britain between 1979-1993 (Cancer Research UK, 2017). This increase has been predicted to increase further in the future (McGlynn and Cook, 2010).

Despite the increasing TGCT incidence, mortality due to TGCTs has reduced dramatically in the UK (Figure 1.2B). In fact TGCT mortality rates have been reduced by 82% since the 1970s (Cancer Research UK, 2017). This has been attributed in part to the introduction of treatment with platinum-based therapy in the late 1970s resulting in a 95% survival rate for TGCT patients and reducing mortality to 0.2-0.3 per 100,000 (Ferlay, 2010; Bosetti *et al.*, 2011; Znaor *et al.*, 2014). The decline in mortality, however, shows geographic variation with high-income countries achieving a greater reduction in mortality in relation to low-income countries (Ferlay, 2010). This indicates that socio-economic factors may be an associated risk factor for the development or diagnosis of TGCTs.

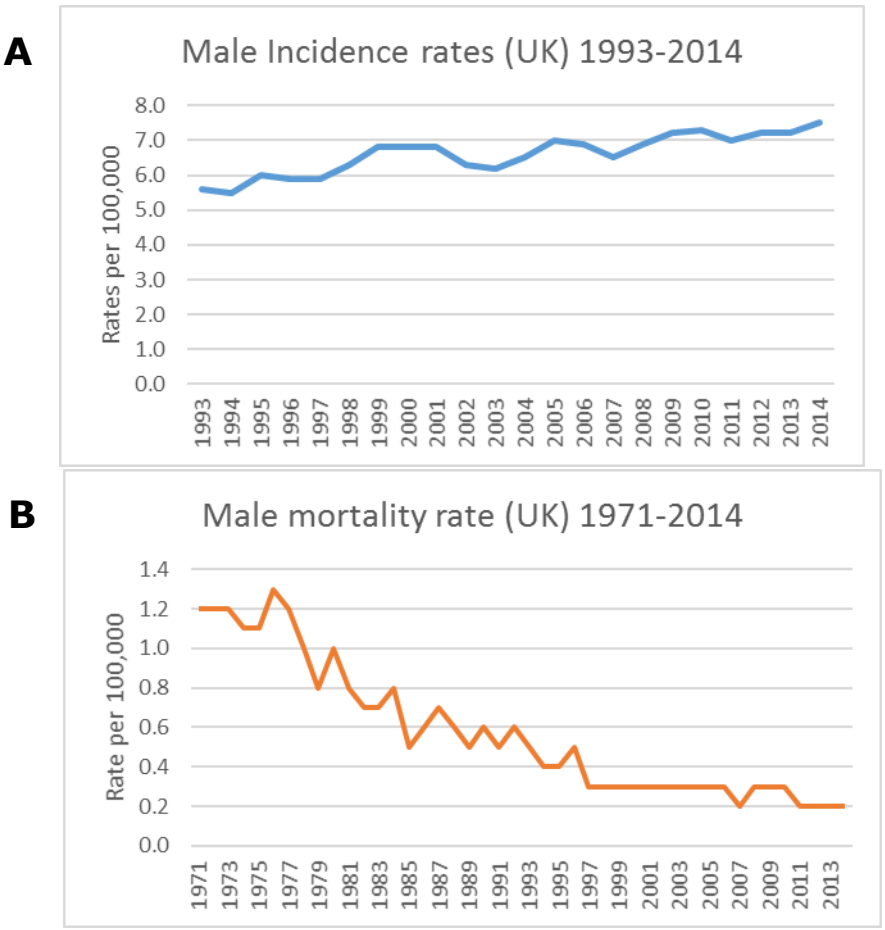


Figure 1.2 (A) A line plot for the male incidence rates in the UK from 1971-2014 and (B) the mortality rates of males in the UK from 1993-2014. Figure adapted from (Cancer Research UK, 2017).

1.2.2 TGCT Aetiology

The aetiology for TGCT is poorly understood. It is currently believed that *in* utero exposure to certain risk factors including maternal lifestyle (e.g. smoking and obesity) in addition to environmental factors (e.g. pollutants and endocrine disrupting chemicals) may perturb the development of the germ cells resulting in increased risk for developing TGCTs (Shanmugalingam *et al.*, 2013).

1.2.3 Risk Factors

Several risk factors have been recognised as associated with the development of TGCTs (McGlynn and Cook, 2009). The identification of risk factors is important, as some of these can be minimised, thus decreasing the prevalence of TGCTs. However some of these risk factors, such as genetic abnormalities (Sheikine *et al.*, 2012) and familial inheritance (Patel, Kvols and Richardson, 1990; Heimdal *et al.*, 1996), may be predictive for the identification of males at high risk of TGCTs thus allowing an early detection of the disease (Albers *et al.*, 2015). Early detection and correct treatment are crucial factors for the successful treatment of TGCTs (Albers *et al.*, 2015).

1.2.3.1 Cryptorchidism

Cryptorchidism is defined by either one or both testes being undescended and absent from the scrotum. Undescended testes are common in young males, affecting approximately 1 in 20 males at birth (Hutson, 2015). Pre-natal exposure to endocrine disruptors is associated with increased risk of developing cryptorchidism (Fernández *et al.*, 2016). In fact, patients with cryptorchidism are four times more likely to develop TGCTs, especially the seminoma subtype

(Reuter, 2005; Fernández *et al.*, 2016). However, the risk of TGCTs can be reduced if the cryptorchidism is corrected before the age of 10 using a surgical procedure called orchidopexy which involves moving the undescended testicle(s) into the scrotum (Walsh *et al.*, 2007). This suggests a role for testicular environment and the development of TGCTs.

1.2.3.2 Testicular Dysgenesis Syndrome

Skakkebaek and colleagues (2001) proposed that four conditions, including poor semen quality, TGCTs, undescended testes and hypospadias, are symptoms of an overarching testicular dysgenesis syndrome (TDS) (Skakkebaek, Meyts and Main, 2001). It is thought that these four conditions share risk factors during fetal development which include maldevelopment of the somatic testes cells, resulting in a disruption in hormone levels in adult life, in addition to endocrine disruptor exposure during germ cell development (McGlynn and Cook, 2009).

1.2.3.3 Family History

It has been shown that approximately 2% of all TGCT patient have a family history of TGCTs suggesting a genetic contribution to TGCT risk (Patel, Kvols and Richardson, 1990; Heimdal *et al.*, 1996; Shanmugalingam *et al.*, 2013). In fact, sons of men who had a farther who developed TGCTs have a 4-fold increase risk of developing TGCTs, whereas men with a brother who developed TGCT are eight times at risk (Cancer Research UK, 2017).

1.2.3.4 Genetics

The chromosome 12p region is the most common form of cytogenetic amplification found in TGCTs but it has also shown that different TGCT subtypes contain particular chromosomal abnormalities (Rodriguez et al. 2003; Looijenga et al. 2003). Non-seminoma TGCTs frequently show amplifications of proximal 17q and loss of 10q, whereas seminomas show high levels of chromosome 12p (Zafarana et al. 2003). Interestingly, these chromosomal regions contain many genes known to positively regulate pluripotency and proliferation in embryonic germ cells and ESCs, including *CCND2*, *NANOG* and *STELLAR* (Korkola et al. 2006), which have all been shown to be highly expressed in TGCTs (Korkola et al. 2006; Ben-David et al. 2014). This suggests that embryonic germ cells *in utero* continue to overexpress pluripotency genes contained within the ectopically amplified chromosomal regions contributing to their transformation into TGCTs. In addition to chromosomal abnormalities, it has been shown that TGCTs also contain gene mutations, the most frequent genes affected are *TP53, N-RAS, KIT, K-RAS* and *B-RAF* (Del-Mazo et al. 2013). Of particular interest are *KIT* and *TP53*. *KIT* has been shown crucial for the migration, proliferation and survival of germ cells during development, whereas p53 is a well characterised tumour suppressor protein and mutations in this gene is associated with a variety of cancers (Hollstein *et al.*, 1991; Olive *et al.*, 2004; Liu, Song and Xu, 2010).

A recent study using genome wide association studies (GWAS) identified 44 risk loci for TGCTs, mainly mapping to non-coding regions (Litchfield *et al.*, 2017). These genomic locations identified to have single nucleotide

polymorphisms (SNPs) which were shown to disrupt cis promoter-enhancer interactions and therefore to disrupt gene expression. Key tumour suppressor and germ cell specific genes were shown to be affected (Litchfield *et al.*, 2017). Of particular interest, SNP '*rs17153755*' is frequently enriched in TGCTs and associated with reduced *GATA4* expression (Litchfield *et al.*, 2017). This is consistent with findings which have shown that GATA4 is a tumour suppressor gene and is silenced in glioblastoma cell (Agnihotri *et al.*, 2011). Other genes identified to be regulated by the GWAS analysis included the germ cell markers *PRDM14, DAZL* and *KITLG* in addition to cell cycle regulators such as *LATS1* (Litchfield *et al.*, 2017).

1.2.3.5 Endocrine Disruptors

Endocrine disruptors (EDs) are a large collection of compounds that act as environmental pollutants which alter cellular homeostasis in living organisms through aberrant modulation of hormone function (Del-Mazo *et al.*, 2013). The main EDs that interact with hormonal pathways and have been linked to TGCTs are phthalates, bisphenol A, MEHP, polychlorinated biphenyls and organochlorides (Mester *et al.*, 2010). Although these are the most investigated compounds, there are hundreds of EDs present in the environment which led the World Health Organisation (WHO) highlighting this of special concern to both humans and wildlife worldwide (Birnbaum, 2013).

Epidemiological studies have identified a correlation between exposure to endocrine disruptors during fetal life with the development of TGCTs (Storgaard, Bonde and Olsen, 2006). During development, PGCs give rise to the

sperm and oocytes within males and females, respectively (Surani, 2015). Exposure to some endocrine disruptors can result in changes in gene expression in exposed PGCs, thereby suggesting an effect on germ cell development (Del-Mazo *et al.*, 2013). Interestingly, it was shown that *MMP2*, a gene implicated in embryonal carcinoma progression, was significantly up-regulated in response to exposure to the endocrine disruptor MEHP (mono-2-ethylhexylphthalate) which is commonly found in indoor dust (Yao, Lin and Richburg, 2012). In addition, when pluripotent embryonic stem cells (ESCs) were exposed to the endocrine disruptor BPA (Bisphenol A) and induced to differentiate, cells failed to repress markers for pluripotency (Chen *et al.*, 2013). This suggests that endocrine disruptors may affect the development and differentiation of embryonic germ cells which may contribute to the risk of TGCTs later in adult life.

1.3 Diagnosis and Staging of TGCTs

As mentioned previously, it is important to identify the types and stages of TGCTs in order to improve patient outcome. According to the EAU guidelines on Testicular Cancer (2016) there are currently two main methods used to identify the presence of TGCTs (Table 1.1). The first diagnostic test is to analyse the levels of circulating biomarkers shown to be elevated in patients with TGCTs. The second is an ultrasound examination of both testis (Albers *et al.*, 2015). In the presence of TGCTs, an orchiectomy is necessary for removal of the tumour. Following this, examination for metastasis is performed, using magnetic resonance imaging (MRI) and computer topography (CT) scans determine whether the disease has spread around the body. Circulating serum markers can also be measured be used to monitor disease progression. These markers (see section 1.3.1) together allow clinical staging of the cancer in order to allow the correct form of treatment (Albers *et al.*, 2015).

Table 1.1 Tests recommended to be performed at diagnosis of TGCTs for staging according to EAU guidelines – 2016 (Albers *et al.*, 2015).

Tests	**Recommendations**
Serum biomarkers	AFP
	β-HCG
	LDH
Bilateral Testis ultrasound	All patients
Chest computed tomography (CT)	All patients
Abdomen and Pelvis CT	All patients
Bone scan Magnetic resonance imaging (MRI)	In patients presenting with metastatic symptoms.
Brain Scan (MRI)	In patients presenting with high metastatic disease and/or a large number of lung metastases and/or high β-HCG.

1.3.1 Biomarkers

A variety of biomarkers have been identified that are increased in TGCT patients, these include beta human chorionic gonadotrophin (β-HCG), alpha-fetoprotein (AFP), placental alkaline phosphatase (PLAP) and lactate dehydrogenase (LDH). Approximately, 80% of TGCT patients with non-seminomas tumours have increased expression of β-HCG and/or AFP at the time of diagnosis whereas in TGCT seminoma patients β-HCG is elevated in ~15% at the time of diagnosis (Trigo *et al.*, 2000; Germa-Lluch *et al.*, 2002). Measurement of these biomarkers are performed both at the time of diagnosis and after orchiectomy in order to stage TGCTs. They also determine the efficacy of treatment and to detect the potential relapses (Kollmannsberger *et al.*, 2014).

1.3.1.1 AFP

AFP is a plasma protein which is highly expressed during fetal development, whereas after birth AFP levels decrease (Terentiev and Moldogazieva, 2013). AFP has been shown to stimulate proliferation and differentiation in cells, including hepatocytes, fibroblasts, ovarian cells and uterine cells (Leffert and Sell, 1974; Toder *et al.*, 1983; Hamel *et al.*, 1987). AFP is elevated in approximately 50-70% cases of non-seminoma tumours whereas there is no AFP present in pure seminomas, thus distinguishing the two subtypes (Sturgeon *et al.*, 2008; Albers *et al.*, 2015).

1.3.1.2 Lactate dehydrogenase (LDH)

LDH is an enzyme found in the majority of living cells and has a role in metabolism converting lactate to pyruvic acid (DeBerardinis *et al.*, 2008). LDH is highly expressed in proliferating cells, under the stimulus of oncogenes such as C-MYC and HER2, in order to metabolise high levels of pyruvate (DeBerardinis *et al.*, 2008). In TGCT patients, although LDH is elevated approximately 40-60% of cases, it is less specific as compared to the biomarkers AFP or β-HCG (Gilligan *et al.*, 2010). This may be a result of increased LDH also being present in patients with tissue damage that is associated with non-malignant diseases such as pancreatitis, liver failure, muscular dystrophies and heart failure (Masterson, Rice and Beck, 2014).

1.3.1.4 PLAP

PLAP is an alkaline phosphatase that is mainly expressed in the placental tissue, however it has been shown to be elevated in 60-70% of seminoma patients (Masterson, Rice and Beck, 2014). Smokers are shown also to have an increased level of PLAP suggesting that this biomarker may result in false positives and low sensitivity for TGCT diagnosis. Therefore, PLAP is not routinely used (Masterson, Rice and Beck, 2014).

1.3.1.5 β-HCG

β-HCG is a hormone that is produced during embryonic development in the post-implantation period by the placenta (Perkins *et al.*, 2003). Increased levels of β-HCG are commonly observed during pregnancy, trophoblastic disease and in patients with TGCTs (Fowler *et al.*, 1982). β-HCG and AFP are used in

combination to diagnose patients with non-seminatonous TGCTs as β-HCG or AFP are elevated in 85% of non-seminoma patients (Perkins *et al.*, 2003). Therefore, these biomarkers can be used in place of biopsies to diagnoses extra-gonadal or metastatic seminoma, however these circulating biomarkers are less effective in stage I patients as only 20% of patients have elevated levels of these markers (Perkins *et al.*, 2003).

1.3.2 Staging

There are a number of factors that are used to stage the progression of TGCTs (Table 1.2). The staging criteria described here has been obtained from the EAU guidelines on testicular cancer, which follows the TNM (tumour/nodes/metastasis) classification (Albers *et al.*, 2015). The staging procedure classically includes a physical examination, serum biomarkers and CT scans of the thorax, pelvis and abdomen (Peckham *et al.*, 1979; Albers *et al.*, 2015). The first parameter describes the primary tumour (pT0-4) and how much local invasion is present, with a score of pT0 to pT4 (Albers *et al.*, 2015). The next parameter describes at the frequency and size of lymph node metastases (N0-3) ranging from no lymph node metastasis (pN0) to advanced lymph node metastasis (pN3) containing multiple lymph node metastases with a size greater than 5cm. In addition, distant metastasis is measured (M0-1b) ranging from none (M0) to distant metastasis to other sites (M1b). Serum biomarkers are also measured for LDH, β-HCG and AFP and are graded depending on how high the levels are in comparison to normal physiological concentrations (S0-S3).

Table 1.2 Guidelines to classify the TGCT using tumour size and distance of metastases, based on guidelines from EAU guidelines on testicular cancer (Albers *et al.*, 2015).

pT	Primary Tumour			
	PTX	Primary cannot be assessed		
	pT0	Evidence for primary tumour cannot be detected		
	pTis	Carcinoma *in situ* (CIS) detected (e.g. histological scar in testis)		
	pT1	Primary tumour is limited to the testis and epididymis without vascular/ lymphatic invasion		
	pT2	Primary tumour is limited to the testis and epididymis with vascular/ lymphatic invasion		
	pT3	Tumour invades the spermatocytic cord - with or without vascular/ lymphatic invasion		
	pT4	Tumour invades the scrotum - with or without vascular/ lymphatic invasion		
N	**Regional lymph node Clinical**			
	NX	Regional lymph nodes cannot be assessed		
	N0	No evidence of regional lymph node metastasis		
	N1	Metastasis with a lymph node with a mass 2cm or less in greatest dimension or multiple lymph nodes but none greater than 2cm		
	N2	Metastasis with a lymph node with a mass greater than 2cm but less than 5cm in greatest dimension or multiple lymph nodes greater than 2cm but less than 5cm		
	N3	Metastasis with a lymph node mass greater than 5cm in its greatest dimension		
M	**Distant Metastasis**			
	MX	Distant metastasis cannot be assessed		
	M0	No evidence of distant metastasis		
	M1	Evidence for distant metastasis		
		M1a Metastasis to regional lymph node(s) or lung		
		M1b Metastasis to other sites		
S	**Serum biomarkers**			
	SX	Serum biomarkers is not available or was not assessed		
	S0	Serum biomarkers present or normal physiological levels		
		LDH1 (U/l)	**β-HCG (mIU/mL)**	**AFP (ng/ml)**
	S1	S1 < 1.5 x N	< 5,000	< 1,000
	S2	S2 1.5-10 x N	5,000-50,000	1,000-10,000
	S3	S3 > 10 x N	> 50,000	>10,000

After the classification based on the guidelines outlined in Table 1.2, the tumour is staged. Stage I TGCTs denotes the earliest stage and presents with normal/ slightly increased levels of serum markers and a tumour limited within the testicle, without metastasis (Table 1.3). This can be separated into Stage IA, which indicates that the primary tumour is limited to the testis and epididymis, and Stage IB which shows more advanced invasion but restricted to the scrotum. Stage II TGCTs also presents a normal to increased level of serum markers but with metastasis to nearby lymph nodes, main the abdomen or pelvis. Stage II A, B and C denote the extent of the lymph node metastasis, with stage Stag IIC indicating the most advanced lymph node metastasis. Stage III TGCTs generally have elevated levels of serum markers and show advanced metastasis to other organs and distant lymph nodes (Table 1.3). The stage and subtype of the TGCT determines which treatment is to be performed.

Table 1.3 Staging of TGCTs, adapted from (Albers *et al.*, 2015).

TGCT Stage	Tumour Class	Lymph Node Metastasis	Distant Metastasis	Serum Biomarker
Stage 0	pTis	N0	M0	S0, SX
Stage I	pT1–T4	N0	M0	SX
Stage IA	pT1	N0	M0	S0
Stage IB	pT2–T4	N0	M0	S0
Stage IS	Any patient/TX	N0	M0	S1–3
Stage II	Any patient/TX	N1–N3	M0	SX
Stage IIA	Any patient/TX	N1	M0	S0, S1
Stage IIB	Any patient/TX	N2	M0	S0, S1
Stage IIC	Any patient/TX	N3	M0	S0, S1
Stage III	Any patient/TX	Any N	M1a	SX
Stage IIIA	Any patient/TX	Any N	M1a	S0, S1
Stage IIIB	Any patient/TX	N1–N3	M0	S2
	Any patient/TX	Any N	M1a	S2
Stage IIIC	Any patient/TX	N1–N3	M0	S3
	Any patient/TX	Any N	M1a	S3
	Any patient/TX	Any N	M1b	Any S

1.4 Treatment of TGCTs

As mentioned previously, TGCT patients in Europe have cure rates of up to 97% for all stages (Hoffmann *et al.*, 2014; Albers *et al.*, 2015). However, the cure rate depends upon early excellent diagnosis in order to inform correct treatment to reduce toxicities and long term disease, such as secondary cancers (Albers *et al.*, 2015). In the presence of a TGCT, orchiectomy (the removal of one or both testicles) is required. This is then followed up with CT imaging in combination with serum biomarkers in order observe the progression of the disease. A large population study calculated that ~75-80% of seminoma patients and ~55% of patients with non-seminomas were diagnosed as stage I TGCTs at the time of diagnosis (Klepp *et al.*, 1990; Heidenreich and Moul, 2002). This indicates that non-seminomas tend to be more advanced by the time of diagnosis, suggesting non-seminomas are more aggressive that seminoma TGCTs.

Approximately 60% of all TGCTs are seminomas, which present at a slightly later mean age in men in the late thirties as compared to other TGCTs. The majority of seminoma patients are commonly diagnosed at stage I and II, whereas advanced disease is much rarer (McGlynn and Cook, 2010). Supporting this, a review of 103 Swedish healthcare registries for years 2009-2013 identified that 92% of seminoma patients had a 'good prognosis' whereas the remaining 8% has an 'intermediate prognosis' (Emilsson *et al.*, 2015), as described in Table 1.4. In fact, this study indicated that there were no cases of poor prognosis seminoma patients and patients had a long-term survival rate of approximately 100% (Emilsson *et al.*, 2015). Without adjuvant treatment, after orchiectomy 13-20% of patients relapse and follow-up treatment resulted in the survival of

almost all patients (Kollmannsberger *et al.*, 2014; Tandstad *et al.*, 2014). Radiotherapy of the retroperitoneal lymph nodes is an option for stage I patients which reduces relapse rates to ~1-3% (Albers *et al.*, 2015).

For patients with stage II metastatic disease, radiotherapy is performed on the ipsilateral and para-aortic nodes at 15 Gy (performed twice) and 18 Gy (performed twice) for stage IIA and stage IIB patients, respectively. This has demonstrated to give a survival rate close to 100% with a 6 year disease free survival for stage IIA and IIB for 95% and 89% of seminoma patients, respectively (Albers *et al.*, 2015). Chemotherapy using 3 cycles of bleomycin, etoposide and cisplatin (BEP) or 4 cycles of etoposide and cisplatin (EP) can achieve similar results, however these cytotoxic chemotherapies are associated with increased toxicity and risk for secondary tumours (Albers *et al.*, 2015). Patients with seminomas at a stage more advanced than stage IIB are treated with three cycles of BEP or four cycles of EP (4 days per cycle) or for patients with a poor prognosis four cycles of BEP (Albers *et al.*, 2015).

Table 1.4 prognosis based staging system, modified from Albers *et al* (2015)

Group	Subtype	5 Year Progression Free Survival	5 Year Survival
Good Prognosis	Non-seminoma (56%)	89%	92%
	Seminoma (90%)	82%	86%
Intermediate Prognosis	Non-seminoma (28%)	75%	80%
	Seminoma (10%)	67%	72%
Poor Prognosis	Non-seminoma (16%)	41%	48%
	Seminoma (0%)	NA	NA

The prognosis is different for non-seminoma TGCTs (Table 1.4). These are heterogeneous and consist of a variety of histological subtypes: seminomas, embryonal carcinomas, yolk-sac tumours and teratomas (Oosterhuis and Looijenga, 2005). Non-seminomas typically occur at a younger age in comparison to seminoma patients at the age of 25-35 years (McGlynn *et al.*, 2003). Approximately 50% of non-seminomas are stage I at the time of diagnosis, whereas 40-45% of non-seminoma patients present with metastasis (Albers *et al.*, 2015), suggesting that non-seminomas are more aggressive then seminomas which show 75-80% of patients to be Stage I at diagnosis. This is reflected in the percentages of non-seminoma patients with an intermediate and poor prognosis which is substantially higher than that of seminoma patients (Table 1.4). In addition to this, a Swedish cohort 19% of non-seminoma patients relapsed in 3 years as compared to 13% of seminoma patients in 2 years (Tandstad *et al.*, 2014). This demonstrated increased malignancy for non-seminoma TGCTs when compared to seminomas.

Approximately 30% of non-seminoma TGCT patients present with sub-clinical metastasis and will relapse unless adjuvant therapy treatment is performed (Klepp *et al.*, 1990, 1991). Therefore, after orchiectomy patients are usually informed about relapses rates and adjuvant therapy offered, including chemotherapy and retroperitoneal node dissection (Albers *et al.*, 2015). In patients with stage IIB advanced disease, patients are recommended for chemotherapy (BEP x 3 cycles; 4 days per cycle) immediately. In the case of stage IIA patients two routes are taken. For patients without elevated AFP/ β-HCG markers or with a pure teratoma, one follow up examination after 6 weeks is conducted; however in the presence on elevated AFP/ β-HCG chemotherapy

should be initiated (BEP x 3; 4 day per cycle) (Albers *et al.*, 2015). In the case of sustained tumour growth, tumour and retroperitoneal node dissection should be the next course of action. In the presence of elevated biomarkers after the follow up, patients should then be subjected to chemotherapy (3xBEP; 4day per cycle). For more advanced non-seminoma disease, the therapy is to increase BEP treatment to 4 cycles. Four cycles of etoposide cisplatin and ifosfamide has also been shown to yield similar outcome, however with increased myelotoxic effects (Nichols *et al.*, 1998).

1.5 Developmental origins of TGCTs

The etiology of TGCTs is not well defined. The precursor cells for all TGCT subtypes are believed to be Carcinoma *in situ* (CIS) cells which can then contribute to the main tumour subtypes: seminomas and non-seminomas (Figure 1.3) (Mostofi, Sesterhenn and Davis Jr, 1987; Stoop *et al.*, 2011). The current hypothesis is that CIS derive from pre-cursor germ cells, called primordial germ cells (PGCs), which fail to differentiate within the developing embryo (Figure 1.3) (Sonne *et al.*, 2009). This section aims to describe distinct TGCT subtypes, and their relationship to their putative cell of origin.

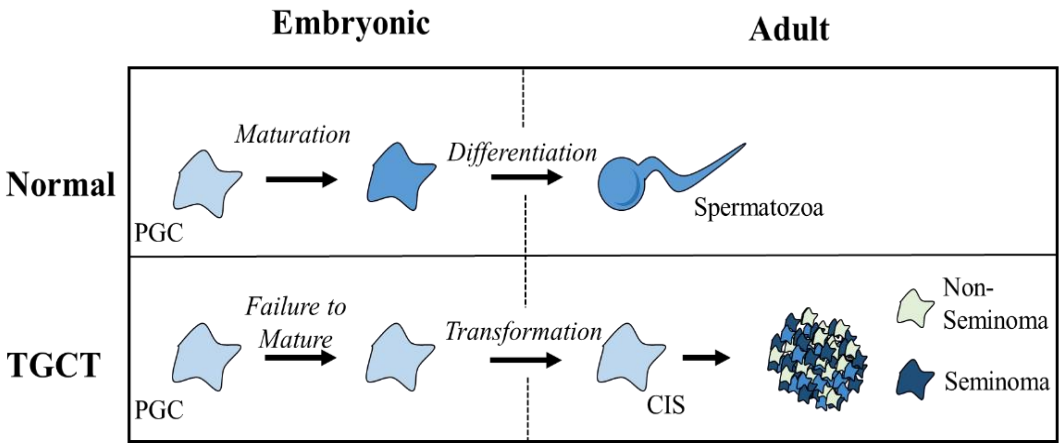


Figure 1.3 A depiction of normal PGC development that successfully differentiate to form spermatozoa, in comparison to PGCs that fail to mature and give rise to TGCTs.

1.5.1 Carcinoma *in situ* (CIS) cells

As mentioned above, it is thought that CIS represents the pre-cursor legion of TGCTs. Studies have identified proteins highly expressed in CIS with the aim to find a marker for early TGCT diagnosis (Meyts *et al.,* 2003). Interestingly, a number of these markers were shown to be highly expressed in CIS and PGCs, whereas expression was undetectable or extremely low in the normal adult testes (Meyts *et al.,* 2003). These markers include PLAP, M2A and TRA-1-60 (Beckstead *et al.,* 1983; Jacobsen *et al* 1984; Giwercman *et al.,* 1991; Müller *et al.,* 1993). This was an indication that TGCT cells are derived from embryonic germ cells and not from the normal differentiated cells within the adult testes. Following this, work conducted by Peter Andrews demonstrated the similarity in global gene expression between CIS and embryonic germ cells (Sonne *et al.,* 2009). This analysis identified that the expression of only five genes differed in these two cell types (*NMNAT2, DEFB119, PTPRZ1, ASXL3* and *AF318333*). Overall this evidence indicated that CIS cells are most likely derived from PGCs that have failed to mature during embryogenesis.

1.5.2 Seminomas

CIS cells can progress into two main types: seminomatous and non-seminomatous TGCTs. Seminomas are histologically uniform and share a high degree of similarity to CIS and PGCs with respect to gene expression (Werrman *et al.*, 2011; Nettersheim *et al.*, 2013; Irie *et al.*, 2015). PGC markers such as SOX17, PRDM1 and LIN28 are genes known to be critical for the specification of PGCs, as knockdown of any one of these genes results in complete loss of PGCs (West *et al.*, 2009; Ohinata *et al.*, 2005). These genes are also highly expressed in seminoma cells (Irie *et al* 2015; West *et al.*, 2009) demonstrating the similarity between PGCs and seminomas. Supporting this, RNA-seq analysis identified that *in vitro* formed PGCs and the seminoma cell line TCam-2 share similar gene expression profiles, therefore TCam-2 cells resemble *in vitro* formed human PGCs (Irie *et al.*, 2015).

Similar to PGCs, seminomas also express the pluripotency markers NANOG and OCT4 (Oosterhui *et al.*, 2005). PRDM1, TFAP2C and SOX17 contribute to their lineage restriction (Weber *et al.*, 2010; Irie *et al.*, 2015) and knockdown of *TFAP2C* or *PRDM1* in the seminoma cell line TCam-2 results in up-regulation of somatic differentiation markers (*HOXA1*, *HOXB1*, *MYOD1*, and *HAND1*) (Weber *et al.*, 2010). Similar to seminomas, the knockdown of *Prdm1* in mouse PGCs results in reduced expression of Tfap2c, while exhibiting an increase in differentiation markers (Weber *et al.*, 2010). This evidence suggests that seminoma and PGCs are similar cell types.

1.5.3 Embryonal Carcinoma Cells

Embryonal carcinoma (EC) cells are considered the pluripotent subtype of TGCTs. ECs represent the pluripotent counterpart of the non-seminoma tumour which are able to differentiate into cells from all three germ layer lineages (teratomas), as well as extra-embryonic differentiation (yolk sack tumours and chorio-carcinomas). Similar to pluripotent embryonic stem cells, EC cells express high levels of the pluripotency markers including NANOG, OCT4, SOX2 (Sperger *et al*., 2003). EC cell lines such as NTERA-2 or NCCIT can differentiate towards the neural lineage with retinoic acid treatment (Andrews *et al.*, 1984; Damjanov *et al.*, 1993). However, not all EC cell lines are capable of differentiation despite expressing pluripotency factors at high levels (Sutiwisesak *et al.*, 2014). Microarray analysis identified that ECs and ESCs clustered most closely when compared to seminomas and other cells within the testes (Sperger *et al*., 2003). This evidence suggested that EC cells are CIS cells which have been reprogrammed back toward a state similar to ESCs losing their germ cell properties. Interestingly, Nettersheim showed that inhibition of BMP signalling using endogenous BMP inhibitor NOGGIN or a chemical BMP inhibitor (LDN193189) reprogrammed the seminoma cells to an EC like state (Nettersheim *et al.*, 2015). This was demonstrated by repression of seminoma (and germ cell) markers *SOX17* and *PRDM1*, whereas EC (and embryonic stem cells) markers including *SOX2* were increased (Nettersheim *et al.*, 2015). This suggests that the microenvironment within the testes may predetermine seminoma and non-seminoma tumours, possibly through differences in BMP signalling.

Altogether, current evidence supports the hypothesis that TGCTs most likely derive from the PGCs. In particular, CIS cells may derive from PGCs which fail to differentiate thus remaining in an embryonic-like state up until adulthood. The CIS cells can give rise to seminoma or EC tumours depending on the signals received in the tumour microenvironment that can either induce the proliferation of immature PGC like cells (seminoma) or reprogramming to an ESC-like state (EC). Therefore, to fully understand how the different subtypes of TGCTs originate it is essential to further our knowledge of germ cell development.

1.6 PGC Development

1.6.1 PGC Specification

In mammals, primordial germ cells (PGCs), the progenitor cells of eggs and sperm, are specified during the peri-gastrulation stage (Tang *et al.*, 2016; Kobayashi *et al.*, 2017). Signals from surrounding embryonic tissues induce the specification of PGCs and following this, PGCs will migrate into the developing gonads where they will mature and give rise to oocyte and spermatozoa in the adult (Ying and Zhao, 2001; Hayashi *et al.*, 2002; de Sousa Lopes *et al.*, 2004). During the migratory stage, PGCs highly express pluripotency factors (such as Nanog and Oct4) while undergoing chromatin remodelling and proliferation (Figure 1.4) (Chambers *et al.*, 2003; Saitou, Kagiwada and Kurimoto, 2012; Gkountela *et al.*, 2015; Kobayashi *et al.*, 2017). Once PGCs reach the gonads, markers of pluripotency are downregulated and male PGCs enter mitotic arrest while females PGCs undergo meiosis (Figure 1.4) (Ewen and Koopman, 2010).

The majority of our knowledge of the mammalian germ line is derived from mouse PGCs studies, however other models including pig and primate embryos have recently been adopted (Kobayashi *et al.*, 2017). In fact, pig (p) and primate (pri) PGCs more closely resemble human PGCs (hPGCs) when compared to mouse, therefore this section will focus mainly on pig PGC development. However, additional mechanisms involved for mouse PGC signalling, when thought similar to humans, will be included also.

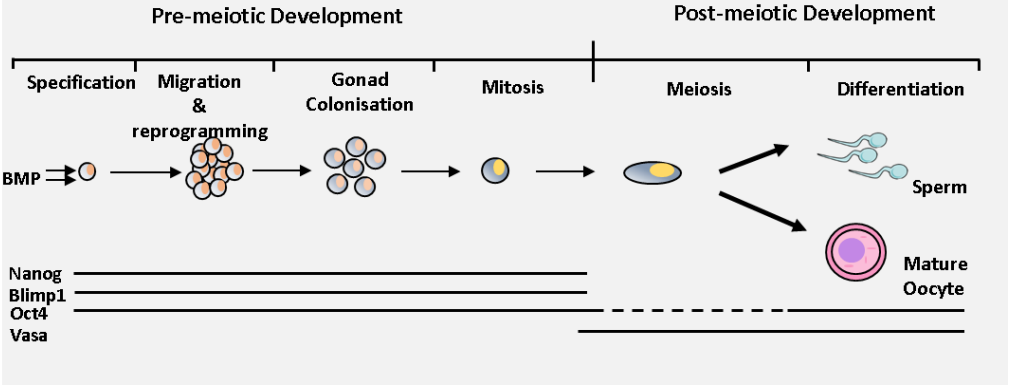


Figure 1.4 Development of PGCs from specification to complete differentiation; in addition to this, expression profiles of genes important for PGC development are shown.

The timing of human PGC specification is less understood compared to mouse and pig (p)PGCs given that the study of early human embryos are impractical due to availability of embryos and ethical considerations. However it is thought that specification of hPGCs occurs around week 2-3 of development (Tang *et al.*, 2016). Similar to humans, pPGCs originate from the posterior pre-primitive streak competent epiblast at approximately embryonic day 12 (Kobayashi *et al.*, 2017). Preceding this, pre-primitive streaks cells exhibit high levels of BMP 2/4 signalling, including their downstream effectors pSMAD1/5/8 (Kobayashi *et al.*, 2017). *Ex vivo* studies have shown that cultured pig epiblast cells can induce pPGCs after 64 hours, whereas *in vivo* inhibition of BMP results in inhibition of PGC specification (Kobayashi *et al.*, 2017). This is important, as it has been shown in mPGC that Bmp and Wnt induce Blimp1 expression a critical determinant for PGCs (Magnúsdóttir *et al.*, 2013). The subsequent activation of Blimp1 inhibits the somatic differentiation programme in PGC founder cells through transcriptional repression at approximately E6.25 (Ohinata *et al.*, 2005;

Saitou *et al.*, 2005). In the midline of the early primitive streak in the pig embryo, where the BMP signalling is highest, a cluster of SOX17 expressing cells at the posterior end are observed (Kobayashi *et al.*, 2017). Following this, the SOX17 positive cells begin to express BLIMP1 followed by concomitant upregulation of NANOG and OCT4 and downregulation of SOX2 (Kobayashi *et al.*, 2017). Mouse PGCs do not express Sox17 where instead Sox2 is highly expressed, thus highlighting a fundamental difference between the two species (Saitou and Yamaji, 2012).

PGCs can be induced *in vitro* from primed hESCs, thus giving further insight into the mechanism of PGC specification. Conventional hESCs when cultured in medium containing the four inhibitors (4i medium) FGF2, TGF-β, LIF and GDK3 β respond to the addition of BMP2 and BMP4 (Figure 1.5) (Irie *et al.*, 2015) and are induced to PGC like cells (PGCLCs) with high efficiency (~50%) (Figure 1.4). These PGCLCs exhibit an expression profile similar to human gonadal PGCs and the seminoma cell line TCam-2 (Irie *et al.*, 2015), which includes BLIMP1, SOX17, TFAP2C, NANOG and OCT4. In addition to this, it is thought that the epigenetic profile of these cells resemble pre-migratory hPGCs given their extensive imprint erasure and demethylation (Tang *et al.*, 2015). Similar to that observed in pPGCs *in vivo*, SOX17 is amongst one of the first transcription factors observed in hPGCLCs (Irie *et al.*, 2015). The knockout of *SOX17* from primed hESCs results in loss of PGCLC specification in addition to compromised BLIMP1 expression, thus suggesting SOX17 as upstream

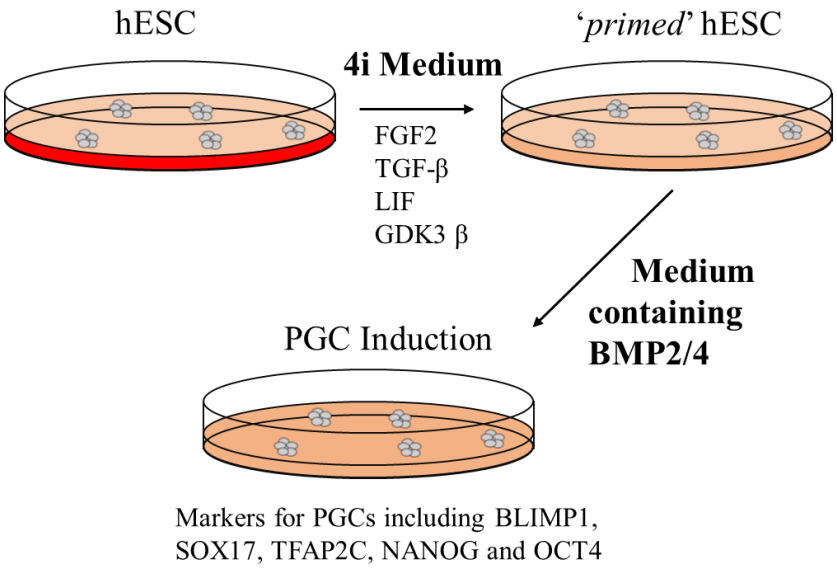


Figure 1.5 An overview for the induction of hPGCLCs *in vitro*. 4i medium is shown to 'prime' cells for PGC competence, exposure of these primed cells to BMP results in the specification of hPGCLCs indicated by an increase in PGC markers (BLIMP1, SOX17, TFAP2C, NANOG and OCT4). Adapted from (Irie *et al.*, 2015).

regulator of BLIMP1 in hPGCs (Figure 1.6) (Irie *et al.*, 2015). The knockout of *Blimp1* in mice results in the repression of PGC specifying genes and the de-repression of developmental genes (Sasaki *et al.*, 2015; Tang *et al.*, 2016; Kobayashi *et al.*, 2017). In mouse PGCLCs, Blimp1 co-operate with Prdm14 and Ap2γ to upregulate genes involved in pluripotency, germ cell specification, germ cell migration and epigenetic remodelling while repressing somatic gene expression (Figure 1.6) (Magnúsdóttir *et al.*, 2013). Using ChIP-seq analysis to map these transcription factors to the genome, Blimp1 binds mainly to the proximal promoters of somatic genes in order to repress expression. Blimp1 will also directly induce the expression of Ap2γ which cooperates with Prdm14 to initiate a PGC specific state (Magnúsdóttir *et al.*, 2013). Prdm14 has been shown to interact with Suz12, a component of the polycomb repressive complex 2,

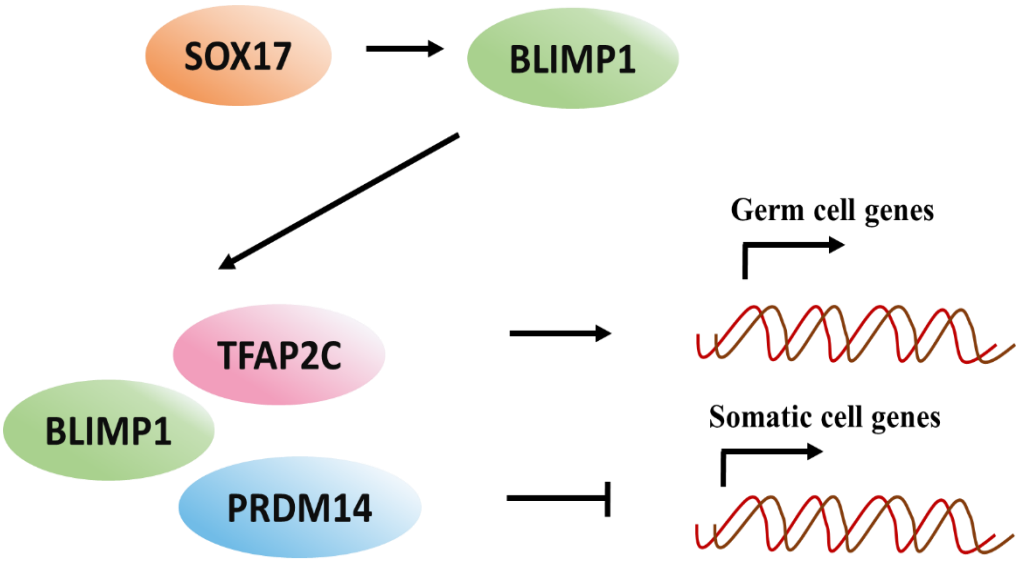


Figure 1.6 A depiction for the specification of hPGCLCs; it is thought that SOX17 will induces BLIMP1 expression. BLIMP1 in combination with TFAP2C and PRDM14 will induce germ cell gene expression will repressing somatic genes. Adapted from (Irie *et al.*, 2015).

which catalyses the deposition of H3K27me3 marks to confer transcriptional repression (Yamaji *et al.*, 2013). In contrast, Prdm1 has been shown to interact with the epigenetic modifiers Hdac2, Kdm1a and Prmt5 to control the expression of target genes (Fog, Galli and Lund, 2012). After the PGCs have been specified they migrate towards the gonads, during this migratory stage they also undergo a process termed 'reprogramming' in order to erase epigenetic marks prior to the cells maturing within the gonads.

1.6.2 PGC Reprogramming

PGCs are required to undergo reprogramming to erase epigenetic memory of differentiation and to prevent inheritance of epimutations down the germline (Hajkova *et al.*, 2002). Epigenetics describes a regulatory system resulting in a change in gene expression without modifying the genetic code (Füllgrabe, Kavanagh and Joseph, 2011). Most epigenetics research related to PGC development has concentrated on histones, histone modifications and the methylation of DNA sequences (Seki *et al.*, 2005, 2007; Gkountela *et al.*, 2015; Tang *et al.*, 2015; Kobayashi *et al.*, 2017).

DNA methylation mediates silencing of transcription by methylating GpC-rich regions in the immediate vicinity of a transcription start site, termed CpG islands (Jones, 2012). In humans, PGC demethylation occurs just after colonisation of the gonads, weeks 7-8 post-conception, completing demethylation at weeks 10-11 weeks and reaching 7.8% and 6.0% of the median methylation levels in male and female PGCs, respectively (Guo *et al.*, 2015). Interestingly, hPGC demethylation occurs simultaneously with increased expression of NANOG (Guo *et al.*, 2015). In addition to this, PGCs remodel their chromatin during reprogramming (Hajkova *et al.*, 2008; Saitou, Kagiwada and Kurimoto, 2012; Gkountela *et al.*, 2013). Genomic DNA is compacted and packaged in chromatin; a structural unit of chromatin called a nucleosome which is ~146bps of DNA wrapped around an octamer of histone proteins (Figure 1.7 A) (Luger *et al.*, 1997). H3 and H4 histone proteins also have protruding strings of amino acids called histone tails. These histone tails, when modified, can influence transcription of nearby genes (Bernstein *et al.*, 2006). Modifications of histone

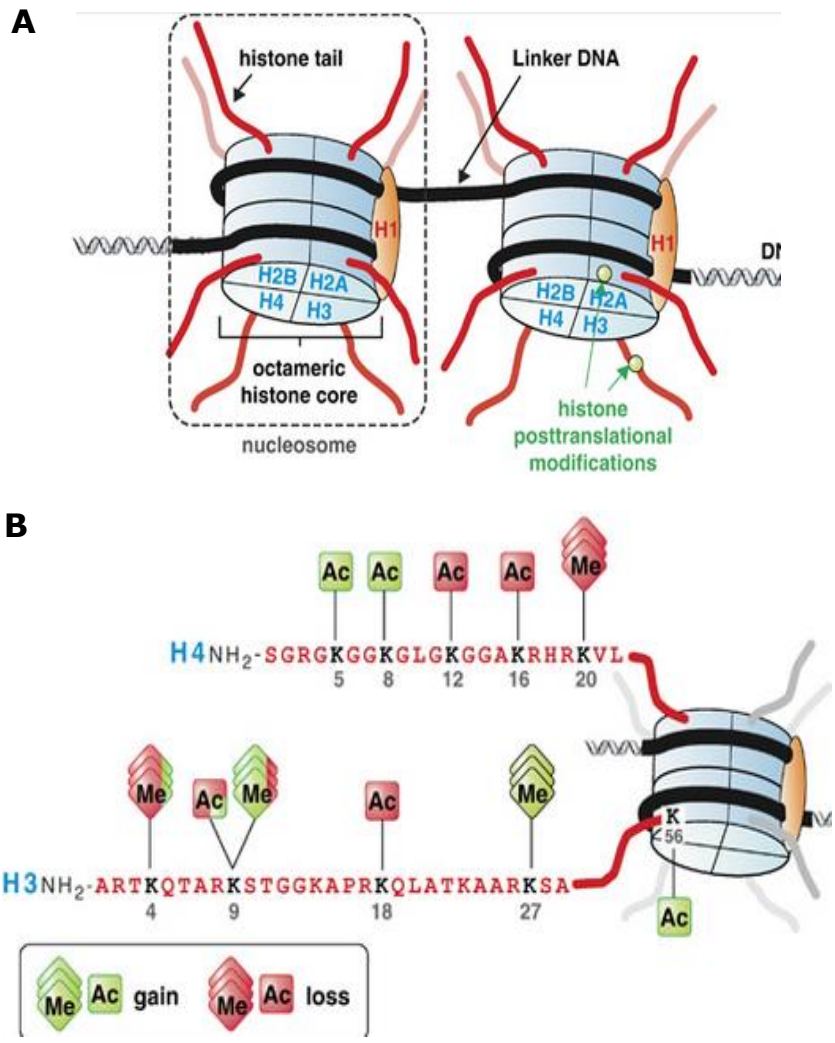


Figure 1.7 (A) shows DNA (black) wrapped around a histone protein complex (blue) attached by strings of amino acids called histone tails (red). **(B)** shows different histone tail modification that can occur, such as methylation (Me, red diamond) or acetylation (Ac, green square) (Füllgrabe et al., 2011).

tails include methylation, ubiquitination, phosphorylation and acetylation; each modification can impact upon transcription depending on their position along their histone tail (Figure 1.7 B) (Cosgrove, Boeke and Wolberger, 2004). The most well-characterised histone modifications are H3K4me3 and H3K27me3 at TSSs which can be used to give indication of a gene activity. For H3K27me3, indicates tri-methylation of lysine 27 on histone H3, whereas H3K4me3 indicates tri-methylation of lysine 4 on histone H3 (Bernstein *et al.*, 2006).

H3K4me3 indicates a gene which is active whereas H3K27me3 indicates an inactive gene (Bernstein *et al.*, 2006). When both histone marks are present at a gene they are termed 'bivalent' and although are generally inactive are poised for activation (Bernstein *et al.*, 2006). During specification of *in vitro* formed mPGC-like cells (mPGCLCs), the promoters of germ cell specifiers such as *Prdm1* become poised for activation (Kurimoto *et al.*, 2015). H3K27me3 levels are transient during migratory stages to be then depleted once PGCs arrive at the genital ridge (Gkountela *et al.*, 2013). H3K27me3 marks are enriched at retrotransposon repeats indicating their repression, whereas H3K4me3 enrichment occurs at genes highly expressed in mPGCs (Ng *et al.*, 2013). Once the cells have reconfigured epigenetic marks and have arrived in the gonads, germ cells undergo a process of maturation.

1.6.3 PGC Maturation

In mice, PGCs arrive at the genital ridge at E10.5 where they find an environment for continued differentiation and specification of germ cells, thus becoming gonocytes (Ewen and Koopman, 2010). Upon arrival, early-PGC markers, such as Ssea1 and Nanos3, are repressed whereas gonocyte-specific genes, such as Dazl, Gcl and Vasa, start to be expressed (Kimura *et al.*, 1999; Saunders *et al.*, 2003; Ohinata *et al.*, 2005; Yamaguchi *et al.*, 2005). Unlike PGCs, which have increased potential for pluripotency, gonocytes harbour a reduced capacity to form pluripotent stem cells in culture (Matsui, Zsebo and Hogan, 1992; Resnick *et al.*, 1992). Nevertheless, gonocytes have the capacity to form teratomas (a tumour made up of different tissues from more or one of the germ layers) and therefore they are still considered multipotent (Kanatsu-Shinohara and Shinohara, 2006). At E14.5, male mouse gonocytes undergo mitotic arrest whereas female mouse PGCs undergo meiosis at E13.5 (Yamaguchi *et al.*, 2005). It is during this maturation stage that pluripotency markers, including Nanog, are down regulated (Yamaguchi *et al.*, 2005). In humans, gonocytes migrate to the basal membrane forming spermatogonia postnatally at approximately 6 months (Zivkovic *et al.*, 2007). It is not until early puberty that spermatogenesis starts (Stukenborg *et al.*, 2013).

1.6.4 Germ Cell Differentiation

Normal spermatogenesis is a complex yet orchestrated process that involves three main aspects; (1) spermatogonia mitotic proliferation and maturation, (2) haploid formation as a result of meiotic cleavage and (3) the final step of maturation of cellular elongation into spermatids (spermiogenesis) (Stukenborg *et al.*, 2013). In humans, spermatogonia resume proliferation just after birth and after sexual maturity cells undergo meiotic division (O'Connor *et al*., 2008). The production of sperm is maintained through the adult lifetime.

Spermatogonia can be categorised into type-A (undifferentiated) and type-B (differentiated). Type-A spermatognia can be further distinguished into cells with either dark or pale nuclei (Waheeb *et al.,* 2013). Type-A cells with dark nuclei do not usually divide mitotically, whereas cells with pale nuclei divide and differentiate to form Type-B cells (Stukenborg *et al.,* 2013). Following this, type-B cells spermatogonia divide to form primary spermatocytes which undergo meiosis post-puberty (Stukenborg *et al.,* 2013). The onset of spermatogenesis is stimulated by pulsing of gonadotropin-releasing hormone (GnRH) secreted by the hypothalamus which in turn induces the release of follicle stimulating hormone (FSH) and luteinising hormone (LH) (Delemarre *et al.,* 2008). FSH stimulates Sertoli cells which, in turn, release androgen binding protein facilitating an increased retention of androgens in the testes, thus initiating spermatogenesis (Delemarre *et al.,* 2008). In addition, LH stimulates Leydig cells to produce testosterone resulting in male secondary sex characteristics (Delemarre *et al.,* 2008).

As previously mentioned, TGCTs present after puberty between the ages 15 and 45 years old (Ferlay *et al.,* 2010). This suggests that the onset of puberty and thus sex hormones regulate TGCT growth. The sex hormone signalling pathways that stimulate cells within the adult testes to grow may be stimulating PGC-like or CIS cells to proliferate. Supporting this, the seminoma cell line TCam-2 expresses high levels of the androgen receptor (AR) protein and the addition of testosterone in culture can stimulate proliferation (Wallacides *et al.,* 2012). In addition, RNA-seq analysis showed that *in vitro* formed PGCLCs also express high levels of AR, which in fact exceeds that of TCam-2 (Irie *et al.,* 2015). Given the similarity between CIS and PGCs (see section 1.5.1), this suggests CIS may express AR highly also. From this, it could be hypothesis that CIS growth stimulated after puberty has been initiated due to increasing testosterone concentrations within the testes. However, contrasting this, patients with collective testicular dysgenesis syndrome (TDS) (see section 1.2.2), are associated with low levels of testosterone (Andersson *et al.,* 2007). In addition, patients with androgen insensitivity syndrome (AIS) have increased risk of developing TGCTs (Rajpert-De Mayts *et al.,* 2006). Patient with complete androgen insensitivity are largely attributed (90-95%) to mutations in the androgen receptor gene (Audi *et al*., 2010; Mongan *et al*., 2015). This, however, suggests that lack of androgen signalling may in fact promote TGCT growth and this further complicates androgens role in TGCT pathogenesis.

This evidence suggests that mal-development of the testes within the developing embryo resulting in the prevention of PGC maturation and a secondary effect of the hormonal or Leydig/Sertoli cell malfunction during male sexual differentiation (Skakkebæk *et al* 2001; Sharpe *et al* 2003).

1.7 The role of NANOG in PGCs

NANOG is a homeobox transcription factor shown to be a key mediator of self-renewal and pluripotency. In mouse embryos, Nanog is expressed very early on within the interior cells of the morulae (Chambers *et al.*, 2003). As the embryo develops and the germ layers start to form, Nanog is downregulated in all tissues apart from the primordial germ cells (Chambers *et al.*, 2003). *Nanog* null mouse embryos result in embryonic lethality and do not progress past the blastocyst stage (Mitsui *et al.*, 2003). In addition to this, conditional knockdown of *Nanog* in PGCs results in a reduction in PGCs that reach the gonads. This reduction is the result of increased apoptosis and decreased proliferation (Yamaguchi *et al.*, 2009). Nanog is also expressed in pluripotent cell lines such as ESCs, EC (embryonal carcinoma), EG (embryonic germ) and SEM (seminoma) cell lines (Chambers *et al.*, 2003; Irie *et al.*, 2015).

The function of Nanog has been most extensively studied in ESCs (Loh *et al.*, 2006; Zhang *et al.*, 2009; Theunissen *et al.*, 2011; Festuccia *et al.*, 2012). NANOG knockdown in hESCs resulted in a reduction of cellular proliferation (Zhang *et al.*, 2009). NANOG directly binds to genomic regulatory regions of the cell cycle regulators *CDK6* and *CDC25A* in order to induce transcription (Zhang *et al.*, 2009). Interestingly, high expression of NANOG and induction of *CDK6* and *CDC25A* result in a cell cycle profile with shortened transition from G1 to S-phase and leads to the accumulation of cells in S-phase, similar to the cell cycle profile of ESCs (Figure 1.8A) (Zhang *et al.*, 2009). Nanog has also been shown to be vital for the maintenance of pluripotency in mESCs, as cells with lower levels of Nanog are more susceptible to differentiation (Abranches

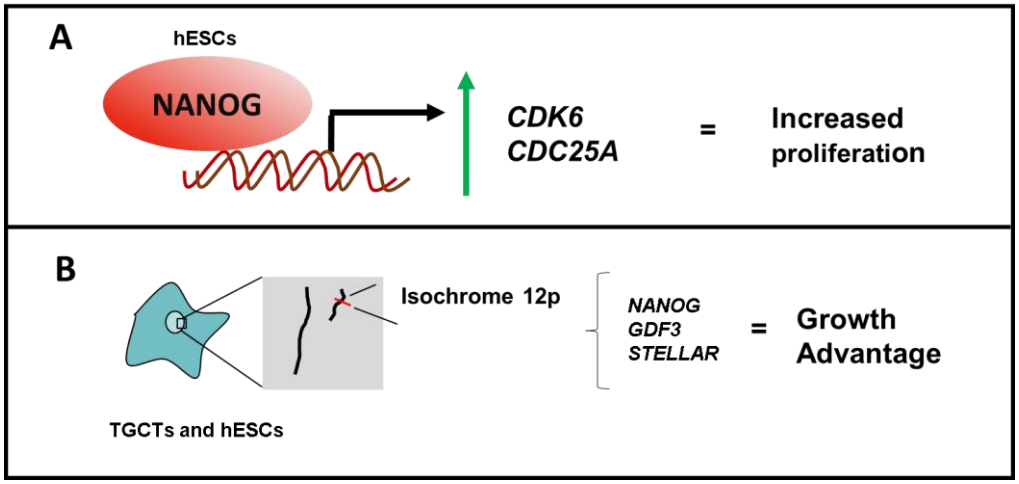


Figure 1.8 (A) NANOG has been shown to positively regulate the transcription of *CDK6* and *CDC25A* in hESCs in order to promoter proliferation (Zhang et al. 2009). **(B)** TGCTs and hESC have been shown to present with amplifications for chromosome 12p which contains genes involved in regulating proliferation and pluripotency in these cells types (Ben-David et al. 2014; Clark et al. 2004).

et al., 2014). ESCs can acquire trisomy of chromosome 12p after extensive culture (Ben-David *et al.*, 2014). Comparisons of hESCs with trisomy in comparison with normal hESCs indicated that trisomy 12 increases the proliferative rate of cells (Figure 1.8 B) (Ben-David *et al.*, 2014). This indicated that gain of chromosome 12p provides a selective advantage over diploid cells and genes within this chromosomal segment are important for promoting proliferation. Chromosome 12p contains genes including *STELLAR*, *GDF3* and *NANOG* which all positively regulate pluripotency (Figure 1.8 B) (Clark *et al.*, 2004; Han *et al.*, 2016). Interestingly, microarray analysis identified that hESCs with trisomy 12 cluster more closely to embryonal carcinoma (EC) cell lines in comparison to diploid hESCs (Ben-David *et al.*, 2014). All the EC cell lines used in this study presented with gain of chromosome 12p suggesting that these EC cell lines are essentially a transformed counterpart of hESCs. In addition to this, the over expression of NANOG in hESCS also resulted in gene expression more similar to EC cells as compared to diploid ESCs (Ben-David *et al.*, 2014).

Therefore, this suggests a central role for NANOG overexpression in cellular transformation.

Pluripotency factors can form multimeric protein complexes with each other; Nanog can form a homodimer in addition to a heteromeric complex with Oct4 and Oct4 with Sox2 as shown in mESCs (Rodda *et al.*, 2005). Nanog, Oct4 and Sox2 are transcription factors that can regulate gene transcription by binding to gene regulatory elements (Loh *et al.*, 2006). They have been shown to bind to a considerable portion of target genes and in fact co-occupy regulatory regions (Boyer *et al.*, 2005; Loh *et al.*, 2006). The introduction of these pluripotency factors (*NANOG*, *OCT4* and *SOX2*), in addition to the introduction of *LIN28* into somatic cells, can induce reprogramming and induction to pluripotency (Yu *et al.*, 2007). Somatic cells which have been reprogrammed back to pluripotency are termed induced pluripotent stem cells (iPSCs) and show similar gene expression profiles to hESCs (Yu *et al.*, 2007). This is of particular interest as Nanog, Oct4, Sox2 and Lin28a are expressed during the migratory stages of mPGCs during epigenetic reprogramming (Schöler *et al.*, 1990; Yamaguchi *et al.*, 2005; West *et al.*, 2009). Interestingly, the knockdown of *Lin28* transcripts using RNA inhibition in cells to be induced to mPGC-like cells *in vitro* resulted loss of PGC specification (West *et al.*, 2009). This work demonstrated an essential role for Lin28 in germ cell specification. Lin28 is an RNA-binding protein that is key for the post-transcriptional regulation of microRNAs (miRNAs). In particular, Lin28 has been shown to repress the processing of miRNA let-7 resulting in its degradation (Heo *et al.*, 2008). Let-7 targets *Blimp1* in PGC cells and in Lin28 deficient mPGCs, the introduction of *Blimp1* rescues PGC formation (West *et al.*, 2009). Interestingly, *let-7* targets include *CDK6* and

CDC25A, the cell cycle promoters positively regulated by NANOG transcription factor binding (Johnson *et al.*, 2007). Completing this regulatory loop, NANOG has also been shown to repress the transcription of *let-7* in ESCs (Marson *et al.*, 2008). In contrast, NANOG can induce the expression of miRNAs from the *miR-302* cluster (Card *et al.*, 2008; Barroso-del Jesus, Lucena-Aguilar and Menendez, 2009). In addition, *miR-302* can be used to reprogram somatic fibroblast into iPSCs in conjunction with *miR-372*. (Subramanyam *et al.*, 2011). In summary, this evidence indicates a role for NANOG in the regulation of pluripotency and proliferation through the transcriptional regulation (induction and repression) of genes, including microRNAs.

1.8 MicroRNA Regulation

1.8.1 MicroRNAs Overview

MicroRNAs (miRNAs) are small non-coding RNAs (~20-21 base pairs) which target mRNAs resulting in the inhibition of transcript translation (Ambros, 2004). MiRNAs are transcribed by RNA-polymerase III which synthesises primary miRNA (pri-miRNA) (Borchert, Lanier and Davidson, 2006). Pri-miRNAs are then processed by the protein complex RNase III enzyme Drosha and a RNA binding protein Dgcr8, to give a cleaved hairpin precursor of 70 nucleotides (70nts) called pre-miRNA (Denli *et al.*, 2004; Landthaler, Yalcin and Tuschl, 2004). The pre-miRNAs are then translocated from the nucleus to the cytoplasm where they are cleaved by the RNase enzyme Dicer at the loop that joins the 5' and 3' arms of the miRNA producing two miRNA mature strands of approximately 20-22 nts (Hammond *et al.*, 2000), as shown in Figure 1.9. One arm, termed the guide arm, is usually more biologically active than the

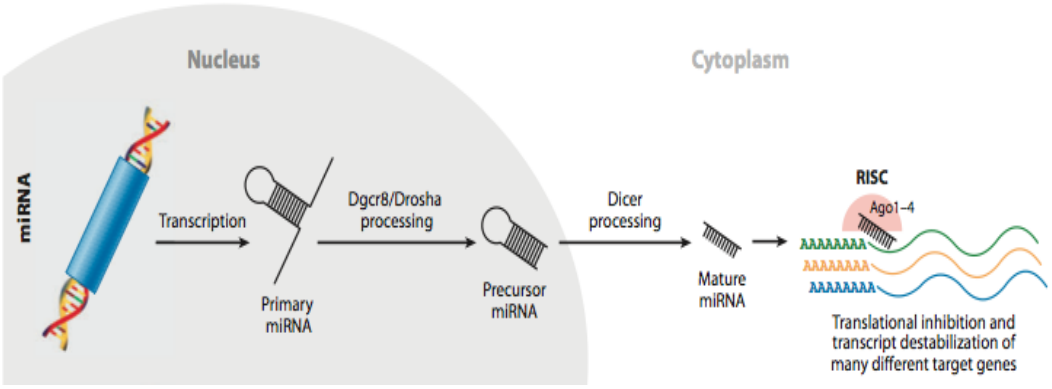


Figure 1.9 The process of miRNA production, from transcription (left) of the primary miRNA and its processing to produce the mature miR (Right). Adapted from (Greve et al., 2013).

other, termed the passenger arm (Okamura *et al.*, 2008). The guide arm is then bound by an Argonaute family RNA binding protein forming the RNA-inducing silencing complex (RISC) (Khvorova, Reynolds and Jayasena, 2003; Schwarz *et al.*, 2003), while the passenger is typically degraded. The RISC complex will predominantly bind to the 3' untranslated region (3' UTR) of the mRNA and functions to inhibit translation (Guo *et al.*, 2010; Bazzini, Lee and Giraldez, 2012; Djuranovic, Nahvi and Green, 2012). MiRNAs are frequently grouped into families or clusters with chromosomal loci; a family of miRNAs relates to their seed sequence, 6-8 base pairs from the 5' end of the miR, whereas the cluster relates to the groups of miRNAs that are expressed from a similar genomic locus (Bartel, 2009). MiRNA families commonly target similar genes, whereas miRNA clusters usually have similar expression patterns (Friedman *et al.*, 2009). Homozygotic deletion of *Dicer* results in complete loss of miRNA production resulting in gross abnormality of mouse embryos at E7.5 and complete resorption of the embryos by E9.5 (Bernstein *et al.*, 2003). Dicer processes miRNAs but is also essential for the processing of other non-coding RNAs such as siRNAs (small interfering RNAs), whereas Dgcr8 is believed to specifically processes miRNAs (Wang *et al.*, 2007; Suh *et al.*, 2010).

