# IDENTIFICATION OF OOCYTE REPROGRAMMING FACTORS AND MECHANISMS

**CHOULIA MOLA, MRes, BSc**

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

**SEPTEMBER 2016**




UNITED KINGDOM · CHINA · MALAYSIA

## Acknowledgments

The work that needs to be carried out in the context of a PhD burdens the shoulders of one person: the PhD candidate. However, in order for a PhD to be successfully completed many more people need to frame and provide a direction in the effort invested by the candidate. The same, of course applied in my case and it is now time to thank and acknowledge the people that assisted, directed, helped and supported me during the ascending and descending phases of my PhD. I will make the best effort so that I do not forget anyone.

Firstly, I want to thank Dr Ramiro Alberio, my supervisor for giving me the opportunity to do my PhD in his lab. Dr Alberio has been a very good supervisor and I consider myself fortunate to have worked under his supervision. At this point I also want to thank Dr Andrew Johnson, my second supervisor, for offering me the chance to work on the *axolotl* and supplying all the relevant material.

Dr Susan Liddell deserves a special mention, as because of her experience in protein research enormously helped in realising the experiments presented in the thesis. Dr Kenneth Davis, the technician of the proteomics lab, was also very helpful and a great joy to work with. I am deeply grateful to the proteomics lab for all the assistance and inspiration it provided over the course of my PhD.

I am grateful to Dr Cinzia Allegrucci for showing me how to use vet school equipment and also sharing methods and protocols with me. Kamila Derecka and Doris Klisch, the molecular lab technicians that I mostly became involved with during the final stages of my PhD, were also very helpful.

I am also deeply grateful to Dr David Boocock and Clare Coveney from the Clinical Proteomics group of Nottingham Trent University. Their help and collaboration was paramount for the realisation of the mass spectrometry experiment presented in my last results' chapter. Dr Matt Loose and Dr Terry Evans also helped in this context, as they prepared and provided me with the

*axolotl* transcriptome. I also need to thank the company Dundee Cell, as I had to resort to them for protein identification needs.

At this point I should not forget the fellow PhD students within the South Lab and QMC that surrounded my PhD and lab life. Sidar Bereketoglu and Darren Crowley from QMC were of particular assistance as they often provided me with a cell line or a reagent that was needed, or showed me a method and shared a protocol. I also want to mention Haixin Zhang who was also supervised by Dr Alberio. It was certainly a pleasure to meet and work with her. Additionally, I want to deeply thank Mattia Piana and Priscila Ramos who helped me print my thesis before my submission.

Last but not least I want to thank my family and parents specifically, Vedat and Bachriye, who have supported and believed in me throughout my life. Alican, my partner, is the person that probably was mostly inflicted with the side effects of the not so pleasant PhD phases. I thank him for his patience and support.

Last but not least I want to thank the University of Nottingham for providing me with the Vice-Chancellor Scholarship for my PhD.

I will end the acknowledgements section thanking everyone mentioned above and hoping that I have not forgotten anyone.

Choulia Mola

## Abstract

The ability of the oocyte to reprogram paternal chromatin is widely accepted. The oocyte cytoplasm has been reported to contain activities that alter the histone composition, demethylate and unwind highly methylated paternal DNA upon fertilization and lead to zygote formation. The same events are mimicked during the course of cell reprogramming, where somatic cells exposed to the oocyte environment acquire a pluripotent character.

Amphibian oocytes have been widely used to study the reprogramming attributes of the female germ cells for over half a century. The size of amphibian oocytes renders them good candidates for technical manipulations, as well as protein biochemistry studies. Oocytes harvested from the amphibian *axolotl* were utilised in the present study to delineate the oocyte reprogramming machinery. Previous studies have reported the presence of key pluripotency factor orthologs, such as Nanog, in the *axolotl* oocyte, thus rendering it an interesting tool for the study of the chromatin remodelling and pluripotency networks as well as their interplay. To study these networks, identification of the physical interactors of the Nanog protein or the Nanog promoter was attempted.

One of the key events taking place during cell and paternal chromatin reprogramming is demethylation. The removal of the methyl mark that is deposited over genetic loci during cell differentiation is linked with the transcriptional activation of the respective loci. Demethylation is therefore fundamental for the activation of pluripotency-associated factors both in fertilisation and cell reprogramming.

The leading theory behind demethylation is the successive oxidation of the methyl mark to formyl and carboxyl and its subsequent removal from Thymine DNA Glycosylase (TDG). Ten Eleven Translocation (TET) family proteins catalyse methyl oxidation and therefore utilise demethylation. Additional demethylation pathways have also been supported by findings, specifically in the context of the oocyte. Base Excision Repair (BER) and

Nucleotide Excision Repair (NER) factors have been associated with DNA demethylation occurring immediately after fertilisation. It is therefore fundamental to delineate the demethylation mechanism facilitated by the oocyte in the early stages of development. The identification of oocyte demethylation factors and mechanisms will greatly improve our understanding of demethylation as well as improve current cell reprogramming methodologies by using the more efficient oocyte demethylation mechanism.

To delineate the demethylation mechanism employed by the oocyte and contribute to the debate of TET oxidation versus NER or BER demethylation via DNA repair in the context of the oocyte, oocyte factors preferentially binding to different cytosine modifications were analysed. The murine Nanog promoter, a genetic locus activated during cell reprogramming, harboured the different cytosine modifications in an attempt to see how each one would affect its transcriptional status. As a result, TET factors were not discovered to bind any of the cytosine modifications while both NER and BER pathway proteins were found to be significantly enriched in the case of methylcytosine. The results obtained therefore support previous findings and advocate for a TET independent DNA demethylation occurring through DNA repair. While teams researching demethylation through DNA repair have supported either NER or BER, the findings unveiled in this thesis advocate for the two pathways acting in concert in the recognition and potential removal of DNA methylation. It also has to be noted that it is the first time an experiment of this kind is carried out in the context of the oocyte and the findings offer a unique insight into oocyte demethylation dynamics. Taking into account the data obtained and previous research, a new DNA demethylation model is proposed. The model advocates for NER taking the lead in DNA demethylation, creating patchy demethylated sites that are subsequently recognised and demethylated via the BER pathway.

CHAPTER 1 8

**Literature review** **8**

**1.1 Pluripotency, its regulatory network and cell reprogramming** **8**
1.1.1 Pluripotency in early development 8
1.1.2 Arrested pluripotency in stem cells 10
1.1.3 The pluripotency network 13
1.1.4 The regulatory network of pluripotency 15
1.1.5 Methods of cell reprogramming towards pluripotency 17
1.1.6 Comparing SCNT and iPS 20

**1.2 Reprogramming with the oocyte** **22**
1.2.1 The oocyte in fertilization and reprogramming 22
1.2.2 Egg and Oocyte: the advantage of immaturity 24
1.2.3 Egg and Oocyte lysate 26
1.2.4 Reprogrammome: the reprogramming fraction of the oocyte 26

**1.3 DNA methylation and demethylation** **30**
1.3.1 Methylated cytosine (5mC), the fifth base and transcriptional repressor 30
1.3.2 DNA methylation reversal: Demethylation 32
1.3.3 DNA demethylation mechanisms 35
1.3.4 The DNA repair pathways BER and NER and their implication in demethylation 37

**1.4 Axolotl as an experimental organism** **42**
1.4.1 Differences between Xenopus and *axolotl*: Preformation and Epigenesis 42
1.4.2 The pluripotency network in the *axolotl* oocyte 44
1.4.3 Aims and objectives 46

CHAPTER 2: MATERIALS AND METHODS 49
2.1 Oocyte extract preparation 49
2.2 Bradford Assay for protein quantification 50
2.3 SDS-PAGE electrophoresis and gel staining 50
2.4 Western Blotting 51
2.5 Protein Immunoprecipitation (IP) 52
2.6 Antibody purification 53
2.7 In vitro translation of mRNA 54
2.8 Recombinant protein overexpression and purification 54
2.9 Immunocytochemistry 57
2.10 End-Point PCR mNanog promoter amplification 57
2.11 Agarose gels and PCR product purification 60
2.11 DNA pull down 61
2.12 Mass spectrometry 62

CHAPTER 3: PROTEIN NETWORK STUDY AND MATERIAL IMPROVEMENT 65

**3.1 Introduction** **65**
3.1.1 Immunoprecipitation in the field of pluripotency and reprogramming 65
3.1.2 Affinity purification: a tool to delineate protein networks and interactions 66

**3.2 Aims and objectives** **67**

**3.3 Materials and methods** **68**

**3.4 Results** **69**
3.4.1 Testing the affinity of the antibody against native and recombinant protein 69
3.4.2 Assessing the efficiency of immunoprecipitation 71
3.4.3 Mass spectrometry analysis of recombinant *axolotl* Nanog 74
3.4.4 Purification of the AxNanog antibody from GroEL immunoglobulins 76

**3.5 Conclusion** **87**

CHAPTER 4: THE REGULATORY NETWORK OF THE NANOG PROMOTER 92

**4.1 Introduction** **92**
4.1.1 Protein-DNA interactions in ESC: Deciphering regulatory pathways 92

**4.2 Aims and objective** **94**

**4.3 Materials and methods** **95**

**4.4 Results** **96**
4.4.1 Murine Nanog promoter and Transcription Factor binding analysis 96
4.4.2 Material preparation and optimization 102
4.4.3 Initial attempt of DNA pull down in *axolotl* oocyte lysate 105
4.4.4 DNA pull down with mESC and mOct4 detection 107
4.4.5 Determination of the protein band identities through mass spectrometry 111
4.4.5 Comparison between *axolotl* embryo and oocyte specific mNanog promoter binders 114

**4.5 Conclusions** **119**

CHAPTER 5: IDENTIFYING THE OOCYTE SPECIFIC DEMETHYLATION MECHANISM AND FACTORS 122

**5.1 Introduction** **122**
5.1.1 5hmC, 5fC and 5caC: 5mC oxidation products and epigenetic markers 122
5.1.2 TET enzymes and their activity 122
5.1.3 Readers and epigenetic significance of the cytosine modifications 124
5.1.4 Demethylation in the oocyte 127

**5.2 Aims and objective** **129**

**5.3 Materials and methods** **130**

**5.4 Results** **131**
5.4.1 Experimental workflow 131
5.4.2 mNanog promoter amplification 132
5.4.3 Optimal protein digestion for mass spectrometry 134
5.4.4 Non-modified C/5mC/5hmC/5fC interactome screen 136
5.4.5 5mC binders: enriched for DNA repair factors, core histones and UHRF1 143

5.4.6 5hmC binders: chromatin remodelers and DNA repair enrichment 146
5.4.7 5fC binders: 5mC resemblance and the methionine cycle 148
5.4.8 Non-modified cytosine binders: low number of binders and absence of PARP 151
5.4.9 BER or NER: Fold enrichment of NER and BER associated factors in C/5mC/5hmC/5fC 153

**5.5 Conclusion** **154**

CHAPTER 6: DISCUSSION 161

**6.1 Protein immunoprecipitation and antibody optimization** **161**

**6.2 Optimization and application of DNA pull down in axolotl oocyte lysates** **162**

**6.3 DNA repair and oocyte demethylation** **163**

APPENDIX 174

Heatmap and identified proteins in the 5mC/5hmC/5fC and non-modified cytosine screen 174

Heat map result statistics and data quality 177

Summary Tables 180

REFERENCES 182

# Chapter 1

## *Literature review*

## *1.1 Pluripotency, its regulatory network and cell reprogramming*

### 1.1.1 Pluripotency in early development

Development could be described as the progressive loss of embryonic totipotency/pluripotency (Nichols and Smith, 2012). Mammalian embryos begin as a cluster of cells (blastomeres) capable of generating all the cell lineages required for the formation of embryonic and extra-embryonic tissue; the cells that are endowed with this capacity, result from the first cell divisions of the zygote and are totipotent. In murine embryos, blastomeres begin to compact at the eight-cell stage while cell divisions continue (Nichols and Smith, 2012). Along with the compaction the cells reorganize and reposition on the inside and outside of the embryo to form the Inner Cell Mass (ICM) and trophoblast. Commitment to the trophoblast is mediated by the transcription factor Tead4 during the cleavage state (Nishioka et al., 2009). Tead4 is activated when the Hippo pathway is inactive and is not expressed in the ICM where the Hippo pathway is active (Wada et al., 2011). Tead4 upregulates the transcription factors Cdx2 and eomesodermin, which in turn create a positive feedback loop with Elf5 and drive the specification of the trophoblast lineage (Ng et al., 2008).

Along with the specification of the trophoblast, a set of transcription factors defines the ICM lineage. As indicated above the ICM does not express Tead4 or any of the other trophoblast transcription factors, but expresses Oct4 (Nichols et al., 1998). The POU-domain transcription factor Oct4 is

fundamental for the establishment of the ICM identity; upon its absence inner cells do form but lack the ICM character and finally differentiate to trophoblast (Nichols et al., 1998).

The trophoblast transfers fluid into the blastocyst, forming the blastocoel cavity. It is after the formation of the blastocoel that the ICM segregates into the hypoblast (primitive endoderm) and epiblast (Nichols and Smith, 2012). Cells expressing PDGFRa and Sox17 segregate from cells expressing epiblast markers and if positioned alongside the blastocoel upregulate Gata4 and Sox7 (Artus et al., 2011). The expression of the latter two factors marks the commitment of the cells to the hypoblast lineage (Artus et al., 2011). The epiblast lineage is marked by the expression of Nanog and also relies on it (Dietrich and Hiiragi, 2007, Plusa et al., 2008). Prior to the segregation of ICM to hypoblast and epiblast, Nanog is found in differing levels across the ICM (Dietrich and Hiiragi, 2007, Plusa et al., 2008). However, upon segregation Nanog-positive cells are restricted to the epiblast, all expressing similar levels of Nanog (Plusa et al., 2008, Dietrich and Hiiragi, 2007). In murine female embryos the silenced X chromosome is activated in the epiblast, whereas it remains inactive in Nanog-null embryos which fail to develop epiblast (Mitsui et al., 2003, Silva et al., 2009). In Nanog-null embryos the hypoblast also fails to develop indicating that hypoblast development depends on signals from the epiblast (Messerschmidt and Kemler, 2010).

The attribute of totipotency is granted to the cells descending from the first divisions of the zygote, prior to the beginning of cell segregation to the ICM and trophoblast. Any cell that is equally capable of differentiating to either trophoblast or ICM is totipotent. Totipotency gives way to pluripotency, which is restricted to the epiblast and is defined as the capacity to generate all three germ layers that will form the embryo proper: the mesoderm, endoderm and ectoderm.

The epiblast state exists only for a short period. In mice it forms shortly before embryo implantation and differentiates within 72 hours of implantation (Kalkan and Smith, 2014). Following implantation, the epiblast undergoes a morphological change and forms a columnar epithelium (Kalkan and Smith,

2014). The change from pre to post-implantation epiblast is marked by the suppression of the naïve epiblast markers and upregulation of Fgf5 (Boroviak et al., 2014, Pelton et al., 2002). The generic epiblast markers Oct4 and Sox2 continue their uniform expression until gastrulation commences (Iwafuchi-Doi et al., 2012).

Upon its differentiation the epiblast will give rise to the three germ layers: ectoderm, mesoderm and endoderm and pluripotency will give its place to cell differentiation.

### 1.1.2 Arrested pluripotency in stem cells

Embryonic Stem Cells (ESC) were first derived in 1981 with the explantation and culture of blastocysts or ICM over a layer of mitotically inactive fibroblast cells (Evans, 1981, Martin, 1981). The culture conditions of the first ESC colonies were based on the prior experience of pluripotent Embryonal Carcinoma (EC) cultures (Martin and Evans, 1975). ESCs contribute to all tissues when injected in blastocysts, including the germ line (Bradley et al., 1984). ESC cultures retain their capacity to contribute to chimaeras even after prolonged culture and do not form tumors, demonstrating their responsiveness to the signals of neighboring cells and tissue in chimaeras (Nichols and Smith, 2012). Successful ESC cultures have been established by the isolation of cells from the epiblast (Brook and Gardner, 1997), from morulae and eight-cell stage blastomeres (Eistetter et al., 1993, Tesar, 2005).

ESC cultures maintain their pluripotency by preventing differentiation and promoting proliferation. Leukemia Inhibitory Factor (LIF) is a growth factor fundamental for preserving ESC in an undifferentiated state (Suda et al., 1987). ESC can be induced to differentiate by the expression of certain transcription factors. Gata6 expression will force ESC to a primitive endoderm cell fate (Fujikura et al., 2002), expression of Cdx2 on the other hand will push ESC towards a trophoectodermal fate (Niwa et al., 2005).

The first stem cell colonies established were cultured on a layer of feeder cells. It was later found that the feeder cells could be replaced by the addition of LIF in the stem cell medium (Smith and Hooper, 1983, Smith and Hooper, 1987, Koopman and Cotton, 1984, Smith et al., 1988, Williams et al., 1988). It has been reported that LIF acts by activating Stat3, a transcription factor (Niwa et al., 1998, Matsuda et al., 1999). LIF is also implicated in the MAP/ERK pathway, which promotes differentiation (Burdon et al., 1999). Fgf4, which also activates the ERK pathway, provokes differentiation (Kunath et al., 2007, Stavridis et al., 2007), however, interestingly inhibition of the ERK pathway alone is not adequate for the maintenance of ESC cultures at an undifferentiated state, unless ESC are cultured at high densities or LIF is supplemented (Ying et al., 2008). Pathway inhibition methods have been widely researched as a mean of understanding and improving ESC culture conditions. The 3i method composes of three inhibitors: an FGF receptor inhibitor, a MAPK ½ inhibitor and a GSK3 inhibitor and has been very efficient in the maintenance of ESC cultures propagating at an undifferentiated state without the need for LIF supplementation (Kiyonari et al., 2010, Yamagata et al., 2010). Because the FGF receptor functions through the ERK pathway, a selective MEK inhibitor can reduce 3i to 2i (Wray et al., 2010).

As a general rule, ESCs derive from ICM or pre-implantation epiblasts; however pluripotent cell colonies were derived from post-implantation epiblasts and survived under different culture conditions (Brons et al., 2007, Tesar et al., 2007). The pre-implantation epiblast represents the ground state of pluripotency because it is capable of generating the entire fetus in the course of development and when a single epiblast cell is transplanted into a different blastocyst it can contribute to all cell lineages (Nichols and Smith, 2009). Nichols and Smith have also used the term naïve to describe the pluripotency of the pre-implantation epiblast (Nichols and Smith, 2009). The ESCs reflect the ground/naïve state of developmental pluripotency in vitro (Nichols and Smith, 2009). Post-implantation epiblast although retaining its pluripotent character, as it is still capable of forming Primordial Germ Cells (PGC) upon induction (Hayashi et al., 2011), it has entered a less potent

pluripotent state. Therefore post-implantation epiblast pluripotency is described as primed (Nichols and Smith, 2009). The pluripotent cells derived from primed epiblast therefore reflect its native attributes. EpiSC (post-implantation epiblast derived stem cells) cannot contribute to chimaeras when injected to blastocysts but can form teratocarcinomas (Brons et al., 2007, Tesar et al., 2007). Upon culture, EpiSCs do not respond to classic ESC culture conditions but depend on FGF2 and activin supplementation (Tesar et al., 2007, Brons et al., 2007). It has to be noted however, that these two different pluripotency stages have been reported in murine development. The pluripotent cell lines that have been isolated from human pre-implantation epiblasts (Thomson et al., 1998) are significantly different from the archetype of murine ESCs and resemble murine EpiSC (Brons et al., 2007, Tesar et al., 2007, Rossant, 2008). Naïve pluripotency might reflect a developmental stage confined to rodent species, since ESCs have only been generated from mouse and rat (Buehr et al., 2008, Li et al., 2008). During the development of most mammalian embryos the epiblast forms a flat layer known as “embryonic disk” in contrast to mouse and rat where the epiblast assumes a cup-shaped form, the “egg cylinder” (Nichols and Smith, 2009). The formation of the egg cylinder is accompanied by directed apoptosis of cell located in the inside of the cell ball that the epiblast is prior to that point (Coucouvanis and Martin, 1999). According to a hypothesis proposed by Nichols and Smith the apoptosis window might be a barrier difficult to surpass for rodent epiblast in vitro, thus extending the time an epiblast can remain at the ground state of pluripotency (Nichols and Smith, 2009). Capturing naïve pluripotency is therefore easier in the rodent species than in species that do not have to undergo cell apoptosis to form an egg cylinder (Nichols and Smith, 2009). The above hypothesis accepts the existence of the naïve pluripotent state as an axiom. A second possible explanation for the inability to capture ground state pluripotency is mammals would be diapause, which is a state of arrested embryonic development in rodents (Nichols and Smith, 2009). The diapause status would again serve as an opportunity to capture and propagate naïve pluripotent cells

in rodents but not in other mammals that do not have diapause in their embryo development (Nichols and Smith, 2009).

The ability of stem cells to renew and maintain their pluripotent character in vitro is a remarkable attribute. Genetic and protein studies have revealed fractions of the molecular network that regulate pluripotency. The next section will present what is currently known of the network that sustains pluripotency in ES cultures.

### 1.1.3 The pluripotency network

**Nanog, Oct4 and Sox2, the core factors of pluripotency**

There are three factors that are primarily implicated in the regulation of pluripotency: Nanog, Oct4 and Sox2 (Young, 2011). The expression of the three above factors marks ICM in early development, while the expression of Nanog becomes heterogeneous at stage E3.5 and cells retaining Nanog expression are the ones reported to complete the epigenetic resetting of the inactive X chromosome (Chazaud et al., 2006, Silva et al., 2009).

**Nanog** (Homeobox protein Nanog) was discovered during screenings for factors that would be able to sustain pluripotency in the absence of LIF (Chambers et al., 2003, Mitsui et al., 2003). Nanog transfected cells were capable of proliferating in the absence of LIF, while also resisting differentiation (Chambers et al., 2003). Although during the first years after its identification, Nanog was perceived to be essential for the maintenance of pluripotency, this notion was challenged when Chambers et al reported that Nanog was not detected in ESC expressing Oct4 (Chambers et al., 2007).

Extensive experimentation carried out by Mitsui et al (Mitsui et al., 2003) showed that Nanog deficient ES cells (disruption of both alleles) expressed differentiation markers. In the same research embryonal Nanog deficiency was also studied and reported to be lethal, as Nanog -/- ICM could not remain

pluripotent in vitro, suggesting that Nanog is essential for efficient ESC derivation (Mitsui et al., 2003).

The concept of Nanog being essential for the maintenance of pluripotency was challenged when ESC not expressing Nanog were detected (Chambers et al., 2007). Furthermore, the Nanog gene was found to be subjected to transient downregulation that might be associated with cells being predisposed to differentiation (Chambers et al., 2007).

While Nanog is dispensable for ESC pluripotency maintenance, the same does not apply for early embryonal development. Disruption of Nanog in embryos leads to lethality (Mitsui et al., 2003), additionally in Nanog-/- embryos reactivation of the inactive X chromosome is not observed, linking Nanog with X chromosome reactivation in females (Silva et al., 2009). Moreover, Nanog-/- embryos demonstrated ICM apoptosis and inability to form hypoblast (Silva et al., 2009). The dispensability of Nanog in the maintenance of ESC pluripotency could potentially be associated with conditions that develop in vitro. However, Nanog in vivo, in the course of embryonal development is crucial for the establishment of ICM which will give rise to the epiblast.

**Oct4**, belonging to the POU domain family of proteins is expressed in the pluripotent epiblast, PGCs and totipotent blastomeres (Rosner et al., 1990, Scholer et al., 1990). Oct4 null embryos, much like Nanog null embryos, were found incapable of forming a pluripotent ICM but rather assumed a trophoctodermal character (Nichols et al., 1998). The expression of Oct4 is fundamental for the maintenance of pluripotency in ESC cultures, however its expression levels have to be regulated. Reduction of Oct4 mRNA levels will lead to the differentiation of the cell towards a trophoectodermal fate, whereas overexpression of Oct4 will lead to differentiation towards mesoderm and endoderm (Niwa et al., 2000).

**Sox2** is expressed in the ICM and extraembryonic ectoderm of blastocysts prior to implantation (Avilion et al., 2003). Sox2 null embryo phenotype resembles that of Oct4 null embryos. Blastocysts deficient of Sox2 cannot develop a pluripotent ICM (Avilion et al., 2003) and differentiate into trophoectoderm (Masui et al., 2007). The similar phenotype of the two

deficiencies is attributed to the fact that Oct4 and Sox2 act in synergy to regulate pluripotency associated genes with their own genes included (Rodda et al., 2005, Kuroda et al., 2005, Nakatake et al., 2006, Okumura-Nakanishi et al., 2005, Tomioka et al., 2002). A validation for this context was the fact that Sox2 deficient mESCs were rescued from differentiation by the ectopic expression of Oct4 (Masui et al., 2007).

### 1.1.4 The regulatory network of pluripotency

Nanog, Oct4 and Sox2 function as the hub of the pluripotency network and instigate a cascade of regulatory events that together with additional factors maintain pluripotency. The core pluripotency factors function together and co-occupy gene regulatory elements including their own promoters, creating dense regulatory loops (Young, 2011). Notably, it has been shown that the pluripotency factors act in concert when regulating gene expression. Notably, Nanog and Oct4 have been reported to co-occupy 44.5% of the total binding loci identified in a Nanog and Oct4 ChIP screen (Loh et al., 2006). The same has also been reported in the context of human stem cells, where Oct4, Nanog and Sox2 were shown to co-occupy gene regulatory loci (Boyer et al., 2005). Interestingly, in the co-occupied loci, the binding sites of Oct4, Sox2 and Nanog were in close proximity to one another, indicating that their proteins physically interact during genetic regulation (Boyer et al., 2005). Moreover, more than half of the bound loci corresponded to transcription factor genes, such as Nanog, Sox2, Oct4 and Rif (Loh et al., 2006). The regulatory activity of the pluripotency triplet involved activating as well as repressing the expression of their genetic targets (Boyer et al., 2005).

Regulatory regions of Oct4, Sox2 and Nanog are bound by Cohesin and Mediator factors that lead to DNA looping between promoter and enhancer regions, resulting in gene transcription (Kagey et al., 2010). Of interest are

recent findings reporting that Oct4, Sox2 and Nanog occupy gene regulatory regions and enhancers in specific in a very dense manner, marking super-enhancer regions (Whyte et al., 2013). Interestingly, super-enhancer regions are also bound by the Mediator, which is a transcription co-activator important for ESC maintenance, potentially recruited by Nanog, Oct4 and Sox2 (Whyte et al., 2013, Kagey et al., 2010). The overall chromatin architecture appears to play a significant role in pluripotency. DNA architectural conformations have been reported to exceed those of promoter-enhancer loops and extend to interactions across chromosomes. Notably, iPS reprogramming has been reported to change the overall chromatin interaction network by establishing DNA loops between the enhancer-promoter regions of Nanog, while also resulting in additional interactions between Nanog and other genes (Apostolou et al., 2013).

RNA interference experiments have expanded the list of known pluripotency factors with the addition of Essrb, Tcl1 and Tbx1 (Ivanova et al., 2006), as well as SetDB1 (Bilodeau et al., 2009) and Tip60-p400 (Fazzio et al., 2008), with the latter two factors being chromatin regulators. Additionally Paf1C (Ding et al., 2009), Trim28 and Cnot3 (Hu et al., 2009), were also determined to be important for murine ESC. In human ESCs, the TAF, COP9, INO80 and mediator complexes were also identified to be important for hESCs (Chia et al., 2010).

Affinity purification methods permitted the confirmation of the physical interaction of Oct4 and Nanog in murine ES cell lysates (Wang et al., 2006, Liang et al., 2008). Following similar experimental methods Sall4 was identified as an interactor of Nanog in mESCs (Wang et al., 2006, Wu et al., 2006) while it was also shown that the two factors regulate the expression of Sall4 and Nanog by binding their enhancer regions (Wu et al., 2006). Along with Sall4, numerous other factors were also identified, with 80% of them being required for early development, as demonstrated by knockout experiments (Wang et al., 2006). Oct4 and Nanog were found to physically interact with transcriptional repressors, namely Pml, Nurd and Sin3A complexes' members in murine stem cells (Liang et al., 2008, Wang et al., 2006). The data Liang et al. obtained showed that Nanog and Oct4 co-purified

with specific members of the NuRD and Sim3A-HDAC, with the list of co-factors not including core factors of the complexes, such as Mbd3 of NuRD (Liang et al., 2008). In the same study, Nanog, Oct4 and repressive factors were found to co-occupy genes targeted by Nanog in mESCs, indicating that the pluripotency factors interconnect with repressors to regulate transcription (Liang et al., 2008).

Van der Berg et al. generated a network of 166 Oct4 co-purifying factors, with the core NuRD factors Mbd3 and Rbbp7 (van den Berg et al., 2010). Multiple research teams have concluded that Oct4 interacts with chromatin remodeling complexes, SWI/SNF, FACT, MutsAlpha, LSD1, histone ubiquitination and NuRD (Pardo et al., 2010, Ding et al., 2012, Liang et al., 2008, van den Berg et al., 2010). Although the co-factors of Oct4 and Nanog have triggered more research, Sox2 has also been attributed its own interactors such as Sox21, which is implicated in differentiation pathways (Mallanna et al., 2010). Sox2 has also been found interact with the chromatin remodelling complexes NuRD and SWI/SNF, DNA repair and replication machinery as well as factors required for ESC self-renewal like Esrrb, Sall4, Mta1 and Mta2 (Gao et al., 2012).

### 1.1.5 Methods of cell reprogramming towards pluripotency

Pluripotency is a very important trait for development, however transient it may be. While naturally cell differentiation and gradual loss of cell potency is a uni-directional pathway is development, it is possible to reprogram differentiated somatic cells to acquire a pluripotent character. The most prominent methods are those of SCNT (Somatic Cell Nuclear Transfer) where reprogramming is carried out by the oocyte cytoplasm, iPS (Induced Pluripotent Stem cells) where reprogramming is initiated with the overexpression of four transcription factors and Stem Cell Fusion where reprogramming is carried out by the fusion of a somatic cell and a stem cell.

**iPS** (Fig. 1.1) cells already amount a decade of research as the method was first reported in 2006 (Takahashi and Yamanaka, 2006). Takahashi and Yamanaka (Takahashi and Yamanaka, 2006) showed that the overexpression of Oct4, Sox2, Klf4 and Myc (OSKM factors) was sufficient to induce reprogramming and generate, after cell divisions, ES-like cells. While the initial experiments had been conducted with murine cells, the OSKM factors proved to be sufficient for human iPS cell generation (Takahashi et al., 2007). At the same time an independent group also reported iPS generation, however this time from a different combination of factors, i.e. Nanog, Oct4, Sox2 and Lin28 (Yu et al., 2007).

Since the discovery of iPS research has shown that Sox2, Klf4 and Myc are redundant and can be substituted by additional factors belonging to the same protein family such as Sox1, Sox3 and Klf2, Klf5 (Nakagawa et al., 2008, Jiang et al., 2008).

**SCNT** (Somatic Cell Nuclear Transfer) (Fig. 1.1) is a method studied for more than half a century and takes place when a somatic cell nucleus is transplanted into an enucleated oocyte or egg (Gurdon, 1962). Upon its insertion into the female germ cell, the nucleus undergoes nuclear remodelling, including histone replacement with oocyte specific histone molecules such as H1foo (Gao et al., 2004), loss of DNA demethylation (Fulka and Fulka, 2007) and transcriptional activation of pluripotency associated genes, such as Oct4 (Wuensch et al., 2007).

**Stem Cell Fusion** (Fig. 1.1) on the other hand consists of the fusion between a somatic and a stem cell into a single cell. Stem cell reprogramming capacity was reported by both human and murine ESC (Tada et al., 1997, Matsumura and Tada, 2008). Direct cell reprograming of somatic cells with ESC lysate has also been reported, further demonstrating the capacity of stem cell to reprogram (Taranger et al., 2005).