1.8.2 MicroRNA regulation in PGCs

In order to investigate the role of miRNAs in mPGCs, *Dicer* was conditionally knocked down specifically in the germline (Hayashi *et al.*, 2008). *Dicer* knockdown PGCs exhibited reduced proliferation resulting in a decreased number of PGCs within the gonads staged E12.5-E13.5. It is worth noting that, in mouse PGCs, Nanog is expressed from E7.5-E16.5 reaching its peak at E13.5 (Yamaguchi *et al.*, 2009). As mentioned previously, Dicer processes other small non-coding RNAs other than miRNAs, such as siRNAs, therefore this experiment alone does not demonstrate that miRNAs are responsible for the reduction in proliferation. However, in mPGCs conditional knockdown of *Ago2,* an enzyme known to specifically process siRNA and not miRNA, showed no difference in PGC proliferation or number (Hayashi *et al.*, 2008). *Dicer* knockout germ cells also express the PGC marker Vasa in late gonadal germ cells, thus suggesting a defect in differentiation related to loss of miRNA function (Hayashi *et al.*, 2008). This evidence suggests that miRNAs are dispensable for early PGC development but are required for later stages when PGCs undergo differentiation and proliferation. *MiR-9, let-7* and *miR-125a* are up regulated in gonadal germ cells suggesting that miRNAs may facilitate the maturation and mitotic arrest observed in germ cells at this stage (Hayashi *et al.*, 2008; Takada *et al.*, 2009). In summary, this evidence suggests that miRNAs play an essential role in the maturation, differentiation and proliferation of embryonic germ cells during development.

1.8.3 MicroRNA regulation in ESCs

MiRNAs are vital for regulating the cell cycle and differentiation in ESCs. In fact, miRNA expression profiles have been instrumental for the distinction between ESCs and differentiated cell types (Wang *et al.*, 2008). Supporting this, over 100 miRNAs show differential expression when comparing hESCs and differentiated embryoid bodies, however the validated targets for the majority of these miRNAs remains to be identified (Morin *et al.*, 2008). The requirement of miRNAs in the maintenance of pluripotency was initially demonstrated in mESC using genetic studies. Genetic deletion of *Dgcr8* or *Dicer* results in defective differentiation and reduced proliferative capacity (Figure 1.10) (Kanellopoulou *et al.*, 2005; Wang *et al.*, 2007). Mutant *Dgcr8* mESCs fail to downregulate pluripotency markers such as Nanog, Oct4, Sox2 and Rex1 during differentiation (Wang *et al.*, 2007). This has been shown to be

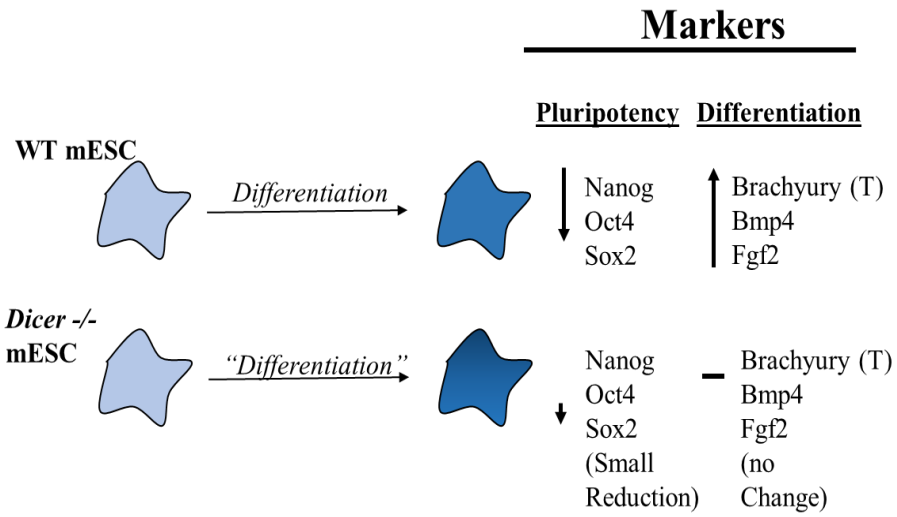


Figure 1.10 The differentiation of mESCs results in a downregulation of pluripotency markers (*Nanog*, *Oct4* and *Sox2*) and up-regulation of differentiation markers (*T*, *Bmp4* and *Fgf2*). The differentiation of Dicer null (KO) mESCs does not result in silencing of pluripotency markers and differentiation markers are not up-regulated. (Kanellopoulou *et al.*, 2005; Wang *et al.*, 2007).

in part due to loss of *miR-145* which binds to the 3' UTR of *OCT4, SOX2* and *KLF4* to repress protein synthesis. *MiR-145* shows a significant upregulation during differentiation and ectopic expression in hESCs disrupts stem cell maintenance and self-renewal (Xu *et al.,* 2009). On the other hand, *Dicer* deficient mESCs lose the ability to methylate the promoter of *Oct4* in order to permanently silence its expression during differentiation (Benetti *et al.,* 2008; Sinkkonen *et al.*, 2008). Transfection of the *miR-290* cluster resulted in the rescue of the *de novo* methylation state of *Dicer* null mESCs (Sinkkonen *et al.*, 2008). However, *miR-290* is the most abundant miRNA in mESC which is repressed during differentiation, suggesting it does not play a key role in the differentiation process. Alternatively, miRNAs from the *miR-290-295* cluster regulate the G1-S transition in order to promote proliferation and repress apoptosis in mESCs (Marson *et al.*, 2008; Zheng *et al.*, 2011).

Interestingly, the pluripotency factors Nanog, Oct4 and Sox2 regulate a subset of miRNA (including *miR-290-295*) in ESCs to modulate the cell cycle (Barroso-delJesus *et al.,* 2008; Card *et al.,* 2008; Marson *et al.,* 2008). This group of miRNAs have been classed as ESC specific cell cycle (ESCC) regulators and they include the miRNA clusters *miR-302* and *miR-371-373* (a human orthologue of *miR-290-295* from the mouse) (Wang and Blelloch, 2011). A similar expression of ESCC miRNAs have also been observed in TGCTs.

1.9 MicroRNAs and TGCTs

Few studies have investigated the expression profiles for miRNAs in TGCTs (Palmer *et al.*, 2010; Dieckmann *et al.*, 2012; Novotny *et al.*, 2012). They have largely utilised high through-put technologies to identify miRNAs which were differentially expressed in tumour samples. One of the first studies investigating miRNAs in tumour cells was performed by Voorhoeve *et al*, 2006, who created a library of plasmid constructs overexpressing a given miRNA. These were then transfected into primary fibroblast cells which were normal or overexpressing RAS^{V12}. Mitogenic signals from RAS^{V12} induced growth arrest in normal fibroblasts and overexpression of the *miR-371/373* overcame growth arrest, thus indicating these miRNAs have a selective growth advantage (Voorhoeve *et al.*, 2006). In support of this, it was shown that mutating the mature sequence of either *miR-372* or *miR-373* abolished this growth advantage. The increase in proliferation observed in cells overexpressing *miR-371/373* was dependant on p53 expression, as *p53* knockdown in *miR-371/373* overexpressing cells exhibited a reduction in proliferative capacity (Voorhoeve *et al.*, 2006). From this, it was hypothesised that tumours which contained wild type *TP53* (the gene encoding p53) would may also overexpress *miR-371/373* in order to promote tumorigenesis. The same study showed that TGCTs have wild type *TP53* and high expression of the *miR-371/373* cluster. In addition to this, *miR-372* and *miR-373* directly bind to the 3' UTR of the transcript encoding the *LATS2* gene in order to repress translation (Voorhoeve *et al.*, 2006). *LATS2* knockdown resulted in the upregulation of CDK2/Cyclin E in order to promote cell growth (Li *et al.*, 2003).

Subsequently, the same research group performed a RT-QPCR array for 156 miRNAs on a variety of TGCT subtypes including seminomas, embryonal carcinomas, teratomas, yolk sac tumours, spermatocytic seminomas and TGCT cell lines (Gillis *et al.*, 2007). Interestingly, a relatively small number of miRNAs analysed by unsupervised hierarchical clustering was able to cluster TGCT samples based on histological subtypes. This study also confirmed that the *miR-371/373* cluster was overexpressed in the seminoma and non-seminoma TGCT subtypes (Gillis *et al.*, 2007). Following this, microarray technology was developed that was able to quantify global miRNA expression in TGCTs (Palmer *et al.*, 2010; Novotny *et al.*, 2012). Novotny *et al*, 2012, performed a microarray analysis on biopsies containing different amounts of Carcinoma *in situ* (CIS) in clinical samples in attempt to elucidate a miRNA expression profile for CIS cells. The microarray identified 12 microRNAs that positively correlated with the amount of CIS in biopsies, these included miRNAs from the *miR-371/373*, *miR-96* and *miR-302* clusters (R>0.93) (Novotny *et al.*, 2012).

In contrast to this, Palmer *et al*, 2010, investigated miRNA expression in malignant and non-malignant TGCTs in both paediatric and adult samples (Palmer *et al.*, 2010). When comparing malignant and non-malignant samples, the top ten differentially expressed miRNAs belonged to two miRNA clusters – *miR-371/373* and *miR-302*. In addition, when unsupervised hierarchical clustering was performed using the expression values for these miRNAs samples could be separated into malignant and non-malignant groups. As *miR-371/373* and *miR-302* clusters are highly expressed in malignant tissues, irrespective of TGCT subtypes, these miRNAs are likely to be involved in promoting disease progression for TGCTs (Palmer *et al.*, 2010). Interestingly, miRNAs from this

cluster could be detected in the serum for TGCT patients. A cohort of 30 age matched patients with TGCTs (n=15 seminoma and n=15 non-seminoma) and 18 healthy patients were analysed for circulating miRNA expression. Mature miRNAs from the *miR-371-373* cluster were highly expressed in TGCTs most consistently (when compared to *miR-302* cluster miRNAs) when compared to the healthy patient controls (Syring *et al.,* 2015). More interestingly, m*iR-367-3p, miR-371a-3p, miR-372-3p* and *miR-373-3p* were significantly higher in malignant TGCT patients when compared to non-malignant disease. MiRNA levels for *miR-367-3p, miR-371a-3p* and *miR-373-3p* showed increased expression in stage pT3 patients when compared to pT1 or pT2, suggesting these miRNAs can be used for determining the progression of TGCT disease. After patients had undergone orchiectomy or carboplatin therapy these circulating miRNAs decreased in expression significantly, thus suggesting these miRNAs could also be used for monitoring relapse (Syring *et al.,* 2015). This study has highlighted a method for diagnosing and classifying disease progression in TGCTs by measuring levels of circulating miRNAs. It has been shown that *miR-371/373* enhances cell cycle progression through targeting *LATS2* (Voorhoeve *et al.,* 2006) while also positively correlating with TGCT malignancy (Syring *et al.,* 2015). The majority of studies have investigated miRNAs that are highly expressed in TGCTs, such as the *miR-371/373* cluster. However, not much work has been conducted on miRNAs that are silenced to enhance TGCT disease progression.

For this reason, the identification and characterisation of miRNAs regulated by NANOG in TGCTs has been the central focus of this dissertation.

Chapter 2: Methods

2.1 Bioinformatics Analyses

2.2.1 ChIP-seq Analysis

ChIP-seq BED files were downloaded from NCBI GEO (GSM1505622; (Tsankov *et al.*, 2015)) for the transcription factors NANOG, OCT4 and SOX2 (HUES64); and from the ENCODE project for the histone modifications H3K4me3 and H3K27me3 in hESCs (H7), shown in Table 2.1. ESCs were used for NANOG ChIP-seq analysis given that TGCTs, like ESCs, express high levels of *NANOG* and the similarity in gene expression between TGCT embryonal carcinoma cells and ESCs (Skotheim *et al.*, 2005; Ben-David *et al.*, 2014). For each transcription factor, the ChIP-seq BED files were intersected using Galaxy

Table 2.1 ChIP-seq datasets used for analysis and their accession numbers.

Sample ID	ChIP	Cell Type
GSM1505695	NANOG	HUES64
GSM1505698	NANOG	HUES64
GSM1505699	NANOG	HUES64
GSM1505700	NANOG	HUES64
GSM1505724	OCT4	HUES64
GSM1505725	OCT4	HUES64
GSM1505728	OCT4	HUES64
GSM1505766	SOX2	HUES64
GSM1505768	SOX2	HUES64
GSM945185	H3K4me3	H7-hESC
GSM945186	H3K4me27	H7-hESC

(https://usegalaxy.org) to give a consolidated BED file for peaks present in all replicates for a given transcription factor. In addition to this, BED files were downloaded using the UCSC browser (http://genome-euro.ucsc.edu; GRCh37 hg19), for sequences 10kb, 5kb, 2kb and 1kb upstream from all known miRNA transcription start sites (TSSs). ChIP-seq and UCSC BED files were uploaded onto Galaxy, the ChIP-seq BED files were intersected with the BED files containing sequences upstream of miRNA TSSs. This produced a list of miRNAs which contained a given transcription factor upstream of their TSS. In order to overlap the lists produced for miRNAs with binding of NANOG, OCT4 and SOX2, Venny (v2.1.0; http://bioinfogp.cnb.csic.es/tools/venny/) a Venn diagram program was used to produce overlaps using miRNA gene names. For the main analysis, miRNAs with NANOG binding within 2kb upstream of the miRNAs main TSS were taken forward. Other analyses included using NANOG BED files for miRNAS and mRNA (2kb upstream of TSSs) for motif analysis using RSAT (http://www.rsat.eu/ (Medina-Rivera *et al.*, 2015)). In addition, GREAT (http://bejerano.stanford.edu/great/public/html/) (McLean *et al.*, 2010)) was used to analyse cis-regulatory regions of DNA binding events across the entire human genome (GRCh37/ hg19) NANOG BED files. The Integrative Genomics Viewer (IGV) was used to visualise ChIP-seq location and enrichment in the human genome (http://software.broadinstitute.org/software/igv/home).

2.2.2 TGCT Microarray Analysis for miRNAs

Datasets from the GEO datasets GSE31824 and GSE18155 ((Novotny et al., 2012) and (Palmer et al., 2010), respectively) were used to investigate differential expressed miRNAs in TGCTs and testis patient samples. GSE31824 contained 3 seminoma, 2 embryonal carcinoma and 3 testes patient samples. GSE18155 contained 4 seminoma, 3 embryonal carcinoma and 2 testes patient samples. The dataset miRNA names were converted to the miRBase version 20 using miRSystem (http://mirsystem.cgm.ntu.edu.tw). The two microarray datasets used different platforms to measure gene expression, therefore the datasets required separate analyses. GSE31824 dataset provided normalized signal for samples; fold change was calculated by dividing TGCT miRNA normalised signal by the testis. MiRNAs with a minimum of 1.5-fold change were determined as differentially expressed in TGCT and testis groups. GSE18155 datasets provided median expression values around corresponding quadruplicate spots for each miRNA after normalisation to the background. Therefore, a different approach was used to calculate differentially expressed miRNAs. MiRNAs with a positive median expression in the TGCT subtype and a negative expression value in the testis tissue were determined as highly expressed miRNAs in TGCTs compared to the control testis. MiRNAs with a negative expression value in TGCT subtypes but positive in the testis were taken as lowly expressed miRNAs in TGCTs. MiRNA lists from both GSE18155 and GSE31824 were consolidated to give a unified list for differentially expressed from the TGCT microarray analysis (full list shown in Table 3.1).

2.2.3 miRNA Target Prediction

In order to produce a list of miRNA predicted targets miRWalk v2.0 (http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk/) was used. MiRwalk allowed several computational algorithms to be used simultaneously producing a list of predicted miRNA targets and which algorithms made this prediction. The algorithms used were: DIANA (http://diana.cslab.ece.ntua.gr/), miRDB (http://mirdb.org/), RNA22-HAS (https://cm.jefferson.edu), TARGETMINER (http://www.isical.ac.in/), TargetScan and PICTAR (http://www.isical.ac.in/). The lists of predicted targets for each algorithm were then overlapped using Venny (http://bioinfogp.cnb.csic.es/tools/venny/). Targets present in 4 or more algorithms were kept for further processing.

2.2.4 TGCT TCGA mRNA-seq and miRNA-seq Analysis

Counts for mRNA and miRNA counts were downloaded from the Cancer Genome Atlas Project (TCGA; https://gdc-portal.nci.nih.gov/) for type II TGCT patient samples (n=156). Both mRNA and miRNA datasets were stratified based on patient clinical data provided using the criteria outlined in Table 2.2 or by *NANOG* expression. Differential expression was determined using Robina (http://mapman.gabipd.org/web/guest/robin) which utilises the EdgeR statistics package (Robinson, McCarthy and Smyth, 2010); genes with a mean log2 (normalized read counts) > 1 and an adjusted p-value <0.05 (FDR) between groups were described as significantly differentially expressed.

Table 2.2 The stratification of the TCGA RNA-seq patient samples.

Data Stratified by:	Methodology definition:
Nanog Expression: High, Medium and Low.	Based on counts divided into terciles.
Patients with an undescended testis.	Patients with an undescended testicle (unilateral or bilateral)
TGCT Subtype.	TGCT subtype were classed based on histological percentage (>80%), all heterogeneous (<80%) samples were not used.
Presence of CIS	Classified by NIH.
History of TGCT	Classified by NIH.
History of Cancer	Classified by NIH.
Cancer Stage	Classified by NIH using Pathological staging.
Cancer Metastasis	Classified by NIH using Pathological staging.

2.2.5 Clustering Analysis

Normalised counts for TCGA TGCT patient samples and RNA-seq samples for TCam-2 Scrambled and TCam-2 *miR-9* were log transformed. Using cluster v3.0 (de Hoon *et al.*, 2004) genes with less than 100 counts and genes where the max count– min count values is no greater than 32 were filtered out. Using the filtered genes, an average linkage hierarchical cluster was produced using centered mean values and a CDT file was produced. Java Treeview v3.0 (Saldanha, 2004) was used to visualise the CDT file to produce an unsupervised hierarchical clustering dendrogram and a gene expression heatmap. This image was then exported as a PNG file.

Using the logged and filtered genes from cluster 3.0, a Pearson correlation was performed in R (v3.4.1) and a correlation heatmap was produced. In addition to this, a PCA plot was also produced using the logged and filtered genes from cluster 3.0. The scripts for creating these plots are shown below:

Heatmap

library(ggplot2)

library(ggfortify)

Heatmap <- qplot(x=Var1, y=Var2, data=melt(cor(*data*, method='pearson'))

Principle Component Analysis

PCA <- autoplot(prcomp(*data*), data='gene ID',)

2.2.6 RNA-seq Analysis

Total RNA was extracted from cell pellets using the Qiagen miREasy kit (as described in section 2.5.1) for TCam-2 Scrambled and TCam-2 *miR-9* samples. The cDNA library preparation and RNA-sequencing was carried out by the Genome Centre at Barts and The London School of Medicine and Dentistry using a standard Illumina library cDNA library preparation kit and sequencing was conducted on an Illumina Next-seq RNA-sequencer. FASTQ files were processed using Trimgalore (Version 0.4.4) to remove adapter sequences and sequence reads were aligned to the human genome (GRCh38.83) using STAR (Dobin *et al.*, 2013). Differential gene expression was determined using by EdgeR (2.4.6) (Robinson, McCarthy and Smyth, 2010), this method produces a list of differentially expressed genes with calculated P-values and adjusted values for false discovery rate (FDR). A FDR adjusted p-value less than 0.05 and a fold change greater than 2 was the chosen level of significance.

2.2.7 Gene ontology and Pathway Analysis

In order to analyse the differentially expressed genes in TCam-2 *miR-9* and TCam-2 scrambled groups, the gene ontology software Webgestalt was used (Zhang, Kirov and Snoddy, 2005) (http://www.webgestalt.org/option.php). The input used the official gene symbols for significantly differentially expressed genes and the reference gene list used all annotated gene used for the RNA-seq analysis. A minimum number of 5 genes and a maximum of 2000 genes was used per category. Bayesian multiple test adjustment was used and the top 10 categories were observed, ranked in order of most significant. A report was generated for biological processes analysis, Reactome pathway analysis, KEGG analysis and network analysis using the TCGA TGCT RNA-seq datasets.

2.3 Testes RNA Samples

Total RNA was purchased from Ambion and Clontec. The Ambion testes sample were derived from one male normal testes aged 38 years old. The Clontec sample was derived from a pool of 5 males aged between 20-87 years of age.

2.4 Cell Culture

2.4.1 TGCT Cell Culture

NTERA2, 2102Ep and 833KE cells were cultured in Dulbecco's modified eagle medium (DMEM; Glucose 4.5g/L with L-glutamine and no sodium pyruvate; 41965-039) containing 10% fetal calf serum (Sigma F7524) and 5% L-glutamine (Sigma G3126). NCCIT and TCam-2 cells were cultured with RPMI medium (RPMI 1640 1x with L-glutamine; 21875-034) containing 10% fetal calf serum (Sigma F7524) and 5% L-glutamine (Sigma G3126). NCCIT were grown on plates with 0.1% gelatin. In order to prepare gelatin plates, 0.1% gelatin was added to culture plate and left at 37^{o}C for 30 minutes and then NCCIT cell suspension was added. Generally, all TGCT cell lines were grown in T25 flasks. 2102Ep, 833KE NCCIT and TCam-2 cells were all passaged at ~90% confluence through the use of 0.05% trypsin (Trypsin-EDTA 1x; Gibco 25300-062). In order to split cell, the medium was first removed, washed with PBS twice and then 1ml of 0.05% trypsin (Trypsin-EDTA 1x; Gibco 25300-062) was added for a maximum of 5 minutes at 37^{o}C. Once cells had detached, 5ml of culture medium was added in order to inactivate the trypsin. In contrast, NTERA2 cells were passaged at ~90% confluence mechanically using a cell scraper (Biofil CSC011025). For all cell types, the cell suspension mixture was transferred into a 15ml falcon and spun at 300g for 3 minutes. The supernatant was aspirated and the pellet was suspended in 3mls of medium. Cell were split

using a 1:3 ratio, therefore 1ml of cell suspension was added to a T25 containing 5mls of medium.

2.4.2 hESC Cell Culture

The hES cell line H9 (WiCell, Wisconsin) was used and analysed. H9 was obtained under the Material Transfer Agreement and the use of these cells were authorised by the MRC Stem Cell Bank Steering Committee (Approvals SCSC 10-16 and SCSC 12-47). In order to culture these cells Vitronectin (Stem cell technologies Catalog #07180) was used to coat the flask using a final concentration of 10µg/ml. Coated flasks were incubated at room temperature for 1 hour before use. hESCs were cultured in Essential 8 (E8) medium. Cells were split at 90%confluence using Tryple (Thermo: 12605036). In order to split cells the medium was removed, the cells were washed twice with PBS and then 1ml of Tryple was added for a maximum of 5 minutes at 37ºC. Once cells had detached, 5ml of culture medium was added in order to inactivate the Tryple. The cell suspension mixture was transferred into a 15ml falcon and spun at 300g for 3 minutes. The supernatant was aspirated and the pellet was suspended in 3mls of medium. Cell were split using a 1:3 ratio, therefore 1ml of cell suspension was added to a T25 containing 5mls of medium.

2.4.3 Counting of cells

In order to determine cell numbers from a cell suspension, a haemocytometer was used (Hausser Scientific, 1490). 10µl of cell suspension was pipetted between a coverslip and the haemocytometer. Four quadrants were counted for cell number to give an average number of cells for one quadrant. The average cell count number was then multiplied by 10^4 to give an estimated cell count (cells/ml).

2.4.4 Cryopreservation of cells

All cell lines were washed with PBS and trypsinised for 5 minutes at 37ºC. Cell suspensions were spun at 300 x g for 3 minutes and re-suspended in a 1:1 ratio of media and 20% dimethyl sulphoxide (DMSO) (Fisher, D/4121/PB08). Cells were transferred into a cryopreservation container and stored at -80ºC.

2.4.5 Thawing of cells

Frozen cells contained cryovials stored in liquid nitrogen were thawed at 37ºC using a water bath. Medium was added slowly to cells to dilute the DMSO and minimize cell death. The cell suspension was spun for 3 minutes at 300 x g, the supernatant was removed and the cell pellet was re-suspended in fresh medium. The cell suspension was added to a 25cm^2 culture flasks (Thermo-Scientific Nunc, 136196) and cultured at 37ºC.

2.5 Generation of Transgenic Cell lines

2.5.1 HEK293T Culture

HEK293T (293T) cells were used as a vector to grow lentivirus particles, these cells were created to over express the large T-antigen which has been shown to improve replication of plasmids containing to viral promoter SV-40.HEK293T cells were defrosted from the liquid nitrogen and warmed until defrosted. Cell suspension was added to 5ml of medium and spun at 300g for 3 minutes. The medium used was DMEM; Glucose 4.5g/L with 5% L-glutamine (Sigma G3126) and no sodium pyruvate; 41965-039) containing 10% fetal calf serum (Sigma F7524). The supernatant was removed, the pellet was re-suspended in 15ml of fresh medium and the cell suspension was added to a T75 flask. To pass cells, cells were washed in PBS and 2ml of trypsin (0.01%) was added to cell for 5 minutes. Following this, 10ml of fresh medium was added to neutralize the trypsin and the cell suspension was spun at 300g for 3 minutes. The supernatant was removed and the cell pellet was re-suspended in 15ml of medium, 1ml of this medium was added to 15ml of fresh medium in a T75 to give a 1:15 split. The flask were then culture at 37^{o}C at until confluent.

2.5.2 Production of Lentiviral particles

In order to create virus particles, HEK293T cells were seeded into a T25 at a density of 2.3 x 10^6 cells/ml. Cell were left to attach overnight at 37^{o}C and then 5ml of fresh medium was added to cells. In order to transfect cells, 200µl of Optimem (Gibco: 31985) was added to a 1.5ml eppendorf followed by 12.5µl of FuGENE 6 (Promega: E2691), the mixture was allowed to equilibrate for 5 minutes at room temperature. The following plasmid were added to the Optimem mixture for lentivirus; 4.6µg of the lentiviral transfer plasmid, 3.8µg of the lentiviral packaging plasmid psPAX2 (Plasmid 12260, Addgene) and 2.5µg of the envelope plasmid pMD2.G (Plasmid 12259, Addgene). For retrovirus production, the following was added to the Optimem mixture; 4.6µg of the lentiviral transfer plasmid, 3.8µg of the retroviral packaging plasmid pUMVC (Plasmid 12260, Addgene) and 2.5µg of the envelope plasmid pMD2.G (Plasmid 12259, Addgene). A list of the transfer plasmids used are described below in Table 2.3. The mixture containing Optimem, FuGENE and plasmids were added

Table 2.3 Transfer plasmids used for viral transduction for over-expression or knockdown of genes.

Plasmid Name	Addgene Ref	Virus	Selection Marker
pSin-EF2-Nanog-Pur	16578	Lentivirus	Puromycin
Tet-pLKO-puro	21915	Lentivirus	Puromycin
MDH1-PGK-GFP_2.0	11375	Retrovirus	GFP

to the HEK293T cells overnight (~12hours) for transfection. After overnight incubation, the transfection mixture was removed from the cells and 5ml of fresh medium containing 8μg/ml of hexadimethrine bromide (polybrene) solution (Sigma: H9268) was added to cells. Polybrene has been shown to increase transduction of viral particles (Davis, Morgan and Yarmush, 2002). The cells were left for 48 hours to allow sufficient viral production. After 48 hour incubation, the medium was passed through a 0.45 μM pore filter and was directly added to cells to be infected.

2.5.3 Antibiotic toxicity testing

Puromycin (Gibco Invitrogen: A11138-03) was used on cells which had been infected with a virus containing a transfer plasmid encoding a puromycin resistance marker. First, a concentration gradient for puromycin was added to wild-type cells in order to determine at which concentration cell without a puromycin resistance gene would be killed. This was performed on cells in a T25 which were at approximately 30% confluence, puromycin was added to cells for a week in total and was replenished with fresh medium containing puromycin every 2 days. After 7 days, the lowest concentration that killed WT cells was the concentration used for selecting cells infected with the virus. The virus infected cells were also subjected to puromycin at 30% confluence and for 7 days. The surviving cells were determined to have successfully integrated the transfer plasmid into the genome.

2.5.3 Flow Cytometry (FACS) analysis

For TCam-2 cells infected with the retrovirus, the selection marker contained GFP. In order to assess that the infection had worked, cells were analysed by microscopy to visualise fluorescence. Cells were grown in T75 flasks until 90% confluence and were washed with PBS and the trypsinised (0.01%) at 37ºC for 5 minutes. 5ml of fresh medium was added to cells to dissociate and then spun at 300g for 3 minutes. Cell were then re-suspended in 500µl of medium and transferred into FACS tubes. The cell suspension was analysed by the MoFlo XDP FACS machine to analyse all cells within the suspension. Data was then analysed using WEASEL (v3.2.1).

2.6 Luciferase Assay

2.6.1 Transfection of cells with pGL3 Plasmid

For plasmid transfection HEK293T cells were seeded in a 24-well at a density of 90,000 cells/ml. The plasmids pGL3 (Promega E1751) and Renilla (Promega E2231) were co-transfected into cells. 1.7 µl of Fugene 6 (Promega E2691) was added to 25µl Optimem (Thermo 31985062) and left to incubate at room temperature for 5 minutes. Following this, the plasmids were added to the mixture giving a working concentration of 150ng/µl for pGL3 and 0.03 ng/µl for Renilla. The mixture was incubated for 15 minutes at room temperature before the mixture was added directly to cells and left for 2-3 days before analysis.

2.6.2 Transfection of cells with psiCHECK-2 Plasmid and miRNA mimics

For plasmid transfection, HEK293T cells were seeded in a 24-well at a density of 90,000 cells/ml. The psiCHECK-2 plasmid (Promega C8011) was co-transfected with miRNA mimics (Scrambled or *miR-9-5p* (Sigma HLMIR0946)) into cells. 3 µl of Lipofectamine 2000 (Thermo 11668027) was added to 25 µl of optimum. The psiCHECK-2 plasmid was added to the mix at a concentration of 300ng/µl and the miRNA mimics at 5pmol/µl and left to equilibrate for 15 minutes and then added directly to the cells and left for 2-3 days before analysis.

2.6.3 Luciferase Analysis

Following the transfection of luciferase plasmids into HEK293T cells, cells were analysed for luciferase expression using the Promega dual-luciferase assay system (E1910). The medium was removed from the cells and 100 µl of passive lysis buffer was added. Plates were then frozen at -20^{o}C for 1 hour to aid lysis of cells. The mixture was defrosted, pipetted up and down to mix and transferred into fresh tubes. To measure luciferase, 50µl of LARII buffer was added to a fresh tube to which 5 µl of lysis mixture was added and measured for luciferase. 50 µl of Stop & Glo buffer was added to the mixture and measured for Renilla luciferase. The luciferase was measure using the Promega Glomax 20/20 single tube luminometer.

2.7 Molecular Techniques and Analyses

2.7.1 RNA Extraction

RNA extraction was performed on cells using Qiagens miRNeasy mini kit (217004) according to the manufacturer's protocol. This procedure involved, taking a cell pellet which had been stored from the -80^{o}C freezer and adding 700µl of fresh culture medium directly to the pellet. Cells were homogenised by pipetting and then left at room temperature for 5 minutes. Following this, 150µl of chloroform was added to the cell mixtures and shaken vigorously for 15 second. After this mixture was left at room temperature for 3 minutes, the mixture was centrifuged at 4^{o}C for 15 minutes at 12,000 x g. The clear aqueous phase was removed and placed into a fresh eppendorf where 525µl of 100% ethanol was added and mixed by pipetting. This mixture was then added to the miRNeasy column and spun at 8,000 x g for 15 seconds. At this point, DNase treatment was carried out on the column. 10µl of DNase I and 70µl of DNase I Buffer was mixed and added to the column followed by a 15 minute wait (Sigma;Dnase70). Next, 500µl of RWT buffer was added directly to column and then centrifuged at 8,000 x g for 15 seconds. 500µl of RPE was then added to the column and centrifuged for 15 seconds, 500µl of RPE was again added but the column was spun for 2 minutes this time round. Following this, the column was placed into a fresh eppendorf and spun at maximum speed for 2 minutes to dry. The column was again placed into a fresh eppendorf, 30µl of RNase free H_20 (ThermoFisher; 10977035) was added and then the column was spun at 8,000 x g for 3 minutes. The contents in this tube contained the extracted RNA. In order to determine the concentration of the RNA samples a Nanodrop machine

was used. RNA with a 260/280 ratio between 1.8 and 2.1 was determined as good quality.

2.7.2 cDNA Synthesis

In order to synthesis cDNA the miRNeasy Reverse Transcription kit (miScript II; 218161) was used. The HiFlex buffer was used which prepared cDNA for the detection of both microRNAs and mRNAs. cDNA synthesis was performed by adding a total of 0.5µg of RNA, 2µl of HiFlex buffer, 1µl of reverse transcriptase and made up to 10µl with RNase free H_20. Samples were then incubated at 37^oC for 60 minutes for the cDNA synthesis and then at 95^oC to inactivate the reverse transcriptase enzyme.

2.7.3 RT-QPCR

MiRNA forward primers were designed based on the miRNA sequences provided by miRBase (http://www.mirbase.org/), primers shown in Table 2.6, whereas the reverse primer was a universal primer complimentary to the universal tag which was purchased from Qiagen (218300). RT-QPCR was performed using Sigma's Lumino Ct SYBR Green qPCR readymix (L6544). RT-QPCR reactions used the following makeup; 7.5µl SYBR green, 1µl Forward primer (5µM), 1 µl Reverse primer (5µM), 3.5 µl ddH_20 and 2 µl cDNA (0.5ng/µl); total 15µl. The following RT-QPCR program was used for miRNA; initial 95°C – 5 minutes and then 40 cycles of 94°C – 15 seconds (s), 55°C – 30s and 72°C – 5s (end cycle) followed by a melting curve. In contrast to this, the RT-QPCR for mRNA was as follows; initial 95°C – 5 minutes and then 40 cycles of 95°C – 3 seconds (s), and 72°C – 30s (end cycle) followed by a melting curve. RT-QPCR was conducted on the following NT2 cDNA concentrations 1ng, 0.01ng, 0.003ng, 0.0001ng in order to calculate the slope and then the primer efficiencies using the formula Efficiency (E) = 10 ^ (-1/slope). Furthering this, the Pflaffl equation for normalising and calculating fold change for RT-QPCR was used, equation shown in Figure 2.1 (Pfaffl, 2001).

$$R = \frac{(E_{target})^{\Delta CPtarget\ (control - sample)}}{(E_{ref})^{\Delta CPref\ (control - sample)}}$$

Figure 2.1 Pflafl equation for calculating fold change normalized for primer efficiency in addition to biological repeat variation (using an endogenous control).

2.7.4 Cloning

In order to insert fragments into a given plasmid, the region of interest was first amplified using Q5 High-Fidelity DNA Polymerase (M0491). For a 25µl reaction, 5µl of 5x reaction buffer mix, 0.5µl of 10mM dNTPs (Invitrogen, 18427-013), 10µM forward primer, 10µM reverse primer, 10ng template DNA, 0.25 µl Q5 high-fidelity DNA polymerase which was made up to 25µl with ddH_20. The PCR mixes were placed on ice and transferred to a thermos-cycling machine to perform the following PCR reaction:

Table 2.4 PCR cycling program for Q5 polymerase PCR.

Program Step	Temperature (°C)	Time	Cycles
Initial denaturation	98	30 secs	1
DNA denaturation	98	5 secs	
Annealing	45-72	30 secs	35
Extension	72	20-30 secs / kb	
Final Extension	72	2 mins	1
Hold	4	-	

The optimum annealing temperatures was calculated using the NEB Tm online calculator (https://tmcalculator.neb.com/#!/). The PCR mixture was run on an agarose gel by gel electrophoresis and purified using nucleospin Gel and PCR extraction kit (Section 2.5.9). Purified amplicons were the digested using restriction enzymes.

Table 2.5 Viral plasmids used in this study, the enzymes and buffers used for restriction digest of plasmids.

Plasmid	Enzyme 1	Enzyme 2	Buffer
psiCHECK-2	XhoI (R0146)	NotI (R0189)	Buffer 3.1
pGL3	NheI-HF (R3131)	SacI-HF (R3156)	Cutsmart
pLKO-Tet	XhoI (R0146)	-	Cutsmart
MDH1-PGK-GFP-miR9	XbaI (R0145)	ClaI (R0197)	Cutsmart

2.7.5 Restriction Enzyme Digest

The PCR amplicons for insertion and the plasmid backbones were cut with the restriction enzymes for the corresponding plasmid (Table 2.5), to produce sticky ends. In a total volume of 30µl, 1µg of DNA with 3µl of buffer and 10 units of enzyme was incubated at 37°C for 1 hour. The digested products were run on agarose gels to separate products and purified using gel extraction (see section 2.5.9). Up to 70% of product was purified from the gel extract. For the ligation of PCR inserts into plasmid back bone a 1:3 ratio was used. In order to calculate this ratio, the below formula was applied:

$$\frac{Vector\ (ng) x\ insert\ size\ (kb)}{Vector\ size\ (kb) x\ insert\ (ng)} = \frac{1}{3}$$

For the ligation reaction, a 10µl reaction containing, 1µl of T4 DNA ligase (M0202), 1µl of 10x T4 DNA ligase buffer, 50ng of DNA insert and 100ng of plasmid made up to 10µl with ddH_2O was mixed, centrifuged and stored at 4°C overnight.

2.7.6 Bacterial Transformation

Competent *E.coli* cells (Stabl3; Invitrogen, C7373-03) were thawed on ice. The ligation reaction was added to the bacteria and incubated on ice for 20 minutes. The bacteria was then heat shocked at 42ºC for 45 seconds and then incubated on ice for 2 minutes. 200µl SOC medium was (Invitrogen, 15544-034) was added to cells and the mixture was incubated at 37ºC and shaken at 150rpm for 1.5 hours to allow the bacteria to recover. This mixture was then spread onto LB plates containing 50µg/mL ampicillin (Sigma: A9518-SG) and incubated at 37ºC overnight. Ampicillin resistant colonies were formed on the plate as a result of successful transformation.

2.7.7 Growing Bacteria

In order to increase the yield of plasmid, the bacterial cells were grown in LB nutrient medium (Oxoid, CM001) containing 50µg/mL ampicillin. Individual colonies were picked and added to 5ml of LB medium containing ampicillin in a universal tube (Barloworld Scientific Ltd, 128A). The mixture was incubated at 37ºC and shaken at 150rpm overnight to allow the bacterial to grow.

2.7.8 Freezing Bacterial Stocks

In order to keep a stocks for plasmids with insertions, 800 µl of LB medium containing bacterial growth was added to 200 µl glycerol and mixed. This mixture was stored at -80ºC.

2.7.9 Plasmid purification by Miniprep

The purification of plasmids from bacterial cultures was performed using a Miniprep kit (Nucleospin: 740615). The 5ml of bacterial LB culture was spun for 5 minutes at 6000 x g in order to pellet the bacteria. The supernatant was removed and the pellet was re-suspended in 250µl of Buffer A1. The mixture was vortexed to ensure the pellet had completely re-suspended. Next, 250 µl of Buffer A2 was added to the mixture and were inverted 6-8 times to mix. The mixture was then incubated at room temperature for 5 minutes. 300 µl of Buffer A3 was added and inverted 6-8 times to ensure mixing. The mixture was centrifuged for 5 minutes at 11,000 x g and the supernatant was then loaded on to a Nucleospin column and centrifuged for 1 minute at 11,000 xg to bind the DNA to the column. To wash the column, 450µl was added to the column and then spun for 3 minutes at 11,000 x g. To dry the column, the column was transferred into a new collection tube and then spun again for 3 minutes at 11,000 x g. The column was transferred to another fresh eppendorf tube and 50 µl of ddH_20 was added to the column to elute the DNA. After a one minute incubation at room temperature, the column was spun at 11,000 x g for 1 minute. The elution liquid in the eppendorf contained the extracted plasmid DNA. In order to determine the concentration of the DNA samples a Nanodrop machine was used. DNA with a 260/280 ratio between 1.8 and 2.0 was determined as good quality.

2.7.10 Gel and PCR Purification

In order to purify PCR products from an agarose gel or directly from the PCR mix the Nucleospin Gel and PCR clean-up kit was used (Nucleospin: 740615). For DNA purification from an agarose gel, the excised band was added to buffer NTI (200µl/100mg) and incubated at 50^{o}C for 10 minutes until the band had completely dissolved. For the purification directly from the PCR mix, 2 volumes of buffer NTI was added to 1 volume of PCR mix. To bind the DNA, the mixture was then transferred to a Nucleospin Gel and PCR clean-up column and spun at 11,000 x g for 30 seconds. To wash the membrane, 700µl of buffer NT3 was added to the column, which was the spun for 30 seconds at 11,000 x g. To dry the membrane, the column was transferred into a new collection tube and spun for 1 minute at 11,000 x g. The column was transferred to another new eppendorf and 20µl of ddH_20 was added to the column. The tube was incubated for 1 minute at room temperature and spun for 1 minute at 11,000 x g to elute the DNA. In order to determine the concentration of the DNA samples a Nanodrop machine was used. DNA with a 260/280 ratio between 1.8 and 2.0 was determined as good quality.

2.7.11 Genomic DNA Extraction

In order to purify genomic DNA from cultures, cell pellets the Qiagen DNeasy blood and tissue kit was used (Qiagen, Cat No. 69504). Cell pellets were re-suspended in 200µl of PBS and 20µl Proteinase K. 200 µl of buffer AL was added to the mixture and incubated for 10 minutes at 56^{o}C. 200 µl of 100% ethanol was added and the mixture was transferred into a DNeasy mini spin column and spun at 6,000 x g for 1 minute. 500 µl of buffer AW1 was added to the column and was spun at 6,000 x g for 1 minute. 500 µl of buffer AW2 was added to the column and was spun for 3 minutes at 17,000 x g for 3 minutes. The column was placed in to a new eppendorf and 100 µl ddH20 was added to the column. The column was incubated for 1 minute at room temperature and spun at 6,000 x g for 1 minute to elute DNA. In order to determine the concentration of the DNA samples a Nanodrop machine was used. DNA with a 260/280 ratio between 1.8 and 2.0 was determined as good quality.

2.7.12 Gel Electrophoresis

In order to separate DNA products on agarose gels, gel electrophoresis was used. For the separation of small DNA products sized (50-150bp) a 2% agarose gel was used, larger DNA fragment (150bp+) were run on a 1% gel for separation. To create the agarose gel, agarose powder (Fisher, BPE1356) was weighed out and added to 100ml of TAE buffer (0.04M Tris-Acetate, 1mM EDTA) in a conical flask and heated in a microwave until the powder had dissolved. The mixture was allowed to cool before the addition of Nancy-520 (Sigma-Aldrich, 01494). The mixture was poured into a gel cassette with a comb and allowed to cool until the gel had set. The gel was added to the gel tank and the tank was filled with fresh TAE until the gel was completely submerged. 5 µl of ladder(1kB DNA ladder (Biolabs, N3232S) or low molecular weight DNA ladder ((Biolabs, N3233S)) was added to the first well followed by the addition of DNA samples (containing 1x loading buffer (NEB #B7025)). A power pack was connected to the gel tank and the gel was run at 100 volts for 1 hour. The gel was then exposed to ultraviolet light and visualised using ImageQuant Capture 300 (GE Healthcare). Bands were then compared to the DNA ladder to calculate approximate size.

2.8 Statistics

Data is expressed as mean ± standard error mean (S.E.M) for biological replicates. A one-way ANOVA test was used to analyse samples with more than two groups for comparison. A two-tailed student t-test was used to analyse samples containing two groups for comparison. These statistical analyses were performed using GraphPad Prism 6.

2.9 PCR Primers

Table 2.6 Primers for miRNA RT-QPCR

miRNA	**Primer Sequence (5' - 3')**
miR-512-3p	AAGTGCTGTCATAGCTGAGGTC
miR-183-3p	GTGAATTACCGAAGGGCCATAA
miR-183-5p	TATGGCACTGGTAGAATTCACT
miR-9-5p	TCTTTGGTTATCTAGCTGTATGA
miR-9-3p	ATAAAGCTAGATAACCGAAAGT
miR-887-5p	CTTGGGAGCCCTGTTAGACTC
miR-887-3p	GTGAACGGGCGCCATCCCGAGG
RNU6B	CGCAAGGATGACACGCAAATTCGTG
MIR-26B-5P	GTAGTGATGAAATTCCACTTCATTG
SNORDU75	GTGGACAGAAGGGATTTCTGAAATTCTA

Universal Primer was purchased from QIAGEN part of the miScript SYBR Green kit (Cat No./ID: 218073).

Table 2.7 Primers for NANOG enriched genomic loci upstream of *miR-9-2* TSS.

pGL3 insertion	
Primer	**Primer Sequence (5' - 3')**
*miR-9*_upstream_F	CTTGCTAGCGAAGACCAATACACTC
*miR-9*_upstream_R	CTCGAGCTCGGCTTTATACATCTC

Table 2.8 Primers for psiCHECK-2 inserts for *miR-9* predicted targets.

For psiCHECK-2 Plasmid Inserts	
Primer	**Primer Sequence (5' - 3')**
*LIN28A*_1_F	TTCTCGAGGGAACTTTTATGCAGGTGCTAG
*LIN28A*_1_R	AGGCGGCCGCCAGGTACAGGCTTTCCTACC
*NR5A2*_1_F	TTTCTCGAGAGGGGAAAGGAAGAGAGTGA
*NR5A2*_1_R	AAGGCGGCCGCCCACATCTGCACAGCTACC
*NR5A2*_2_F	TTTCTCGAGGGCGGCACAATCACTTTG
*NR5A2*_2_R	AAGGCGGCCGCCCTTTGAGCCACGAAGATAACAC
*ESRRB*_F	TTTCTCGAGGTGACAAGAAGACTGAAGGG
*ESRRB*_R	AAGGCGGCCGCACAAGTAGGTGATGCTGCA
*CCNE1*_F	TTTCTCGAGCAGCTTGGATTTGCTGGAC
*CCNE1*_R	AAGGCGGCCGCGCAACAGACAGAAGAGAACG
*KDM4C*_F	TTTCTCGAGGCTGAGCTCCAGGTTTAAG
*KDM4C*_R	AAGGCGGCCGCGTGGCTTTTCCTTGTGGG
*KDM7A*_F	TTTCTCGAGCCACTGATCCAGTGACTCC
*KDM7A*_R	AAGGCGGCCGCCCAATGGATCCCAGCTGAAC

Table 2.9 RT-QPCR primers for predicted *miR-9* targets

RT-QPCR for *miR-9* Targets	
Primer	**Primer Sequence (5' - 3')**
βACTIN F	CCTCGCCTTTGCCGATCC
βACTIN R	CGCGGCGATATCATCATCC
GAPDH F	CGCTTCGCTCTCTGCTCCTCCTGT
GAPDH R	CGCTTCGCTCTCTGCTCCTCCTGT
LIN28A For	AAGCGCAGATCAAAAGGAGA
LIN28A Rev	CTGATGCTCTGGCAGAAGTG
CCNE1 F	AAAGTTGCACCAGTTTGCGT
CCNE1R	AAGGCCGAAGCAGCAAGTAT
KDM7A F	CGAGTGCGATATCTGCAAGG
KDM7A R	GTCATGTCTGTGCCAGTTCCT
LIN28B For	TGATAAACCGAGAGGGAAGC
LIN28B Rev	TGTGAATTCCACTGGTTCTCC
KDM4C F	CGAGGTGGAAAGTCCTCTGAA
KDM4C R	GGGCTCCTTTAGACTCCATGTAT
NR5A2 F	CCATTATGGGCTCCTCACC
NR5A2 R	CGTGCTTGCTTCGGTTAGC

Chapter 3: Bioinformatics analysis of TGCT transcriptome and ChIP-Seq datasets

3.1 Introduction

This chapter aimed to use available bioinformatic datasets in an attempt to interrogate the relationship between NANOG and miRNA regulation in TGCTs. Nanog is essential for ESC pluripotency and PGC development (Loh *et al.*, 2006; Chambers *et al.*, 2007; Yamaguchi *et al.*, 2009). In mouse ESCs, low Nanog expression is associated with increased cellular differentiation (Chambers *et al.*, 2007) and decreased proliferation (Zhang *et al.*, 2009). NANOG forms an auto-regulatory forward feedback loops with the pluripotency master regulators OCT4 and SOX2 to enhance ESC self-renewal (Boyer *et al.*, 2005). Mouse ESCs with low levels of *Nanog* show an increased propensity to differentiate but are still able to be maintained indefinitely in culture (Chambers *et al.*, 2007). NANOG also transcriptionally represses *CDK6* and *CDC25A* in hESCs in order to lengthen the S-phase and shorten the G1-phase of the cell cycle (Neganova *et al.*, 2009; Zhang *et al.*, 2009). This cell cycle profile is associated with pluripotency and increased proliferation, whereas the lengthening of the G1-phase is associated with differentiation and a decrease in proliferative capacity (Calder *et al.*, 2012). Mouse PGCs express Nanog while proliferating during their migratory stages (Yamaguchi *et al.*, 2005) and Nanog knockdown results in a reduction of PGCs within the developing gonad (Chambers *et al.*, 2007; Yamaguchi *et al.*, 2009). These Nanog knockdown mouse PGCs also exhibit a reduction in proliferation (Yamaguchi *et al.*, 2009), suggesting a similar role for Nanog in PGCs and ESCs. This evidence gives a clear role for Nanog in

regulating proliferation and pluripotency. There is now increasing understanding that NANOG regulation of miRNAs is an important mechanism involved in the maintenance of pluripotency (Kashyap *et al.*, 2009).

MiRNAs are processed by two main enzymes, Dicer and Dgcr8 (Ketting *et al.*, 2001; Han *et al.*, 2004). Functional targeting of these enzymes have been used to investigate miRNA function in mESCs (Greve, Judson and Blelloch, 2013) and when either of these proteins are deleted, cellular morphology remains normal and cells continue to express pluripotency genes. Upon differentiation of Dicer knockout (KO) or Dgcr8 KO mESCs, both genotypes fail to repress markers for pluripotency (Kanellopoulou *et al.*, 2005; Wang *et al.*, 2007). In addition to this, both knockout mice exhibit a reduction in key differentiation markers during embryoid body formation (Wang et al. 2007; Kanellopoulou et al. 2005) thereby indicating a clear role for miRNAs in silencing pluripotency in mESCs.

Consistent with this, the introduction of the miRNA let-7 into Dicer KO mESCs downregulates numerous pluripotency factors including Nanog, Oct4 and Sox2 indicating a role for let-7 in cell differentiation (Melton, Judson and Blelloch, 2010). Inhibition of pluripotency factors by let-7 does not occur when let-7 is co-introduced with a group of miRNAs that have been termed embryonic stem cell–specific cell cycle–regulating (ESCC) miRNAs (Greve, Judson and Blelloch, 2013). The ESCC miRNAs include miRNAs from two main families, the *miR-302* family and the *miR-371*/373 (*miR-290*/295 in the mouse), and have been shown to be important for the regulation of the cell cycle in ESCs (Greve, Judson and Blelloch, 2013). In addition to the effect on pluripotency, Dicer KO and Dgcr8 KO mESCs also exhibit a reduction in proliferation (Wang *et al.*,

2008). These knockout mESCs have an abnormal cell-cycle which results in the accumulation of ESCs during the G1 phase of the cell cycle (Wang *et al.*, 2008). This shows that miRNAs are important for the ESC cell-cycle profile. The ESCC miRNAs target inhibitors of the G1 to S-phase transition, such as *Cdkn1a*, *Lats2* and *Rbl2* (Wang *et al.*, 2008). Interestingly, the *miR-302* cluster (part of the ESCC miRNAs) is positively regulated by NANOG (Card *et al.*, 2008; Barroso-del Jesus, Lucena-Aguilar and Menendez, 2009), whereas let-7g (part of the let-7 family) is transcriptionally repressed by NANOG (Marson *et al.*, 2008). Therefore, NANOG can transcriptionally activate miRNA that sustain proliferation and pluripotency and repress miRNAs that induce differentiation by binding to their promoter regions.

Chromatin immunoprecipitation followed by DNA sequencing (ChIP-seq) is a technique used to identify protein-DNA interactions (Park, 2009). Loh and colleagues, 2006, used Nanog ChIP-PET (paired end ditag) in mESCs which involved inserting the Nanog enriched DNA into plasmids followed by sequencing and alignment to the genome. Nanog co-localises with Oct4 at many gene regions to activate stem-cell associated genes while repressing genes promoting differentiation (Loh *et al.*, 2006). Advances in technology produced ChIP-seq which uses antibodies to bind specific proteins bound to DNA regions which are then sequenced directly and aligned to the genome thus indicating genomic regions of protein binding (Park, 2009; Thomas-Chollier *et al.*, 2012). In mESCs, similar work has been conducted in hESCs in order to identify genomic binding sites for the transcription factors NANOG, OCT4 and SOX2 together with the histone marks H3K4me3 and H3K27me3 (Tsankov *et al.*, 2015). H3K4me3 is associated with active chromatin whereas H3K27me3

associates with a repressive state of chromatin (Bernstein *et al.*, 2006). Therefore, this can be used to identify genomic regions associated with expression. Currently, there are no NANOG ChIP-seq data from TGCTs or PGCs, however hESCs offer a viable alternative, as embryonal carcinoma (EC) TGCT cells show a similar mRNA expression profile to hESCs (Skotheim *et al.*, 2005). Furthermore, ECs, seminomas (SEMs) and ESCs express similar levels of *NANOG* (Skotheim *et al.*, 2005). Therefore it is believed this ChIP-seq dataset can be used for the identification of miRNA putative regulatory regions which are transcriptionally regulated by NANOG binding in TGCTs.

In addition to this, datasets using two other technologies (microarray and RNA-Seq) can be used to investigate expression profiles of miRNAs and their potential targets in TGCTs. Microarray technology uses a chip containing a series of DNA probes, which can bind complementary cDNA fragments fluorescently labelled that can then be quantified and related to gene expression levels (Thomson *et al.*, 2004). Novotny (2012) and Palmer (2010) both used microarrays to determine differential miRNA expression in TGCTs and testes controls to identify miRNAs important for TGCT development. Palmer used a mixture of germ cell tumour (GCT) samples obtained from both sexes at different stages of progression in order to identify miRNAs associated with GCT malignancy (Palmer *et al.*, 2010). This study identified the *miR-371*/373 cluster of miRNAs which has subsequently been shown to be useful for identifying TGCT patients using serum samples, thus providing a platform for further development of TGCT diagnostics (Syring *et al.*, 2015). Following this, Novotny sought to elucidate the expression profiles for miRNAs for the carcinoma *in situ* (CIS) counterpart of the TGCT tumour. As CIS cells only constitutes a small

percentage of the tumour, the amounts of CIS in biopsies were determined using histological evaluation for placental alkaline phosphatase protein. It was shown that the expression of *miR-302*/367 and *miR-371*/373 positively correlated with the amount of CIS in samples (Novotny *et al.*, 2012).

Another source of TGCT gene expression data is the TCGA dataset (https://portal.gdc.cancer.gov/projects/TCGA-TGCT). The TCGA (The Cancer genome Atlas) is a collaboration of the National Cancer Institute and the National Human Genome Research Institute which has generated RNA-seq datasets spanning 33 types of cancer (https://cancergenome.nih.gov). Within the TCGA, 156 TGCT patient samples were analysed for mRNA and miRNAs expression which included clinical information for each individual patient. Other studies have used the TCGA data analysis to further cancer diagnosis and treatment. For instance, Verhaak (2010) utilised the RNA-seq, somatic mutation and clinical data from the glioblastoma TCGA (Verhaak *et al.*, 2010). Using the sample's expression profiles, consensus average linkage hierarchical clustering identified four groups which were then classified into clinically relevant subtypes of glioblastoma, two of which were previously unclassified. Using available clinical data, the newly defined glioblastoma subtypes were analysed for treatment type and survival analysis in order to identify improve treatment efficacy. This study showed how the TCGA can be used to advance understanding of cancer, improve diagnosis and transform disease characterisation, diagnosis and treatment for improved patient outcome, but also showed how these datasets can be used to further the understanding of the biology behind a given disease.

In summary, both NANOG and miRNAs have a role in regulating proliferation and pluripotency in pluripotent cells (Wang *et al.*, 2008; Zhang *et al.*, 2009). It has also been shown that NANOG can control the regulation of miRNAs (Card *et al.*, 2008). ChIP-seq data sets for NANOG and other pluripotency markers are available for hESCs which can be used for further investigation of NANOG binding. In addition to this, microarray and TCGA (RNA-seq) datasets that have quantified the expression of miRNAs in TGCTs are available to establish a relationship between NANOG and miRNA regulation in TGCTs (Palmer *et al.*, 2010; Novotny *et al.*, 2012; Weinstein *et al.*, 2013).

Aims and Objectives

The aim of this chapter was to use available bioinformatics datasets in order to produce a list of miRNAs which might be transcriptionally regulated by NANOG in TGCTs. The experimental strategy is summarised in Figure 3.1.

The main objectives were:

1. To analyse hESC NANOG ChIP-seq datasets in order to produce a list of miRNAs with NANOG binding upstream of miRNA TSSs.

2. To identify differentially expressed miRNAs from available microarray published datasets comparing TGCT and Testes groups.
3. To characterise the TGCT TCGA dataset using unsupervised clustering for gene expression to define cancer subtypes.
4. To stratify TCGA TGCT samples into *NANOG* high, medium and low tertiles of expression and analyse to give differentially expressed miRNAs in these groups.
5. To overlap the lists of miRNAs identified from (1) NANOG ChIP-seq analysis, (2) differentially expressed miRNAs from the microarray analysis (TGCT vs Testes) and (3) differentially expressed miRNAs from *NANOG* high, medium and low groups from TCGA TGCT datasets; in order to give a consensus list of candidate NANOG-regulated miRNAs possibly involved in TGCTs.

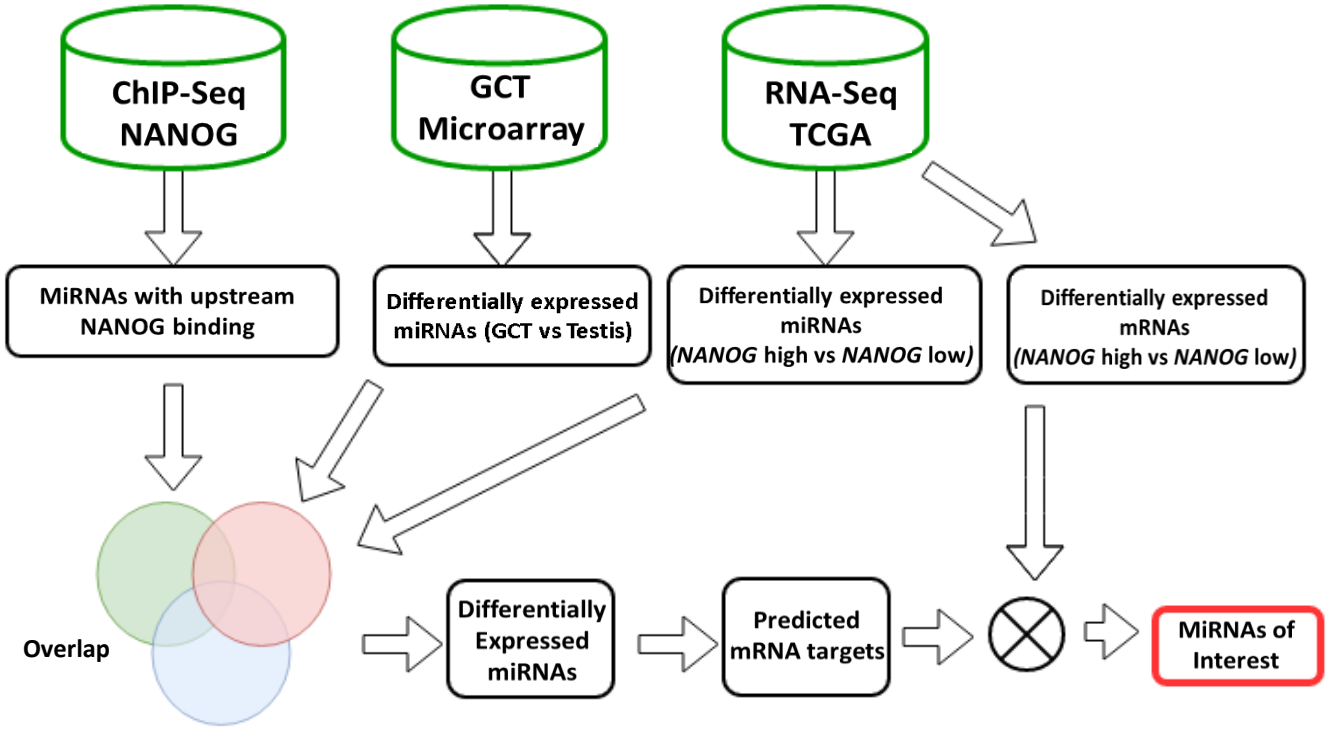


Figure 3.1 A flow chart depicting the bioinformatics analysis conducted in order to produce a list of miRNAs that may be regulated by NANOG and their predicted targets. Green cylinders represent data sources, black boxes represent processes, black circle with a cross represents a summing junction and the red box represents the end of the flow diagram.

3.2 Results

3.2.1 Analysis of ChIP-Seq data

In order to identify a list of miRNAs that may be regulated by NANOG, hESC ChIP-seq peak data (bed files) were downloaded for NANOG, OCT4 and SOX2 (Tsankov *et al.*, 2015). The NANOG ChIP-seq was first characterised for functional significance of cis-regulatory regions. The first analysis used GREAT (v3.0.0) which investigated the genomic contexts for NANOG binding (McLean *et al.*, 2010). A small percentage of NANOG peaks were positioned upstream of gene TSSs (2.6%) whereas 37% of NANOG peaks were contained within genomic introns (Figure 3.2A). Following this, motif analysis was conducted using NANOG binding sites upstream of mRNA or miRNA TSSs. This analysis identified that the NANOG motif at mRNA TSSs (CCAATCA) was different to the motif at miRNA TSSs (CCTCCC) (Figure 3.2B). Following this, NANOG binding was analysed based on the distance from miRNA TSSs; 37 microRNAs are bound by NANOG within a putative promoter (as define being 2kb upstream the TSS) (Figure 3.2C). Following this, OCT4 and SOX2 binding was analysed based on the distance from miRNA TSSs. A list of miRNAs with transcription factor binding was produced for NANOG, OCT4 and SOX2; these were then overlapped to produce a Venn diagram (Figure 3.2D). A total of 18 of 305 microRNAs had evidence of all 3 core pluripotency factors bound (5.9%); while those NANOG co-bound with by OCT4 or SOX2 were 28 miRNAs (9.2%) and 7 miRNAs, respectively (Figure 3.2D; Table S3.1).

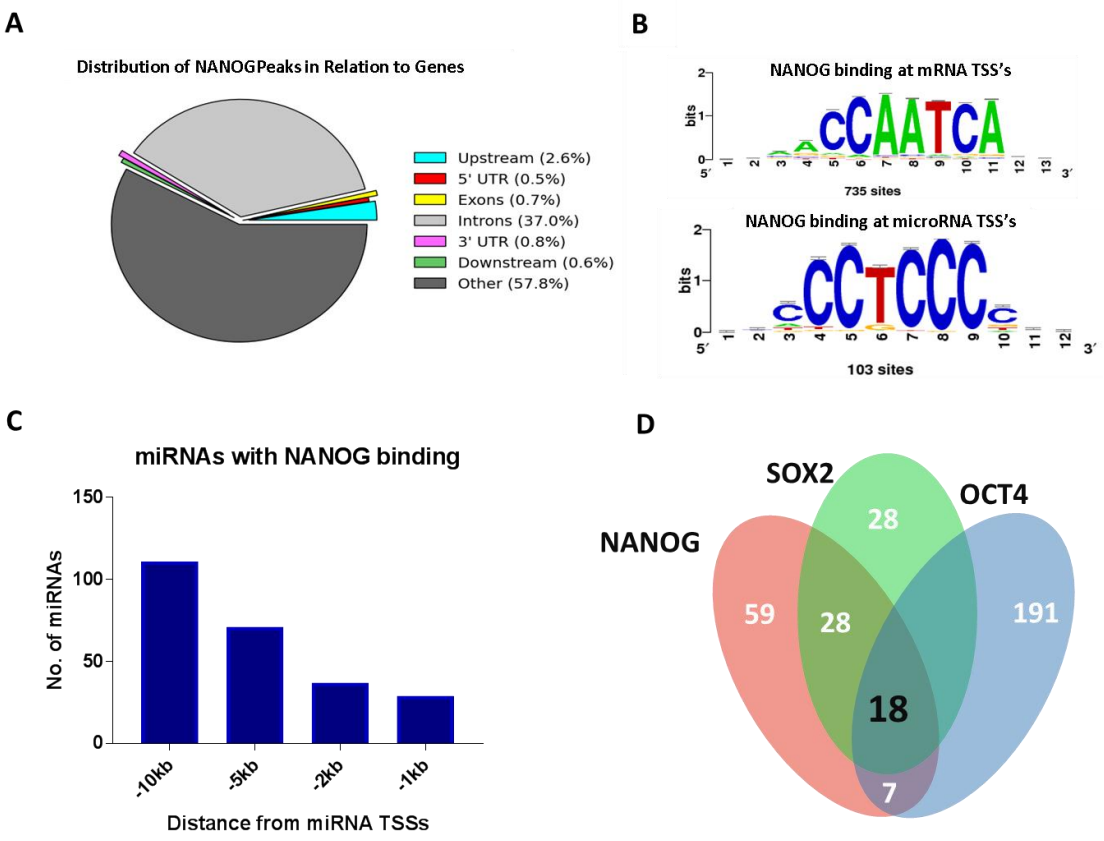


Figure 3.2 (A) a pie chart showing the distribution of NANOG peaks in relation to genes (coding and non-coding; GRCh37/hg19). (B) NANOG motif analysis for mRNAs (top) and microRNAs (bottom). (C) A bar plot showing the number of miRNAs with NANOG binding upstream from miRNA TSSs (10kb-1kb). (D) A Venn diagram overlapping binding sites for NANOG, OCT4 and SOX2 upstream of microRNA TSSs (10kb).

3.2.2 Analysis of Microarray Data

Differential miRNA expression was determined using available microarray data, comparing the groups TGCTs and normal testes. Using the Murray dataset, a total of 166 microRNAs were differentially expressed in TGCTs compared to normal testes; 92 miRNAs were higher and 74 lower in TGCTs when compared to the normal testes. Following this, the Novotny *et al* 2012 dataset analysis identified 168 miRNAs differentially expressed in TGCTs compared to normal testes; 87 miRNAs were higher and 81 miRNAs lower. A total of 51 miRNAs (39.8%) were shown to be highly expressed in TGCTs in both studies (Figure 3.3A), whereas 35 miRNAs were shown to be lowly expressed (Figure 3.3B); Table 3.1 outlines the full list of differentially expressed miRNAs and their expression.

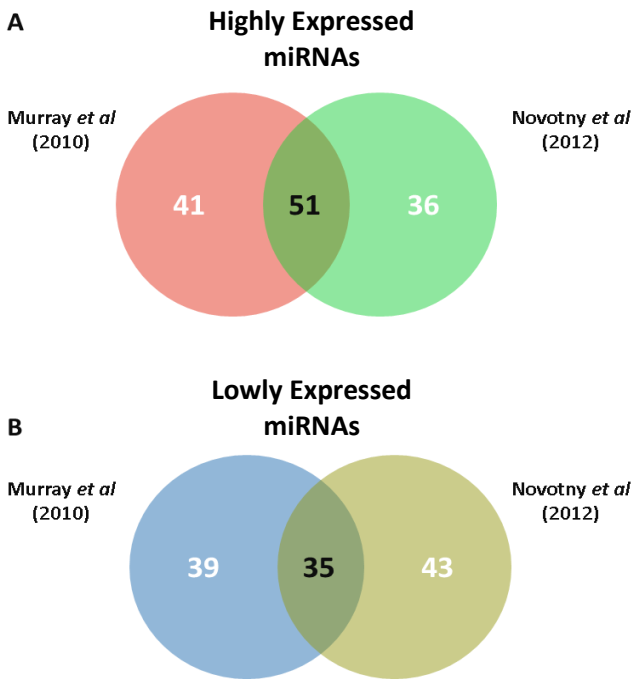


Figure 3.3 microRNAs common in both studies GSE18155 (Murray *et al* 2010) and GSE31824 (Novotny *et al* 2012) that are upregulated (A) or downregulated (B) in TGCTs.

Table 3.1 A heatmap of microRNAs differentially expressed in TGCTs compared to the testis control; green represent miRNAs that are lowly expressed and red which are highly expressed in TGCTs compared to the testis control.

	GSE31824		GSE18155	
microRNAs	**SEM**	**EC**	**SEM**	**EC**
hsa-let-7d-3p				
hsa-let-7f-2-3p				
hsa-let-7g-3p				
hsa-miR-105-5p				
hsa-miR-10a-3p				
hsa-miR-130b-3p				
hsa-miR-130b-5p				
hsa-miR-141-3p				
hsa-miR-142-5p				
hsa-miR-154-3p				
hsa-miR-15a-3p				
hsa-miR-17-5p				
hsa-miR-182-5p				
hsa-miR-183-3p				
hsa-miR-183-5p				
hsa-miR-18b-5p				
hsa-miR-198				
hsa-miR-211-5p				
hsa-miR-212-3p				
hsa-miR-216b-5p				
hsa-miR-29b-1-5p				
hsa-miR-302c-3p				
hsa-miR-302d-3p				
hsa-miR-323b-5p				
hsa-miR-330-5p				
hsa-miR-335-3p				
hsa-miR-338-5p				
hsa-miR-339-5p				
hsa-miR-33a-3p				
hsa-miR-362-5p				
hsa-miR-371a-5p				
hsa-miR-373-5p				
hsa-miR-380-3p				
hsa-miR-422a				
hsa-miR-433-3p				

hsa-miR-449b-5p				
hsa-miR-454-3p				
hsa-miR-454-5p				
hsa-miR-485-3p				
hsa-miR-486-3p				
hsa-miR-487a-3p				
hsa-miR-487b-3p				
hsa-miR-490-3p				
hsa-miR-490-5p				
hsa-miR-491-3p				
hsa-miR-491-5p				
hsa-miR-500a-3p				
hsa-miR-501-3p				
hsa-miR-502-3p				
hsa-miR-506-3p				
hsa-miR-508-3p				
hsa-miR-512-3p				
hsa-miR-514a-3p				
hsa-miR-515-3p				
hsa-miR-517-5p				
hsa-miR-518a-3p				
hsa-miR-518d-3p				
hsa-miR-518e-3p				
hsa-miR-518f-5p				
hsa-miR-520a-5p				
hsa-miR-520e				
hsa-miR-523-3p				
hsa-miR-524-5p				
hsa-miR-525-3p				
hsa-miR-526b-3p				
hsa-miR-542-3p				
hsa-miR-542-5p				
hsa-miR-604				
hsa-miR-605-5p				
hsa-miR-617				
hsa-miR-631				
hsa-miR-654-5p				
hsa-miR-657				
hsa-miR-659-3p				
hsa-miR-660-5p				
hsa-miR-767-3p				
hsa-miR-769-5p				
hsa-miR-873-5p				
hsa-miR-887-3p				

hsa-miR-891a-5p				
hsa-miR-891b				
hsa-miR-934				
hsa-miR-9-3p				
hsa-miR-9-5p				
hsa-miR-99a-3p				
hsa-miR-99a-5p				

3.2.3 Analysis of The Cancer Genome Atlas TGCT dataset

3.2.3.1 Characterisation of TGCT Subtypes

TGCT TCGA samples were analysed to understand how the patient subtype data clustered based on miRNAs and transcripts expression. Unsupervised hierarchical clustering of global mRNA gene expression indicated that the majority of seminoma (SEM) samples clustered together, whereas the embryonal carcinoma (EC), yolk sack tumours (YST) and teratoma (Tera) samples clustered together in a separate branch (Figure 3.4A). Consistent with this, there was a higher correlation of mRNA gene expression in samples which formed clusters with each other when compared between clusters. The heatmap

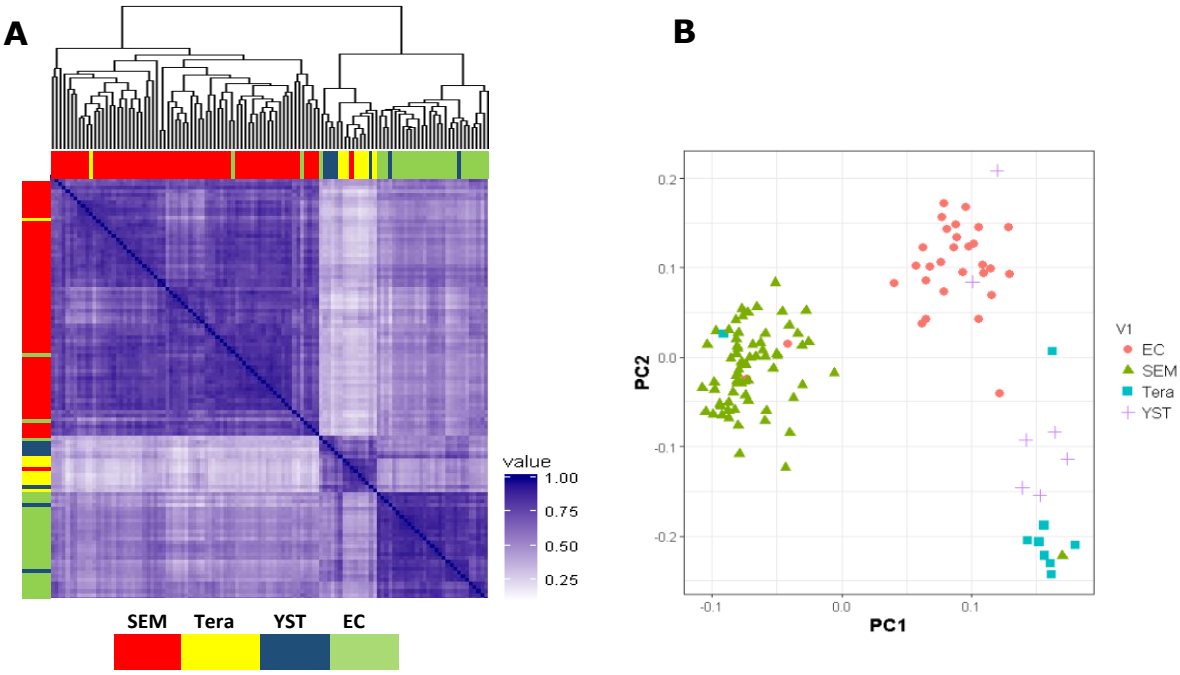


Figure 3.4 (A) Pearsons correlation heatmap of mRNA gene expression (log_2 (normalised read counts)) in TGCT samples with unsupervised hierarchical clustering. The blue colour key indicates the correlation coefficient within the heatmap whereas the TGCTs samples are colour coded as follows; SEM (red; n=68), Tera (yellow; n=9), YST (Blue; n=7) and EC (green; n=30). (B) Principle Component Analysis (PCA) of mRNA gene expression (log_2 (normalised read counts)), PC1 versus PC2, with the sample key indicated on the right hand side of the plot.

showed 3 groups of patients which were SEM, EC and MIX (tumour containing a mixture of YST and Tera). This was further supported by the principle component analysis (PCA) which indicated the same three groups (Figure 3.4B).

Unsupervised hierarchical clustering using miRNA expression also separated the TGCT samples into 2 main clades, however unlike the mRNA expression data SEM clustered with YST, whereas Tera and EC clustered in its own individual branch (Figure 3.5A). Supporting this, the heatmap indicated there was a higher correlation of miRNA gene expression in samples which formed clusters with each other when compared between clusters, Figure 3.5A. In addition to this, PCA analysis also indicated a distinction in expression profile between SEM, EC and Tera samples (Figure 3.5B).

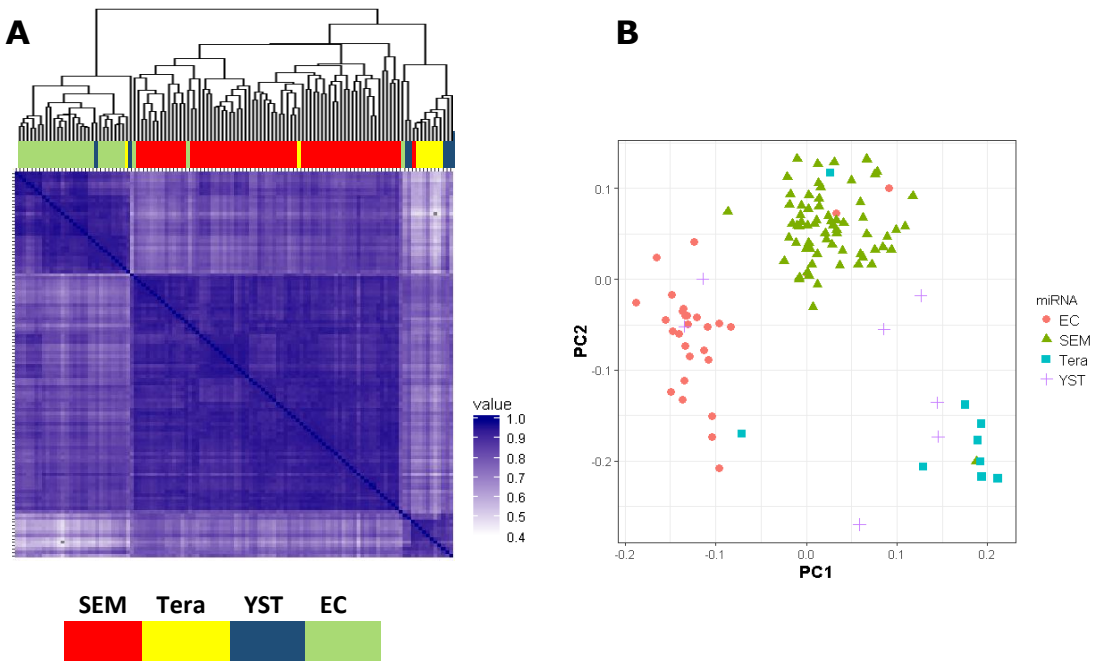


Figure 3.5 (A) Persons correlation heatmap of miRNA gene expression (log_2 (normalised read counts) in TGCT samples with unsupervised hierarchical clustering. The blue colour key indicates the correlation coefficient within the heatmap whereas the TGCTs samples are colour coded as follows; SEM (red; n=68), Tera (yellow; n=9), YST (Blue; n=7) and EC (green; n=30). (B) Principle Component Analysis (PCA) of microRNA gene expression (log_2 (normalised read counts)), PC1 versus PC2, with the sample key indicated on the right hand side of the plot.

Expression of key pluripotency genes was also analysed across TGCT subtypes. *NANOG* expression was shown to have higher expression in EC when compared to other TGCT subtypes ($p<0.05$), whereas SEM showed higher expression when compared to Tera and YST (Figure 3.6A). SEM had significantly lower expression for *SOX2* when compared to EC but more highly expressed when compared to Tera groups (Figure 3.6B). EC exhibited higher *OCT4* expression when compared to Tera and YST (Figure 3.6B). *SOX2* showed the lowest expression in SEM when compared to all other TGCT subtypes, whereas EC exhibited the highest expression of *SOX2* (Figure 3.6C). In contrast, *SOX17* expression showed highest expression in SEM samples and was significantly higher than EC and Tera TGCT subtypes (Figure 3.6D). YST *SOX17* expression had higher expression in YST compared to EC (Figure 3.6D).

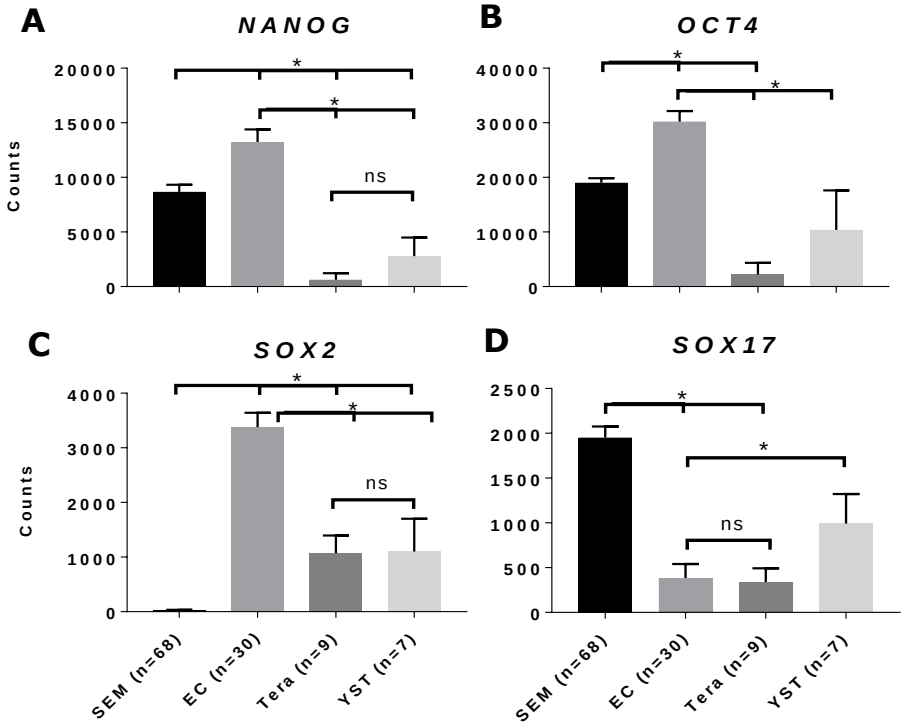


Figure 3.6 shows the expression profiles of *NANOG* (A), *OCT4* (B), *SOX2* (C) and *SOX17* (D) for TGCT samples from the TCGA; bars represent transcript counts with standard error of the mean while * indicates a significant difference between groups ($p<0.05$). EdgeR was used to calculate statistical significance.

3.2.3.2 Characterisation of groups based on Nanog Expression

Following the characterisation of TCGA samples, they were next separated into groups based on *NANOG* expression in attempt to identify miRNAs that correlate with *NANOG* expression. Samples were separated into tertile groups based on *NANOG* expression; *NANOG* high, *NANOG* medium (med) and *NANOG* low. Across samples the Max count reading for *NANOG* was 33,114 whereas its lowest count reading was 2; within groups the *NANOG* mean count was 14426, 7184 and 1432 for *NANOG* high, *NANOG* med and *NANOG* low, respectively (Table 3.2).

A total of 420 miRNAs were differentially expressed across groups (Figure 3.7). *NANOG* high groups were shown to highly express pluripotency-related miRNAs (*miR-302*b, *miR-302*c and *miR-302*d) whereas these miRNAs were expressed lowly in *NANOG* low groups. Unsupervised hierarchical clustering using miRNAs differentially expressed in *NANOG* expression groups (high, med and low) showed some heterogeneity (Figure 3.7); this may be a result of TGCT subtype heterogeneity as samples appeared to cluster into TGCT subtypes when looking at miRNA expression (Figure 3.5A).

Table 3.2 describes how TCGA TGCT samples were separated based on NANOG expression.

***NANOG* Groups**	**Tertile**	**Number of Samples**	***NANOG* Counts**	***NANOG* Count Mean**
***NANOG* high**	3/3	52	>9524	14426
***NANOG* medium**	2/3	51	4927-9524	7184
***NANOG* low**	1/3	53	0-4927	1432

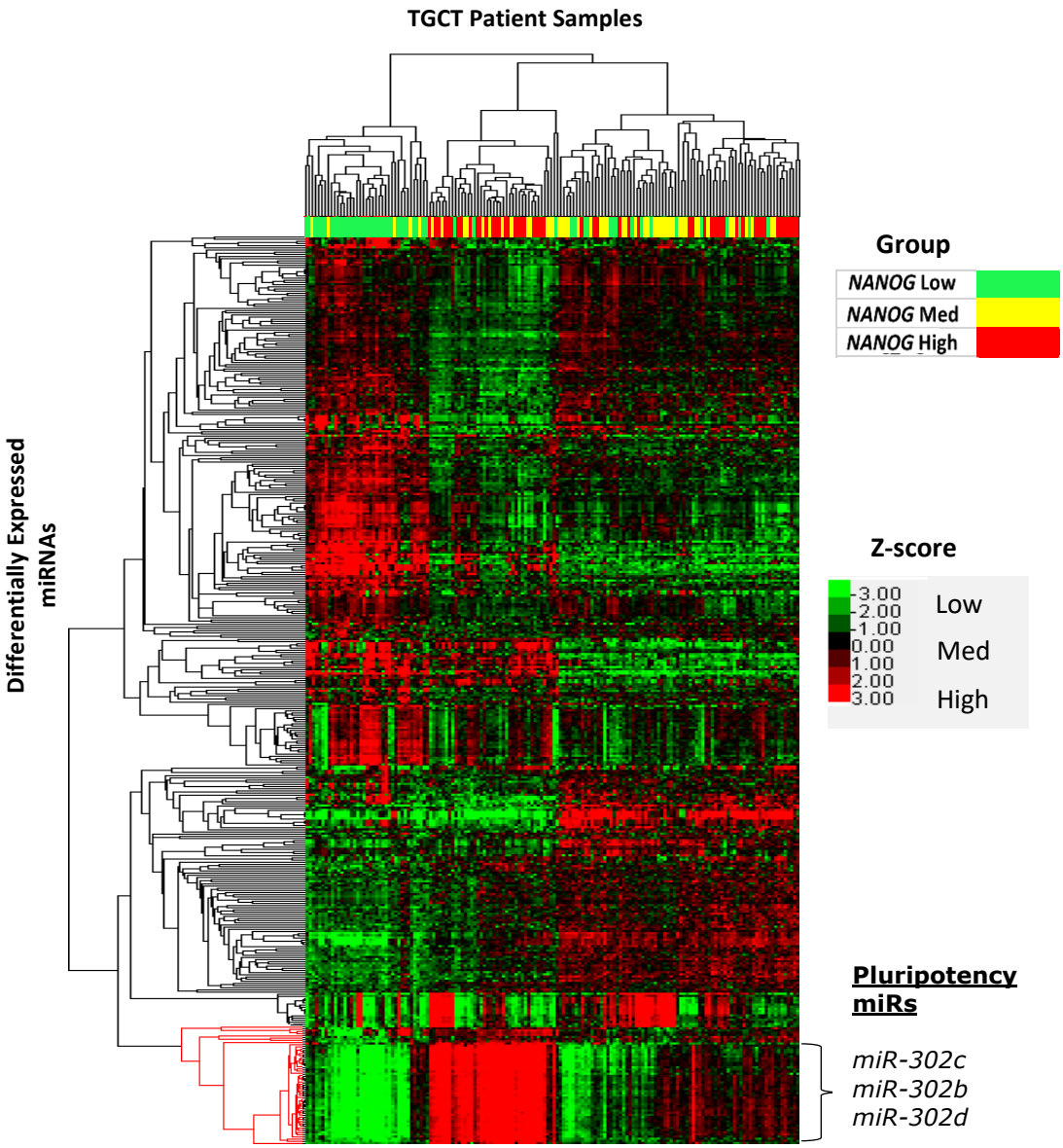


Figure 3.7 shows a heatmap of 420 differentially expressed microRNAs for *NANOG* High, *NANOG* Med and *NANOG* Low TGCT groups. The top dendrogram shows TGCT samples subject to unsupervised hierarchical clustering; the vertical dendrogram shows differentially expressed miRNAs. Patient samples are colour coded as follows; *NANOG* high (red; n=52), *NANOG* med (yellow; n=51) and *NANOG* low (green; n=53). EdgeR was used to calculate statistical significance for differentially expressed genes.

3.2.4 Overlapping of Bioinformatics Analyses

In order to form a comprehensive list of NANOG regulated miRNAs, the bioinformatic analyses described previously (ChIP-Seq, microarray and TCGA RNA-seq data) were combined. This analysis included: (1) miRNAs which had NANOG binding 10kb upstream of TSSs (n=111), (2) differentially expressed miRNAs by microarray TGCT vs samples (n=86) and (3) differentially expressed miRNAs in *NANOG* High, Med and Low groups (n=401). This analysis identified a total of 9 microRNAs that were differentially expressed across all three analyses (Figure 3.8 A and B, respectively). There were 4 miRNAs with NANOG binding <2kb upstream of TSSs (*miR-512*, *miR-183*, *miR-9* and *miR-887*); of these *miR-512* and *miR-183* had other miRNAs in close proximity suggesting that they may be part of a co-regulated cluster (Table 3.3). Proximal promoters are generally ~2kb upstream of TSSs (Kim *et al.*, 2005), therefore to make the analysis more stringent, miRNAs with NANOG binding within 2kb of TSSs were chosen. *MiR-512* and *miR-183* were shown to be highly expressed in TGCT samples with high *NANOG* expression, whereas *miR-9* and *miR-887* showed low expression in samples (Figure 3.9). This is consistent with NANOG binding directly and regulating *miR-9* and *miR-887* loci. All four chosen miRNAs contained binding for NANOG, SOX2 and OCT4 at similar loci further verifying the NANOG ChIP-seq as a control regulatory module (Figures 3.10-13) Both *miR-9* and *miR-183* (Figure 3.10 and 3.13, respectively) showed bivalent enrichment for H3K27me3 and H3K4me3 marks indicating these regions were transcriptionally repressed but poised for transcription.

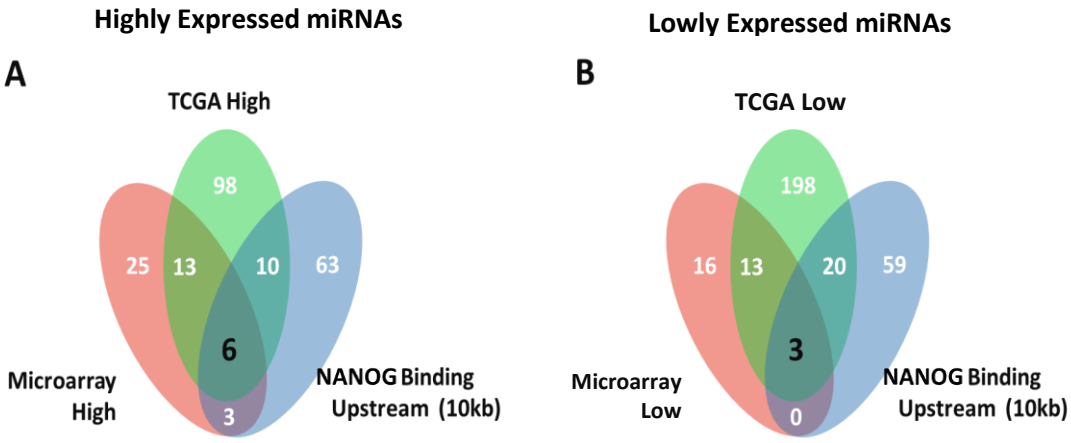


Figure 3.8 The overlap for differentially expressed microRNAs in given groups (TCGA RNA-Seq, microarray and NANOG ChIP-seq data) identifying miRNAs with NANOG binding 10kb upstream of TSSs.

Table 3.3 shows descriptions for the microRNAs identified after bioinformatics overlaps. UP= upregulation; DR=downregulation.

microRNA	Predicted regulation by Nanog	TSS distance from Nanog Binding	Closest miRNA to Nanog Binding	MiRNAs chosen for further Analysis?
miR-512	UR	<1kb	Yes	Yes
miR-183	UR	1 kb	Yes	Yes
mir-520e	UR	6.7kb	No	No
mir-182	UR	5.7kb	No	No
mir-515	UR	9.7kb	No	No
mir-373	UR	5kb	No	No
miR-9	DR	<1kb	Yes	Yes
miR-887	DR	1.5kb	Yes	Yes
mir-216b	DR	8.3kb	Yes	No

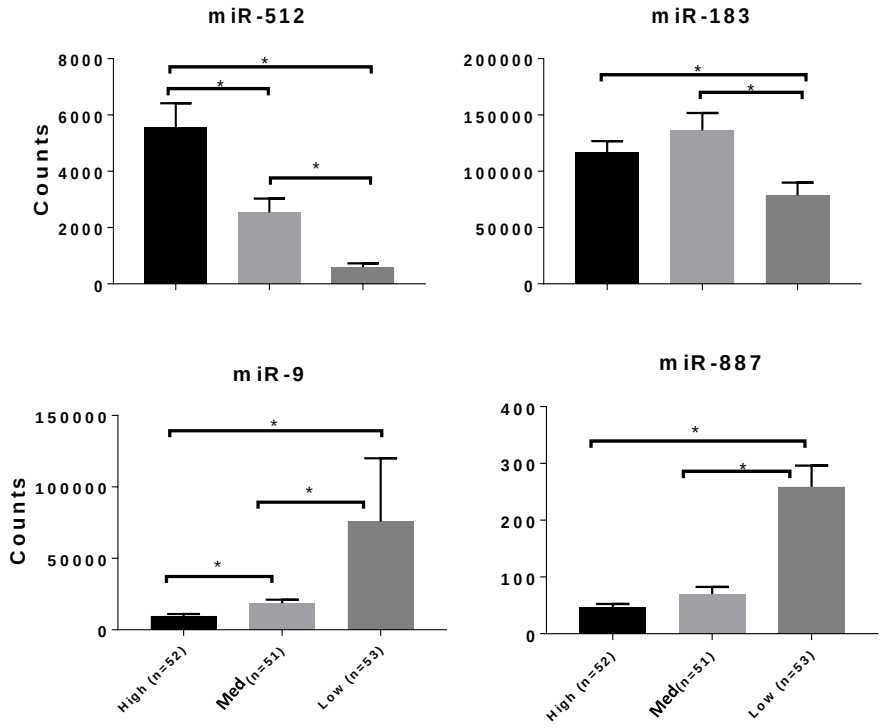


Figure 3.9 shows the expression of chosen miRNAs that have NANOG binding 2kb upstream of TSSs in TCGA that have High, Medium (Med) or Low *NANOG* expression; * denotes differential expression greater than 2 fold ($p<0.05$). Error bars indicate the mean ± SEM for samples. EdgeR was used to calculate statistical significance (section 2.2.6).

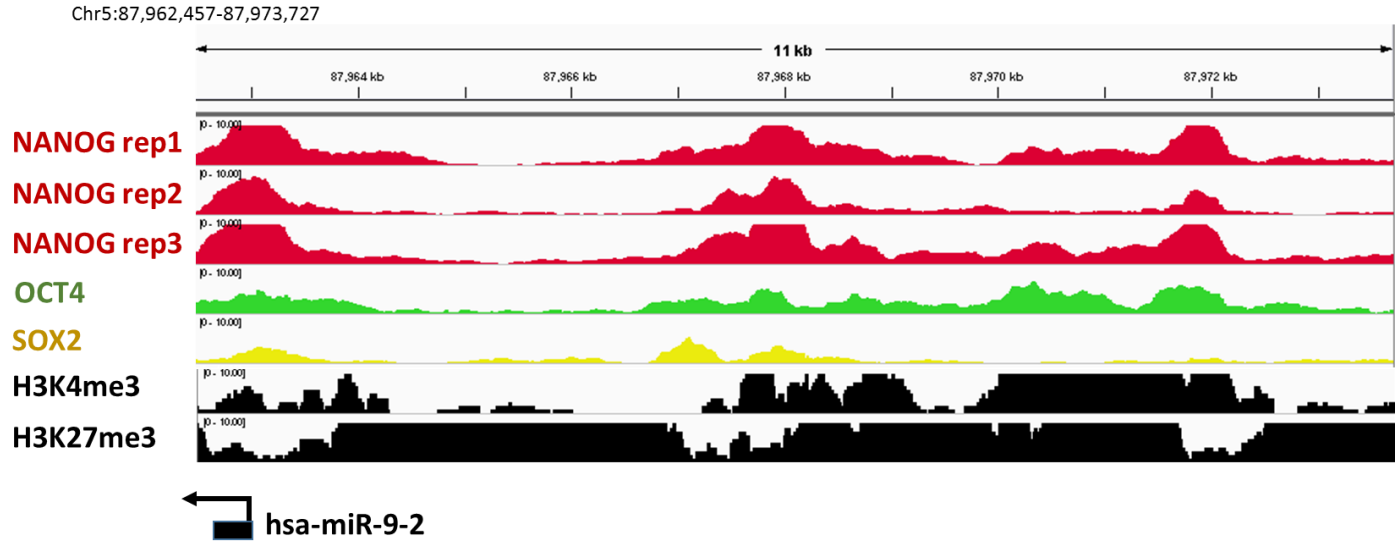


Figure 3.10 shows the miR-9-2 binding profile for NANOG (red; n=3), OCT4 (green; n=1), SOX2 (yellow; n=1), H3K4me3 (top black; n=1) and H3K27me3 (black bottom; n=1) using ChIP-seq from hESCs. All ChIP-seq signal intensities were set to 0-10.

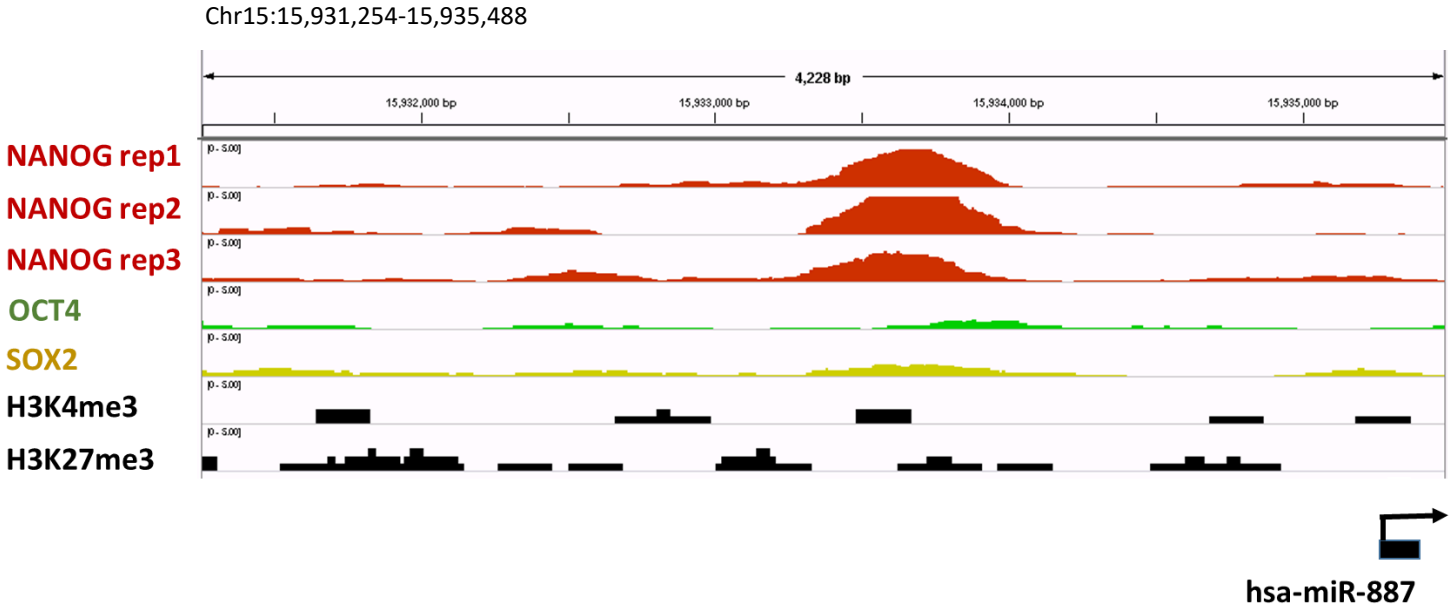


Figure 3.11 shows the miR-887 binding profile for NANOG (red; n=3), OCT4 (green; n=1), SOX2 (yellow; n=1), H3K4me3 (top black; n=1) and H3K27me3 (black bottom; n=1) using ChIP-seq from hESCs. All ChIP-seq signal intensities were set to 0-5.

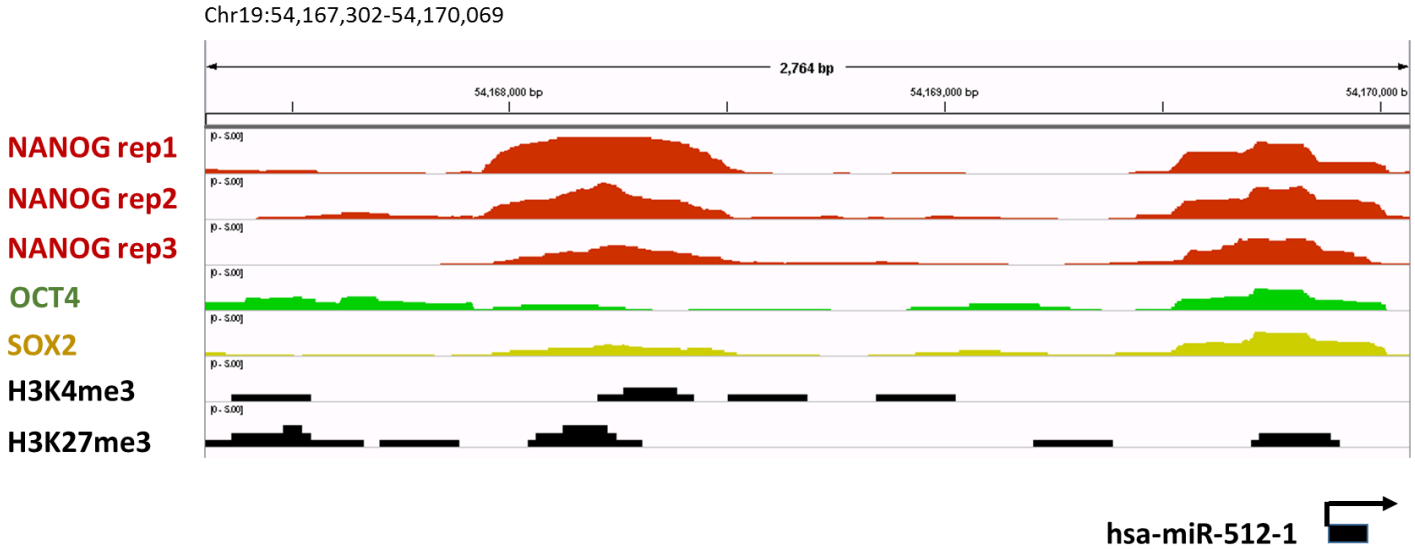


Figure 3.12 shows the miR-512-1 binding profile for NANOG (red; n=3), OCT4 (green; n=1), SOX2 (yellow; n=1), H3K4me3 (top black; n=1) and H3K27me3 (black bottom; n=1) using ChIP-seq from hESCs. All ChIP-seq profiles signal intensities were set to 0-5.

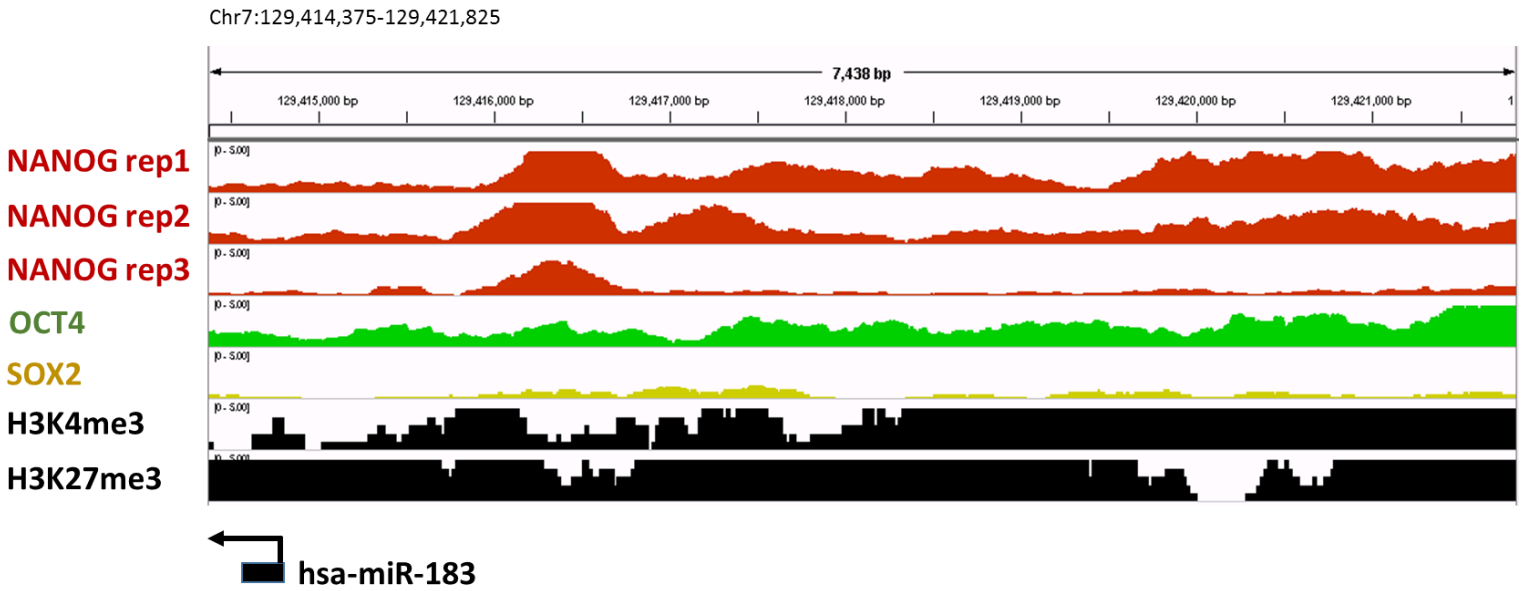


Figure 3.13 shows the miR-183 binding profile for NANOG (red; n=3), OCT4 (green; n=1), SOX2 (yellow; n=1), H3K4me3 (top black; n=1) and H3K27me3 (black bottom; n=1) using ChIP-seq from hESCs. All ChIP-seq signal intensities were set to 0-5.

3.2.5 Analysis of TCGA clinical data

The TCGA includes patient clinical information from which the samples were obtained. The data can be utilised to investigate how gene expression is related to disease progression and patient outcome. Therefore, data was categorised into groups based on the clinical information given in order to elucidate the clinical significance of NANOG and the miRNAs identified using the following criteria: TGCT subtype, presence of carcinoma *in situ* (CIS), stage of cancer and family history of TGCT. *MiR-512* was highly expressed in EC TGCT subtypes when compared to SEM and Tera ($p<0.05$), whereas *miR-183* showed no significant differences among groups (Figure 3.14). *MiR-9* showed higher expression in SEM groups when compared to EC and YST, whereas *miR-887* showed no significant differences among groups (Figure 3.14).

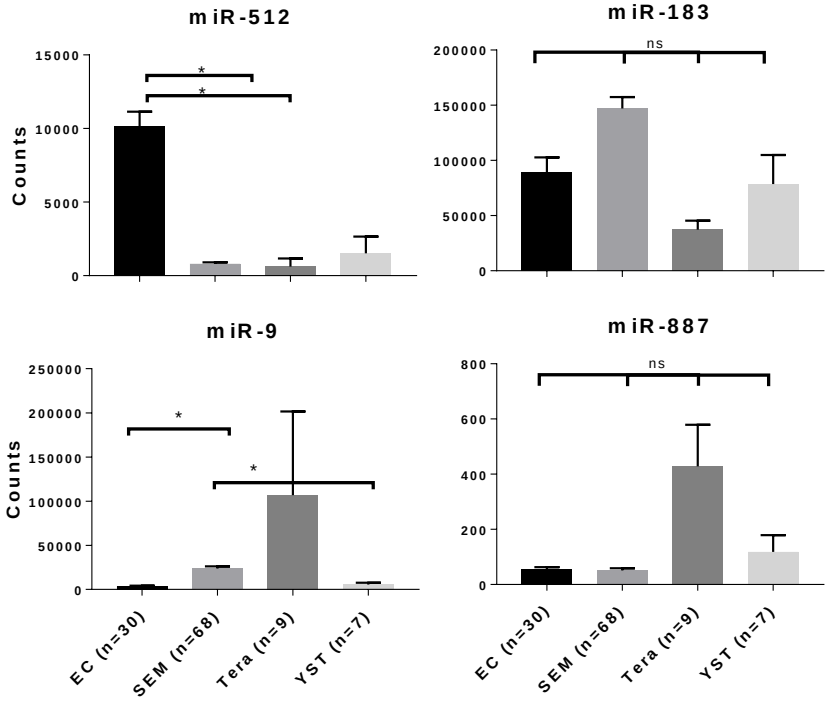


Figure 3.14 Expression of chosen miRNAs in TCGA in patient samples with different TGCT subtypes; * denotes differential expression greater than 2 fold ($p<0.05$). Error bars indicate the mean ± SEM for samples. EdgeR was used to calculate statistical significance (section 2.2.6).

Following this, patient samples were grouped based on their presence or absence of Carcinoma *in situ* (CIS). *NANOG* expression did not differ significantly between CIS groups (Figure 3.15). *MiR-512* showed higher expression in patients with CIS compared to those without while *miR-183* showed no differences between groups. *MiR-9* had higher expression in TGCT samples that had no CIS compared to those with a CIS present, whereas *miR-887* showed no significant difference between groups (Figure 3.15).

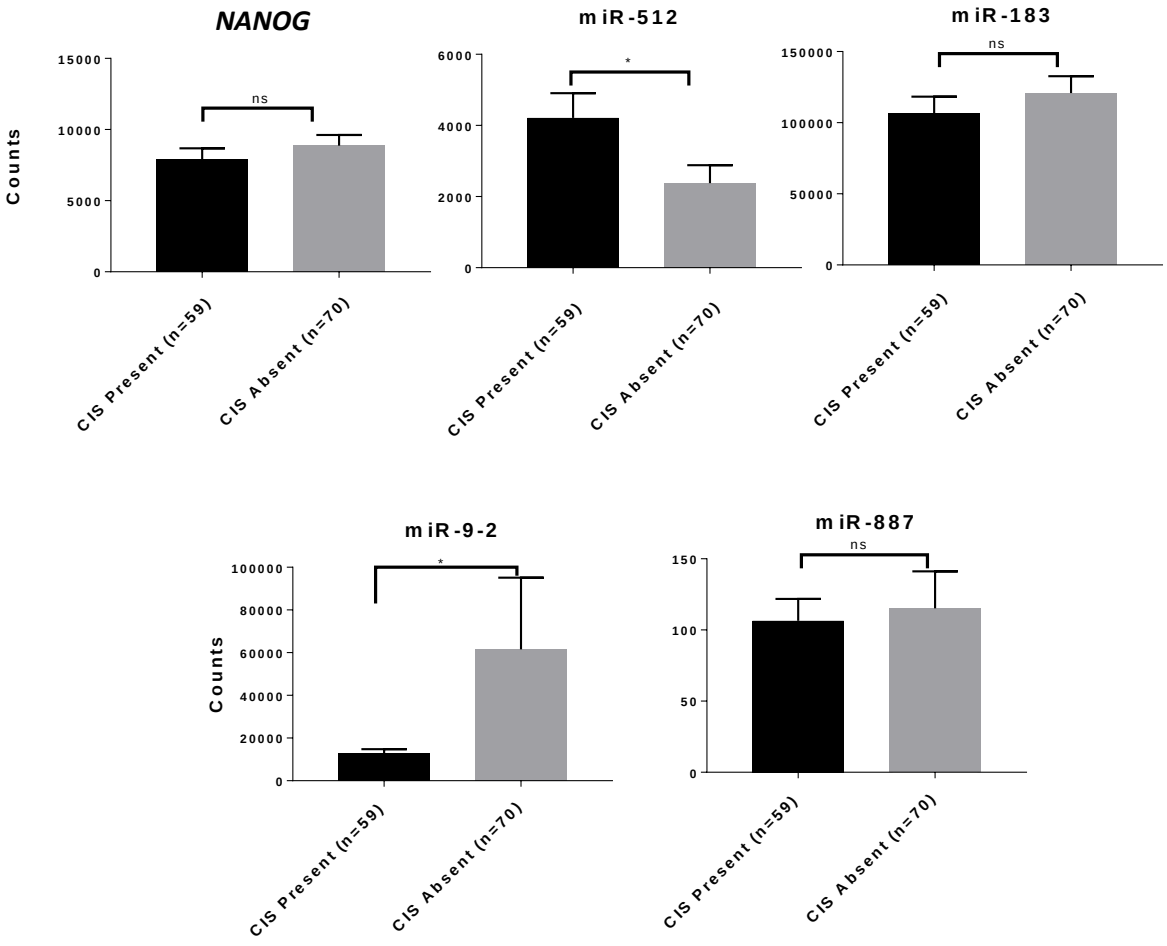


Figure 3.15 The expression of *NANOG* and chosen miRNAs in the TCGA in patients which had carcinoma *in situ* presence/ absence; * denotes differential expression greater than 2 fold ($p<0.05$). Error bars indicate the mean ± SEM for samples. EdgeR was used to calculate statistical significance (section 2.2.6).

When TGCT samples were grouped by cancer stage, *NANOG* expression did not differ significantly between groups (Figure 3.16). *MiR-512* showed a higher expression in stage I compared to stage II, whereas *miR-183* showed no significant change in expression across groups. *MiR-9* had higher expression in stage I compared to stages II and III, whereas *miR-887* showed no difference across cancer stages (Figure 3.16).

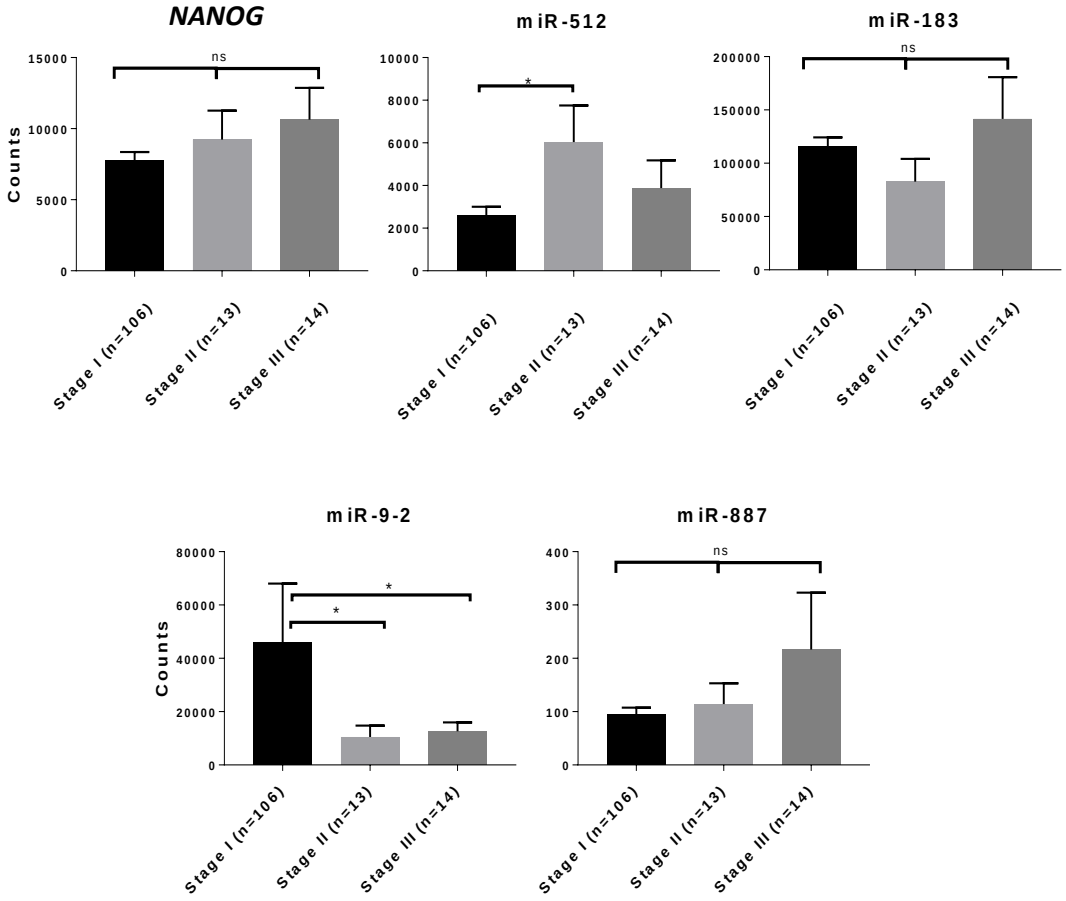


Figure 3.16 Expression of *NANOG* and detected miRNAs in the TCGA in patients samples which were at different stages of TGCT progression; * denotes differential expression greater than 2 fold ($p<0.05$). Error bars indicate the mean ± SEM for samples. EdgeR was used to calculate statistical significance (section 2.2.6).

In addition to this, expression of NANOG, *miR-512*, *miR-183*, *miR-9* and *miR-887* were not significantly differentially expressed in TGCT patient samples that had a family history of TGCT as compared to those without (Figure 3.17).

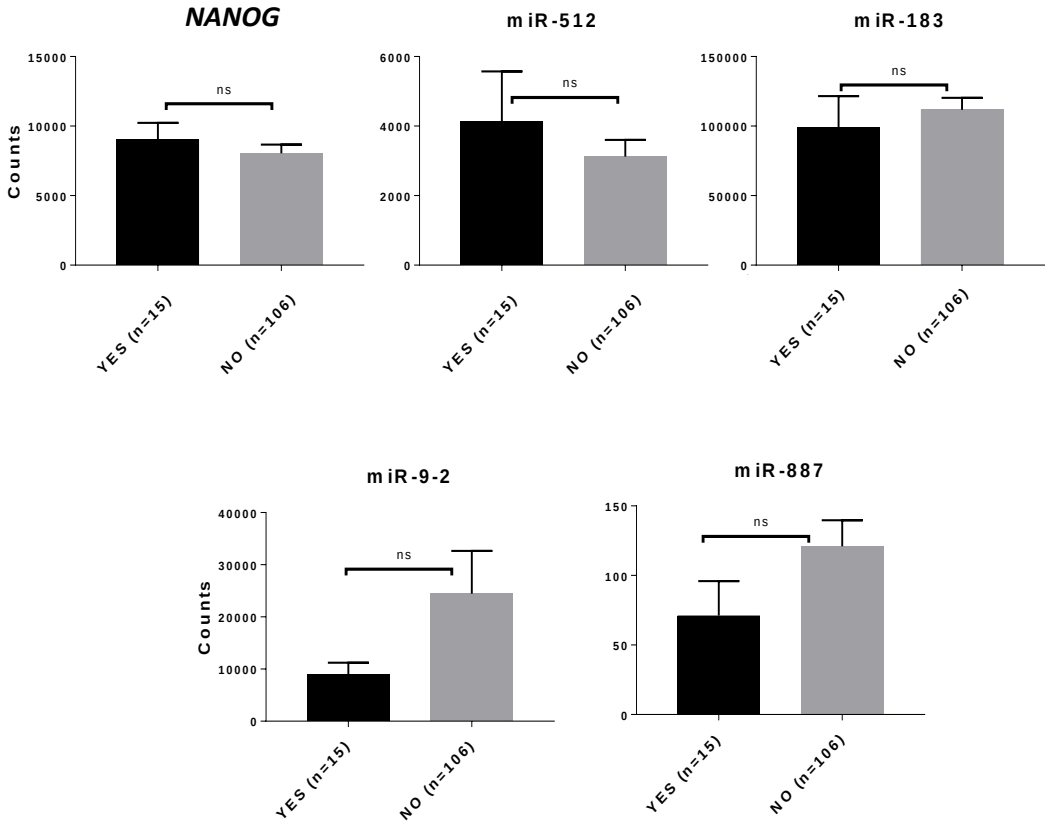


Figure 3.17 Expression of *NANOG* and chosen miRNAs in the TCGA in patients samples which had a history of TGCT in the family; * denotes differential expression greater than 2 fold ($p<0.05$). Error bars indicate the mean ± SEM for samples. EdgeR was used to calculate statistical significance (section 2.2.6).

3.3 Discussion

3.3.1 Identification of miRNAs potentially regulated by NANOG

The purpose of this chapter was to identify a set of NANOG regulated miRNAs important for the regulation of pluripotency and proliferation in TGCTs and PGC development. To do this, ChIP-seq data for NANOG binding in hESCs was used (Tsankov *et al.*, 2015). Analysis of the ChIP-seq data showed global NANOG binding at TSSs comprised only 2.6% of binding sites, whereas 37% of NANOG binding sites occurs within introns (Figure 3.2 A). This suggested a role for NANOG in regulating genes through binding introns or long-range regulation. Applications of HiC may address this in the future. Given that NANOG, OCT4 and SOX2 co-occupy promoter regions to control transcription, ChIP-seq for OCT4 and SOX2 were also included in the analysis. The initial analysis used the NANOG ChIP-seq in order to produce a list of NANOG-bound miRNA loci, as overlapping miRNAs bound by NANOG, OCT4 and SOX2 greatly reduced the possible NANOG-bound miRNAs from 111 to 18 (Figure 3.2 D). Initially, the analysis was restricted to miRNAs with NANOG binding <2kb upstream of TSSs; however, this criterion reduced the number of miRNAs from 111 to 37 and as such miRNAs with NANOG binding 10kb upstream of TSSs were also considered (Figure 3.2 C). This method maximised the number of candidate miRNAs as increasing the stringency at the start of the analysis could give too little or no results to work with.

Following this, gene expression datasets for TGCTs were analysed using available microarray and RNA-seq datasets. The microarray analysis used two different datasets which used different platforms (Palmer *et al.*, 2010; Novotny *et al.*, 2012). Different gene quantification platforms use different normalisation protocols, therefore instead of pooling the two datasets, the datasets were processed and analysed separately. This resulted in a list of differentially expressed miRNAs comparing TGCT to the testes for each of the microarray datasets. The list of differentially expressed miRNAs from each dataset was overlapped to yield a consensus of differentially expressed miRNAs in both datasets (Figure 3.3 A and B). A total of 168 miRNAs were differentially expressed in both Novotny and Murray datasets for TGCTs compared to the testes; 51 miRNAs were higher and 35 were lower relative to the testes (Figure 3.3; the full list of miRNAs are shown in Table 3.1). This, therefore, gave a list of miRNAs differentially expressed in TGCT versus a 'normal' tissue control. Using tumour versus control analysis indicated whether miRNAs could be involved in the development of TGCT but did not help to elucidate their relationship with NANOG. To address this, TGCT mRNA expression from the TCGA was next analysed in the context of *NANOG* expression.

The first analysis performed was unsupervised hierarchical clustering which clustered data based solely on the expression profile for TGCT samples. For both mRNA and miRNA datasets two main clusters were evident (Figure 3.4 A and Figure 3.5 A) and when this was integrated with the TCGA clinical data it appeared that the clusters were formed based on histological subtypes. For mRNA, the embryonal carcinoma (EC) group clustered more closely to that of

the teratoma (Tera) and yolk sac tumour (YST) samples, whereas the Seminoma samples (SEM) clustered in a group of its own (Figure 3.3A). However, the miRNA expression profile showed SEM, YST and Tera samples clustering together and EC samples clustering furthest in its own group. This data suggested that the mRNA expression profile from YST and Tera samples was more similar to EC samples, whereas for miRNA expression profiles YST and Tera have greater similarity to SEM samples. However, the small sample size for YST and Tera reduces the confidence in this distinction. Following this, a Pearson's correlation coefficient heatmap using ordered expression datasets was performed, while keeping the samples order from the cluster analysis (Figure 3.4A and Figure 3.5A). In both mRNA and miRNA data sets, TGCT samples had a higher correlation with those in the same histological subtype (Figure 3.4A and Figure 3.5A). The correlation heatmap for the mRNA dataset showed a higher level of correlation within groups when compared to miRNA thus suggesting that the mRNA expression profile may better define TGCT histological subtypes compared to the miRNA expression profile. For both expression datasets, the correlation heatmap separated samples into three main correlative groups; (1) SEM, (2) EC and (3) YST with Tera (Figure 3.4A and 3.5A). Principle component analysis (PCA) also showed that samples separated into these same three groups (Figure 3.3B and 3.4B).

TGCTs were next separated into histological subtypes and analysed for *NANOG, OCT4*, *SOX2* and *SOX17* expression (Figure 3.6). *NANOG*, and *OCT4* had significantly higher expression in SEM and EC samples when compared to Tera (Figure 3.6 A and B). Tera samples represent the differentiated counterpart of

TGCTs derived from EC cells and, therefore, Tera samples express lower levels of pluripotency factors (Palumbo *et al.*, 2002; Hart *et al.*, 2005). *SOX2* was most highly expressed in EC samples when compared to SEM, Tera and YST (Figure 3.6 C). This is consistent with previous studies showing that *SOX2* expression is higher in EC samples than SEM samples (De Jong *et al.*, 2008). De Jong also showed that *SOX17* is highly expressed in SEM samples and TCam-2, a seminoma derived cell line, whereas EC samples do not express *SOX17* (De Jong *et al.*, 2008). In agreement with this, the TCGA analysis shows *SOX17* has significantly higher expression in SEM when compared to EC (Figure 3.6 D).

After characterising the TCGA data for subtypes, the TGCT samples were stratified based on *NANOG* expression. In order to create groups, samples were grouped into tertiles based on *NANOG* high, *NANOG* medium and *NANOG* low groups (Table 3.2). *NANOG* was differentially expressed in *NANOG* high, medium and low groups (Figure S3.1). A total of 420 miRNAs were differentially expressed between *NANOG* high, medium and low groups (Figure 3.7). Unsupervised clustering for differentially expressed miRNAs using high, medium and low *NANOG* groups showed that miRNAs cluster into two main groups; (1) samples with low *NANOG* expression (2) and samples with a mixture of medium and high levels of *NANOG* expression (Figure 3.7). However, there was a large degree of heterogeneity in the clusters which may be a result of the samples clustering into TGCT subtypes, rather than by *NANOG* expression as shown previously in Figure 3.4 and Figure 3.5. This would explain the large proportion of samples which clustered in the *NANOG* high and medium cluster, as the number of SEM and EC samples exceeded the number of YST and Tera

samples (Figure 3.4). There was a group of highly expressed miRNAs in the *NANOG* high groups and, interestingly, this groups contains the *miR-302* cluster (Figure 3.7). This is in agreement with previous studies which show that NANOG, OCT4 and SOX2 bind to the *miR-302* promoter region to enhance its expression (Card *et al.*, 2008; Barroso-del Jesus, Lucena-Aguilar and Menendez, 2009). Therefore, from the data obtained it could be extrapolated that *miR-302* is also regulated by NANOG in TGCTs.