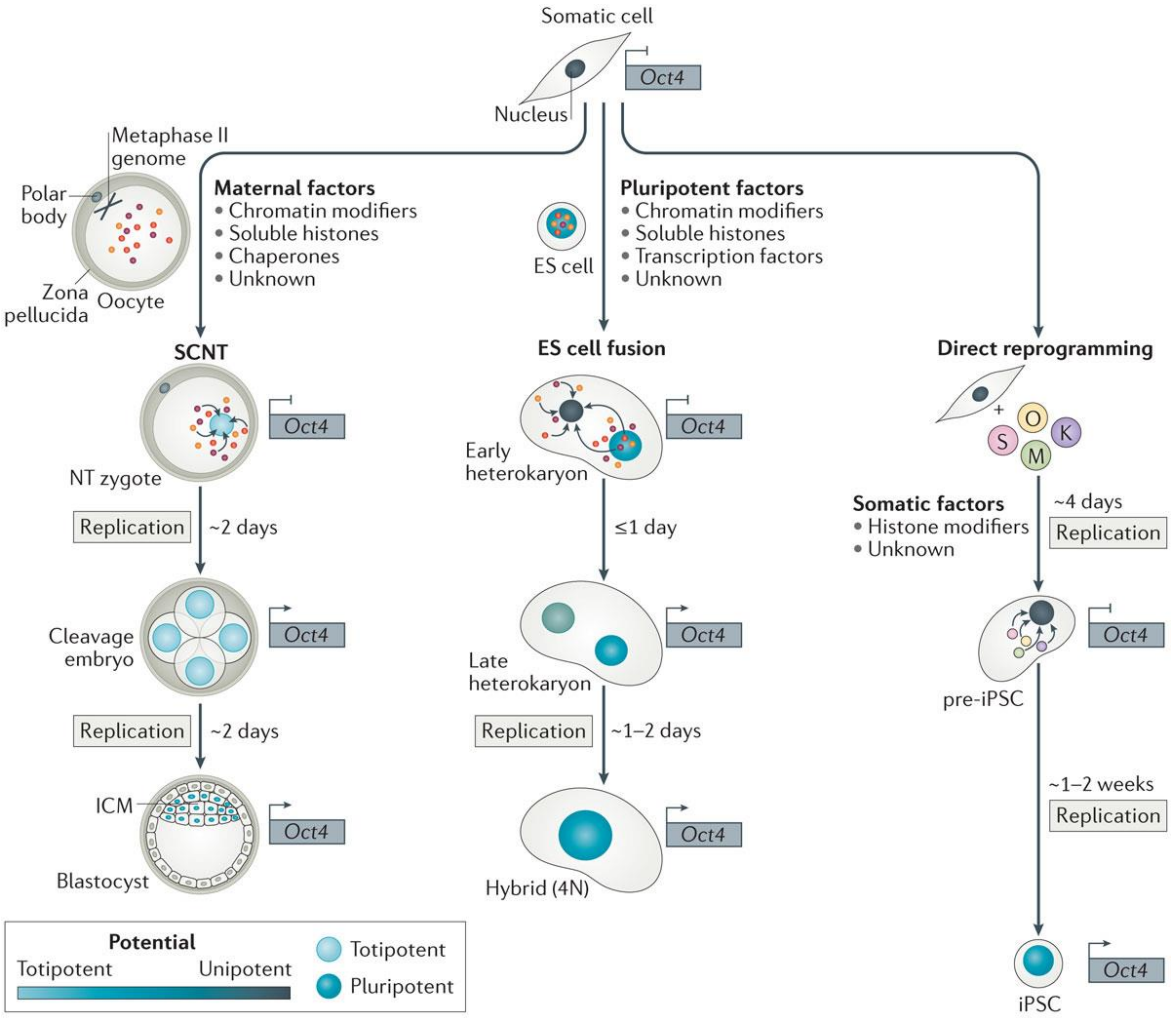



**Figure 1.1. The three methods whereby a somatic cell can be directed to an embryonic stem cell state: SCNT (Somatic Cell Nuclear Transfer), ES Cell Fusion and Direct Reprogramming (iPS). During cell differentiation pluripotency associated genes become silenced, mainly through epigenetic modifications. During development the potential of a cell ranges from totipotent (able to generate embryonic and extraembryonic tissue) to unipotent (engaged to one cell fate and unable to generate other cell types). During SCNT somatic nuclei are transferred into enucleated oocytes. Maternal factors globally and rapidly remodel somatic chromatin giving rise to the pluripotent character via embryo formation. During ES Cell Fusion, a somatic cell is fused with a stem cell. Upon cell fusion a heterokaryon is formed and global chromatin reprogramming is initiated. The pluripotent and somatic nuclei fuse after cell division creating a tetraploid nucleus. During iPS direct reprogramming Oct4, Sox2, Klf4 and Myc (OSKM) are introduced into somatic cells. The transition from the non-pluripotent state to the induced pluripotent cell state (iPSC) requires multiple divisions and extended latency (Image obtained from (Smith et al., 2016)).**

### 1.1.6 Comparing SCNT and iPS

The iPS method was a breakthrough in the field as it showed that a small number of factors can initiate and lead to the establishment of pluripotency. Up until the development of iPS, the existing cell reprogramming methods were either technically demanding, as was the case with SCNT, or resulted in the formation of polykaryons in the case of Stem Cell Fusion. Therefore the ability to generate pluripotent cells by factor overexpression, without disturbing their composition was received with excitement. iPS also paved the way for the clinical application of reprogrammed cells, as the generation of patient specific pluripotent cells was made more feasible. Moreover, the iPS methodology solved ethical problems arising with SCNT. Research and clinical applications demanded oocyte harvesting and experimentation with human oocytes was almost prohibitive.

Today, a decade after the method was reported, iPS has seen dramatic improvement. iPS cells have been obtained through the overexpression of a single gene, Oct4, as opposed to the initial OSKM tetrad (Li et al., 2011). Synthetic mRNA has also been employed for the generation of iPS cell with high efficiency and without the need to overexpress factors, thus omitting cellular antiviral response (Warren et al., 2010). microRNAs and direct delivery of reprogramming proteins has been employed to further advance the efficiency of iPS (Anokye-Danso et al., 2011, Kim et al., 2009). Additionally iPS cells have been obtained from a wide range of cell types body fluids, for example blood cells (Haase et al., 2009), keratinocytes (Aasen and Izpisua Belmonte, 2010) and urine (Zhou et al., 2011).

Despite all the improvements, SCNT is still considered more advantageous compared to iPS. SCNT exploits the native capacity of the oocyte to reprogram: when differentiated nuclei are injected in enucleated oocytes, reprogramming occurs in only 3.5 days leading in 12-20% stem cell derivation (Alberio, 2012). iPS has been reported to generate stem cells at a range of 0.001-3% in 3-4 weeks and with the analogous number of cell divisions

(Alberio, 2012). The longer time that is required for iPS cells to commence might allow the accumulation of genetic mutations or harmful epigenetic changes. Notably when compared to ESC, iPS cells demonstrated differences in their global gene expression and epigenetic profiles (Chin et al., 2009, Lister et al., 2011). Moreover, research has reported the introduction of de novo mutations and copy number variation over the course of iPS (Hussein et al., 2011, Gore et al., 2011). Additionally, iPS cells have been found to retain epigenetic traits of their donor cells, thus not being able to fully erase the epigenetic memories of somatic predecessors (Kim et al., 2010). On the other hand pluripotent murine cells generated from SCNT did not show retention of epigenetic memory when compared to stem cells (Brambrink et al., 2006).

The difference in the efficiency between the two methods triggered debate and research, while Jullien et al proposed that SCNT reprogramming was deterministic and its iPS counterpart was stochastic (Jullien et al., 2011). The term deterministic refers to events that happen with a certain order and are connected to one another, i.e. one leads to next. Stochastic, on the other hand, refers to events happening at a random order and without connection to one another (Jullien et al., 2011). However, the accumulation of iPS related findings showed that there is a series of events that appears to gradually establish the pluripotent character and rule out cells failing to follow the same pathway. Notably, the downregulation of the fibroblast associated marker Thy1 and the transition to the expression of the embryonic marker SSEA1 are the events that mark the initial stages of iPS reprogramming (Papp and Plath, 2011). With event sequences appearing to operate in both SCNT and iPS, it is probable that the main difference between the two methods lies in the extent of epigenetic reprogramming. More recent findings highlight the efficiency of SCNT reprogramming over iPS in terms of reprogramming parental methylation patterns. Notably, nuclear transfer generated ES cells correlated with IVF (In Vitro Fertilization) stem cells in terms of DNA methylation patterns (Ma et al., 2014).

The oocyte contains a depot of factors with many of them still unexplored. It is very likely that molecules participating in chromatin reprogramming, demethylation and wider gene regulation exist among the currently unknown factors of the oocyte. Therefore oocyte research will not only broaden our understanding of fertilisation and early development, but will also assist us in improving iPS and develop its clinical application.

## *1.2 Reprogramming with the oocyte*

### 1.2.1 The oocyte in fertilization and reprogramming

The event of fertilization is marked by remarkable epigenetic changes. Sperm chromatin is condensed in a manner only observed in sperm cells potentially for protective reasons (Ward, 2010). Over the course of fertilization sperm chromatin undergoes drastic remodelling (Teperek and Miyamoto, 2013). Unlike in somatic cells, sperm DNA is not bound to histones and organized in nucleosomes. Sperm DNA is tightly packed with protamines which are replaced by oocyte histones at the onset of fertilization. For example oocyte nusleoplasmin molecules (NPM) are directly involved in sperm decondensation by replacing protamines with histones (Ramos et al., 2010, Inoue et al., 2011a, McLay and Clarke, 2003). At the same time sperm chromatin reaches 10-fold decondensation and additional ooplasm proteins interact with paternal chromatin (McLay and Clarke, 2003). Additionally paternal chromatin, hypermethylated in its compacted state, undergoes significant demethylation upon fertilization (Oswald et al., 2000). Reprogramming of the paternal genome begins soon after fertilization, indicating that maternal factors contain chromatin remodelling activities.

The changes triggered during fertilization are orchestrated by an assortment of oocyte factors, the list of which is not yet complete; however research has revealed some of its members. Chaperone proteins have been reported to assemble histones on paternal chromatin after fertilization. Oocyte specific chaperones such as nucleoplasmin, NAP-1, NLP and TAP/p32 were shown to be responsible for the removal of sperm associated proteins and assembly of the maternal histones H2A and H2B on sperm chromatin in Xenopus and Drosophila (Philpott and Leno, 1992, Emelyanov et al., 2014). Maternal Hira, the chaperone of the histone variant H3.3 has been found to be of great importance in murine zygote development as it is involved in the reprogramming of the paternal genome (Lin et al., 2014). The findings from a wide range of species are indicative of how well conserved oocyte reprogramming is (Jullien et al., 2010).

Much like with sperm, donor nuclei swell when transferred into an oocyte and their size can increase up to 30-fold (Byrne et al., 2003, Gao et al., 2007, Gurdon, 1968). A result of the nucleus swelling which equals to chromatin decondensation is the increase in transcription which is observed in all cases of oocyte reprogramming (Gurdon et al., 1976, Gurdon, 1968, Murata et al., 2010). Somatic chromatin components are replaced by oocyte molecules in a procedure that very much mimics the events following fertilization. Oocyte specific histones, such as murine H1foo and Xenopus laevis B4 protein, replace somatic histone during reprogramming (Jullien et al., 2010, Gao et al., 2004, Teranishi et al., 2004). It has been shown that the replacement leads to the transcriptional activation of core pluripotency factors such as Oct4 and Sox2 (Jullien et al., 2010). Aside from the replacement of histones, other events complement the changes taking place on the donor chromatin. HP1 and BMI1 for example, proteins associated with heterochromatin are loaded onto chromatin and either supplement or replace somatic factors (Jullien et al., 2010, Teranishi et al., 2004). Maternal actin also accumulates in donor nuclei contributing to reprogramming (Miyamoto et al., 2011).

Donor somatic DNA comes with a distinct methylation profile which is a result of cell differentiation. Female germ cells possess demethylation enzymes that act upon sperm and donor chromatin and promote its demethylation (Bian et al., 2009, Simonsson and Gurdon, 2004).

Interestingly, the chromatin remodeling taking place during SCNT is potentially preparatory for transcriptional activation. As reported by Julien et al (Jullien et al., 2010), the oocyte histone B4 has been shown to accumulate at the re-activated promoters of Nanog, Oct4 and Sox2, while B4 incorporation was showed to be necessary for gene reactivation. Direct activation of Oct4 as a result of oocyte reprogramming has also been reported by Byrne et al (Byrne et al., 2003).

Moreover, it was demonstrated that different levels of epigenetic modifications, even though finally remodeled by the oocyte, had the capacity to stall Oct4 gene activation. More specifically, methylated and chromatinized DNA (i.e. containing histones) required double the time to be transcriptionally reactivated in Xenopus oocytes as compared to methylated DNA, devoid of histones. DNA, devoid of both histones and methylation marks was rapidly reactivated (Simonsson and Gurdon, 2004). The transcriptional activation capacity of the oocyte targets other genes as well, as demonstrated recently when MyoD was found to be reactivated in somatic cells by Xenopus oocytes (Biddle et al., 2009).

### 1.2.2 Egg and Oocyte: the advantage of immaturity

The terms egg and oocyte both describe and refer to the female germ cell, at different developmental stages though. Egg, is the mature (Metaphase II) female germ cell ready to receive a sperm cell and become fertilized. The term oocyte on the other hand refers to the immature egg cell (Germinal Vesicle stage) which needs to undergo maturation to reach the stage of

fertilization. Germinal Vesicle (GV) oocytes are also known as Prophase I oocytes. Both amphibian and mammalian oocytes are arrested at Meiotic Prophase I. Although oocytes at both developmental stages can be used in SCNT, the events that unfold after nuclei transfer are significantly different in the two oocyte stages.

In egg nuclear transfer a zygote forms bearing the DNA from the transplanted nucleus. The newly formed embryo undergoes cell divisions recapitulating the events that take place after fertilization. The events leading to the formation of a blastocyst are dependent on DNA replication (Gurdon and Wilmut, 2011). In oocyte nuclear transfer, on the other hand, no cell division takes place and no zygote is formed. The donor chromatin decondenses and the genome is demethylated, much like in egg nuclear transfer (Gurdon and Wilmut, 2011), however the step of DNA replication is omitted. The difference is associated with the developmental stage that the two oocyte types are arrested at. The difference between the two types of nuclear transfer experiments apart from reflecting the biological differences between mature and immature oocytes has also been exploited to provide a tool for the better understanding of the mechanisms underlying oocyte reprogramming.

The multiple cell divisions taking place during egg nuclear transfer obscure the reprogramming events that take place after the nuclear transfer. Oocyte nuclear transfer, on the other hand, omits cellular division and provides an advantage for the study of reprogramming. While DNA replication and cell division are absent in oocytes, transcription still takes place (Jullien et al., 2011). In both egg and oocyte nuclear transfer there is evidence of transcriptional activation (Halley-Stott et al., 2010, Jullien et al., 2010). However, the absence of cellular division in the case of the oocyte, allows for the study of the direct reprogramming events occurring before any DNA replication commences.

### 1.2.3 Egg and Oocyte lysate

Egg and oocyte lysates have been reported to retain their reprogramming capacity (Hansis et al., 2004) and enhance nuclear transfer success (Bui et al., 2008, Rathbone et al., 2010). Lysis of the female germ cell releases its reprogramming compounds thus creating a cocktail capable of chromatin remodelling. Oocyte lysate can therefore be used to study how reprogramming ensues in a DNA replication independent context (Alberio et al., 2005). Using cell lysates has the advantage of treating multiple cells at the same time and also omitting the complicated manipulation that oocyte/egg enucleation and nuclear transfer requires. Aside from that it also stretches the potential of research and valuable findings as it makes biochemical studies a reality. Egg/oocyte lysates can be studied biochemically for the identification of the fraction which possesses reprogramming activity; thence the composition of the fraction can be further delineated to lead to the identification and characterization of the molecules that compose it. The reprogrammome, the fraction of the oocyte that encapsulates its reprogramming activity has drawn the attention of researches attempting to identify oocyte specific reprogramming factors.

### 1.2.4 Reprogrammome: the reprogramming fraction of the oocyte

From the perspective of cell reprogramming, the oocyte has attracted much interest because of its reprogramming capacity. Between the iPS and SCNT, the oocyte has consistently been the most efficient reprogramming method and its efficiency has been attributed to the native capacity of the oocyte to remodel chromatin (Gurdon and Wilmut, 2011, Awe and Byrne, 2013). The identification of the oocyte reprogramming factors would not only delineate

the mechanism of chromatin remodelling during fertilization and SCNT but could also potentially improve the efficiency of iPS by pointing out molecules to be used in iPS trials.

The aim to identify molecules constituting the reprogrammome of the oocyte is challenging. It is not a task that can be approached through classical molecular biology; proteomic technologies are a lot more suitable to tackle the identification of the reprogramming fraction. The nature of the material itself adds another level of challenge to oocyte research. Retrieving oocytes, especially from mammals, can be challenging and requires technical expertise. Moreover, because of the oocyte numbers that are required for experimentation, proteomic research becomes almost prohibitive for mammals.

Over the past 10 years proteomic studies of the various developmental stages of the oocyte have emerged (Ma et al., 2008, Zhang et al., 2009, Yurttas et al., 2010, Wang et al., 2010). Ma et al identified a large number of murine oocyte proteins participating in cell cycle regulation, RNA processing and early embryo development. Additionally, the identified proteome was found to significantly overlap with mouse and human ESC proteomes (Ma et al., 2008).

DNMT1 (DNA methyltransferase 1), a methyltransferase primarily associated with the maintenance of methylation was found to be highly expressed in a proteome screen of MII murine oocytes (Zhang et al., 2009). Interestingly oocyte DNMT1 (DNMT1O) has been reported to be a smaller protein compared to somatic cell DNMT1 and also have an oocyte specific exon (Yurttas et al., 2010). Notably, a significant number of proteins in the oocyte were found to be liable to alternative splicing and post-translational modifications reflecting the functional complexity of the oocyte proteome (Ma et al., 2008). This is an important consideration for the search of oocyte specific factors, as the identified proteins should not only be assessed for their oocyte exclusiveness, but for differences in splicing, exons and post translational modifications.

Zhang et al also reported an intersection between the identified oocyte and mESC proteins, with the factor DPPA5 (Developmental Pluripotency

Associated 5) present in the overlapping fragment (Zhang et al., 2009). At the same time the maternal factors MATER and TLE6 were identified in the oocyte protein screen (Zhang et al., 2009). However, it is worthwhile to say that even though in both cases over 300 proteins were identified a significant number of the identified molecules has not yet been characterised. It is therefore possible that proteins with oocyte specific attributes might reside among the non-characterised factors.

The proteomic comparison of murine GV, MII stage oocytes and zygotes has been also carried out exploiting more up to date technical manipulations which led to the identification of more oocyte proteins compared to previous attempts (Wang et al., 2010). According to the data and the conclusions of the team (Wang et al., 2010), MII stage oocytes express higher amounts of transcription factors, epigenetic regulators and DNA repair proteins. However, the differences between the two oocyte stages appear to be quantitative rather than qualitative (Wang et al., 2010). Wang et al (Wang et al., 2010) suggested that MII stage oocytes might be more appropriate for reprogramming due to their elevated expression of transcription factors and transcriptional activators. Biological trials on the other hand, have demonstrated that GV stage oocytes can reprogram (Bui et al., 2008, Miyamoto et al., 2011, Kong et al., 2013) and therefore it would be safer to suggest that both immature and mature oocytes are capable of reprogramming, but during maturation there is an increase in the expression of certain factors. Additionally GV stage oocytes, as previously discussed, do not promote DNA replication and therefore reflect reprogramming carried out on chromatin prior to any transcriptional activation over the course of oocyte reprogramming.

Oocytes and ES cells both contain factors that have the ability to reprogram as shown by nuclear transfer and stem cell fusion experiments (Han et al., 2008). The proteome of ES cells has been widely studied and the core pluripotency network has been validated through mass spectrometry. Teams studying the oocyte proteome have used the ES proteome as a reference for the proteins identified in oocytes. Specifically, Pfeifer et al (Pfeiffer et al., 2011) coined the

term *reprogrammome* and then attempted to decipher it by comparing oocyte and ES through mass spectrometry. During the comparison the team took into account only proteins that were identified in both ES and oocytes and complied with three conditions: to be nucleolar, possess catalytic activity and also be associated with chromatin modification. 28 factors were found to satisfy the criteria and were presented as the basis to understand and further investigate the reprogramming fraction of the oocyte. RUVBL1 and RUVBL2, implicated in transcriptional regulation and DNA damage response, SMARCA4, member of the BAF chromatin remodelling complex, as well as DNMT1 and DNMT3A methyltransferases were present among the 28 factors reported as potential reprograming candidates (Pfeiffer et al., 2011). Proteomic studies carried out earlier also compared oocyte with ES proteome based on the same rational (Ma et al., 2008, Zhang et al., 2009) and identified significant overlap between oocytes and ES cells, however Pfeiffer et al identified candidate reprogramming factors by establishing the aforementioned criteria (Pfeiffer et al., 2011). Fractionation experiments have been carried out to isolate molecules associated with chromatin and have successfully identified MLL2 to be required in oocyte epigenetic reprogramming, as MLL2 (Histone lysine Methyltransferase 2D) knockout oocytes were not able to establish transcriptional repression, while resulting in abnormal histone tail modifications (Andreu-Vieyra et al., 2010). Recently, an additional factor was found to be required for reprogramming. VIM (Vimentin), was identified by mass spectrometry and upon its inhibition led to lower numbers of SCNT blastocysts (Kong et al., 2014).

## 1.3 DNA methylation and demethylation

### 1.3.1 Methylated cytosine (5mC), the fifth base and transcriptional repressor

DNA methylation occurs with the deposition of a methyl mark on the fifth carbon of cytosine, which predominantly happens within CpG dyads. The positioning of CpG dinucleotides is not even across the genome; CpG islands (CGIs) harbour higher number of CpGs in comparison to the rest of the DNA. CGIs are grouped into higher and lower methylation bearing regions, with lower methylation regions overlapping with gene regulatory loci (Stadler et al., 2011). The low methylation regions (LMRs) represent loci with lower CpG numbers and this is where DNA methylation has been reported to have gene significant transcriptional impact (Schubeler, 2015).

The functional characterisation of DNA methylation has been marked by experiments reporting that methylated DNA is transcriptionally inactive when injected into Xenopus oocytes (Vardimon et al., 1982) and mammalian cell cultures (Vardimon et al., 1982, Stein et al., 1982). During the last four decades more research has showed that DNA methylation correlates with transcriptional repression, thus corroborating the initial findings (Shen et al., 2007, Hsieh, 1994, Wolffe and Matzke, 1999, Riggs, 1975).

The modification upon cytosines is introduced by DNA methyltransferases (DNMTs), a family of highly conserved enzymes between animals and plants (Law and Jacobsen, 2010). DNA methylation, as an epigenetic mark, combines stability and flexibility. It can be deposited over non-methylated cytosines and maintained over cell divisions in descending cells, retaining however its reversibility. The DNMT enzymes are specialized accordingly as DNMT3A and DNMT3B, which are de novo DNA methyltransferases (Okano et al., 1998, Okano et al., 1999), i.e. involved in the initial deposition of the methyl mark when the methylation profile becomes established early in development, while DNMT1 is a maintenance methyltransferase (Hermann et al., 2004).

Additional members of the DNMT family are DNMT3-like (DNMT3L), a nuclear protein that does not contain methyltransferase activity but interacts with DNMT3A and DNMT3B, therefore playing an indirect role in de novo methylation (Suetake et al., 2004). DNMT2 has been described as lacking the N-terminal domain of the DNMT family, even though it possesses a catalytic domain (Turek-Plewa and Jagodzinski, 2005). Lacking the N-terminus translates into incapacity of DNA binding and further inability to methylate DNA; however, DNMT2 appears to play a role in tRNA methylation (Goll et al., 2006).

Both de novo and maintenance methyltransferases are very important developmentally as their deletion or deficiency causes lethality at the embryo stage or later (Okano et al., 1999, Li et al., 1992). Maintenance methyltransferases preserve DNA methylation between DNA replication cycles by relying on the semi-conservative nature of replication. Methylated DNA duplexes unzip and replicate; however, the newly synthesized strand will not bear any methyl marks. DNMT1 is then recruited to hemimethylated DNA by UHRF1 (Ubiquitin like with PHD and RING finger domains 1), a nuclear protein capable of recognising and binding hemimethylated DNA (Law and Jacobsen, 2010, Avvakumov et al., 2008). Models advocating for DNA methylation as a cell memory transmittance unit have been published as early as 1975 (Riggs, 1975, Holliday and Pugh, 1975).

De novo DNA methyltransferases on the other hand, deposit the methyl mark thus establishing methylation profiles early in development. Loss of de novo methyltransferases in males leads to spermatocyte apoptosis and meiotic arrest in spermatogonia (Bourc'his and Bestor, 2004). Oocytes lacking de novo methylation activity are impacted similarly as they result in embryonic fatality at E10.5 (Bourc'his et al., 2001).

The expression of de novo methyltransferases climaxes during the waves of de novo methylation occurring in germ cells and early embryos (Sakai et al., 2004, La Salle et al., 2004, Watanabe et al., 2002). De novo DNA methylation occurs in PGCs and establishes germ-cell specific methylomes; it, however, commences at different developmental stages in females and males. Both

genders have fully established their methylome pattern at birth, which is then maintained through cell divisions (Davis et al., 2000).

DNA methylation also plays a role in pluripotency. In global methylation analysis experiments comparing ES cells with cells of differentiated lineages, pluripotency associated genes as well as chromatin remodelling genes were found to be hypomethylated in ES cells (Farthing et al., 2008). Along with the gene bodies, promoters were found to be differentially methylated as well (Farthing et al., 2008). Nanog, a core pluripotency factor demonstrates significant methylation differences in differentiated and pluripotent cells. The Nanog promoter was found to be methylated in sperm DNA and demethylated in ES cells and the zygote, indicating that the demethylation occurred over the course of fertilization, with demethylation being crucial for Nanog expression (Farthing et al., 2008). Further investigation led to the realization that the Nanog promoter becomes demethylated during iPS (Park et al., 2008) as well as oocyte reprogramming (Bian et al., 2009).

### 1.3.2 DNA methylation reversal: Demethylation

While DNA demethylation can be transmitted from parental to daughter cells it is marked by reversibility and can be removed, resulting in non-modified cytosine. Two global demethylation waves have been described during the course of development (Fig. 1.2). One of these occurs in the preimplantation zygote, where sperm chromatin is subjected to wide epigenetic remodelling. Sperm DNA is tightly packed prior to fertilization and reorganizes upon oocyte penetration. After the replacement of sperm protamines with oocyte histones, sperm DNA is subjected to extensive DNA demethylation (Mayer et al., 2000, Oswald et al., 2000). The post fertilization demethylation of sperm DNA is completed before the first zygotic cell division and is observed in mammalian and non-mammalian animal models, with some exceptions (Beaujean et al., 2004, Jiang et al., 2013).

Another wave of demethylation is observed during the specification of the PGC line, when a number of cells originating from the posterior epiblast are induced to become Primordial germ cells (PGC). The cells that will form the germ cells of the mature organism migrate to reach the gonadal ridge, the precursor of the adult gonad. PGCs are globally demethylated during their migration (Seisenberger et al., 2012).

DNA demethylation during preimplantation development and PGC specification are extensive epigenetic modifications, altering the methylome globally or at a large scale. However site specific demethylation has also been observed affecting distinct genomic loci and under certain stimuli. T-lymphocytes undergo locus specific demethylation at the regulatory region of interleukin-2 after activation (Bruniquel and Schwartz, 2003). The regulatory regions of BDNF (Brain Derived Neurotrophic Factor) have also been reported to be subjected to demethylation upon neuronal stimulation (Martinowich et al., 2003).

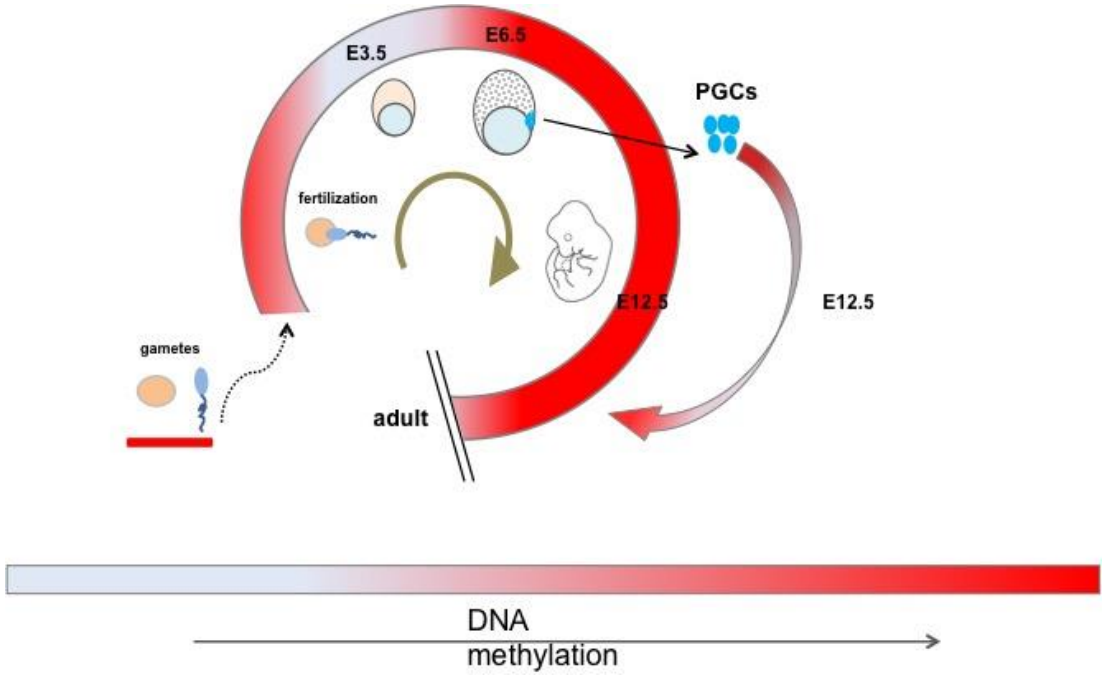



**Figure 1.2. Methylation and demethylation waves during development. The male and female gametes enter the cycle of reproduction bearing germ cell associated methyl marks. Demethylation commences upon fertilisation and climaxes at stage E3.5. De novo methylation begins at stage E6.5 and is maintained through development with PGCs being the only exception. At stage E13.5 PGCs undergo a wave of demethylation and are subsequently remethylated to acquire a PGC specific methylation profile. PGC methylome is then maintained through cell divisions.**

### 1.3.3 DNA demethylation mechanisms

Active and passive mechanisms have been proposed as potential DNA demethylation pathways. Passive DNA demethylation refers to the involvement of DNA replication and thereby gradual dilution of methylation. Active DNA demethylation, on the other hand, refers to DNA replication independent erasure of the methyl mark.

Various active mechanisms have been proposed to explain active demethylation. Successive oxidation of 5mC, direct excision of the methyl mark from the affected cytosine and removal of the methylated cytosine through the BER pathway are among the proposed models.

**Oxidative demethylation** is the leading model for active demethylation and relies on the successive oxidation of 5mC to 5hmC, 5fC and 5caC. The existence of such a pathway was first indicated in bacteria (Falnes et al., 2002, Trewick et al., 2002) and yeast (Smiley et al., 2005), which led to the identification of mammalian homologue molecules participating in the demethylation pathway (Tahiliani et al., 2009). TET1, the first TET (Ten Eleven Translocation) protein family molecule to be identified, was shown to oxidize 5mC to 5hmC (5-hydroxy-methylcytosine) (Tahiliani et al., 2009). It later became clear that along with TET1, TET2 and TET3 also possess the same activity (Ito et al., 2010). Interestingly triple TET knockouts ESCs were found to be devoid of the 5hmC mark, indicating that no other factors are involved in the oxidation of 5mC to 5hmC (Dawlaty et al., 2014). 5mC has been widely associated with transcriptional repression, but the epigenetic significance of the oxidative derivatives is still elusive. What is of outmost interest is that while 5mC is distributed among all cell lineages, the rest of the four modified cytosines show an uneven pattern of distribution. 5hmC is present at 40% in Purkinje neurons (Kriaucionis and Heintz, 2009), while in ESC it is lower than 0.1% (Tahiliani et al., 2009). According to oxidative demethylation where 5mC is oxidised to 5hmC, the latter is oxidised to 5fC (5-formyl-cytosine) and 5fC

can be oxidised to 5caC (5-carboxyl-cytosine), while all the oxidation steps are carried out by TET proteins. The last oxidation products, 5fC and 5caC can be removed by a DNA glycosylase with the resulting gap subsequently repaired by the Base Excision Repair (BER) pathway (Xu and Wong, 2015). The DNA glycosylase that has mostly been related to the removal of 5fC and 5caC from DNA is TDG (Thymine DNA Glycosylase). Reportedly TDG recognises 5fC AND 5caC and its knockout in iPS cells was reported to lead to the accumulation of 5caC (He et al., 2011, Maiti and Drohat, 2011). Several studies have supported BER mediated oxidative DNA demethylation (Shen et al., 2013, Wu et al., 2014, Song et al., 2013).

An alternative active demethylation pathway proposed involves Activation Induced Deaminase (AID) and Apolipoprotein B mRNA Editing Enzyme Catalytic polypeptide (APOBEC) which deaminate methylated cytosine (5mC) to thymine (T) creating a mismatch that will trigger DNA repair mechanisms (Morgan et al., 2004). However recent finding have challenged the AID/APOBEC model while strengthening TET dependent DNA demethylation (Hogenbirk et al., 2013, Arioka et al., 2012).