Following the analysis of these datasets there were three separate lists of miRNAs were regulated; (1) miRNAs with NANOG binding 10kb upstream of TSSs (2) miRNAs differentially expressed in TGCT vs Testes samples from the microarray analysis and (3) miRNAs differentially expressed in *NANOG* high, medium and low groups from TCGA RNA-seq data. The three datasets were then overlapped yielded a consensus list of miRNAs, in attempt to identify miRNAs that were transcriptionally regulated by NANOG. The overlap analysis identified nine miRNAs in total; six upregulated and three downregulated (Figure 3.8 A and B). In order to increase stringency, only miRNAs with NANOG binding within the <2kb region upstream of TSSs were included. This increased stringency resulting in two miRNAs which were highly expressed in TGCT groups (*miR-512* and *miR-183*) and two lowly expressed (*miR-9* and *miR-887*) (Table 3.3; Figure 3.9). All four miRNAs identified also contained enrichment for OCT4 and SOX2 upstream of TSSs at similar sites to NANOG (Figure 3.10-3.13). It is known that NANOG, OCT4 and SOX2 co-localise together to regulate gene expression (Rodda *et al*., 2005), therefore verifying the NANOG ChIP-seq data sets. In addition to this, both *miR-183* and *miR-9* showed H3K4me3 and H3K27me3 marks within the putative promoter regions (Figure

3.10 and 3.13, respectively). The bivalency of the two epigenetic marks suggests that genes, and thus these miRNAs, are poised for transcription but are not transcribed (Bernstein *et al*., 2006). This is consistent with *miR-9* given that it is lowly expressed in EC and SEM samples (Figure 3.10) but is not in agreement with *miR-183* which is highly expressed in EC cells (Figure 3.14). For *miR-183* this may be a discrepancy between ESC and EC cell types and therefore may contain differences in epigenetic marks. This may also be true for *miR-887* and *miR-512* which are shown to have little enrichment for H3K4me3 or H3K27me3 (Figure 3.11 and 3.12). In addition, this analysis has comprehensively included all epigenetic marks such as H3K9me3 or H3K36me3 which are known to mark repressed and active genes, respectively (Martens *et al*., 2005; Bannister *et al*., 2003).

3.3.2 Conclusion

The goal of this chapter was to utilise TGCT expression datasets (microarray and RNA-seq (TCGA)) and NANOG ChIP-seq to produce a list of miRNAs that may be regulated by NANOG. The microarray analyses produced a list of miRNAs that were differentially expressed in TGCTs compared to a testes control (Figure 3.2). Therefore, this analysis provided miRNAs which had high or low expression in the tumours and may support oncogenic or tumour suppressive properties in TGCTs, respectively. In addition to this, samples stratified by *NANOG* expression produced a list of miRNAs that were differentially expressed in groups with high, medium or low expression of *NANOG* (Figure 3.9). NANOG ChIP-seq was also included in the analysis which allowed the identification of miRNA that contained NANOG binding upstream

of their TSSs (Figure 3.1). Overlapping of all three datasets resulted in a list of 4 NANOG bound miRNAs of interest; *miR-512* and *miR-183* were up-regulated in TGCTs, whereas *miR-887* and *miR-9* were down-regulated in TGCTs. Following this, TCGA RNA-seq data was stratified based on clinical information for the patient samples from which no defined conclusions were made. Now that a list of miRNAs had been produced using available bioinformatics datasets, the next chapter focused on validating these results.

3.4 Supplementary

Table S3.1 miRNAs with upstream binding for NANOG, OCT4 or SOX2 (10kb upstream of TSSs).

NANOG Only		NANOG, SOX2 and OCT4	NANOG and SOX2	NANOG and OCT4
ACA26	*mir-4269*	*ACA27*	*mir-216b*	*ACA6*
HBII-135	*mir-4289*	*mir-124-2*	*mir-4288*	*E2*
HBII-166	*mir-4290*	*mir-1323*	*mir-515-1*	*HBII-296A*
mir-106a	*mir-4325*	*mir-135b*	*mir-548f-2*	*HBII-296B*
mir-1182	*mir-548f-3*	*mir-21*	*mir-548h-2*	*mir-1203*
mir-1184-1	*mir-548p*	*miR-302a*	*U58A*	*mir-1227*
mir-1202	*mir-551b*	*miR-302b*	*U58B*	*mir-1244-1*
mir-1226	*mir-568*	*miR-302c*		*mir-1275*
mir-125b-2	*mir-569*	*miR-302d*		*mir-148a*
mir-1265	*mir-577*	*mir-3119-2*		*miR-183*
mir-1269	*mir-578*	*mir-367*		*mir-1915*
mir-128-2	*mir-582*	*mir-498*		*mir-2276*
mir-182	*mir-620*	*miR-512*		*mir-3175*
mir-18b	*mir-675*	*miR-512*		*mir-3191*
mir-193a	*mir-720*	*mir-520e*		*mir-3193*
mir-19b-2	*miR-887*	*miR-9*		*miR-371*
mir-20b	*miR-92a-2*	*miR-9-3*		*mir-372*
mir-2114	*SNORA38B*	*U102*		*mir-373*
mir-219-2	*U13*			*mir-4258*
mir-22	*U3*			*mir-4313*
mir-26a-1	*U3-2*			*mir-550-1*
mir-301a	*U3-3*			*mir-607*
miR-302f	*U3-4*			*mir-634*
mir-30a	*U37*			*mir-885*
mir-3131	*U49A*			*miR-96*
mir-3143	*U49B*			*snR39B*
mir-3169	*U8*			*U3-2B*
mir-320c-1	*U94*			*U70D*
mir-326	*U97*			
mir-363				

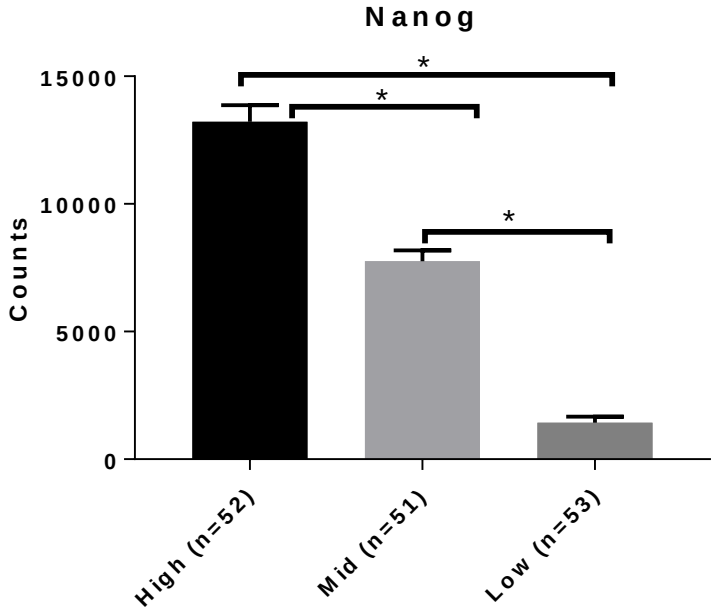


Figure S3.1 The expression of *NANOG* for TCGA samples for the following groups; High, Medium (Mid) or Low *NANOG* expression; * denotes differential expression greater than 2 fold ($p<0.05$). Error bars indicate the mean ± SEM for samples. EdgeR was used to calculate statistical significance.

Chapter 4: Validation of selected TGCT miRNAs

4.1 Introduction

In the previous chapter, four miRNA were identified as putative direct targets of NANOG in TGCT cells. This analysis aimed to answer two questions; (1) which miRNAs expressed in pluripotent cells have a NANOG binding site upstream of their transcriptional start sites (TSSs) and (2) are these miRNAs differentially expressed in the *NANOG* high and *NANOG* low groups identified by analysis of TGCT microarray and TCGA data. Out of the four miRNAs identified in Chapter 3, two were upregulated in TGCTs (*miR-512* and *miR-183*) and two were downregulated (*miR-887* and *miR-9*). This chapter seeks to validate NANOG regulation of these miRNAs.

MiR-512 is located on chromosomal band 19q13.4 which contains the largest number of miRNA genes in humans; these miRNAs have been termed as the C19MC (Chromosome 19 microRNA cluster) (Flor & Bullerdiek 2012). It is thought that miRNAs contained within C19MC originate from the neighbouring *miR-371-373* cluster (Zhang et al. 2008). These miRNAs have been used as a marker of malignant TGCTs (Murray et al. 2011) and is associated with cell proliferation and tumorigenesis through inhibition of the cell cycle regulator LATS2 (Voorhoeve et al. 2006). In addition to this, many of the C19MC miRNAs harbour the same 5' AAGUGC 3' seed sequence, including *miR-512-3p*. The pluripotency-associated miRNA clusters *miR-302a/b/c/d* and miR-372/373 also share this seed sequence (Nguyen et al. 2017). This suggests that *miR-512* may have targets in common with the *miR-302* and *miR-372/373*

clusters and may be important for the regulation of pluripotency and proliferation of ESCs and TGCTs.

The other upregulated miRNA, *miR-183*, is involved in the carcinogenesis (Mihelich *et al.*, 2011; Lee *et al.*, 2012). Indeed, *miR-183* expression is significantly lower in normal endometrium cells as compared with endometrial hyperplasia and carcinoma (Lee *et al.*, 2012). In addition, Novotny also showed high expression of *miR-183* in seminoma patient samples (Novotny *et al.*, 2012). This suggested that *miR-183* may also promote carcinogenesis in TGCTs. In contrast to this, a study comparing miRNA expression of human pluripotent stem cells versus differentiated endothelial cells reported that *miR-887* is highly expressed in differentiated cells (Lee *et al.*, 2017). In fact, *miR-887* had an inverse relationship with the pluripotency miRNAs, *miR-302* and *miR-367*, further suggesting a role for *miR-887* and differentiation (Lee *et al.*, 2017). Therefore, loss of *miR-367* expression in TGCTs could favour cells proliferation instead of differentiation. On the other hand, *miR-9* is important for neural differentiation in certain contexts as it is highly expressed in the central nervous system of the zebrafish, chick and mouse (Darnell *et al.*, 2006; Leucht *et al.*, 2008; Zhao *et al.*, 2009). In addition to this, it has been shown in mESCs that Nanog and Oct4 co-localised upstream of *miR-9-2* suggesting that it may control the expression of this miRNA in these cells (Rodda *et al.*, 2005). Collectively, this evidence indicates that the miRNAs selected are implicated in regulating pluripotency, proliferation, differentiation and carcinogenesis.

Meta-analyses of large genome-wide studies generally require further experimentation to validate the extrapolated relationships. Therefore, in this chapter four miRNA were chosen to be validated in TGCT cell lines.

TGCT cell lines are commonly used to analyse gene expression and to perform functional studies aimed at addressing biological questions related to TGCTs. Most of available TGCT cell lines including (TCam-2, NT2, NCCIT, 833KE, 2102Ep and Tera-1) have been shown to express high levels of NANOG (Schorle and Nettersheim, 2012), and therefore represent a good model to study the function of miRNA regulated by NANOG. All EC cell lines available for this project were derived from teratocarcinomas of the testis, apart from NCCIT (a male patient) which was derived from an extra-gonadal germ cell tumour within the mediastinum (Sperger *et al.*, 2003). These EC cell lines vary in their differentiation capacity. For instance, 833KE, 2102Ep and Tera1 are nullipotent and are unable to differentiate, whereas NTERA2 (NT2) and NCCIT retain the ability to differentiate (Matthaei et al. 1983; Andrews 1984). Both NT2 and NCCIT can be induced to form embryoid bodies (EBs) *in vitro* containing cells from all three germ layers, a characteristic also observed in teratomas and embryonic stem cells (Pal and Ravindran, 2006; Simões and Ramos, 2007; Mandal *et al.*, 2016). Despite the ability of NT2 to form all three lineages in EBs, these cells have a limited capacity for mesodermal differentiation when compared to hESCs (Simões and Ramos, 2007). A study by Sperger *et al*, 2003, conducted a microarray experiment including cell lines from ECs, somatic tumours, hESCs as well as patient samples for SEM, EC, Testis and yolk sac tumours (Sperger *et al.*, 2003). Unsupervised hierarchical clustering showed that

the EC cell lines clustered most closely with hESCs, indicating that their expression profiles were similar. This suggests that the EC compartment of TGCTs represents cells which have been reverted, or reprogrammed, to a state which more closely resembles hESCs.

TCam-2 and JKT-1 are two cell lines derived from seminoma tumours (Eckert *et al.*, 2008). JKT-1 does not express OCT4 or NANOG whereas TCam-2 and seminomas express high levels of these pluripotency genes (Eckert *et al.*, 2008). Therefore, it has been suggested that JKT-1 does not represent a suitable model of seminoma tumours (Eckert *et al.*, 2008). Like seminomas, TCam-2 cells express high levels of the gonadal PGC markers *VASA* and *DAZL*, suggesting TCam-2 as a suitable model of seminomas (Irie *et al.*, 2015). Both TCam-2 and seminomas express high levels of SOX17 which has been shown to be crucial for the specification of hPGC like cells (hPGCLCs) (Irie et al. 2015). This data strengthens evidence suggesting TGCTs are derived from PGCs and that TCam-2 and seminomas are important for modelling germ cell biology *in vitro*. The cell lines used in this project harbour duplications of chromosome 12p which is common in invasive TGCTs (Rosenberg et al. 2000). It has been observed that hESCs acquire amplification of chromosome 12p in culture and demonstrate increased proliferation and resistance to differentiation (Ben-David *et al.*, 2014). As noted earlier, *NANOG* resides on the chromosome 12p and has been shown to be significantly over-expressed in pluripotent cells with 12p amplifications (Ben-David *et al.*, 2014). As previously mentioned, NANOG regulates pluripotency and proliferation and has been shown to contribute to the phenotypic effects seen in 12p trisomic pluripotent cells (Ben-David *et al.*,

2014). (Ben-David *et al.*, 2014). In addition to this, Palmer used a mixture of TGCT cell lines to quantify miRNA expression using a microarray. The TGCT cell lines included those derived from ECs (Tera-2 and 2102ep), yolk sac tumours (GCT44 and 1411H) and seminomas (TCam-2). It was shown by unsupervised hierarchical clustering using miRNA expression that these cell lines clustered more closely to SEM, EC and YST patient samples when compared to teratoma samples and normal testis/ovary controls (Palmer *et al.*, 2010). This study demonstrated that TGCT cell lines *in vitro* can be used to quantify miRNA expression to model TGCT subtypes from patient samples. Therefore, based on the cell of origin and molecular characteristics, TGCT cell lines represent a good model to test the hypothesis of this project in order to find a causal association between miRNA involved in TGCTs and their transcriptional regulation by NANOG.

Aims and Objectives

The aim of this chapter was to validate the relationship inferred by the bioinformatics analysis described in Chapter 3 for selected miRNAs and NANOG.

The main objectives were:

1. To analyse TGCT cell lines and testes samples for expression of candidate miRNAs and correlate this with the expression profile obtained from the bioinformatics.

2. To validate that NANOG binds to and transcriptionally regulates genomic regions upstream of selected miRNAs using a luciferase assay.

3. To knockdown NANOG in TGCT cell lines to show the causal relationship between NANOG regulation and miRNA expression in TGCT cell lines.

4. To over-express the chosen NANOG-regulated miRNA in TGCT cell lines.

4.2 Results

4.1.1 Characterisation of TGCT cell lines

Gene expression analyses were performed using six culture TGCT cell lines including five embryonal carcinoma cell lines and one seminoma cell line (Table 4.1). NCCIT and 2102Ep have a flat cell morphology, especially when approaching confluence, whereas NTERA2 cl.D1 (NT2/D1) and 833KE cells exhibit a rounded morphology. TCam-2 cells exhibit enlarged nuclei with flat morphology (Figure 4.1). These cell lines were used to model TGCTs but were first validated for the expression of known markers characteristic of different TGCT subtypes using quantitative RT-PCR (RT-QPCR). *NANOG* and *OCT4* are known to be expressed in all TGCT subtypes whereas *SOX2* and *SOX17* are known to be specific for EC and SEM, respectively (Irie et al. 2015; Nonaka 2009).

Table 4.1 Description of the TGCT cell lines which were used for RT-QPCR analysis of miRNAs expression.

Name	Sample	Origin	Metastasis	Chr12 Status	References
NTERA2 cl.D1	Cell line	EC	-	+ 12p	(Mostert *et al.*, 1996; Ben-David *et al.*, 2014)
2102Ep	Cell line	EC	-	+ 12p	(Mostert *et al.*, 1996)
833KE	Cell line	EC	Abdomen	+ 12p	(Ben-David *et al.*, 2014)
NCCIT	Cell line	EC	-	+ 12p	(Ben-David *et al.*, 2014)
TERA1	Cell line	EC	Lung	+ 12p	(Fogh and Trempe, 1975)
TCam-2	Cell line	SEM	-	+ 12p	(de Jong *et al.*, 2008)

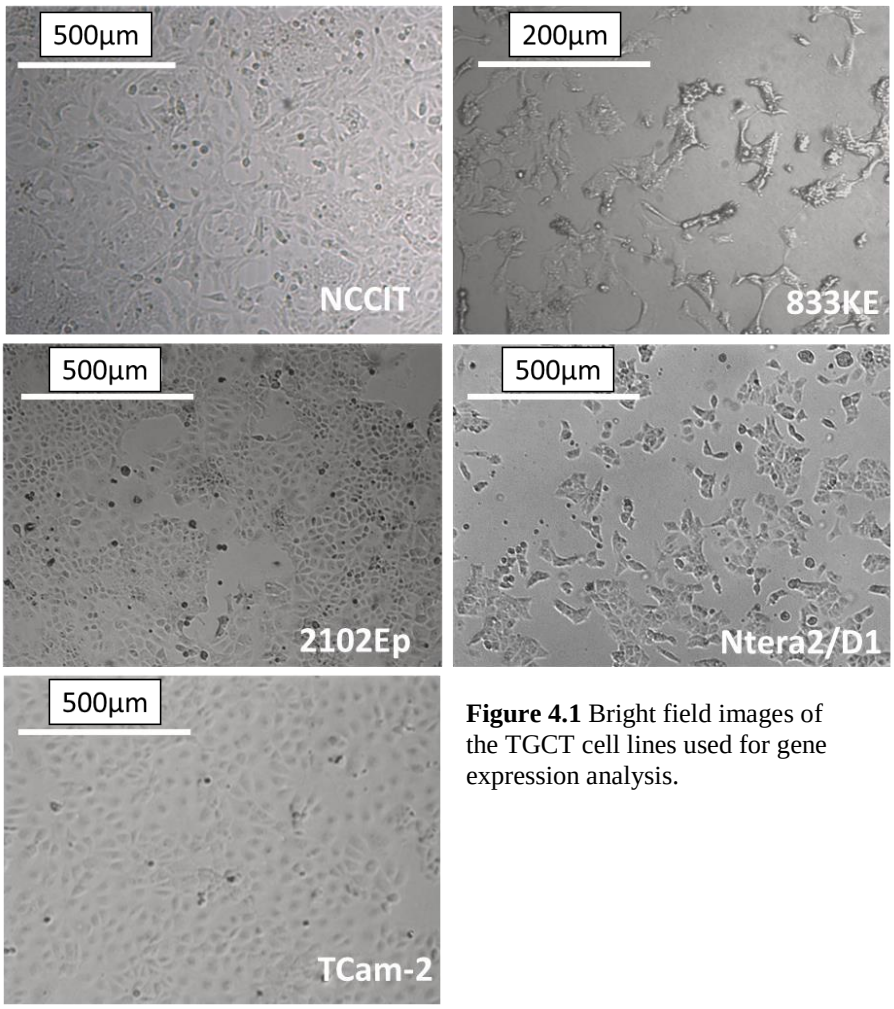


Figure 4.1 Bright field images of the TGCT cell lines used for gene expression analysis.

Prior to gene expression analysis of the genes of interest, the expression of three housekeeping genes were measured to determine which was most stable across samples and suitable for relative quantification. For all primers, the efficiency was calculated (Figure 4.5 and 4.6) and were shown to produce one product using melting curve analysis (Figures 4.2 and 4.3). The 'GeNorm' software (v3) was then used to analyse the stability of the endogenous controls across all samples to calculate an M-score (Vandesompele *et al.*, 2002). The M-score gives a value indicating the stability of an endogenous control, the lower the value the more stable the control. The M-score and the standard deviation was lowest for *β-ACTIN (ACTB)*, whereas *18S* had the highest M-score and standard deviation. Supporting this, the high-density intervals (HDIs) were smaller for *β-ACTIN* Ct values when compared to *18S* and *GAPDH* indicating a lower distribution cross all samples for *β-ACTIN* (Figure 4.6). The bean plots for *18S* and *GAPDH* indicated two populations in the samples, whereas *β-ACTIN* indicated a single population (Figure 4.6). This data indicated that *β-ACTIN* was the most stable across samples, therefore was used as a housekeeping gene for relative quantification for *NANOG*, *OCT4*, *SOX2*, and *SOX17*.

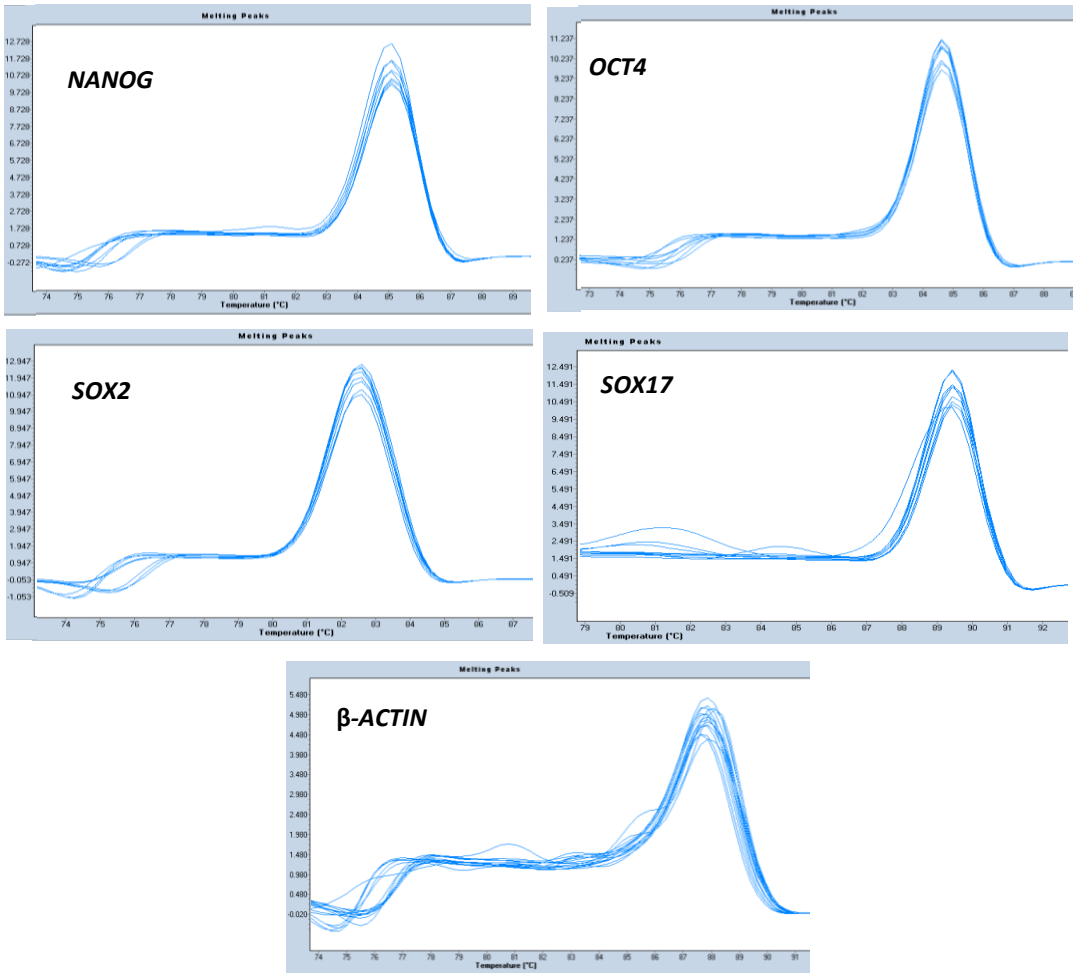


Figure 4.2 Melting curves for the following gene amplicons; *NANOG*, *OCT4*, *SOX2*, *SOX17* and *β-ACTIN*. A single peak indicates a single product amplified.

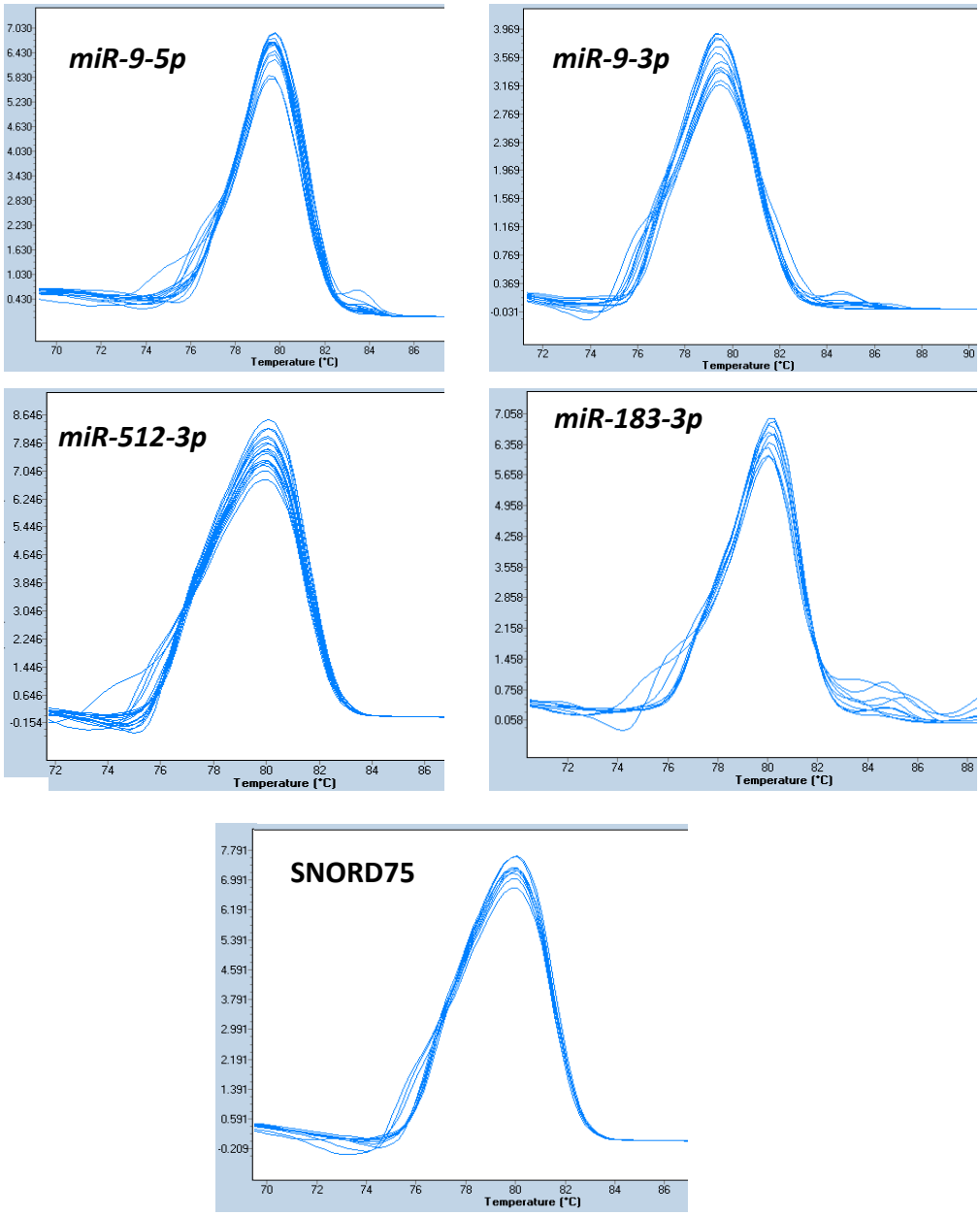
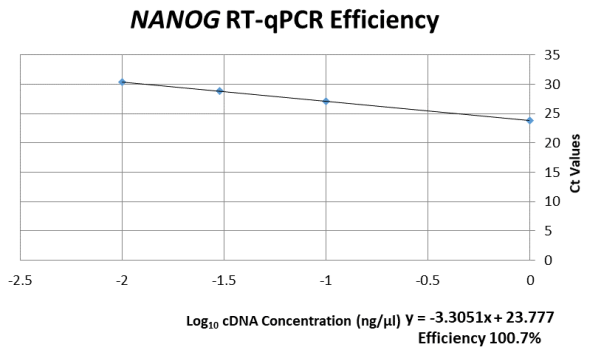
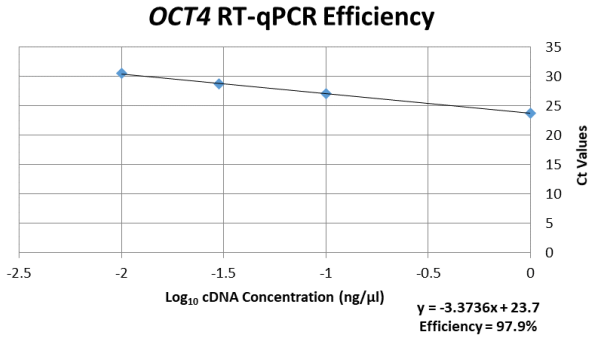
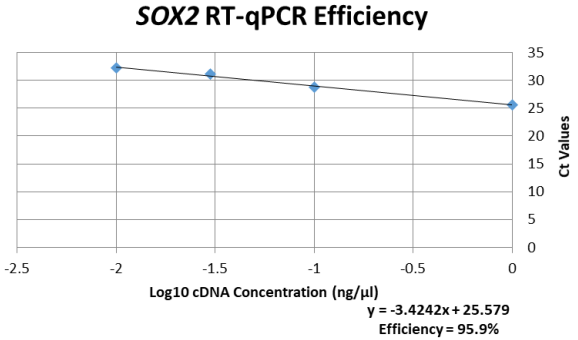
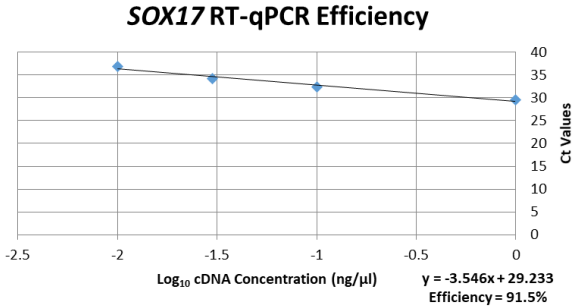


Figure 4.3 Melting curves for the following gene amplicons; *miR-9-5p*, *miR-9-3p*, *miR-512-3p*, *miR-183-3p* and SNORD75 (U75). A single peak indicates a single product amplified.









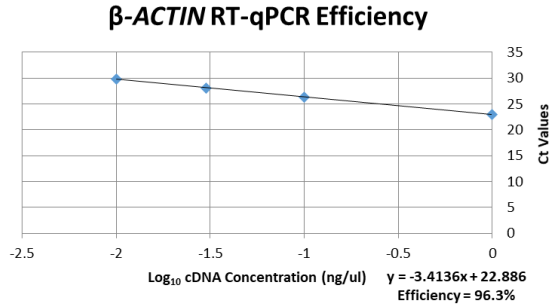


Figure 4.4 A scatterplot used to calculate the efficiency of *NANOG* (A), *OCT4* (B), *SOX2* (C) *SOX17* (D) and *β-ACTIN* (E) primers. RT-QPCR was performed on the following cDNA concentrations; 1, 0.1, 0.03 and 0.01 ng/ul. Ct values were plotted against the Log_{10} cDNA concentrations and the slope was calculated. The Thermo-Fisher qPCR efficiency calculator was used to calculate efficiencies.

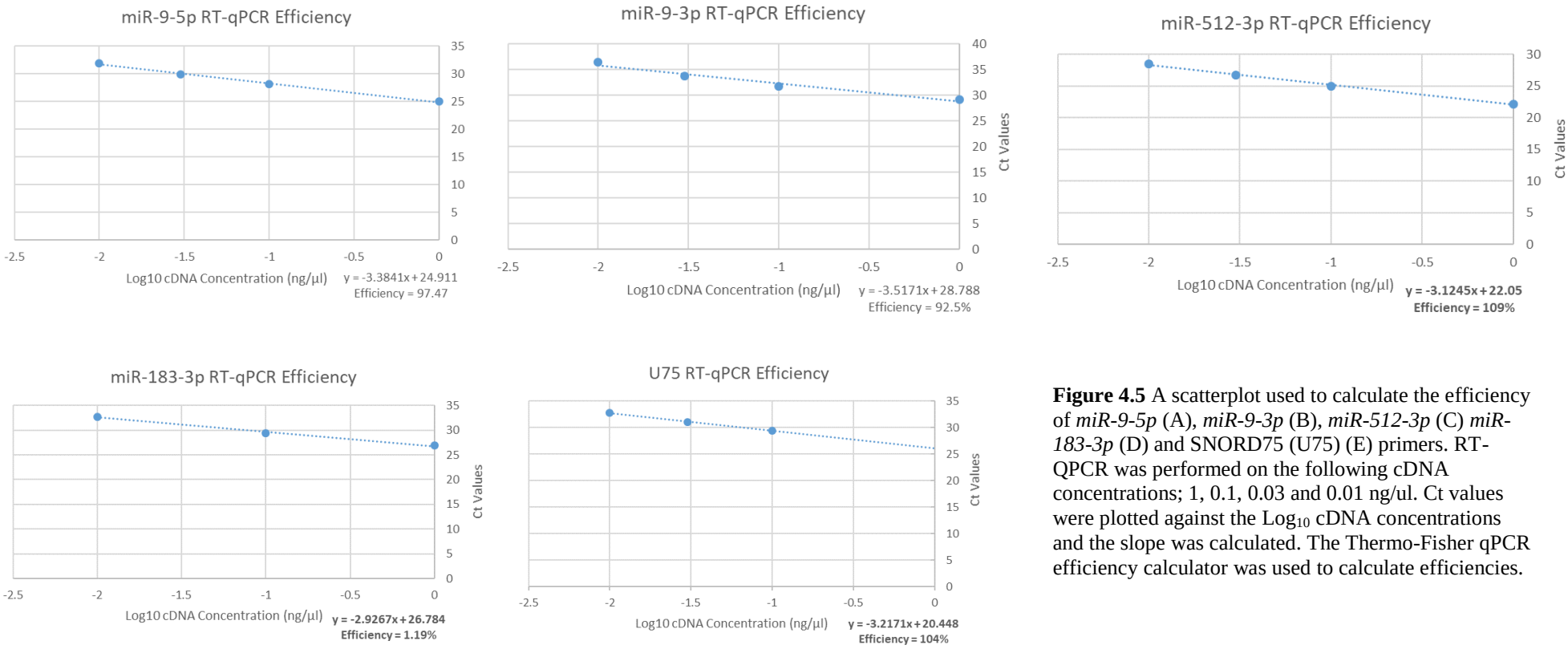
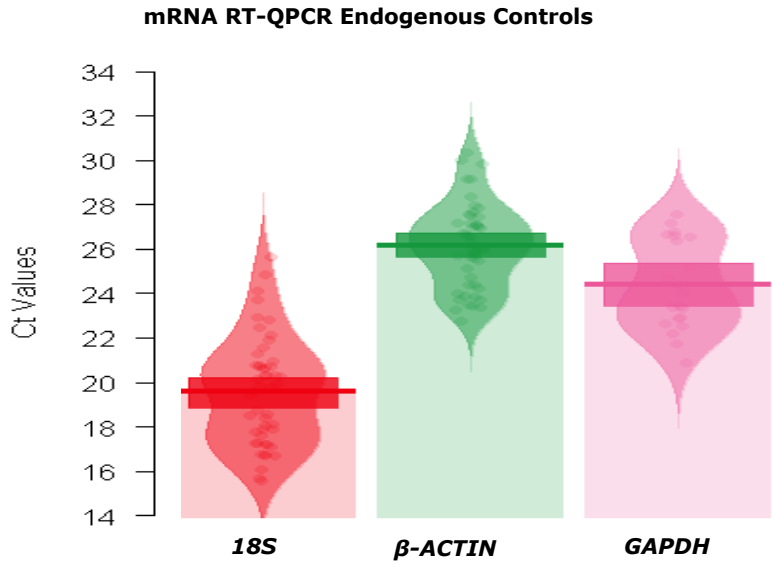


Figure 4.5 A scatterplot used to calculate the efficiency of *miR-9-5p* (A), *miR-9-3p* (B), *miR-512-3p* (C) *miR-183-3p* (D) and SNORD75 (U75) (E) primers. RT-QPCR was performed on the following cDNA concentrations; 1, 0.1, 0.03 and 0.01 ng/ul. Ct values were plotted against the Log_{10} cDNA concentrations and the slope was calculated. The Thermo-Fisher qPCR efficiency calculator was used to calculate efficiencies.



M-Score	0.240	0.184	0.185
St-Dev	2.1	1.6	2.1

Figure 4.6 A pirate plot showing the Ct values across all TGCT samples for *18S* (red), *β-ACTIN* (green) and *GAPDH* (pink). The points indicates each samples Ct (n=51), solid lines are the average Ct values and the larger solid box are the high density intervals (95% HDI) around the mean. The beans plots around the bars indicate point density. GeNorm was used to calculate the stability of endogenous controls; the lower the M-scores indicates increased stability (Vandesompele *et al.*, 2002). Variation of Ct values was also calculated using standard deviation.

The expression for *NANOG*, *OCT4* and *SOX2* in TGCTs was compared to the hESC line H9, as it expresses these pluripotency genes and therefore was used as a positive control (Figure 4.7 A-C). In contrast, *SOX17* expression was compared to TCam-2 as this cell line was known to express high levels. All TGCT cell lines express similar levels of *NANOG* and *OCT4* as compared to H9 (Figure 4.7 A-C). *SOX2* expression was similar in EC TGCT cell lines whereas TCam-2 expressed very low levels of *SOX2*, but high levels of *SOX17* expression. This was low in all EC samples and in H9 (Figure 4.8). *NANOG*, *OCT4* and *SOX2* expression was very low in testes samples when compared to H9 (Figure 4.7) and also low for *SOX17* when compared to TCam-2 (Figure 4.4). All RT-QPCR amplicons were sequenced to ensure the right gene was amplified (Supplementary S4.1).

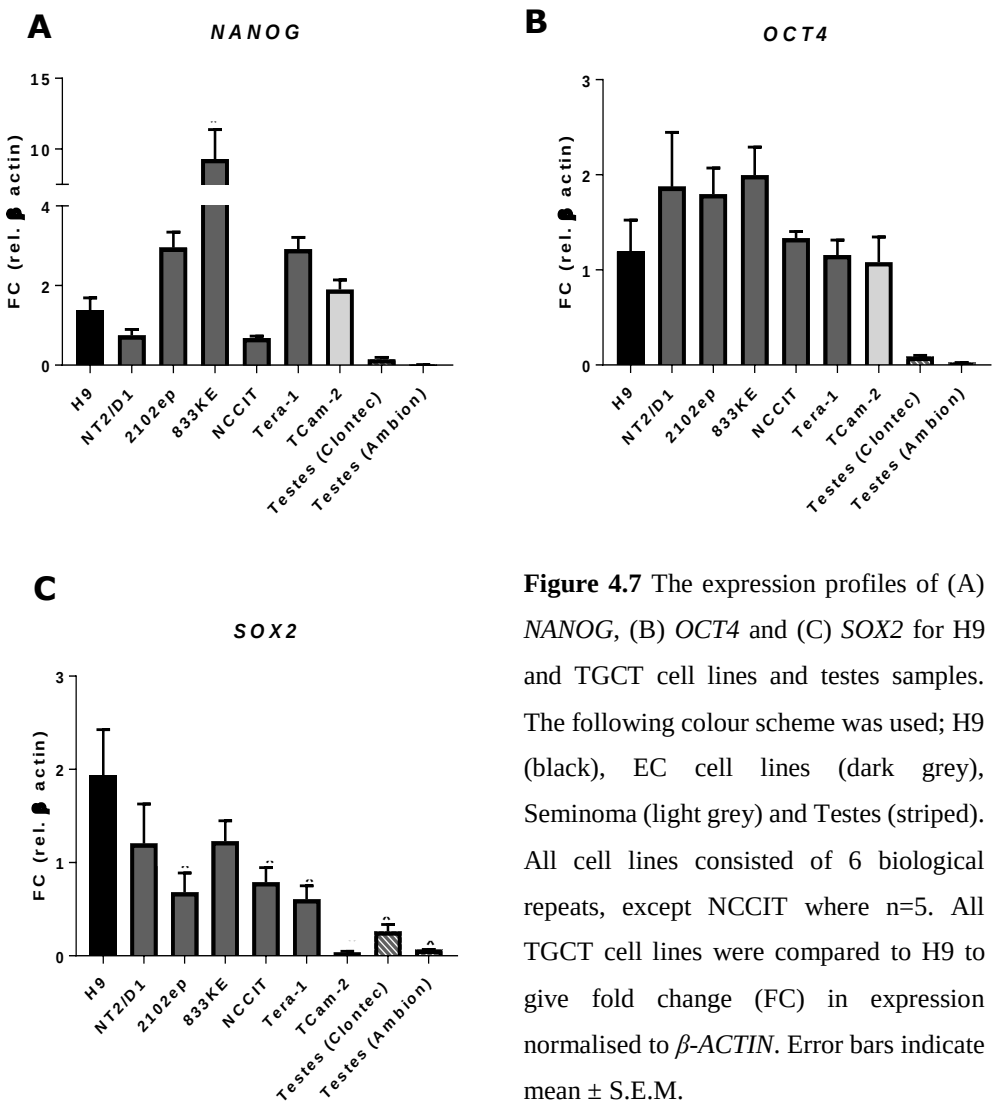


Figure 4.7 The expression profiles of (A) *NANOG*, (B) *OCT4* and (C) *SOX2* for H9 and TGCT cell lines and testes samples. The following colour scheme was used; H9 (black), EC cell lines (dark grey), Seminoma (light grey) and Testes (striped). All cell lines consisted of 6 biological repeats, except NCCIT where n=5. All TGCT cell lines were compared to H9 to give fold change (FC) in expression normalised to *β-ACTIN*. Error bars indicate mean ± S.E.M.

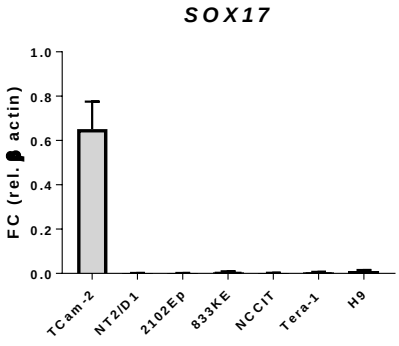
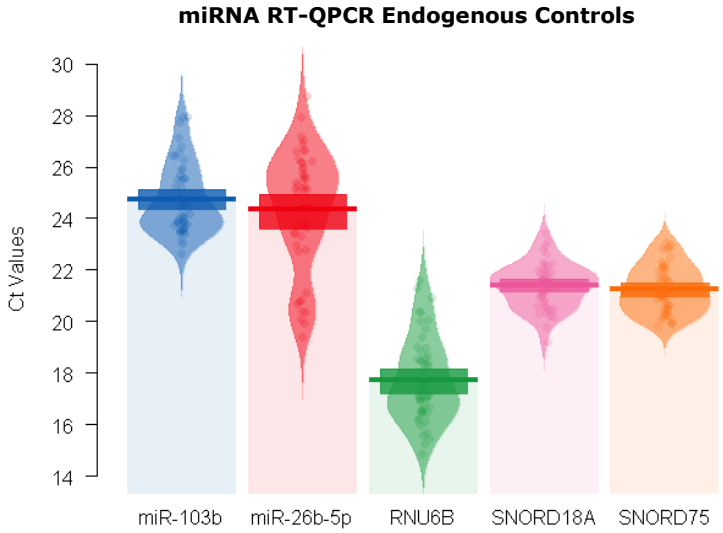


Figure 4.8 The expression profiles of *SOX17* for TGCT cell lines and Testes samples. The following colour scheme was used; H9 (black), EC cell lines (dark grey), Seminoma (light grey) and testes (striped). All cell lines consisted of 6 biological repeats, except NCCIT where n=5. All TGCT cell lines were compared to TCam-2 to give fold change (FC) expression normalised to *β-ACTIN*. Error bars indicate mean ± S.E.M.

4.2.2 Expression analysis of miRNAs in TGCT cell lines

For the RT-QPCR of miRNA genes, five endogenous controls and four candidate miRNA were analysed by GeNorm to elucidate the stability of the endogenous controls across all TGCT samples. *SNORD75* had the lowest M-score, whereas *miR-26b-5p* had the highest (Figure 4.9). In agreement with this, *miR-26b-5p* also had the highest standard deviation (SD) whereas *miR-26B-5p* had the lowest SD followed by *SNORD75*. MiR-26-5p had the largest HDI, whereas *miR-26B-5p* and *SNORD75* had the smallest. The bean plots for *miR103b*, *miR-26b-5p* and *RNU6B* indicated a large distribution in Ct values across samples (Figure 4.9). Bean plots for *miR-26B-5p* and *SNORD75*, however, were the most consistent indicating a smaller distribution in Ct values across samples. This data indicated that *SNORD75* was the most stable across samples and was used as the endogenous control for miRNA expression analysis.



M-Score	0.126	0.158	0.151	0.116	0.104
St-Dev	1.12	1.84	1.36	0.69	0.77

Figure 4.9 A pirate plot showing the Ct values across all TGCT samples *for miR-103b* (blue), *miR-26*-5p (red), *RNU6B* (green), *SNORD18A* (pink) and *SNORD75* (orange). The points represent the samples Ct value for each gene (n=51), solid lines are the average Ct values and the larger solid box are the high density intervals (95% HDI). The bean plots around the bars indicate point density. GeNorm was used to calculate the stability of endogenous controls; the lower the M-scores indicates increased stability (Vandesompele *et al.*, 2002). Variation of Ct values was also calculated using standard deviation.

Following the validation of TGCT cell lines for expression markers characteristic of each TGCT subtype, the miRNAs chosen from the bioinformatics analysis (Chapter 3; Table 3.3) were next analysed by RT-QPCR. All samples were compared to the normal testes sample (Ambion), as this would give a comparison of tumour versus normal tissue similar to the microarray analysis analysed in Chapter 3 (Figure 3.2). *SNORD75* (U75) was used as an internal control as it was shown to be the most stable all sample (Figure 4.9).

MiR-512-3p had lower expression ($p<0.05$) in all TGCT cell lines when compared to the testes control (Figure 4.10A). *MiR-183-3p* also had a lower expression ($p<0.05$) in TGCT cell lines when compared to the testes control (Figure 4.10B). Both *miR-9-5p* and *miR-9-3p* showed lower expression ($p<0.05$) in all TGCT samples (except 2102Ep) when compared to the testes control (Figure 4.10 B and C). In the microarray analysis in Chapter 3 it was shown that pre-miRNAs *miR-183* and *miR-512* were highly expressed in TGCT samples, however RT-QPCR data failed to detect *miR-183-5p* and *miR-887-3p* in cultured TGCT cell lines. However, RT-QPCR results for *miR-9-5p* and *miR-9-3p* were in agreement with the bioinformatics analysis which indicated a low expression of these miRNAs in TGCTs. Given this result, *miR-9* was chosen for further analysis.

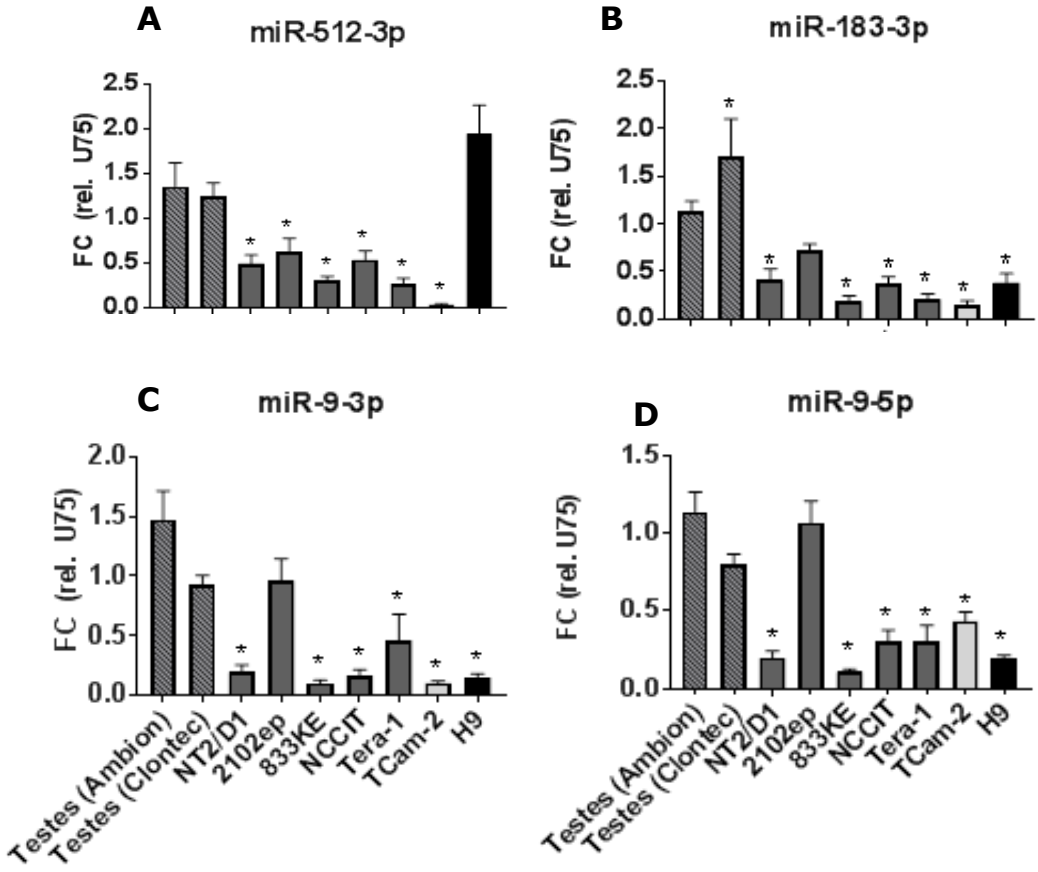


Figure 4.10 Expression profiles of (A) *miR-512-3p*, (B) *miR-183-3p*, (C) *miR-9-5p* and (D) *miR-9-3p* for TGCT cell lines and testes. The following colour scheme was used; testes (diagonal pattern), EC cell lines (dark grey) and TCam-2 a seminoma cell line (light grey). One way ANOVA was used to compare samples, * indicated a significant level ($p<0.05$) compared to testes (Ambion) to give fold change (FC) expression normalised to SNORD75 (U75). Error bars indicate mean ± S.E.M.

4.2.3 Characterising NANOG binding to the *miR-9* promoter

Following the RT-QPCR analysis of TGCT cell lines, the next step was to examine potential NANOG regulation of *miR-9*. In order to do this, NANOG was ectopically overexpressed in HEK293T cells via lentiviral transduction. To achieve this, three viral plasmids were first transfected into HEK293T cells to produce the lentivirus; pPAX2 (the packaging vector), pMD2.G (the envelop vector) and pSIN-EF2-NANOG-Pur (the transfer vector) shown in Figure 4.7A. The pSIN vector contains an EF1α promoter, the coding sequence for *NANOG* and a puromycin resistance cassette (Figure 4.11B). Once the lentivirus was produced, it was used to infect fresh HEK293T cells. Uninfected HEK293T cells (HEK293T WT), with no puromycin resistance, were killed at a concentration of 1.5µg/ml of puromycin (Figure 4.12). Cells infected with the lentivirus were puromycin resistant at a concentration of 1.5µg/ml for 7 days (Figure 4.12), indicating that the *NANOG* expressing plasmid had successfully integrated into the genome of the HEK293T cells. RT-PCR confirmed that control wild type (WT) HEK293T cells did not express *NANOG* whereas puromycin-resistant cells showed a high level of *NANOG* expression (Figure 4.13). From this point forward puromycin resistant cells were referred to as HEK293T-NANOG-OE.

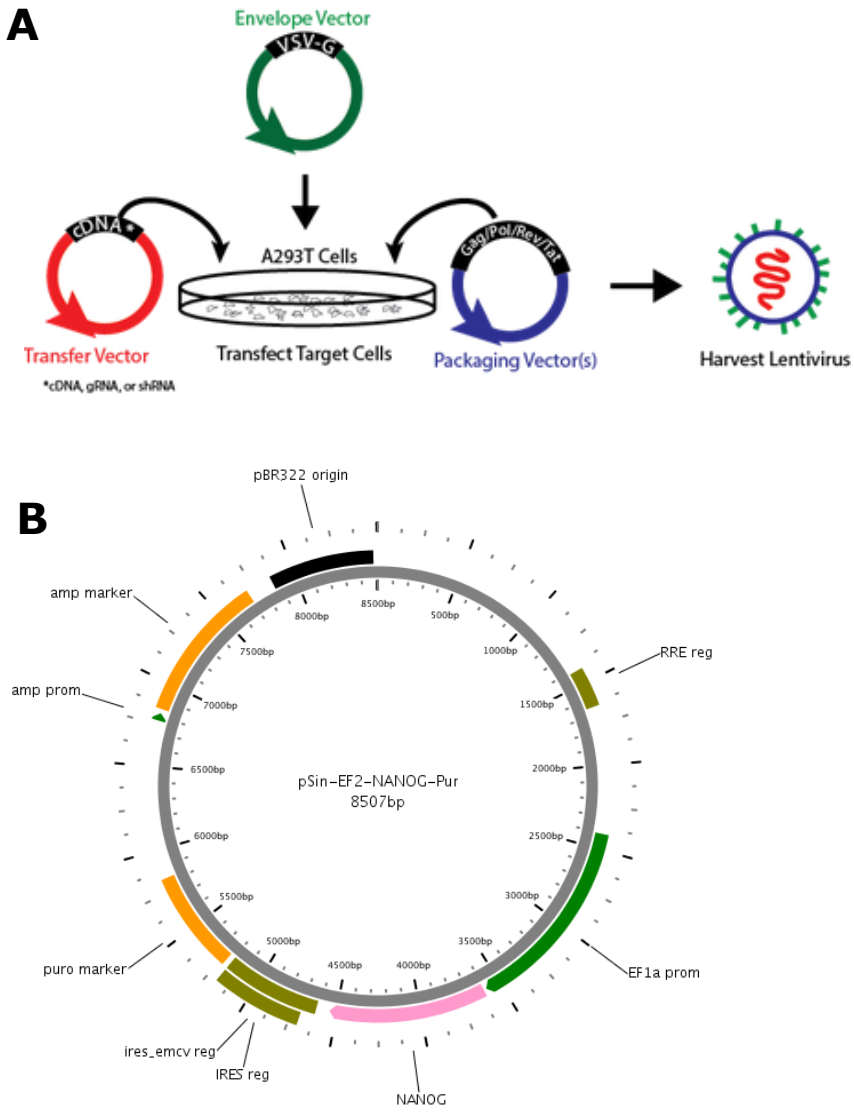


Figure 4.11 (A) schematic showing the procedure by which lentiviral particles were created; three plasmids were used including the packaging vector (blue), envelope vector (green) and the transfer vector (red). These plasmids were transfected into HEK293T cells allowing lentivirus manufacture. This schematic was adapted from Addgene protocols ((http://www.addgene.org/lentiviral/protocols-resources/). (B) the map for pSIN-EF2-NANOG-Puro, the transfer plasmid (created using PlasMapper (Dong *et al.*, 2004).

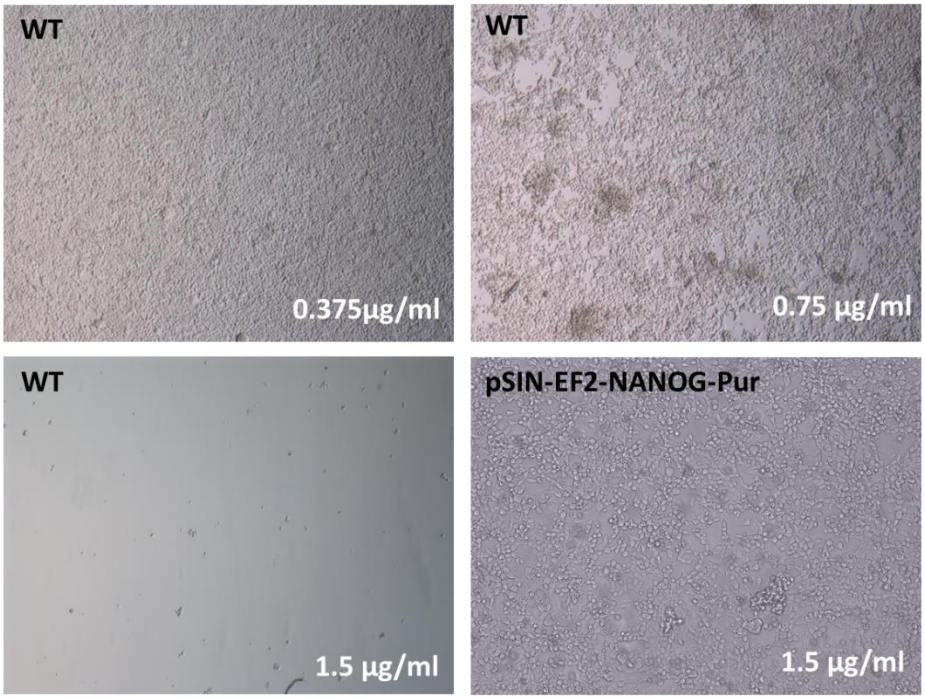


Figure 4.12 Puromycin selection experiment for HEK293T cells; increasing concentrations of puromycin were added to medium at 0.375, 0.75, 1.5µg/ml. 1.5µg/ml of puromycin was added to HEK293T cells transfected with the puromycin resistance gene (pSIN-EF2-NANOG-Pur).

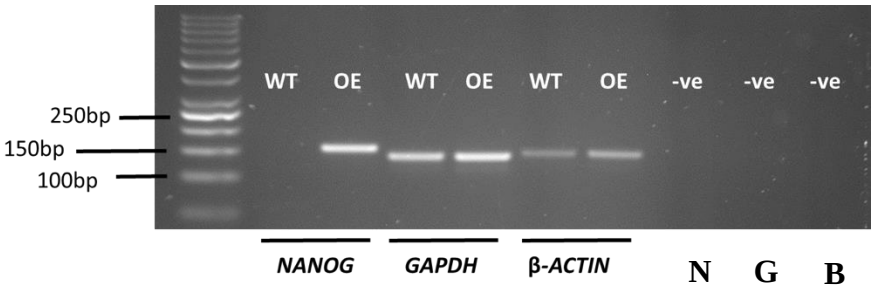


Figure 4.13 RT-PCR in HEK293T-WT (WT) and HEK293T-NANOG-OE (OE) cells for *NANOG, GAPDH* and *β-ACTIN*; expected band sizes were 155bp, 139bp and 144bp respectively. Water negative control (-ve) are shown for the primer *NANOG* (N), *GAPDH* (G) and *β-ACTIN* (B).

Following the creation of HEK293T-NANOG-OE, the next step was to produce a vector which contained the genomic sequence upstream from the transcriptional start site (TSS) of *miR-9* which is believed to harbour a NANOG regulated region NANOG ChIP-seq dataset from Chapter 3 was used to identify the *miR-9* promoter (Figure 3.1). The genomic region 464bp upstream of *miR-9* was shown to be bound by NANOG enrichment (Figure 4.14).

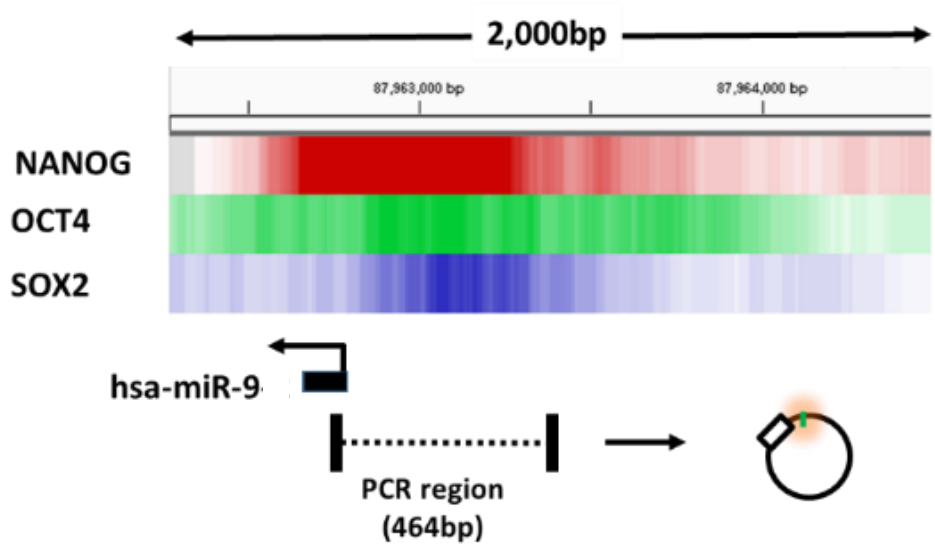


Figure 4.14 A schematic illustrating the genomic region with NANOG enrichment that was amplified by RT-PCR and inserted into pGL3 plasmids. NANOG (red), OCT4 (green) and SOX2 (blue) enrichment is shown upstream of *miR-9* TSS using the ChIP-seq data from Chapter 3 (Figure 3.1).

This genomic region was amplified from TCam-2 using primers flanked with the restriction enzyme sequences for NheI and SacI which are present in the multiple cloning site (MCS) of the pGL3 plasmids (Figure 4.15). Two pGL3 plasmids were used, the pGL3-Promoter and pGL3-Basic. The pGL3 promoter plasmid carries a SV40 promoter of *firefly* luciferase, in contrast to pGL3 basic which lacks a constitutive promoter upstream of the *firefly* luciferase gene (Figure 4.15). Both plasmids also carry a resistance ampicillin cassette. The purified 464bp PCR amplicon and pGL3 plasmids were digested using NheI and SacI. The genomic region upstream of *miR-9* was then sub-cloned into the pGL3-Basic and pGL3-Promoter plasmids and transformed into bacterial cells. Ampicillin resistant colonies were selected and grown in overnight cultures. The pGL3 plasmids were then verified to contain insertions through restriction digest using NheI and SacI to show a band of ~464bp and sequenced to show the insert was correct (Figure 4.16; Figure S4.3). The pGL3 plasmids are referred to as pGL3-Basic-*miR-9* and pGL3-Promoter-*miR-9*.

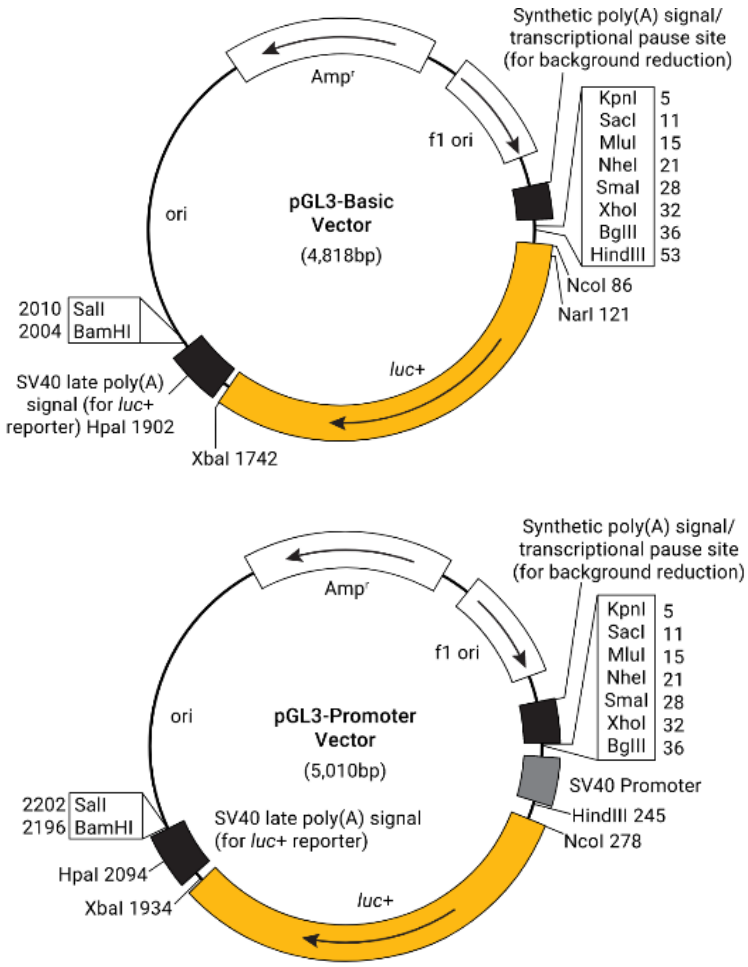


Figure 4.15 The map for pGL3-Basic (top) and pGL3-Promoter (bottom); vectors maps were adapted from the Promega Technical manual 'pGL3 Luciferase Reporter Vectors' (https://www.promega.com/-/media/files/resources/protocols/technical-manuals/0/pgl3-luciferase-reporter-vectors-protocol.pdf).

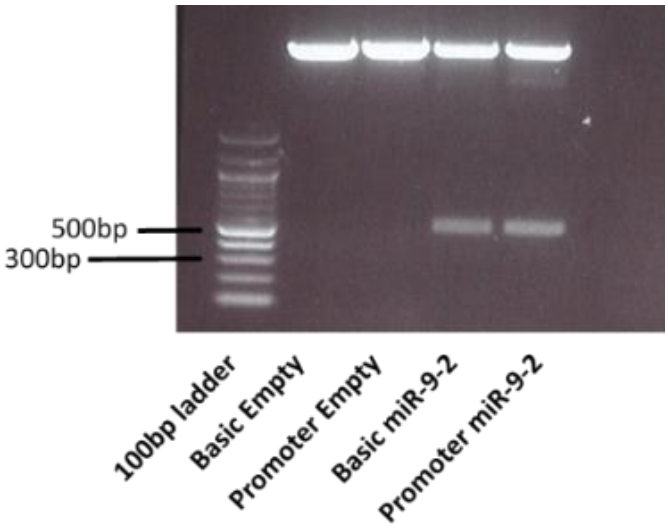


Figure 4.16 pGL3 plasmids were cut with NheI and SacI to show inserts at the correct size (464bp) were successfully ligated.

Once the pGL3-*miR-9* plasmids had been created they were then transfected intoHEK293T WT and HEK293T NANOG OE cells. Either pGL3-Basic-*miR-9* or pGL3-Promoter-*miR-9* was transfected into cells in combination with a plasmid encoding the *Renilla* luciferase gene. *Renilla* was used for the normalisation of luciferase activity. HEK293T NANOG OE cells transfected with pGL3-Basic-*miR-9* showed a significantly lower expression of *firefly* luciferase when compared to HEK293T WT (Figure 4.17 A). NANOG OE HEK293T cells transfected with pGL3-Promoter-*miR-9*-3 had no difference in luciferase activity when comparing HEK293T WT and HEK293T NANOG cells (Figure 4.17 B). The pGL3 Basic suggested that the genomic region upstream of *miR-9* contained a promoter sequence that is negatively regulated by NANOG.

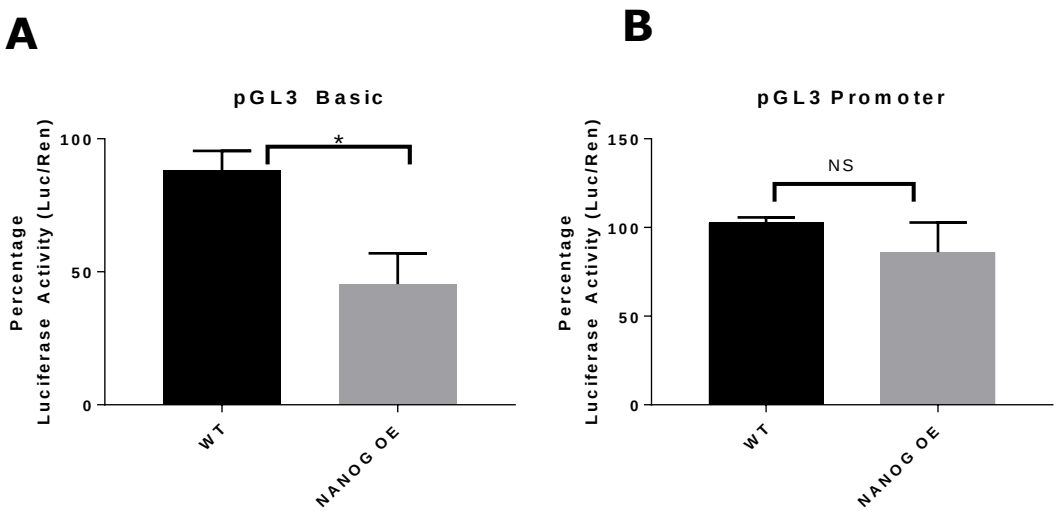


Figure 4.17 *MiR-9*-pGL3-Basic was transfected into HEK293T WT (WT) and HEK293T cells over-expressing NANOG (NANOG OE), this experiment consisted of 6 biological replicates (n=6). *MiR-9*-pGL3-Promoter was transfected into HEK293T WT (WT) and HEK293T cells over-expressing NANOG (NANOG OE). A two-tailed student t-test was used to compare samples, * indicated a significant level $p<0.05$. Error bars indicated mean ± standard error mean.

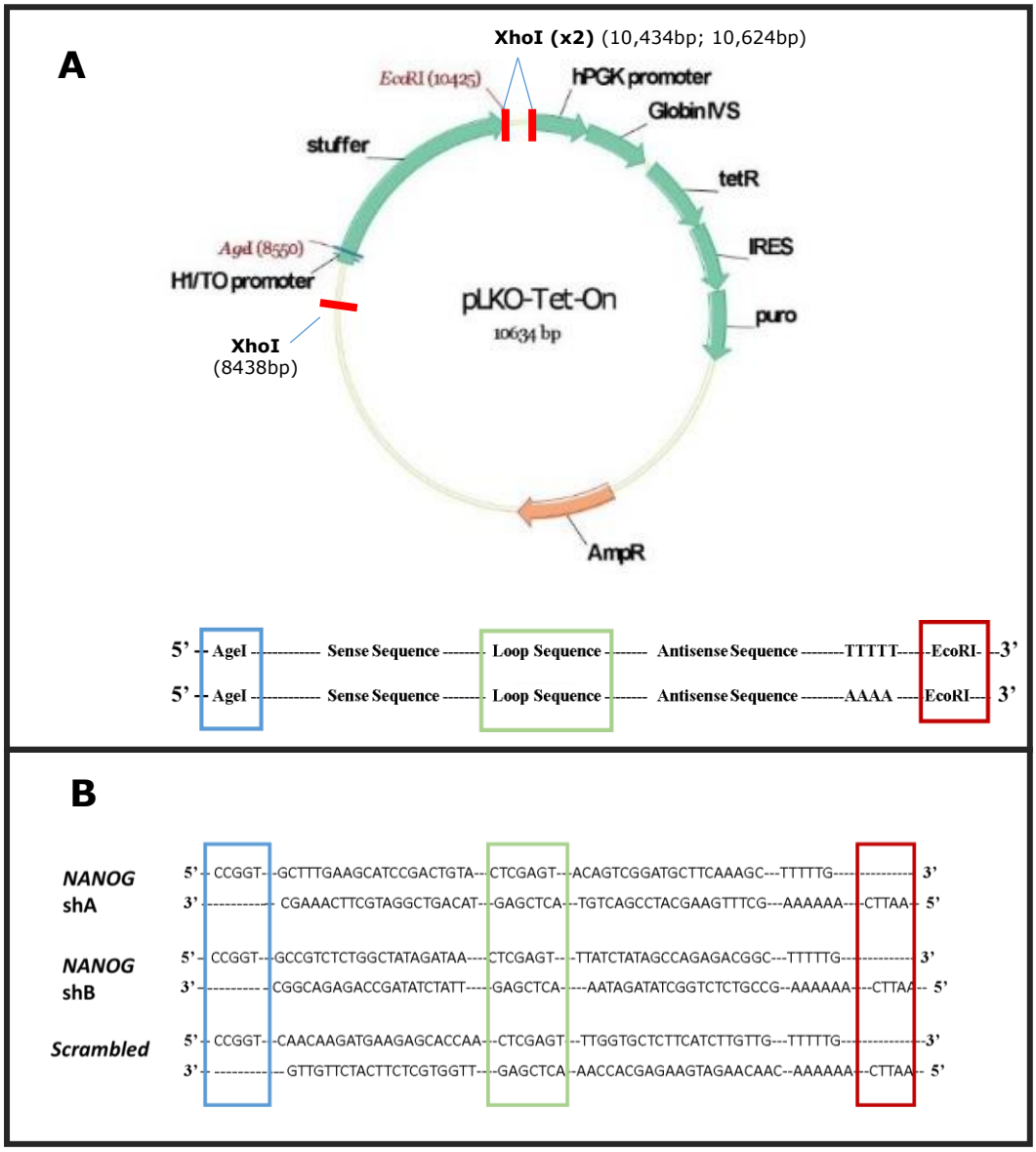


Figure 4.18 (A) A plasmid map for pLKO-Tet-On (top) and an example of an annealed oligonucleotide that were inserted into the pLKO plasmid (bottom). The vector map was adapted from AddGene (https://media.addgene.org/data/41/67/165920fc-af64-11e0-90fe-003048dd6500.pdf). (B) The oligonucleotides for *NANOG* shA, *NANOG* shB and *Scrambled*. The boxes indicate the AgeI sites (blue), the loop sequences (green) and the EcoRI sequences (red).

In order to validate the regulation of *miR-9* by NANOG in TGCTs, shRNA was used to knockdown the expression of *NANOG* in TCam-2 cells and the effect on *miR-9* expression was analysed. TCam-2 cells were chosen given their similarity in gene expression to human PGCs and their significance in the cellular origin of TGCTs. To knockdown NANOG, a short hairpin (sh)RNA approach using the Tet-pLKO-Puro (pLKO) lentivirus vector was used (Figure 4.18 A).

The *NANOG* shRNAs were designed to bind to complementary sequences for the *NANOG* transcript. In addition to this, a '*Scrambled'* shRNA was also created to be used as a non-targeting control (Figure 4.18 B). To allow the ligation of the oligonucleotides into the transfer vector pLKO, the sense and antisense oligonucleotides were annealed together leaving sticky overhangs (AgeI at the 5' end and EcoRI at the 3' end). The pLKO plasmid was digested using AgeI and EcoRI restriction enzymes and purified. The annealed oligonucleotides were then ligated into the linearised plasmid; this was performed for all three *NANOG* targeting oligonucleotides (*NANOG* shA, *NANOG* shB and *Scrambled*). Plasmids were transformed into bacterial cells, ampicillin resistant colonies were selected and grown in mini cultures. The pLKO plasmids were then verified to contain insertions through restriction digest using XhoI to confirm an insertion into the pLKO plasmid. As expected, clones with an insertion cut with XhoI produced three bands sized 8447, 200bp and 170bp; whereas the unmodified plasmid gave three bands 8447bp, 1996bp and 170bp (Figure 4.19). This confirmed that the correct insertion was present in NANOG shA, NANOG shB and Scrambled pLKO plasmids.

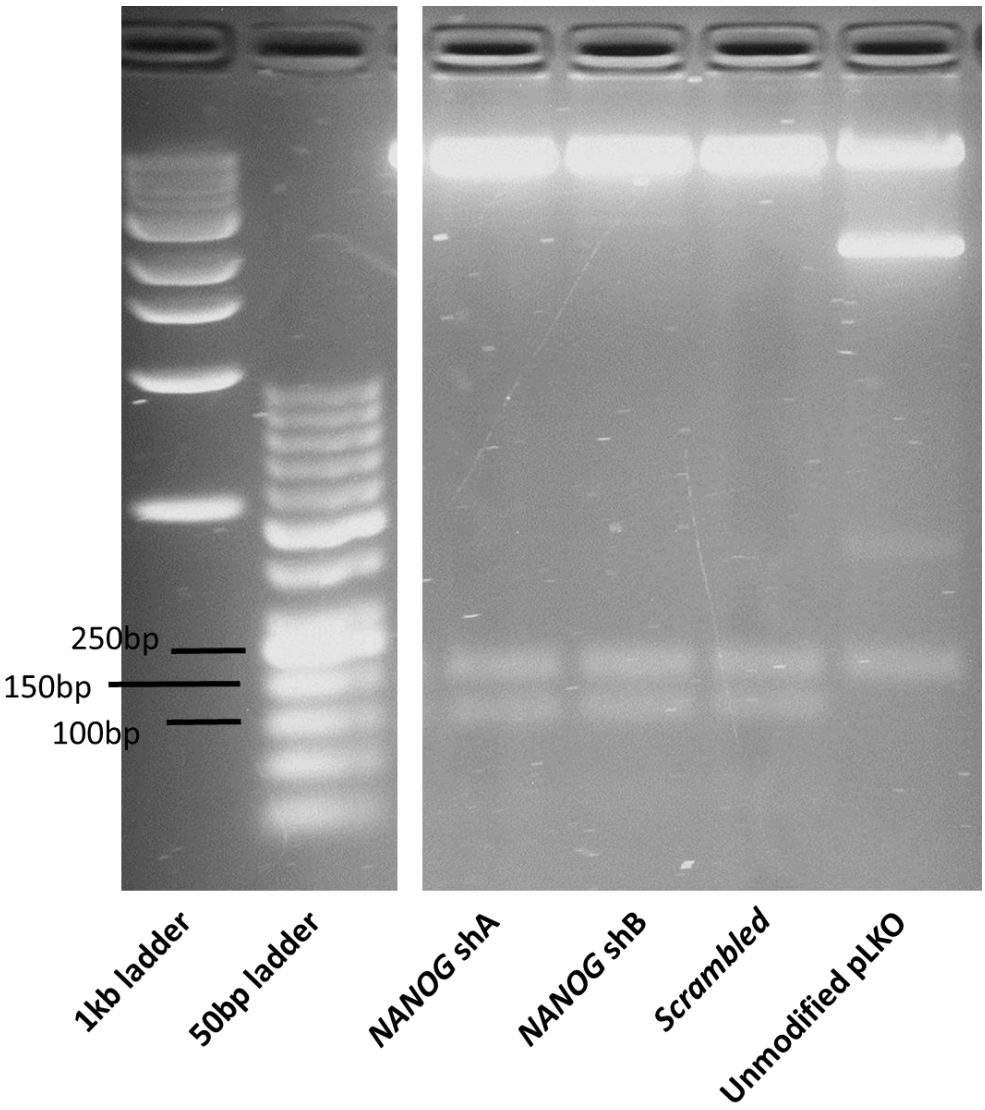


Figure 4.19 Enzyme digestions using XhoI for the plKO plasmids containing the inserts *NANOG* shA, *NANOG* shB and *Scrambled* in addition to the unmodified pLKO plasmid. pLKO plasmids with an insert when cut with XhoI gives three bands sized 8447, 200bp and 170bp; whereas the unmodified plasmid gives three bands 8447bp, 1996bp and 170bp.

The NANOG targeting lentivirus was then prepared using the same protocol as described for the overexpression of NANOG in HEK293T cells (Figure 4.11A); once prepared, the virus was then used to transduce TCam-2 cells. A death curve for puromycin showed that 0.5µg/ml of puromycin after 7 days killed uninfected TCam-2 cells (WT) (Figure 4.20). However, cells transduced with the lentivirus

for *NANOG* shA, *NANOG* shB and Scrambled survived the 7 days puromycin treatment (0.5µg/ml) (Figure 4.20). Both *NANOG* shA and shB transduced TCam-2 cells showed a reduction in *NANOG* expression when compared the *Scrambled* control ($p<0.05$) (Figure 4.21). As *NANOG* shA cells showed the greatest reduction in *NANOG* knockdown (69%) when compared to *NANOG* shB cells, *NANOG* shA cells were analysed for *miR-9* expression. *NANOG* shA TCam-2 cells exhibited an increase in *miR-9* expression when compared to the *Scrambled* control ($p<0.05$) (Figure 4.22).

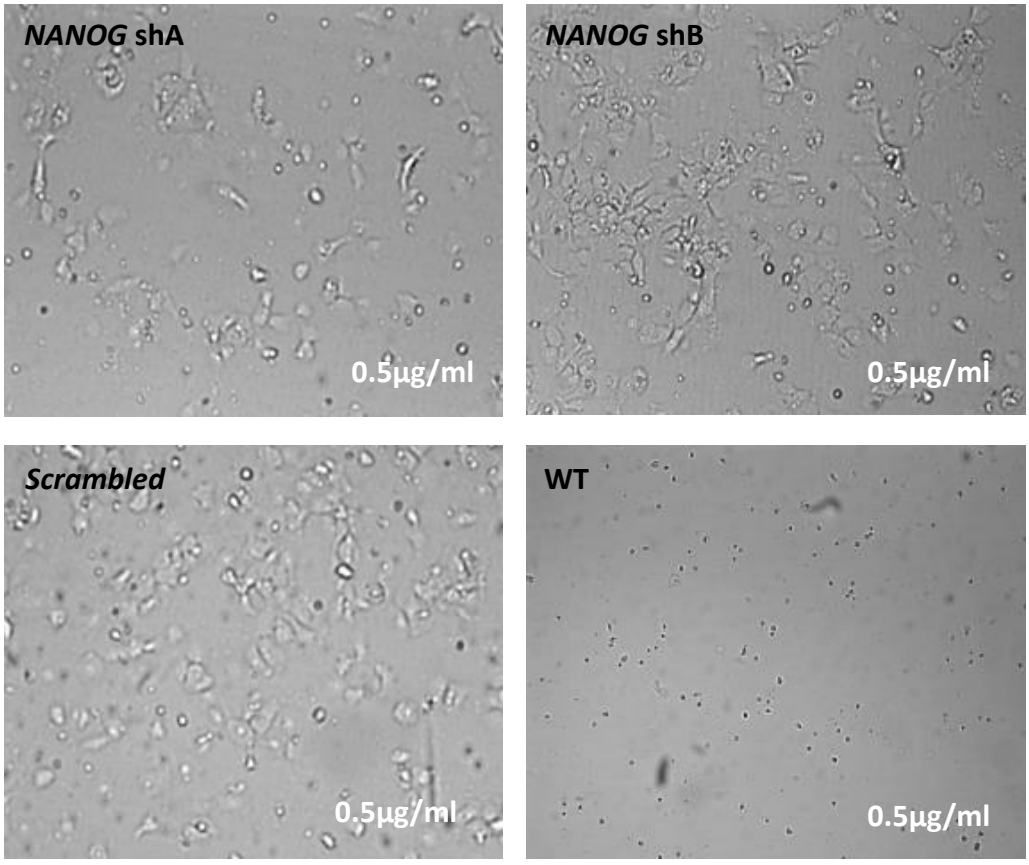


Figure 4.20 Puromycin was added at a concentration of 0.5µg/ml to TCam-2 cells infected with lentivirus containing the pLKO transfer plasmids *NANOG* shA (top left), *NANOG* shB (top right), *Scrambled* (bottom left) and uninfected TCam-2 WT cells (bottom right).

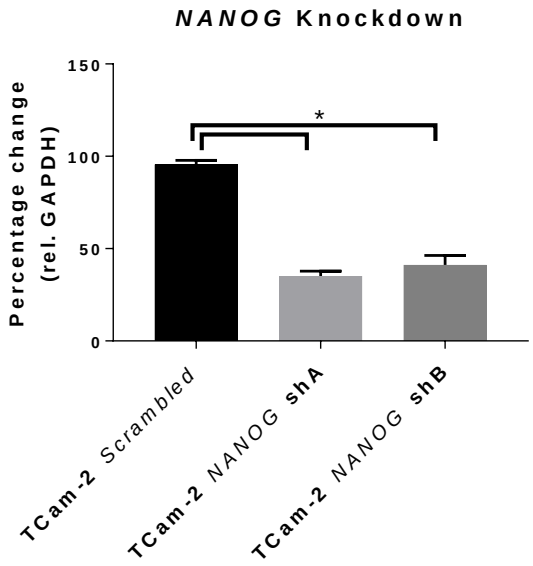


Figure 4.21 RT-QPCR for the expression of *NANOG* in TCam-2 *Scrambled*, *NANOG*-shA and *NANOG*-shB samples; samples were normalised to *GAPDH* expression. ANOVA was used to compare samples, * indicated a significant level ($p<0.05$) when compared to *Scrambled*. This experiment consisted of two biological repeats (n=2). Error bars indicated mean ± standard error mean. Analysis confirms successful knockdown of *NANOG* mRNA.

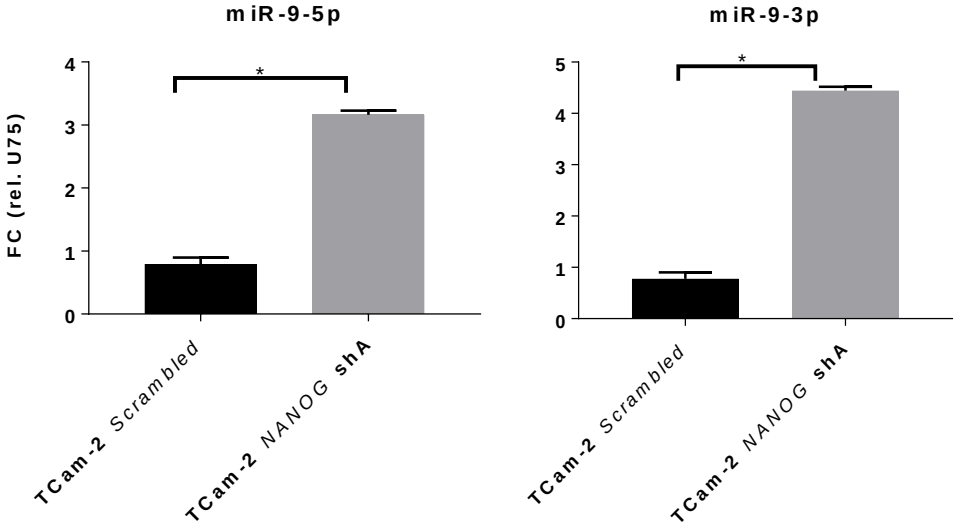


Figure 4.22 RT-QPCR for TCam-2 *Scrambled* and *NANOG* shA for the expression of *miR-9-5p* (A) and *miR-9-3p* (B); samples were normalised to SNORD75 (U75) expression. A two-tailed student t-test was used to compare samples, * indicated a significant level $p<0.05$. This experiment consisted of two biological repeats (n=2). Error bars indicated mean ± standard error mean.

4.2.4 Overexpression of *miR-9* in TGCTs

As it was shown that NANOG negatively regulated the expression of *miR-9* in TGCTs, the next step was to create a TGCT stable cell line overexpressing *miR-9* so that its function in TGCTs could be tested. Similar to NANOG knockdown experiments, *miR-9* overexpression experiments were also conducted in TCam-2 using lentiviral transduction as described previously for pSIN-EF2-NANOG-Pur (Figure 4.11 A). The Tet-pLKO-puro plasmid used for NANOG knockdown (Figure 4.18 A) was also used to express a synthetic miRNA mimic for *miR-9-5p*. As described previously, the oligonucleotides were ligated into the linearised pLKO plasmid (Figure 4.23 A) and transformed into bacterial cells. Ampicillin resistant colonies were selected and grown in overnight cultures. The pLKO plasmids were then verified to have insertions through restriction digest using XhoI to confirm an insertion into the pLKO plasmid. The pLKO-*miR-9-5p* and pLKO-Scrambled plasmids cut with XhoI gave three bands sized 8447, 200bp and 170bp; whereas the unmodified plasmid gives three bands 8447bp, 1996bp and 170bp (Figure 4.23 B). This confirmed that the correct insertion was present in *miR-9-5p* and *Scrambled* pLKO plasmids.

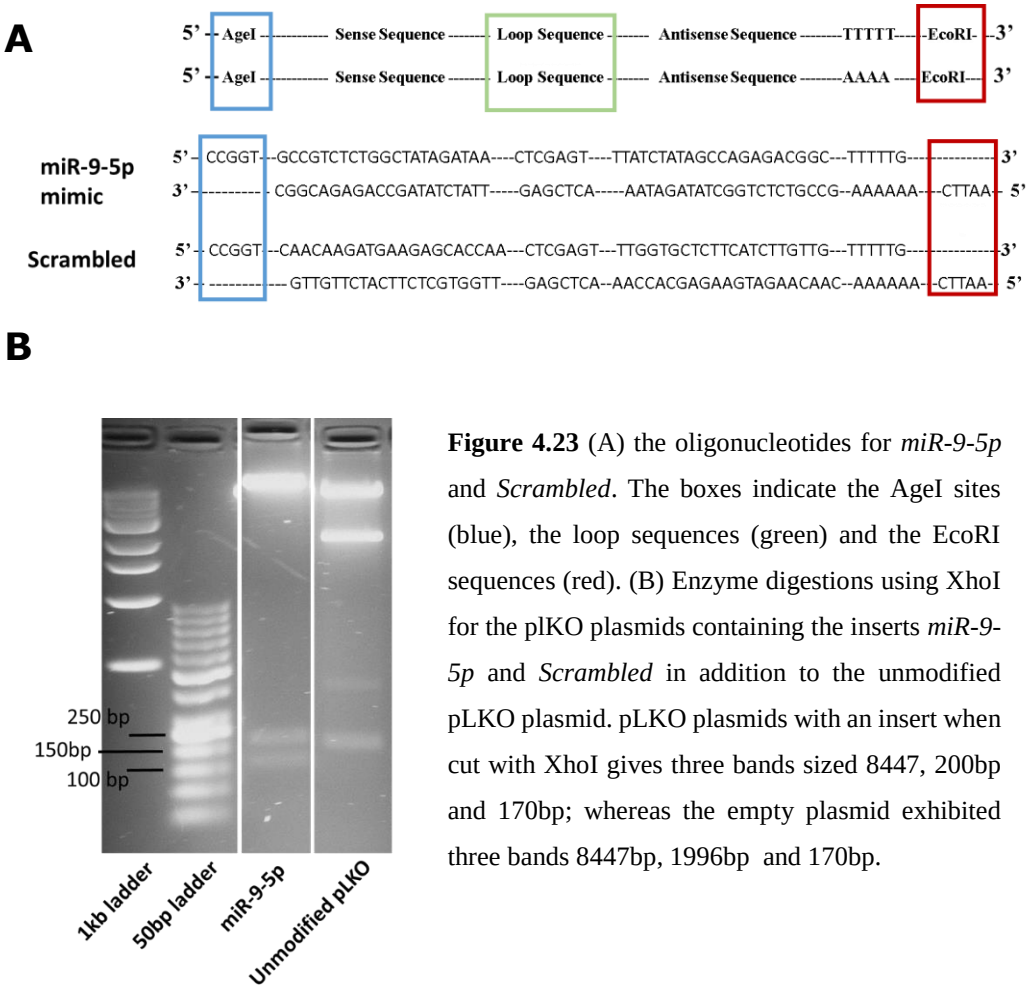


Figure 4.23 (A) the oligonucleotides for *miR-9-5p* and *Scrambled*. The boxes indicate the AgeI sites (blue), the loop sequences (green) and the EcoRI sequences (red). (B) Enzyme digestions using XhoI for the plKO plasmids containing the inserts *miR-9-5p* and *Scrambled* in addition to the unmodified pLKO plasmid. pLKO plasmids with an insert when cut with XhoI gives three bands sized 8447, 200bp and 170bp; whereas the empty plasmid exhibited three bands 8447bp, 1996bp and 170bp.

The lentivirus was prepared using the same protocol as described for the over-expression of NANOG in HEK293T cells (Section 4.2.5) and once created, the virus was then used to infect TCam-2 cells. As shown previously, TCam-2 WT cells were killed with puromycin at a concentration of 0.5µg/ml (Figure 4.20), this experiment was repeated using TCam-2 WT and TCam-2 infected with lentivirus for *miR-9-5p* pLKO. TCam-2 WT cells were killed with medium containing 0.5µg/ml of puromycin whereas the infected cells remained alive, indicating that the transfer plasmids containing Scrambled and *miR-9-5p* had successfully integrated into the TCam-2 cells genome and expressed the gene for puromycin resistance (Figure 4.24).

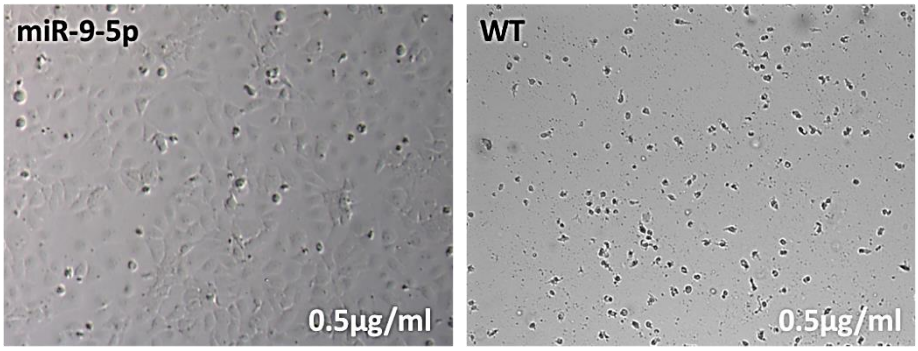


Figure 4.24 Puromycin was added at a concentration of 0.5µg/ml to TCam-2 cells infected with lentivirus containing the pLKO transfer plasmids *miR-9-5p* (left) and and uninfected TCam-2 WT cells (right).

Following this, the puromycin resistant cells were analysed by RT-QPCR for *miR-9-5p* and *miR-9-3p* expression. Cells transduced with the pLKO-*miR-9-5p* had significantly higher expression of *miR-9-5p* when compared to cells transduced with pLKO-Scrambled (Figure 4.25). *MiR-9-5p* expression showed no significant change in pLKO-Scrambled cells when compared to TCam-2 WT cells. *MiR-9-3p* expression showed no significant difference in expression across all three groups (Figure 4.25). The increase seen in *miR-9-5p* indicated that the plasmid was functional, however the increase in expression was modest at ~2.5 fold increase. As this increase in expression might not be sufficient to downregulate targets and produce functional effects, these modified TCam-2 cells were not used for further analysis.

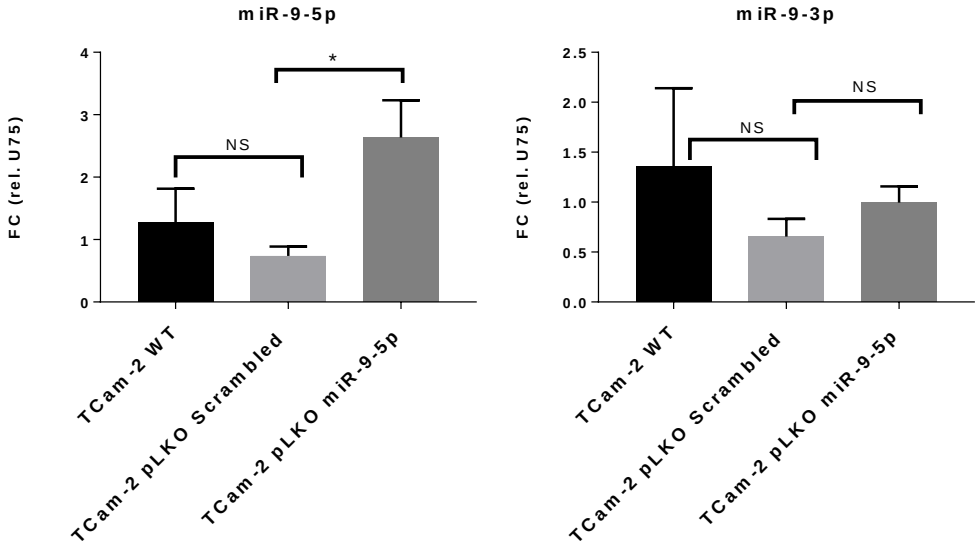


Figure 4.25 RT-QPCR for *miR-9-5p* (left) and *miR-9-3p* (right) for TCam2 WT (Black), TCam-2 pLKO Scrambled (light grey) and TCam-2 *miR-9-5p* (dark grey). A two-tailed student t-test was used to compare samples, * indicates a significant level $p<0.05$. This experiment consisted of three biological repeats (n=3). Error bars represent transcript counts with standard error of the mean.

A second plasmid was next used for the over expression of *miR-9* (MDH1-PGK-GFP-miR9). MDH1-PGK-GFP-miR9 contained the genomic region for *miR-9* thus has the ability to produce both *miR-9-5p* and *miR-9-3p* (Ma *et al.*, 2010). As this plasmid already contained the coding sequence for *miR-9*, a control MDH1-PGK-GFP was made using a scrambled sequence. The sense and antisense oligonucleotides for scrambled were annealed leaving the sticky ends (EcoRI at the 5' end and XhoI at the 3' end). The MDH1-PGK-GFP-miR9 plasmid was digested with EcoRI and XhoI which linearised the plasmid and removed the *miR-9*-3 sequence. The annealed oligonucleotide for scrambled was then ligated into the empty MDH1-PGK-GFP plasmid and transformed into

bacterial cells. Ampicillin resistant colonies were selected and grown in overnight cultures. To validate the insertion was successful, plasmids were cut with XbaI and ClaI. MDH1-PGK-GFP-miR9 positive control produced a large band (715bp) when compared to MDH1-PGK-GFP clones one and two which had a smaller bottom band (427bp) (Figure 4.26). This indicated that in clones one and two the scrambled sequence had been successfully inserted into the MDH1-PGK-GFP plasmid.

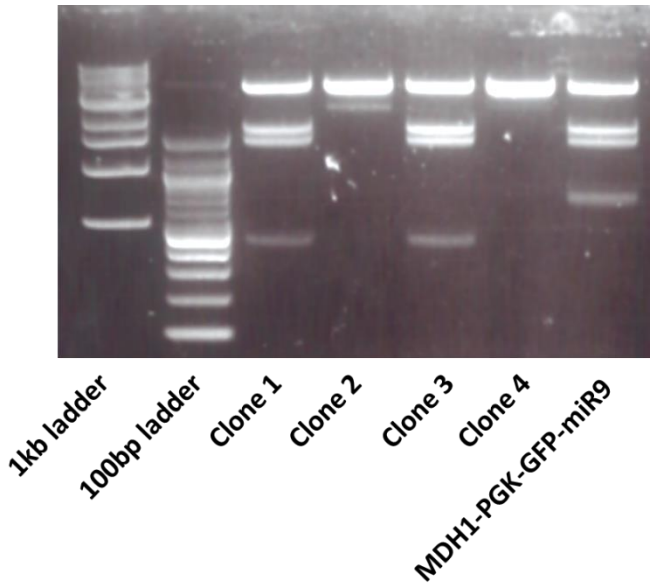


Figure 4.26 Enzyme digestions for the MDH1-PGK-GFP plasmids using XbaI and ClaI on ligated plasmids (clones 1-4) and the unmodified MDH1-PGK-GFP-miR9 plasmid. Positive clones are indicated with a bottom band of 427bp (clones 1 and 3).

The MDH1-PGK-GFP plasmid was a retroviral vector therefore uses different packaging plasmids as than used for lentivirus methods, although the procedure was similar. In order to prepare the retrovirus, three viral plasmids were transfected into HEK293T cells to produce the virus; pUMVC (the packaging vector), pMD2.G (the envelop vector) and MDH1-PGK-GFP (the transfer vector). Once the retrovirus was produced, it was added to TCam-2 cells. As the

MDH1-PGK-GFP plasmid contained a GFP expression cassette, both *miR-9* and Scrambled MDH1-PGK-GFP infected cells were shown to express GFP (Figure 4.27). As shown in the fluorescent images for TCam-2-*miR-9* and TCam-2-Scrambled, there was a heterogeneous population of GFP positive and GFP negative cells (Figure 4.27).

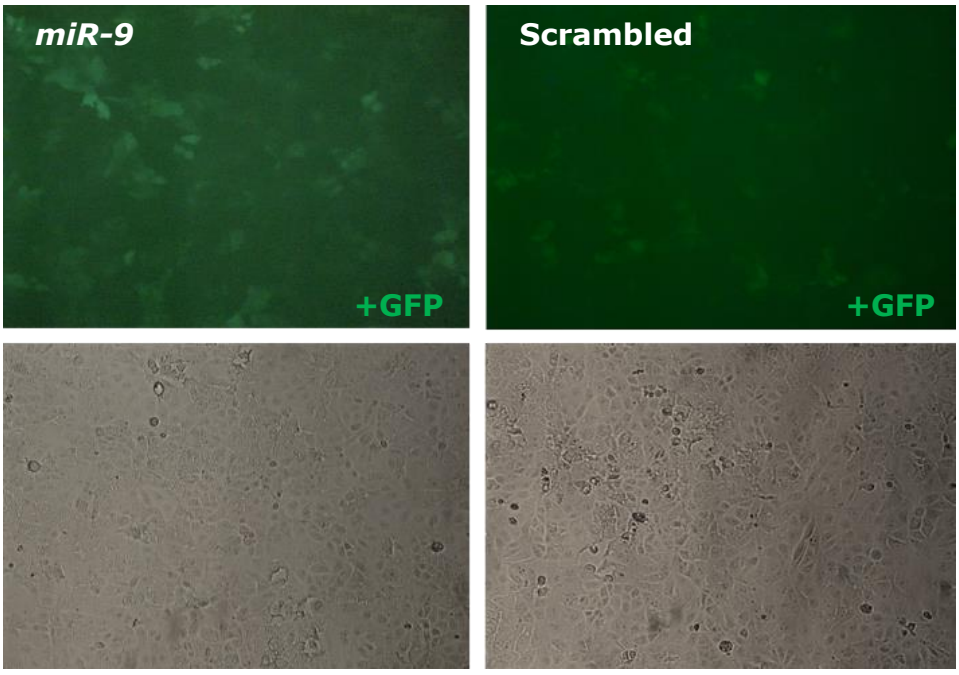
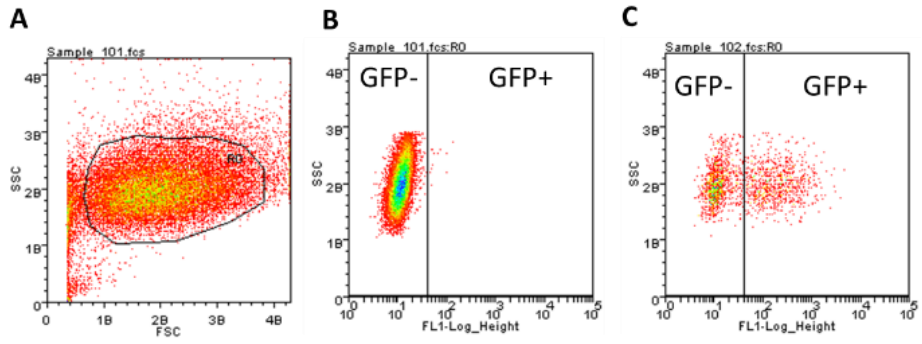


Figure 4.27 GFP is detected in both TCam-2 MDH1-PGK-GFP-miR9 and MDH1-PGK-GFP-*Scrambled* infected TCam-2 cells.

In order to separate out GFP negative cells fluorescence-activated cell sorting (FACS) was used. FACS is a method by which cells are passed through a capillary tube using flow cytometry and sorted based on fluorescence. In order to assess cell singularity and cell viability, the forward and side scatter of the light was used, respectively. Using forward and side scatter analysis on a sample population of cells, gating was created for the population of cells which consisted of cells both singular and viable (Figure 4.28 A). Following this, non-

labelled cells were analysed for of GFP to allow the distinction between GFP positive and GFP negative cells (Figure 4.28 B). Using a sample population from the MDH1-*miR-9* infected cells, it was shown this gate gave two cell populations GFP positive and GFP negative (Figure 4.28 C). Three biological replicates were used for each treatment type (n=3 *miR-9* overexpression; n=3 scrambled control); all 6 samples were analysed by FACS using the parameters described. On average, 49.9% of cells for *miR-9* TCam-2 cells were GFP+, whereas Scrambled TCam-2 cells had an average of 14.5% of GFP+ cells (Figure 4.24 D). From this point forward the cells infected with the MDH1-PGK-GFP-miR9 and MDH1-PGK-GFP-*Scrambled* were called TCam-2-*miR-9* and TCam-2-*Scrabmled.*



D

Sample Type	% GFP Positive	No. Cell Sorted
MiR-9 rep 1	50.5	188,542
MiR-9 rep 2	51.2	212,521
MiR-9 rep 3	48	187,372
Scrambled rep 1	15.5	55,556
Scrambled rep 2	13.7	50,498
Scrambled rep 3	14.2	46,309

Figure 4.28 Gating for the FACS analysis was based on (A) Forward and side scatter analysis and (B) fluorescence analysis on TCam-2 WT cells which were GFP negative (GFP-). (C) Using this gate, infected cells indicated two populations of cells (GFP + and GFP -), (D) *miR-9* (n=3) and *Scrambled* (n=3) infected TCam-2 cells were analysed and the percentage and number of GFP + cells sorted was calculated.

Sorted cells were then analysed for *miR-9* gene expression; the expression of *miR-9-5p* and *miR-9-3p* was significantly higher in TCam-2-*miR-9* when compared to TCam-2-*Scrambled* and TCam-2-WT (Figure 4.29). The levels of over-expression were greater than those previously obtained using the pLKO vector and to a sufficient level that the cells were subsequently analysed by RNA-seq to identify *miR-9* targets in TGCTs.

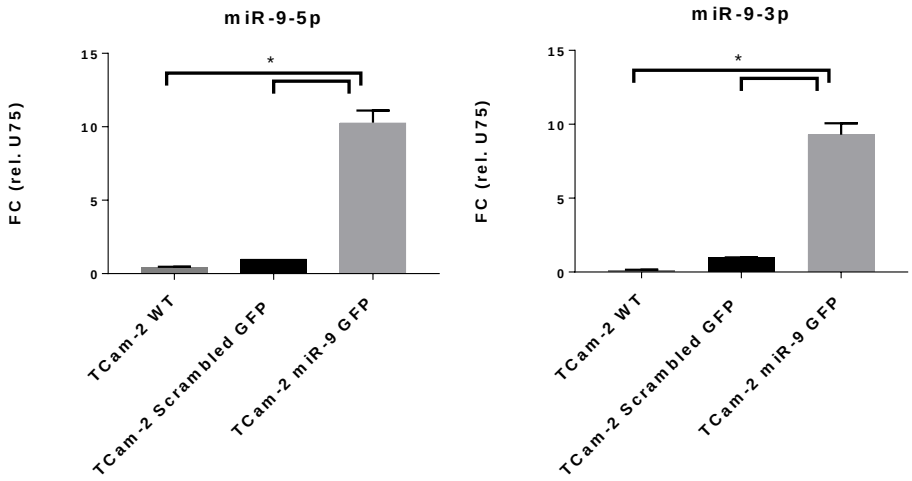


Figure 4.29 RT-QPCR for *miR-9-5p* (left) and *miR-9-3p* (right) for TCam2 WT (dark grey), TCam-2 MDH1-*Scrambled* (black) and TCam-2 MDH1-*miR-9* (light grey). A two-tailed student t-test was used to compare samples, * indicates a significant level $p<0.05$. This experiment consisted of three biological repeats (n=3). Error bars represent transcript counts with standard error of the mean.

4.3 Discussion

The purpose of this chapter was to validate those miRNAs that were identified in Chapter 3 as potentially involved in TGCTs and regulated by NANOG. In order to do this, five embryonal carcinoma and one seminoma cell line cell lines were used (Table 4.1). All TGCT cell lines presented an undifferentiated morphology in culture, suggesting that they could potentially retain the ability to differentiate (i.e. Ntera2 and NCCIT (Matthaei et al. 1983; Andrews 1984)). Following this, RT-QPCR was used to analyse the expression for mRNA transcripts. In order to do this, a stable endogenous control was required. β-*ACTIN* showed the lowest standard deviation and 'M-score' across all TGCT samples when compared to *18S* and *GAPDH* (Figure 4.6). As it has been shown that the lowest M-score value gives indication of the most stable endogenous control across samples (Vandesompele *et al.*, 2002), β-*ACTIN* was used as the endogenous control when normalising RT-QPCR for mRNA transcripts. RT-QPCR for the expression of the pluripotency marker *NANOG* indicated that TGCTs expressed levels either similar to or higher than hESCs (H9 cell line). *OCT4* had an expression similar to hESCs in all TGCT cell lines instead. It has been previously shown that both *NANOG* and *OCT4* are highly expressed in EC and SEM, both tumours and cell lines, therefore the data obtained in this study confirmed what was known in the literature (Sperger *et al.*, 2003; Clark *et al.*, 2004; Ezeh *et al.*, 2005; Hart *et al.*, 2005; Ben-David *et al.*, 2014; Irie *et al.*, 2015).

Interestingly *OCT4* is one of the top two significantly expressed genes in seminomas, embryonal carcinomas and hESCs when compared to somatic cells, normal testis and yolk-sac tumours (Sperger *et al.*, 2003). This data suggests the

similarity in gene expression across these samples which is in agreement with the RT-QPCR data (Figure 4.7 B). Although *NANOG* shows some variation, all TGCT cell lines express high levels of *NANOG* (Figure 4.7A), as previously shown in both SEM and EC TGCT patient samples (Ezeh *et al.,* 2005; Hart *et al.*, 2005; Hoei-Hansen *et al.*, 2005). Contrasting this, *SOX2* expression was similar in EC and hESCs, but not in the seminoma cell line TCam-2 (Figure 4.7 C). On the other hand, *SOX17* was highly expressed in TCam-2 (Figure 4.8), in agreement with previous studies which have shown that *SOX17* and *SOX2* can be used to distinguish SEM and EC samples as SEM and EC exclusively express high levels of *SOX17* and *SOX2*, respectively (De Jong *et al.,* 2008). This expression analysis identified that TGCT cell lines express high levels of the pluripotency markers *NANOG* and *OCT4* and that SEM and EC express their respective markers (*SOX17* and *SOX2*). As the TCGT cell lines were shown to express markers at the levels expected, they could be used for further experiments.

Therefore, TGCT cell lines were next analysed for miRNA expression analysis. Similar to the microarray analysis in Chapter 3 (Figure 3.1), TGCT cell lines were related to a testes control in order to make a tumour versus normal control comparison. A different endogenous control was required in order to analyse miRNA expression. Therefore, a total of five small non-coding RNA endogenous controls were analysed in TGCT and testes samples (Figure 4.9). *SNORD75* (U75) had the lowest M-score when compared to all other small non-coding RNAs in all TGCT and testes samples (Figure 4.9) and was subsequently used as an endogenous control when analysing miRNA expression. The analysis

of miRNA expression gave some unexpected results. Indeed, *miR-512-3p* and *miR-183-3p* expression was significantly lower in the majority of TGCT cell lines in comparison to the testes control (Figure 4.10 A and B). This is the opposite of what was observed in the microarray data, which showed that TGCT SEM and EC patient samples exhibited high levels of expression in comparison to the testes (Table 3.1). However, *miR-9-5p* and *miR-9-3p* expression was significantly lower in the majority of TGCTs when compared to the testes control (Figure 4.10 C and D). This expression profile was in agreement with the microarray data indicating low expression of *miR-9-5p* and *miR-9*-3 in TGCT patient samples (SEM and EC) compared to the testes control. Although *miR-887-5p* and *miR-183-5p* were chosen as candidate miRNA for validation, they could not be successfully amplified using RT-QPCR. One possible explanation could be that the expression of *miR-887* could have been too low to be detected, as the TCGA RNA-seq counts were very low for this miRNA (an average of 126 counts in comparison to *miR-512* which had an average count of 2882). A different explanation could be responsible in the failure of amplification for *miR-183-5p*. *MiR-183* transcript showed a mean count of 110,113 in the RNA-seq data, which was reflected by successful amplification of *miR-183-3p* (Figure 4.10 B). However, it has been shown that, when the miRNA duplex associates with the RISC complex, only one arm is preferentially bound while the other is degraded (Khvorova, Reynolds and Jayasena, 2003). This could explain why miR183-3p is detectable in TGCT cell lines whereas *miR-183-5p* is undetectable. However, it was shown from the microarray result that both *miR-183-5p* and *miR-183-3p* was highly expressed in TGCT patient samples,

therefore this may suggest that the RT-QPCR was unable to amplify *miR-183-5p* due to technical reasons.

Overall, the miRNA RT-QPCR analysis showed that only *miR-9* expression in TGCT cell lines was in agreement with bioinformatics analysis presented in Chapter 3. Therefore, *miR-9* was chosen for further analysis.

It was therefore necessary to address NANOG regulation at the *miR-9* locus. In order to do this, HEK293T cells were infected with a lentivirus containing a *NANOG* overexpression vector (Figure 4.11 A). The *NANOG* plasmid carried puromycin resistance therefore after cell transduction cells were selected for using puromycin (Figure 4.12). RT-PCR indicated that *NANOG* was overexpressed in puromycin resistant cells, indicating that the *NANOG* plasmid has successfully integrated into theHEK293T host genome (Figure 4.13). In order to test the regulation of NANOG on the *miR-9* promoter, the NANOG bound genomic region upstream of *miR-9* (Figure 4.14) was cloned into a luciferase plasmid (Figure 4.16). Two types of luciferase plasmids were used; pGL3-basic which lacks a constitutive promoter and pGL3-promoter which contained a viral promoter (SV40) this enables analysis of potential positive and negative regulation of the *miR-9* luciferase reporter by NANOG. Following this, both pGL3 plasmids were transfected into HEK293T wild type (WT) cells and HEK293T cells that overexpressed NANOG (NANOG OE). The pGL3-basic plasmid expressed luciferase indicating that the genomic sequence upstream of *miR-9* inserted contained a promoter region (Figure 4.17 A). In addition to this, the pGL3-basic plasmid containing the *miR-9* genomic sequence expressed a reduced level of luciferase in NANOG OE HEK293T cell when compared to

HEK293T WT cells (Figure 4.17 A). This indicated that NANOG bound to this genomic region and to represses transcription of *miR-9*. In contrast to this, the pGL3-promoter plasmid containing the *miR-9* genomic sequence showed no significant difference between WT and NANOG OE HEK293T cells (Figure 4.17 B). HEK293T cells have been transformed by expression of the large T-antigen and it has been shown that the T-antigen binds to the SV40 to stimulate expression (Linzer and Levine, 1979). As both cell lines express high levels of the T-antigen, it is possible that the SV-40 promoter could stimulate luciferase expression to such an extent that NANOG mediated repression is rendered ineffective. Overall, this luciferase reporter experiment indicated that NANOG bound to the *miR-9* promoter in order to repress its expression in HEK293T cells. Corroborating with this, in mESCs that Nanog binds upstream of the *miR-9*-2 transcriptional start sites, although its effects on transcription were not investigated (Loh *et al.*, 2006).

In order to validate NANOG regulation of *miR-9* in TGCT cells, the next step was to knockdown NANOG in a TGCT cell line and measure the effect on *miR-9* expression. To knockdown NANOG, a short-hairpin (sh)RNA approach utilising the pLKO-Tet plasmid was used. Two *NANOG* targeting shRNA sequence and a 'scrambled' sequence (control) were successfully cloned into the pLKO-Tet plasmid (Figure 4.19) and TCam-2 cells were infected with lentivirus under puromycin selection (Figure 4.20). As a result, *NANOG* expression was reduced by 65% and 59% for NANOG shA and NANOG shB respectively, when compared to the scrambled control (Figure 4.21). Given that NANOG shA resulted in the largest reduction in *NANOG* expression, this sample was chosen

for subsequent *miR-9* expression analysis. *MiR-9-5p* and *miR-9-3p* showed an increase in expression of 3 and 4 fold, respectively, when compared to the scrambled control (Figure 4.22). This result demonstrated that NANOG binds to the *miR-9* promoter in order to downregulate its expression in TGCTs. The seminoma cell lines TCam-2 was selected for the NANOG knockdown experiment as *miR-9* has been extensively studied in EC and ES cells. In particular, the addition of retinoic acid to NT2/D1 cells induces high levels of *miR-9* expression (Sempere *et al.*, 2004) and the overexpression of *miR-9*, in combination with *miR-124a*, promotes neural differentiation in ESCs (Lagos-Quintana *et al.*, 2002; Krichevsky *et al.*, 2006). In contrast to this, *miR-9* is expressed in late gonadal PGCs at the time when the PGCs are starting the maturation process (Hayashi *et al.*, 2008). However, the loss or over expression of *miR-9* as well as *miR-9* targets in PGCs remain investigated. Given the similarity of gene expression profiles of that seminomas and gonadal PGCs (Irie *et al.*, 2015), the TCam-2 cell line represented the best choice for the investigation of the role on *miR-9* in the origin of TGCTs.

After showing that NANOG regulates the expression of *miR-9*, the next step was to over express *miR-9* in order to study its function in TCam-2. In order to overexpress *miR-9*, the pLKO-tet plasmid from *NANOG* knockdown (KD) experiment was used. Following this cells were infected and selected for puromycin resistance (Both pLKO-*miR-9-5p* and pLKO-Scrambled). The

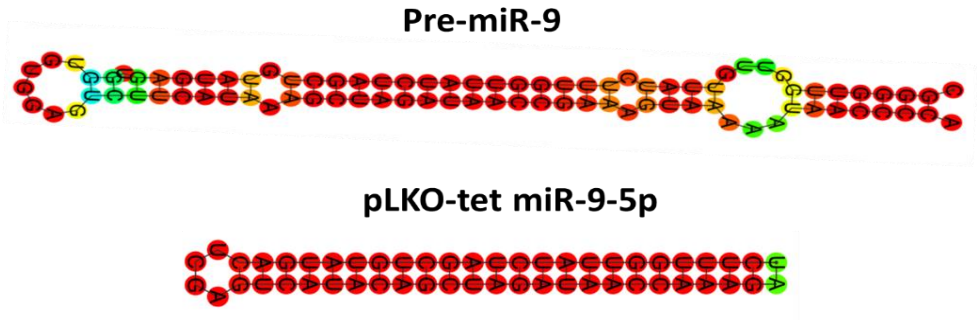


Figure 4.30 is a depiction of the pre-*miR-9*, which contains both *miR-9-5p* and *miR-9-3p* sequences, and the annealed oligos that were cloned into the pLKO-tet plasmid. Colour indicate the binding dynamic for the hairpin loop using the following legend; Red (tightly bound), Orange, Yellow and (green not bound) (Mathews *et al.*, 2004).

expression for miR-9-3p showed no change. This was expected as the *miR-9-5p* oligos are complementary to the *miR-9-5p* sequence but to that of *miR-9-3p*. The *miR-9-5p* expression was instead significantly higher in pLKO-tet-*miR-9-5p* TCam-2 cells when compared to the pLKO-tet-scrambled control (Figure 4.25). However, this represented only a modest 2.5 fold increase. When comparing TGCT cell lines to normal testes (Ambion) that on average *miR-9-5p* and *miR-9-3p* exhibited a 6.5 fold and 8.6 fold reduction in expression, respectively (Figure 4.10 C and D). Therefore, it was determined that a higher level of over expression was required in order to investigate downstream target mRNAs for *miR-9*.

The relatively modest overexpression of *miR-9-5p* achieved could be a result of the hairpin not being processed as effectively as endogenous *miR-9*. It is shown in Figure 4.30 that the pre-*miR-9* (which contains *miR-9-5p* and *miR-9-3p*) shows a series of mismatches within the hairpin loop. Contrasting this, the *miR-9-5p* cloned into the pLKO-tet plasmid has a completely complementary sequence with no mismatches. Therefore, a second approach used an

overexpression vector that expressed the pre-*miR-9* hairpin with endogenous like mismatch containing sequence. A validated retroviral plasmid was used for the expression of *miR-9* (MDH1-PGK-GFP-*miR-9*). This plasmid contained the genomic region for *miR-9* allowing the expression of the pre-*miR-9*, resulting in the overexpression of *miR-9-5p* and *miR-9-3p* (Ma *et al.*, 2010). In order to create a control vector, *miR-9* was removed from the plasmids and a scrambled sequence inserted (Figure 4.26). The plasmid selection marker GFP was expressed in both *miR-9* and scrambled plasmids, indicating the plasmid had successfully integrated into the genome of TCam-2 (Figure 4.27). However, transduced TCam-2 cells showed a heterogeneous population of GFP positive and GFP negative cells. Therefore, it was necessary to use FACS to obtain a homogenous GFP +ve population and by inference *miR-9* overexpressing cells. This proved successful as the RT-QPCR analysis identified that GFP+ *miR-9*-MDH1-TCam-2 cells expressed both *miR-9-5p* and *miR-9-3p* 10 times higher when compared to scrambled-MDH1-TCam-2 and TCam-2 WT (Figure 4.29). These cells were used in the RNA-sequencing experiments described in the next Chapter.

4.3.1 Conclusion

The goal of this chapter was to validate putative NANOG regulated miRNAs identified in Chapter 3. The RT-QPCR expression analysis was performed on TGCT cell lines and compared to a testes control to give a tumour versus control comparison (Figure 4.10), similar to the microarray analysis of Chapter 3 (Figure 3.3). Only *miR-9* expression (Figure 4.10 C and D) was shown consistent with the expression profile identified by the microarray analysis (Table 3.1). In addition to this, the *miR-9* promoter was negatively regulated by NANOG in HEK293T cells (Figure 4.17 A). This was validated in the seminoma cell line TCam-2 as *NANOG* KD resulted in the upregulation for both *miR-9-5p* and *miR-9-3p* (Figure 4.22). This data demonstrate a clear role for NANOG in the transcriptional repression of *miR-9* in TGCTs. In order to investigate *miR-9* targets, a retrovirus vector for the over expression of *miR-9* was used to infect TCam-2 cells. The following chapter used these *miR-9* overexpressing cells in order to investigate *miR-9* targets.

4.4 Supplementary

MIR-26B-5P

```
TTTTTTTTTTTCATCAAAACATCCGAAAAAATCATGTGGTTCAAAAACACGGACCAATGA
-------------------------------------------------------------------------CAATGA
                                                                         ******

AGTGGAATTTCATCACTACAATCGAATTCCCGCGGCCGCCATGGCGGCCGGGAGCATGCG
AGTGGAATTTCATCACTAC ------------------------------------------
*******************
```

SNORD75

```
TTCGGAGTCTTATCTTAAAGTCTTTAGGGAAGACAGGTGTTAGCTTAAGGGCGCCG
-TCGGAGTCTTATCTTAAAGTCTTTAGGGAAGACAGGTG-----------------
  ************************************
```

miR-9-3p

```
CGGCCGCGAATTCACTAGTGATTATAAAGCTAGATAACCGAAAGTAAAAAAAAAAAAAAA
-----------------------ATAAAGCTAGATAACCGAAAGT---------------
                        **********************
```

miR-9-5p

```
CCGAGGTCGACTTCCTAGATTTTTTTTTTTTTTTTTATACAGCTAGATAACCAAAGATAT
-------------------------------------atacagctagataaccaaaga---
                                     ********************
```

miR-512-3p

```
CCTAGATTTTTTTTTTTTTTTTTATTTTTTTGACCTCAGCTATGACAGCACTTAATCAAA
--------------------------------GACCTCAGCTATGACAGCACTT-------
                                 **********************
```

Figure S4.1 Sequencing analysis results for the RT-QPCR amplicon for *SNORD18A*, *SNORD75*, *miR-9-5p*, *miR-9-3p* and *miR-512-3p*.

NANOG

```
ACGGAGACTGTCTCTCCTCTTCCTTCCTCCATGGATCTGCTTATTCAGGACAGCCCTGAT
----------------------------------------------------CAGCCCTGAT
                                                     **********

TCTTCCACCAGTCCCAAAGGCAAACAACCCACTTCTGCAGAGAAGAGTGTCGCAAAAAAG
TCTTCCACCAGTCCCAAAGGCNAACAACCCACTTCTGCAGAGAAGAGTGTCGC-------
******************** ******************************
```

SOX2

```
GAGAGAGATCCTGGACTTCTTTTTGGGGGACTATTTTTGTACAGAGAAAACCTGGGGAGG
---------CCTGGACTTCTTTTTGGGGGACTATTTTTGTACAGAGAAAACCTGGG----47
         *********************************************
```

LIN28A

```
CTACAACTGTGGAGGTCTAGATCATCATGCCAAGGAATGCAAGCTGCCACCCCAGCCCAA
----------------------CNTCATGCCNAGGAATGCAAGCTGCCACCCCAGCCCAA
                      * ******* ****************************

GAAGTGCCACTTCTGCCAGAGCATCAGCCATATGGTAGCCTCATGTC
GAAGTGCCACTTCTGCCAGAGCATCAG--------------------
***************************
```

KDM4C

```
ACCATCATGGAGGTGGCCGAGGTGGAAAGTCCTCTGAACCCCAGCTGTAAGATAATGACC
------------------AGGTGGAAAGTCCTCTGAACCCCAGCTGTAAGATAATGACC
                   ***************************************

TTCAGACCCTCCATGGAGGAGTTCCGGGAGTTCAACAAATACCTTGCATACATGGAGTC
TTCAGACC-----------------------------------------------------
********
```

GAPDH

```
GTGTCTCAGCGATGTGGCTCGGCTGGCGACGCAAAA--AGATGCGGCTGACTGTCGAACA

GTGTCTGAGCGATGTGGCTCGGCTGGCGACGCAAAAGAAGATGCGGCTGACTGTCGAACA
**********************************  **********************


GGAGGAGCAGAGAGCGAAG

GGAGGAGCAGAGAGCGAAG
*******************
```

Figure S4.2 Sequencing results for the RT-QPCR amplicon for *NANOG*, *SOX2*, *LIN28A*, *KDM4C* and *GAPDH*.

```
pGL3 basic sec        TCGATAGGTACCGAGCTCGGCTTTATACATCTCTATAGAGAGAGGCTAAAGAGTCTTGCG
miR-9-2 amplicon      -----------------TGGCTTTATACATCTCTATAGAGAGAGGCTAAAGAGTCTTGCG
                                       ****************************************

pGL3_basic_sec         TGAGAAAGAAGCAATTTACAATGCAAATAGAGGCCTGTGAAGGGACGGATGCTGCTGCTC
miR-9-2_amplicon      TGAGAAAGAAGCAATTTACAATGCAAATAGAGGCCTGTGAAGGGACGGATGCTGCTGCTC
                      ************************************************************

pGL3_basic_sec        ACTCGTAACAGCTTGCTGCACCTTAATAGCTGGTTTTATTTTCTGATATTAATAGATGAA
miR-9-2_amplicon      ACTCGTAACAGCTTGCTGCACCTTAATAGCTGGTTTTATTTTCTGATATTAATAGATGAA
                      ************************************************************

pGL3_basic_sec        TTGTTGGCTAACGCTGCCGGAGATTACTTGCTGAAAGGAAAAAGAAAAGTTCGGAATAAA
miR-9-2 amplicon      TTGTTGGCTAACGCTGCCGGAGATTACTTGCTGAAAGGAAAAAGAAAAGTTCGGAATAAA
                      ************************************************************

pGL3_basic_sec        CTCTGAAGGCAACAGATCTCGTTTGGATGTTTTAGCCGCGCACACACACAAACAGAAACC
miR-9-2_amplicon      CTCTGAAGGCAACAGATCTCGTTTGGATGTTTTAGCCGCGCACACACACAAACAGAAACC
                      ************************************************************

pGL3 basic sec        AGGTCATATTTACATAAAAGGAGGAATCTTAAGCGCGGCAAGTATCCTGGACGACCACTC
miR-9-2_amplicon      AGGTCATATTTACATAAAAGGAGGAATCTTAAGCGCGGCAAGTATCCTGGACGACCACTC
                      ************************************************************

pGL3_basic_sec        TTCGGTACTGCCAGAAAGGATCAGGACCTGGAGTCTGGCAAGAGGAAGACAGAGGCCTGT
miR-9-2_amplicon      TTCGGTACTGCCAGAAAGGATCAGGACCTGGAGTCTGGCAAGAGGAAGACAGAGGCCTGT
                      ************************************************************

pGL3 basic sec        GTGGGAAGCGAGTTGTTATCTTTGGTTATCTAGCTGTATGAGTGTATTGGTCTTCGCTAG
miR-9-2_amplicon      GTGGGAAGCGAGTTGTTATCTTTGGTTATCTAGCTGTATGAGTGTATTGGTCTTC-----
                      ********************************************************



pGL3_promoter_seq      NNNNNNNNNNANNTCTATCGATAGGTACCGAGCTCGGCTTTATACATCTCTATAGAGAG
miR-9-2 amplicon       ----------------------------------TGGCTTTATACATCTCTATAGAGAG
                                                          *************************

pGL3_promoter_seq      GGCTAAAGAGTCTTGCGTGAGAAAGAAGCAATTTACAATGCAAATAGAGGCCTGTGAAGG
miR-9-2_amplicon       GGCTAAAGAGTCTTGCGTGAGAAAGAAGCAATTTACAATGCAAATAGAGGCCTGTGAAGG
                       ************************************************************

pGL3_promoter_seq      GACGGATGCTGCTGCTCACTCGTAACAGCTTGCTGCACCTTAATAGCTGGTTTTATTTTC
miR-9-2 amplicon       GACGGATGCTGCTGCTCACTCGTAACAGCTTGCTGCACCTTAATAGCTGGTTTTATTTTC
                       ************************************************************

pGL3_promoter_seq      TGATATTAATAGATGAATTGTTGGCTAACGCTGCCGGAGATTACTTGCTGAAAGGAAAAA
miR-9-2_amplicon       TGATATTAATAGATGAATTGTTGGCTAACGCTGCCGGAGATTACTTGCTGAAAGGAAAAA
                       ************************************************************

pGL3 promoter seq      GAAAAGTTCGGAATAAACTCTGAAGGCAACAGATCTCGTTTGGATGTTTTAGCCGCGCAC
miR-9-2_amplicon       GAAAAGTTCGGAATAAACTCTGAAGGCAACAGATCTCGTTTGGATGTTTTAGCCGCGCAC
                       ************************************************************

pGL3_promoter_seq      ACACACAAACAGAAACCAGGTCATATTTACATAAAAGGAGGAATCTTAAGCGCGGCAAGT
miR-9-2_amplicon       ACACACAAACAGAAACCAGGTCATATTTACATAAAAGGAGGAATCTTAAGCGCGGCAAGT
                       ************************************************************

pGL3_promoter_seq      ATCCTGGACGACCACTCTTCGGTACTGCCAGAAAGGATCAGGACCTGGAGTCTGGCAAGA
miR-9-2_amplicon       ATCCTGGACGACCACTCTTCGGTACTGCCAGAAAGGATCAGGACCTGGAGTCTGGCAAGA
                       ************************************************************

pGL3 promoter seq      GGAAGACAGAGGCCTGTGTGGGAAGCGAGTTGTTATCTTTGGTTATCTAGCTGTATGAGT
miR-9-2 amplicon       GGAAGACAGAGGCCTGTGTGGGAAGCGAGTTGTTATCTTTGGTTATCTAGCTGTATGAGT
                       ************************************************************

pGL3_promoter_seq      GTATTGGTCTTCGCTAGCCCGGGCTCGAGATCTGCGATCTGCATCTCAATTAGTCAGCAA
miR-9-2_amplicon       GTATTGGTCTTC------------------------------------------------
                       ************
```

Figure S4.3 Sequencing analysis results for the amplicon for the pGL3-Basic-*miR-9*-2 and pGL3-Promoter plasmid identifying the NANOG enriched sequence upstream from *miR-9*-2 TSS was correct.