The **direct removal of the methyl mark** from methylated cytosines requires the existence of an enzyme capable of catalysing the dismantling of the carbon-carbon bond formed between the base and the methyl additive. MBD2 (Methyl CpG Binding Domain protein 2) was associated with the excision of the methyl mark due to its ability to read and bind methylated cytosines (Bhattacharya et al., 1999). However, the role of MBD2 in direct demethylation is controversial due to the inability of other laboratories to repeat the results. Regardless of the MBD2 controversy, enzymes capable of cleaving the methyl mark have been described in the context of histone demethylation (Klose et al., 2006, Klose and Zhang, 2007), and therefore suggest that the breakage of the carbon-carbon bond is enzymatically feasible.

The **Base Excision Repair (BER)** pathway has been proposed as a possible mechanism for DNA demethylation. The BER mechanism initiates demethylation by cleaving the methylated cytosine of the DNA backbone,

thus leading to the creation of an apurinic/apyrimidic gap that triggers DNA gap repair pathways (Sancar et al., 2004). The DNA repair mechanism as a demethylation apparatus was first isolated and described in chicken embryo extracts (Jost, 1993, Jost et al., 1995). Thymine DNA Glycosylase (TDG) enzymatic activity was isolated as part of the demethylation complex in the chicken embryo (Zhu et al., 2000) and was considered to be the enzyme that initiates demethylation by creating the nucleotide gap to be corrected by the BER pathway.

**Nucleotide Excision Repair (NER)** has also been implicated in DNA demethylation, mostly in the context of the oocyte. More specifically GADD45a has been linked with the NER pathway in oocyte demethylation (Barreto et al., 2007). However similar functions of GADD45a were not observed in somatic, potentially indicating that the mediation of GADD45a/NER demethylation is oocyte exclusive (Jin et al., 2008, Engel et al., 2009).

### 1.3.4 The DNA repair pathways BER and NER and their implication in demethylation

The two major DNA repair pathways residing in the cell are the Base Excision Repair (BER) and the Nucleotide Excision Repair (NER) pathways (Fig. 1.3).

The **BER** pathway takes over to repair non-bulky nucleobase lesions that do not distort the DNA helix (Zharkov, 2008). The first step in the BER pathway comprises of the creation of an AP (Apurinic/Apyrimidic, also known as abasic) site, when a DNA glycosylase removes the damaged base. DNA glycosylases involved in BER are NTHL1, MPG, TDG etc. The AP site, which is a substrate for AP endonucleases such as APEX1, is then subjected to hydrolysis thus resulting in a single strand break. From that point onward BER branches to two BER sub-pathways, based on the size of DNA that will be synthesized and therefore repaired. Short patch BER (SP-BER) refers to a single base being

excised and repaired. After the repair has commenced and the new base has been deposited by POLB, the remaining nick is sealed by a DNA ligase I or III. Long patch BER (LP-BER) on the other hand, repairs a larger DNA fragment and a larger DNA patch is synthesised by POLB (or POLD) displaces the respective DNA piece. Flap endonuclease FEN1 then removes the DNA flap created by the displacement and the resulting nick is subsequently ligated by DNA ligase I or III (Zharkov, 2008) (Fig. 1.3).

The **NER** pathway repairs bulky helix distorting lesions. Unlike BER, in the NER pathway the DNA duplex has to first be unwound before the damaged DNA sequence is excised from the locus. Much like BER, NER branches in two sub-pathways. Global Genome NER (GG-NER) and Transcription Coupled NER (TC-NER). GG-NER is implicated in the repair of lesions throughout the genome, while TC-NER repairs lesions on the coding strand of actively transcribed genes (Shuck et al., 2008). Double helix distortions play an important role in the recognition of the damaged locus during GG-NER. The proteins involved in the recognition of the lesion are XPC, RAD23B, CETN2 that compose the trimeric complex of XPC (Shuck et al., 2008). UV-damage is also repaired by the NER pathway, after the contribution of auxiliary factors such as DDB1 and XPE (DDB2). These auxiliary proteins augment the helix distortion and render the DNA fragment recognisable for the aforementioned recognition factors. In the case of TC-NER, when an RNA polymerase II actively transcribing a gene stalls at a damaged locus, the factors CSA and CSB (also known as ERCC8 and ERCC6, respectively) disassemble the transcriptional machinery on site and allow for repair factors to access the locus. After the recognition of the lesion, in both TC and GG-NER sub-pathways, the TFIIH complex is recruited on the site and the DNA duplex is subjected to unwinding. Subsequently, the DNA strand region containing the damage is excised by endonucleases at the 5’ and 3’ ends, leading to the removal of approximately 30 nucleotides. DNA polymerases such as POLB, POLD or POLE, using the undamaged DNA strand as template, synthesize its complementary strand and the nicks at the 3’ and 5’ ends are ligated by DNA ligases (Shuck et al., 2008) (Fig. 1.3). The proteins implicated in the NER and BER pathways can be seen in Table 1.1.

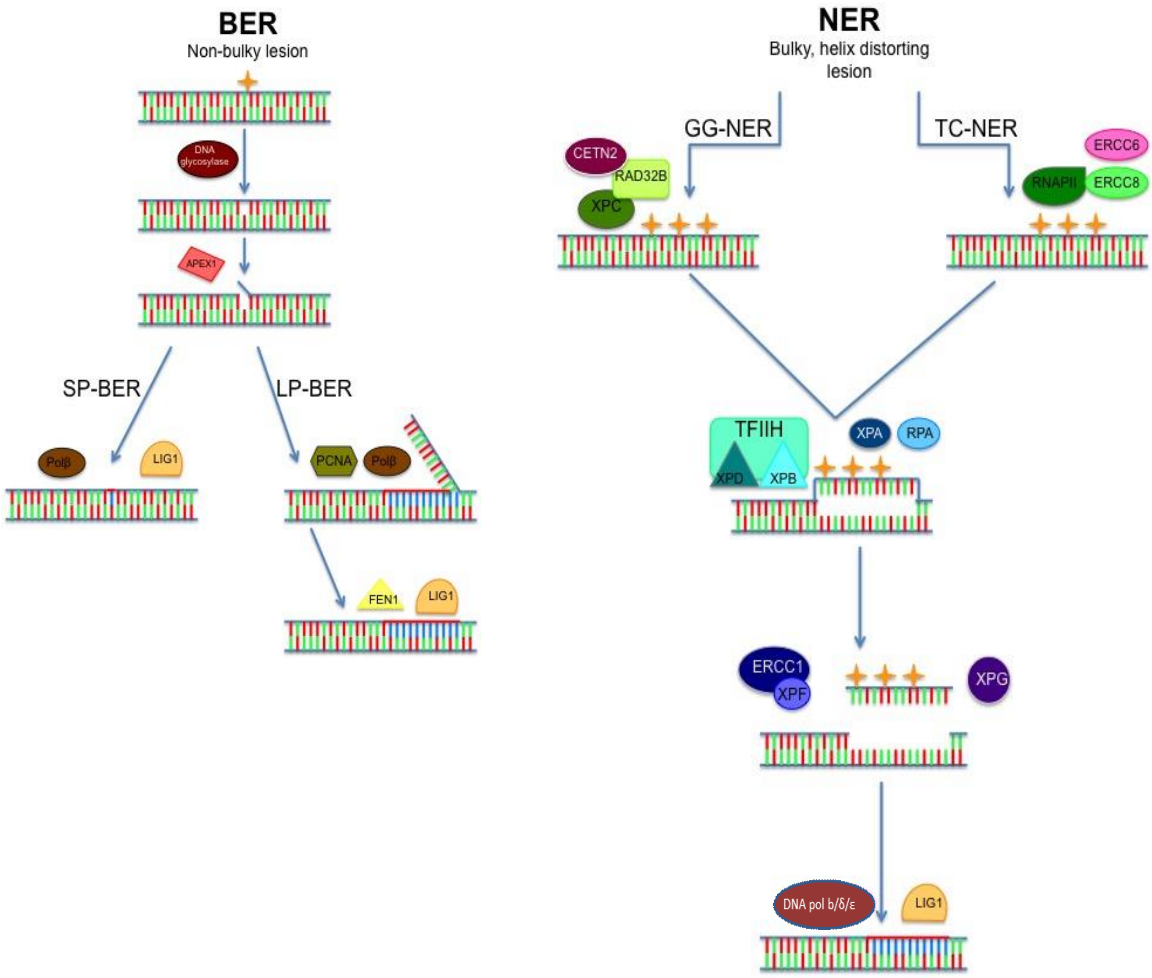



**Figure 1.3. The NER and BER DNA repair pathways. The BER pathway begins with the recognition of the lesion by a DNA glycosylase and the creation of an AP site, after the cleavage of the target base. APEX1 recognises the AP site and creates a single strand break at the 5' end of the AP site. Subsequently, DNA Polβ repair the gap and this is where the pathway branches to SP-BER and LP-BER. In SP-BER a single nucleotide gap is filled by the complementary base, while in LP-BER more nucleotides are inserted dislocating the damaged strand. In both cases, after the DNA is repaired LIG1 (or LIFIII) ligates remaining nicks and FEN1 removes the dislocated DNA flap in LP-BER. The NER pathway branches at the stage of lesion recognition as different factors identify lesions during transcription and globally. RNAPII, ERCC8 and ERCC6 recognize lesions on actively transcribed genomic loci while XPC-RAD23B-CETN2 recognise damaged DNA globally. After the recognition both GG-NER and TC-NER converge to the same NER mechanism with the recruitment of the TFIIH multi-subunit factor, XPA and RPA that unwind the damaged locus thus creating a nucleotide bubble. The damaged sequence is subsequently removed by ERCC1, XPF and XPG and at the final step DNA polδ or ε fill in the gap by synthesizing the complementary strand. Any remaining nicks are ligated by DNA ligase.**

| Base Excision Repair | Nucleotide Excision Repair |
|---|---|
| UNG, TDG, SMUG1, NTHL1, MUTYH, OGG1, MBD4, NEIL1, NEIL2, NEIL3, MPG, FEN1, XRCC1, PARP1, PCNA, RAD9B, RAD1, HUS1, RFC, TDP1, LIGI, LIGIII, Polγ, Polδ, Polε, Polα, Polζ, Polβ, Polλ, APEX1, APEX2, PALF, PARP1 | XPA, p36, XPB, TFIIH, XPC, RAD23B, CETN2, DDB1, DDB2, XPD, XPE, XPF, ERCC1, XPG, RPA, p70/p34/p14, Polε, LIGI, Polδ, XRCC1, LIGIIIa, CSA, CSB, PCNA, RFC3, Polβ |

**Table 1.1. Proteins implicated in BER and NER pathways (Zharkov, 2008)**

Interestingly while TET3 is highly expressed in oocytes and at the same time 5hmC, the oxidative product of 5mC, was found to be present in murine zygotes (Wossidlo et al., 2011, Gu et al., 2011), findings emerge that suggests alternative DNA demethylation mechanisms in the context of the oocyte. Additionally, 5fC and 5caC were also identified in the paternal pronucleus following fertilization, while their removal was linked with dilutive mechanisms dependent on DNA replication (Inoue et al., 2011b). Two individual groups have reported that the demethylation of the paternal DNA commences in two distinct phases after fertilization (Amouroux et al., 2016, Santos et al., 2013). In the first phase of demethylation that happens in the early embryo stages of PN0-PN2, the loss of the 5mC mark is not linked to 5hmC increase or TET activity. TET dependent 5hmC increase marks the later stages of PN3-PN4, when the 5mC mark has already been depleted from the paternal pronucleus (Amouroux et al., 2016). The first phase of demethylation is linked with AID and the UNG2 glycosylase that initiates Long Patch BER to complete demethylation (Santos et al., 2013).

The NER repair pathway has been implicated with oocyte DNA demethylation with the mediation of GADD45a that was shown to interact with XPG, a core contributor of the NER pathway (Barreto et al., 2007). GADD45a has also been implicated in the demethylation of the Oct4 promoter in Xenopus oocytes, indicating gene regulation specific mode of function (Schafer et al., 2010, Barreto et al., 2007).

Interestingly, additional BER associated factors have been reported to have a participation in oocyte demethylation. PARP1 (Poly-ADP-ribose polymerase 1) which catalyses parylation (poly-ADP-ribosylation), a post-translational modification was found to be present in high levels in zygotic pronuclei (Hajkova et al., 2010). At the same time, PAR (poly-ADP-ribose) accumulation was also observed, highlighting the enzymatic activity of PARP1 (Hajkova et al., 2010). It has been reported that PARP1 is associated with single strand DNA breaks and upon the recognition of one it auto-modifies itself and other DNA related proteins by adding ADP-ribose groups (De Vos et al., 2012). Protein parylation results in recruitment of additional factors, necessary for the completion of DNA repair and also mark the size of the DNA lesion (De Vos et al., 2012). PARP1 has been reported to parylate histones and potentially result in chromatin loosening and allow further DNA repair factor binding (Krishnakumar and Kraus, 2010). The chromatin remodelling factor ALC1 (Amplified in Liver Cancer 1) of the SNF2 ATPase superfamily, has been found responsive to paryliation, which results in its recruitment on chromatin to promote chromosome sliding, thus rendering DNA more accessible (Gottschalk et al., 2009).

Additional data have linked PARP1 with the NuRD (Nucleosome Remodeling and Deacetylase complex) and PRC1 (Polycomb Repressive Complex 1) complexes (Chou et al., 2010, Polo et al., 2010). Specifically, the teams found the recruitment of the aforementioned complexes to be PAR-dependent and together their data suggested that PARP1 potentially recruits inhibitory complexes to lesion sites to inhibit and dislocate RNA Pol II (Chou et al., 2010, Polo et al., 2010).

Additionally, PARP1 was reported to form a heterodimeric complex with DNMT1, which resulted in the latter's catalytic function inhibition (Reale et al., 2005). For the complex formation PARP1 has to be auto-parylated as DNMT1 recognises and binds poly-ADP-ribose (Reale et al., 2005). More recent findings want PARP1 to operate during PGC demethylation wave as PAR and PARP1 were found highly expressed at that stage. Additionally, inhibition of PARP activity impaired DNA demethylation in PGCs (Ciccarone et al., 2012).

Interestingly, PARP1 has been found to interact with UHRF1, which is implicated in DNA methylation maintenance along with DNMT1 (De Vos et al., 2014). Moreover, UHRF1 was found to physically interact with the DNA glycosylase MPG, indicating potential association of UHRF1 to the BER pathway (Liang et al., 2013).

## *1.4 Axolotl as an experimental organism*

### 1.4.1 Differences between Xenopus and *axolotl*: Preformation and Epigenesis

Ambystoma mexicanum, widely known as *axolotl*, is an amphibian urodele, used in some classical experiments that marked modern developmental biology. *Axolotl*s are neotenic animals: they reach adulthood without undergoing metamorphosis; they are able to heal without scarring and regenerate body parts. Experimenting with *axolotl* oocytes/embryos and observing vertebrate embryo development is easy due to the large size of their eggs. Currently, the *axolotl* is being used as a research model in various branches of biology, as a result of these attributes. However, the most interesting feature of the axolotl is its difference from frogs in terms of germ cell specification. While Xenopus oocytes have been widely used in SCNT experiments, research results argue that pluripotency is compromised in the lineage of anurans (frogs).

In anurans, Primordial Germ Cells (PGC), originate by a mechanism known as preformation. Preformation refers to the specification of germ cells by maternal determinants, which are localized in what is also known as germplasm. Germplasm, however, is absent in *axolotl*s. In this organism PGCs are induced by signaling molecules during gastrulation. Since the ’60 there is evidence that frogs and urodeles, though belonging to the same taxonomic

class, exhibit different mechanisms of germ cells specification. Niewkoop (Nieuwkoop, 1969) found that *axolotl* animal cap cells could be induced to give rise to PGCs, while Xenopus animal caps did not give such a response (Michael, 1984). Xenopus germplasm forms the Mitochondial Cloud (MC), that then locates germplasm to the vegetal cortex of the egg (Houston and King, 2000). The cells that inherit germplasm produce PGCs, that will then give rise to the germ cells (Wylie, 1999). A comparison between *axolotl* and xenopus stage l previtellogenic oocytes showed that a mitochondrial cloud is not present in *axolotl* oocytes (Johnson et al., 2001), which serves as another argument in favor of the regulative mode of PGC derivation.

Moreover, *axolotl* Dazl was shown to lack the localization of its xenopus counterpart, Xdazl (Johnson et al., 2001). Xenopus DAZ-like, Xdazl is present in germplasm at all stages of oogenesis and embryogenesis and is therefore a marker of germplasm in development (Houston et al., 1998). More recently, further research corroborated that *axolotl* PGCs result from inductive signals. More specifically, FGF signaling was found to induce PGCs, as well as Brachyury and BMP4 that were able to increase the expression of PGC markers when injected in animal caps (Chatfield et al., 2014).

Germ plasm does not form in mouse embryos, whose germ cells arise from cells clonally identical to somatic cells (Lawson and Hage, 1994). According to Tam and Zhou (Tam and Zhou, 1996), somatic cells can enter germ cell fate if translocated to specific embryonic regions. Various research teams have found that signaling molecules belonging to the BMP family are needed for the production of germ cells in mice (Lawson et al., 1999, Ying and Zhao, 2001, Ying et al., 2000). In addition, when comparing embryos of *axolotl* and mammalian embryos, such as mice, similarities in terms of PGC development were reported (Johnson et al., 2003).

Regulative germ cell specification has pluripotency as a prerequisite, i.e. embryos from species with regulative germ cell specification need to have pluripotent cells as template in order to successfully generate PGCs. While with preformation of PGCs pluripotency is compromised and limited only to the generation of somatic cell lineages. Therefore it can be said that

epigenesis (regulative mode of PGC specification) is concomitant with pluripotency. Indeed, the gene regulatory networks governing pluripotency are conserved in *axolotl*s and mammals (Dixon et al., 2010, Swiers et al., 2010). On the other hand, core factors of the pluripotency network have been deleted in frogs (Dixon et al., 2010, Frankenberg and Renfree, 2013).

### 1.4.2 The pluripotency network in the *axolotl* oocyte

*Axolotl* Nanog, homologue of mammalian Nanog, expression was reported in *axolotl* embryos (Dixon et al., 2010). Up till then, Nanog was considered a trait of the mammalian lineage, as there was no previous record of Nanog in non-mammalian animal models. The finding of Nanog in *axolotl* is consistent with the regulative mode of PGC specification in urodelels, while it also corroborates pluripotency.

Apart from Nanog, Sox2 and Oct4 are also present in *axolotl* oocytes, unlike Xenopus where Oct4 is absent from the Xenopus genome and is replaced by other Pou-domain factors (Unpublished data, Loose M., Johnson A.). Pou5f1 (Oct4) is also non-existent in Xenopus and the POU protein family is represented by three POU genes that recapitulate the expression pattern of Oct4 (Livigni et al., 2013). Along with Nanog, the presence of Oct4 and Sox2 could suggest the existence of the pluripotency network, thus far described in ESC, in the *axolotl* oocyte. Therefore the existence of the Oct4-Sox2-Nanog network in *axolotl* is indicative of evolutionary differences between species of amphibians (Bachvarova et al., 2009).

Additionally, research has demonstrated that *axolotl* oocyte extracts are more efficient in reprogramming mammalian cells than Xenopus oocyte extracts (Allegrucci et al., 2011). The enhanced efficiency of the *axolotl* oocyte is possibly related to the conserved pluripotency network that resides in its oocyte. Members of the TET protein family, TET2 and TET3 in specific, have been found to be expressed in *axolotl* oocytes (Unpublished data, Loose M.,

Johnson A.) Even though both Xenopus and *axolotl* oocytes can reprogram somatic cells (Allegrucci et al., 2011, Byrne et al., 2003), the enhanced success of the *axolotl* oocyte extract potentially reflects the intersection between chromatin remodeling and pluripotency networks in a manner that is exclusive to the *axolotl* oocyte. A large amphibian oocyte that contains the pluripotency network is a promising tool for the research of the pluripotency network and its interacting co-factors in the context of cell reprogramming. Amphibian oocytes, primarily from Xenopus, have been widely used to reprogram differentiated cells (Gurdon, 1968, Gurdon et al., 1958, Biddle et al., 2009, Briggs and King, 1952, Byrne et al., 2003). Their size, ease of obtaining, and manipulation has been advantageous for SCNT experiments with donor nuclei from mammalian somatic cells. However, Xenopus lacks core pluripotency factors in its genome. Would an oocyte in possession of core pluripotency factors have an advantage in SCNT reprogramming, is one of the questions that could be asked.

### 1.4.3 Aims and objectives

The oocyte possesses the capacity to reprogram chromatin as demonstrated in the case of paternal chromatin during fertilization. This strategic oocyte attribute has been taken advantage of for SCNT and the study of fate reversal towards pluripotency. The iPS, now a decade old, was a breakthrough in the field as it showed that the establishment of pluripotency was feasible by the overexpression of a limited number of factors, such as the OSKM tetrad.

When comparing the two methods, SCNT is the one involving more technical manipulations and ethical constrains. iPS on the other hand came to lift such barriers. While iPS is currently more fitted to clinical research applications, it pales in comparison with the efficiency of SCNT. The question that arises is how and whether the natural attributes of the oocyte can be translated to the iPS methodology.

One of the key events taking place during fertilization is demethylation. Cell fate reversal towards pluripotency is also marked by demethylation that subsequently leads to gene activation. While the leading theory for demethylation advocates for successive oxidation of the methyl mark, DNA repair pathways have been reported to act in demethylation specifically in the context of the oocyte. With demethylation playing an important role in chromatin reprogramming, delineating its mechanism would significantly influence cell reprogramming methodologies. Factors identified to participate in the oocyte demethylation mechanism could be employed in the iPS context to improve the current method.

While Xenopus oocytes have been widely used for SCNT research, *axolotl* oocytes were used to conduct the experiments presented in the following chapters. *Axolotl* oocytes have been reported to contain an ortholog of Nanog, as well as Oct4 and Sox2, thus recapitulating the core pluripotency network within the oocyte. This attribute offers the opportunity to study the pluripotency and chromatin reprogramming network, as well as their intersection in the oocyte. Factors that are likely to be involved in reprogramming might potentially be interacting with Nanog, or Nanog might

physically interact with and recruit factors required for pluripotency associated reprogramming to occur.

The hypothesis behind the research presented in the next chapters is that Nanog, a core pluripotency factor, interacts with oocyte factors crucial for the establishment of pluripotency. The physical interactors of Nanog in the *axolotl* oocyte potentially include pluripotency, as well as chromatin remodeling factors. Nanog is therefore likely to be a central node between the pluripotency and chromatin reprogramming networks. *Axolotl* Nanog can therefore reveal factors to be utilized in iPS improvement. Additionally, it is hypothesized that delineating the oocyte demethylation mechanism of the oocyte will also provide candidates for iPS trials, as efficient demethylation is crucial for reprogramming towards pluripotency.

Aims

- To identify *axolotl* Nanog interacting proteins and determine candidates for iPS assays.
- To delineate the demethylation mechanism operating in the *axolotl* oocyte and identify candidates for iPS assays.
- To improve the efficiency of the iPS methodology by overexpressing the factors identified and tested in the iPS assays.

Objectives

- To identify factors directly interacting with *axolotl* Nanog and potentially involved in the establishment of pluripotency, *axolotl* Nanog will be used in co-immunoprecipitation assays to purify its interactors.
- The purified proteins will be identified by mass spectrometry.
- After validating their interaction with *axolotl* Nanog, researching their role and determining their suitability, the selected candidates will be used in iPS trials.

- A regulatory region of Nanog will be employed to determine the demethylation mechanism employed by the oocyte for the establishment of pluripotency.
- The regulatory region will be assayed for proteins binding its different methylation profiles.
- The binding proteins will again be identified with mass spectrometry.
- After shortlisting candidates, the most suitable factors will be employed in iPS trials.

# Chapter 2: Materials and methods

## 2.1 Oocyte extract preparation

*Low speed extract (LS)*

Ovaries from adult female *axolotl*s were treated with collagenase (8 mg/ml Collagenase Type II, Sigma) for 2-3 h, agitating on rocker at room temperature. Free oocytes were removed every 15-20 min and washed in saline buffer (0.9% NaCl (w/v), Sigma). After the removal of necrotic eggs (usually discoloured and misshapen), the oocytes were washed thrice in extraction buffer (20 mM Hepes, Sigma, 100 mM KCl, Sigma, 5 mM MgCl2, Sigma) and Sigma protease inhibitors (β-Mercaptoethanol 2mM, Leupeptin 3µl/ml, Aprotinin 1µg/ml, Pepstatin A 1µg/ml). The oocytes were packed into centrifuge tubes and centrifuged at 9000rpm for 10 min at 4$^{o}$C. The centrifugation resulted in an upper layer of lipids, a middle layer of ooplasm and a bottom layer of egg yolk to be visible. The ooplasmic layer was obtained and centrifuged at 9000rpm for 10 min at 4$^{o}$C, to remove debris. Protease inhibitors (Aprotinin 1µg/ml, Sigma) and glycerol (5%) (Sigma) were added to the supernatant after the centrifugation, which was then aliquoted and frozen in cryotubes in liquid nitrogen at -80$^{o}$C.

*High speed extract (HS)*

In order to prepare the High Speed Supernatant, the ooplasmic layer from the Low Speed spin (LS), was ultracentrifuged at 35000rpm for 45 min at 4$^{o}$C (OPTIMA TLX ultracentrifuge). The resulting supernatant was obtained, aliquoted and stored with 5% glycerol (Sigma) at -80$^{o}$C.

## 2.2 Bradford Assay for protein quantification

The Bradford assay was performed in microplates with each sample dilution occupying one well. For the standard curve, BSA (Sigma) standards in the concentration range of 0.6, 0.3, 0.15, 0.075, 0.0375, 0.01875 mg/ml were prepared and assayed in triplicates on the same microplate as the samples of interest. Several dilutions (usually 1:5, 1:10, 1:20, 1:40) of the sample to be quantified were assayed in triplicates. The diluent used to prepare the BSA curve and sample dilutions was $H_2O$. 10 µl of BSA standard/sample were pipetted to the bottom of the well. Subsequently, 250 µl of 1xDye reagent (Bio-Rad Protein Assay Dye Reagent Concentrate) were added in each well ensuring that no bubbles were formed. The microplate was then incubated at room temperature for 10 min. Protein concentration was measured through absorbance at 595nm (Biochrom WPA Biowave II). In the case of sample readings falling outside the standard curve, the assay was repeated with more diluted or concentrated samples to ensure reliable protein quantification.

## 2.3 SDS-PAGE electrophoresis and gel staining

*1D-mini gels*

10% resolving and 5% stacking gels were prepared as follows.

10% resolving acrylamide gel was left to polymerize in stacking cassettes (1 mm) and was overlaid with water saturated butanol. After its polymerization, 5% stacking acrylamide gel was added on top of the resolving gel and left to polymerize as well. The samples to be loaded were prepared in Laemmli loading buffer (Bio Rad) and boiled for 5 min (95$^{o}$C) in a heating block. The gels were run in the Bio Rad Mini-Protean III electrophoresis chamber with 1L of 1xTGS buffer. The molecular weight markers used were either Bio Rad or New England Biolabs prestained protein markers. Gels run at 135 V until the bromophenol blue reached the bottom of the gel plates.

*Coomassie stain*

SDS-PAGE gels were washed in ultrapure water to remove impurities and then submerged to Imperial protein stain (Thermo Scientific), and left to stain overnight, agitating gently on rocker. Gels were washed ($H_2O$) several times for the background to be sufficiently destained and then scanned on the Bio Rad GS-800 densitometer.

*Silver stain (Yan et al., 2000)*

The procedure was carried out automatically using the Hoefer Processor Plus, according to the Yan-Wait silver staining protocol. Gels were fixed with 40% methanol (v/v) 10% acetic acid solution (v/v) twice for 15 min each. Sensitizer solution (0.2% STS (v/v), 68 g/L sodium acetate (w/v), 30% methanol (v/v)) was incubated for 30 min. Three washes with 18 M Ω water, lasting 5 min each followed, before incubating the gels in the impregnating solution (2.5 g/L silver nitrate) for 20 min. The gels were washed twice with 18 M Ω water for 1 min each. The developing solution (25 g/L Na2CO3, 0.4% HCHO (v/v), 0.27% STS (v/v)) was then applied for enough time to deliver sufficiently stained gels. In order to stop the developing reaction, 14.6 g/L EDTA was applied to the gel. The gel was then again washed with 18 M Ω water twice for 5 min each. The liquid volume used for each step in the case of mini gels, was 225 ml.

## 2.4 Western Blotting

SDS-PAGE gels run under the conditions stated above, the proteins were then transferred onto PVDF membrane (Bio Rad) through the application of electric current. Before the transfer, the membrane was submerged in ethanol for 1 min and then equilibrated in transfer buffer along with the components of the transfer (2xsponges, 2xblotting paper, SDS-PAGE gel). The protein transfer run at 70 V for approximately 2 h. After the transfer the membrane was washed with PBS (phosphate buffered saline)-Tween20 0.1% and then incubated for 1 h at room temperature in blocking solution (5% milk powder (w/v) in PBS-

Tween20 0.1% (v/v)). The membrane was then washed in PBS-Tween20 0.1% thrice for 10 min each, and incubated in primary antibody solution (AxNanog Ab diluted 1/2000 in blocking solution, concentration of primary AxNanog Ab stock 1.1 mg/ml) overnight at 4$^{o}$C agitating gently on rocker. After the primary antibody incubation the membrane was washed in PBS-Tween20 0.1% four times for 10 min each. The incubation in the secondary antibody (Horseradish Peroxidase labelled rabbit IgG diluted in PBS-Tween20 0.1% at 1/10000) lasted 1 h at room temperature and on the rocker. The membrane was again washed four times for 10 min each with PBS-Tween20 0.1%. The resulting signal was viewed with the Amersham ECL plus Western Blotting detection kit.

## 2.5 Protein Immunoprecipitation (IP)

In order to obtain Nanog protein complexes, AxNanog Ab was bound on Dynal magnetic beads (Invitrogen, protein G coated dynabeads). 50 µl of Dynal beads coated with protein G were washed twice with CPB buffer. AxNanog Ab (40 µg) and 300 µl of CPB buffer were added to the beads and incubated for 40 min at room temperature, on rocker. Supernatant was removed and beads were washed again twice with CPB (25mM Citric Acid, 25mM Disodium Phosphate) buffer (1 ml each time). To crosslink the antibody, the beads were washed twice with 1 ml of 0.2M triethanolamine-HCl (pH 8.2) (Sigma) and incubated for 30 min on rocker at room temperature in fresh DMP (Sigma) crosslinking solution (20 mM DMP in triethanolamine pH 8.2). Supernatant was removed and the beads were incubated in 50 mM Tris-HCl pH 7.5 for 15 minutes at room temperature on rocker. The beads were washed thrice in 1 ml PBS-Tween20 0.05%. For the elution of unbound Ab, the beads were washed in 30 µl of 0.1 M citrate buffer pH2-3 twice and then equilibrated immediately in 0.1 M sodium phosphate buffer pH 8. Two 1ml washes of PBS-Tween20 0.05% followed. The beads were incubated in 100 µl of LSS *axolotl* oocyte extract for 1 h at 4$^{o}$C on rocker. After the incubation the LSS supernatant was removed and stored at -80$^{o}$C for further analyses and the

complexes bound on the antibodies were eluted with 1 ml of 0.1 mM citrate buffer pH 2-3, the elution was done twice. In order to preserve the beads, they were washed immediately with 1 ml of 0.1 mM sodium phosphate buffer pH 8. The beads were then washed twice in 1 ml PBS-Tween20 0.05%, and stored at 4$^{o}$C in 100µl of PBS-Tween20 0.05%. Where antibody crosslinking was omitted, precipitated protein was eluted in 2xLaemmli buffer (54mg/ml DTT).

## 2.6 Antibody purification

Serum from rabbit immunized against *axolotl* Nanog protein was passed through a Glutathione-Sepharose column. *Axolotl* Nanog recombindant protein (GST tagged) was bound on the beads. The column was prepared in house. 7ml of rabbit serum were diluted to 10ml with PBS and circulated overnight through the column at 4$^{o}$C. Before eluting the antibody, the column was washed with TBS buffer (10mM Tris, 150mM NaCl, pH 7.5) to remove unspecifically bound material. Antibody was eluted with 0.1M Glycine, pH 2.5. Collected fractions were collected in eppendorfs containing 120ul of 1M Tris pH 8, in order to immediately equilibrate its pH. The collected fractions were of 1ml total volume (120ul Tris buffer and eluted Glycine). The protein concentration of the eluted fractions was quantified through OD measurement on a spectrophotometer set at 280nm (Biochrom WPA Biowave II). The fractions forming a protein concentration curve were pooled, dialyzed and concentrated to a final antibody stock.

It needs to be noted that the rabbit immunisation stage along with the purification of the first antibody batch was carried out by R. Alberio, C. Allegrucci, J. Dixon and A. Johnson. Following the discovery of the antibody inconsistency, all antibody purification was carried out by the author, C. Mola.

## 2.7 In vitro translation of mRNA

Rabbit reticulocyte cell kit for in vitro mRNA translation was purchased from Promega (L4960). *Axolotl* Nanog mRNA was supplied by Andrew Johnson's lab. The translation reaction was assembled as follows. 35 µl Reticulocyte lysate, 0.5 µl Amino acid mixture minus methionine (1mM), 0.5 µl Amino acid mixture minus leucine (1mM), 1µl *axolotl* Nanog mRNA (1µg/µl), 13 µl Nuclease free water (Total volume of reaction 50 µl).

The reaction was assembled in an Eppendorf tube, while the mRNA was denatured at 65$^{o}$C for 5mins prior to its addition. The reaction was subsequently incubated at 30$^{o}$C for 1.5hr. 10 µl of the translation reaction was then loaded on SDS-PAGE gel with 10 µl of 2xLaemmli buffer with DTT (54 mg/ml) and western blotted for the presence of the respective protein. The remaining reaction was frozen at -80$^{o}$C.

## 2.8 Recombinant protein overexpression and purification

*E. Coli transfection*

50 µl BL21 (DE3) pLysS competent cells (Promega, L1195) were thawed on ice and incubated with 50ng of PGEX-6 plasmid transformed for AxNanog for 30 min on ice. Subsequently, BL21 cells were heat-shocked in a water bath of 42$^{o}$C for 30seconds and then immediately incubated in ice for 2 min. 450 µl of SOC medium (0.5% Yeast Extract, 2% Tryptone, 10mM NaCl, 2.5mM KCl, 10mM MgCl2, 10mM MgSO4, 20mM Glucose) were added and the bacteria were incubated at 37$^{o}$C, for 1 hr, shaking. The bacteria were then centrifuged, the SOC medium was removed and the pellet was resuspended in 50 µl of fresh SOC medium. Cells were plated on agar plate with Ampicillin (50 µg/ml) and Chloramphenicol (34 µg/ml) and incubated at 37$^{o}$C overnight. The following day a single colony was picked to inoculate 50 ml of LB medium (10g Tryptone, 5g Yeast Extract, 10g NaCl in 1L deionized H20, pH 7) containing

Ampicillin (50 μg/ml) and Chloramphenicol (34 μg/ml) and was left to incubate overnight at 37$^{o}$C, shaking.

*Protein overexpression*

The 5ml colony culture was then used to inoculate 50ml of LB broth with the aforementioned antibiotics and left at 37$^{o}$C, overnight, shaking. This culture was used to inoculate 1L of LB medium which was left to grow until it reached an OD of 0.4-0.6 (at 600nm). Protein overexpression was induced with IPTG for 4hrs at a concentration of 1mM, at 37$^{o}$C. Bacteria were harvested through centrifugation and lysed through sonication (amplitude set at 6, cells sonicated for 6 rounds of a 30 second pulse followed by 30 second rest period, sonication was carried out on ice) in lysis buffer (20mM Tris-HCl pH 7.5, 100mM NaCl 5mM MgCl2, 0.5% NP-40, 1mM PMSF, 1mM DTT). After the sonication, lysozyme (SIGMA) was added to the lysed bacteria to break down cell membrane. The concentration of lysozyme was 0.1mg/ml and lysed bacteria were incubated in room temperature for 20 minutes. The sonicated bacteria were centrifuged at 6000rpm and the supernatant was incubated overnight, at 4$^{o}$C, rotating with Glutathione-Sepharose beads for the protein of interest to be purified.

*Inclusion body purification*

For the inclusion body purification, lysed *E. Coli* were centrifuged at 6000rpm to separate the soluble and insoluble fractions. The soluble fraction is the aquatic top layer which contains soluble and therefore natively folded protein. Insoluble is the pellet fraction that forms containing insoluble cell material and proteins. The insoluble fraction was salvaged and washed with PBS containing 0.1mM PMSF to remove traces of the soluble fraction. The pellet was subsequently resuspended in lysis buffer (20mM Tris-HCl pH 7.5, 100mM NaCl 5mM MgCl2, 0.5% NP-40, 1mM PMSF, 1mM DTT) by vortexing and again centrifuged at 6000rpm to further clean the pellet. At this stage the pellet can be frozen for preservation if desired.

*Urea solubilisation and refolding by dialysis of inclusion body protein*

Urea at a concentration gradient of 2M-4M-6M-8M was used to solubilise the inclusion body pellet. The urea washes were carried out sequentially so that the lower concentration urea buffers would further purify the inclusion bodies and the highest concentration would solubilise the inclusion body protein.

The inclusion body pellet was incubated in 2ml of each urea wash at room temperature for 15mins, mixing, while the pellet was crashed with tissue grinder in the urea wash prior to each incubation. After the incubation, the inclusion bodies were centrifuged at 15000rpm, for 20mins. The supernatant was retained for analysis and the pellet was passed on to the next urea wash, until full solubilisation. The fully solubilised urea wash (8M) was step-wise dialysed sequentially in 3L of 6M-4M-2M-PBS of urea to promote protein re-folding. The solubilised fractions (supernatant of washes) and the final product of the dialysis (re-folded protein) were analysed on SDS-PAGE gel.

*Detergent solubilisation and refolding of inclusion bodies*

Inclusion bodies were incubated with buffer B (50mM Tris, 300mM NaCl, 10mM b-mercaproethanol) for 30mins after the pellet was crushed with tissue grinder. The inclusion bodies were then centrifuged at 15000rpm, for 20mins and the supernatant was removed. The resulting pellet was treated with buffer A (50mM Tris, 300mM NaCl, 10mM b-mercaptoethanol, 10% sarkosyl) and crusher in tissue grinder. The solution was subsequently centrifuged at 15000rpm, for 20mins. The supernatant was salvaged and its sarkosyl concentration was diluted to 2% with buffer B containing 4% Triton X100 and 40mM of CHAPS. The solubilised protein as analysed on SDS-PAGE gel.

*Preparation of AxNanog antibody purification column with solubilised inclusion body protein*

0.5ml Glutathione sepharose beads 4B (GE Healthcare) were incubated with the 8M solubilised inclusion body protein for 1hr at room temperature, mixing.

The same volume of GST beads was also incubated in the detergent solubilised inclusion bodies at the same conditions as above. The GST beads were retrieved with centrifugation at 3000rpm for 5mins.
In both cases the beads were washed with PBS 4 times and centrifuged at the same conditions to remove carry-over binding. The beads were subsequently analysed on SDS-PAGE gel.

## 2.9 Immunocytochemistry

Cells were cultured to 80% confluency on gelatinized coverslips. Medium was removed and cells were washed twice with washing buffer (1% BSA – 0.1% Tween20 in PBS). Cell fixation commenced in 2.5% PFA (Paraformaldehyde), for 15 minutes at room temperature. Cells were subsequently washed again twice with washing buffer. Cells were permeabilized with 0.1% Triton100 in PBS for 15 minutes at room temperature. After washing buffer applications to remove Triton100 residues, blocking solution (5% BSA in PBS) was applied for 1 hour at room temperature. Primary antibody (AxNanog 0.7mg/ml and IgG 1mg/ml) used at a concentration of 1/500 in blocking solution and left to incubate overnight at 4 C. Primary antibody removed with three washing buffer applications and secondary antibody (donkey anti rabbit IgG, Abcam) was incubated at a concentration of 1/200 diluted in washing buffer for 1 hour at room temperature. Excessive secondary antibody was removed with three washing buffer applications and DAPI was mounted subsequently. Cell slides were stored at $4^{o}$C until microscope observation. The slides were observed under a fluorescence microscope (Leica DMR, Heidelberg, Germany) with a digital camera (Hammamatsu, ORCA-ER, Japan). Simple PCI was the software used (Compix Image System, USA).

## 2.10 End-Point PCR mNanog promoter amplification

Primers

F: ACTCAGATCCCCCACTTGAC
R: CTACCCACCCCCTATTCTCC

Both forward and reverse primers were biotinylated on the 5' end (purchased from Eurofins). Primers were resuspended in TE buffer (10mM Tris pH 8.0, 1mM EDTA) at final concentrations of 100pmol/µl.

*Non-modified cytosine mNanog-promoter amplification*

The ReadyMix (SIGMA-ALDRICH P4600) used contained dNTPs and therefore there was no need for additional dNTP to be supplied. The amplification was carried out in 40 cycles.

| Reagent | Volume (µl) | Final concentration |
|---|---|---|
| 2xReadyMix Taq PCR | 12.5 | 1x |
| Primer mix | 1 | 0.16µM |
| Template DNA | 0,12 | |
| $H_2O$ | 11.38 | |
| Total volume | 25 | |

**Table 2.1. End-Point PCR reaction mix reagents and respective volumes for biotinylated mNanog promoter amplification.**

| Temperature (C) | Duration | PCR step |
|---|---|---|
| 94 | 1 min | Initial denaturation |
| 94 | 1 min | Denature |
| 58 | 1 min | Anneal |
| 72 | 1 min | Extension |
| 72 | 2 min | Final extension |
| 10 | | End |

**Table 2.2. Cycling conditions for the amplification of non-modified cytosine mNanog promoter over 40 cycles.**

*5mC and 5hmC-mNanog promoter amplification*

5-Methylcytosine dNTP mix (D1030) and 5-Hydroxymethylcytosine dNTP mix (D1040) were purchased from Zymo Research. ZymoTaq DNA polymerase (E2001) from Zymo Research was used to amplify 5mC and 5hmC-mNanog promoter. The 5mC/5hmC-mNanog promoter was amplified over 40 PCR cycles.

| Reagent | Volume (µl) | Final concentration |
|---|---|---|
| 2x Reaction buffer | 12.5 | 1x |
| 5mC or 5hmC dNTPs mix | 2.5 | 1mM |
| Primer mix | 1 | 0.16µM |
| Zymo Taq polymerase | 0.2 | 1 unit/25µl |
| Template DNA | 0.12 | |
| $H_2O$ | 8.68 | |
| Total volume | 25 | |

**Table 2.3. End-Point PCR reaction mix reagents and respective volumes for biotinylated 5mC/5hmC-mNanog promoter amplification.**

| Temperature (C) | Duration | PCR step |
|---|---|---|
| 95 | 10 mins | Initial denaturation |
| 95 | 30 sec | Denaturation |
| 58 | 40 sec | Annealing |
| 72 | 1 min | Extension |
| 72 | 7 min | Final extension |
| 4 | | End |

**Table 2.4. Cycling conditions for the amplification of 5mC/5hmC-mNanog promoter over 40 cycles.**

*5fC-mNanog promoter amplification*

5fC-dNTP was purchased from TriLink (N2064) and the dNTPs mix containing 5fC was prepared at home to a concentration of 10mM. The 5fC-mNanog promoter amplification was carried out with DreamTaq DNA polymerase from Thermo Scientific (EP0702). Amplification carried out over 40 PCR cycles.

| Reagent | Volume (µl) | Final concentration |
|---|---|---|
| 10x DreamTaq buffer | 2.5 | 1x |
| 5fC dNTPs mix | 1 | 1mM |
| Primer mix | 1.25 | 0.2µM |
| Dream Taq polymerase | 0.125 | 0.625 unit/25µl |
| Template DNA | 0.12 | |
| $H_2O$ | 20.005 | |
| Total volume | 25 | |

**Table 2.5. End-Point PCR reaction mix reagents and respective volumes for biotinylated 5fC-mNanog promoter amplification.**

| Temperature (C) | Duration | PCR step |
|---|---|---|
| 95 | 3 mins | Initial denaturation |
| 95 | 30 sec | Denaturation |
| 56 | 30 sec | Annealing |
| 72 | 1 min | Extension |
| 72 | 15 min | Final extension |
| 4 | | End |

**Table 2.6. Cycling conditions for the amplification of 5fC-mNanog promoter over 40 cycles.**

## 2.11 Agarose gels and PCR product purification

1% agarose gels were prepared with 1xTAE buffer and gel electrophoresis commenced in 1xTAE buffer. Ethidium bromide at a concentration of 1µg/ml was used to visualise nucleic acid. 1kb DNA markers (NEB) for bp size estimation, while samples were mixed with loading buffer prior to gel loading and electrophoresis. The gel was run to assess for efficient amplification and specificity.

The amplified biotinylated promoters (C, 5mC, 5hmC and 5fC) were purified from the reaction after the completion of PCR cycling with the kit and protocol supplied with the Nucleospin Gel and PCR clean-up kit, by Mackerey-Nagel (740609). After purification, the amplified promoter was stored at -20$^{o}$C.

## 2.11 DNA pull down

Magnetic Dynabeads M-280 Streptavidin beads, from Invitrogen (11205, 11206) were used as the resin to receive amplified biotinylated mNanog-promoter. Beads were prepared as the product manual suggested prior to DNA coupling.

200µl of M-280 beads were incubated with 7µg of mNanog promoter. Where 60µl of beads were used they were incubated with 2.1µg mNanog promoter. The incubation was carried out in PBS buffer, overnight at 4$^{o}$C, mixing on rotator. Unbound DNA was removed and the beads were washed thrice in PBT buffer (0.1% Triton x100, in PBS) and twice in D-T buffer (0.2mM EDTA, 20% glycerol, 20mM Hepes pH 7.9, 0.1M KCl, 1mM DTT, 1mM PMSF added fresh, 0.1% Triton x100). 2mg of *axolotl* oocyte lysate was incubated with the bead-DNA conjugate in incubation buffer (0.05mM EDTA, 5% glycerol, 5mM Hepes pH 7.9, 150mM KCl, 1mM DTT, 1mM PMSF added fresh, 0.025% Triton x100, in PBS) for 2hrs at 4 C. Unbound lysate was removed and the beads were washed 6 times in D-T buffer and once in PBS as described previously.

For SDS-PAGE gel analysis the beads were boiled for 5min at 95$^{o}$C in 2xLaemmli loading buffer with DTT (54mg/ml). The eluate was then loaded to the acrylamide gel after separating the beads with a magnet.

For mass spectrometry analysis without SDS-PAGE resolution, the samples were not stored, but were processed immediately after the final washes.

## 2.12 Mass spectrometry

*In gel digestion*

Excised gel pieces were prepared for mass spectrometry with the Proteoms Works Mass Prep Station (WATERS). Gels bands were de-stained with 15mM potassium ferricyanate and 50mM sodium thiosulphate (Silver staining) or with ammonium bicarbonate and 50% methanol (Coomassie staining). The gel pieces were then dehydrated in ammonium bicarbonate and 50% acetonitrile. The proteins were subsequently reduced in 10mM DTT in 100mM ammonium bicarbonate for 20mins at 60 C. The proteins were then alkylated with 55mM iodoacetamide (IAA) in 100mM ammonium bicarbonate, for 20mins. Gel pieces were finally dehydrated with acetonitrile. Digestion with Trypsin Gold (Promega, V5280) followed at 4ng/µl in 25mM ammonium bicarbonate, for 4hrs at 40 C (samples incubated in the fridge for 30mins prior to the 4hr incubation). The samples were made up to 0.1% formic acid.

*Rapigest digestion*

After the DNA pull down the magnetic beads were retrieved and washes six times with D-T buffer and once with 50mM ammonium bicarbonate. Precipitated protein were eluted thrice in 0.2% Rapigest (Waters) dissolved in 50mM AMBIC (ammonium bicarbonate) and 5mM DTT after heating the beads at 60$^{o}$C for 20mins. Beads were separated from the eluate with a magnet and all eluates were pooled together and allowed to cool. Protein were subsequently alkylated with 15mM iodoacetamide (IAA) and incubated for 30 minutes. The proteins were then digested with Trypsin Gold (Promega) at a concentration of 4ng/µl for 4hrs at 40$^{o}$C. 0.5% formic acid was added to stop the digestion and acidify the detergent (Rapigest) and incubated at 37$^{o}$C for 45mins. The protein were subsequently centrifuged at 13000rpm, for 10mins and the supernatant containing the tryptic peptides was moved to a fresh Eppendorf tube. C18 (ThermoScientific) peptide clean up followed as described by the product manual. Finally tryptic peptides were eluted in 70%

acetonitrile (ACN) and concentrated until dry in the speed vacuum. At this stage samples were frozen at -80$^{o}$C until mass spectrometry analysis, when each sample was resuspended in 0.1% FA.

*On bead digestion protocol*

After the DNA pull down the magnetic beads were retrieved and washes six times with D-T buffer. Subsequently the precipitated proteins were incubated in Buffer 1 (2M urea, 50mM Tris-HCl pH 7.5, 4ng/ul Trypsin) and digested for 1hr at 40$^{o}$C. The beads were then separated from the supernatant with a magnet the supernatant was transferred into a fresh Eppendorf tube. Beads were then washed twice in buffer 2 (2M urea, 50mM Tris-HCl pH 7.5, 1mM DTT) and both washes were pooled with the previous supernatant and left to digest for 4hrs at 40$^{o}$C. Alkylation followed with 15mM IAA for 30mins. 0.5% Formic Acid (FA) was added to stop trypsinization. The tryptic peptides were transferred to a clean Eppendorf tube. C18 (ThermoScientific) peptide clean up followed as described by the product manual. Finally tryptic peptides were eluted in 70% acetonitrile (ACN) and concentrated until dry in the speed vacuum. At this stage samples were frozen at -80$^{o}$C until mass spectrometry analysis, when each sample was resuspended in 0.1% FA.

*C18 Clean up*

C18 clean-up was carried out to remove salt and other contaminant prior to mass spectrometry and therefore enhance protein identification. Fast Flow C18 Tips were used (SPE tip HyperSep SpinTip C18 10 to 200µL Thermo Scientific 10404603), one for each replicate. The C18 tips were activated by pipetting 200 µl 50% ACN (Acetonitrile) though the resin. After activation the resin was equilibrated with 200 µl of 5% ACN and 0.1% Formic Acid. The sample was then pipetted through the resin for peptide binding. Flow through sample was pipetted through the resin once more to ensure maximum peptide binding on the resin. To wash the resin 200 µl of 5% ACN and 0.1% Formic Acid was pipetted through the resin thrice. Bound peptides were

eluted with 50 µl of 70% ACN. The eluate was pipetted through the resin twice to ensure maximum elution of tryptic peptides.

*Mass spectrometry injection*

Samples were directly injected onto the analytical column (YMC Triart C18, 15 cm, 3 um particle size, 300um internal diameter) using an Eksigent 400 nanoLC system and autosampler at 5 uL/min flow rate. They were injected onto a Sciex Tripletof 6600 mass spectrometer. IDA (Information dependent acquisition): HPLC gradient length 120 min. Time 0 min 98% A; 110 min 60% A; 115 min 20% A; 117 min 20% A; 118 min 98 % A; 120 min 98% A). Mobile phase A 2% Acetonitrile, 5 % DMSO, 0.1% formic acid; Mobile Phase B same but 2% water instead. Autosampler kept at 8$^{o}$C, column at 50$^{o}$C. Samples analysed in IDA mode on Sciex TripleTof 6600. IDA was carried out over the m/z range from 350-1800. Charge state 2+ to 4+ with dynamic exclusion of fragmented precursors for 10 seconds using rolling collision energy (Sciex parameters). Precursor ions were analysed in an MS survey scan for 250 milliseconds, then top 30 ions fragmented, each with an accumulation time of 60 milliseconds giving a total cycle time of 2.1 seconds (m/z 100-2000). IDA files were searched in either Protein Pilot 5.0 or PEAKS studio against the custom database comprised of the *axolotl* transcriptome (Matt Loose). Label-free quantification was carried out on the PEAKS-Q (Waters). For the generation of the heatmap proteins with at least one unique peptide were listed. The fold change threshold was set at 2 and the protein significance threshold was set at 15. The detailed search parameters can be seen in the appendix (Table 14-16).

# Chapter 3: Protein network study and material improvement

## 3.1 Introduction

### 3.1.1 Immunoprecipitation in the field of pluripotency and reprogramming

Affinity purification has been widely used to investigate the pluripotency network in ES cells. The pioneering study of Wang et al. (Wang et al., 2006) identified a network of 17 factors directly or indirectly associated with Nanog in murine ES cells. The team exploited the technique of in vivo biotinylation and subsequent purification of the biotinylated molecules via streptavidin coated beads. Among the factors identified as interactors of Nanog were Oct4, Sall1, Sall4 and Dax1 (Wang et al., 2006). Liang et al. showed that Nanog directly interacts with proteins implicated in repressive pathways, such as members of the NuRD and Sin3A-HDAC complexes (Liang et al., 2008). The team also identified members of the nucleosome remodelling SWI/SNF complex to co-immunoprecipitate with Nanog (Liang et al., 2008). Further advances extended the list of Nanog interacting molecules to 27 high confidence candidates, with TET1 reported to enhance the efficiency of reprogramming in synergy with Nanog (Costa et al., 2013). Gagliardi et al. extended the list even further by identifying a plethora of transcription factors, members of the NuRD, p400, Polycomb, MLL chromatin remodelling complexes (Gagliardi et al., 2013). Oct4 affinity purification methods have also generated large numbers of candidate Oct4 interactors with Sall4, Sox2, Nanog, Dax1 as well as subunits of the NuRD complex amongst them (Pardo et al., 2010). Novel Oct4 interactors were identified by Ding et al. that included members of the SWI/SNF complex (Ding et al., 2012). Sox2 interactors were identified in murine ES cells engineered for inducible

expression of the OSKM factors, resulting in a list of over 70 proteins associated with Sox2 (Gao et al., 2012).

### 3.1.2 Affinity purification: a tool to delineate protein networks and interactions

Affinity purification is a method that exploits the affinity of a target molecule for another factor or resin. A resin is used as a binding surface for and to pull-down the target molecule via centrifugation or magnetic force. Various methods have been developed relying on the principle of affinity purification. The target protein can be conjugated to a specific tag which can serve as the connector between the protein and a resin with affinity for the tag. Technical manipulation is a prerequisite for expressing a tagged protein; however affinity purification can be attempted with native proteins as well. Immunoprecipitation is an alternative to affinity purification, exploiting the specificity of an antibody against its antigen. The antigen can be an entire protein, a peptide or even a chemical compound. An antibody designed to recognise an antigen in its native three-dimensional conformation is capable to specifically bind that antigen in crude cell/tissue lysates. Thus, targeting and isolating specific proteins becomes tangible by the use of an antibody. While the variable regions of an antibody are responsible for the recognition and binding of the antigen, the constant region of the heavy chain is biochemically conjugated to a resin. Affinity purification was mainly applicable for the purification of target molecules from complex lysates. However the principle has evolved and is employed for the study of protein interactions. In tissue/cell lysates native proteins interact with their co-factors and therefore affinity purification attempts can capture molecules physically interacting with a target protein. When an antibody is employed to investigate for co-factors the term co-immunoprecipitation is used. Affinity purification has been widely used to study protein networks with the target protein as their core

component, or to determine whether two (or more) proteins physically interact.

## 3.2 Aims and objectives

Previously, a Nanog ortholog capable of regulating pluripotency in murine ESC was isolated in *axolotl* (Dixon et al., 2010). Transcriptomic analysis also revealed the presence of Oct4 and Sox2 orthologs in *axolotl* oocytes (unpublished data, Johnson A, Loose M).

Interactome studies in ESC have revealed that the pluripotency network is interlinked with chromatin remodelling and epigenetic factors, as well as molecules implicated in transcriptional regulation. Based on these prior interactome findings, the aims of the present chapter are:

- To investigate the physical interactors of axolotl Nanog in the oocyte.
- To identify axolotl Nanog interacting molecules implicated in the pluripotency and/or chromatin reprogramming network.
- To shortlist candidates for biological assays.

To address the issue, axolotl Nanog antibody will be employed in co-immunoprecipitation experiments. Along with the axolotl Nanog protein that will be purified, Nanog interacting factors will also precipitate during the experiment. Protein identification will commence through mass spectrometry. After testing the antibody previously generated against recombinant and native *axolotl* protein to confirm its efficiency, immunoprecipitation and co-immunoprecipitation experiments were carried out.

## 3.3 Materials and methods

Western blot, SDS-PAGE and Protein immunoprecipitation were carried out as described in Chapter 2, Sections 2.4, 2.3 and 2.5, respectively. E. Coli protein overexpression and inclusion body purification and solubilisation were also performed according to the protocol in Chapter 2, Section 2.8. Changes in reagent amounts and slight procedure alterations are stated. For the mass spectrometry run in 3.4.2 the samples were prepared with the in gel digestion method and sent to Dundee Cell Products. The identification searches were run against the NCBI database with a chordata filter, followed by a specific search against the Uniprot axolotl database. Lastly, a local peptide search against the provided axolotl Nanog sequence was conducted. All searches were run through MASCOT. For the mass spectrometry run in 3.4.3 and 3.4.4 the samples were prepared according to the in gel digestion method and analysed by S. Liddell in the Proteomics Laboratory (School of Biosciences, University of Nottingham). After tryptic digestion the resulting peptides were delivered via nanoLC- to the Q-ToF2 for ESI-MSMS analysis. An automated experiment (DDA, Data Dependent Acquisition) was run where selected peptides automatically enter MSMS for fragmentation. The data was searched against the custom database comprised of AXOLOTL EST sequences using MS/MSIONS search on the local MASCOT server using standard default settings (Pept Tol +-1.2Da, MS/MS Tol 0.6 Da, Pept Charge 2+). For the immunocytochemistry experiment, Nanog null pre-iPS cells overexpressing axolotl Nanog were used. The cells were provided by Darren Crowley and Andrew Johnson.

## 3.4 Results

### 3.4.1 Testing the affinity of the antibody against native and recombinant protein

The *axolotl* Nanog specific polyclonal antibody was generated previously by laboratory members via the immunization of a rabbit with recombinant *axolotl* Nanog, overexpressed in *E. Coli*. The recombinant Nanog contained the GST tag which was exploited for its subsequent purification with Glutathione Sepharose resin. In order for the recombinant protein to be used for immunization purposes the GST tag was cleaved off. After sufficient exposure to the antigen, the animal's serum was isolated. The polyclonal antibody was purified from the serum via affinity chromatography, where the overexpressed *axolotl* Nanog protein was conjugated to Glutathione Sepharose resin through the GST-tag.

The purified antibody was tested against recombinant *axolotl* Nanog and *axolotl* oocyte lysate. To confirm the observations of the previous lab members on the affinity of the antibody against recombinant and native protein, SDS-PAGE Western blot experiments were conducted, prior to further utilization of the antibody in different experiments. The results can be seen in Fig. 3.1. The recombinant *axolotl* Nanog tested was expressed and purified from *E. Coli*, as described above. The *axolotl* oocyte lysate was generated from GV (Germinal Vesicle) stage oocytes that were lysed in mild buffer with mechanical methods. The lysate was fractionated to Low Speed (LS) and High Speed (HS), through ultra-centrifugation. Amphibian oocyte lysate tends to be dense as it contains not only proteins, but organelles, lipids and other cellular components. Ultra-centrifugation (75K RPM) is employed to clear lipids, organelles, and membranes, but retains the oocyte protein content. The LS fraction contains dense particles like membranes and organelles as well as the lipid phase of the lysate. On the contrary, the HS fraction contains the clarified

oocyte proteome. Both LS and HS fractions were assayed for the presence of *axolotl* Nanog in the SDS-PAGE Western blots.

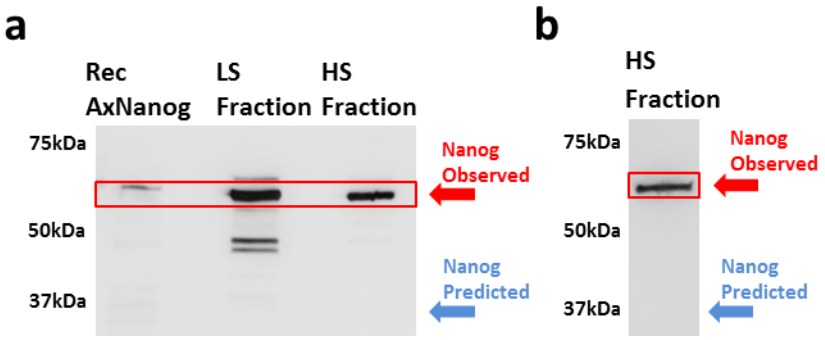



**Figure 3.1. Testing polyclonal anti-*axolotl* Nanog antibody against recombinant and native *axolotl* Nanog. a) Western blot for Recombinant as well as LS and HS oocyte fraction AxNanog. The predicted and observed molecular weight for AxNanog is shown with blue and red arrows, respectively. b) Western blot of HS *axolotl* oocyte fraction after antibody titration to decrease background binding. In both a and b figures, molecular weight markers appear on the left side of the image.**

The results obtained showed a predominant band at 65kDa in all lanes. However, higher and lower molecular weight bands were also observed specifically in the LS *axolotl* oocyte fraction. It was assumed that the additional molecular weight bands represented either unspecific binding or, in the case of the lower molecular weight bands, degradation products. The additional bands were more prominent in the LS fraction, a fact that lead to the assumption that proteins that have entered the degradation pathway and are hence bound by large protein complexes, such as the proteasome, sediment with the larger molecules and particles during ultracentrifugation. However the prevalent 65kDa protein band is also enriched in the case of the LS oocyte extract fraction, indicating that amounts of the intact native protein are lost with ultracentrifugation. The protein lost could be marked for degradation and involved in proteasome complexes that sediment during ultracentrifugation.

Interestingly, the observed molecular weight for *axolotl* Nanog differs from the predicted. According to calculations made by the UniProt tool, ExPasy and based on the protein sequence of *axolotl* Nanog, its molecular weight was

predicted to be 34kDa. However, neither the recombinant nor what appears to be native Nanog, were of that size. In order to validate the Western blot

results and rule out the possibility of the signal being unspecific, c-Myc-tagged *axolotl* Nanog mRNA was in vitro translated with reticulocyte lysates, and subsequently immunoblotted for *axolotl* Nanog as seen in Fig. 3.2.
In this case, a band at the region of 40kDa responded positively to both the anti-*axolotl* Nanog antibody and the anti-Myc antibody. The recombinant and native *axolotl* Nanog protein appeared at 65 kDa.

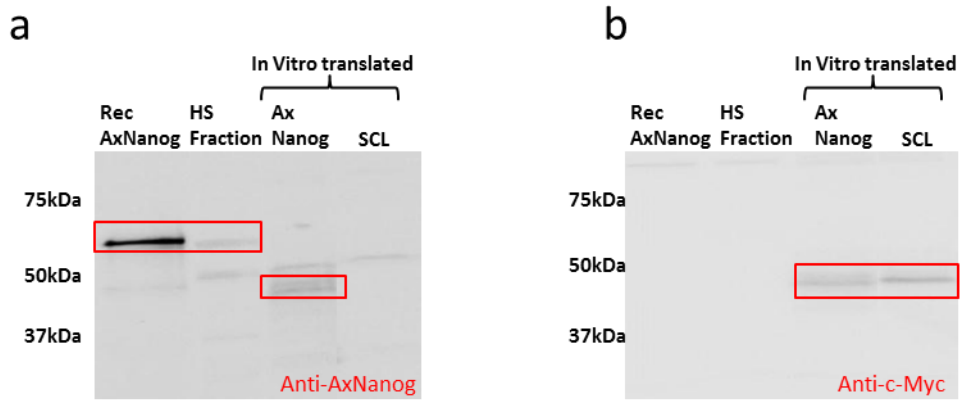



**Figure 3.2. In vitro translation of *axolotl* Nanog mRNA. a) Western blot for *axolotl* Nanog. Lanes loaded with recombinant Nanog protein, HS *axolotl* oocyte lysate, in vitro translated c-Myc tagged AxNanog and SCL (SCL) as a negative control of translation. b) Western blot for c-Myc, the tag of the in vitro translated mRNA that was used to assess false positive bands.**

### 3.4.2 Assessing the efficiency of immunoprecipitation

Prior to attempting the identification of AxNanog co-factors in the *axolotl* oocyte, it was crucial that the efficiency of the immunoprecipitation was verified. In order to do so AxNanog immunoprecipitation was attempted. The polyclonal antibody raised against the protein was conjugated to magnetic beads. The antibody-bead conjugate was incubated in *axolotl* oocyte lysate to capture the protein and physically interacting molecules. The beads were subsequently retrieved via a magnet and washed to remove unspecific

binders. The proteins specifically bound by the antibody were eluted and the eluates were analysed on SDS PAGE.

A way to assess the success of an immunoprecipitation attempt is to measure the depletion of the protein of interest from the lysate used in the experiment. In the present case depletion of native Nanog was observed in the immunoprecipitated lysate, as seen in Fig. 3.3a. More specifically, in order to validate the efficiency of the IP, sequential amounts of antibody were used to immunoprecipitate *axolotl* Oocyte lysate samples of the same protein concentration. As seen on the Western blot in Fig. 3.3a, the higher the amount of antibody the more significant the depletion of AxNanog in the lysate.