Chapter 5: Analysis of chosen miRNA

5.1 Introduction

The previous chapter demonstrated that *miR-9* transcription is repressed by NANOG in TCam-2 cells and therefore proliferation in these cells may be regulated through this mechanism. This chapter aimed to identify and investigate *miR-9* targets in TGCTs. The *miR-9* gene is a highly conserved miRNA that has been extensively studied for its role in neural differentiation (Coolen, Katz and Bally-Cuif, 2013). The conservation of *miR-9* extends to *Drosophila* with the *Drosophila miR-9a* gene being identical to those in humans, suggesting an evolutionary conserved role for this gene (Coolen, Katz and Bally-Cuif, 2013). The expression of *miR-9*/9* was high in the developing central nervous system of zebrafish, chicken and mouse (Darnell *et al.*, 2006; Leucht *et al.*, 2008; Zhao *et al.*, 2009; Yoo *et al.*, 2011). In adult mice *miR-9* expression is limited to the brain (Sempere *et al.*, 2004). *In vitro*, the overexpression of *miR-9*/9* in combination with *miR-124* differentiates fibroblasts into neurons, thus supporting the importance of *miR-9* in neurogenesis (Yoo *et al.*, 2011). In fact, when NT2 cells are induced towards neuronal differentiation *miR-9* expression is up-regulated (Sempere *et al.*, 2004). The majority of studies have focused on *miR-9* in neural differentiation, however *miR-9* was also upregulated in late gonadal PGCs, suggesting a role in differentiation of PGCs (Hayashi *et al.*, 2008). Interestingly, *miR-9* has been shown to target *PRMD1* transcripts in targets genes shown to be vital for the PGC development in Hodgkin's lymphoma cells (Nie *et al.*, 2008). This may suggest that *miR-9* may target *PRDM1* in other cell types including PGCs. PRDM1 is key for the specification of mPGCs and hPGCLCs, as *PRDM1* knockdown resulting in complete loss of

PGCs (Ohinata *et al.*, 2005; Saitou *et al.*, 2005; Irie *et al.*, 2015; Kobayashi *et al.*, 2017). In addition to this, *miR-9* targets *Lin28a* and *Lin28b* (Eda *et al.*, 2009; La Torre, Georgi and Reh, 2013), an RNA binding protein expressed in early mouse PGCs but repressed in gonadal PGCs (West *et al.*, 2009).

LIN28 is an important regulator of pluripotency, as its overexpression in combination with NANOG, SOX2 and OCT4 can be used to reprogram somatic fibroblasts into induced pluripotent stem cells (iPSCs) (Yu *et al.*, 2007). In addition, Lin28 also regulates *miR-9* post-transcriptionally in the mouse embryonal carcinoma cell lines (P19) (Nowak *et al.*, 2014). Using SILAC (stable isotope labelling by/with amino acids in cell culture) this study showed that Lin28a was one of the most enriched proteins bound to the *miR-9* sequence (Nowak *et al.*, 2014). Conversely, the knockdown of *Lin28a* using siRNA increased *miR-9* expression indicating that Lin28a negatively regulates *miR-9* transcripts (Nowak *et al.*, 2014). This evidence suggests a role for regulating pluripotency and germ cell genes, which may be important for the development and progression of TGCTs. Both LIN28 and PRDM1 are highly expressed in TGCTs, therefore this data suggests that loss of *miR-9* in TGCTs could be responsible for the acquisition of pluripotency and loss of differentiation.

MiR-9 has also been extensively studied in the context of tumour progression for a variety of cancer types (Coolen, Katz and Bally-Cuif, 2013). *MiR-9* inhibits *CDH1*, the E-cadherin encoding mRNA, to enhance metastasis in breast cancer cells (Ma *et al.*, 2010). However, it also exhibits tumour suppressor characteristics in ovarian serous carcinoma (Tang *et al.*, 2013), collectively this

demonstrated it has oncogenic and tumour suppressive characteristics in different cancer types. However, it is particularly interesting that *miR-9* directly targets *CXCR4* in oral squamous cell carcinoma (OSCC) and its expression is associated with decreased metastasis (Yu *et al.*, 2014). CXCR4 promotes tumour progression in a variety of cancers by influencing metastasis, proliferation and angiogenesis (Knaut *et al.*, 2003; Meier *et al.*, 2007) but is also required for the migration and survival of mouse and zebrafish PGCs (Knaut *et al.*, 2003; Molyneaux *et al.*, 2003).

In this Chapter, bioinformatic methods were used to investigate candidate *miR-9* targets. As mentioned previously, miRNAs contain a 'seed' sequence at the 5' end of the miRNA sequence that has been shown to be energetically favourable for miRNA-target interaction (Dweep *et al.*, 2011). Supporting this, mutations of the miRNA 'seed' sequence results in an inactivated interaction (Doench and Sharp, 2004). Binding of miRNAs to a target sequence results in a reduction of translation through a number of mechanisms including inhibition of translational initiation and elongation. It also induces RNA adenylation which results in increased RNA destabilisation and degradation (Filipowicz, Bhattacharyya and Sonenberg, 2008). The majority of miRNA studies have investigated the targeting of 3' untranslated region (3' UTR). However, miRNAs can target the transcripts 5' UTR and coding sequence (CDS) as well as the gene promoter to regulate gene expression (Lytle, Yario and Steitz, 2007; Tay *et al.*, 2008). Given that the majority of work has focused on miRNAs targeting the 3' UTR, this mechanism was the focus of the analysis in this chapter.

Therefore, the aim of this chapter was to identify and validate *miR-9* in TGCT cell lines. The objectives were:

1. To produce a list of *miR-9* predictive targets using multiple bioinformatics test to yield a consensus *miR-9* target list.

2. To validate *miR-9* binding and inhibition of translation using a luciferase model for predicted candidate *miR-9* targets.

3. To validate *miR-9* targets in TGCT cell lines using gene expression analyses.

5.2 Results

5.2.1 Identifying miRNA Predicted Targets

In order to identify a list of predicted mi-9 targets, the online program miRWalk (v2.0) was used (Dweep and Gretz, 2015). MiRWalk allows the selection of the number of miRNA target prediction tools used for the analysis and performs an overlap between miRNA targets lists for each algorithm in order to give a consensus list of miRNA targets. Therefore, the following miRNA target prediction algorithms were used: miRWalk (Dweep and Gretz, 2015), miRanda (John *et al.*, 2004) miRDB (Wong and Wang, 2014), RNA22 (Miranda *et al.*, 2006) and TargetScan (Lewis *et al.*, 2003). In order to maximise the number of miRNA targets, targets that were identified in at least 3 prediction tools were chosen for further analysis. Using this criterion, there were 2,843 *miR-9-5p* and 1,776 *miR-9-3p* predicted targets. (Figure 5.1 A and B).

My previous results showed that *miR-9* expression is the lowest in TGCTs with high *NANOG* expression (Chapter 3; Table 3.3). For this reason, *miR-9* predicted targets were overlapped with those genes that showed high expression in tumour samples with high *NANOG* expression. This identified 305 and 445 predicted genes for *miR-9-5p* and *miR-9-3p*, respectively (Figure 5.2 A and B). Predicted targets were then selected based on their reported importance for TGCT and/or normal PGC development. From this list, *LIN28A*, *CCNE1, ESRRB, KDM7A, NR5A2* and *KDM4C* were selected for further investigation.

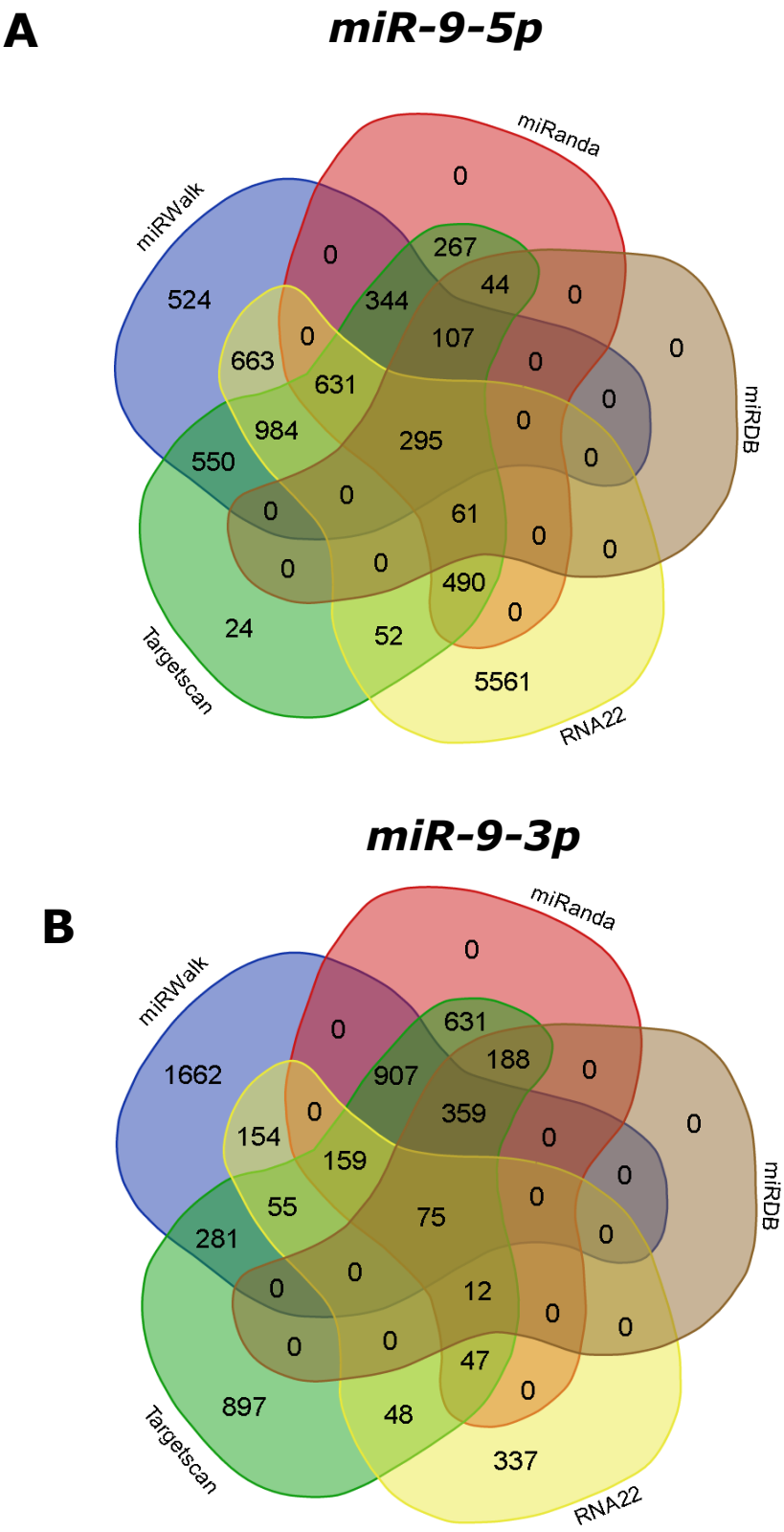


Figure 5.1 Overlap of miRNA targets using five different prediction algorithms for (A) *miR-9-5p* and (B) *miR-9-3p*. Venn diagrams created using Bioinformatics & Evolutionary Genomic Venns (http://bioinformatics.psb.ugent.be/webtools/Venn/).

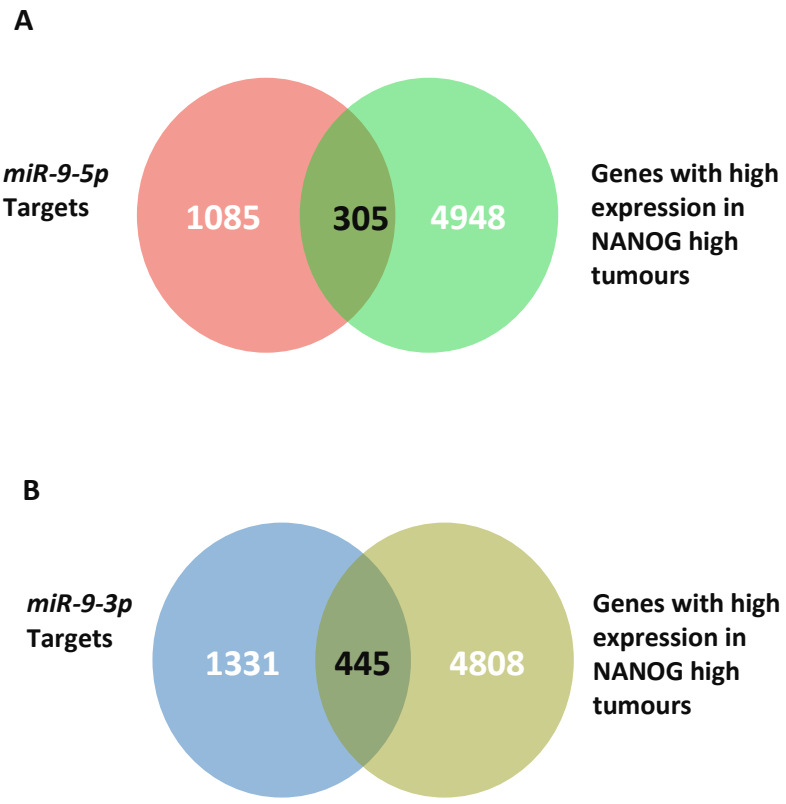


Figure 5.2 Overlap of *miR-9* predicted targets with genes that were highly expressed inTGCT patient samples with high *NANOG* expression; (A) *miR-9-5p* and (B) *miR-9-3p*.

5.2.2 Validation of *miR-9* Targets (293T)

As mentioned previously, a predicted target list for six *miR-9* targets were chosen for further analysis. In order to test whether *miR-9* binds target sequences, a miRNA luciferase reporter assay was performed using HEK293T cells. To perform this experiment, the 3'UTRs from each of the transcripts predicted to be targeted by *miR-9* were amplified using PCR with added restriction sites for XhoI and NotI and the fragment was sub-cloned into psiCHECK-2 plasmid (Figure 5.3 A and B). The psiCHECK-2 plasmid carries a SV40 promoter followed by *Renilla* luciferase which can be tagged with the miRNA target sequences, together with a separate HSV-TK promoter followed by the *firefly* luciferase (Figure 5.3). The psiCHECK-2 plasmid also carries a gene for ampicillin resistance. The purified PCR product for the predicted *miR-9* targets were digested using XhoI and NotI and the digest fragment was sub-cloned into the psiCHECK-2 plasmid and transformed into bacterial cells. Ampicillin resistant colonies were selected and grown in overnight cultures. The psiCHECK-2 plasmids were verified to have insertions by restriction enzyme digestion using XhoI and NotI. Predicted 3' UTR *miR-9* targets were amplified and individual psiCHECK-2 reporter constructs were generated for each gene. *KDM4C, NR5A2_1, NR5A2_2, ESRRB, KDM7A, CCNE1* and *LIN28A* plasmids contained the correct sized fragment, as shown in Table 5.1 and Figure 5.4. Plasmids were sequenced to confirm the correct sequence was inserted into the plasmid (Figure S5.1).

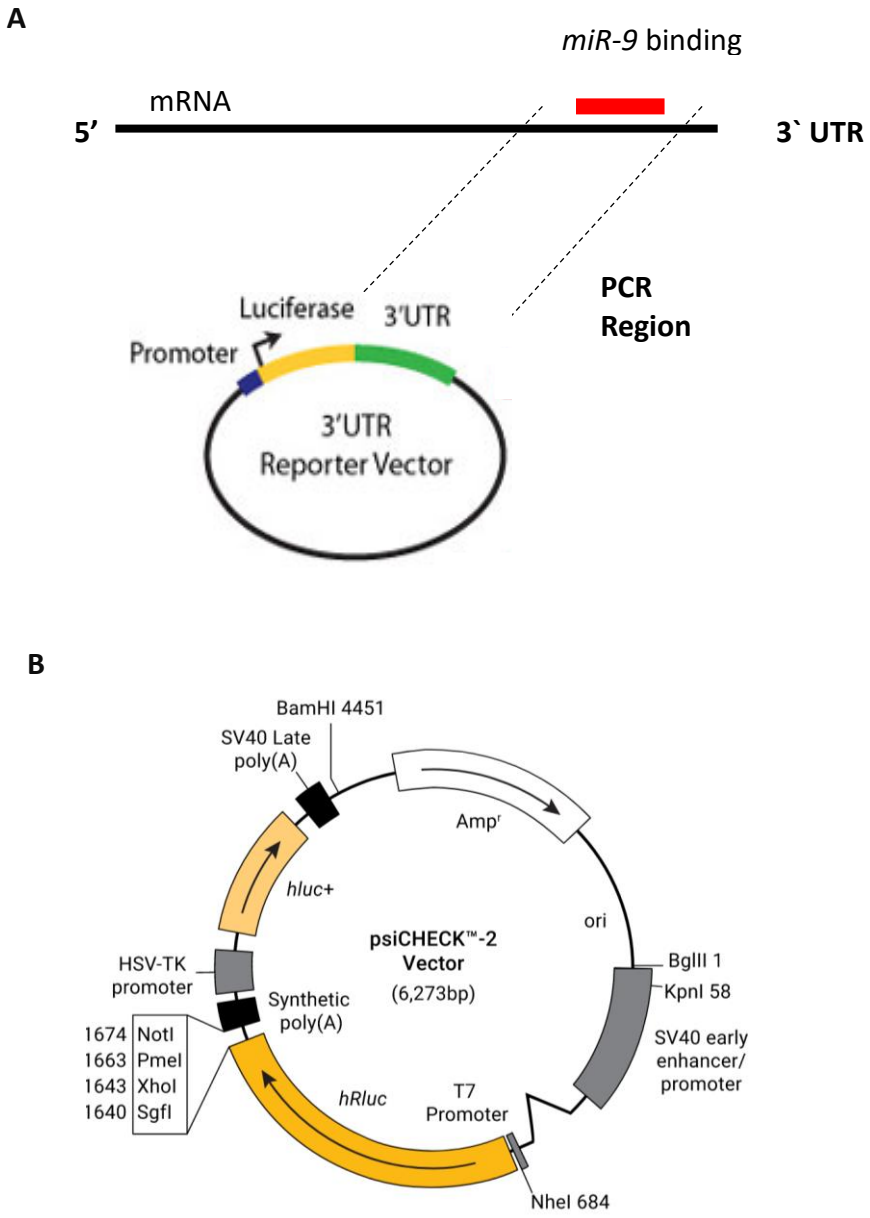


Figure 5.3 (A) A depiction of amplifying the *miR-9* target region using PCR and inserting this downstream of a luciferase gene. (B) The map for the PSICHECK-2 plasmid; vector maps were adapted from the Promega technical manual 'siCHECK Vectors' https://www.promega.co.uk/-/media/files/resources/protocols/technical-bulletins/101/sicheck-vectors-protocol.pdf?la=en.

Table 5.1 The predicted *miR-9* target gene list and the expected fragment size cloned into psiCHECK-2.

Gene of Interest	Expected Size
KDM4C	399bp
NR5A2_1	189bp
NR5A2_2	601bp
ESRRB	173bp
KDM7A	317bp
CCNE1	203bp
LIN28A	158bp

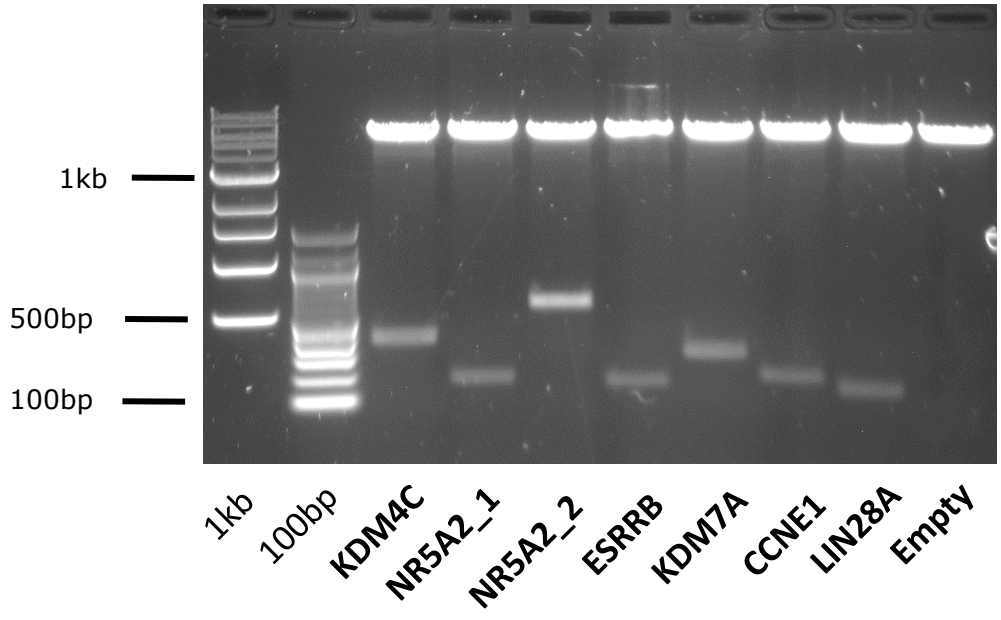


Figure 5.4 PsiCHECK-2 plasmids were cut with XhoI and NotI to show the inserts of correct sizes were successfully ligated.

Once the identity of the psiCHECK-2 plasmids were confirmed to contain the 3' UTR for the chosen *miR-9* predicted targets, they were transfected into 293T cells in combination with *miR-9-5p* or *Scrambled* oligonucleotides using lipofectamine 2000. The *firefly* luciferase was used to normalise luciferase activity in this experiment. Two different 3' UTR target sites for *NR5A2* (*NR5A2*_1 and *NR5A2*_2) were downregulated by approximately 70% when *miR-9-5p* was transfected when compared to the control (*Scrambled*) ($p<0.05$) (Figure 5.5). *MiR-9-5p* also reduced luciferase expression for the *KDM4C* plasmid by approximately 25% ($p<0.05$), whereas *miR-9* had was no significant effect on the *ESRRB* reporter (Figure 5.5). *MiR-9-5p* targeted *LIN28A* to give a 70% reduction ($p<0.05$) in addition to *KDM7A* and *CCNE1* giving a 50% reduction in expression ($p<0.05$) (Figure 5.6).

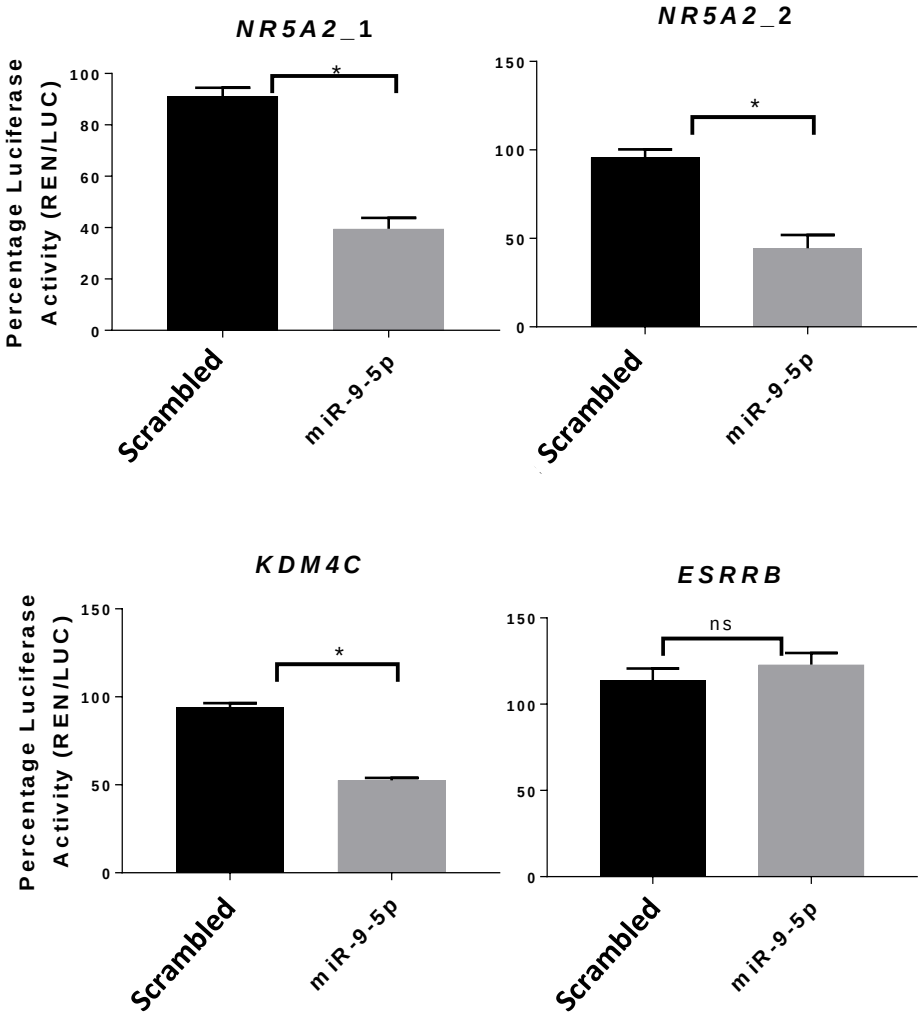


Figure 5.5 PsiCHECK-2 plasmids containing the *miR-9* predicted target sequence for *NR5A2* (two sites *NR5A2_1* and *NR5A2_2*), *KDM4C* and *ESRRB* were transfected into 293T cells in combination with either a *Scrambled* sequence (control) or with *miR-9-5p*. This experiment consisted of 6 biological replicates (n=6). A two-tailed student t-test was used to compare samples, * indicates $p<0.05$. Error bars indicated mean ± standard error mean.

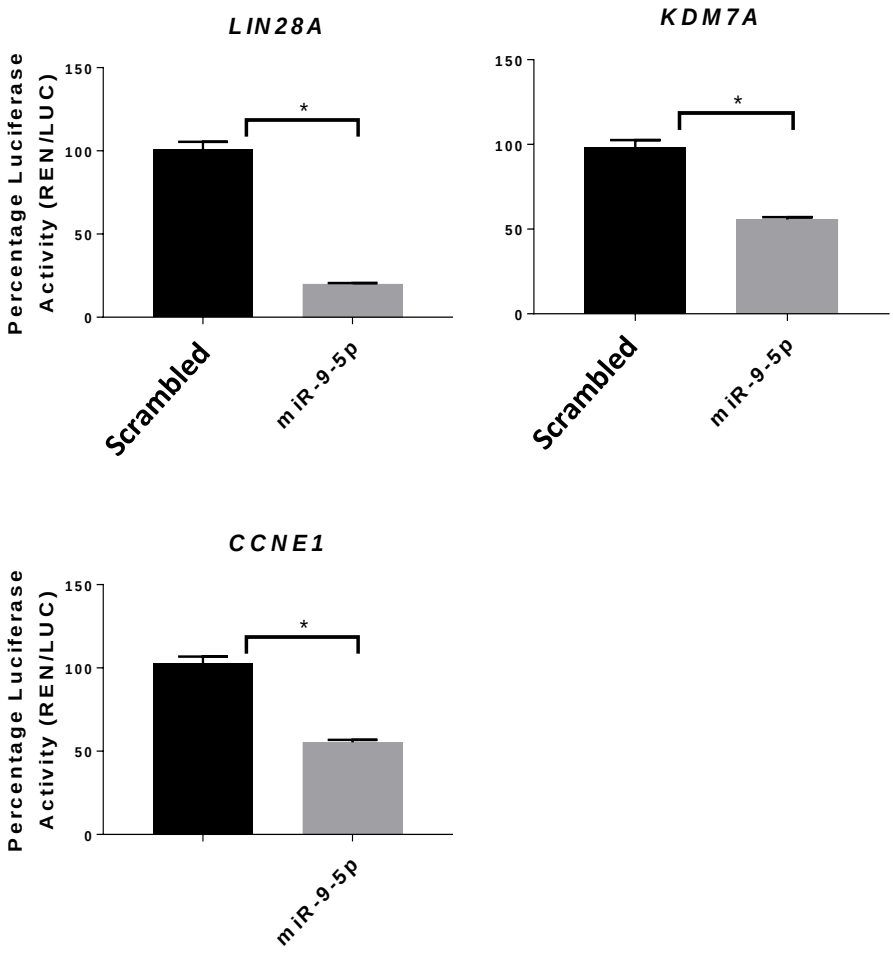


Figure 5.6 PsiCHECK-2 plasmids containing the *miR-9* predicted target sequence for *LIN28A* (n=6), *KDM7A* (n=6) and *CCNE1* (n=3) were transfected into 293T cells in combination with either a *Scrambled* sequence (control) or with *miR-9-5p*. A two-tailed student t-test was used to compare samples, * indicates $p<0.05$. Error bars indicated mean ± standard error mean.

5.2.3 Validation of *miR-9* Targets in TGCTs

Following the analysis of *miR-9* targets in 293T cells, the next step was to validate these results in TGCT cell lines. To do this, TCam-2 cells over expressing *miR-9* (as described Chapter 4, Figure 4.25) were analysed for expression of the targets using RT-QPCR. *LIN28A* and *KDM4C* showed no significant differences in expression when comparing TCam-2 over expressing *miR-9* to TCam-2 Scrambled (control) cells (Figure 5.7). Expression analysis was also performed for *NR5A2*, *CCNE1*, *KDM7A* and *ESRRB*, however RT-QPCR detected no expression of these genes (Figure S4.1).

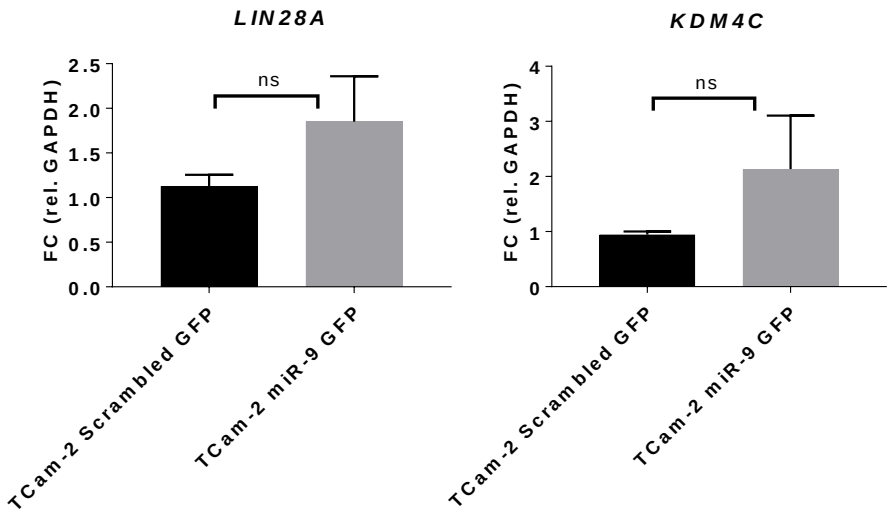


Figure 5.7 Expression analysis for *LIN28A* and *KDM4C* in TCam-2 cells that overexpressed *miR-9* (grey) compared to a scrambled control (black). This experiment consisted of two biological replicates (n=2). A two-tailed student t-test was used to compare samples. Error bars indicated mean ± standard error mean.

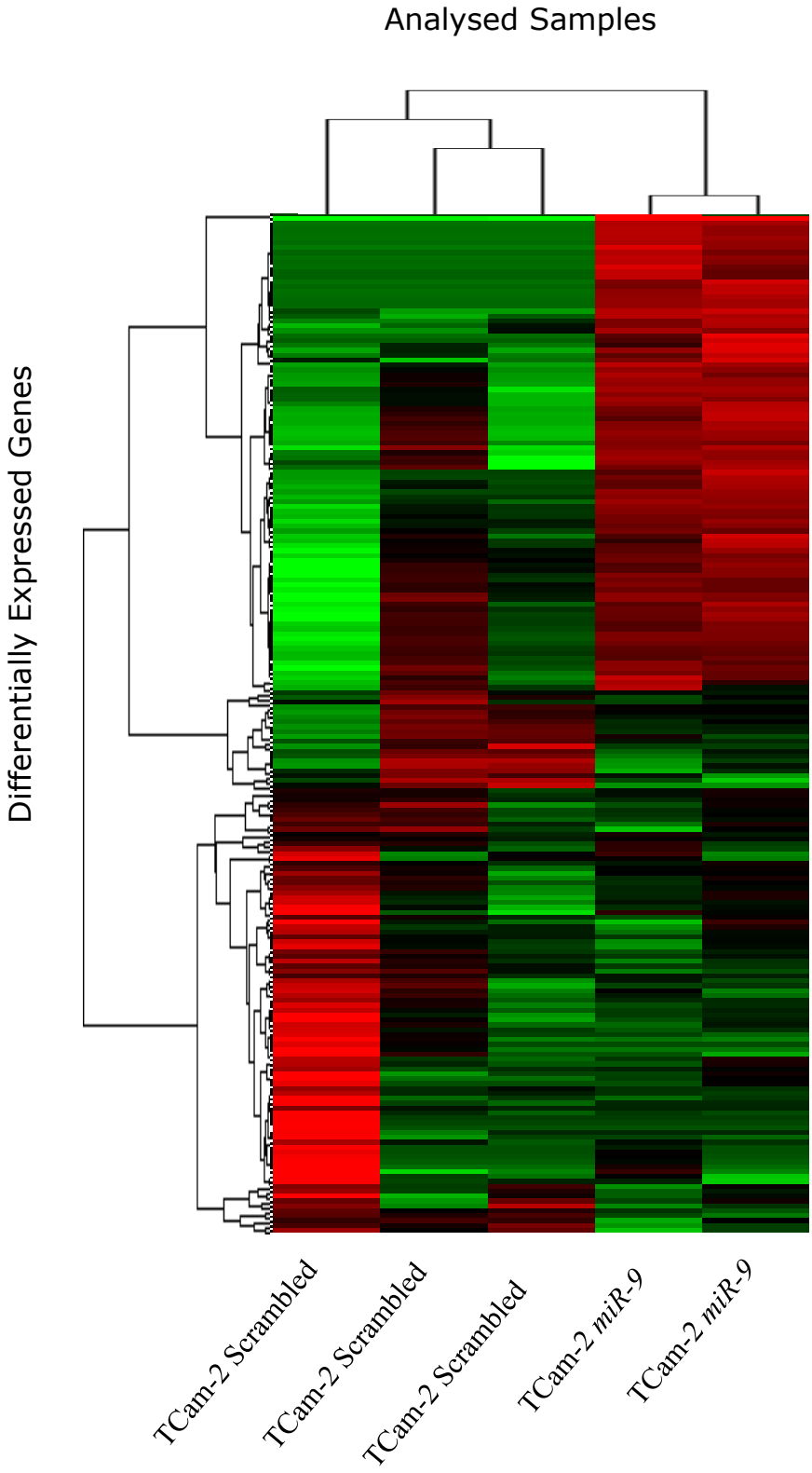


Figure 5.8 A heatmap of 209 differentially expressed genes identified in TCam-2 Scrambled (n=3) and TCam-2 *miR-9* (n=2) groups. The top dendrogram shows TCam-2 samples subject to unsupervised hierarchical clustering; the vertical dendrogram shows differentially expressed genes. EdgeR was used to calculate statistical significance for differentially expressed genes.

Table 5.2 Top 20 significantly differentially expressed genes when comparing TCam-2 Scrambled and TCam-2 *miR-9*.

Ensembl ID	Gene symbol	Scrambled vs *miR-9*_OE (LogFC)	False Discovery Rate (FDR)	
ENSG00000207819	*miR-9-3*	-9.26	0.00000	Genes higher in *miR-9* OE TCam-2
ENSG00000196337	*CGB7*	-6.65	0.00034	
ENSG00000186973	*FAM183A*	-6.39	0.00320	
ENSG00000279653	*AC004678.2*	-6.32	0.00320	
ENSG00000185267	*CDNF*	-6.32	0.00624	
ENSG00000125122	*LRRC29*	-6.32	0.00624	
ENSG00000150873	*C2ORF50*	-6.24	0.00624	
ENSG00000170893	*TRH*	-6.24	0.00624	
ENSG00000204950	*LRRC10B*	-6.24	0.00624	
ENSG00000250892	*AC108475.1*	-6.16	0.01197	
ENSG00000255356	*AC007536.1*	4.69	0.00000	Genes lower in *miR-9* OE TCam-2
ENSG00000205609	*EIF3CL*	4.71	0.00000	
ENSG00000204382	*XAGE1B*	4.87	0.00000	
ENSG00000249264	*EEF1A1P9*	5.18	0.00000	
ENSG00000240087	*RPSAP12*	5.54	0.00000	
ENSG00000235508	*RPS2P7*	5.57	0.03068	
ENSG00000242858	*AC020899.1*	5.57	0.04009	
ENSG00000263887	*AC053481.2*	5.98	0.00396	
ENSG00000189051	*RNF222*	6.61	0.00777	
ENSG00000215093	*EEF1A1P29*	6.72	0.00000	

5.2.4 Transcriptomic Analysis

As the experiments presented above demonstrated that none of the predicted targets were repressed in TCam-2 cells overexpressing *miR-9*, transcriptomic analysis was next performed to identify global transcriptomic changes occurring in response to *miR-9* overexpression.

Using next generation sequencing RNA-sequencing (RNA-seq), the global gene expression was compared in WT TCam-2 and *miR-9* overexpressing TCam-2. To this end, three biological replicates of each *miR-9* overexpressing and scrambled control cells were analyzed by RNAseq. However, one of the *miR-9* samples failed the sequencing and therefore could not be included in the analysis. The analysis demonstrated a total of 209 genes differentially expressed between the TCam-2 Scrambled and TCam-2 *miR-9* samples (Figure 5.8 and Table 5.2). The top 20 differentially expressed genes are shown in Table 5.2. Importantly, *miR-9* was the most highly differentially expressed in TCam-2 *miR-9* cells when compared to TCam-2 Scrambled (Table 5.2). Unsupervised hierarchical clustering of the differentially expressed genes separated the samples into two main branches, TCam-2 Scrambled and TCam-2 *miR-9* (Figure 5.8) but did not identify a particular subset of genes that clustered differently in these two sample types.

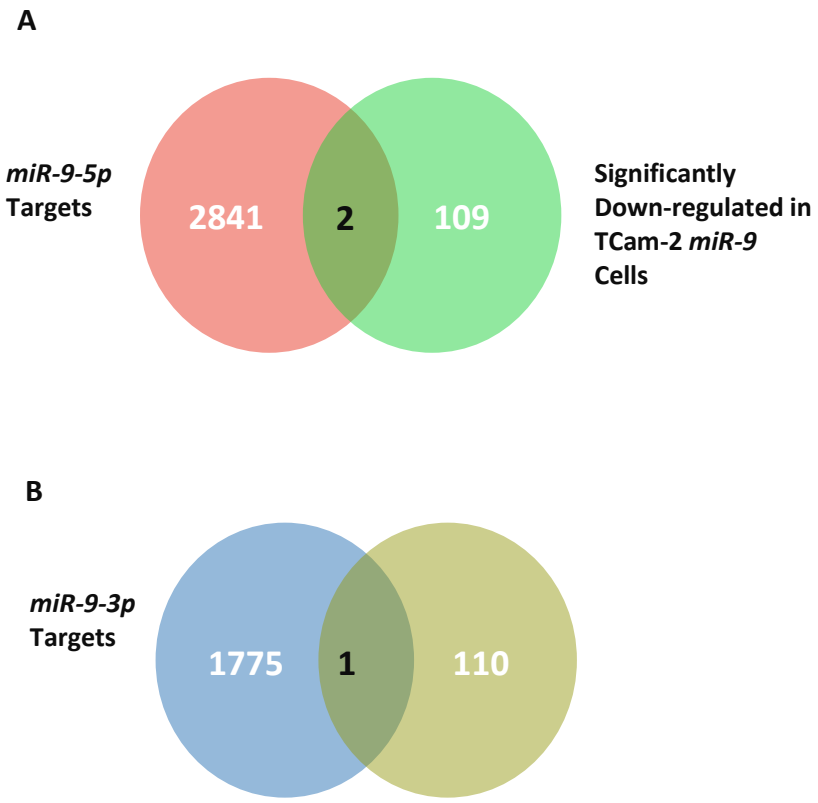


Figure 5.9 The overlap of *miR-9-5p* (A) and *miR-9-3p* (B) predicted targets with genes that were lowly expressed TCam-2 *miR-9* cells.

Following the analysis of the RNA-seq samples, the next step was to overlap genes that were downregulated in TCam-2 *miR-9* cells with the putative *miR-9* targets identified by the targets prediction tools (Figure 5.9). Only 3 predicted genes were significantly downregulated in TCam-2 *miR-9* compared to the Scrambled control (Figure 5.9). These were *SIK1*, *CD34* and *SMN2*, as described in Table 5.3.

Table 5.3 Genes downregulated in TCam-2 *miR-9* cells that were also predicted targets of *miR-9-5p* or *miR-9-3p*.

Ensembl ID	miRNA Target	Gene Symbol	Description
ENSG00000142178	*miR-9-5p*	*SIK1*	Salt inducible kinase 1
ENSG00000174059	*miR-9-5p*	*CD34*	CD34 molecule
ENSG00000205571	*miR-9-3p*	*SMN2*	Survival of motor neuron 2, centromeric

To determine whether the list of differentially expressed genes were part of defined biological processes, molecular functions and gene interactions, an Overrepresentation Enrichment Analysis was performed using the WebGestalt software (Zhang, Kirov and Snoddy, 2005). First, a gene ontology analysis was next used to investigate the functions of the differentially expressed genes. The top 3 ontologies for biological processes were matched to genes which regulate the transport of proteins to the Golgi apparatus (Table 5.4). The *RGPD3, RGPD6* and *RGPD2* were represented in all three biological processes. In addition to this, two genes involved in retinal cone cell development and retinoid differentiation were also identified (Table 5.4).

Next, pathway analysis was used to identify genes associated with well characterized molecular pathways. Using the Reactome Pathway Database identified several pathways which were involved in hormonal signalling. The key genes involved were *CYP19A1, CGA* and *POMC* (Table 5.5). KEGG pathway analysis also highlighted genes that were enriched for hormonal signalling (Table 5.6). The top pathway identified using KEGG analysis was the hematopoietic cell lineage which was enriched for genes that encoded cell surface antigens (*CD38, CD24* and *CD14*) (Table 5.6).

Table 5.4 Gene ontology (GO) for biological processes for genes differentially expressed in TCam-2 Scrambled and TCam-2 *miR-9*.

Gene Ontology	Name	Class	Observed	Expected	Enrichment	P value	FDR	Genes
BP:GO:0000042	protein targeting to Golgi	18	3	0.06	50.18	<0.0001	1.89E-01	RGPD3;RGPD6;RGPD2
BP:GO:0072600	establishment of protein localization to Golgi	21	3	0.07	43.01	<0.0001	1.89E-01	RGPD3;RGPD6;RGPD2
BPGO:0000301	retrograde transport, vesicle recycling within Golgi	24	3	0.08	37.63	<0.0001	1.90E-01	RGPD3;RGPD6;RGPD2
BP:GO:0034067	protein localization to Golgi apparatus	34	3	0.11	26.57	<0.0001	3.20E-01	RGPD3;RGPD6;RGPD2
BP:GO:0051270	Regulation of cellular component movement	790	10	2.62	3.81	<0.0001	3.20E-01	CGA;COL1A1;ADRA2A;CYP19A1;GNRH1;HAS2;PGAM4;PDE4D;CEMIP;THY1
BP:GO:0032845	negative regulation of homeostatic process	174	5	0.58	8.65	<0.0001	3.20E-01	ADRA2A;PDE4D;CEMIP;THY1;CD38
BP:GO:0042670	retinal cone cell differentiation	8	2	0.03	75.27	<0.0001	3.20E-01	DIO3;THY1
BP:GO:0046549	retinal cone cell development	8	2	0.03	75.27	<0.0001	3.20E-01	DIO3;THY1
BP:GO:0006891	intra-Golgi vesicle-mediated transport	47	3	0.16	19.22	<0.0001	4.58E-01	RGPD3;RGPD6;RGPD2
BP:GO:1903169	regulation of calcium ion transmembrane transport	113	4	0.38	10.66	<0.0001	4.58E-01	ADRA2A;PDE4D;CEMIP;THY1

Table 5.5 Pathways identified using the Reactome Pathway Database for genes differentially expressed in TCam-2 Scrambled and TCam-2 miR-9.

Gene Ontology	Name	Class	Observed	Expected	Enrichment	P value	FDR	Genes
R-HSA-2142850	Hyaluronan biosynthesis and export	5	2	0.02	114.7	0.0001	0.139	*HAS2;CEMIP*
R-HSA-196071	Metabolism of steroid hormones	31	3	0.11	27.75	0.0002	0.139	*CGA;CYP19A1;POMC*
R-HSA-209968	Thyroxine biosynthesis	10	2	0.03	57.35	0.0005	0.239	*CGA;DIO3*
R-HSA-193048	Androgen biosynthesis	11	2	0.04	52.14	0.0006	0.175	*CGA;POMC*
R-HSA-375281	Hormone ligand-binding receptors	12	2	0.04	47.79	0.0008	0.239	*CGA;GNRH1*
R-HSA-209952	Peptide hormone biosynthesis	14	2	0.05	40.96	0.001	0.273	*CGA;POMC*
R-HSA-211897	Cytochrome P450 - arranged by substrate type	65	3	0.23	13.23	0.0015	0.303	*CYP19A1;POMC;CYP2W1*
R-HSA-2142845	Hyaluronan metabolism	17	2	0.06	33.73	0.0016	0.303	*HAS2;CEMIP*
R-HSA-209776	Amine-derived hormones	18	2	0.06	31.86	0.0017	0.303	*CGA;DIO3*
R-HSA-211976	Endogenous sterols	26	2	0.09	22.06	0.0036	0.614	*CYP19A1;POMC*

Table 5.6 Pathway identified using the KEGG Pathway Database for genes differentially expressed in TCam-2 Scrambled and TCam-2 miR-9.

Name	Class	Observed	Expected	Enrichment	P value	FDR	Genes
Hematopoietic cell lineage - Homo sapiens (human)	94	4	0.44	9.06	0.001	1.60E-01	*HLA-DRB5;CD14;CD34;CD38*
RNA transport - Homo sapiens (human)	171	5	0.8	6.22	0.001	1.60E-01	*RGPD3;SMN2;EIF3CL;RGPD2;RGPD5*
Glycine, serine and threonine metabolism - Homo sapiens (human)	40	2	0.19	10.64	0.015	1.00E+00	*CBSL;PGAM4*
Ovarian steroidogenesis - Homo sapiens (human)	50	2	0.23	8.51	0.023	1.00E+00	*CGA;CYP19A1*
Autoimmune thyroid disease - Homo sapiens (human)	50	2	0.23	8.51	0.023	1.00E+00	*CGA;HLA-DRB5*
Biosynthesis of amino acids - Homo sapiens (human)	74	2	0.35	5.75	0.047	1.00E+00	*CBSL;PGAM4*
Salivary secretion - Homo sapiens (human)	90	2	0.42	4.73	0.066	1.00E+00	*CST1;CD38*
GnRH signaling pathway - Homo sapiens (human)	92	2	0.43	4.63	0.069	1.00E+00	*CGA;GNRH1*
Amoebiasis - Homo sapiens (human)	99	2	0.47	4.3	0.078	1.00E+00	*COL1A1;CD14*
Glucagon signaling pathway - Homo sapiens (human)	102	2	0.48	4.17	0.082	1.00E+00	*SIK1;PGAM4*

Table 5.7 Differentially expressed genes (TCam-2 Scrambled vs TCam-2 *miR-9*) that were also enriched in TCGA TGCT patient samples.

Gene Symbol	Random Walk Probability	Gene Symbol	Random Walk Probability	Gene Symbol	Random Walk Probability
RGPD3	0.012457795	*GFRA1*	0.00873514	*LRRC29*	0.008743032
RGPD6	0.01143915	*CYS1*	0.008715704	*POMC*	0.008655178
RGPD5	0.010814264	*FAM167A*	0.008701902	*HIST1H2AE*	0.008653203
HLA-DRB5	0.009351256	*SLC43A1*	0.00869856	*RCSD1*	0.008652696
NOMO3	0.009285313	*HAND1*	0.008697185	*PRODH*	0.008650505
CST1	0.009262145	*TRH*	0.008697082	*ADRA2A*	0.008648462
RGS7	0.009162544	*CDHR1*	0.008696998	*CYP2W1*	0.008645162
KIAA1024	0.009073871	*PDE4D*	0.008689967	*CD14*	0.008636607
LRRN4	0.009072931	*LCA5L*	0.00868276	*PWP2*	0.008630784
TRPC5	0.009031232	*RASL12*	0.008674295	*COL1A1*	0.008622119
CGB7	0.009017402	*SMN2*	0.008661604	*CD34*	0.008617282
FOSB	0.008987837	*FHL1*	0.00865532	*DIO3*	0.008611163
LANCL3	0.008937134	*HAS2*	0.00880989	*ZMAT1*	0.008606495
GNRH1	0.008925414	*MAGED4B*	0.008800959	*PGAM4*	0.008603878
FIGNL2	0.008900705	*HVCN1*	0.008800641	*FILIP1L*	0.008597562
GTF2IP1	0.008874583	*DES*	0.008788739	*ZNF483*	0.008593204
FAM156A	0.008854833	*SYTL5*	0.008781617	*POU5F1B*	0.00858897
THY1	0.008840817	*GOLT1A*	0.008765504	*CHST7*	0.00857764
CD38	0.008831939	*CDNF*	0.008749259	*EIF3CL*	0.008571551
SUMO1P3	0.008821116	*LRRC10B*	0.008745842		

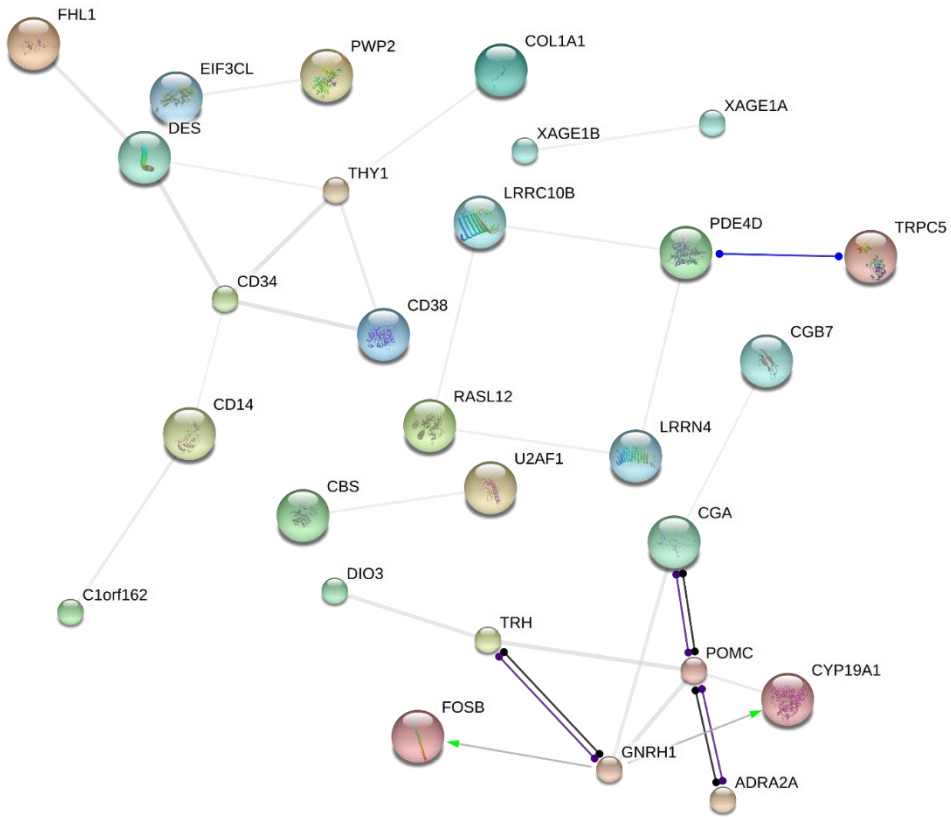


Figure 5.10 STRING analysis (Szklarczyk *et al.*, 2014) for genes differentially expressed between TCam-2 Scrambled and TCam-2 *miR-9* groups. Green arrows indicate activation, purple lines indicate catalysis, black lines indicate a reaction between proteins and blue lines indicate similarity in amino acid sequences.

The next level of analysis was to investigate possible interactions among genes that were differentially expressed. Gene network analysis using the TCGA TGCT functional network database showed that 59 differentially expressed genes were also enriched in TGCT patient samples (Table 5.6). *GNRH1* and *POMC* which were identified by the gene ontology analyses (Table 5.4 and Table 5.5) were also enriched in the TCGA TGCT patient samples (Table 5.7). Interestingly, a pseudogene for POU5F1 (OCT4) called *POU5F1B* was also enriched in TCGA samples (Table 5.6).

Following this, differentially expressed genes were analysed by STRING (v10.5) in attempt to elucidate protein-protein interactions (Szklarczyk et al., 2014). This analysis identified a collection of nodes which comprised 7 genes. These genes were *FOSB, TRH, GNRH1, POMC, CYP19A1, ADRA2A* and *CGA* (Figure 5.10). Following this, principle component analysis (PCA) using RNA-seq data from this study (TCam-2 Scrambled and TCam-2 *miR-9*) in addition to samples derived from gonadal human PGCs (hPGCs), embryonic somatic cells (Soma), hESCs (cultured in 4i medium) and TCam-2 (WT) (Figure 5.11) from a study conducted by Irie and colleagues (Irie *et al.*, 2015). Given that TCam-2 cells resembles hPGCs (Irie *et al.*, 2015), TCam-2 cells overexpressing *miR-9* were compared to hPGCs, ESCs and somatic cells to investigate if *miR-9* overexpression induced global gene expression towards any of these cell types. The soma and hPGCs clustered furthest away from TCam-2 WT, Scrambled and *miR-9* cells, whereas hESCs clustered more closely (Figure 5.11). It was also shown that TCam-2 WT, TCam-2 Scrambled and TCam-2 *miR-9* clustered most closely, indicating that their gene expression profiles were the most similar (Figure 5.11).

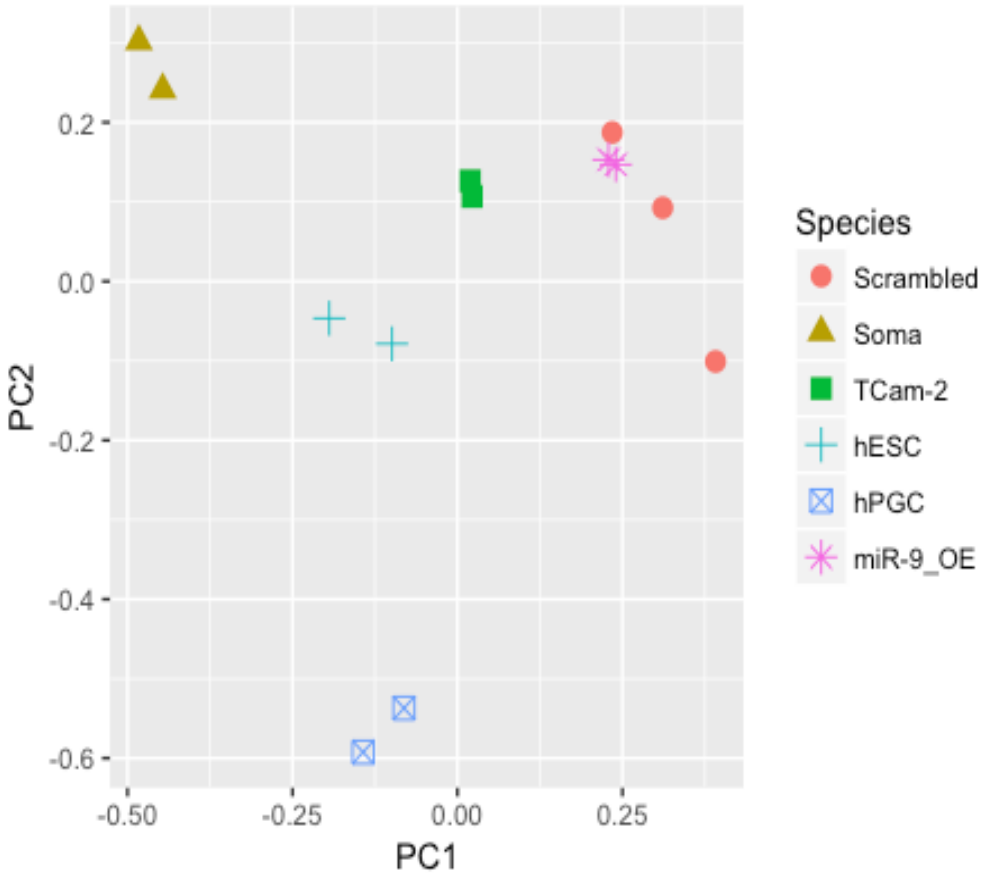


Figure 5.11 A principle component analysis (PCA) plot for gene expression using RNA-seq samples for TCam-2 Scrambled (red circle) and TCam-2 *miR-9* (pink star) in addition to available RNA-seq datasets (GSE60138) for the samples from embryonic somatic tissue (yellow triangle), WT TCam-2 (green square), hESCs (blue cross) and human gonadal PGCs (blue square cross).

5.3 Discussion

This chapter aimed to identify novel targets for *miR-9* in TCam-2 TGCTs and to validate these targets for TGCTs. In order to identify *miR-9* targets, a number of available tools were used in combination to generate a consensus list of *miR-9-5p* and *miR-9-3p* predictive targets: miRWalk (Dweep and Gretz, 2015), miRanda (John *et al.*, 2004) miRDB (Wong and Wang, 2014), RNA22 (Miranda *et al.*, 2006) and TargetScan (Lewis *et al.*, 2003) were used to calculate *miR-9* predicted targets, a total of 295 and 75 predicted targets for *miR-9-5p* and *miR-9-3p*, respectively, were verified by all 5 algorithms used (Figure 5.1 A and B). *MiR-9* targets that were identified by 3 or more algorithms were taken for further analysis this resulted in 2,843 predicted targets for *miR-9-5p* and 1,776 for *miR-9-3p* (Figure 5.1 A and B). This list of predicted *miR-9* targets was overlapped with genes that were highly expressed in TCGA TGCT samples with *NANOG* high groups (Figure 2.5 A and B). This method was used as it has been shown in Chapter 3 that samples with high *NANOG* expression show a significantly lower expression of *miR-9* (Figure 3.9). Therefore, as miRNAs are known to target transcripts for degradation it is expected that samples with low *miR-9* expression would have a higher expression of the predicted targets. Overlapping predicted *miR-9* targets with genes which are highly expressed in *NANOG* groups resulted in 305 and 445 genes that were present in both groups (Figure 5.2 A and B). This list was examined manually to identify genes that may be important for germ cell development, TGCTs, pluripotency or proliferation.

The predicted *miR-9* targets chosen for validation were *LIN28A, CCNE1, ESRRB, KDM7A, NR5A2* and *KDM4C,* these transcripts were predicted to be targeted by *miR-9-5p*. These genes were chosen based on their known function in pluripotency and proliferation. As mentioned previously, *Lin28a* has been shown to be important for ESC pluripotency and the maintenance of mouse PGCs during development (Eda *et al.,* 2009; West *et al.,* 2009). In fact, it has been shown that *miR-9* targets the 3' UTR of *Lin28a* in the mouse embryonal carcinoma cell line P19, resulting in its downregulation (Eda *et al.*, 2009). In addition to this, both *ESRRB* and *NR5A2* are important for pluripotency as the overexpression of either will reprogram somatic cells to iPSCs (Feng *et al.*, 2009; Heng *et al.*, 2010). In addition to this, in mESCs Nanog binds directly to the promoter of Esrrb to promote its expression (Festuccia *et al.*, 2012). Esrrb over expression can rescue and stabilize pluripotency and self-renewal in mESCs for which *Nanog* expression was disrupted. This indicates that Esrrb is a key downstream regulator for Nanog involved in self-renewal and pluripotency in mESCs (Festuccia *et al.*, 2012). Supporting this, Esrrb, as well as Nanog, can bind to the proximal 5' UTR of Oct4 in mESCs in order to activate its transcription (Zhang *et al.,* 2008). In contrast to this, Ccne1 is a cyclin that functions as a regulatory subunit for Cdk2 and, in mESCs, is shown to be constitutively expressed in order to keep the cell cycle primed for DNA replication and sustaining a high proliferative index (Stead *et al.*, 2002). It has also been shown that in a variety of cancers that CDK2 and CCNE1 are highly expressed and associated with promoting proliferation in these cells (Bai *et al.*, 2017; Shi *et al.*, 2017). In contrast to this, KDM7A and KDM4C are enzymes that catalyse histone demethylation, more specifically KDM7A is specific for

K3K9me2 and H3K27me2 demethylation, whereas KDM4C is specific H3K9me2/me3 and H3K36me2/me3 demethylation (Berry and Janknecht, 2013; Pan *et al.*, 2015). Exit from the self-renewing state in ESCs is accompanied with an induction of repressive H3K9me2/me3 marks, Kdm4C will reverse these epigenetic modifications at the promoters for key pluripotency genes including *Nanog* to promote self-renewal in mESCs (Loh *et al.*, 2007). Therefore, KDM4C and KDM7A may be important for the activation and maintenance of expression of key pluripotency genes. In support of this, it has been shown in human oesophageal squamous cell carcinoma that KDM4C epigenetically enhances the expression of *SOX2.* While KDM7A is relatively less well investigated however has been shown to regulate differentiation in mESCs and promote proliferation in prostate cancer (Huang *et al.*, 2010; Kang *et al.*, 2017). From this evidence, it was determined that these targets were to be investigated further.