Of course the presence of the antigen itself in the IP eluate lane is itself a robust verification of the experiment's success. In the present case, identifying Nanog in the IP eluate was challenging as the heavy IgG chain created a smear covering up the molecular weight region of Nanog, obfuscating any other bands of similar molecular weight (Fig. 3.3b). The IgG bands obscure not only visualisation of the antigen band by protein staining, but in Western blotting as well. In fact the effect IgG bands have on Western blots is more significant. The IgG chains will occupy two molecular weight regions, one round the 50-60kDa and another at the 20-30kDa. The secondary antibody used in the Western blot will react not only with the primary antibody detecting the protein of interest, but will also bind to the IgG chains that were resolved in the SDS-PAGE gel. Upon the reaction of the HRP substrate with the secondary antibody a diffused signal is produced because of the larger quantities of antibody binding to the membrane. Occasionally the signal from the 20-30kDa band and the 50-60kDa regions will merge causing greater obfuscation.

Because the aforementioned issue is trivial in IP experiments, light chain specific antibodies have been generated, that only interact with the light chain in its native conformation. Thus, only the protein of interest that is bound by the primary antibody is detected and the 50-60kDa heavy IgG band does not interfere with its visualisation in the Western blot. I decided to try

the same method in order to verify the presence of AxNanog in the IP eluate lane, however the results obtained were not permissive of such a verification. When the Western blotting method proved to be inconclusive I decided to employ mass spectrometry in order to determine the existence of Nanog beneath the smear created by the heavy IgG chain. Specifically, in order to enhance the possibility of identifying Nanog beneath the smear I used sequential amounts of antibody (3μg, 5μg and 10μg) and carried out three individual IP experiments. As seen in Fig. 3.3b, more antibody can deplete more AxNanog. Of course more antibody also comes with the repercussion of a larger smear, however the presence of additional AxNanog might be beneficial for its identification. The eluates were resolved on SDS-PAGE gel which was subsequently Coomassie stained to visualise the bands. The heavy IgG chain smear bands that were potentially concealing AxNanog were cut from all three SDS-PAGE gels, put in individual Eppendorf tubes with $H_2O$ to prevent gel shrinking and forwarded to the mass spectrometry company. The results from the mass spectrometry analysis showed high amounts of rabbit IgG, as expected, but no traces of AxNanog were found in all three replicates, despite the stringency of the analysis. These results created the grounds to question the affinity of the antibody against native *axolotl* Nanog.

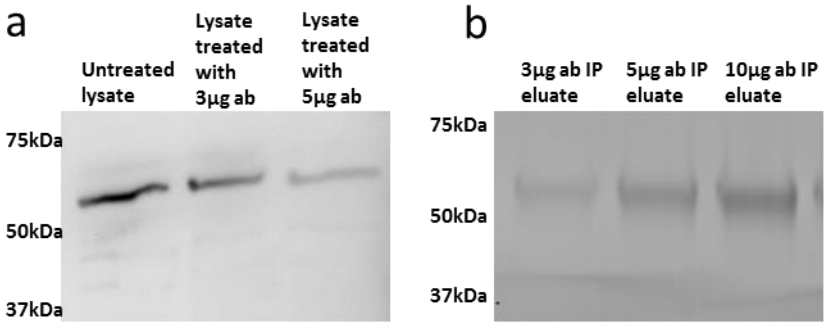



**Figure 3.3. a) Western blot assaying the depletion rate of AxNanog on oocyte lysate after IP experiments with 3μg and 5μg of anti-AxNanog antibody. b) Coomassie stained acrylamide gel, lanes loaded with the eluates of three individual IPs carried out with 3μg, 5μg and 10μg of anti-AxNanog ab in *axolotl* oocyte lysate. The bands seen mainly represent the heavy IgG chain which might be overlapping with native Nanog.**

### 3.4.3 Mass spectrometry analysis of recombinant *axolotl* Nanog

To investigate the quality of the overexpressed recombinant *axolotl* Nanog, the 65kDa band corresponding to the recombinant protein was analysed with mass spectrometry (Fig. 3.4). The results, surprisingly, showed that the 65kDa band, believed to be Nanog, was in fact a bacterial chaperone, GroEL. Along with the 65kDa band, fainter bands that were showing on the SDS-PAGE gel of the recombinant protein solution were also assayed with mass spectrometry. In total seven bands were processed ranging from 65 to just below 20kDa. An ID was obtained for all the assessed bands and reagents used in the overexpression of the protein were identified, regardless of their scarcity. For example GST, the tag attached to Nanog to be exploited in the protein purification step, as well as thrombin, the protease that was used to cleave the GST tag off the recombinant protein, were identified in their predicted molecular weight regions. Nanog, as mentioned earlier, was not identified in the 65kDa band, as we would expect, judging by the Western blot experiments. However, Nanog was identified in the band dissected from below the 20kDa region. The peptides identified to match the *axolotl* Nanog amino-acid sequence, originated from the middle and the N-terminus of the protein, suggesting that Nanog was degraded in the solution. The presence of *axolotl* Nanog below 20kDa is by itself an indication that the protein is degraded, as it contrasts both the observed and the predicted molecular weight of *axolotl* Nanog, at 34kDa.

The results demonstrate that the recombinant protein solution is not purely Nanog, but a mixture of bacterial products, with GroEL prevailing in quantity as shown by SDS-PAGE (Fig. 3.4). The degradation of Nanog indicates its sensitivity and has probably occurred after its purification. It is obvious that the antibody reacts with GroEL and to further prove that the signal seen is not Nanog relevant, wild type *E. Coli* cells were cultured, lysed and Western blotted using the same antibody (Fig. 3.4b). The results clearly showed that it

is the bacterial GroEL that generates the signal seen at 65kDa. The 65kDa molecule that generates the Western blot signal in the case of the *axolotl* oocyte lysate is the equivalent of the bacterial GroEL in higher animals, Hsp60 which also weights 65kDa. GroEL and Hsp60 are both protein chaperones and share similarities in terms of function and amino acid sequence.

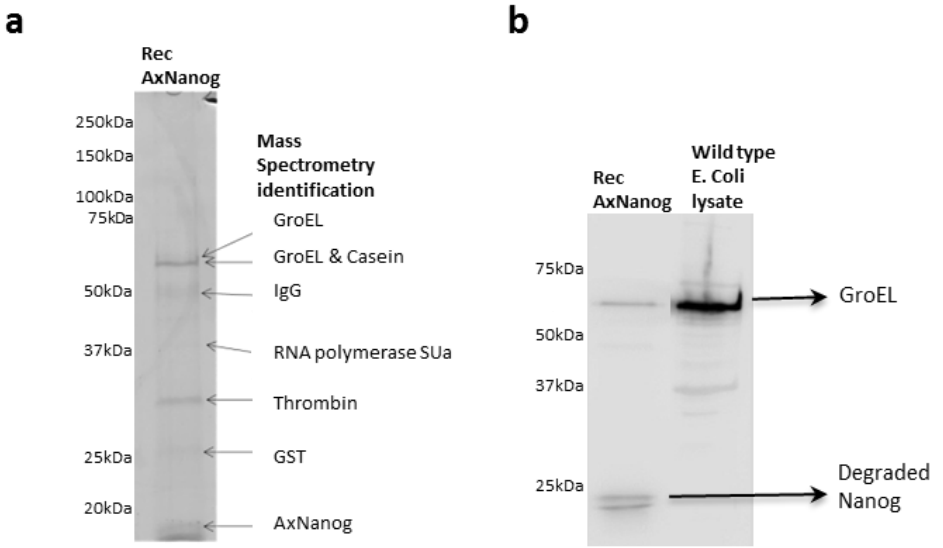



**Figure 3.4. Recombinant *axolotl* Nanog contaminated by bacterial GroEL. a) Bands of recombinant AxNanog protein analysed with mass spectrometry and their respective identities b) Western blot of recombinant AxNanog and non-transformed (wild type) *E. Coli* lysate. The results demonstrate that the antibody reacts with wild type *E. Coli* bacteria, containing no foreign plasmids. The EC signal clearly coincides with the 65kDa band observed in RN and lysate lanes**

### 3.4.4 Purification of the AxNanog antibody from GroEL immunoglobulins

Large quantities of serum were retrieved from the AxNanog immunized rabbit and only a portion of it was purified to generate AxNanog antibody. Having an excess of serum containing the specific antibody and at the same time having shown that the AxNanog antibody is capable of detecting in vitro translated AxNanog (Fig. 3.2) and degraded recombinant AxNanog (Fig. 3.4b), I reasoned to attempt a new AxNanog antibody purification from the remaining serum, using the classical method of affinity chromatography. In order to succeed in such an attempt and not generate additional batches of AxNanog/GroEL antibody mixtures, the affinity column had to comprise of recombinant AxNanog, devoid of any GroEL chaperone.

Initially, my plan was to overexpress recombinant AxNanog and purify it with glutathione sepharose beads. Naturally, AxNanog would co-purify with bacterial GroEL. At the final stage of affinity purification and prior to reduced glutathione cleavage of the GST-tag, a significant quantity of recombinant protein will be bound and immobilised on the resin. I reasoned that attempting sequential concentration NaCl or urea washes would remove GroEL and result in pure recombinant AxNanog, still bound on the resin. SDS-PAGE and Western blot experimentation would show the effectiveness of the method. If successful, the column would be used for antibody purification and to generate pure recombinant protein solution the last step would be to release the protein from the resin and the GST tag with reduced glutathione.

*E. Coli* bacteria were transformed with the plasmid containing the Nanog sequence and induced to overexpress Nanog. After cell harvest and lysis, the necessary controls were run on SDS-PAGE to assess the success of the induction (Fig. 3.5). The induction proved to be successful, but only a part of the expressed protein was present in the soluble fraction after bacterial lysis. It has to be noted that successful recombinant AxNanog overexpression was confirmed by mass spectrometry (data not shown). The majority of the overexpressed protein was in the insoluble fraction/pellet after the

centrifugation to separate soluble from insoluble fraction (Fig. 3.5a). The presence of Nanog in the pellet indicates that it is aggregated in inclusion bodies, which are often formed in bacteria and contain either overexpressed or viral proteins.

In the current case, the presence of inclusion bodies was advantageous, as they are not expected to contain any chaperone proteins and should primarily consist of recombinant Nanog. Because of this finding, that was not reported by other laboratory members, I decided to exploit the inclusion bodies as they would lack any GroEL and modified my initial plan. I assumed that if purified successfully, inclusion bodies could release high amounts of recombinant Nanog, with a low contamination from host proteins and specifically GroEL. To release the protein, the inclusion bodies were purified from cell debris and the aggregated protein was solubilized. The solubilisation step was necessary to disrupt the protein aggregates and was followed by a protein-folding step in order for the protein to assume its correct folding.

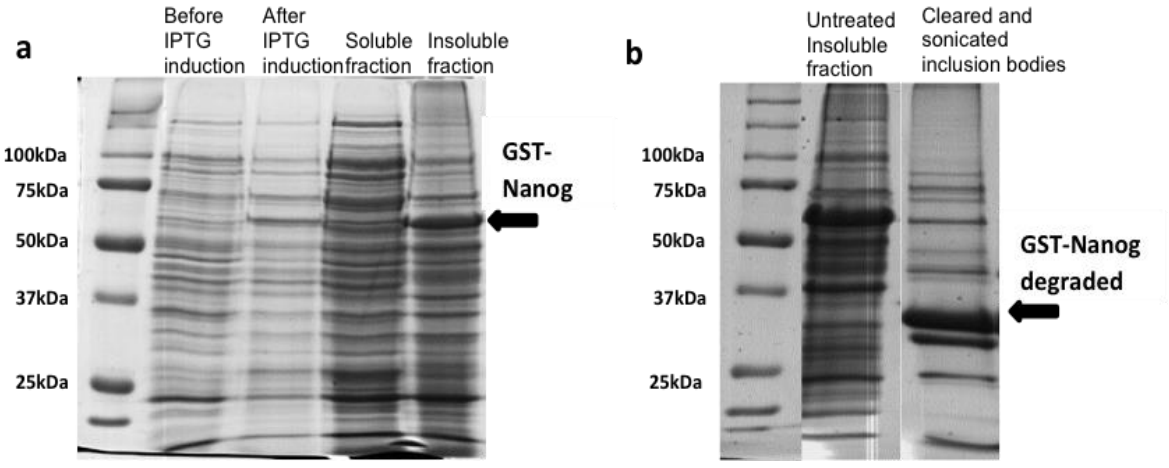



**Figure 3.5. Coomassie stained SDS-PAGE gels of the recombinant *axolotl* Nanog overexpression in *E. Coli*. a) SDS-PAGE gel comparing the different stages of the protein overexpression. The band highlighted by the arrow is the recombinant Nanog, it has to be noted that Nanog in this case appears as a 65kDa because it is GST-tagged. GST is a tag of 26kDa and it significantly affects the molecular weight of any protein. b) Attempt to purify inclusion bodies with two sequential sonication-centrifugation steps.**

Sequential sonication rounds were employed to purify the inclusion bodies from irrelevant bacterial material. This first inclusion bodies' purification method proved to be a harsh treatment for the recombinant protein that degraded after the sonication steps, as shown by the SDS-PAGE gels (Fig. 3.5b). Urea, a protein denaturant, was used in the next attempt to purify and solubilize inclusion bodies. Buffers of 2M-8M urea gradient were used to treat sequentially or non-sequentially the inclusion bodies (Fig. 3.6a and b). Sequential washes proved to be more efficient in the purification of the inclusion bodies, as unspecific material was gradually removed with the lower molarity urea treatments, leaving recombinant Nanog purer. Insoluble inclusion bodies were expected to be fully solubilized in the high molarity urea buffers, but solubilisation was not observed until the inclusion bodies were treated with 8M urea buffer overnight (Fig. 3.6a) .

After the solubilisation of the inclusion bodies the released Nanog protein has to assume its spatial conformation to remain in solution. This is an important step, as it will determine the success of the antibody purification. The first approach used was dialysis to assist protein folding. With dialysis, the urea residues that maintain proteins in a denatured state will be gradually removed, to allow for a gradual native folding (Yamaguchi and Miyazaki, 2014). It has to be noted that a step-wise dialysis protocol was followed, with the solubilised protein (8M urea) being first dialysed in a 6M urea buffer, then in a 4M urea buffer, subsequently in a 2M urea buffer to a final dialysis in PBS. Unfortunately, precipitation was observed during the dialysis steps indicative of the limited success of dialysis in a large amount of protein assuming a proper/soluble conformation.

A different protein solubilisation and refolding protocol was tried in order to maximise the amounts of refolded protein, avoid protein precipitation and also assess the ability of the retrieved recombinant protein to bind on glutathione Sepharose beads. The additional method that was chosen for the solubilisation and folding of the protein was based on a publication exploiting the attributes of three detergents (Tao et al., 2010). To have a thorough

impression of how the solubilisation and refolding method impacts the binding of the GST tagged recombinant protein to the glutathione Sepharose beads, the urea solubilisation/dialysis refolding and detergent solubilisation/refolding methods were compared via SDS-PAGE (Fig. 3.7).

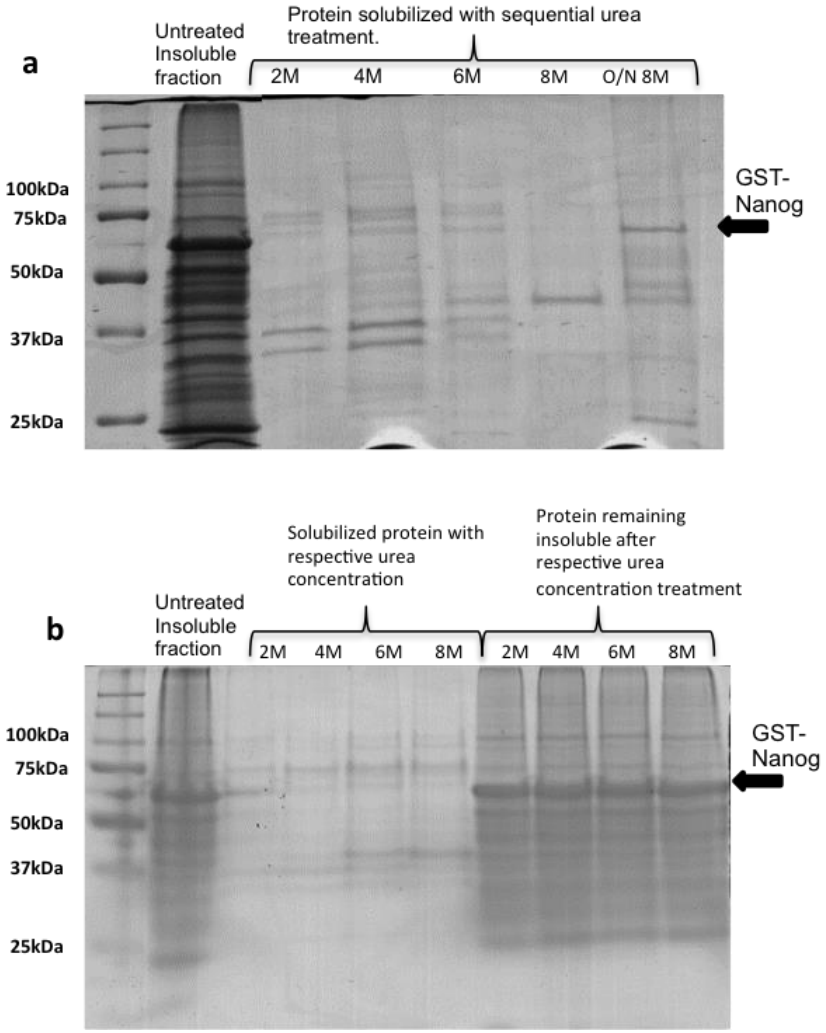



**Figure 3.6. Coomassie stained SDS-PAGE gels of the inclusion bodies' solubilisation attempts. a) Attempt to solubilize inclusion bodies with sequential urea treatments. After the isolating and clearing the insoluble fraction containing the recombinant AxNanog protein, the inclusion bodies were treated with mild and harsh urea buffers in a sequential manner. Anything that was not solubilised with a lower urea concentration buffer was subsequently treated with a higher urea concentration buffer. Urea concentration solutions used: 2M, 4M, 6M and 8M. An overnight 8M urea treatment (O/N 8M) was also employed. b) SDS-PAGE gel of non-sequential urea treatments for the solubilisation of inclusion bodies. Here batches of inclusion bodies were treated with a single urea buffer of the respective concentration and solubilisation was assayed by separating the soluble and insoluble fractions. Urea buffer concentrations used: 2M, 4M, 6M and 8M.**

As shown in Fig. 3.7, both detergent and urea methods were effective in the solubilisation of inclusion bodies. When the purified protein was incubated with glutathione Sepharose beads, detergent solubilised protein did bind the beads more effectively in comparison to urea solubilised *axolotl* Nanog. At the same time, the urea solubilised inclusion bodies demonstrated higher amounts of non-bound GST tagged recombinant *axolotl* Nanog, compared to detergent solubilised inclusion bodies. This possibly indicates that in the case of urea solubilisation, where the protein is refolded via step-wise dialysis, the protein is not refolded successfully and the GST-tag is potentially engulfed and inaccessible. On the other hand, detergent solubilisation that is followed by detergent dilution to assist in protein folding, showed a lower amount of non-bound *axolotl* Nanog, indicative of a higher rate of properly folded protein. Regarding background binding on the glutathione Sepharose beads, as seen in Fig. 3.7, urea solubilised/dialysed protein bound with a much cleaner profile while detergent solubilised protein showed higher amounts of background. The background binding might be associated with degraded *axolotl* Nanog, or irrelevant proteins. However, because of the large amount of *axolotl* Nanog binding on the beads after detergent solubilisation of the inclusion bodies that was the method chosen for the reconstitution of recombinant AxNanog and preparation of the affinity purification column. Because the ultimate aim is to successfully purify the *axolotl* Nanog antibody from the anti-GroEL contaminant, it was important to have an excess of *axolotl* Nanog on the affinity purification column to ensure purification of as much AxNanog specific antibody as possible. I was not alerted by the presence of background binding, as none of the background bands correlated with GroEL and should therefore not interfere with the purification of the AxNanog antibody.

To generate the affinity purification column, a larger batch of inclusion bodies was prepared and solubilised with the chosen method. The antibody was purified from the rabbit serum with the conventional antibody purification method. After its purification the new antibody batch was validated for its affinity to native and recombinant *axolotl* Nanog via Western blotting and immunocytochemistry. The old (anti-GroEL contaminated) and new AxNanog

antibody batches were compared by Western blotting to see how the purification step benefited the function of the antibody (Fig. 3.8a and b). *Axolotl* oocyte and embryo lysates, along with in vitro translated recombinant Nanog were tested and a profound improvement was seen between the old and new antibody batch. Between the oocyte and embryo lysates, the old antibody primarily detects its GroEL equivalent; however, the new batch is less responsive to the 65kDa band and reveals additional bands in the regions of 50 and 37kDa with the latter one corresponding likely to native Nanog. Regarding recombinant and in vitro translated Nanog, the old antibody is more prone in detecting the GroEL contaminant in comparison with the purified antibody that results in a much fainter 65kDa band. Interestingly both antibodies still recognise the GroEL contaminant in the recombinant Nanog lane. The bands below the 25kDa region in the RN lanes (Fig. 3.8a and b) correspond to degraded recombinant Nanog as indicated by mass spectrometry (Fig. 3.4a). The purified antibody was more efficient in detecting the degradation products, which indicates its enhanced specificity. In vitro translated Nanog was also better detected with the purified antibody.

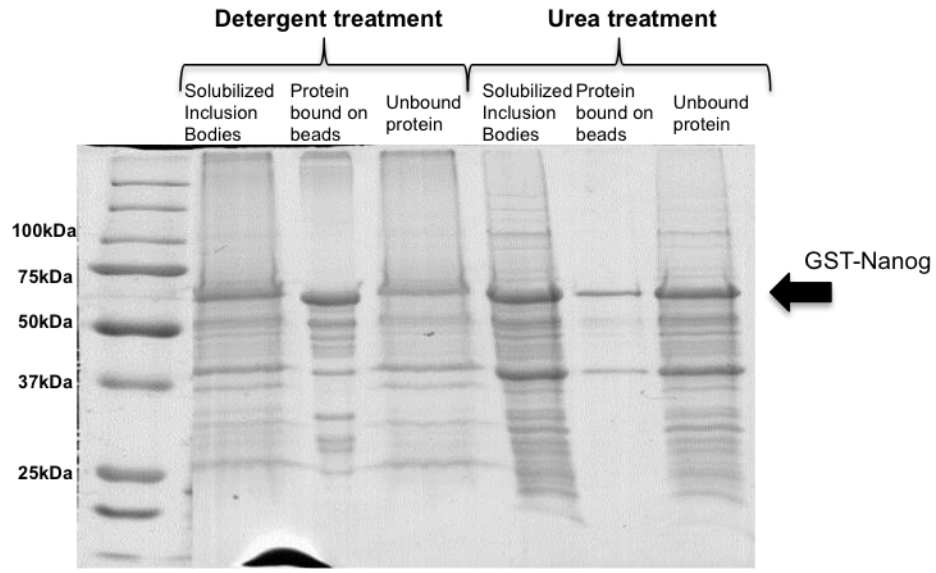



**Figure 3.7. Coomassie stained SDS-PAGE gel comparing detergent and urea treated inclusion bodies in terms of AxNanog solubilization and glutathione sepharose bead binding. The total protein bound on the beads as well as the protein that failed to bind served as measures of AxNanog folding efficiency in each condition.**

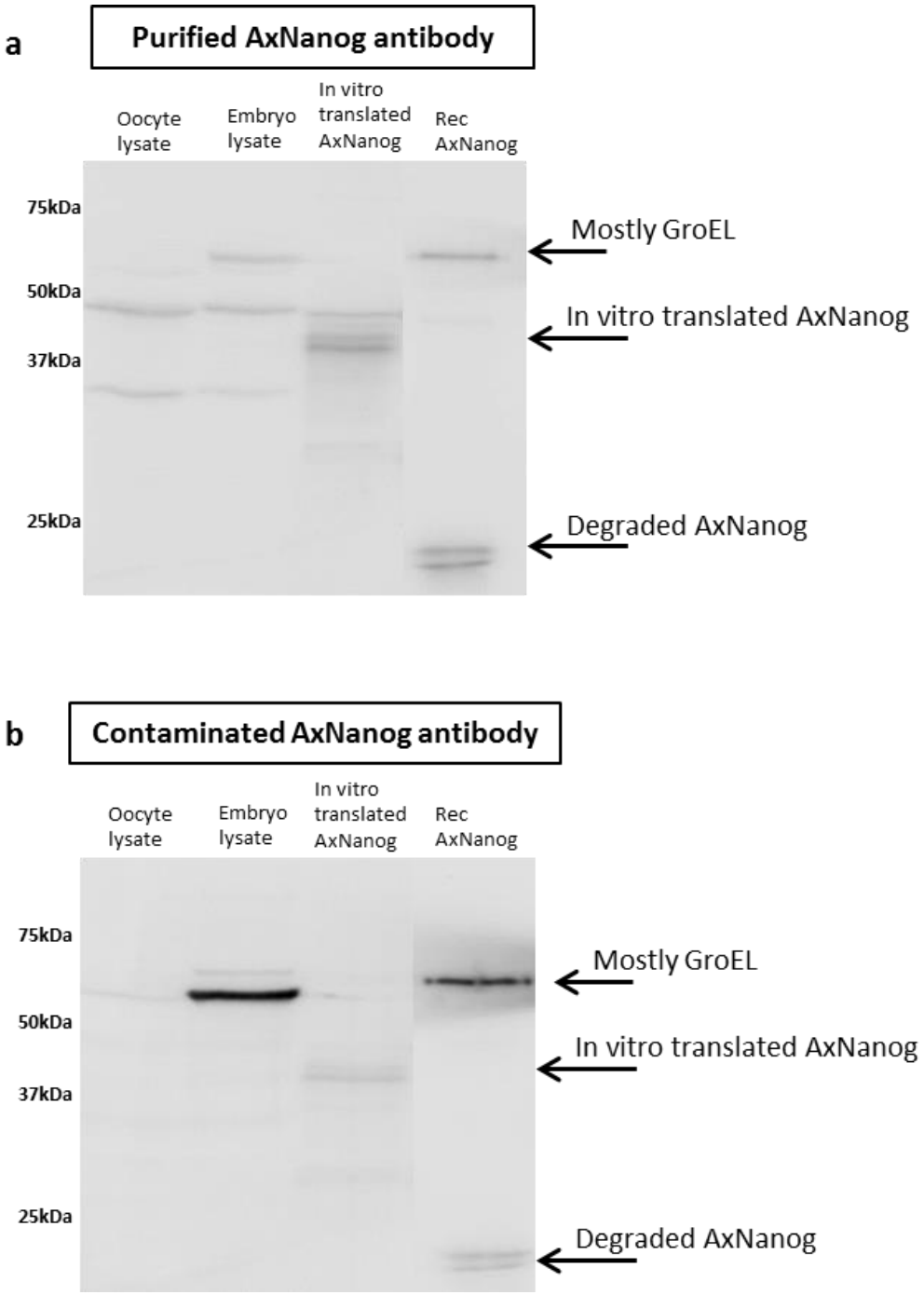



**Figure 3.8. Western blots comparing the newly purified to the GroEL contaminated antibody. a) Western blot assessing the affinity of the newly purified Axnanog antibody against native, in vitro translated and recombinant AxNanog. b) Western blot of the contaminated AxNanog antibody against native, in vitro translated and recombinant AxNanog. For the recombinant AxNanog loaded lane in both images, a batch of GroEL contaminated recombinant protein solution was used.**

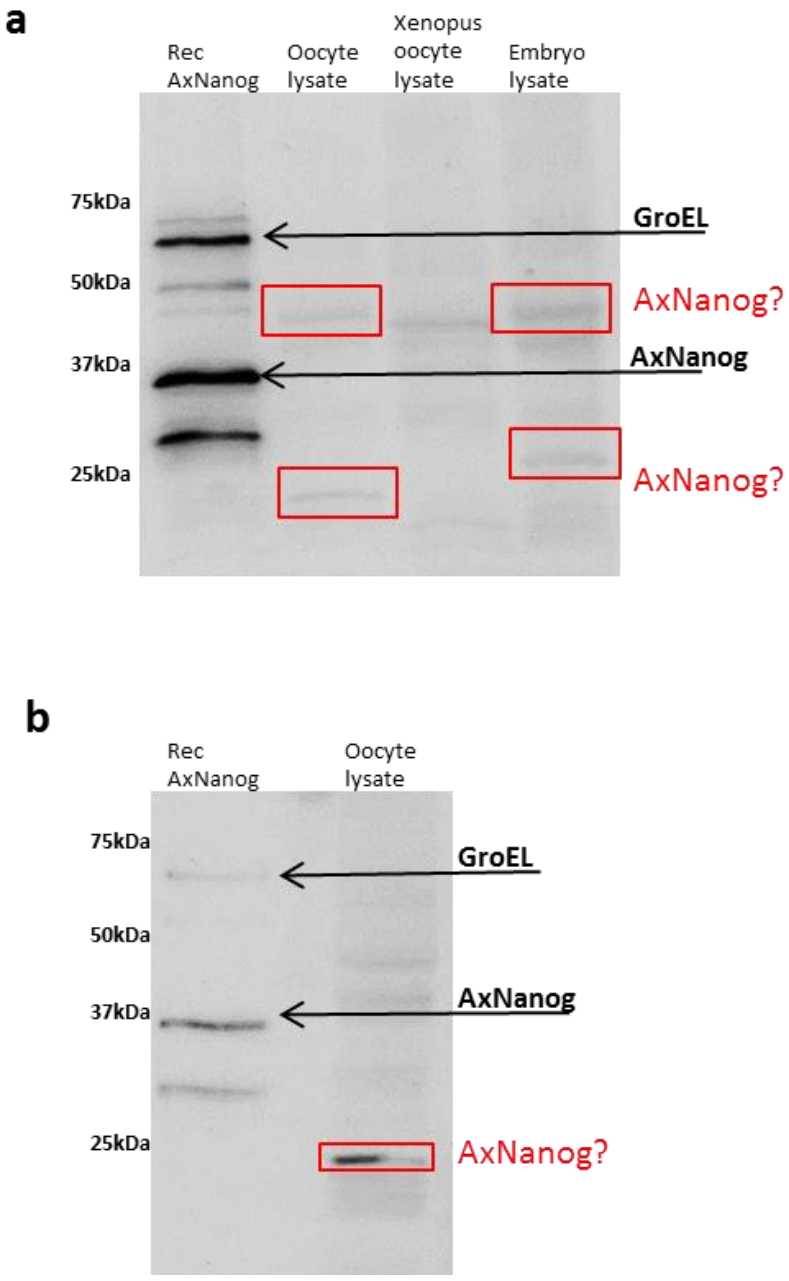



**Figure 3.9. Western blots comparing the newly purified to the GroEL contaminated antibody. a) Western blot assessing the affinity of the newly purified Axnanog antibody against native and recombinant AxNanog. Xenopus oocyte lysate was employed as negative control as it does not express any Nanog orthologs. b) Western blot testing newly purified AxNanog antibody at a lower titer against recombinant and native AxNanog.**

In Fig. 3.9a and b, freshly made recombinant Nanog was immunoblotted along with lysates in another attempt to test the purified antibody. The antibody could clearly recognise and bind recombinant Nanog which in this case is located at the 37kDa region. The molecular weight was the appropriate, judging by the predictions. There was still a band of 65kDa on the Western blots, possibly indicating that GroEL was still purified along with the recombinant protein. However the same band could also correspond to the version of recombinant Nanog still conjugated with the GST tag. In order to generate recombinant Nanog, the GST tag that was exploited as a means of conjugation to a resin, needed to be cleaved off with the use of reduced glutathione. The recombinant protein was then released from the resin, resulting in a recombinant protein solution. However, the cleavage step is a challenging step. It is often difficult to cleave off significant amounts of recombinant protein and there is no reassurance that the buffers used cannot elute traces of intact GST tagged recombinant protein. It needs to be noted that even though the bands appearing on the Western blots were profound and appeared to represent significant amounts of protein that was not the case with the respective Coommasie stained SDS-PAGE gels (image not shown). The bands appearing on the SDS-PAGE for recombinant Nanog were fainter and did not correspond to the respective size of the Western blot bands. This was indicative firstly of the cleavage inefficiency and secondly of the high specificity of the antibody, capable of binding to its antigen even in scarce amounts.

Even though there was no ambiguity regarding the ability of the antibody to recognise *axolotl* Nanog in its in vitro translated and recombinant form, in the case of the lysates, the antibody did not demonstrate reproducibility. As seen in Fig. 3.8a the oocyte and embryo lysate lanes showed two additional bands at the 50 and 37kDa regions; however the same pattern did not reproduce in Fig. 3.9a. On the contrary additional bands appeared mostly resembling background binding, as a similar profile was observed in the Xenopus oocyte lane. In this case the Xenopus oocyte lysate was used as a negative control, because its genome is devoid of a Nanog gene. In the same figure, two bands

each appearing in the oocyte and the embryo lysate lanes (highlighted by black squares) were found in a region where the Xenopus oocyte extract did not give any bands and could potentially correspond to Nanog. The same band for the case of the *axolotl* oocyte lysate was observed in Fig. 3.9b giving a bolder signal. However, from the above figures and from additional data (not shown) we cannot conclude that the antibody definitely identified Nanog in the lysates. In order to obtain an additional perspective, murine pre-iPS cells null of endogenous Nanog but modified to express *axolotl* Nanog were assayed via immunocytochemistry (Fig. 3.10).

Rabbit IgG was used as a negative control and to identify background binding. DAPI was used to stain nuclei in both cases. Comparing the cells stained for AxNanog with the ones stained for IgG, it is obvious that in the case of Nanog the signal is strong and nuclear as opposed to the faint signal in the IgG group.

The results obtained suggest that the AxNanog antibody generated is capable of recognising the AxNanog antigen in its native form, however in its first purification it was contaminated by anti-GroEL antibody. The purification was successful as it enhanced the detection of recombinant and in vitro translated AxNanog and minimized detection of GroEL and its equivalents. However, the antibody did not result in a band for the case of *axolotl* oocyte lysate. This contrasted the immunocytochemistry results where a nuclear signal was observed for AxNanog.

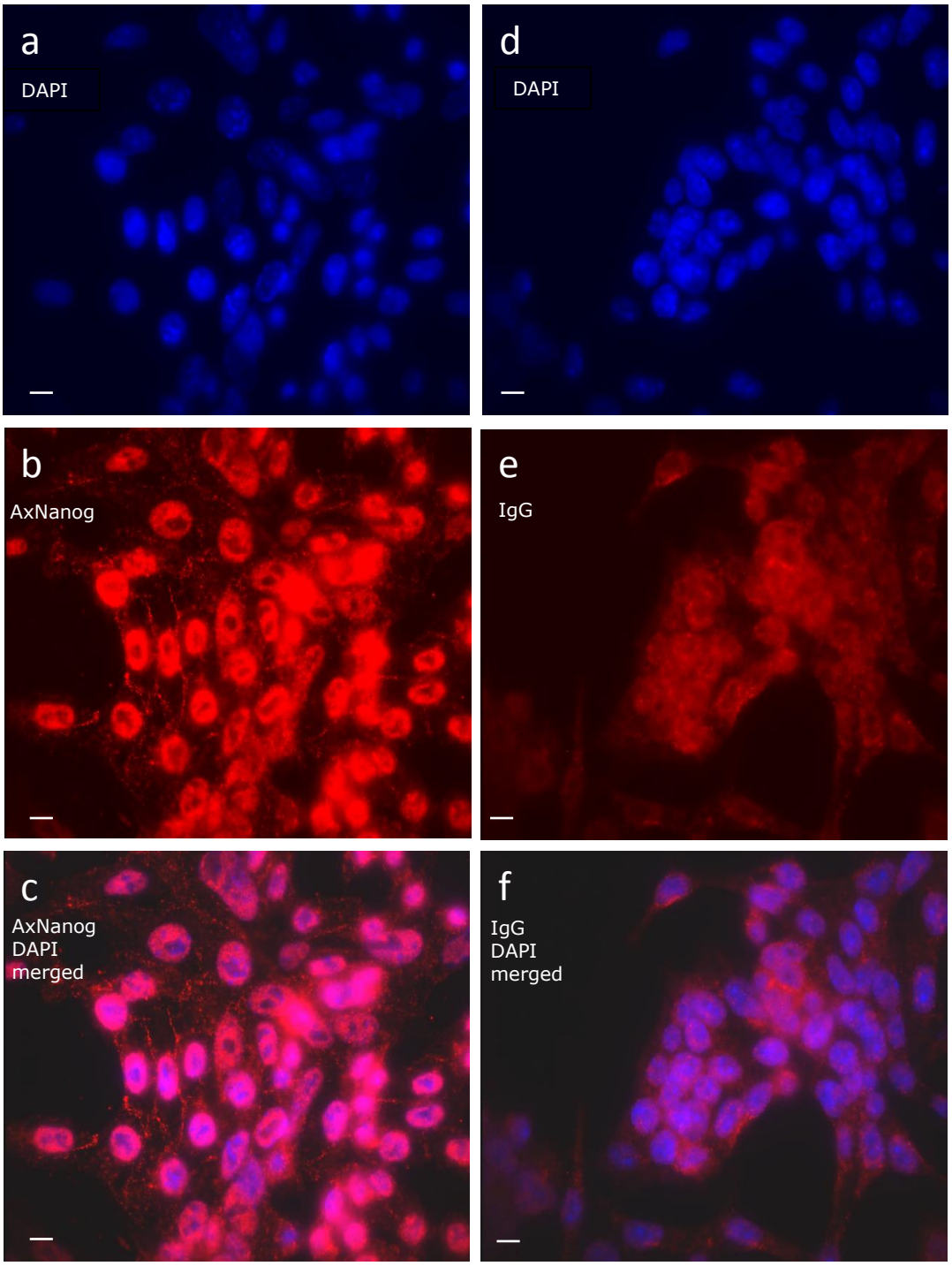



**Figure 3.10. Immunocytochemistry with the newly purified *axolotl* Nanog antibody on Nanog null murine pre-iPS cell expressing *axolotl* Nanog. a, b, c) cells stained for Nanog. d, e, f) cells stained for IgG. a, d) DAPI staining of the nuclei. b, e) anti-Nanog (b) and anti-rabbit IgG (e) staining. c, f) Merged image. Scale bars 10µm.**

## 3.5 Conclusion

The present chapter was initially aimed at identifying physical interactors of *axolotl* Nanog in *axolotl* oocytes. The tool I used was an AxNanog antibody, generated and tested by previous lab members, that was employed in protein immunoprecipitation and co-immunoprecipitation experiments. Prior to conducting any immunoprecipitation experiment, I tested the AxNanog antibody against native *axolotl* Nanog in oocytes. A band at 65kDa was observed that coincided with previous observations. While this was not the predicted molecular weight of *axolotl* Nanog based on its amino acid sequence, I reasoned that this might be because of post-translational modification occurring on the surface of *axolotl* Nanog in the oocyte. Extensive post-translational modifications are frequently observed in the oocyte (Ma et al., 2008). Additionally, the antibody was also tested against in vitro translated *axolotl* Nanog protein and resulted in a band of molecular size close to the predicted, however equally specific as the band observed in the oocyte.

Upon experimenting with protein immunoprecipitation I realised that native Nanog was not precipitated, as should be the case for an efficient immunoprecipitation experiment. This initiated a round of experiments that effectively showed that the AxNanog antibody was contaminated with GroEL antibody that was significantly hindering detection of AxNanog. The contamination has potentially resulted by the contamination of the recombinant *axolotl* Nanog by bacterial GroEL in the first place. GroEL is a bacterial chaperone assisting in protein folding which can often contaminate recombinant proteins overexpressed in bacterial systems (Rohman and Harrison-Lavoie, 2000). The schematic model for the contamination of recombinant *axolotl* Nanog by GroEL is presented is figures 3.11 and 3.12. GroEL, being a chaperone, binds unfolded protein and promotes its correct folding and therefore proteins overexpressed in bacteria will assume their native conformation with the help of GroEL. During antibody production, if

GroEL contamination preventing methods are not employed, the chaperone will be present in the purified recombinant protein solution used for the immunization step, and therefore antibodies for the chaperone will be generated as well. For the purification of the antibody, the protein-antigen has to be overexpressed in bacteria and immobilized on resin, but again GroEL will be present as a contaminant leading in this way to the co-purification of both the antigen specific and GroEL antibody (Fig. 3.11 and 3.12). While attempting to overexpress AxNanog in *E. Coli* in order to purify the contaminated AxNanog antibody, I realised that the recombinant protein was forming inclusion bodies in the bacteria as a result of its overexpression. I decided to exploit the inclusion bodies to my advantage because they almost exclusively contain overexpressed protein and no bacterial chaperones, such as GroEL. The reason why overexpressed proteins might end up in these aggregates is linked with bacterial chaperones. Proteins will assume the appropriate folding after interacting with chaperones. Proteins that fail to fold successfully become insoluble and if in large amounts, they form inclusion bodies. When overexpressing proteins in bacterial systems, the amount of bacterial chaperones is only sufficient to bind and promote the folding of a proportion of the overexpressed protein. The rest remain unfolded and aggregate in inclusion bodies. To use the inclusion bodies for the purification of AxNanog antibody from the GroEL antibody contaminant, I purified, solubilized and re-folded inclusion body AxNanog. After AxNanog antibody purification (from immunized rabbit serum prepared by previous lab members) I tested the new AxNanog antibody and found that a definite improvement was observed regarding the GroEL contaminant and detection of recombinant and in vitro translated AxNanog, as well as native AxNanog in immunocytochemistry. However, the antibody failed to demonstrate reproducibility with native oocyte AxNanog in Western blot experiments. . It is possible that the levels of protein might be too low to be detected in the lysates because its expression and existence in the lysate is transient. Additionally, it needs to be noted that AxNanog is a nuclear protein as predicted and observed in Fig. 3.10. The concentration of the protein in the

relatively small space of the nucleus might be beneficial for its detection. Alternatively, in the case of oocyte lysate the protein becomes defused in the lysate, something that could be impeding its detection in the case it is expressed in lower amounts. It is likely that higher amounts of the AxNanog protein might be expressed upon oocyte maturation and only traces of it reside in the immature oocyte.

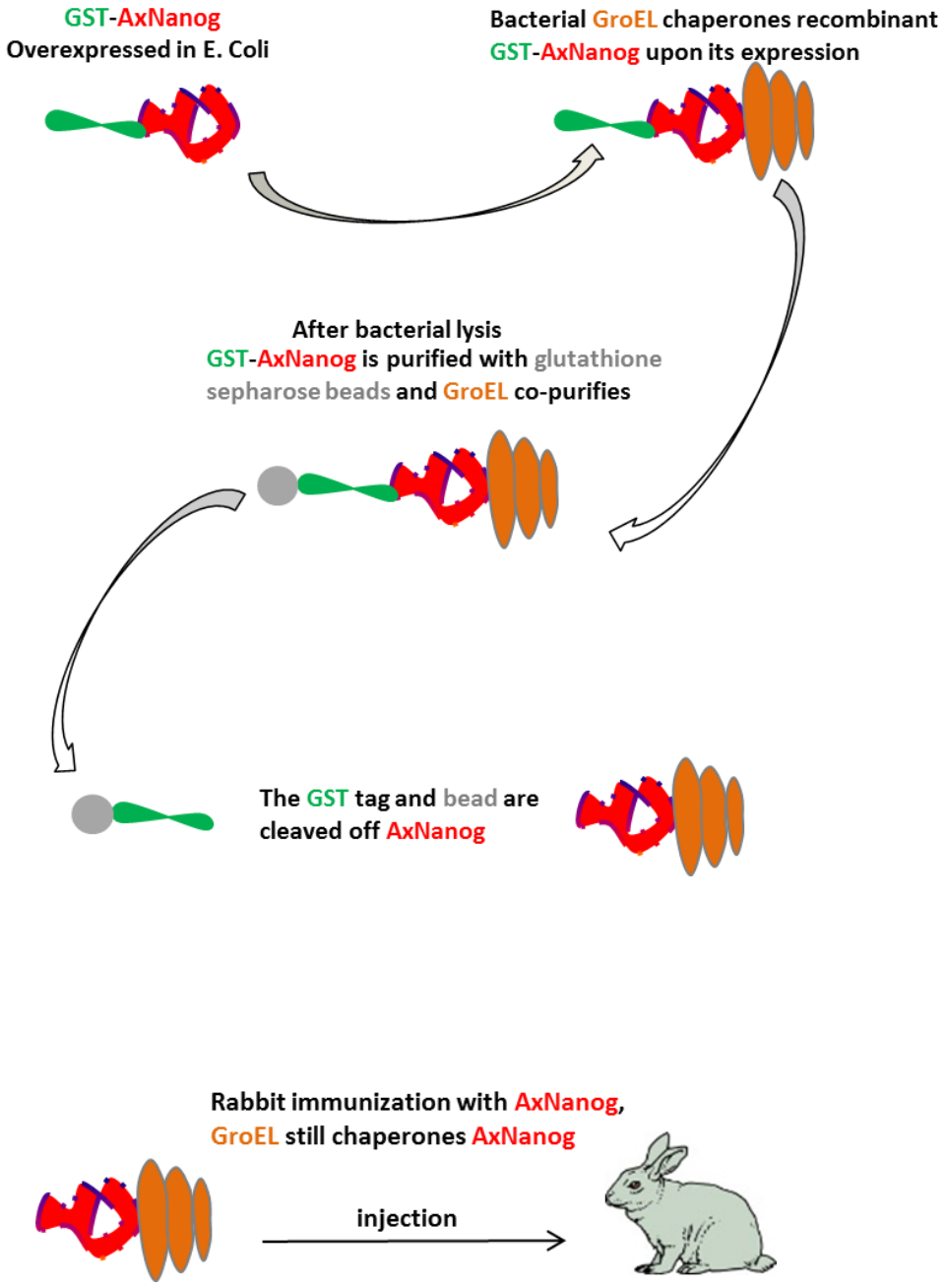



**Figure 3.11. Model depicting the contamination of recombinant AxNanog with bacterial GroEL, The overexpression of GST-tagged AxNanog is induced, however GroEL as a chaperone, binds to the synthesized protein in order to assist its folding. The purification of recombindant AxNanog from the bacterial lysate is carried out via the GST-tag that is bound by glutathione sepharose coated beads. Upon its purification, the GroEL chaperone remains bound on recombinant AxNAnog. Prior to the immunization of the rabbit, recombinant AxNanog is cleaved off the GST-tag, leaving the tag still bound to the beads. However, since the GroEL chaperone is bound to AxNanog, it is retained in the recombinant AxNanog protein solution. The recombinant AxNanog protein solution (and GroEL) was then used for the immunization of a rabbit. Note that the recombinant AxNAnog solution is used as a control for AxNanog in SDS-PAGE gels and Western blots.**

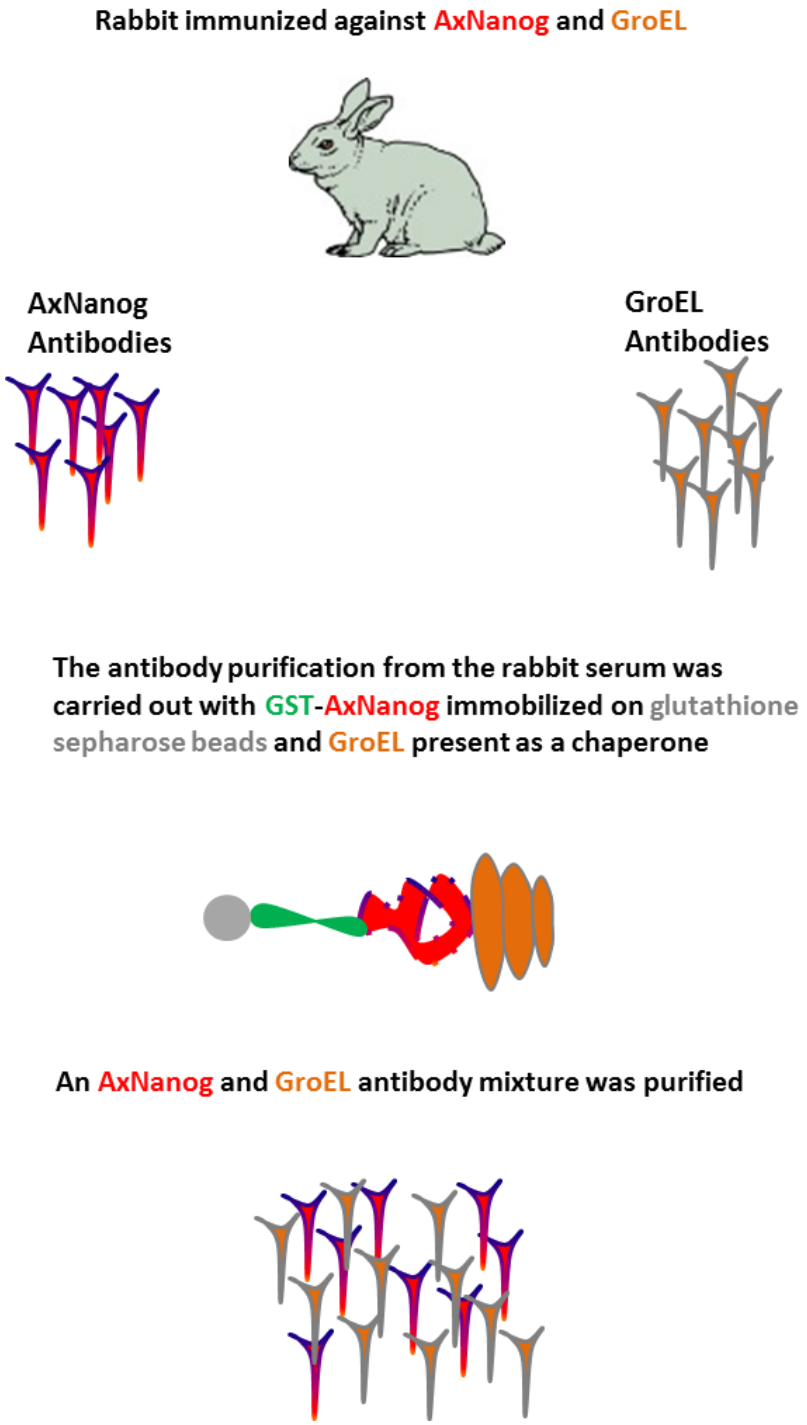



**Figure 3.12. A continuation from Figure 9. Model depicting the contamination of recombinant AxNanog with bacterial GroEL. Because of the presence of GroEL along with AxNanog, the rabbit was immunized against both proteins. The purification of AxNanog specific antibodies was carried out with the use of GST-tagged AxNanog overexpressed in *E. Coli*, bound to glutathione sepharose beads. Hence, GroEL was again present as a chaperone on the surface of AxNanog. Therefore both AxNanog and GroEL specific antibodies were purified from the rabbit serum.**

# Chapter 4: The regulatory network of the Nanog promoter

## *4.1 Introduction*

### 4.1.1 Protein-DNA interactions in ESC: Deciphering regulatory pathways

Protein-DNA interactions govern transcriptional regulation in all cell lineages. Proteins recognise and bind genetic domains and thus impact the expression of the respective gene. The regulatory regions of pluripotency associated genes are targeted by transcription factors as discussed in the introduction (Rodda et al., 2005). Reportedly, two core pluripotency factors, Oct4 and Sox2 aside from physically interacting with one another, also bind to the respective oct-sox elements of their own regulatory domains (promoter and enhancer), forming a positive loop that maintains the expression of Sox2 and Oct4 (Chew et al., 2005). The transcription factor Sall4 (Sal-like protein 4) was found to be an interacting partner of Nanog but also to positively regulate Nanog expression by binding to its enhancer (Wu et al., 2006). Additionally, TCF3 (Transcription Factor 3) was reported to occupy Nanog promoter and repressing its transcription (Pereira et al., 2006).

Transcription factors do not solely operate through binding and impacting regulatory regions of genes. They can also involve in the recruitment and in interaction with additional molecules, thus extending the activity they can confer on a genetic locus. The interactomes (interacting protein networks) of the core pluripotency factors have been studies revealing a multitude of associations.

Oct4 was found to physically interact with the transcription factors Nanog, Sall4, SP1 (Specificity Factor 1) and Requiem, while the histone deacetylase

HDAC2 was also identified in the Oct4 complex (Wang et al., 2006). Sox2 was reported to have its own protein interaction network, overlapping to an extent with the reported networks for Oct4 and Nanog. More specifically Sox2 was found to interact with Sall4, Lin28, H2A histones, FACT complex subunits (SUPT16h and SSRP1), PARP proteins etc. reflecting cellular multitasking (Mallanna et al., 2010). A far larger network of interactors was reported by Gagliardi et al including transcription factors such as ESRRB, REX1 and 2, UBTF, Smarca5, NuRD subunits (CHD4, MBD3 and 2) as well as chromatin remodellers (Gagliardi et al., 2013).

Recently, Beyer et al identified the core pluripotency factors Nanog, Oct4 and Sall4 along with members of the Hippo pathway (TEAD1, 3, 4) and SMAD2/3 as binders of the Nanog promoter in hESC (Beyer et al., 2013). The Hippo pathway participates in the cell proliferation and apoptosis while SMAD2/3 are transcription factors and downstream effectors of the TGF-b signalling pathways.

The above findings collectively show that the core pluripotency factors reside in protein interaction as well as gene regulation networks. While a protein network study can begin with the protein of interest as bait to generate its physical interactors, a gene regulatory network can be identified through the study of a regulatory region's binders. Often a gene regulatory and protein network will dissect, as factors physically interacting in the nucleus (or cytoplasm) will interact on the surface of DNA. Both studies can generate candidate molecules for further studies and biological assays. More specifically, in the context of ESC and reprogramming towards pluripotency factors identified in protein or protein-DNA interaction screens can be employed to assess their impact on iPS overexpression trials.

## 4.2 Aims and objective

The Nanog gene regulatory regions are recipients of a number of regulatory binders and a recent publication reported an efficient method for the identification of human Nanog promoter binders (Beyer et al., 2013). Based on the procedure described by Beyer et al (Beyer et al., 2013), the equivalent murine genetic region (mouse Nanog promoter) was incubated in *axolotl* oocyte lysate to identify its binding factors. Identified binders will potentially play a regulatory role in the transcriptional activation of Nanog during reprogramming.

Aims

- To identify pluripotency associated factors, preferentially binding to the murine Nanog promoter in the context of the oocyte.
- To shortlist pluripotency associated candidates among the identified factors.
- To delineate the specific role of each candidate in reprogramming by additional biological assays.

Objectives

- Employ the DNA pull down method to precipitate mNanog promoter binders in the axolotl oocyte.
- Identify binding factors with mass spectrometry.
- Compare the factors binding to the murine Nanog promoter in the oocyte versus the embryo to identify factors likely to be involved in the regulation of pluripotency genes in the early stages of development.

## 4.3 Materials and methods

DNA pull down was employed to identify proteins interacting with mNanog promoter. The method and materials are presented in detail in Chapter 2, Section 2.11. The primers, PCR reaction and cycling conditions can be found in Chapter 2, Section 2.10. To determine the right protocol and conditions for the purification of the PCR product for the DNA pull down three different procedures were tested as presented in Section 4.4.2. The transcription factors binding analysis was carried out with the commercially available software TRANSFAC. The DNA pull down results were analyses on SDS-PAGE and stained with either Coomassie or silver stain, as described in Section 2.3. The western blotting was carried out as described in Section 2.4. The antibodies used are stated in the respective figure and within the respective section. For the mass spectrometry samples were first analysed on SDS-PAGE gel, stained accordingly and then processed with in gel tryptic digestion as described in Section 2.12. After C18 clean-up of the samples, they were processed in the Nottingham Trent University facilities. The mass spectrometry instrument used was ABSCIEX6600 and the data was obtained after a 45 minute LS-MSMS run (Section 2.12, Mass spectrometry injection). The obtained data was searched through MASCOT against the custom database comprised of AXOLOTL EST (Matt Loose) sequences using MS/MSIONS search on the local MASCOT server using standard default settings (Pept Tol +-1.2Da, MS/MS Tol 0.6 Da, Pept Charge 2+). Significance threshold set at 0.05.

## 4.4 Results

### 4.4.1 Murine Nanog promoter and Transcription Factor binding analysis

Based on the recent study of Beyer et al (Beyer et al., 2013) where the human Nanog promoter is employed in DNA pull down experiments to identify its interactors, I decided to study the equivalent murine DNA sequence. The murine Nanog promoter was first identified and described by Wu and Yao (Wu and Yao, 2005) (Fig. 4.1). The equivalent of the region termed *Nanog80* by the research team of Beyer et al, where the binding loci of the Oct4, Sox2 and SMAD factors are concentrated (Beyer et al., 2013), is located in the -211 to -127 region (Fig. 4.1). Notably, the *Nanog80* region is well conserved between human and mouse as judged by Clustal Omega sequence alignment (data not shown).

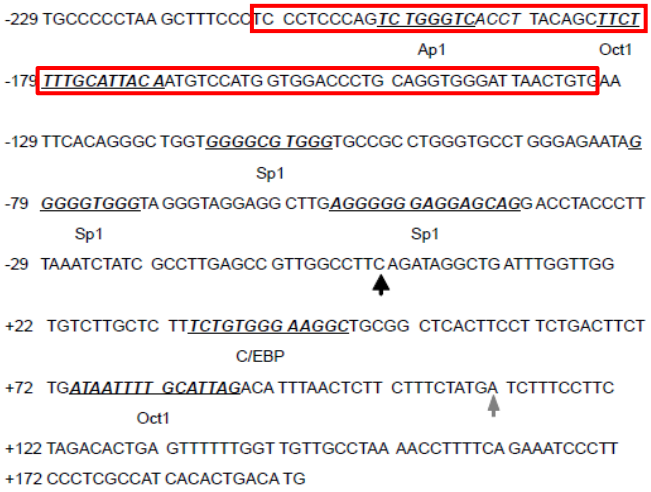



**Figure 4.1. The mouse Nanog promoter as reported by Wu and Zao, depicting the 5' flanking region of the Nanog gene, containing the nucleotide sequence of -229 to +193. The black arrow shows the transcription start site (TSS). The underlined regions represent putative binding regions for the respective elements. The grey arrow points another minor TSS. The red rectangle area marks the Nanog80 equivalent in the murine Nanog promoter sequence.**

In order to centralize the *Nanog80* equivalent, I chose to extend the genetic sequence analysed to contain the 5' flanking region starting from position -542 to +121 (Fig. 4.2). I wanted the *Nanog80* locus to be central as it concentrates the binding loci of Oct4 and Sox2 and these factors could lead to the recruitment of additional factor upstream or downstream *Nanog80.* Additionally the transcription start site of Nanog is located downstream of the *Nanog80* locus, which might also affect the binding of some factors. The extended promoter was analysed for putative transcription factor binders in silica, with the commercially available software TRANSFAC (BIOBASE, Qiagen). The transcription factor binding analysis carried out by the software was based on experimentally characterised binding sites by a range of DNA-transcription factor interaction experiments.

TGCGTTAAAAAGCCGCACTTTTGGAGGGAAGATTTCCCCAGGTTTCCCAATGTGAAG
AGCAAGCAAGAAACGCTGAGTGCTGAAAGGAAAGCCGTGTATAAACAGAGACTTTGG
CAGCAAGGTCTGACTCTTTCATGTCTGTAGAAAGAATGGAAGAGGAAACTCAGATCC
CCCACTTGACCTGAAACTTCCCACTAGAGATCGCCAGGGTCTGGAGGTGCAGCCGTG
GTTAAAAGATGAATAAAGTGAAATGAGGTAAAGCCTCTTTTTGGGGGGGGGGGATGT
CTTTAGATCAGAGGATGCCCCCTAAGCTTTCCC**TCCCTCCCAGTCTGGGTCACCTTA**
**CAGCTTCTTTTGCATTACAATGTCCATGGTGGACCCTGCAGGTGGGATTAACTGTG**A
ATTCACAGGGCTGGTGGGGCGTGGGTGCCGCCTGGGTGCCTGGGAGAATAGGGGGTG
GGTAGGGTAGGAGGCTTGAGGGGGGAGGAGCAGGACCTACCCTTTAAATCTATCGCC
TTGAGCCGTTGGCCTTCAGATAGGCTGATTTGGTTGGTGTCTTGCTCTTTCTGTGGG
AAGGCTGCGGCTCACTTCCTTCTGACTTCTTGATAATTTTGCATTAGACATTTAACT
CTTCTTTCTATGATCTTTCCTTC

**Figure 4.2. Murine Nanog promoter sequence subjected to TRANSFAC transcription factor binding analysis. The underlined sequence is the equivalent of human Nanog80 region.**

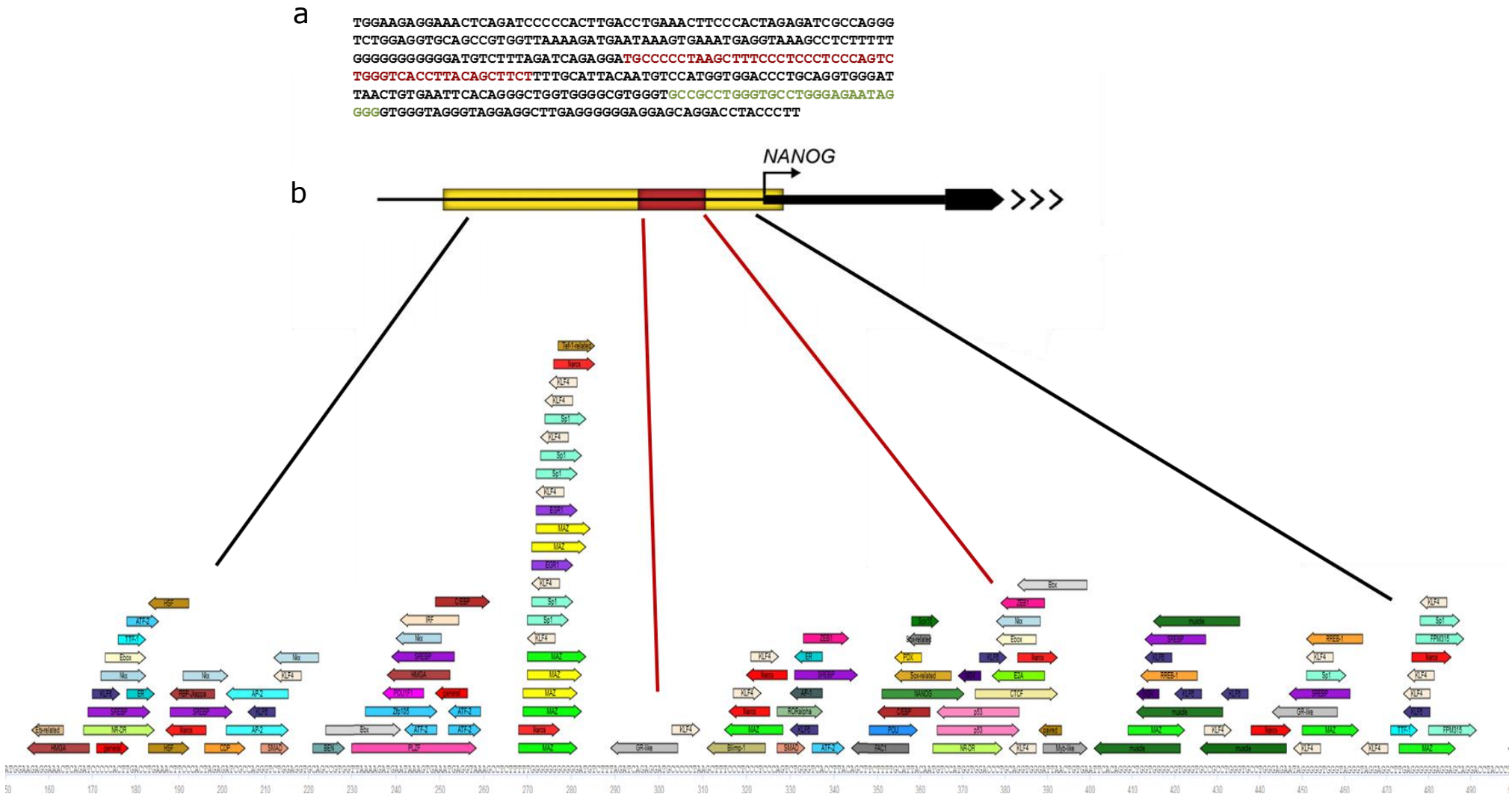



**Figure 4.3. The 350 nucleotide long region of murine Nanog promoter determined as the regions of interest after TRANSFAC analysis of the 650 nucleotide long murine Nanog promoter sequence. Highlighted in bold red is the region containing the Nanog, POU and Sox binding domains, highlighted in green is the transcription start site and exon b) Transcription factor binding site analysis of the 350 nucleotide long murine Nanog promoter sequence with the TRANSFAC software. The arrows depict the binding locus of the respective transcription factor, arrows pointing to the left depict transcriptions factors binding on the complementary DNA strand, whereas arrows pointing to the right depict transcription factors binding to the coding DNA strand.**

Initially a 650 nucleotide long genetic sequence was analysed for putative TFs. To avoid complications by using such a long sequence, I decided to minimize its length by concentrating on the genetic region that assembled the most interesting putative TF binders and kept the *Nanog80* domain central. The TSS was included in the new 350 nucleotide long sequence as it could potentially have an impact on the binding of cis acting transcription factors and omitting the TSS could potentially affect the physical interaction of molecules with their respective sequences (Fig. 4.3). A total of 65 putative binders were identified, 33 of them binding on the antisense DNA strand and 32 binding on the coding strand (Fig. 4.3 and Table 4.1). Nanog, POU and Sox related proteins, SP1, AP2 and CTCF were predicted as putative binders of the analysed murine Nanog promoter domain.