To test these predicted targets, the predicted *miR-9* binding sites for *miR-9* in the 3' UTR for *LIN28A, CCNE1, ESRRB, KDM7A, NR5A2* and *KDM4C* were inserted into a plasmid (psiCHECK-2) downstream of a luciferase reporter (Figure 5.4). These plasmids were co-transfected with either *miR-9-5p* or *Scrambled* into HEK293T cells. Plasmids containing *LIN28A, CCNE1, KDM7A, NR5A2* and *KDM4C* 3' UTRs exhibited a significant reduction ($p<0.05$) in luciferase expression in cells co-transfected with *miR-9-5p* when compared to the Scrambled control (Figure 5.5 and 5.6). This data showed that *miR-9-5p* repressed the translation of these targets by binding to the 3' UTR sequence into the plasmid in this 293T model system, however this interaction required

validation in TGCTs. In order to test this, TCam-2 cells which over expressed *miR-9* from Chapter 4 (Figure 4.25) were analysed for gene expression using RT-QPCR. *LIN28A* and *KDM4C* showed no significant differences between TCam-2 Scrambled (control) when compared to TCam-2 *miR-9* (Figure 5.7). RT-QPCR was also performed for *KDM7A, CCNE* and *NR5A2*, however no amplification was detected. This data suggested that *miR-9-5p* does not target the predicted genes in these cells. However, it has been shown that miRNAs may repress protein synthesis without the destruction of the targeted transcript (Filipowicz, Bhattacharyya and Sonenberg, 2008). Considering this, the overexpression of *miR-9* may be inhibiting the translation of these proteins without interfering with the transcript levels. This may explain why there is no significant reduction in mRNA expression for these genes (Figure 5.7), whereas a reduction in luciferase reporters is observed (Figure 5.5 and 5.6).

In attempt to identify and validate transcripts that are targeted and degraded by *miR-9*, RNA-seq was performed on TCam-2 Scrambled and TCam-2 *miR-9* cells and global differential expression analysis was performed. A total of 209 genes were differentially expressed when comparing the two groups (Figure 5.8). Unsupervised hierarchical clustering, using the 209 differentially expressed genes, resulted in samples being separated into two main branches, (1) TCam-2 Scrambled and (2) TCam-2 *miR-9* (Figure 5.8). This data indicated that overexpression of *miR-9* resulted in a change in global gene expression which permitted the separation of samples into scrambled and *miR-9* overexpressing clusters. *MiR-9* was in the top 20 differentially expressed genes and was significantly up-regulated in TCam-2 *miR-9* cells thus confirming these cells

overexpressed *miR-9* (Table 5.2). Within the list of top 20 differentially expressed genes, no genes were known to be essential for pluripotency and proliferation (Table 5.2). To investigate which genes *miR-9* may be targeting directly, genes that were downregulated in the TCam-2 *miR-9* samples that were also predicted *miR-9* targets were identified (Figure 5.9). This identified 3 predicted *miR-9* targets which were also downregulated in TCam-2 cells over expressing *miR-9*: *SIK1, CD34* and *SMN2* (Figure 5.9; Table 5.3). SIK1 is a member of the AMP-activated protein kinase (AMPK) family which play key roles in both metabolism and growth (Shackelford and Shaw, 2009). Low expression of *SIK1* has been associated with poor outcome in breast cancer patients and has been associated with tumour suppressive qualities including reduced proliferation and migration in adenocarcinoma and pancreatic cancer (Chin *et al.*, 2006; Lu *et al.*, 2008; Cheng *et al.*, 2009; Ren *et al.*, 2016). In contrast to this, CD34 is a cell surface glycoprotein that is highly expressed in early hematopoietic, epithelial and stromal cells (Sidney *et al.*, 2014). The function of CD34 remains largely unknown, however a link has been established between CD34 expression and stem cell activity, as expressing CD34 have an increased differentiation capacity (Sidney *et al.*, 2014). SMN2 encodes a protein that is ubiquitously expressed and is involved in a variety of tissue-specific functions that includes mRNA processing, axonal transport and homeostasis of ubiquitination (Wadman *et al.*, 2016). The majority of research on SMN2 has concentrated on the role in spinal muscular dystrophy, as low expression has been associated with this disease (Wadman *et al.*, 2016; Li *et al.*, 2017). Interestingly, *miR-9* expression inversely correlated with *SMN1* expression for patients with spinal muscular dystrophy (Wang *et al.*, 2014). The SMN1 and

SMN2 genes are identical except for single nucleotide differences on exon 7 (Lefebvre *et al.*, 1995), therefore this may suggest that *miR-9* may be important for the regulation of both *SMN1* and *SMN2*. Altogether, these 3 predicted genes appear perform diverse roles that may not be connected. It has been shown that SIK1 interacts with the histone deacetylase, HDAC5, to regulate MEF2 expression during muscle differentiation (Berdeaux *et al.*, 2007). In addition to this, SMN2 binds to SNW1, a coactivator that enhance transcription at pol II transcribed loci (Lleres *et al.*, 2010). Given that that HDAC5 and SNW1 can control the expression of many genes, the repression of *SIK1* and *SMN2* by *miR-9* may result in the changes observed in 209 genes in TCam-2 Scrambled and TCam-2 *miR-9* samples.

In attempt to elucidate the downstream effects of *miR-9* on these cells, gene ontology analysis was used. The first gene ontology analysis identified biological processes by which were enriched for the 209 differentially expressed genes identified in TCam-2 *miR-9* samples (Table 5.4). Genes enriched in this ontology were RGPD3, RGPD6 and RGPD2 which, although not much is known about their function, they exhibit high similarity in sequence to *RANBP2* (Nothwang *et al.*, 1998; Fagerberg *et al.*, 2014). RAN protein has been shown to bind to RANBP2 and RGPD2 (Wan *et al.*, 2015; Huttlin *et al.*, 2017). RAN is a small GTPase plays a key role in the process for spindle formation (Quimby and Dasso, 2003) and has been shown to be broadly over expressed in cancer cells when compared to normal cells (Xia, Lee and Altieri, 2008). RAN is shown regulate the cell cycle through regulation of the mitotic spindle formation (Clarke and Zhang, 2008) and is highly expressed in poor prognosis serous

epithelial ovarian cancer by enhancing cellular proliferation (Barrès *et al.*, 2010). This may implicate the RAN binding proteins RGPD3, RGPD6 and RGPD2 for the development and progression of TGCTs. In addition to this, genes involved in retinal cone differentiation and development were enriched (Table 5.4). This is interesting as *miR-9* is known to be highly expressed in neural cells during development (Darnell *et al.*, 2006; Leucht *et al.*, 2008) and *miR-9* expression is induced in NT2 cells when cell differentiate towards the neural lineage (Sempere *et al.*, 2004). One of these genes in this gene ontology was DIO3, this gene encodes a core protein that is subunit of the small nuclear ribonucleoprotein (snRNP) complex. SnRNPs associates with small nuclear RNAs (snRNAs) in order to form a spliceosome (Matera, Terns and Terns, 2007). The spliceosome complex is responsible for the splicing of pre-mRNA by removing introns from the transcripts, therefore the spliceosome is important for the generation of alternate splice forms of genes (Wahl, Will and Lührmann, 2009) SMN proteins are also part of the spliceosome complex and are required for the recognition of snRNA sequences (Fischer, Liu and Dreyfuss, 1997). As mentioned previously, *SMN2* is a predicted target of *miR-9* and is also downregulated in TCam-2 *miR-9* cells (Table 5.4). It has also been shown that the germ cell specific marker PRMT5 is responsible for modifying DIO3 to allow the incorporation of snRNPs into the spliceosome complex (Gonsalvez *et al.*, 2007), suggesting that DIO3 may also be important for germ cell spliceosome regulation. Smn protein has also been shown to be enriched in germline stem cells in *Drosophila* with the over overexpression of *Smn* resulting in increased spermatocyte proliferation (Grice and Liu, 2011). SMN is also enriched in highly proliferative mESCs and the reduction of SMN impairs pluripotency through a reduction in pluripotency

markers including *Sox2, Klf4* and *Rex1* (Chang *et al.*, 2015). It could be speculated that constitutive *miR-9* over expression in TCam-2 cells resulted in targeting of *SMN2*, and in turn *DIO3*, resulting in a reduction of spliceosome complex formed to repress pluripotency in these cells.

Gene ontology analysis was next performed using the Reactome Pathway Database (Croft *et al.*, 2010) for genes differentially expressed in TCam-2 Scrambled and TCam-2 *miR-9* (Table 5.5). This analysis indicated that genes differentially expressed in TCam-2 Scrambled vs TCam-2 *miR-9* samples were enriched for the hormonal pathway. Pathways included metabolism of steroid hormones, androgen biosynthesis and hormone-ligand receptors (Table 5.5). This is of particular interest given that the testes produce androgens and TGCTs are most prominent in males after the onset of puberty, thereby suggesting a hormonal regulation for tumour growth (Oosterhuis and Looijenga, 2005). The genes enriched in the ontologies related to hormone regulation included *CGA, POMC, GNRH1, CYP19A1* and *DIO3* (Table 5.5). Interestingly, STRING analyses identified that CGA, POMC and CYP19A1 proteins may functionally interact (Figure 5.10). POMC is a prohormone that has been shown to be most highly expressed in the pituitary (Clark, 2016). POMC is cleaved forming multiple peptide hormones including adrenocorticotrophic hormone (ACTH), this mammalian pituitary hormone in adults is shown to drive steroidogenesis in the adrenal cortex (Clark, 2016). Interestingly, ACTH has also bene shown to play a role in prenatal development as the melanocortin receptors (Mc2R and Mc5R) are highly expressed in a wide variety of mouse embryonic tissues including the progenitor germ cells within the developing testes (Nimura *et al.*,

2006). O'Shaughnessy and colleagues also showed that incubation of mouse neonatal testes with ACTH *in vitro* stimulated testosterone production (O'Shaughnessy *et al.*, 2003). CGA encodes a subunit of human chorionic gonadotropin (hCG) which is a protein that is expressed during embryonic development (Jurisicova et al., 1999) that has also been shown to promote testosterone production (Kirk *et al.*, 1994). This is particular interest as the treatment of TCam-2 cells with testosterone results in an increase in proliferation rate (Wallacides *et al.*, 2012). This may suggest a positive feedback loop for TGCTs that express high levels of POMC and CGA, may produce high levels of ACTH and hCG to drive increased steroidogenesis and as a result enhanced tumour growth. In contrast to this, *CYP19A1* encodes a protein that is a member of the cytochrome P450 family, which are involved in hormone metabolism and detoxification Interestingly, CYP19A1 converts androgens to oestrogens (mainly testosterone to estradiol) and CYP19A1 is up-regulated in hESCs when differentiated towards granulosa-like cells (Lan *et al.*, 2013). Thus the data obtained here identified a role for *miR-9* in indirectly regulating genes involved in hormone pathways.

It has been shown in mouse gonadal PGCs that *miR-9* is up-regulated (Hayashi *et al.*, 2008), in addition to this *miR-9* is known to target genes important for germ cell development such as *Blimp1* (Huang *et al.*, 2012; Nowak *et al.*, 2014). This data suggested that *miR-9* is upregulated in late gonadal cells in order to aid the maturation process through targeting early PGC genes. It has been shown by gene expression analysis (RNA-seq) that the seminoma cell line TCam-2 resemble PGCs that are half way between early PGC development (i.e. time of

specification) and late stage gonadal cells that have begun the maturation process (Irie *et al.*, 2015). Given that *miR-9* is expressed during the late stages of germ cell development and is known to target early germ cells genes (Hayashi *et al.*, 2008; Huang *et al.*, 2012), we hypothesised that in TCam-2 cells that over-expressed *miR-9* may result in global gene expression to change to resemble later staged germ cells. In order to test this hypothesis, RNA-seq samples derived from hESCs (cultured in 4i medium), somatic cells, hPGCs and TCam-2 (WT) from the study conducted by Irie *et al*, 2015, in addition to the TCam-2 Scrambled and TCam-2 *miR-9* samples created in this study were analysed using PCA (Figure 5.11). TCam-2 (WT) and the samples from this study (scrambled and *miR-9*) clustered most closely (Figure 5.8). This data confirmed that the TCam-2 Scrambled samples are a viable control as they are not significantly different from unmodified TCam-2 cells from the Irie study. However, TCam-2 *miR-9* did not cluster more closely to hPGCs than either TCam-2 WT or TCam-2 scrambled. This suggested that *miR-9* overexpression does not change gene expression to a profile that resembles late stage gonadal PGC (hPGC) or towards a more differentiated state (Soma) (Figure 5.11).

Conclusion

The goal of this Chapter was to identify novel *miR-9* targets and to validate these in TGCTs. Using the miRWalk software, a list for *miR-9-5p* and *miR-9-3p* predicted targets was created (Figure 5.1) which were then overlapped with genes that were lowly expressed in *miR-9* high TCGA patient samples (Figure 5.2). It was shown that in 293T cells that *miR-9-5p* repressed the expression for *LIN28A, CCNE1, KDM7A, NR5A2* and *KDM4C* (Figure 5.4). However, when the expression of these genes were analysed in TCam-2 cells that over expressed *miR-9 LIN28A* and *KDM4C* were not shown to be downregulated (Figure 5.7). In attempt to elucidate transcripts that were targeted by *miR-9*, TCam-2 cells that overexpressed *miR-9* or a scrambled sequence were analysed by RNA-seq analysis and it was shown a total of 209 genes were differentially expressed (Figure 5.9). It was shown that 3 genes that were downregulated in TCam-2 *miR-9* cells were also *miR-9* targets (Figure 5.10), these genes were *SIK1*, *SMN2* and *CD34* (Table 5.3). From this, we hypothesised that repression of SIK1 and SMN2 by miR-9 results in the differentially expressed genes when comparing TCam-2 *miR-9* and TCam-2-Scrambled cells (Figure 5.12). Gene Ontology analysis, using the differentially expressed genes, identified processes involving the regulation of the Golgi apparatus in addition to hormonal pathways to change in response to *miR-9* over expression.

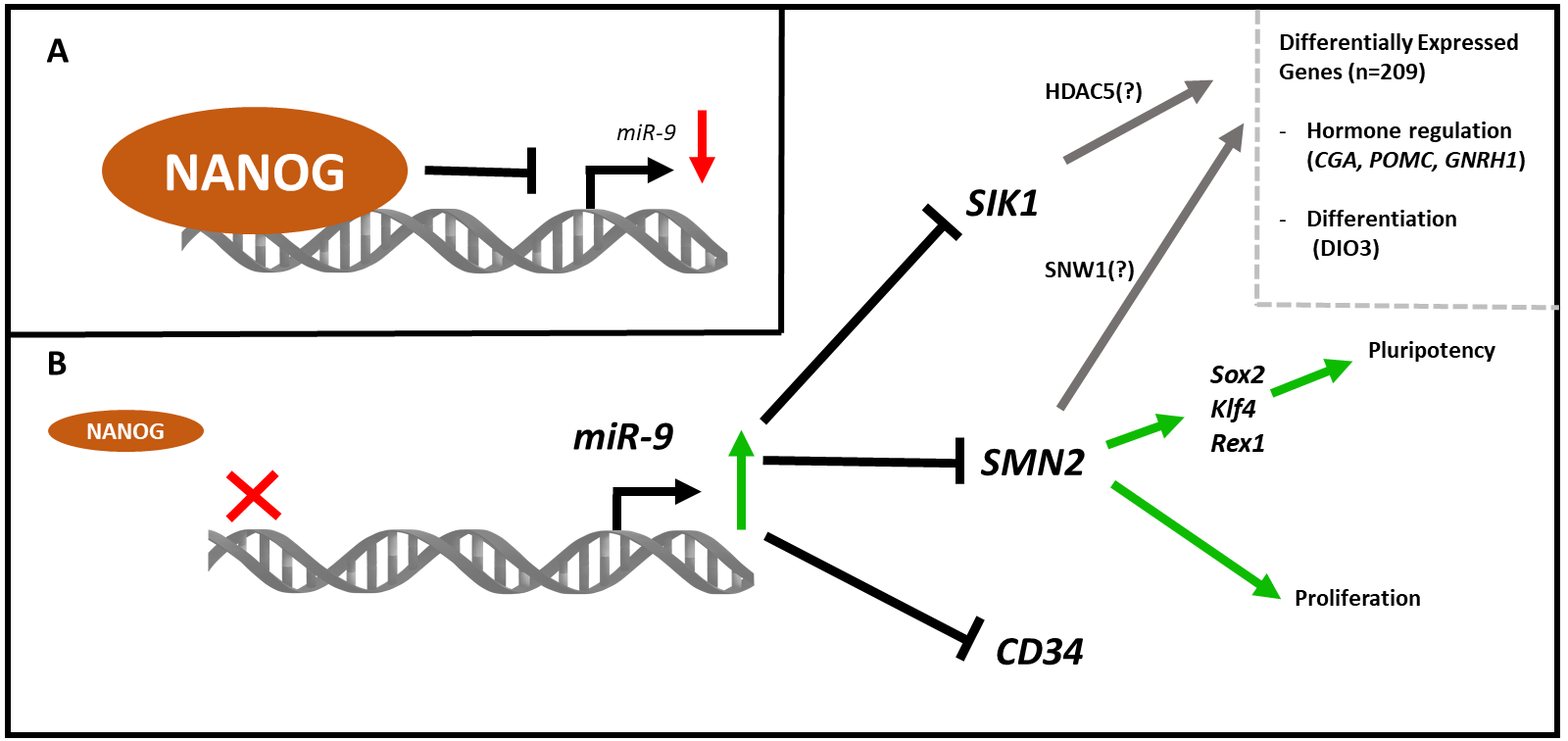


Figure 5.12 A depiction of NANOG regulating miR-9 in TGCTs. MiR-9 predicted targets which are also downregulated in TCam-2 cells overexpressing miR-9 (*SIK1*, *SMN2* and *CD34*) are hypothesised to regulate the 209 differentially expressed genes in TCam-2-miR-9 and TCam-2-Scrambled groups.

5.4 Supplementary

KDM4C (99.7% similarity)

```
KDM4C_PSI_Plasmid_Sequencing      GTGCTTAAGAGGATCGCGGCGCTCTTCTGCGTGCCCGTGGCCACCAAGACCAAACCCCGA
KDM4C_amplicon                    ------------------------------------------------------------

KDM4C_PSI_Plasmid_Sequencing      GCCGCCTCCGAATGAGAGTGTTTCGTTCCTTCCCCCTCCCCCCGCGTCAGACAAACCCTA
KDM4C_amplicon                    ------------------------------------------------------------

KDM4C_PSI_Plasmid_Sequencing      ACCACCGCTTAAGCGGCCCCCGCGAGGTCCGAAGACTCATTTAGATCCTCACACAAAAAA
KDM4C_amplicon                    ------------------------------------------------------------

KDM4C_PSI_Plasmid_Sequencing      CCAACACACAGATGTAATGAAAATAAAGATATTTTATTGCGGCCAGCGGCCGCGTGGCTT
KDM4C_amplicon                    ----------------------------------------------------------GTGGCTT
                                                                                             *******

KDM4C_PSI_Plasmid_Sequencing      TTCCTTGTGGGTGATTCACTAGTATTTCCTTTCACTTAGAAAACTTCTTTGGGATCATAG
KDM4C_amplicon                    TTCCTTGTGGGTGATTCACTAGTATTTCCTTTCACTTAGAAAACTTCTTTGGGATCATAG
                                  ************************************************************

KDM4C_PSI_Plasmid_Sequencing      AAGACAACTGCTCACAAGTGAAAGCAAGACTAGAGATGATTTCAGATTGCTGCGTCCAAT
KDM4C_amplicon                    AAGACAACTGCTCACAAGTGAAAGCAAGACTAGAGATGATTTCAGATTGCTGCGTCCAAT
                                  ************************************************************

KDM4C_PSI_Plasmid_Sequencing      TCATTGTGTGTATTTTGTATTTTTGGTGACCAAACTCTGGAAACCCGATTTACCACTGTC
KDM4C_amplicon                    TCATTGGGTGTATTTTGTATTTTTGGTGACCAAACTCTGGAAACCCGATTTACCACTGTC
                                  ****** ***************************************************

KDM4C_PSI_Plasmid_Sequencing      AGAGACACACCCTGTCACCTATGCCAGCAAAACTCACGGCACAGCCCCAAGGATGTCCCT
KDM4C_amplicon                    AGAGACACACCCTGTCACCTATGCCAGCAAAACTCACGGCACAGCCCCAAGGATGTCCCT
                                  ************************************************************

KDM4C_PSI_Plasmid_Sequencing      TCATCTTCTCTTTTCCAACCCAAGCTGCTCTGTGGCCTGCAGCGATGTATGCAGACTACT
KDM4C_amplicon                    TCATCTTCTCTCTTCCAACCCAAGCTGCTCTGTGGCCTGCAGCGATGTATGCAGACTACT
                                  ********** ***************************************************

KDM4C_PSI_Plasmid_Sequencing      GTCTCTTCTGGCACTTCTTCTGGAAAGAGCTCTTCAAAAAAGTGCGGTATACAGGGTCGG
KDM4C_amplicon                    GTCTCTTCTGGCACTTCTTCTGGAAAGAGCTCTTCAAAAAAGTGCGGTATACAGGGTCGG
                                  ************************************************************

KDM4C_PSI_Plasmid_Sequencing      CCACATATTCATTCTTAAACCTGGAGCTCAGCCTCGAGCGATCGCCTAGAATTACTGCTC
KDM4C_amplicon                    CCACATATTCATTCTTAAACCTGGAGCTCAGC----------------------------
                                  ********************************

KDM4C_PSI_Plasmid_Sequencing      GTTCTTCAGCACGCGCTCCACGAAGCTCTTGATGTACTTACCCATTTCATCTGGAGCGTC
KDM4C_amplicon                    ------------------------------------------------------------

KDM4C_PSI_Plasmid_Sequencing      CTCCTGGCTGAAGTGGAGGCCCTTCACCTTCACGAACTCGGTGTTAGGGAACTTCTTAGC
KDM4C_amplicon                    ------------------------------------------------------------
```

NR5A2_1

```
NR5A2_1_PSI_Plasmid_Sequencing    AACCACCGCTTAAGCGGCCCCCGCGAGGTCCGAAGACTCATTTAGATCCTCACACAAAAA
NR5A2_amplicon                    ------------------------------------------------------------

NR5A2_1_PSI_Plasmid_Sequencing    ACCAACACACAGATGTAATGAAAATAAAGATATTTTATTGCGGCCAGCGGCCGCCCACAT
NR5A2_amplicon                    ----------------------------------------------------------CCACAT
                                                                                           ******

NR5A2_1_PSI_Plasmid_Sequencing    CTGCACAGCTACCTTTTAAAATTGTGAAATAAGGCAAGACTTTCATACCTGTGGCATCGC
NR5A2_amplicon                    CTGCACAGCTACCTTTTAAAATTGTGAAATAAGGCAAGACTTTCATACCTGTGGCATCGC
                                  ************************************************************

NR5A2_1_PSI_Plasmid_Sequencing    ATGTTAATAATACTGATAATAGTCATTTGAAGAATTTGCGATTCTTTGTCCTCCCATATT
NR5A2_amplicon                    ATGTTAATAATACTGATAATAGTCATTTGAAGAATTTGCGATTCTTTGTCCTCCCATATT
                                  ************************************************************

NR5A2_1_PSI_Plasmid_Sequencing    TGTTCTACAGCAGACTTTGGTCTATGACTTAAAAAGGTCAGTATCACTCTCTTCCTTTCC
NR5A2_amplicon                    TGTTCTACAGCAGACTTTGGTCTATGACTTAAAAAGGTCAGTATCACTCTCTTCCTTTCC
                                  ************************************************************

NR5A2_1_PSI_Plasmid_Sequencing    CCTCTCGAGCGATCGCCTAGAATTACTGCTCGTTCTTCAGCACGCGCTCCACGAAGCTCT
NR5A2_amplicon                    CCT---------------------------------------------------------
                                  ***

NR5A2_1_PSI_Plasmid_Sequencing    TGATGTACTTACCCATTTCATCTGGAGCGTCCTCCTGGCTGAAGTGGAGGCCCTTCACCT
NR5A2_amplicon                    ------------------------------------------------------------
```

NR5A2_2

```
NR5A2_2_Sequencing_Result      AACCAACACACAGATGTAATGAAAATAAAGATATTTTATTGCGGCCAGCGGCCGCCCTTT
NR5A2_2_Amplicon               ------------------------------------------------------------

NR5A2_2_Sequencing_Result      GAGCCACGAAGATAACACGGGTGACTGAGATTGCGCATGAGTCAACGCCCAGCTTACCTA
NR5A2_2_Amplicon               ------------------------------------------------------------

NR5A2_2_Sequencing_Result      ATGTGTACCCTTCATGTGGATGAGTATTAGAGGTGATTAATAAACATAATAAAATAATCC
NR5A2_2_Amplicon               -------CCCTTCATGTGGATGAGTATTAGAGGTGATTAATAAACATAATAAAATAATCC
                                      **************************************************

NR5A2_2_Sequencing_Result      TCTTTGTACACACAGAAGTAATAACTAATATCTTTGGTATACAATTTACAATGCTTGGAA
NR5A2_2_Amplicon               TCTTTGTACACACAGAAGTAATAACTAATATCTTTGGTATACAATTTACAATGCTTGGAA
                               ************************************************************

NR5A2_2_Sequencing_Result      AGGAAAACCAGGTAAGACTCCACCTTGTTGAGTTTACTTGAAAACCAGAGGAAGCTACAA
NR5A2_2_Amplicon               AGGAAAACCAGGTAAGACTCCACCTTGTTGAGTTTACTTGAAAACCAGAGGAAGCTACAA
                               ************************************************************

NR5A2_2_Sequencing_Result      GTCACATAATTTACAATAGATATTAAATAAACTACAGTAGGAATGATGAAGATTTACAGA
NR5A2_2_Amplicon               GTCACATAATTTACAATAGATATTAAATAAACTACAGTAGGAATGATGAAGATTTACAGA
                               ************************************************************

NR5A2_2_Sequencing_Result      CATTTAATCAAGTGAATGATTTTAATTCATTAAGACTCTTCTATGTCCCTGAACTAAGTT
NR5A2_2_Amplicon               CATTTAATCAAGTGAATGATTTTAATTCATTAAGACTCTTCTATGTCCCTGAACTAAGTT
                               ************************************************************

NR5A2_2_Sequencing_Result      CCACTAATTTGAAGACATGCTACAGTAAACACATATATTTGGTTTCTTCAGTCTTGGCTA
NR5A2_2_Amplicon               CCACTAATTTGAAGACATGCTACAGTAAACACATATATTTGGTTTCTTCAGTCTTGGCTA
                               ************************************************************

NR5A2_2_Sequencing_Result      TATCTTATACAAAGGTGAAACCAAGAAACAATAAATTAGCATCATAATTGTCTACAGTTG
NR5A2_2_Amplicon               TATCTTATACAAAGGTGAAACCAAGAAACAATAAATTAGCATCATAATTGTCTACAGTTG
                               ************************************************************

NR5A2_2_Sequencing_Result      AAAGTATGATATACAATATAAATACACTACTCAAAAGCCAGAAATATTTATGAAAAAATT
NR5A2_2_Amplicon               AAAGTATGATATACAATATAAATACACTACTCAAAAGCCAGAAATATTTATGAAAAAATT
                               ************************************************************

NR5A2_2_Sequencing_Result      AAGGGACTAGGAGATACTTTTATTTTTTACAGTTTCTACAAAGTGATTGTGCCGCCCTCG
NR5A2_2_Amplicon               AAGGGACTAGGAGATACTTTTATTTTTTACAGTTTCTA----------------------
                               ************************************

NR5A2_2_Sequencing_Result      AGCGATCGCCTAGAATTACTGCTCGTTCTTCAGCACGCGCTCCACGAAGCTCTTGATGTA
NR5A2_2_Amplicon               ------------------------------------------------------------

NR5A2_2_Sequencing_Result      CTTACCCATTTCATCTGGAGCGTCCTCCTGGCTGAAGTGGAGGCCCTTCACCTTCACGAA
NR5A2_2_Amplicon               ------------------------------------------------------------

NR5A2_2_Sequencing_Result      CTCGGTGTTAGGGAACTTCTTAGCTCCCTCGACAATAGCGTTGGAAAAGAACCCAGGGTC
NR5A2_2_Amplicon               ------------------------------------------------------------
```

ESRRB

```
ESRRB_Sequencing_Result        CGAGCCGCCTCCGAATGAGAGTGTTTCGTTCCTTCCCCCTCCCCCCGCGTCAGACAAACC
ESRRB_2_Amplicon               ------------------------------------------------------------

ESRRB_Sequencing_Result        CTAACCACCGCTTAAGCGGCCCCCGCGAGGTCCGAAGACTCATTTAGATCCTCACACAAA
ESRRB_2_Amplicon               ------------------------------------------------------------

ESRRB_Sequencing_Result        AAACCAACACACAGATGTAATGAAAATAAAGATATTTTATTGCGGCCAGCGGCCGCACAA
ESRRB_2_Amplicon               ------------------------------------------------------------ACAA
                                                                                            ****

ESRRB_Sequencing_Result        GTAGGTGATGCTGCAATGGGGAACTGCCATTTTAAAAAATATATACATCTTAAGAAGCCA
ESRRB_2_Amplicon               GTAGGTGATGCTGCAATGGGGAACTGCCATTTTAAAAAATATATACATCTTAAGAAGCCA
                               ************************************************************

ESRRB_Sequencing_Result        TTCCAGCCATATTATCATTATTAGAAAATAAAACAATCTTTCCACATAGTGCTTTTCTTT
ESRRB_2_Amplicon               TTCCAGCCATATTATCATTATTAGAAAATAAAACAATCTTTCCACATAGTGCTTTTCTTT
                               ************************************************************

ESRRB_Sequencing_Result        GGTGCTGTTCCGTGAGCATGGACTCTGGACCCTTCAGTCTTCTTGTCACCTCGAGCGATC
ESRRB_2_Amplicon               GGTGCTGTTCCGTGAGCATGGACTCTGGACCCTTCAGTCTTCTTGTCAC-----------
                               ********************************************

ESRRB_Sequencing_Result        GCCTAGAATTACTGCTCGTTCTTCAGCACGCGCTCCACGAAGCTCTTGATGTACTTACCC
ESRRB_2_Amplicon               ------------------------------------------------------------

ESRRB_Sequencing_Result        ATTTCATCTGGAGCGTCCTCCTGGCTGAAGTGGAGGCCCTTCACCTTCACGAACTCGGTG
ESRRB_2_Amplicon               ------------------------------------------------------------
```

KDM7A_2

```
KDM7A-2_Sequencing_Result      CCGCCTCCGAATGAGAGTGTTTCGTTCCTTCCCCCTCCCCCCGCGTCAGACAAACCCTAA
KDM7A_2_Amplicon               ------------------------------------------------------------

KDM7A-2_Sequencing_Result      CCACCGCTTAAGCGGCCCCCGCGAGGTCCGAAGACTCATTTAGATCCTCACACAAAAAAC
KDM7A_2_Amplicon               ------------------------------------------------------------

KDM7A-2_Sequencing_Result      CAACACACAGATGTAATGAAAATAAAGATATTTTATTGCGGCCAGCGGCCGCCCAATGGA
KDM7A_2_Amplicon               ----------------------------------------------------------CCAATGGA
                                                                                         ********

KDM7A-2_Sequencing_Result      TCCCAGCTGAACAGTCACTTATTTGAAACAATACAGGAAAATTATCATCTATTCCTATTA
KDM7A_2_Amplicon               TCCCAGCTGAACAGTCACTTATTTGAAACAATACAGGAAAATTATCATCTATTCCTATTA
                               ************************************************************

KDM7A-2_Sequencing_Result      CTAAAGTTGCATATGTTAATCATAATTCTCAATTTTACATATTTGTATTTGGTTACAACT
KDM7A_2_Amplicon               CTAAAGTTGCATATGTTAATCATAATTCTCAATTTTACATATTTGTATTTGGTTACAACT
                               ************************************************************

KDM7A-2_Sequencing_Result      AGCTAAATGCTATCAGTGCTGTCTTCCTAAGTTGCTGTGTCTGTATGGTATAAATGAGGT
KDM7A_2_Amplicon               AGCTAAATGCTATCAGTGCTGTCTTCCTAAGTTGCTGTGTCTGTATGGTATAAATGAGGT
                               ************************************************************

KDM7A-2_Sequencing_Result      TCTCTATTACTGTTAAGCCCAGGTACTCGCAGCAGCTACCACAAAGTTCAGCCATTTCCA
KDM7A_2_Amplicon               TCTCTATTACTGTTAAGCCCAGGTACTCGCAGCAGCTACCACAAAGTTCAGCCATTTCCA
                               ************************************************************

KDM7A-2_Sequencing_Result      TGAGATAAAGGGATGCATTATGGCCAGTAATGTTATAAGACGTTTGACCAGGAGTCACTG
KDM7A_2_Amplicon               TGAGATAAAGGGATGCATTATGGCCAGTAATGTTATAAGACGTTTGACCAGGAGTCACTG
                               ************************************************************

KDM7A-2_Sequencing_Result      GATCAGTGGCTCGAGCGATCGCCTAGAATTACTGCTCGTTCTTCAGCACGCGCTCCACGA
KDM7A_2_Amplicon               GATCAGTGG---------------------------------------------------
                               *********

KDM7A-2_Sequencing_Result      AGCTCTTGATGTACTTACCCATTTCATCTGGAGCGTCCTCCTGGCTGAAGTGGAGGCCCT
KDM7A_2_Amplicon               ------------------------------------------------------------
```

CCNE1

```
Sequencing_CCNE_21      TCAGCGGCCCGCCAGTTACCACCCGCCGACCAAACCCCCGCCTCCACGGAGGGCGGGGGG
Amplicon_CCNE1          ------------------------------------------------------------

Sequencing_CCNE_21      GTGCTTAAGAGGATCGCGGCGCTCTTCTGCGTGCCCGTGGCCACCAAGACCAAACCCCGA
Amplicon_CCNE1          ------------------------------------------------------------

Sequencing_CCNE_21      GCCGCCTCCGAATGAGAGTGTTTCGTTCCTTCCCCCTCCCCCCGCGTCAGACAAACCCTA
Amplicon_CCNE1          ------------------------------------------------------------

Sequencing_CCNE_21      ACCACCGCTTAAGCGGCCCCCGCGAGGTCCGAAGACTCATTTAGATCCTCACACAAAAAA
Amplicon_CCNE1          ------------------------------------------------------------

Sequencing_CCNE_21      CCAACACACAGATGTAATGAAAATAAAGATATTTTATTGCGGCCAGCGGCCGCGCAACAG
Amplicon_CCNE1          ------------------------------------------------------------ACAG
                                                                                     ****

Sequencing_CCNE_21      ACAGAAGAGAACGTGGAGCAGGCGCGCAACTGTCTTTGGTGGAGAAGGATGGGGTGGTCA
Amplicon_CCNE1          ACAGAAGAGAACGTGGAGCAGGCGCGCAACTGTCTTTGGTGGAGAAGGATGGGGTGGTCA
                        ************************************************************

Sequencing_CCNE_21      CGCCATTTCCGGCCCGCTGCTCTGCTTCTTACCGCTCTGTGGCGGGGTGAGGAGCCCACT
Amplicon_CCNE1          CGCCATTTCCGGCCCGCTGCTCTGCTTCTTACCGCTCTGTGGCGGGGTGAGGAGCCCACT
                        ************************************************************

Sequencing_CCNE_21      GGGGAGAGGAGAAGCCCTATTTTGTTCAGACAACATGGCTTTCTTTGCTCGGGCTTTGTC
Amplicon_CCNE1          GGGGAGAGGAGAAGCCCTATTTTGTTCAGACAACATGGCTTTCTTTGCTCGGGCTTTGTC
                        ************************************************************

Sequencing_CCNE_21      CAGCAAATCCAAGCTGCTCGAGCGATCGCCTAGAATTACTGCTCGTTCTTCAGCACGCGC
Amplicon_CCNE1          CAGCAAATCCAAGCTG--------------------------------------------
                        ***************

Sequencing_CCNE_21      TCCACGAAGCTCTTGATGTACTTACCCATTTCATCTGGAGCGTCCTCCTGGCTGAAGTGG
Amplicon_CCNE1          ------------------------------------------------------------

Sequencing_CCNE_21      AGGCCCTTCACCTTCACGAACTCGGTGTTAGGGAACTTCTTAGCTCCCTCGACAATAGCG
Amplicon_CCNE1          ------------------------------------------------------------

Sequencing_CCNE_21      TTGGAAAAGAACCCAGGGTCGGACTCGATGAACATCTTAGGCAGATCGTCGCTGGCCCGA
Amplicon_CCNE1          ------------------------------------------------------------

Sequencing_CCNE_21      AGGTAGGCGTTGTAGTTGCGGACAATCTGGACGACGTCGGGCTTGCCTCCCTTAACGAGA
Amplicon_CCNE1          ------------------------------------------------------------
```

Figure S5.1 Sequencing results for the amplicon for psiCHECK-2 plasmids for *miR-9* binding sites for the genes *KDM4C*, *NR5A2_1*, *NR5A2_2*, *ESRRB*, *KDM7A* and *CCNE1* confirming the correct genomic sequences were inserted.

Chapter 6: General Discussion

The aim of this project was to study the role of miRNAs involved in TGCTs that were regulated by NANOG. To do this, available bioinformatics datasets were used to predict candidate NANOG regulated miRNAs. TGCT cell lines were used as a model to validate NANOG regulation of miRNAs in TGCTs and to identify miRNAs downstream targets that may be important for TGCT development and progression. This chapter will review the experimental chapters and how they relate to current literature, their strength and limitations, their possible applications, and future direction.

The results from chapter 3 identified a list of miRNAs that could be transcriptionally regulated by NANOG. Analysis of TGCT gene expression datasets then produced a list of differentially expressed miRNAs for the groups *NANOG* high vs *NANOG* low TCGA germ cell tumour samples and tumour vs normal tissue. These miRNA lists were overlapped in order to give a candidate list of miRNAs that could be regulated by NANOG and that could be involved in TGCTs. Chapter 4 analysed the expression of chosen miRNAs comparing TGCT cell lines and normal testes to validate those miRNAs that were found differentially expressed in TGCT vs normal testes by the microarray analysis performed in Chapter 3. This analysis identified *miR-9* to be downregulated in TGCT cell lines. Luciferase and NANOG knockdown experiments demonstrated that *miR-9* was transcriptionally repressed by NANOG. In Chapter 5, a list of predicted targets for *miR-9-5p* and *miR-9-3p* was produced and genes involved in regulating pluripotency and proliferation were selected. *MiR-9-5p*

targeted the 3'UTR for these transcripts but the overexpression of *miR-9* in the TGCT cell line TCam-2 did not result in increased degradation of gene transcripts. TCam-2 cells that overexpressed *miR-9* were next analysed by RNA-seq and this identified 3 genes significantly downregulated that were also *miR-9* predicted targets. Therefore, these targets could represent important genes involved in TGCTs.

6.1 Importance of the study: role of miRNA in TGCTs

MiRNAs are an abundant class of small non-coding RNA that function as negative regulators of gene expression and protein translation (Esquela-Kerscher and Slack, 2006). Mature miRNAs bind to an Argonaute RNA-binding protein to form a RNA inducing silencing complex (RISC). The RISC complex then binds to the 3' UTR of target mRNA sequences resulting in transcript degradation and/or repression of translation (Khvorova, Reynolds and Jayasena, 2003; Schwarz *et al.*, 2003). Using bioinformatics analyses, it has been shown that miRNAs can directly target and repress hundreds of genes, resulting in changes in both the transcriptome and proteome (Baek *et al.*, 2008). MiRNAs function as transcriptional and translational regulators, generally exerting their effects through subtle modulation of a number of targets and are often thought to function as fine tuners of expression. MiRNAs can eliminate low level transcripts which may be inappropriate in a given cellular context thereby giving a robust cellular phenotype (Bartel, 2009; Liu and Olson, 2010). Through the regulation of multiple proteins involved in different steps of a biological pathway, miRNAs have been shown to exert powerful effects in regulating biological processes including metabolism, growth, differentiation and apoptosis

(Cheng *et al.*, 2005; Kloosterman and Plasterk, 2006; Liu and Olson, 2010). This highlights miRNAs importance in the regulation of complex biological process in normal physiological conditions. Abnormal expression of miRNAs plays a crucial role in the development of disease (Kloosterman and Plasterk, 2006). Indeed, miRNA are frequently mutated or aberrantly expressed in cancer, indicating their importance for suppressing or promoting carcinogenesis (Iorio *et al.*, 2005; Lu *et al.*, 2005; Esquela-Kerscher and Slack, 2006; Novotny *et al.*, 2012).

It has been shown that the expression profiles for miRNAs can define a variety of tumour types compared to normal tissue (Lu *et al.*, 2005). Indeed, a study that investigated expression of 217 miRNAs from normal and tumour samples derived from a number of different tissues (colon, kidney, prostate, lung, breast and the uterus) (Lu *et al.*, 2005) demonstrated that 129 of the 217 miRNA were downregulated in tumour samples when compared to normal samples. This suggested that the downregulation of these miRNAs was associated with tumour development and that the expression profile of miRNAs could be used to distinguish tumour samples from normal. Further work investigated the miRNA expression profile of 400 paraffin embedded and fresh-frozen samples derived from primary tumours and metastases for cancer samples from different origins (Rosenfeld *et al.*, 2008). MiRNA expression profiles could be used to identify with high confidence which tissue the samples originated from (Rosenberg *et al.*, 2000), thus highlighting the effectiveness of miRNAs as potential biomarkers for tracing the origin of cancer (Rosenberg *et al.*, 2000). Therefore,

miRNA expression profiles can be used to distinguish between normal and tumour samples as well as classify which origin the cancer is derived from.

The majority of research into the role of miRNAs in TGCTs has concentrated on global gene expression profiles (Palmer *et al.*, 2010; Novotny *et al.*, 2012). Several studies have investigated the expression of miRNAs in TGCTs in order to identify miRNAs that can be used as biomarkers to aid diagnosis of the disease (Murray *et al.*, 2011; Dieckmann *et al.*, 2012; Syring *et al.*, 2015). Novotny and colleagues (2012) performed a microarray analysis for TGCTs with differing amounts of carcinoma *in situ* (CIS) cells. MiRNAs from *miR-371-373* and *miR-302* clusters were highly expressed in samples containing CIS cells (Novotny *et al.*, 2012). This was further validated by *in situ* hybridisation using a probe against the *pre-miR-371-373* cluster showing these miRNAs were highly expressed in CIS cells when compared to the surrounding testis tissue (Novotny *et al.*, 2012). In addition to this, Palmer *et al*, 2010, demonstrated that miRNAs from the *miR-302* and *miR371-373* cluster were also overexpressed in TGCTs regardless of the histological subtype, histological site or patient age (Palmer *et al.*, 2010), thus highlighting the importance for these miRNAs for the development and progression in TGCTs. Subsequently, a study conducted by Dieckmann and colleagues (2012) identified that miR-373-3p to be highly expressed in the serum of TGCT patient's samples when compared to patients without TGCTs. *MiR-371, miR-372* and *miR-373* expression was dramatically reduced in TGCTs patients after orchiectomy surgery (Dieckmann *et al.*, 2012). This work demonstrated that miRNA from the *miR-371-373* cluster could be

used successfully as a biomarker for the diagnosis of TGCTs and to monitor patient post-orchiectomy.

The *miR-302* and *miR-371-373* cluster has been shown to target genes that regulate the cell cycle to enhance pluripotency and proliferation in ESCs (Barroso-delJesus *et al.*, 2008; Barroso-del Jesus, Lucena-Aguilar and Menendez, 2009; Lin *et al.*, 2010; Subramanyam *et al.*, 2011). However, similar experiment have not been conducted in TGCTs. This is also true for other miRNAs that have been identified to be highly or lowly expressed in TGCTs (Palmer *et al.*, 2010; Novotny *et al.*, 2012). For example, Palmer showed that miRNAs from the *miR-183-182-96* cluster are highly expressed in TGCTs, however these miRNAs are yet to be investigated for their role in TGCT progression (Palmer *et al.*, 2010). The overexpression of miRNAs from this cluster have been associated with carcinogenesis and disease progression in both breast and prostate cancer, through targeting of *FOXO1* and *ZIP1*, respectively (Guttilla and White, 2009; Mihelich *et al.*, 2011). Therefore, there are many miRNAs that have been identified in TGCTs that are yet to be investigated in the context of their transcript targets in order to help understanding of the mechanisms at the basis of TGCT carcinogenesis.

6.2 The role of *miR-9* in TGCTs

This project focused on identifying miRNAs differentially expressed in TGCTs whose expression is regulated by the transcription factor NANOG. Among the miRNA studied, *miR-9* was discovered to be negatively regulated by NANOG. As it has been shown that increased *NANOG* expression is associated with increased self-renewal and proliferation in ESCs (Torres and Watt, 2008; Zhang *et al.*, 2009; Abranches *et al.*, 2014), this suggests that *miR-9* expression may exert a negative effect on these processes.

It has been shown that *miR-9* is highly expressed in embryonal carcinoma cells that have been induced towards neural differentiation (Sempere *et al.*, 2004). During retinoic acid (RA) induced neural differentiation of NT2 cells, *NANOG* expression is dramatically reduced 2 days after the addition of RA (Deb-Rinker *et al.*, 2005), indicating an inverse correlation between *miR-9* and *NANOG* expression. This suggests that the downregulation of *NANOG* allows the de-repression of *miR-9* that is shown in differentiated NT2 cells. In addition to this, mouse embryonal carcinoma cells (F9) that after 1 day of RA treatment the SMN protein is reduced dramatically, indicating a role for SMN in pluripotency (Sharpe, Williams and Latchman, 1990). Supporting the high expression of SMN in *Drosophila* spermatocytes increases cellular proliferation (Grice and Liu, 2011) and is highly enriched in mESCs (Chang *et al.*, 2015). The knock down of *SMN* in mESC also resulted in a reduction in pluripotency markers and knockdown teratomas exhibit reduced proliferation and abnormal differentiation, especially for neural differentiation (Chang *et al.*, 2015). This is of particular interest as in TCam-2 cells overexpressing *miR-9*, SMN2 is

downregulated and it is also a predicted target for *miR-9-3p* (Chapter 5; Table 5.3). This perhaps suggests that *miR-9* targets *SMN2* in TGCT cell in order to promote differentiation.

It has also been shown that *miR-9* targets *SOX2* which resulted in decreased proliferation in glioma stem cells (Jeon *et al.*, 2011). This is of interest as NANOG and SOX2 have been shown to co-localise at many gene regions to induce pluripotency associated genes while repressing somatic gene expression (Loh *et al.*, 2006). In addition to this, NANOG and SOX2 have been shown to bind to each other's promoter in order to enhance expression and promote pluripotency (Boyer *et al.*, 2005). This may indicate that both NANOG and SOX2 binding can repress *miR-9* expression, creating a negative feedback loop between *miR-9* and SOX2. Similar to this, *miR-9* has been shown to target and repress *LIN28* (Eda *et al.*, 2009; La Torre, Georgi and Reh, 2013). Conversely, Lin28 in mouse embryonal carcinoma cells (P19) has been shown to induce *pre-miR-9* degradation (Nowak *et al.*, 2014). NANOG works in combination with both LIN28 and SOX2 to reprogram somatic fibroblast into induced pluripotent stem cells (Yu *et al.*, 2007). This suggests that there is an inverse relationship between *miR-9* and pluripotency factors, which is agreement with the data obtained in this project which showed NANOG repressing the transcription of *miR-9* in TGCT cells.

PGCs also highly express pluripotency factors during embryonic development (Irie *et al.*, 2015; Tang *et al.*, 2015) which have been shown to be critical for PGC formation (Kehler *et al.*, 2004; Yamaguchi *et al.*, 2009; Magnúsdóttir *et*

al., 2013). In particular, *Nanog* knockdown in mouse PGCs results in decreased PGC proliferation and increased apoptosis (Yamaguchi *et al.*, 2009). Nanog has been shown to be highly expressed during the migratory stage in mouse PGCs, but it is then downregulated after the PGCs colonise the gonads and start to mature (Yamaguchi *et al.*, 2005). Interestingly, *miR-9* expression follows the opposite pattern, being lowly expressed during PGC migratory stages and 5 fold higher as the cells start to colonise the gonads (Hayashi *et al.*, 2008). As mentioned previously, the seminoma cell line TCam-2 and gonadal PGCs share a similar mRNA expression profile (Irie *et al.*, 2015).

Results presented in this project demonstrated that NANOG repressed the transcription of *miR-9* in TCam-2 cells. Therefore, it could be speculated that the downregulation of Nanog in gonadal mPGC and concomitant upregulation of *miR-9* may be as a result of direct Nanog repression. *MiR-9* has also been shown to target two genes shown to be critical for the specification of PGCs in mice, *Lin28* and *Prdm1* (Huang *et al.*, 2012; Nowak *et al.*, 2014). Consequently, it is possible that *miR-9* upregulation in gonadal PGCs may be induced to enhance the maturation of PGCs by targeting genes associated with early PGC development. In addition to this, *PRDM1* and *LIN28* are highly expressed in seminoma TGCTs (West *et al.*, 2009; Gillis *et al.*, 2011; Nettersheim *et al.*, 2015), suggesting a role for these genes in TGCT malignancy. This study also demonstrated that *miR-9* was repressed in both SEM and EC TGCTs suggesting that *miR-9* may repress tumorigenicity (Figure 4.18). Supporting this, this study identified *miR-9* to be transcriptionally repressed by NANOG in TGCTs and high *NANOG* expression has been associated with enhanced malignancy in other

cancer types including oral squamous carcinoma (Chiou *et al.*, 2008), prostate cancer (Jeter *et al.*, 2011), colorectal cancer (Meng *et al.*, 2010) and breast cancer (Han *et al.*, 2012).

The tumour suppressive role of *miR-9* as has been shown in other cancer types. For instance, high *SOX2* expression promoted tumour growth in stem cells of glioblastoma multiforme (GBM), an aggressive form of brain malignancy (Vescovi, Galli and Reynolds, 2006). However, *miR-9* directly targets *SOX2* in these cells thereby repressing its expression and reducing tumour growth. This evidence supports a role for *miR-9* as a potential tumour suppressor and supports the notion of *miR-9* also promoting differentiation given that *SOX2* is a key pluripotency gene in ESCs (Chew *et al.*, 2005). Ectopic *miR-9* overexpression would induce anti-invasive, anti-proliferative and pro-apoptotic activity in breast cancer cell lines through the targeted downregulation of *MTHD2* (Selcuklu *et al.*, 2012). Sun and colleagues, 2017, also identified *miR-9* to play a tumour repressive role in breast cancer through the targeting of *HER2* (Sun *et al.*, 2017). Overall, this evidence suggests that NANOG mediated *miR-9* repression may perform tumour suppressive qualities in TGCTs.

Conversely, it has also been shown in breast cancer cells that *miR-9* expression is promoted by the oncogenes MYC and MYCN to enhance tumour metastasis (Ma *et al.*, 2010). More specifically, *miR-9* targeted *CDH1*, the gene that encodes the protein E-Cadherin, in order to promote invasion and motility (Ma *et al.*, 2010). In this study, the overexpression of *miR-9* in TCam-2 down-regulated *SIK1*, a gene which was also a predicted target of *miR-9-5p* (Chapter 5).

Interestingly, the downregulation of *SIK1* by miR-203 promoted migration and invasion in pancreatic cancer (Ren *et al.*, 2016). This also indicates a role for *miR-9* may regulate metastasis in TGCTs through *SIK1* targeted downregulation.

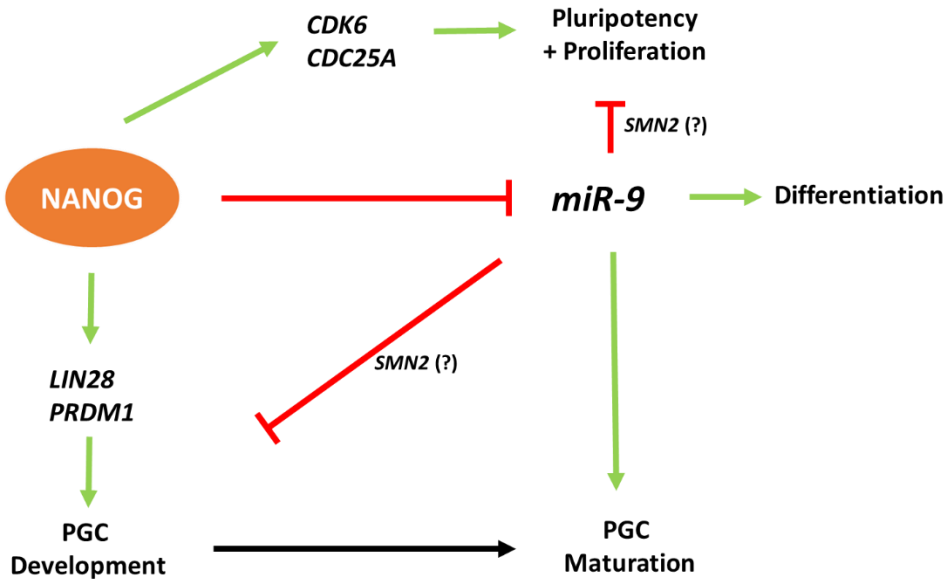


Figure 6.1 An overview for NANOG regulation of *miR-9* and the downstream effects in different cell types. It has been shown that NANOG transcriptionally represses *miR-9* expression. *MiR-9* represses proliferation and proliferation in ESCs and PGCs but will promote ESC differentiation and PGC maturation.

6.3 Limitations of this Study

This project represents an advance in the understanding on the molecular events involved in TGCT development as for the first time it defines a role for miRNA regulated by the transcription factor NANOG in both SEM and EC tumours, However, several limitations must be acknowledged.

In order to identify miRNAs that were transcriptionally regulated by NANOG, available hESC ChIP-seq datasets were utilised (Tsankov *et al.*, 2015). As both EC and SEM patient samples express high levels of pluripotency genes, including NANOG (Sperger *et al.*, 2003; Irie *et al.*, 2015), therefore it was reasonable to use hESC ChIP-seq data for NANOG binding. It has been shown through microarray analysis that ESCs samples and EC cell lines share a similar gene expression (Sperger *et al.*, 2003). This therefore suggests that hESC and EC NANOG binding sites may be comparable. However, Nanog binding sites are different in ESCs compared to epiblast like cells (EpiLCs) induced to become PGCLCs in mice (Murakami *et al.*, 2016). Furthermore, Nanog and Sox2 promote the pluripotent state in ESCs, whereas in EpiLCs Nanog would enhance the expression of germ cell specific genes for PGCLC induction through enhancer regulatory elements, while Sox2 would inhibit this induction (Murakami *et al.*, 2016). Therefore, NANOG regulation in ESCs and PGCs are different. Given that seminomas and PGCs share a similar expression profiles (Irie *et al.*, 2015), NANOG may also show differences for genomic binding in seminomas when compared to hESCs. This suggests that the hESC ChIP-seq used may be reductive when attempting to elucidate NANOG enrichment for both SEM and EC subtypes of TGCTs. On the other side, it has been

demonstrated that the knockdown of *NANOG* in either ESC or PGCs results in a reduction in proliferation (Yamaguchi *et al.*, 2009; Zhang *et al.*, 2009), suggesting that similar role for NANOG in both cell types.

Another limitation of this project was then use of TGCT cell lines. These were used to validate differentially expressed miRNAs in tumours vs normal testes. As mentioned previously, it is thought that TGCTs are derived from early PGCs that have failed to differentiate (Andrews *et al.*, 2005; Almstrup *et al.*, 2010). Therefore, using human PGCLCs, which model early PGCs just after specification (Irie *et al.*, 2015), as a 'normal' control would most likely be more suitable compared to using testes. Using this analysis may have resulted in the exclusion of potentially important miRNAs. PGCLCs were not available for this study, as the methodology had not yet been developed at the time. In addition to this, testis samples were required as a control to validate the expression profiles obtained from the microarray data. RT-QPCR was used to validated miRNA expression in TGCT cell lines, only *miR-9* expression was shown to corroborate with the microarray analysis. Focusing on a single miRNA is limitation as it is likely that NANOG is regulating more than one miRNA in TGCTs. This could be addressed by performing a miRNA transcriptomic analysis on *NANOG* KD TGCT cell lines and differentially expressed miRNAs with miRNAs that show NANOG binding upstream of TSSs from the ChIP-seq analysis (Chapter 4). This would most likely produce a more comprehensive list of miRNAs to be analysed for expression in TGCT patient samples.

Once *miR-9* had been confirmed to have low expression in TGCTs and to be transcriptionally repressed, the next step was to identify and validate *miR-9* targets. In order to ascertain if the *miR-9* targets were down-regulated in cell that overexpressed *miR-9*, RT-QPCR and RNA-sequencing was used. Only 3 *miR-9* predicted targets were downregulated in TCam-2 cells that overexpressed *miR-9*. This may be a result of the miRNA-target prediction software not producing a comprehensive or accurate prediction list or mRNA transcripts not being targeted for degradation. MiRNAs can target transcripts for degradation or inhibit translation without the destruction of the transcript (Filipowicz, Bhattacharyya and Sonenberg, 2008). This suggested that *miR-9* may be targeting the genes of interest in the TCam-2 cells overexpressing *miR-9* and inhibiting translation without the destruction of transcripts. RT-QPCR was used as it was the most reliable method to assess mRNA levels. In contrast, in TCam-2 cells overexpressing *miR-9* could be analysed using SILAC followed by mass-spec in order to identify changes in the proteome when compared to TCam-2 scrambled control cells (Nowak *et al.*, 2014). This would then be compared to a list of predicted *miR-9* targets which would then be validated using western blot. This would identify if the targets have been downregulated at the protein level. This however this process relies on the accuracy of bioinformatics tools to calculate predicted targets for *miR-9*. It was shown for TCam-2 cells overexpressing *miR-9* that 209 genes are differentially expressed, yet the prediction software only identified 3. This may suggest that the prediction software may have missed targets which are targeted and downregulated in TGCTs.

6.4 Future work

This project mainly focussed in the study of *miR-9* in SEM. Although SEM represents the most prevalent type of TGCTs, it would important also to study its role in EC tumours.

Therefore, future work would include performing NANOG knockdown in other TGCT EC cell lines in order to show that *miR-9* is also regulated by NANOG in EC cells. In addition to this, *miR-9* overexpression would also be performed in EC cell lines which have been shown to express low levels of *miR-9* such as 833KE and Tera-1. These cell lines are nullipotent and therefore it would be interesting to test if *miR-9* overexpression would induce differentiation, consistent with its role in neural differentiation (Sempere *et al.*, 2004). This could lead to repression of pluripotency and proliferation genes as a result of differentiation and not the inhibition of translation by *miR-9*. The next step would be to validate the 3 *miR-9* predicted targets that were identified to be downregulated in the *miR-9* over expressing TCam-2 cells from the RNA-seq. This would include a luciferase assay to validate that *miR-9* binds to the 3' UTR of the target genes in 293T cells, as shown in Chapter 5 for other genes (Figure 5.3). These genes would then be analysed for gene expression using RT-QPCR to validate their downregulation by *miR-9*.

In addition to this, RT-QPCR for *miR-9* and validated targets in TGCT patient's samples compared to patient matched testes normal would be performed. This would validate the expression profile observed in TGCT cell lines, in patient

derived samples. Once the *miR-9* predicted targets have been validated, siRNAs would then be made to downregulate the *miR-9* target genes individually using TGCT cell lines. The cell lines would then be analysed for changes in proliferation, pluripotency markers and apoptosis. In addition to this, as *miR-9* has now been validated to target *SMN2*, a component of the spliceosome complex, TCam-2 *miR-9* RNA-seq dataset could be analysed for differences in alternative splicing. Genes shown to be alternatively spliced would be validated using RT-QPCR in TCam-2 *miR-9* and TCam-2 scrambled, but also in TCam-2 cells which have *SMN2* knockdown.

Conclusions

In conclusion, this study used available bioinformatics datasets in order to identify miRNAs regulated by NANOG in TGCTs. In total, 4 miRNAs were identified that were differentially expressed in TGCTs when compared to normal testes and contained NANOG binding <2kb upstream of miRNA TSSs. The expression of these miRNAs were then analysed in TGCT cell lines. Of these *miR-9* demonstrated the same expression profile observed in the bioinformatics analyses. In addition, NANOG transcriptionally repressed *miR-9* expression in TGCTs. This indicated that *miR-9* may play a role in repressing the tumorigenesis and progression in TGCTs. TCam-2 cells that overexpressed *miR-9* were analysed using transcriptomic analysis. This identified 3 genes (*SIK1*, *CD34* and *SNM2*) that were predicted to be direct targets of *miR-9*. These transcripts need to be validated as *miR-9* targets and their function in TGCTs requires further investigation. This study has identified *miR-9* being regulated by the master regulator NANOG. Therefore, *miR-9* targets identified may

advance our knowledge of how NANOG may regulate proliferation and TGCT tumorigenesis. In turn, using TGCT cell lines that resemble PGCs, this project could help to understand the role of NANOG in PGC development.

Bibliography

Abranches, E. *et al.* (2014) 'Stochastic NANOG fluctuations allow mouse embryonic stem cells to explore pluripotency', *Development*. Oxford University Press for The Company of Biologists Limited, 141(14), pp. 2770–2779.

Agnihotri, S. *et al.* (2011) 'A GATA4-regulated tumor suppressor network represses formation of malignant human astrocytomas', *Journal of Experimental Medicine*. Rockefeller University Press, p. jem-20102099.

Albers, P. *et al.* (2015) 'Guidelines on testicular cancer: 2015 update', *European Urology*. Elsevier, 68(6), pp. 1054–1068.

Almstrup, K. *et al.* (2010) 'Carcinoma in situ testis displays permissive chromatin modifications similar to immature foetal germ cells', *British Journal of Cancer*. Nature Publishing Group, 103(8), pp. 1269–1276.

Ambros, V. (2004) 'The functions of animal microRNAs', *Nature*, 431(7006), pp. 350–355.

Andrews, P. W. *et al.* (2005) 'Embryonic stem (ES) cells and embryonal carcinoma (EC) cells: opposite sides of the same coin'. *Biochemical Society Transactions* , 33 (6) pp. 1526-1530

Baek, D. *et al.* (2008) 'The impact of microRNAs on protein output', *Nature*. NIH Public Access, 455(7209), p. 64-71.

Bai, Y. *et al.* (2017) 'Stanniocalcin 1 promotes cell proliferation via cyclin E1/cyclin-dependent kinase 2 in human prostate carcinoma', *Oncology Reports*. Spandidos Publications, 37(4), pp. 2465–2471.

Barrès, V. *et al.* (2010) 'An essential role for Ran GTPase in epithelial ovarian cancer cell survival', *Molecular Cancer*. BioMed Central, 9(1), p. 272.

Barroso-del Jesus, A., Lucena-Aguilar, G. and Menendez, P. (2009) 'The miR-302-367 cluster as a potential stemness regulator in ESCs', *Cell Cycle*. Taylor & Francis, 8(3), pp. 394–398.

Barroso-delJesus, A. *et al.* (2008) 'Embryonic stem cell-specific miR302-367 cluster: human gene structure and functional characterization of its core promoter', *Molecular and Cellular Biology*. Am Soc Microbiol, 28(21), pp. 6609–6619.

Bartel, D. P. (2009) 'MicroRNAs: target recognition and regulatory functions', *Cell*, 136(2), pp. 215–233.

Bazzini, A. A., Lee, M. T. and Giraldez, A. J. (2012) 'Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish', *Science*, 336(6078), pp. 233–237.