Following the transcription factors analysis, I compare list of predicted TFs (transcription factors) with the *axolotl* oocyte transcriptome (generated by Matt Loose, University of Nottingham). Of all the unique 48 predicted binding transcription factors, 27 were found to be present in *axolotl* oocytes as mRNA transcripts, suggesting that the respective proteins are translated in the *axolotl* oocyte (Table 4.2 and 4.3). The above fact indicates that a potential DNA pull down attempt with the analysed genetic sequence in *axolotl* oocyte extract, could result in the precipitation and identification of either one of the 27 transcription factors. However additional factors might also be identified.

| TFs binding on antisense DNA strand<br>3' ➔ 5' | TFs binding on sense DNA strand<br>5' ➔ 3' |
|---|---|
| ATF-2, TTF-1, Ebox, Nkx, KLF6, ER, SREBP, NR-DR, general, AP-2, SMAD, SDP, HSF, C/EBP, Zfp105, Bbx, BEN, PLZF, Tef-1-related, Ikaros, SP1, EGR1, MAZ, KLF4, ZEB1, Ror-alpha, Sox10, Nanog, POU, CTCF, p53, Myb-like, FPM315 | Ets-related, HMGA, HSF, Nkx, Klf4, AP-2, KLF6, Ikaros, RBP-JKappa, IRF, SREBP, POU1F1, general, ATF-2, GR-like, MAZ, Blimp-1, ER, AP-1, FAC1, C/EBP, Sox-related, PDX, Bbx, ZEB1, Ebox, ING4, E2A, p53, paired, Muscle, RREB-1 |

**Table 4.1. Results of TF binding analysis in silica with the TRANSFAC software. 65 putative binders were predicted, 33 binding on the antisense and 32 binding on the sense strand.**

| Putative Binders Expressed In *axolotl* Oocytes | | | | Putative Binders Not Found in *axolotl* Oocytes | | |
|---|---|---|---|---|---|---|
| Ets-related | TTF-1 | NANOG | ZEB1 | SREBP | RBP-Jkappa | E2A |
| Nkx | ATF-2 | AP-2 | AP-1 | ZFP105 | Ikaros | paired |
| ROR-alpha | ING4 | SMAD | Tef-1-related | PLZF | Sox10 | Muscle-related |
| KLF6 | HSF | KLF4 | HMGA | FAC1 | p53 | NR-DR |
| Myb-like | Sox-related | BEN | IRF | Bbx | ER | RREB-1 |
| CTCF | PDX | POU | C/EBP | CDP | Ebox | FPM315 |
| MAZ | SP1 | EGR1 | | POU1F1 | GR-like | Blimp-1 |

**Table 3.2. The 48 putative transcription factors predicted to bind the murine Nanog promoter. 27 of them have been identified in the transcriptome of *axolotl* oocytes.**

**Predicted mNanog promoter transcription factors expressed in *axolotl* oocytes**

| Transcription factor | Role | References |
|---|---|---|
| Nanog, Oct4 (POU), Sox2 (Sox-related) | Regulation of pluripotency | (Loh et al., 2006) |
| SP1 (Specificity protein 1) | Regulation of Nanog transcription | (Wu and Yao, 2006) |
| AP2 (Activator protein 2) | Regulation of the trophoectoderm, potential Nanog inhibitor | (Auman et al., 2002) |
| CTCF (CCCTC Binding factor) | Chromatin insulator, interacts with Nanog | (Apostolou et al., 2013) |
| C/EBP (CCAAT enhancer binding proteins) | Transcriptional activators | (Kovacs et al., 2003) |
| IRF (Interferon regulatory factor) | Antiviral defence, immune response and cell growth regulation | (Paun and Pitha, 2007) |
| HMGA (High mobility group AT) | Promoting the alteration of enhancer/promoter architecture to enhance transcription | (Ozturk et al., 2014) |
| TEF-1 related (TEAD, TEA domain transcription factor) | Participates in the Hippo pathway and activates transcription | (Meng et al., 2016) |
| ZEB1 (Zinc finger E-box binding homeobox 1) | Participates in epithelial to mesenchymal transition (EMPT) | (Zhang et al., 2015) |
| SMAD | ESC self-renewal | (Xu et al., 2008) |
| KLF4 (Kruppel like factor 4) | Transcription repressor/activator, one of the four OSKM factors | (Takahashi and Yamanaka, 2006) |
| KLF6 (Kruppel like factor 6) | Embryo development in drosophila | (Weber et al., 2014) |
| ATF-2 (Activating transcription factor 2) | Stimulates CRE-dependent transcription, acts as a histone acetyltransferase at the same time | (Bhoumik and Ronai, 2008) |
| ING4 (Inhibitor of growth family member 4) | Interacts with histone acetyltransferase complexes, capable of interacting with chromatin and transcription factors | (Mathema and Koh, 2012) |

**Table 4.3. Putative binders of the murine Nanog promoter predicted with the TRANSFAC analysis that are also found in the *axolotl* oocyte transcriptome and could potentially be precipitated in the DNA pull down.**

### 4.4.2 Material preparation and optimization

To assess for the factors binding in vivo to the murine Nanog promoter, the genetic sequence analysed for transcription factor binders in silica was incubated in *axolotl* oocyte extract. Prior to its incubation, the genetic sequence was amplified via standard End Point PCR. The region that was amplified by the primers was 300bp long.

a Forward Nanog Promoter
ACTCAGATCCCCCACTTGAC

Reverse Nanog Promoter
CTACCCACCCCCTATTCTCC

b
TGGAAGAGGAA**ACTCAGATCCCCCACTTGAC**CTGAAACTTCCC
ACTAGAGATCGCCAGGGTCTGGAGGTGCAGCCGTGGTTAAAAG
ATGAATAAAGTGAAATGAGGTAAAGCCTCTTTTTGGGGGGGGG
GGATGTCTTTAGATCAGAGGATGCCCCCTAAGCTTTCCC**TCCC**
**TCCCAGTCTGGGTCACCTTACAGCTTCTTTTGCATTACAATGT**
**CCATGGTGGACCCTGCAGGTGGGATTAACTGTG**AATTCACAGG
GCTGGTGGGGCGTGGGTGCCGCCTGGGTGCCTG**GGAGAATAGG**
**GGGTGGGTAG**GGTAGGAGGCTTGAGGGGGGAGGAGCAGGACCT
ACCCTT

**Figure 4.4. a) Forward and Reverse primer for the murine Nanog promoter. b) The 350 nucleotide long murine Nanog promoter region used for Transcription Factor binding analysis. The underlined sequences in bold black represent the forward and reverse primer regions respectively. The underlined sequence in red represents the Nanog80 locus equivalent.**

The 5' end of the primers used for the amplification (Fig. 4.4) was biotinylated. The biotin mark was exploited to conjugate the PCR products on streptavidin coated resin. The mNanog promoter-streptavidin bead conjugate was used to precipitate mNanog promoter specific interactors.

**Establishing optimal PCR product purification and quantification methods**

After the PCR amplification, the resulting PCR reaction was purified to retrieve the DNA product. The PCR product purification step is particularly important and the method that ensures high purity and maximum retrieval should be determined and used. In my case, three different methods were tested and compared and the one generating the highest yield of pure PCR product was chosen. It has to be noted that along with the preferred PCR purification method, the optimal DNA concentration measuring method was determined and applied in all DNA quantifications to minimise errors.

The first method used for the purification of DNA was the Silica Bead DNA Gel Extraction Kit from *ThermoScientific* (K0513), where the PCR reaction was first run on agarose gel and the desired band was subsequently cut and subjected to the purification protocol. The purified DNA was measured with Nanodrop. The kit yielded low concentration of DNA, suggesting that for the purification of sufficient amounts of DNA the procedure had to be repeated numerous times, which from a practical point of view was not desirable.

The second method employed was the ethanol precipitation protocol which directly purifies DNA from the PCR reaction by omitting the step of agarose gel resolution. The purified DNA was quantified with Nanodrop and Qubit Fluorimetry. There was a discrepancy between the two quantification methods with Nanodrop resulting in 10x the quantification given by Qubit. The yield was higher compared to the previous method. However the high concentration quantified by Nanodrop suggested that the ethanol DNA precipitation method did not only purify the relevant PCR product, but precipitated dNTPs and primer remnants etc. The Qubit Fluorimetric Quantitation specifically measures double stranded DNA in a solution, thus avoiding “background” quantitation of single stranded primers that might have been co-purified with the DNA product of interest, or even non-incorporated dNTPs that might interfere and obscure accurate quantification. Nanodrop on the other hand is a less stringent quantification method that

might allow for single stranded primers and dNTPs to be quantified along with the double stranded DNA PCR product. The high quantification demonstrated by Nanodrop was therefore indicative of precipitation not restricted to double stranded DNA (PCR product). Ethanol DNA precipitation in our case would prohibit the saturation of streptavidin beads with the mNanog promoter not only by overestimating the DNA concentration but by including 5' biotinylated primers in the purified PCR product. The biotinylated primers would antagonise with the mNanog promoter in streptavidin binding, thus prohibiting the saturation of the beads with the relevant genetic sequence. This would interfere with the subsequent steps of the experiment and would result to low amounts of mNanog promoter specific binders.

Finally, the third method tested relied on the protocol supplied with the commercially available kit for Gel and PCR clean-up by *Mackerey-Nagel* (740609). The amplified DNA was purified directly from the PCR reaction. The DNA was measured via Qubit Fluometry and Nanodrop to assess for co-purified irrelevant material. In this case both methods resulted in similar quantification indicating that only double stranded DNA was purified. The latter DNA precipitation method also yielded high amounts of DNA, rendering itself the preferred method for PCR product purification.

To have an accurate DNA quantitation is not a trivial matter as an inaccurate quantitation would interfere with the downstream application if DNA-bead conjugation. The saturation of the beads with the DNA sequence of interest is important as there is a potential analogy between the amount of DNA bound to the beads and the amount of precipitated binding factors. Lower amounts of DNA bound to the resin could compromise the precipitation of significant binding factor amounts and therefore hinder their identification potential via mass spectrometry.

### 4.4.3 Initial attempt of DNA pull down in *axolotl* oocyte lysate

The purified DNA was then conjugated to magnetic Streptavidin beads (Dynabeads M-280 Streptavidin, *ThermoScientific*). To capture mNanog promoter specific binders the bead-DNA conjugate was incubated in *axolotl* oocyte extracts. The incubation of the bead-DNA conjugate with the cell extract was carried out in the presence of a blocker compound (poly-dIdC), which by mimicking the double-stranded conformation of the DNA, is capable of attracting unspecific binders.

An important issue to address here is that of the control. The scientific control for a DNA pull down experiment would rationally be a random DNA sequence that would be conjugated to a resin in the same manner as the sequence of interest. The control and experimental sequences would then be incubated in cell extract and the binding molecules would be identified by mass spectrometry. The molecules identified as binders of the control sequence would be subtracted from the identified binders for the genetic sequence of interest, thus revealing the specific binders for the region of interest.

However that is not the case for DNA pull down. Research teams reporting findings via DNA-pull down do not use such a control (Beyer et al., 2013, Zhang et al., 2008, Jutras et al., 2012). The reason for avoiding such a control is that DNA binding proteins demonstrate specificity to small sequences, sometimes composing of only a few bases. It is therefore possible that a random DNA strand could harbour loci attracting DNA binding molecules that are at the same time specific binders of the genetic sequence of interest. That way one could end up excluding as irrelevant molecules that are in fact specific.

In experiments were the primary interest is to study the binding preferences of a known protein, sequence mutation can serve as a confirmatory tool to prove the binding of the given factor to the respective genetic locus (Beyer et al., 2013). Additionally, in experiments where the binding profiles of a given sequence are compared between different cell lysates, the resulting protein IDs can be quantified. Quantification can reveal more or less specific binders,

while disregarding factors that uniformly bind the given sequence in both cases, which is potentially the case with unspecific binders. The resin used to immobilise the bait (DNA or protein) for protein precipitation experiments can often times attract unspecific binders. Proteins precipitated as a result of resin binding are therefore termed “beadome”. To screen for such background binding and confirm that additional proteins are binding to the genetic sequence of interest, naked streptavidin beads (without mNanog promoter) were included in the experiment to assess for molecules precipitating with the streptavidin coated magnetic beads rather than the DNA sequence. To assess the results of this initial attempt, the factors purified as a result of the DNA pull down were visualised on SDS-PAGE with subsequent silver-staining (Fig. 4.5).

The Input and Flow through lanes represent *axolotl* oocyte lysate prior and after the DNA pull down with the murine Nanog promoter. There was no qualitative difference between the input and flow through Ax. oocyte lanes, potentially because the amount of DNA-conjugated beads was not sufficient to cause the depletion of any protein. There was a profound difference in the profiles of the mNanog promoter specific and the beadome background binders. This rendered the DNA pull down experiment particularly promising.

To obtain identities for the mNanog promoter specific binders, the bands present in the respective lane of Fig. 4.5 were cut and prepared for mass spectrometry. A total of 12 bands were picked from the gel, however when mass spectrometry analysis was attempted with the band abundant in protein (as judged by the staining results), no identity was obtained indicating scarcity of protein in the respective bands.

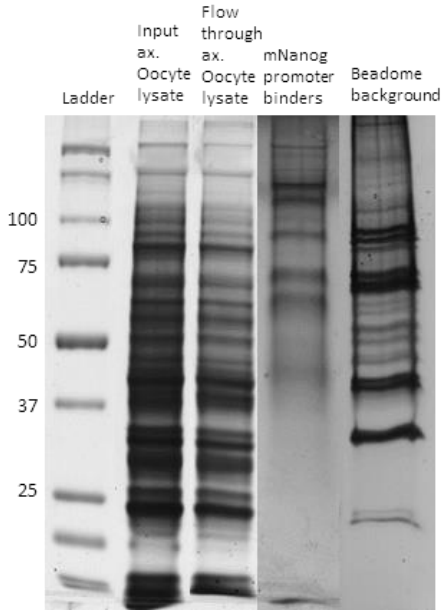



**Figure 4.5. Silver stained image of the first DNA pull down attempted. *Axolotl* oocyte was used in this case. The input and flow through lysates were used to view depletion of certain bands after the pull down. There are colour differences between the bands as they had to be separated over the course of silver staining to prevent overstaining and also allow for fainter bands to be visualised.**

### 4.4.4 DNA pull down with mESC and mOct4 detection

After the initial *axolotl* oocyte lysate DNA pull down, I decided to use mESC lysate to optimise the conditions and volumes/amounts needed for protein identification. Additionally, I was interested in replicating the experiment conducted by Beyer et al in hESC (Beyer et al., 2013). The volumes and amounts used in this case were the same described by the Beyer research team (Beyer et al., 2013). For this experiment I opted for western blotting the resulting purified proteins after the DNA pull down for the presence of murine Oct4 amongst them. According to Beyer et al (Beyer et al., 2013), who did the same experiment using the human Nanog promoter and hESC lysate, Oct4

was one of the molecules that was purified as hNanog promoter specific. I therefore ventured to look whether murine Oct4 is pulled down in murine ESC extracts.

The experiment composed of the mNanog promoter along with a beadome control pull down. The purified proteins were resolved on SDS-PAGE, followed by a western blot (Fig. 4.6).

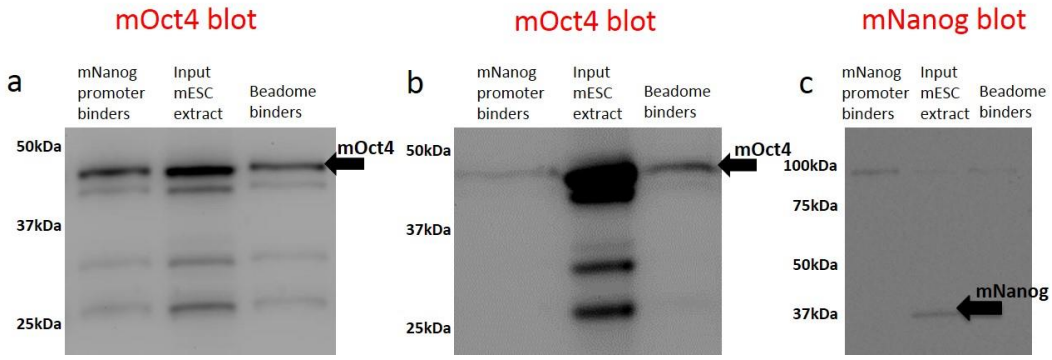



**Figure 4.6. Western blot images for the detection of mOct4 and mNanog after a DNA pull down experiment with the murine Nanog promoter, in comparison with unspecific bead binders. a) First attempt to western blot for mOct4. The antibody detects mOct4 in the expected molecular weight, however additional bands can be seen below the main band. The signal of the main and additional bands is uniform between the promoter specific and beadome binders. b) Second attempt, after optimisation, to detect mOct4. The additional bands have disappeared form the pull down lanes, however the additional bands are still present in the non-treated input lane. c) The blot seen in b) stripped and probed for mNanog. The antibody successfully detects the respective protein in the input sample; however an additional band is seen above the main band which is more prominent in the pull down lanes.**

The expected molecular weight of murine Oct4 is 45-52kDa, as specified by the manufacturer of the antibody used for the detection of mOct4. As seen in Fig. 4.6a, even though the antibody successfully detected mOct4 in the predicted size, additional bands can be seen below the mOct4 band. From the data available on the commercial website (Santa Cruz sc-8628), fainter bands below the main band are visualised without any further information as to what they represent. Strikingly the profiles of the three lanes are very similar, with the bands seen in the pull down lanes being fainter. There is also minimal differentiation in the band intensities between the mNanog promoter and beadome lanes. The expected result would be to see the mOct4 specific band appear only in the case of the mNanog promoter lane, as the findings of Beyer

et al (Beyer et al., 2013) suggested. The uniformity of the band profiles and signal intensities in the pull down lanes (Fig. 4.6a) is indicative of either background binding or protein carry-over, rather than specific binding in any of the pull down cases (mNanog promoter or beadome).

I reasoned that the use of the same Eppendorf tube could potentially be causing protein carry-over from one step of the experiment to the next one. Therefore, I opted to repeat the DNA pull down, this time renewing the Eppendorf tubes after each bead treatment (lysate incubation and bead washes), without altering the actual pull down protocol. The samples were then resolved on acrylamide gel, followed by western blotting. The results of this attempt are depicted in Fig. 4.6b. The carry-over in the pull down lanes was significantly minimised, if not depleted. The mOct4 and additional bands still appeared in the input lane, but only the main mOct4 band could be seen in the pull down lanes. The signal intensity was higher for the beadome control pull down, suggesting that the mOct4 band is a result of unspecific binding.

Nanog has also been reported to bind its own promoter and I therefore decided to look for the presence of Nanog between the purified molecules (Fig. 4.6c). The mNanog specific antibody (Peprotech 500-p236) signals a band of 37kDa in the input lane, which corresponds to the expected molecular weight for murine Nanog (Fig. 4.6c). The mNanog promoter and beadome lanes do not demonstrate any relevant band, except for a faint band at the molecular weight of 100kDa (Fig. 4.6c).

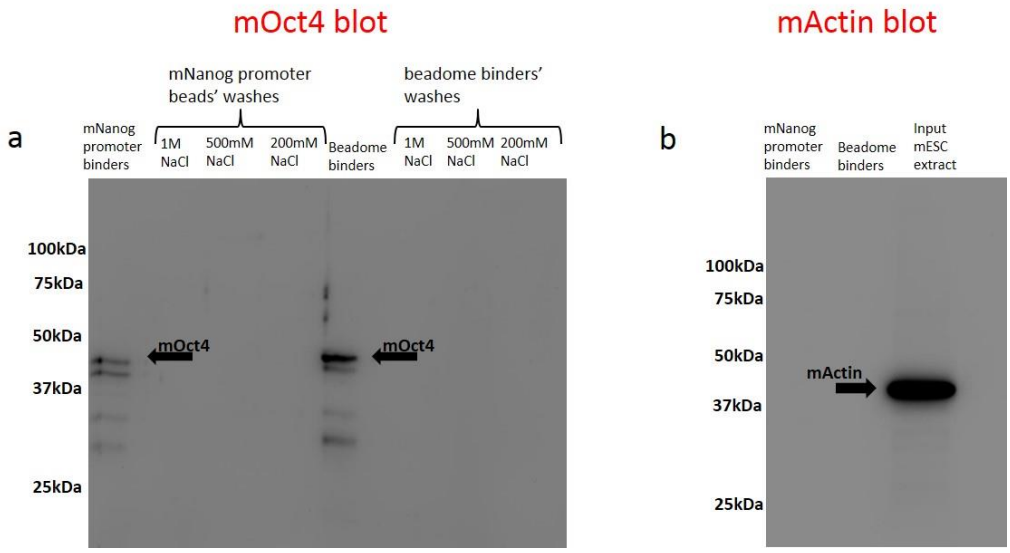



**Figure 4.7. Western blot images of mOct4 and mActin from the newly implemented pull down protocol. a) In this case washes of high, medium and low NaCl concentration were applied to the beads to sequentially remove background/unspecific binding. b) Western blot for mActin in the purified material after the pull down to check for homogeneous background binding.**

After the results seen in Fig. 4.6, I decided to further experiment with the bead/cell lysate protein concentration analogies and searched for alternative DNA pull down protocols in the available literature. After finding the paper by Jutras et al (Jutras et al., 2012), I opted to follow their protocol and incorporate 1M, 500mM and 200mM NaCl washes to sequentially remove unspecific binding and look at the binding affinity of the mNanog promoter binders. The different NaCl concentrations would test the affinity of the binders. The low NaCl concentration would remove binders of lower affinity, while the medium and high salt concentrations would wash off molecules of respective affinities. In the western blot for mOct4 (Fig. 4.7a), no mOct4 band was observed in the sequential NaCl washes. The mOct4 band however, was still predominant in both pull down lanes (promoter and beadome) suggesting that we are dealing with a rather persistent unspecific binder. Interestingly, the signal of mOct4 for the beadome lane was more intense compared to the mNanog promoter pull down. Western blotting for mActin was carried out in an attempt to see whether more prevalent cell proteins behaved in a manner similar to mOct4 (Fig. 4.7b). A mActin specific band is only detected in the

non-treated input mESC lysate, but not in the pull down lanes, suggesting that the mOct4 profile is not related to generic background binding (Fig. 4.7b).

### 4.4.5 Determination of the protein band identities through mass spectrometry

After the western blot experiments presented previously, I conducted two separate DNA pull down experiments with the mNanog promoter and beadome control, in mESC lysate. Since I was not able to validate the DNA pull down experiment by reproducing the results obtained by Beyer et al (Beyer et al., 2013), I decided to focus on determining that the DNA pull down was reproducible. My objective was to assess the reproducibility of the DNA pull down in terms of SDS-PAGE protein band profile. The results (Fig. 4.8) showed that the band profiles obtained were reproducible; however the intensity differed between some bands.

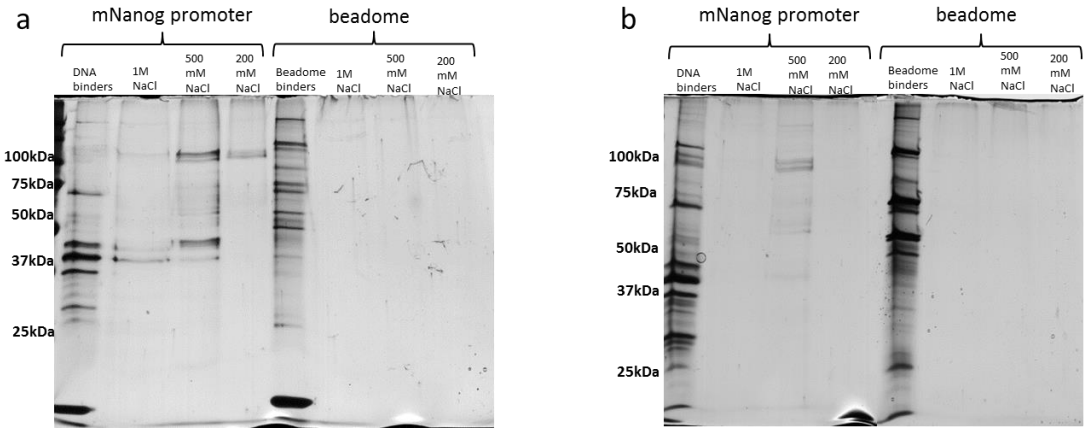



**Figure 4.8. Silver stained SDS-PAGE gels of mNanog promoter and beadome pull down in mESC lysate. Each image represents a different set of experimental replicates. The experiment was conducted twice to test the reproducibility of the band profiles on the SDS-PAGE gels.**

The gels in Fig. 4.8 were silver stained. Bearing in mind the sensitivity of the method in visualising low amounts of protein, I dissected the most prominent bands from both gels to obtain their protein identities (Fig. 4.9). The same bands were picked from both gels, and the equivalent bands were merged to enhance the protein amount and chance of protein identification. Seven and six bands were dissected from gels a and b in Fig. 4.9, respectively. Band G was not found in gel b and was only dissected from gel a. The proteins identified are presented in Table 4.4.

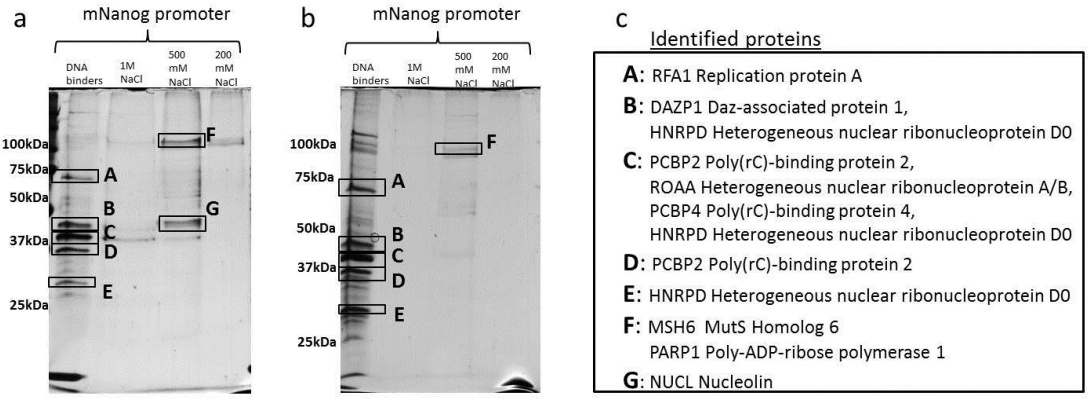



**Figure 4.9. Dissected protein bands for mass spectrometry analysis, and their respective protein identities. a) and b) represent two replicates for mNanog promoter pull down in mESC lysate. The gels can also be seen in Fig. 8. The black rectangles in a) and b) mark the bands picked from mass spectrometry identification. c) The proteins identified in the respective band.**

| Proteins identified after mass spectrometry and their function | | |
|---|---|---|
| **Factor** | **Role** | **References** |
| Nucleolin (C23) | Facilitator of chromatin transcription | (Angelov et al., 2006, Tajrishi et al., 2011) |
| PCBP2 (PolyRC Binding Protein 2) and PCBP4 (PolyRC Binding Protein 4) | RNA binding proteins | (Leidgens et al., 2013) |
| ROAA (Heterogeneous Nuclear Ribonucleoprotein A/B) | RNA binding proteins, potentially associated with neural development | (Sinnamon et al., 2012) |
| HNRPD (Heterogeneous Nuclear Ribonucleoprotein D) | RNA binding protein | (Dempsey et al., 1998) |
| RFA1 (RPA1, Replication Protein A1) | DNA repair, replication and recombination | (Lin et al., 1998, Aboussekhra et al., 1995, Sleeth et al., 2007) |
| DAZP1 (Deleted in Azoospermia associated protein 1) | Germ cell development | (Kee et al., 2009) |
| MSH6 (Muts Homolog 6) | DNA mismatch repair | (Edelbrock et al., 2013) |
| PARP1 (Poly-ADP-ribose polymerase 1) | Catalyses PARyliation, associated with DNA repair and gene expression modulation | (Jiang et al., 2015) |

**Table 4.4. The factors identified after mass spectrometry analysis and their respective role.**

### 4.4.5 Comparison between *axolotl* embryo and oocyte specific mNanog promoter binders

Having established a reproducible DNA pull down protocol capable of precipitating protein amounts sufficient for mass spectrometry identification, I next decided to compare the binders of mNanog promoters between *axolotl* embryo and oocyte. This was a small scale experiment, containing only a single replicate for the embryo and oocyte DNA pull downs. Prior to conducting a large scale experiment with multiple replicates of both embryo and oocyte DNA pull down samples, I decided to run a much simpler version to roughly see how differently the two samples (oocyte and embryo) would behave in relation to mNanog promoter. Additionally, in this case I was aiming to identify the entire network of proteins precipitating with mNanog in each case (oocyte and embryo). Therefore instead of dissecting bands from the SDS-PAGE gel, this time I would analyse the entire SDS-PAGE lane. I was thus introducing a new treatment to the experiment and a small scale experiment would hint potential alterations that would be needed to improve protein identification and differentiation between the DNA pull down of the embryo and oocyte. The experimental workflow remained the same; however this time along with the *axolotl* oocyte DNA pull down, an *axolotl* embryo DNA pull down was also carried out. The oocyte possesses an assortment of maternal proteins as well as proteins generated from maternal mRNA. The blastula stage embryos that were used in the experiment however have undergone MBT (Mid Blastula Transition) where the maternal mRNA was degraded and embryonic gene expression was initiated. Therefore it is very likely that the embryo might harbour additional gene expression regulators. The precipitated proteins and sequential NaCl washes were analysed on SDS-PAGE gel (Fig. 4.10).

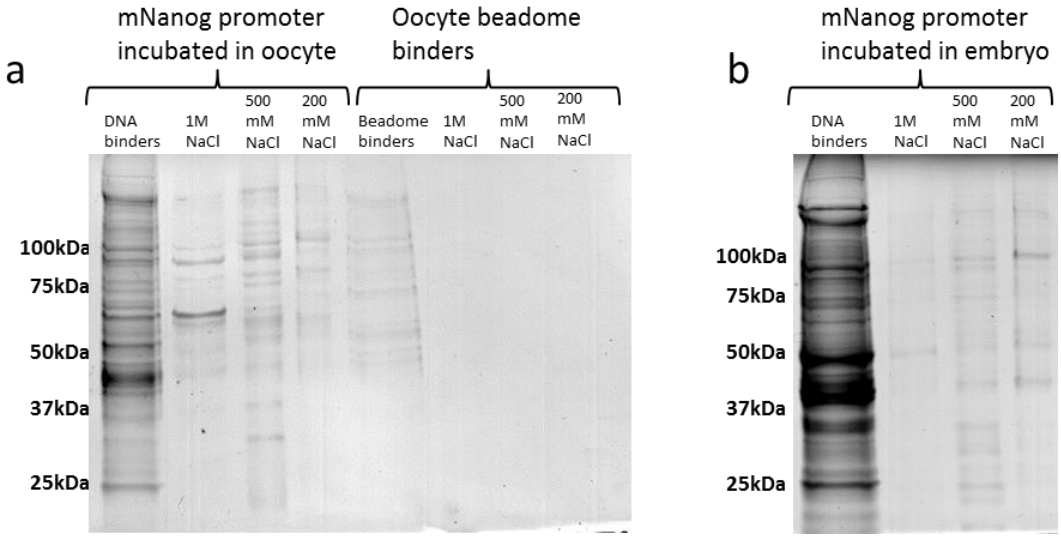



**Figure 4.10. Coomassie stained SDS-PAGE gel comparing oocyte versus embryo specific binders of mNanog promoter in DNA pull down experiments. The beads were washed sequentially with different NaCl concentration buffers prior to the resolution of the binders on the acrylamide gel. a) mNanog and beadome pull down in *axolotl* oocyte lysate. b) mNanog pull down in *axolotl* embryo lysate.**

Due to *axolotl* embryo lysate scarcity, a beadome binding control was only carried for the oocyte lysate. The beadome profile was different compared to that of mNanog binders (Fig. 4.10a). The embryo DNA pull down precipitated higher amounts of protein compared to the DNA pull down carried out with oocyte lysate (Fig. 4.10a and b). Because a large number of bands are present in each mNanog promoter binders' lane, it is difficult to conclude whether the two profiles contain equivalent proteins. In this case instead of single bands, the entire DNA binders' lane (oocyte and embryo specific as well as beadome binders) was dissected and analysed through mass spectrometry.