Ben-David, U. *et al.* (2014) 'Aneuploidy induces profound changes in gene expression, proliferation and tumorigenicity of human pluripotent stem cells', *Nature Communications*. Nature Publishing Group, 5.

Benetti, R. *et al.* (2008) 'A mammalian microRNA cluster controls DNA methylation and telomere recombination via Rbl2-dependent regulation of DNA methyltransferases', *Nature Structural & Molecular Biology*, 15(3), pp. 268–279.

Berdeaux, R. *et al.* (2007) 'SIK1 is a class II HDAC kinase that promotes survival of skeletal myocytes', *Nature Medicine*. Nature Publishing Group,

13(5), p. 597-603.

Bernstein, B. E. *et al.* (2006) ‘A bivalent chromatin structure marks key developmental genes in embryonic stem cells’, *Cell*. Elsevier, 125(2), pp. 315–326.

Bernstein, E. *et al.* (2003) ‘Dicer is essential for mouse development’, *Nature Genetics*, 35(3), pp. 215–217.

Berry, W. L. and Janknecht, R. (2013) ‘KDM4/JMJD2 histone demethylases: epigenetic regulators in cancer cells’, *Cancer Research*. AACR, 73(10), pp. 2936–2942.

Biermann, K. *et al.* (2007) ‘Genome-wide expression profiling reveals new insights into pathogenesis and progression of testicular germ cell tumors’, *Cancer Genomics-Proteomics*. International Institute of Anticancer Research, 4(5), pp. 359–367.

Birnbaum, L. S. (2013) ‘State of the science of endocrine disruptors’, *Environmental health perspectives*. National Institute of Environmental Health Science, 121(4), p. a107.

Borchert, G. M., Lanier, W. and Davidson, B. L. (2006) ‘RNA polymerase III transcribes human microRNAs’, *Nature Structural & Molecular Biology*. Nature Publishing Group, 13(12), p. 1097-1101.

Bosetti, C. *et al.* (2011) ‘Trends in mortality from urologic cancers in Europe, 1970–2008’, *European Urology*. Elsevier, 60(1), pp. 1–15.

Boyer, L. A. *et al.* (2005) ‘Core transcriptional regulatory circuitry in human embryonic stem cells’, *Cell*, 122(6), pp. 947–956.

Calder, A. *et al.* (2012) 'Lengthened G1 phase indicates differentiation status in human embryonic stem cells', *Stem Cells and Development*. Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA, 22(2), pp. 279–295.

Cancer Research UK (2017) *Testicular cancer statistics, http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/testicular-cancer#heading-Zero*.

Card, D. A. G. *et al.* (2008) 'Oct4/Sox2-regulated miR-302 targets cyclin D1 in human embryonic stem cells', *Molecular and Cellular Biology*. American Society for Microbiology, 28(20), pp. 6426–6438.

Chambers, I. *et al.* (2003) 'Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells', *Cell*, 113(5), pp. 643–655.

Chambers, I. *et al.* (2007) 'Nanog safeguards pluripotency and mediates germline development', *Nature*, 450(7173), pp. 1230–1234.

Chang, W.-F. *et al.* (2015) 'SMN is required for the maintenance of embryonic stem cells and neuronal differentiation in mice', *Brain Structure and Function*. Springer, 220(3), pp. 1539–1553.

Chen, X. *et al.* (2013) 'Effect of bisphenol A on pluripotency of mouse embryonic stem cells and differentiation capacity in mouse embryoid bodies', *Toxicology in vitro*. Elsevier, 27(8), pp. 2249–2255.

Cheng, A. M. *et al.* (2005) 'Antisense inhibition of human miRNAs and indications for an involvement of miRNA in cell growth and apoptosis',

Nucleic Acids Research. Oxford University Press, 33(4), pp. 1290–1297.

Cheng, H. *et al.* (2009) 'SIK1 couples LKB1 to p53-dependent anoikis and suppresses metastasis', *Science Signaling*. NIH Public Access, 2(80), p. ra35.

Chew, J.-L. *et al.* (2005) 'Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells', *Molecular and Cellular Biology*. Am Soc Microbiol, 25(14), pp. 6031–6046.

Chin, K. *et al.* (2006) 'Genomic and transcriptional aberrations linked to breast cancer pathophysiologies', *Cancer Cell*. Elsevier, 10(6), pp. 529–541.

Chiou, S.-H. *et al.* (2008) 'Positive correlations of Oct-4 and Nanog in oral cancer stem-like cells and high-grade oral squamous cell carcinoma', *Clinical Cancer Research*. AACR, 14(13), pp. 4085–4095.

Clark, A. J. L. (2016) '60 YEARS OF POMC: The proopiomelanocortin gene: discovery, deletion and disease', *Journal of Molecular Endocrinology*. Society of Endocrinology, 56(4), pp. T27–T37.

Clark, A. T. *et al.* (2004) 'Human STELLAR, NANOG, and GDF3 genes are expressed in pluripotent cells and map to chromosome 12p13, a hotspot for teratocarcinoma', *Stem Cells*, 22(2), pp. 169–179.

Clarke, P. R. and Zhang, C. (2008) 'Spatial and temporal coordination of mitosis by Ran GTPase', *Nature Reviews: Molecular Cell Biology*. Nature Publishing Group, 9(6), p. 464-477.

Coolen, M., Katz, S. and Bally-Cuif, L. (2013) 'miR-9: a versatile regulator of neurogenesis', *Frontiers in Cellular Neuroscience*, 7, p.220.

Cosgrove, M. S., Boeke, J. D. and Wolberger, C. (2004) 'Regulated

nucleosome mobility and the histone code.', *Nature Structural & Molecular Biology*, 11(11), p.1037-1043.

Croft, D. *et al.* (2010) 'Reactome: a database of reactions, pathways and biological processes', *Nucleic Acids Research*. Oxford University Press, 39(issue_suppl_1), pp. D691–D697.

Darnell, D. K. *et al.* (2006) 'MicroRNA expression during chick embryo development', *Developmental Dynamics*. Wiley Online Library, 235(11), pp. 3156–3165.

Davis, H. E., Morgan, J. R. and Yarmush, M. L. (2002) 'Polybrene increases retrovirus gene transfer efficiency by enhancing receptor-independent virus adsorption on target cell membranes', *Biophysical Chemistry*. Elsevier, 97(2), pp. 159–172.

Deb-Rinker, P. *et al.* (2005) 'Sequential DNA methylation of the Nanog and Oct-4 upstream regions in human NT2 cells during neuronal differentiation', *Journal of Biological Chemistry*. ASBMB, 280(8), pp. 6257–6260.

DeBerardinis, R. J. *et al.* (2008) 'The biology of cancer: metabolic reprogramming fuels cell growth and proliferation', *Cell Metabolism*. Elsevier, 7(1), pp. 11–20.

Del-Mazo, J. *et al.* (2013) 'Endocrine disruptors, gene deregulation and male germ cell tumors', *International Journal of Developmental Biology*. UPV/EHU Press, 57(2-3–4), pp. 225–239.

Denli, A. M. *et al.* (2004) 'Processing of primary microRNAs by the Microprocessor complex', *Nature*, 432(7014), pp. 231–235.

Dieckmann, K. P. *et al.* (2012) 'MicroRNAs miR-371-3 in serum as diagnostic tools in the management of testicular germ cell tumours', *British Journal of Cancer*. Nature Publishing Group, 107(10), p. 1754-1760.

Djuranovic, S., Nahvi, A. and Green, R. (2012) 'miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay', *Science*, 336(6078), pp. 237–240.

Dobin, A. *et al.* (2013) 'STAR: ultrafast universal RNA-seq aligner', *Bioinformatics*. Oxford University Press, 29(1), pp. 15–21.

Doench, J. G. and Sharp, P. A. (2004) 'Specificity of microRNA target selection in translational repression', *Genes & Development*. Cold Spring Harbor Lab, 18(5), pp. 504–511.

Dong, X. *et al.* (2004) 'PlasMapper: a web server for drawing and auto-annotating plasmid maps', *Nucleic Acids Research*. Oxford University Press, 32(suppl_2), pp. W660–W664.

Dweep, H. *et al.* (2011) 'miRWalk–database: prediction of possible miRNA binding sites by "walking" the genes of three genomes', *Journal of Biomedical Informatics*. Elsevier, 44(5), pp. 839–847.

Dweep, H. and Gretz, N. (2015) 'miRWalk2. 0: a comprehensive atlas of microRNA-target interactions', *Nature Methods*. Nature Publishing Group, 12(8), p. 697.

Eckert, D. *et al.* (2008) 'TCam-2 but not JKT-1 cells resemble seminoma in cell culture', *Cell and Tissue Research*. Springer, 331(2), pp. 529–538.

Eda, A. *et al.* (2009) 'Systematic gene regulation involving miRNAs during

neuronal differentiation of mouse P19 embryonic carcinoma cell', *Biochemical and Biophysical Research Communications*. Elsevier, 388(4), pp. 648–653.

Emilsson, L. *et al.* (2015) 'Review of 103 Swedish healthcare quality registries', *Journal of Internal Medicine*. Wiley Online Library, 277(1), pp. 94–136.

Esquela-Kerscher, A. and Slack, F. J. (2006) 'Oncomirs--microRNAs with a role in cancer', *Nature Reviews. Cancer*. Nature Publishing Group, 6(4), p. 259-269.

Ewen, K. A. and Koopman, P. (2010) 'Mouse germ cell development: from specification to sex determination', *Molecular and Cellular Endocrinology*. Elsevier, 323(1), pp. 76–93.

Ezeh, U. I. *et al.* (2005) 'Human embryonic stem cell genes OCT4, NANOG, STELLAR, and GDF3 are expressed in both seminoma and breast carcinoma', *Cancer*, 104(10), pp. 2255–2265.

Fagerberg, L. *et al.* (2014) 'Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics', *Molecular & Cellular Proteomics*. ASBMB, 13(2), pp. 397–406.

Feng, B. *et al.* (2009) 'Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb', *Nature Cell Biology*. Nature Publishing Group, 11(2), pp. 197–203.

Ferlay, J. (2010) 'Globocan 2008, Cancer incidence and mortality worldwide: IARC CancerBase No. 10', *http://globocan. iarc. fr*. International Agency for Research on Cancer.

Fernández, M. F. *et al.* (2016) ‘Bisphenol A and other phenols in human placenta from children with cryptorchidism or hypospadias’, *Reproductive Toxicology*. Elsevier, 59, pp. 89–95.

Festuccia, N. *et al.* (2012) ‘Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells’, *Cell Stem Cell*. Elsevier, 11(4), pp. 477–490.

Filipowicz, W., Bhattacharyya, S. N. and Sonenberg, N. (2008) ‘Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight?’, *Nature Reviews. Genetics*. Nature Publishing Group, 9(2), p. 102-114.

Fischer, U., Liu, Q. and Dreyfuss, G. (1997) ‘The SMN–SIP1 complex has an essential role in spliceosomal snRNP biogenesis’, *Cell*. Elsevier, 90(6), pp. 1023–1029.

Fog, C. K., Galli, G. G. and Lund, A. H. (2012) ‘PRDM proteins: important players in differentiation and disease’, *Bioessays*. Wiley Online Library, 34(1), pp. 50–60.

Fogh, J. and Trempe, G. (1975) ‘New human tumor cell lines’, in *Human tumor cells in vitro*. Springer, pp. 115–159.

Fowler, J. E. *et al.* (1982) ‘Commercial radioimmunoassay for beta subunit of human chorionic gonadotropin: falsely positive determinations due to elevated serum luteinizing hormone’, *Cancer*. Wiley Online Library, 49(1), pp. 136–139.

Friedman, R. C. *et al.* (2009) ‘Most mammalian mRNAs are conserved targets of microRNAs’, *Genome Research*, 19(1), pp. 92–105.

Füllgrabe, J., Kavanagh, E. and Joseph, B. (2011) ‘Histone onco-modifications’, *Oncogene*. Nature Publishing Group, 30(31), p. 3391-3403.

Germa-Lluch, J. R. *et al.* (2002) ‘Clinical pattern and therapeutic results achieved in 1490 patients with germ-cell tumours of the testis: the experience of the Spanish Germ-Cell Cancer Group (GG)’, *European Urology*. Elsevier, 42(6), pp. 553–563.

Gilligan, T. D. *et al.* (2010) ‘ASCO clinical practice guideline on uses of serum tumor markers in adult males with germ cell tumors’, *Journal of Oncology Practice*. American Society of Clinical Oncology Alexandria, VA, 6(4), pp. 199–202.

Gillis, A. J. M. *et al.* (2007) ‘High-throughput microRNAome analysis in human germ cell tumours’, *The Journal of Pathology*. Wiley Online Library, 213(3), pp. 319–328.

Gillis, A. J. M. *et al.* (2011) ‘Expression and interdependencies of pluripotency factors LIN28, OCT3/4, NANOG and SOX2 in human testicular germ cells and tumours of the testis’, *International Journal of Andrology*. Wiley Online Library, 34, pp. e160-e174

Gkountela, S. *et al.* (2013) ‘The ontogeny of cKIT+ human primordial germ cells proves to be a resource for human germ line reprogramming, imprint erasure and in vitro differentiation’, *Nature Cell Biology*, 15(1), pp. 113–122.

Gkountela, S. *et al.* (2015) ‘DNA demethylation dynamics in the human prenatal germline’, *Cell*. Elsevier, 161(6), pp. 1425–1436.

Gonsalvez, G. B. *et al.* (2007) ‘Two distinct arginine methyltransferases are

required for biogenesis of Sm-class ribonucleoproteins', *Journal of Cellular Biology*. Rockefeller University Press, 178(5), pp. 733–740.

Greve, T. S., Judson, R. L. and Blelloch, R. (2013) 'microRNA control of mouse and human pluripotent stem cell behavior', *Annual Review of Cell and Developmental Biology*, 29, pp. 213–239.

Grice, S. J. and Liu, J.-L. (2011) 'Survival motor neuron protein regulates stem cell division, proliferation, and differentiation in Drosophila', *PLoS Genetics*. Public Library of Science, 7(4), p. e1002030.

Guo, F. *et al.* (2015) 'The transcriptome and DNA methylome landscapes of human primordial germ cells', *Cell*. Elsevier, 161(6), pp. 1437–1452.

Guo, H. *et al.* (2010) 'Mammalian microRNAs predominantly act to decrease target mRNA levels', *Nature*, 466(7308), pp. 835–840.

Guttilla, I. K. and White, B. A. (2009) 'Coordinate regulation of FOXO1 by miR-27a, miR-96, and miR-182 in breast cancer cells', *Journal of Biological Chemistry*. ASBMB, 284(35), pp. 23204–23216.

Hajkova, P. *et al.* (2002) 'Epigenetic reprogramming in mouse primordial germ cells', *Mechanisms of Development*. Elsevier, 117(1), pp. 15–23.

Hajkova, P. *et al.* (2008) 'Chromatin dynamics during epigenetic reprogramming in the mouse germ line', *Nature*, 452(7189), pp. 877–881.

Hamel, S. *et al.* (1987) 'Phenotype and function of bone marrow-derived T and non-T cells activated in vitro by alpha-fetoprotein', *Biological Activities of Alpha-fetoprotein*. Florida, pp. 167–177.

Hammond, S. M. *et al.* (2000) 'An RNA-directed nuclease mediates post-

transcriptional gene silencing in Drosophila cells', *Nature*, 404(6775), pp. 293–296.

Han, J. *et al.* (2004) 'The Drosha-DGCR8 complex in primary microRNA processing', *Genes & Development*. Cold Spring Harbor Lab, 18(24), pp. 3016–3027.

Han, J. *et al.* (2012) 'RNA interference-mediated silencing of NANOG reduces cell proliferation and induces G0/G1 cell cycle arrest in breast cancer cells', *Cancer Letters*. Elsevier, 321(1), pp. 80–88.

Han, M.-H. *et al.* (2016) 'Growth and differentiation factor 3 Is transcriptionally regulated by OCT4 in human embryonic carcinoma cells', *Biological and Pharmaceutical Bulletin*. The Pharmaceutical Society of Japan, 39(11), pp. 1802–1808.

Hart, A. H. *et al.* (2005) 'The pluripotency homeobox gene NANOG is expressed in human germ cell tumors', *Cancer*. Wiley Online Library, 104(10), pp. 2092–2098.

Hayashi, K. *et al.* (2002) 'SMAD1 signaling is critical for initial commitment of germ cell lineage from mouse epiblast', *Mechanisms of Development*. Elsevier, 118(1), pp. 99–109.

Hayashi, K. *et al.* (2008) 'MicroRNA biogenesis is required for mouse primordial germ cell development and spermatogenesis', *PloS One*. Public Library of Science, 3(3), p. e1738.

Heidenreich, A. and Moul, J. W. (2002) 'Contralateral testicular biopsy procedure in patients with unilateral testis cancer: is it indicated?', in *Seminars*

in Urologic Oncology, pp. 234–238.

Heimdal, K. *et al.* (1996) 'Familial testicular cancer in Norway and southern Sweden.', *British Journal of Cancer*. Nature Publishing Group, 73(7), p. 964-969.

Heng, J.-C. D. *et al.* (2010) 'The nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells', *Cell Stem Cell*. Elsevier, 6(2), pp. 167–174.

Heo, I. *et al.* (2008) 'Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA', *Molecular Cell*. Elsevier, 32(2), pp. 276–284.

Hoei-Hansen, C. E. *et al.* (2005) 'Stem cell pluripotency factor NANOG is expressed in human fetal gonocytes, testicular carcinoma in situ and germ cell tumours', *Histopathology*, 47(1), pp. 48–56.

Hoffmann, R. *et al.* (2014) 'Innovations in health care and mortality trends from five cancers in seven European countries between 1970 and 2005', *International Journal of Public Health*. Springer, 59(2), pp. 341–350.

Hollstein, M. *et al.* (1991) 'p53 mutations in human cancers', *Science*. JSTOR, 253(5015), pp. 49–53.

de Hoon, M. J. L. *et al.* (2004) 'Open source clustering software', *Bioinformatics*. Oxford University Press, 20(9), pp. 1453–1454.

Huang, C. *et al.* (2010) 'Dual-specificity histone demethylase KIAA1718 (KDM7A) regulates neural differentiation through FGF4', *Cell Research*. Nature Publishing Group, 20(2), pp. 154–165.

Huang, X. *et al.* (2012) 'CD99 triggers upregulation of miR-9-modulated

PRDM1/BLIMP1 in Hodgkin/Reed–Sternberg cells and induces redifferentiation', *International Journal of Cancer*. Wiley Online Library, 131(4). pp E382-E394.

Hutson, J. M. (2015) 'Cryptorchidism and hypospadias'. MDText. com, Inc.

Huttlin, E. L. *et al.* (2017) 'Architecture of the human interactome defines protein communities and disease networks', *Nature*. Nature Research, 545(7655), pp. 505–509.

Iorio, M. V *et al.* (2005) 'MicroRNA gene expression deregulation in human breast cancer', *Cancer Research*. AACR, 65(16), pp. 7065–7070.

Irie, N. *et al.* (2015) 'SOX17 is a critical specifier of human primordial germ cell fate', *Cell*. Elsevier, 160(1), pp. 253–268.

Jeon, H.-M. *et al.* (2011) 'ID4 imparts chemoresistance and cancer stemness to glioma cells by derepressing miR-9*–mediated suppression of SOX2', *Cancer research*. AACR, 71(9), pp. 3410–3421.

Jeter, C. R. *et al.* (2011) 'NANOG promotes cancer stem cell characteristics and prostate cancer resistance to androgen deprivation', *Oncogene*. NIH Public Access, 30(36), p. 3833-3845.

John, B. *et al.* (2004) 'Human microRNA targets', *PLoS Biology*. Public Library of Science, 2(11), p. e363.

Johnson, C. D. *et al.* (2007) 'The let-7 microRNA represses cell proliferation pathways in human cells', *Cancer Research*. AACR, 67(16), pp. 7713–7722.

Jones, P. A. (2012) 'Functions of DNA methylation: islands, start sites, gene bodies and beyond', *Nature Reviews Genetics*. Nature Publishing Group, 13(7),

pp. 484–492.

de Jong, J. *et al.* (2005) ‘Diagnostic value of OCT3/4 for pre-invasive and invasive testicular germ cell tumours’, *The Journal of Pathology*. Wiley Online Library, 206(2), pp. 242–249.

de Jong, J. *et al.* (2008) ‘Further characterization of the first seminoma cell line TCam-2’, *Genes, Chromosomes and Cancer*. Wiley Online Library, 47(3), pp. 185–196.

De Jong, J. *et al.* (2008) ‘Differential expression of SOX17 and SOX2 in germ cells and stem cells has biological and clinical implications’, *The Journal of Pathology*. Wiley Online Library, 215(1), pp. 21–30.

Kanatsu-Shinohara, M. and Shinohara, T. (2006) ‘The germ of pluripotency’, *Nature Biotechnology*. Nature Publishing Group, 24(6), pp. 663–664.

Kanellopoulou, C. *et al.* (2005) ‘Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing’, *Genes & Development*, 19(4), pp. 489–501.

Kang, M. *et al.* (2017) ‘Histone demethylase KDM7a to control androgen receptor activity in hormone-sensitive prostate cancer.’ American Society of Clinical Oncology. 35(6), pp. 262-262.

Kashyap, V. *et al.* (2009) ‘Regulation of stem cell pluripotency and differentiation involves a mutual regulatory circuit of the NANOG, OCT4, and SOX2 pluripotency transcription factors with polycomb repressive complexes and stem cell microRNAs’, *Stem cells and development*. Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA, 18(7), pp.

1093–1108.

Kehler, J. *et al.* (2004) ‘Oct4 is required for primordial germ cell survival’, *EMBO Reports*. EMBO Press, 5(11), pp. 1078–1083.

Ketting, R. F. *et al.* (2001) ‘Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans’, *Genes & Development*. Cold Spring Harbor Lab, 15(20), pp. 2654–2659.

Khvorova, A., Reynolds, A. and Jayasena, S. D. (2003) ‘Functional siRNAs and miRNAs exhibit strand bias’, *Cell*, 115(2), pp. 209–216.

Kim, T. H. *et al.* (2005) ‘A high-resolution map of active promoters in the human genome’, *Nature*. Nature Publishing Group, 436(7052), pp. 876–880.

Kimura, T. *et al.* (1999) ‘Molecular cloning and genomic organization of mouse homologue of Drosophila germ cell-less and its expression in germ lineage cells’, *Biochemical and Biophysical Research Communications*. Elsevier, 262(1), pp. 223–230.

Kirk, J. M. W. *et al.* (1994) ‘Gonadal function and response to human chorionic and menopausal gonadotrophin therapy in male patients with idiopathic hypogonadotrophic hypogonadism’, *Clinical Endocrinology*. Wiley Online Library, 41(1), pp. 57–63.

Klepp, O. *et al.* (1990) ‘Early clinical stages (CS1, CS1Mk+ and CS2A) of non-seminomatous testis cancer: Value of pre-and post-orchiectomy serum tumor marker information in prediction of retroperitoneal lymph node metastases’, *Annals of Oncology*. Oxford University Press, 1(4), pp. 281–288.

Klepp, O. *et al.* (1991) ‘Early Clinical Stages of Nonseminomatous Testis

Cancer Evaluation of the Primary Treatment and Follow-up Procedures of the Swenoteca Project', *Scandinavian Journal of Urology and Nephrology*. Taylor & Francis, 25(3), pp. 179–190.

Kloosterman, W. P. and Plasterk, R. H. A. (2006) 'The diverse functions of microRNAs in animal development and disease', *Developmental Cell*. Elsevier, 11(4), pp. 441–450.

Knaut, H. *et al.* (2003) 'A zebrafish homologue of the chemokine receptor Cxcr4 is a germ-cell guidance receptor', *Nature*. Nature Publishing Group, 421(6920), pp. 279–282.

Kobayashi, T. *et al.* (2017) 'Principles of early human development and germ cell program from conserved model systems', *Nature*. Nature Research, 546(7658), pp. 416–420.

Kollmannsberger, C. *et al.* (2014) 'Patterns of relapse in patients with clinical stage I testicular cancer managed with active surveillance', *Journal of Clinical Oncology*. American Society of Clinical Oncology, 33(1), pp. 51–57.

Krichevsky, A. M. *et al.* (2006) 'Specific microRNAs modulate embryonic stem cell–derived neurogenesis', *Stem Cells*. Wiley Online Library, 24(4), pp. 857–864.

Kurimoto, K. *et al.* (2015) 'Quantitative dynamics of chromatin remodeling during germ cell specification from mouse embryonic stem cells', *Cell Stem Cell*. Elsevier, 16(5), pp. 517–532.

Lagos-Quintana, M. *et al.* (2002) 'Identification of tissue-specific microRNAs from mouse', *Current Biology*. Elsevier, 12(9), pp. 735–739.

Lan, C.-W. *et al.* (2013) 'Differentiation of human embryonic stem cells into functional ovarian granulosa-like cells', *The Journal of Clinical Endocrinology & Metabolism*. Oxford University Press, 98(9), pp. 3713–3723.

Landthaler, M., Yalcin, A. and Tuschl, T. (2004) 'The human DiGeorge syndrome critical region gene 8 and its D. melanogaster homolog are required for miRNA biogenesis', *Current Biology*, 14(23), pp. 2162–2167.

Lee, H. *et al.* (2012) 'Expression of miRNAs and PTEN in endometrial specimens ranging from histologically normal to hyperplasia and endometrial adenocarcinoma', *Modern Pathology*. Nature Publishing Group, 25(11), pp. 1508–1515.

Lee, Y.-H. *et al.* (2017) 'Low-cell-number, single-tube amplification (STA) of total RNA revealed transcriptome changes from pluripotency to endothelium', *BMC Biology*. BioMed Central, 15(1), p. 22.

Lefebvre, S. *et al.* (1995) 'Identification and characterization of a spinal muscular atrophy-determining gene', *Cell*. Elsevier, 80(1), pp. 155–165.

Leffert, H. L. and Sell, S. (1974) 'Alpha1-fetoprotein biosynthesis during the growth cycle of differentiated fetal rat hepatocytes in primary monolayer culture', *The Journal of Cell Biology*. The Rockefeller University Press, 61(3), p. 823-829.

Leucht, C. *et al.* (2008) 'MicroRNA-9 directs late organizer activity of the midbrain-hindbrain boundary', *Nature Neuroscience*. Nature Publishing Group, 11(6), pp. 641–648.

Lewis, B. P. *et al.* (2003) 'Prediction of mammalian microRNA targets', *Cell*.

Elsevier, 115(7), pp. 787–798.

Li, L. *et al.* (2017) 'Evaluation and comparison of three assays for molecular detection of spinal muscular atrophy', *Clinical Chemistry and Laboratory Medicine (CCLM)*, 55(3), pp. 358–367.

Li, Y. *et al.* (2003) 'Lats2, a putative tumor suppressor, inhibits G1/S transition', *Oncogene*. Nature Publishing Group, 22(28), p. 4398.

Lin, S.-L. *et al.* (2010) 'MicroRNA miR-302 inhibits the tumorigenecity of human pluripotent stem cells by coordinate suppression of the CDK2 and CDK4/6 cell cycle pathways', *Cancer Research*. AACR, 70(22), pp. 9473–9482.

Linzer, D. I. H. and Levine, A. J. (1979) 'Characterization of a 54K dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells', *Cell*. Elsevier, 17(1), pp. 43–52.

Litchfield, K. *et al.* (2017) 'Identification of 19 new risk loci and potential regulatory mechanisms influencing susceptibility to testicular germ cell tumor', *Nature Genetics*. Nature Research, 49(7), pp. 1133–1140.

Liu, D., Song, H. and Xu, Y. (2010) 'A common gain of function of p53 cancer mutants in inducing genetic instability', *Oncogene*. NIH Public Access, 29(7), p. 949-956.

Liu, N. and Olson, E. N. (2010) 'MicroRNA regulatory networks in cardiovascular development', *Developmental Cell*. Elsevier, 18(4), pp. 510–525.

Lleres, D. *et al.* (2010) 'Direct interaction between hnRNP-M and

CDC5L/PLRG1 proteins affects alternative splice site choice', *EMBO reports*. EMBO Press, 11(6), pp. 445–451.

Loh, Y.-H. *et al.* (2006) 'The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells', *Nature Genetics*. Nature Publishing Group, 38(4), pp. 431–440.

Loh, Y.-H. *et al.* (2007) 'Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells', *Genes & Development*. Cold Spring Harbor Lab, 21(20), pp. 2545–2557.

Looijenga, L. H. J. *et al.* (2003) 'POU5F1 (OCT3/4) identifies cells with pluripotent potential in human germ cell tumors', *Cancer Research*. AACR, 63(9), pp. 2244–2250.

Lu, J. *et al.* (2005) 'MicroRNA expression profiles classify human cancers', *Nature*. Nature Publishing Group, 435(7043), p. 834.

Lu, X. *et al.* (2008) 'Predicting features of breast cancer with gene expression patterns', *Breast Cancer Research and Treatment*. Springer, 108(2), p. 191-201.

Luger, K. *et al.* (1997) 'Crystal structure of the nucleosome core particle at 2.8 Å resolution', *Nature*. Nature Publishing Group, 389(6648), pp. 251–260.

Lytle, J. R., Yario, T. A. and Steitz, J. A. (2007) 'Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5′ UTR as in the 3′ UTR', *Proceedings of the National Academy of Sciences*. National Acad Sciences, 104(23), pp. 9667–9672.

Ma, L. *et al.* (2010) 'miR-9, a MYC/MYCN-activated microRNA, regulates E-

cadherin and cancer metastasis', *Nature Cell Biology*. Nature Publishing Group, 12(3), pp. 247–256.

Magnúsdóttir, E. *et al.* (2013) 'A tripartite transcription factor network regulates primordial germ cell specification in mice', *Nature Cell Biology*. Nature Publishing Group, 15(8), pp. 905–915.

Mandal, C. *et al.* (2016) 'RNA sequencing reveals the alteration of the expression of novel genes in ethanol-treated embryoid bodies', *PloS One*. Public Library of Science, 11(3), p. e0149976.

Marson, A. *et al.* (2008) 'Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells', *Cell*, 134(3), pp. 521–533.

Masterson, T. A., Rice, K. R. and Beck, S. D. W. (2014) 'Current and future biologic markers for disease progression and relapse in testicular germ cell tumors: a review', in *Urologic Oncology: Seminars and Original Investigations*. Elsevier, pp. 261–271.

Matera, A. G., Terns, R. M. and Terns, M. P. (2007) 'Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs', *Nature Reviews: Molecular Cell Biology*. Nature Publishing Group, 8(3), p. 209.

Mathews, D. H. *et al.* (2004) 'Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure', *Proceedings of the National Academy of Sciences of the United States of America*. National Acad Sciences, 101(19), pp. 7287–7292.

Matsui, Y., Zsebo, K. and Hogan, B. L. M. (1992) 'Derivation of pluripotential

embryonic stem cells from murine primordial germ cells in culture', *Cell*, 70(5), pp. 841–847.

McGlynn, K. A. *et al.* (2003) 'Trends in the incidence of testicular germ cell tumors in the United States', *Cancer*. Wiley Online Library, 97(1), pp. 63–70.

McGlynn, K. A. and Cook, M. B. (2009) 'Etiologic factors in testicular germ-cell tumors', *Future oncology*. Future Medicine, 5(9), pp. 1389–1402.

McGlynn, K. A. and Cook, M. B. (2010) 'The epidemiology of testicular cancer', in *Male Reproductive Cancers*. Springer, pp. 51–83.

McLean, C. Y. *et al.* (2010) 'GREAT improves functional interpretation of cis-regulatory regions', *Nature Biotechnology*. Nature Research, 28(5), pp. 495–501.

Medina-Rivera, A. *et al.* (2015) 'RSAT 2015: regulatory sequence analysis tools', *Nucleic Acids Research*. Oxford University Press, 43(W1), pp. W50–W56.

Meier, R. *et al.* (2007) 'The chemokine receptor CXCR4 strongly promotes neuroblastoma primary tumour and metastatic growth, but not invasion', *PLoS One*. Public Library of Science, 2(10), p. e1016.

Melton, C., Judson, R. and Blelloch, R. (2010) 'Opposing microRNA families regulate self-renewal in mouse embryonic stem cells', *Nature*. NIH Public Access, 463(7281), p. 621-626.

Meng, H.-M. *et al.* (2010) 'Over-expression of Nanog predicts tumor progression and poor prognosis in colorectal cancer', *Cancer Biology & Therapy*. Taylor & Francis, 9(4), pp. 295–302.

Mester, B. *et al.* (2010) ‘Occupational causes of testicular cancer in adults’, *The International Journal of Occupational and Environmental Medicine*, 1(4 October). pp160-170.

Mihelich, B. L. *et al.* (2011) ‘miR-183-96-182 cluster is overexpressed in prostate tissue and regulates zinc homeostasis in prostate cells’, *Journal of Biological Chemistry*. ASBMB, 286(52), pp. 44503–44511.

Miranda, K. C. *et al.* (2006) ‘A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes’, *Cell*. Elsevier, 126(6), pp. 1203–1217.

Mitsui, K. *et al.* (2003) ‘The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells’, *Cell*. Elsevier, 113(5), pp. 631–642.

Molyneaux, K. A. *et al.* (2003) ‘The chemokine SDF1/CXCL12 and its receptor CXCR4 regulate mouse germ cell migration and survival’, *Development*. The Company of Biologists Ltd, 130(18), pp. 4279–4286.

Morin, R. D. *et al.* (2008) ‘Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells’, *Genome Research*. Cold Spring Harbor Lab, 18(4), pp. 610–621.

Mostert, M. M. C. *et al.* (1996) ‘Fluorescence in situ hybridization-based approaches for detection of 12p overrepresentation, in particular i (12p), in cell lines of human testicular germ cell tumors of adults’, *Cancer Genetics and Cytogenetics*. Elsevier, 87(2), pp. 95–102.

Mostofi, F. K., Sesterhenn, I. A. and Davis Jr, C. J. (1987) ‘Immunopathology

of germ cell tumors of the testis.', in *Seminars in Diagnostic Pathology*, pp. 320–341.

Murakami, K. *et al.* (2016) 'NANOG alone induces germ cells in primed epiblast in vitro by activation of enhancers', *Nature*. Nature Publishing Group. 529(7586), p.403-407.

Murray, M. J. *et al.* (2011) 'Identification of microRNAs from the miR-371~373 and miR-302 clusters as potential serum biomarkers of malignant germ cell tumors', *American Journal of Clinical Pathology*. The Oxford University Press, 135(1), pp. 119–125.

Neganova, I. *et al.* (2009) 'Expression and functional analysis of G1 to S regulatory components reveals an important role for CDK2 in cell cycle regulation in human embryonic stem cells', *Oncogene*. Nature Publishing Group, 28(1), pp. 20–30.

Nettersheim, D. *et al.* (2011) 'The seminoma cell line TCam-2 is sensitive to HDAC inhibitor depsipeptide but tolerates various other chemotherapeutic drugs and loss of NANOG expression', *Genes, Chromosomes and Cancer*. Wiley Online Library, 50(12), pp. 1033–1042.

Nettersheim, D. *et al.* (2015) 'BMP inhibition in seminomas initiates acquisition of pluripotency via NODAL signaling resulting in reprogramming to an embryonal carcinoma', *PLoS Genetics*. Public Library of Science, 11(7), p. e1005415.

Ng, J.-H. *et al.* (2013) 'In vivo epigenomic profiling of germ cells reveals germ cell molecular signatures', *Developmental Cell*. Elsevier, 24(3), pp. 324–333.

Nichols, C. R. *et al.* (1998) ‘Randomized comparison of cisplatin and etoposide and either bleomycin or ifosfamide in treatment of advanced disseminated germ cell tumors: an Eastern Cooperative Oncology Group, Southwest Oncology Group, and Cancer and Leukemia Group B Study.’, *Journal of Clinical Oncology*, 16(4), pp. 1287–1293.

Nie, K. *et al.* (2008) ‘MicroRNA-mediated down-regulation of PRDM1/Blimp-1 in Hodgkin/Reed-Sternberg cells: a potential pathogenetic lesion in Hodgkin lymphomas’, *The American Journal of Pathology*. Elsevier, 173(1), pp. 242–252.

Nimura, M. *et al.* (2006) ‘Spatial and temporal patterns of expression of melanocortin type 2 and 5 receptors in the fetal mouse tissues and organs’, *Anatomy and Embryology*. Springer, 211(2), pp. 109–117.

Nothwang, H. G. *et al.* (1998) ‘Identification of a novel Ran binding protein 2 related gene (RANBP2L1) and detection of a gene cluster on human chromosome 2q11–q12’, *Genomics*. Elsevier, 47(3), pp. 383–392.

Novotny, G. W. *et al.* (2012) ‘MicroRNA expression profiling of carcinoma in situ cells of the testis’, *Endocrine-Related Cancer*, 19(3), pp. 365–379.

Nowak, J. S. *et al.* (2014) ‘Lin28a regulates neuronal differentiation and controls miR-9 production’, *Nature Communications*. Europe PMC Funders, 5, p. 3687.

O’Shaughnessy, P. J. *et al.* (2003) ‘Adrenocorticotropic hormone directly stimulates testosterone production by the fetal and neonatal mouse testis’, *Endocrinology*. Oxford University Press, 144(8), pp. 3279–3284.

Ohinata, Y. *et al.* (2005) ‘Blimp1 is a critical determinant of the germ cell lineage in mice’, *Nature*, 436(7048), pp. 207–213.

Okamura, K. *et al.* (2008) ‘The regulatory activity of microRNA* species has substantial influence on microRNA and 3′ UTR evolution’, *Nature Structural & Molecular Biology*. Nature Publishing Group, 15(4), pp. 354–363.

Olive, K. P. *et al.* (2004) ‘Mutant p53 gain of function in two mouse models of Li-Fraumeni syndrome’, *Cell*. Elsevier, 119(6), pp. 847–860.

Oosterhuis, J. W. and Looijenga, L. H. J. (2005) ‘Testicular germ-cell tumours in a broader perspective’, *Nature Reviews Cancer*, 5(3), pp. 210–222.

Pal, R. and Ravindran, G. (2006) ‘Assessment of pluripotency and multilineage differentiation potential of NTERA-2 cells as a model for studying human embryonic stem cells’, *Cell Proliferation*. Wiley Online Library, 39(6), pp. 585–598.

Palmer, R. D. *et al.* (2010) ‘Malignant germ cell tumors display common microRNA profiles resulting in global changes in expression of messenger RNA targets’, *Cancer Research*. AACR, 70(7), pp. 2911–2923.

Palumbo, C. *et al.* (2002) ‘Expression of the PDGF α-receptor 1.5 kb transcript, OCT-4, and c-KIT in human normal and malignant tissues. Implications for the early diagnosis of testicular germ cell tumours and for our understanding of regulatory mechanisms’, *The Journal of Pathology*. Wiley Online Library, 196(4), pp. 467–477.

Pan, M.-R. *et al.* (2015) ‘G9a orchestrates PCL3 and KDM7A to promote histone H3K27 methylation’, *Scientific Reports*. Nature Publishing Group, 5, p.

18709.

Park, P. J. (2009) ‘ChIP–seq: advantages and challenges of a maturing technology’, *Nature Reviews Genetics*. Nature Publishing Group, 10(10), pp. 669–680.

Patel, S. R., Kvols, L. K. and Richardson, R. L. (1990) ‘Familial testicular cancer: report of six cases and review of the literature’, in *Mayo Clinic Proceedings*. Elsevier, pp. 804–808.

Peckham, M. J. *et al.* (1979) ‘Combined management of malignant teratoma of the testis’, *The Lancet*. Elsevier, 314(8137), pp. 267–270.

Perkins, G. L. *et al.* (2003) ‘Serum tumor markers’, *American family physician*. WB Saunders Co, 68(6), pp. 1075–1088.

Quimby, B. B. and Dasso, M. (2003) ‘The small GTPase Ran: interpreting the signs’, *Current Opinion in Cell Biology*. Elsevier, 15(3), pp. 338–344.

Rapley, E. A. *et al.* (2009) ‘A genome-wide association study of testicular germ cell tumor’, *Nature Genetics*. Nature Publishing Group, 41(7), pp. 807–810.

Rashid, H. H. *et al.* (2004) ‘Testicular microlithiasis: a review and its association with testicular cancer’, in *Urologic Oncology: Seminars and Original Investigations*. Elsevier, pp. 285–289.

Ren, Z.-G. *et al.* (2016) ‘miR-203 promotes proliferation, migration and invasion by degrading SIK1 in pancreatic cancer’, *Oncology Reports*. Spandidos Publications, 35(3), pp. 1365–1374.

Resnick, J. L. *et al.* (1992) ‘Long-term proliferation of mouse primordial germ

cells in culture'. Nature, 359(6395), p.550-551.

Reuter, V. E. (2005) 'Origins and molecular biology of testicular germ cell tumors.', *Modern Pathology*, 18. pp. S51–S60.

Ries, L. A. G. *et al.* (2008) 'Surveillance, Epidemiology, and End Results (SEER) Program SEER* Stat Data base: Incidence—SEER 9 Regs Public-Use, Nov 2004 Sub (1973–2002). National Cancer Institute, Division of Cancer Control and Population Sciences, Surveillance Research Program, Cancer Research Statistics Branch'. Released April 2005.

Robinson, M. D., McCarthy, D. J. and Smyth, G. K. (2010) 'edgeR: a Bioconductor package for differential expression analysis of digital gene expression data', *Bioinformatics*. Oxford University Press, 26(1), pp. 139–140.

Rodda, D. J. *et al.* (2005) 'Transcriptional regulation of NANOG by OCT4 and SOX2', *Journal of Biological Chemistry*. ASBMB, 280(26), pp. 24731–24737.

Rosen, A. *et al.* (2011) 'Global trends in testicular cancer incidence and mortality', *European Urology*. Elsevier, 60(2), pp. 374–379.

Rosenberg, C. *et al.* (2000) 'Overrepresentation of the short arm of chromosome 12 is related to invasive growth of human testicular seminomas and nonseminomas', *Oncogene*. Nature Publishing Group, 19(51), p. 5858-5862.

Rosenfeld, N. *et al.* (2008) 'MicroRNAs accurately identify cancer tissue origin', *Nature Biotechnology*. Nature Publishing Group, 26(4), p. 462-469.

Saitou, M. *et al.* (2005) 'Blimp1 and the emergence of the germ line during development in the mouse', *Cell Cycle*. Taylor & Francis, 4(12), pp. 1736–

1740.

Saitou, M., Kagiwada, S. and Kurimoto, K. (2012) ‘Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells’, *Development*, 139(1), pp. 15–31.

Saitou, M. and Yamaji, M. (2012) ‘Primordial germ cells in mice’, *Cold Spring Harbor Perspectives in Biology*. Cold Spring Harbor Lab, 4(11), p. a008375.

Saldanha, A. J. (2004) ‘Java Treeview—extensible visualization of microarray data’, *Bioinformatics*. Oxford University Press, 20(17), pp. 3246–3248.

Sasaki, K. *et al.* (2015) ‘Robust in vitro induction of human germ cell fate from pluripotent stem cells’, *Cell Stem Cell*. Elsevier, 17(2), pp. 178–194.

Saunders, P. T. *et al.* (2003) ‘Absence of mDazl produces a final block on germ cell development at meiosis’, *Reproduction*. Societ of Reproduction and Fertility, 126(5), pp. 589–597.

Schöler, H. R. *et al.* (1990) ‘Oct-4: a germline-specific transcription factor mapping to the mouse t-complex.’, *The EMBO journal*. European Molecular Biology Organization, 9(7), p. 2185-2195.

Schorle, H. and Nettersheim, D. (2012) ‘NANOG (Nanog homeobox)’. *Atlas of Genetics and Cytogenetics in Oncology and Haematology*. Atlas Database. 16(10). pp 727-731.

Schwarz, D. S. *et al.* (2003) ‘Asymmetry in the assembly of the RNAi enzyme complex’, *Cell*, 115(2), pp. 199–208.

Seki, Y. *et al.* (2005) ‘Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development

of germ cells in mice', *Developmental Biology*. Elsevier, 278(2), pp. 440–458.

Seki, Y. *et al.* (2007) 'Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice', *Development*. The Company of Biologists Ltd, 134(14), pp. 2627–2638.

Selcuklu, S. D. *et al.* (2012) 'MicroRNA-9 inhibition of cell proliferation and identification of novel miR-9 targets by transcriptome profiling in breast cancer cells', *Journal of Biological Chemistry*. ASBMB, 287(35), pp. 29516–29528.

Sempere, L. F. *et al.* (2004) 'Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation', *Genome Biology*. BioMed Central, 5(3), p. R13.

Sesterhenn, I. A. and Davis Jr, C. J. (2004) 'Pathology of germ cell tumors of the testis', *Cancer Control*, 11(6), pp. 374–387.

Shackelford, D. B. and Shaw, R. J. (2009) 'The LKB1-AMPK pathway: metabolism and growth control in tumor suppression', *Nature Reviews: Cancer*. NIH Public Access, 9(8), p. 563-575.

Shanmugalingam, T. *et al.* (2013) 'Global incidence and outcome of testicular cancer', *Clinical epidemiology*. Dove Press, 5, p. 417-427.

Sharpe, N. G., Williams, D. G. and Latchman, D. S. (1990) 'Regulated expression of the small nuclear ribonucleoprotein particle protein SmN in embryonic stem cell differentiation.', *Molecular and Cellular Biology*. American Society of Microbiology, 10(12), pp. 6817–6820.

Sheikine, Y. *et al.* (2012) 'Molecular genetics of testicular germ cell tumors',

American journal of cancer research. e-Century Publishing Corporation, 2(2), p. 153-167.

Shi, R. *et al.* (2017) 'CDCA2 promotes lung adenocarcinoma cell proliferation and predicts poor survival in lung adenocarcinoma patients', *Oncotarget*. Impact Journals, LLC, 8(12), p. 19768-19799.

Sidney, L. E. *et al.* (2014) 'Concise review: evidence for CD34 as a common marker for diverse progenitors', *Stem Cells*. Wiley Online Library, 32(6), pp. 1380–1389.

Simões, P. D. and Ramos, T. (2007) 'Human pluripotent embryonal carcinoma NTERA2 cl. D1 cells maintain their typical morphology in an angiomyogenic medium', *Journal of Negative Results in Biomedicine*. BioMed Central, 6(1), p. 5-14.

Sinkkonen, L. *et al.* (2008) 'MicroRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells', *Nature Structural & Molecular Biology*. Nature Publishing Group, 15(3), pp. 259–267.

Skakkebæk, N. (1972) 'Possible carcinoma-in-situ of the testis', *The Lancet*. Elsevier, 300(7776), pp. 516–517.

Skakkebaek, N., Meyts, E. R. and Main, K. M. (2001) 'Testicular dysgenesis syndrome: an increasingly common developmental disorder with environmental aspects', *Apmis*. Wiley Online Library, 109(S103).

Skotheim, R. I. *et al.* (2005) 'Differentiation of human embryonal carcinomas in vitro and in vivo reveals expression profiles relevant to normal

development', *Cancer Research*. AACR, 65(13), pp. 5588–5598.

Sonne, S. B. *et al.* (2009) 'Analysis of gene expression profiles of microdissected cell populations indicates that testicular carcinoma in situ is an arrested gonocyte', *Cancer Research*. AACR, 69(12), pp. 5241–5250.

de Sousa Lopes, S. M. C. *et al.* (2004) 'BMP signaling mediated by ALK2 in the visceral endoderm is necessary for the generation of primordial germ cells in the mouse embryo', *Genes & Development*. Cold Spring Harbor Lab, 18(15), pp. 1838–1849.

Sperger, J. M. *et al.* (2003) 'Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors', *Proceedings of the National Academy of Sciences*. National Acad Sciences, 100(23), pp. 13350–13355.

Stead, E. *et al.* (2002) 'Pluripotent cell division cycles are driven by ectopic Cdk2, cyclin A/E and E2F activities', *Oncogene*. Nature Publishing Group, 21(54), p. 8320-8333.

Stoop, H. *et al.* (2011) 'Diagnosis of testicular carcinoma in situ "(intratubular and microinvasive)"seminoma and embryonal carcinoma using direct enzymatic alkaline phosphatase reactivity on frozen histological sections', *Histopathology*. Wiley Online Library, 58(3), pp. 440–446.

Storgaard, L., Bonde, J. P. and Olsen, J. (2006) 'Male reproductive disorders in humans and prenatal indicators of estrogen exposure: a review of published epidemiological studies', *Reproductive Toxicology*. Elsevier, 21(1), pp. 4–15.

Sturgeon, C. M. *et al.* (2008) 'National Academy of Clinical Biochemistry

laboratory medicine practice guidelines for use of tumor markers in testicular, prostate, colorectal, breast, and ovarian cancers', *Clinical Chemistry*. Clinical Chemistry, 54(12), pp. e11–e79.

Subramanyam, D. *et al.* (2011) 'Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells', *Nature Biotechnology*. Nature Publishing Group, 29(5), pp. 443–448.

Suh, N. *et al.* (2010) 'MicroRNA function is globally suppressed in mouse oocytes and early embryos', *Current Biology*, 20(3), pp. 271–277.

Sun, G. *et al.* (2017) 'Her-2 expression regulated by downregulation of miR-9 and which affects chemotherapeutic effect in breast cancer', *Cancer Gene Therapy*. Nature Publishing Group, 24(5), pp. 194–202.

Surani, M. A. (2015) 'Human germline: a new research frontier', *Stem Cell Reports*. Elsevier, 4(6), pp. 955–960.

Syring, I. *et al.* (2015) 'Circulating serum miRNA (miR-367-3p, miR-371a-3p, miR-372-3p and miR-373-3p) as biomarkers in patients with testicular germ cell cancer', *The Journal of Urology*. Elsevier, 193(1), pp. 331–337.

Szklarczyk, D. *et al.* (2014) 'STRING v10: protein–protein interaction networks, integrated over the tree of life', *Nucleic Acids Research*. Oxford University Press, 43(D1), pp. D447–D452.

Takada, S. *et al.* (2009) 'Potential role of miR-29b in modulation of Dnmt3a and Dnmt3b expression in primordial germ cells of female mouse embryos', *RNA*, 15(8), pp. 1507–1514.

Tandstad, T. *et al.* (2014) 'Management of clinical stage I seminomatous

testicular cancer: A report from SWENOTECA.' American Society of Clinical Oncology. (May 2014). pp4508-4508.

Tang, H. *et al.* (2013) 'miR-9 functions as a tumor suppressor in ovarian serous carcinoma by targeting TLN1', *International Journal of Molecular Medicine*. Spandidos Publications, 32(2), pp. 381–388.

Tang, W. W. C. *et al.* (2015) 'A unique gene regulatory network resets the human germline epigenome for development', *Cell*. Elsevier, 161(6), pp. 1453–1467.

Tang, W. W. C. *et al.* (2016) 'Specification and epigenetic programming of the human germ line', *Nature Reviews Genetics*. Nature Research. 17(10), p.585-600.

Tay, Y. *et al.* (2008) 'MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation', *Nature*, 455(7216), pp. 1124–1128.

Terentiev, A. A. and Moldogazieva, N. T. (2013) 'Alpha-fetoprotein: a renaissance', *Tumor Biology*. Springer, 34(4), pp. 2075–2091.

Theunissen, T. W. *et al.* (2011) 'Nanog overcomes reprogramming barriers and induces pluripotency in minimal conditions', *Current Biology*. Elsevier, 21(1), pp. 65–71.

Thomas-Chollier, M. *et al.* (2012) 'RSAT peak-motifs: motif analysis in full-size ChIP-seq datasets', *Nucleic Acids Research*. Oxford Univ Press, 40(4), pp. e31–e31.

Thomson, J. M. *et al.* (2004) 'A custom microarray platform for analysis of

microRNA gene expression', *Nature Methods*. Nature Publishing Group, 1(1), pp. 47–53.

Toder, V. *et al.* (1983) 'The effect of alpha-fetoprotein on the growth of placental cells in vitro', *Placenta*. Elsevier, 4(1), pp. 79–85.

La Torre, A., Georgi, S. and Reh, T. A. (2013) 'Conserved microRNA pathway regulates developmental timing of retinal neurogenesis', *Proceedings of the National Academy of Sciences*. National Academy Sciences, 110(26), pp. E2362–E2370.

Torres, J. and Watt, F. M. (2008) 'Nanog maintains pluripotency of mouse embryonic stem cells by inhibiting NF [kappa] B and cooperating with Stat3', *Nature Cell Biology*. Nature Publishing Group, 10(2), p. 194-201.

Trigo, J. M. *et al.* (2000) 'Tumor markers at the time of recurrence in patients with germ cell tumors', *Cancer*. Wiley Online Library, 88(1), pp. 162–168.

Tsankov, A. M. *et al.* (2015) 'Transcription factor binding dynamics during human ES cell differentiation', *Nature*. Nature Research, 518(7539), pp. 344–349.

Vandesompele, J. *et al.* (2002) 'Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes', *Genome Biology*. BioMed Central, 3(7), p. 95-98.

Verhaak, R. G. W. *et al.* (2010) 'Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1', *Cancer Cell*. Elsevier, 17(1), pp. 98–110.

Vescovi, A. L., Galli, R. and Reynolds, B. A. (2006) 'Brain tumour stem cells',

Nature Reviews: Cancer. Nature Publishing Group, 6(6), p. 425-436.

Voorhoeve, P. M. *et al.* (2006) 'A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors', *Cell*. Elsevier, 124(6), pp. 1169–1181.

Wadman, R. I. *et al.* (2016) 'A Comparative Study of SMN Protein and mRNA in Blood and Fibroblasts in Patients with Spinal Muscular Atrophy and Healthy Controls', *PloS One*. Public Library of Science, 11(11), p. e0167087.

Wahl, M. C., Will, C. L. and Lührmann, R. (2009) 'The spliceosome: design principles of a dynamic RNP machine', *Cell*. Elsevier, 136(4), pp. 701–718.

Wallacides, A. *et al.* (2012) 'Estrogens promote proliferation of the seminoma-like TCam-2 cell line through a GPER-dependent ERα36 induction', *Molecular and Cellular Endocrinology*. Elsevier, 350(1), pp. 61–71.

Walsh, T. J. *et al.* (2007) 'Prepubertal orchiopexy for cryptorchidism may be associated with lower risk of testicular cancer', *The Journal of Urology*. Elsevier, 178(4), pp. 1440–1446.

Wan, C. *et al.* (2015) 'Panorama of ancient metazoan macromolecular complexes', *Nature*. Nature Research, 525(7569), pp. 339–344.

Wang, L.-T. *et al.* (2014) 'Survival of motor neuron protein downregulates miR-9 expression in patients with spinal muscular atrophy', *The Kaohsiung Journal of Medical Sciences*. Elsevier, 30(5), pp. 229–234.

Wang, Y. *et al.* (2007) 'DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal', *Nature genetics*. Nature Publishing Group, 39(3), pp. 380–385.

Wang, Y. *et al.* (2008) ‘Embryonic stem cell–specific microRNAs regulate the G1-S transition and promote rapid proliferation’, *Nature Genetics*. Nature Publishing Group, 40(12), pp. 1478–1483.

Wang, Y. and Blelloch, R. (2011) ‘Cell cycle regulation by microRNAs in stem cells’, in *Cell Cycle in Development*. Springer, pp. 459–472.

Weber, S. *et al.* (2010) ‘Critical function of AP-2gamma/TCFAP2C in mouse embryonic germ cell maintenance’, *Biology of Reproduction*. Oxford University Press, 82(1), pp. 214–223.

Weinstein, J. N. *et al.* (2013) ‘The cancer genome atlas pan-cancer analysis project’, *Nature Genetics*. Nature Research, 45(10), pp. 1113–1120.

West, J. A. *et al.* (2009) ‘A role for Lin28 in primordial germ-cell development and germ-cell malignancy’, *Nature*. Nature Publishing Group, 460(7257), pp. 909–913.

Wong, N. and Wang, X. (2014) ‘miRDB: an online resource for microRNA target prediction and functional annotations’, *Nucleic Acids Research*. Oxford University Press, 43(D1), pp. D146–D152.

Xia, F., Lee, C. W. and Altieri, D. C. (2008) ‘Tumor cell dependence on Ran-GTP–directed Mitosis’, *Cancer Research*. AACR, 68(6), pp. 1826–1833.

Xu, N. *et al.* (2009) ‘MicroRNA-145 regulates OCT4, SOX2, and KLF4 and represses pluripotency in human embryonic stem cells’, *Cell*. Elsevier, 137(4), pp. 647–658.

Yamaguchi, S. *et al.* (2005) ‘Nanog expression in mouse germ cell development’, *Gene Expression Patterns*. Elsevier, 5(5), pp. 639–646.

Yamaguchi, S. *et al.* (2009) ‘Conditional knockdown of Nanog induces apoptotic cell death in mouse migrating primordial germ cells’, *Development*. The Company of Biologists Limited, 136(23), pp. 4011–4020.

Yamaji, M. *et al.* (2013) ‘PRDM14 ensures naive pluripotency through dual regulation of signaling and epigenetic pathways in mouse embryonic stem cells’, *Cell Stem Cell*. Elsevier, 12(3), pp. 368–382.

Yao, P.-L., Lin, Y.-C. and Richburg, J. H. (2012) ‘Mono-(2-ethylhexyl) phthalate (MEHP) promotes invasion and migration of human testicular embryonal carcinoma cells’, *Biology of Reproduction*. Oxford University Press, 86(5). pp 1-10.

Ying, Y. and Zhao, G.-Q. (2001) ‘Cooperation of endoderm-derived BMP2 and extraembryonic ectoderm-derived BMP4 in primordial germ cell generation in the mouse’, *Developmental Biology*. Elsevier, 232(2), pp. 484–492.

Yoo, A. S. *et al.* (2011) ‘MicroRNA-mediated conversion of human fibroblasts to neurons’, *Nature*. Nature Publishing Group, 476(7359), pp. 228–231.

Yu, J. *et al.* (2007) ‘Induced pluripotent stem cell lines derived from human somatic cells’, *Science*. American Association for the Advancement of Science, 318(5858), pp. 1917–1920.

Yu, T. *et al.* (2014) ‘MicroRNA-9 inhibits the proliferation of oral squamous cell carcinoma cells by suppressing expression of CXCR4 via the Wnt/β-catenin signaling pathway’, *Oncogene*. Nature Publishing Group, 33(42), pp. 5017–5027.

Zhang, B., Kirov, S. and Snoddy, J. (2005) 'WebGestalt: an integrated system for exploring gene sets in various biological contexts', *Nucleic Acids Research*, 33(suppl 2), pp. W741–W748.

Zhang, X. *et al.* (2008) 'Esrrb activates Oct4 transcription and sustains self-renewal and pluripotency in embryonic stem cells', *Journal of Biological Chemistry*. ASBMB, 283(51), pp. 35825–35833.

Zhang, X. *et al.* (2009) 'A role for NANOG in G1 to S transition in human embryonic stem cells through direct binding of CDK6 and CDC25A', *The Journal of Cell Biology*. Rockefeller University Press, 184(1), pp. 67–82.

Zhao, C. *et al.* (2009) 'A feedback regulatory loop involving microRNA-9 and nuclear receptor TLX in neural stem cell fate determination', *Nature Structural & Molecular Biology*. Nature Publishing Group, 16(4), pp. 365–371.

Zheng, G. X. Y. *et al.* (2011) 'A latent pro-survival function for the mir-290-295 cluster in mouse embryonic stem cells', *PLoS Genetics*. Public Library of Science, 7(5), p. e1002054.

Znaor, A. *et al.* (2014) 'International variations and trends in testicular cancer incidence and mortality', *European Urology*. Elsevier, 65(6), pp. 1095–1106.

Rights and permission to use Figures

License Number	4185891487732
License date	Sep 11, 2017
Licensed Content Publisher	Elsevier
Licensed Content Publication	European Urology
Licensed Content Title	Global Trends in Testicular Cancer Incidence and Mortality
Licensed Content Author	Alexandre Rosen,Gautam Jayram,Michael Drazer,Scott E. Eggener
Licensed Content Date	Aug 1, 2011
Licensed Content Volume	60
Licensed Content Issue	2
Licensed Content Pages	6
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
Title of your thesis/dissertation	Defining the role of miRNAs regulated by the pluripotency gene NANOG in the development and progression of Testicular Germ Cell Tumours.
Expected completion date	Sep 2017
Estimated size (number of pages)	170
Requestor Location	1 cary rd London, Leytonstone e113lg United Kingdom Attn: University of Nottingham
Publisher Tax ID	GB 494 6272 12
Total	0.00 USD

License Number	4185900563147
License date	Sep 11, 2017
Licensed Content Publisher	Nature Publishing Group
Licensed Content Publication	Oncogene
Licensed Content Title	Histone onco-modifications
Licensed Content Author	J Fullgrabe, E Kavanagh and B Joseph
Licensed Content Date	Apr 25, 2011
Licensed Content Volume	30
Licensed Content Issue	31
Type of Use	reuse in a dissertation / thesis
Requestor type	academic/educational
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Figures	figure 1
Author of this NPG article	no
Your reference number	
Title of your thesis / dissertation	Defining the role of miRNAs regulated by the pluripotency gene NANOG in the development and progression of Testicular Germ Cell Tumours.
Expected completion date	Sep 2017
Estimated size (number of pages)	170
Requestor Location	1 cary rd London, Leytonstone e113lg United Kingdom Attn: University of Nottingham
Billing Type	Invoice
Billing address	University of Nottingham , United Kingdom Attn:
Total	0.00 USD

Professional Internships for PhD Students

Reflection Form

Name of Organization

NHiS

Details of Placement
Please describe your main activities during the placement (150-200 words)

NHiS is a healthcare informatics company who are dedicated to using existing Hospital Episode Statistics to improve patient flows and outcomes for the benefit of the NHS. The majority of my time was spent on a computer using Microsoft excel to navigate my way through pivot tables. These tables were used to interrogate ICD10 codes for codes, costs and patient flows in order to create figures that the company can use. I worked on two main projects. The first was in partnership with Nigel Mongan at the University of Nottingham to interrogate two separate diseases in attempt to elucidate causation between the two. The second project lasted the largest amount of time which was investigating the cost of COPD for inpatient and outpatient patients for the past 10 years and to make tables and graphs in a presentable for the data produced. In addition to this, a lot of training in data protection was also included.

Placement Achievements
Please detail all outcomes from the placement, including any publications, presentations given and reports written etc. (150-200 words)

This was an interesting experience as NHiS is a commercial company which focuses on making profits, therefore giving indication of what working in industry may resemble. During this placement two large documents were created. The work performed was in collaboration with the University of Nottingham in order to create a scientific document was produced to send my supervisor. This data is to be part of a larger study which is to be published in a scientific journal. The second was to investigate COPD across different CCGs in the united kingdom, working out costs and patient flows in order to make graphs and figures. In addition to this a literature review mainly referencing governmental documents was written to be used in combination with the graphs and tables to produce a document for COPD. This work is to be published by NHiS and will be available to the public via their website.

Skill development
Has this Placement helped you developed any new skills or enhanced your previous skill set? (100-150 words)

During this placement I used excel sheets the majority of my time. Therefore I learned how to navigate around pivot tables and create tables that look professional. In addition to this, I did not know much about how the healthcare system had worked, for example, how the districts are separated into CCGs and disease / symptoms into ICD10 codes. This was particularly interesting in informing myself how much data is collected by the healthcare system and how, possibly, in the future as part of other project it could be used to better inform us about disease and treatments. In addition to this, this placement has showed me how a corporate working environment operates and how different it is to working in academia.

Future Work
Has this Placement influenced your future career aspirations? If so, in what way? (150-200 words)

As mentioned it has shown me how disease and treatment are monitored by the current national health service, so this is something to keep in mind as NHiS may be a potential collaborator for using their data in the future. In addition to this, it has shown me that although a company which is, in part, interested in making the health service better NHiS It is still money focused and this was more important than doing something interesting or creative with the data. This was an eye opening experience and has indicated that I am more interested in continuing to work within academia and if I end up working for a corporation I would be more interested in a company which was more interested in understanding things or helping people. Also, working at the NHiS I was part of a multidisciplinary team which consisted of people from HR, programmers and admin staff. This placement has enhanced my ability to be part of a multidisciplinary team in order to work on a common goal.