The mass spectrometry analysis identified 721 and 445 protein IDs in the cases of *axolotl* embryo and *axolotl* oocyte mNanog promoter DNA pull down, respectively. 286 proteins overlapped, while 59 and 434 proteins were identified as *axolotl* oocyte and embryo exclusive binders (Fig. 4.11). Transcription factors, chromatin associated factors and histones were identified as mNanog promoter binders in *axolotl* oocyte and embryo (Table 4.5).

a




b

| Fraction | No of proteins |
|---|---|
| Embryo only | 434 |
| Oocyte only | 59 |
| Embryo-Oocyte overlap | 286 |

c

| | Embryo | Oocyte | Common |
|---|---|---|---|
| Transcription factors | SMARCE1, MEIS2, SMARCC1, NRF1, STAT1, SSRP1 | YBX2 | YBX1, DNAJC2 |
| Chromatin remodellers | RUVBL1, RBBP7 | | SMARCA5, SMARCA4, RBBP4 |
| Nucleosome assembly | CHAF1B | CHAF1A | NAP1L1, NAP1L4 |
| Histones | H1H4i, H2H2aa1, H1H2al, H1H2bn, H2H3c1, H1H3h, H1H2bo | H1H2ad | H1H2ao, H1H2ab |
| Additional factors | DNMT1, UHRF1 | UHRF2 | RARP1 |

**Figure 4.11. Comparison of mNanog promoter binders, for *axolotl* embryo and oocyte lysate. a) Venn diagram showing embryo and oocyte specific, as well as overlapping mNanog promoter binders. b) Numbers of embryo and oocyte specific as well as the oocyte/embryo overlapping proteins. c) Proteins specific to the *axolotl* embryo and oocyte DNA pull down as well as overlapping molecules between the two, grouped according to their molecular function.**

## Oocyte and Embryo mNanog promoter binders

| Transcription Factors | | | |
|---|---|---|---|
| **Factor** | **Role** | **References** | **Oocyte/Embryo/Common** |
| YBX1 (Y Box binding protein 1) | mRNA regulation | (Lyabin et al., 2014) | Common |
| DNAJC2/ZRF1 (Zuotin related factor 1) | Associated with the transcription regulation complex, Mediator | (Papadopoulou et al., 2016) | Common |
| YBX2 (Y Box binding protein 2) | Germ cell specific, mRNA regulation | (Zheng and Dean, 2007) | Oocyte |
| SMARCE1 (SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily E) | Regulation of Nanog expression in mESC | (Schaniel et al., 2009) | Embryo |
| MEIS2 (Meis homeobox 2) | Neural crest development | (Machon et al., 2015) | Embryo |
| SMARCC1 (SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily C member 1) | Regulation of Nanog expression in mESC | (Schaniel et al., 2009) | Embryo |
| NRF1 (Nuclear respiratory factor 1) | Cell homeostasis, NRF1 knockouts are lethal | (Leung et al., 2003) | Embryo |
| STAT1 (Signal transducer and activator of transcription 1) | Forms heterodimers with STAT3, which supports self-renewal in mESC | (Durbin et al., 1996) | Embryo |
| SSRP1 (Structure specific recognition protein 1) | Subunit of the FACT complex, SSRP1 deficient murine embryos are lethal | (Cao et al., 2003) | Embryo |

| Additional factors | | | |
|---|---|---|---|
| **Factor** | **Role** | **References** | **Oocyte/Embryo/Common** |
| DNMT1 (DNA methyltransferase 1) | Maintenace of DNA methylation | (Hermann et al., 2004) | Embryo |
| UHRF1 (Ubiquitin like with PHD and Ring finger domains 1) | Binding of hemimethylated DNA | (Liu et al., 2013) | Embryo |
| UHRF2 (Ubiquitin like with PHD and Ring finger domains 2) | Recognition of 5hmC DNA modification | (Liu et al., 2016) | Oocyte |
| PARP1 (Poly-ADP-ribose polymerase 1) | Catalyses PARyliation, associated with DNA repair and gene expression | (Beck et al., iang et al., 2015) | Common |

| Chromatin and nucleosome associated factors | | | |
|---|---|---|---|
| **Factor** | **Role** | **References** | **Oocyte/Embryo/Common** |
| CHAF1A and CHAF1B (Chromatin assembly factors 1A and 1B) | CAF-1 subinit, nucleosome assembly with H3/H4 histones, associated with inaccessible chromatin | (Cheloufi et al., 2015) | Oocyte and Embryo, respectively |
| NAP1L1 and NAP1L4 (Nucleosome assembly protein 1 like 1 and like 4) | Nucleosome assembly with H2A/H2B histones, chromatin remodelling | (Chang et al., 1997, Ito et al., 2000) | Common |
| RUVBL1 (RuvB like AAA ATPase 1) | Associated with chromatin remodelling complexes | (Jha et al., 2008, Mizuguchi et al., 2004) | Embryo |
| RBBP4 and RBBP7 (Retinoblastoma binding protein 4 and 7) | Transcriptional repressors as members of the NuRD complex, histone chaperones | (Kloet et al., 2015, Balboula et al., 2015) | Oocyte and Embryo, respectively |
| SMARCA4 and SMARCA5 (SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily A, members 4 and 5) | Subunits of the SWI/SNF chromatin remodelling complex, which assists ESC self-renewal | (He et al., 2012, Kidder et al., 2009) | Common |

**Table 4.5. Table summarising the Transcription Factors and chromatin associated factors that were identified in the *axolotl* oocyte and embryo mNanog promoter DNA pull down.**

## 4.5 Conclusions

DNA pull down was chosen as the method to identify factors that specifically bind the murine Nanog promoter. Mass spectrometry is the de facto tool that is involved in all studies aiming to identify proteins in a given sample or after a line of experiments. DNA pull down is highly advantageous for enriching the specific binders and therefore enhancing protein identification rates with mass spectrometry, where protein abundance and purity are beneficial.

The Nanog promoter concentrates characteristics that render it a very interesting substrate for a DNA pull-down experiment as it is a) a locus with many putative binding sites (Beyer et al., 2013, Papp and Plath, 2011), b) a locus that is epigenetically remodelled over the course of reprogramming (Papp and Plath, 2011), c) involved in the regulation of pluripotency by being implicated in chromatin interactions (Papp and Plath, 2011).

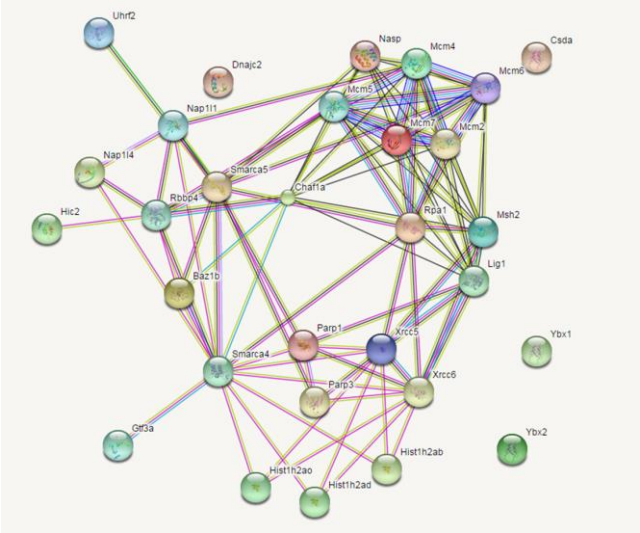
In the present chapter I attempted to repeat an experimental procedure conducted in human stem cell by Beyer et al (Beyer et al., 2013). While using the murine Nanog promoter sequence that was equivalent to that of Beyer et al, I optimised the volumes and protein amount for mass spectrometry protein identification using mouse stem cell lysate and subsequently conducted a single replicate comparison between *axolotl* embryo and *axolotl* oocyte mNanog promoter binders. Upon mES lysate optimization, I was not able to replicate mOct4 binding to the murine Nanog promoter in the manner that Beyer et al (Beyer et al., 2013) reported human Oct4 to bind Oct4 promoter in human stem cells. In my case Oct4 appeared to bind the magnetic streptavidin beads as a background binder. The comparison between *axolotl* embryo and oocyte binders of mNanog promoter would potentially highlight differences in the mNanog binding molecules in each case. The oocyte has a depot of maternal factors that facilitate the initial stages following fertilization, while the blastula stage embryos that were used have initiated zygotic transcription. The single replicate experiment that was carried out identifies an abundance of proteins; however conclusions cannot

be drawn from it. While a large scale experiment comparing the two (embryo and oocyte) was planned, it was not conducted as new research inspired me to change my experimental design and initiate a new project. The experiments and optimization carried out in the present chapter however were useful as I gained experience with DNA pull down. Notably, the same experimental procedure (DNA pull down) was followed with the new project, while an epigenetic perspective was added to it and *axolotl* embryo screening were omitted, focusing only on the oocyte. This will be discussed and presented in detail in the following chapter.

While not many conclusions can be drawn from the present chapter in terms of oocyte and embryo specific mNanog promoter binders, I was still able to see how a DNA fragment is treated upon oocyte and embryo lysate incubation. Upon lysate treatment mNanog promoter appeared to become chromatinized, i.e. wound on histones and this was potentially carried out by the nucleosome assembly factors that were identified (CHAF1A, CHAF1B, NAP1L1, NAP1L4). UHRF1 and DNMT1 were identified as mNanog promoter binders in *axolotl* embryo, potentially reflecting initiation of DNA methylation and its maintenance as these molecules are implicated in DNA methylation maintenance (Liu et al., 2013). PARP1, a parylating protein implicated in DNA repair, was found among the common (oocyte and embryo) binders of mNanog promoter. Interestingly PARP1 has been reported to bind the Nanog promoter over the course of iPS reprogramming with a potential transcriptional activation impact (Doege et al., 2012). Whether PARP1 binding in the present screen is specifically related to mNanog promoter and it activation needs further experiments for a conclusion. However, SMARCE1 and SMARCC1, previously reported to implicate in ESC Nanog expression were also identified as mNanog promoter binders in the oocyte (Schaniel et al., 2009). When the identified binders of mNanog promoter in the oocyte and embryo were assayed for potential network associations with the protein network tool STRING (available online), well connected network were revealed in both cases. This potentially indicates that the precipitated factors

not only interact with the mNanog promoter, but interact with each other as well (Fig. 4.12).

**a**




**b**

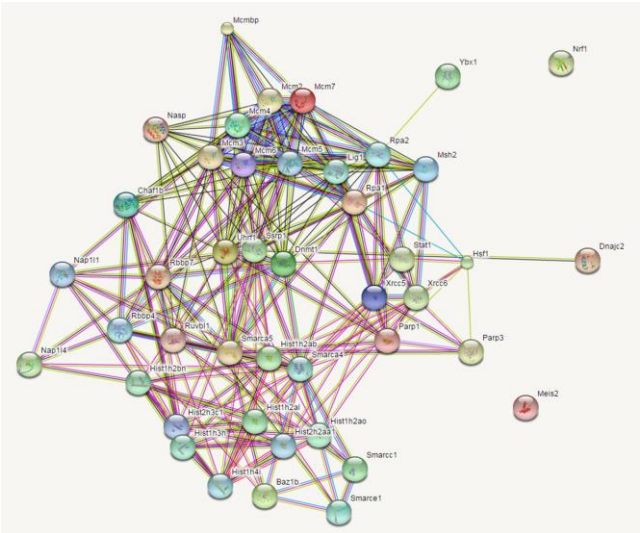



**Figure 4.12. a) Protein network of identified mNanog promoter binding factors in *axolotl* oocyte. b) Protein network of identified mNanog promoter binding factors in *axolotl* embryo. Both images were prepared with the online protein network tool STRING.**

# Chapter 5: Identifying the oocyte specific demethylation mechanism and factors

## *5.1 Introduction*

### 5.1.1 5hmC, 5fC and 5caC: 5mC oxidation products and epigenetic markers

Alternative cytosine modifications were identified recently, offering a deeper insight into the complex field of epigenetic transcriptional regulation. 5-hydroxymethylcytosine (5hmC) was identified as the product of the oxidation of 5-methylcytosine (5mC) (Tahiliani et al., 2009). Furthermore, the modifications 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) were identified as the oxidation products of 5hmC and 5fc respectively (Ito et al., 2011). The initial perception of the new cytosine modifications was that they served as demethylation intermediates, as 5fC and 5caC could suffice as substrates for TDG (Thymine DNA Glycosylase) (Maiti et al., 2013). The base-excision-repair mechanism then repairs nicks in the DNA created by TDG substituting the excised cytosine with its non-methylated version, thus forming a DNA demethylation pathway (Wu and Zhang, 2014).

### 5.1.2 TET enzymes and their activity

The Ten-Eleven-Translocation proteins convert 5mC into 5hmC, 5fC and 5caC in consecutive steps (Pfaffeneder et al., 2011, He et al., 2011, Ito et al., 2011, Munzel et al., 2010, Ito et al., 2010, Tahiliani et al., 2009, Kriaucionis and

Heintz, 2009). TET enzymes were primarily associated with active demethylation of 5mC through its oxidation to the sequential intermediates 5hmC, 5fC and 5caC. The oxidative intermediates are three additional cytosine modifications and therefore might play a role in gene regulation. Recent findings have demonstrated that 5hmC and 5fC are stable modifications in murine DNA (Bachman et al., 2014, Bachman et al., 2015). This evidence challenges what was proposed initially for the additional cytosine modifications in the context of oxidative demethylation. 5hmC and 5fC are now proposed as stable epigenetic modifications, products of 5mC, but with potentially stand-alone significance.

Studies on the kinetics of TET enzymes delineate their catalytic activity and also provide indications on how oxidative demethylation is carried out. The main question behind this research was whether the oxidative activity of the TET enzymes is processive or distributive. Processive oxidation denotes that the TET enzyme remains bound to DNA through the sequential oxidation steps and is released when 5caC is formed. On the other hand in the distributive oxidation model TET enzymes are released after each oxidation step. This processive mode of action for the TET family was demonstrated by Crawford et al (Crawford et al., 2016) who showed that murine TET2 remained bound on methylated DNA through the formation of 5caC. The distributive oxidation model has been supported by Xu et al (Xu et al., 2014) and Tamanaha et al (Tamanaha et al., 2016). The first group studied the catalytic domain of murine TET1 while the latter group studied murine TET1, TET2 as well as a distant TET1 homologue, NgTET1 from the single cell organism *Naegleria gruberi* (Tamanaha et al., 2016). A distributive mode of action corroborates the regulative role of the oxidation derivatives. It was also reported that TET enzymes after a modified cytosine turnover, release the oxidized product and dissociate from DNA in order to find a new substrate (Tamanaha et al., 2016).

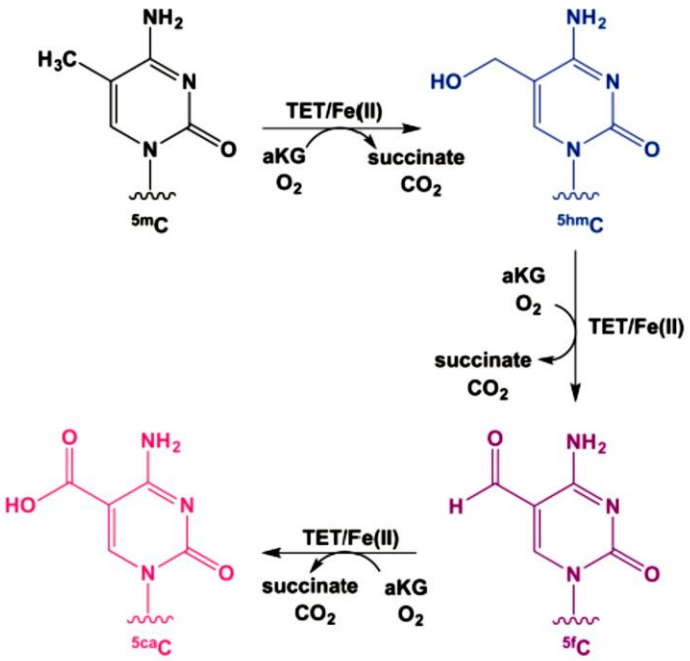



**Figure 5.1. Oxidation of 5mC by the TET enzymes. In each oxidation step, aKG is converted to succinate and $CO_2$.**

### 5.1.3 Readers and epigenetic significance of the cytosine modifications

As discussed previously, with distributive oxidation 5mC is sequentially oxidised and TETs are released after each turnover. After its oxidation the cytosine is released and if additional oxidation is to follow the TET enzyme has to bind to the base again and carry out a single oxidation step. The alternative model, suggesting that TETs act in a processive manner, indirectly suggests that the 5mC oxidation intermediates are transient modifications, only serving as passage points between 5mC and 5caC. In a processive mode of action TETs would not release the oxidised derivatives of 5hmC and 5fC and these would not be exposed to the nuclear environment, they would thus bear no epigenetic significance. With the distributive model, however, additional roles for the oxidation derivatives are not excluded. The exposure of 5hmC, 5fC and

5caC to the nuclear environment suggests their participation/involvement in different biological pathways.

The discovery of additional cytosine modifications triggered the research of the 5hmC/5fC/5caC binders. The interactome of the cytosine modifications could delineate the role each one plays in the demethylation pathway and also reveal the factors that participate in or trigger demethylation. Among the factors identified as 5mC/5hmC/5fC/5caC binders, factors potentially associated with proposed DNA demethylation models arose, however the identification of additional molecules corroborated the hypothesis of a cytosine epigenetic code.

MBD family proteins have been found to extensively bind 5mC. Examples include MeCP2, MBD4 (Spruijt et al., 2013, Bartke et al., 2010), MBD1 (Spruijt et al., 2013), and MBD2 (Bartke et al., 2010). The aforementioned proteins are directly associated with transcriptional repression and nucleosome remodelling (Hendrich and Bird, 1998, Du et al., 2015).

Other 5mC interactors include UHRF1 (Spruijt et al., 2013, Bartke et al., 2010) and KLF4 (Spruijt et al., 2013). UHRF1 is related to the maintenance of the DNA methylation mark by recognising hemimethylated DNA and interacting with the DNA methyltransferase DNMT1 (Bostick et al., 2007, Hashimoto et al., 2009). KLF4, on the other hand, is one of the OSKM factors that when overexpressed in somatic cells can induce the pluripotent character (Takahashi and Yamanaka, 2006).

MBD proteins such as MeCP2 and MBD4, and UHRF1 also bind 5hmC, although not with the same affinity as 5mC (Spruijt et al., 2013). UHRF2 was also identified as a 5hmC binder (Spruijt et al., 2013) while it has recently been associated with the regulation of Epithelial-Mesenchymal Transition transcription factors (Lai et al., 2016). Additional attributes of UHRF2 link it with the histone code (Pichler et al., 2011), while both UHRF1 and UHRF2 interact with DNA repair mechanisms through their association with the MPG (N-Methylpurine DNA Glycosylase) (Liang et al., 2013). MPG, NEIL3, NEIL1 are glycosylases identified as 5hmC binders and along with RECQL, a DNA helicase, further link 5hmC with DNA repair mechanisms (Spruijt et al., 2013).

Glycosylases such as TDG and NEIL1 were shown to bind both 5fC and 5caC, while interestingly p53, a major DNA repair factor and tumour suppressor (Williams and Schumacher, 2016), was found to specifically bind 5fC (Spruijt et al., 2013). In the case of 5caC, DNMT1 and BAF170, a subunit of the SWI/SNF chromatin remodelling complex, were identified as its specific binders (Spruijt et al., 2013).

These findings support the idea that DNA demethylation occurs through oxidation and DNA repair, however the array of molecules identified (as binders of each cytosine modification) also reflects additional levels of functional/regulatory complexity in the context of epigenetics. It is important to note that cytosine modifications are interpreted in relation to the genetic sequence they mark, according to the findings of Iurlaro et al, where the entire NuRD complex was enriched for the 5fC at the Fgf15 promoter but not for the Pax6 promoter (Iurlaro et al., 2013). The presence of the NuRD complex potentially indicates a repressive function for the 5fC mark in the context of 5fC-Fgf15 in ESC. Interestingly, there was an overlap between the 5fC and 5mC binders identified in the case of the Fgf15 promoter (Iurlaro et al., 2013). Additionally, the array of each cytosine modification's binders is affected by cell identity, as reported by Spruijt et al (Spruijt et al., 2013). Iurlaro et al examined the binders of two different promoters in the context of ESC nuclear lysate, while Spruijt et al examined the crosstalk of cell identity with the binding profile of cytosine modifications. Klf4 for example was only identified as a 5mC binder in the context of mESC, but not in Neuronal Progenitor Cells (NPC) or brain tissue (Spruijt et al., 2013). On the other hand Dlx5 and Dlx6, genes expressed in specific regions of the brain and playing a role in brain development were identified as 5mC specific binders only in the context of adult brain tissue (Spruijt et al., 2013). Thap11, reported as highly expressed in Purkinje cells, as well as replication factor C subunits, were identified as 5hmC specific binders in brain tissue (Spruijt et al., 2013). The above findings are a reflection of the dynamic nature of the cytosine modification associated interactomes, impacted by DNA sequence and cell identity.

### 5.1.4 Demethylation in the oocyte

Extensive demethylation takes place on the paternal DNA after fertilization (Reik and Surani, 2015). The male and female gametes enter the cycle of embryogenesis highly methylated and the paternal gamete undergoes drastic demethylation upon fertilization, excluding paternal imprinted genomic loci (Reik and Surani, 2015). The discovery of the 5mC oxidation products triggered research on the presence of such cytosine modifications during zygote formation. The findings would be indicative of the demethylation mechanism residing in the oocyte.

The presence of 5hmC in the paternal genome, in the course of zygote formation, was reported corroborating TET activity in the oocyte (Iqbal et al., 2011, Wossidlo et al., 2011). 5hmC removal however was linked to dilution by DNA replication (Inoue and Zhang, 2011). 5fC and 5caC were also discovered in paternal pronuclei, as further proof of oocyte TET dependent demethylation (Inoue et al., 2011b). The 5fC/5caC modifications increased concomitantly with 5mC reduction and were reported to be relatively stable, removed by dilutive mechanisms similar to 5hmC (Inoue et al., 2011b).

Recent findings provided further insight into the oocyte demethylation mechanisms by reporting that oocyte demethylation is not TET dependent. More specifically, Amouroux et al observed that inhibition of TET3 or 5hmC formation with the use of DMOG (Dimethyloxallyl Glycine) did not affect the loss of the 5mC mark (Amouroux et al., 2016). Additionally, Amouroux et al reported a link between DNMT3A and DNMT1 with 5hmC formation in the oocyte, i.e. the lack of DNMT3A and DNMT1 impacted negatively 5hmC concentration on paternal chromatin after fertilisation (Amouroux et al., 2016). The team also reported that while 5mC loss occurs by early PN3 stage (murine zygote), 5hmC formation was not concomitant with 5mC loss but lagged and increased in PN4. Interestingly, the same phenomenon has been observed by Bachman et al, in the context of cancer cell lines (HCT116 and MCF7) and mES. The team determined that 5hmC is a stable DNA

modification, and reports that 5hmC formation (during DNA replication) requires 2h-4h after the deposition of 5mC (Bachman et al., 2014). According to the observations of Bachman et al 5hmC is a result of 5mC oxidation, however it occurs in a TET1 independent manner, as the absence of TET1 does not impair 5mC conversion into 5hmC (Bachman et al., 2014). However, whether TET2/3 or a TET independent oxidation pathway carries out the formation of 5hmC needs to be further researched. These findings are suggestive of alternative demethylation mechanisms residing in the oocyte and somatic cells. However, the questions regarding the demethylation mechanism employed by the oocyte to reprogram DNA, and the interpretation of the different cytosine modifications by the oocyte remain to be answered.

## 5.2 Aims and objective

Specific demethylation pathways have been reported to operate in the oocyte, while the different cytosine modifications attract binders and potentially have an impact in gene regulation. The oxidative derivatives of 5mC might be of specific significance on pluripotency gene regulatory regions that need to be active during early development. Here, I investigated the demethylation mechanism operating for the activation of the Nanog in the context of the oocyte.

Aims

- To identify oocyte factors specifically binding cytosine modifications
- To delineate the demethylation pathway operating in the context of the oocyte
- To identify pluripotency associated factors that are linked with the demethylation pathway
- To generate candidates for the improvement of iPS

Objectives

- To employ the mNanog promoter, incorporate modified cytosines and conduct a DNA pull down assay in the context of the oocyte
- To employ mass spectrometry for the identification of cytosine specific binders
- To conduct biological assays with the identified factors to delineate their function

## 5.3 Materials and methods

DNA pull down was employed to identify proteins interacting with the epigenetically modified mNanog promoter. The method and materials are presented in detail in Chapter 2, Section 2.11. The primers, PCR reaction and cycling conditions for the incorporation of the modified cytosine residues can be found in Chapter 2, Section 2.10. Amplified mNanog promoter was purified with and as suggested by the manual of the Gel and PCR clean up kit from Mackerey-Nagel (740609). To optimize the digestion conditions for optimal protein identification three different digestion protocols were tested. The digestion materials and methods are described in Chapter 2, Section 2.12. The mass spectrometry analysis in this case was carried out without prior SDS-PAGE analysis. After the DNA pull down, the samples were not stored, but were processed immediately after the final washes. Finally, identification for precipitated proteins was obtained through mass spectrometry with an ABSCIEX TripleTof 6600 mass spectrometer in the Nottingham Trent University facilities. Detailed description of the injection conditions can be found in Chapter 2, Section 2.12. Protein identification was obtained through PEAKS (Waters) and by search against the custom database comprised of the *axolotl* transcriptome (Matt Loose). Label-free quantification was carried out on the PEAKS-Q (Waters). The search parameters and results statistics can be found in the Appendix, Fig. 7.1-7.4 and Tables 7.1-7.3. Notably, the significance threshold for the proteins was set at 15, while only proteins with a fold change equal or above to 2 were listed. A non-modified cytosine sample was set as the reference sample for the label-free quantification.

## 5.4 Results

### 5.4.1 Experimental workflow

In the previous chapter we presented the DNA pull down as a suitable method for identifying DNA binders, associated with a specific genetic locus.

The method was chosen as a tool to identify pluripotency related molecules associated in the activation of the Nanog gene during cell reprogramming to pluripotency. While the DNA pull down method was effective for identifying factors associated with the chosen genetic locus, we decided to broaden the spectrum of experiments and involve epigenetic readers to the screen.

Prior to that decision I based my experimental plan on the publication by Beyer et al (Beyer et al., 2013), where the promoter region of Nanog was assayed for specific binders and genetic mutations of its sub-loci served as a means to validate the binding of the identified factors in the given sequence. After setting up the conditions and protocol for the DNA pull down experiment, two papers were published by Spruijt et al (Spruijt et al., 2013) and Iurlaro et al (Iurlaro et al., 2013), which gave an additional context to my research. By incorporating cytosine modifications to the Nanog promoter locus used in Chapter 4, I could not only screen for mNanog promoter associated binders but also differentiate between binders in each cytosine modification status. In that manner, I could potentially also obtain an insight into the demethylation and/or activation dynamics of Nanog over the course of demethylation and cell reprogramming towards pluripotency.

The methyl marks are relevant to the Nanog promoter because it contains CpG dyads, whose methylation has been reported to negatively impact the expression of Nanog (Deb-Rinker et al., 2005). To screen for binding factors for the alternative cytosine modifications on the mouse Nanog promoter, the same procedure that was described in Chapter 4 was followed. However in this case, along with the version of the mNanog promoter containing non-

modified cytosines, the additional cytosine modifications were incorporated in to the genetic sequence individually (5mC/5hmC/5fC/5caC). The different cytosine modifications were introduced during the course of standard End-Point PCR amplification with the substitution of non-modified cytosine with its modified counterparts. The amplified mNanog promoter was purified and conjugated to streptavidin coated magnetic beads, as described in Chapter 4. The bead-DNA conjugates were subsequently incubated in *axolotl* oocyte lysate and the associated binders were identified through mass spectrometry.

### 5.4.2 mNanog promoter amplification

For the amplification of non-modified cytosine (C), a commercially available End-Point PCR reaction buffer supplied with dNTPs was used. To successfully incorporate the modified cytosines a PCR reaction buffer already supplied with dNTPs could not be used. 5mC and 5hmC dNTP mixes were purchased from Zymo Research, while 5fC-dCTP and 5caC-dCTP were purchased from TriLink. dNTP mixtures were prepared with the latter two replacing dCTP. With this method the substitution of non-modified cytosine with its modified version was close to 100%, excluding the cytosines present on the primers. The same biotinylated primers used in Chapter 4 were used in this case as well and therefore the amplification product was the murine Nanog promoter locus with a 5' biotin mark to be used for the streptavidin bead-DNA coupling.
The 5mC and 5hmC-mNanog promoter sequences were amplified with ZymoTaq DNA polymerase, while 5fC was amplified with the DreamTaq polymerase from ThermoScientific (Fig. 5.2).
I was unable to amplify the 5caC-mNanog promoter with any of the polymerases used. Altering the cycling conditions with different polymerases did not lead to any improvement; therefore I had to exclude the 5caC modification from my screen.

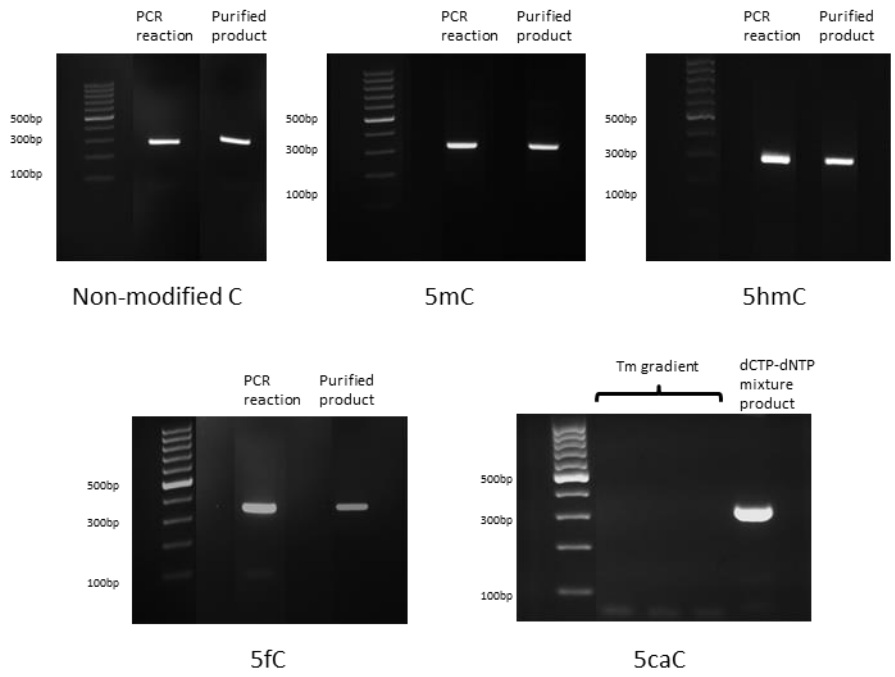



**Figure 5.2. Amplification of mNanog promoter from murine DNA with the incorporation of one of the modified cytosine residues of C/5mC/5hmC/5fC/5caC. The desired product is of 300bp and is presented prior to and after purification from the PCR reaction. For the case of 5caC that was not responsive with any of the polymerases used, the gel shows a Tm gradient for the 5caCTP containing dNTPs mixture whereas the non-modified dNTPs result in the desired product.**

5fC and 5caC-mNanog promoters failed to amplify with the ZymoTaq polymerase. Phusion High Fidelity (NEB) polymerase was also tested in Tm gradients for both 5fC and 5caC, however no improvement was observed. In both cases, however, non-modified dNTP mixtures successfully resulted in mNanog promoter amplification. Potentially the polymerases find the incorporation of some base modifications challenging. Upon testing the DreamTaq polymerase, I found that it was efficient for the case of 5fC, but 5caC remained unresponsive (Fig. 5.2). DMSO and formamide were also added to the PCR reactions (individually) to check if they would lead to amplification, however this did not improve the result. The C/5mC/5hmC/5fC mNanog promoter was purified directly from the PCR reactions with the Nucleospin kit (Mackerel-Nigel).

### 5.4.3 Optimal protein digestion for mass spectrometry

In this case I screened for a wide range of molecules that bind on the DNA locus, therefore I wanted to be able to extract and identify as many proteins as possible through mass spectrometry. The method I employed up to that point was in gel digestion, after I had resolved the DNA pull down eluate in an SDS-PAGE gel and separated the entire lane into several pieces, which were then prepared for mass spectrometry. The above is a standard method; however, alternative methods have been developed and employed in pull down experiments.

I compared three digestion methods for their protein identification rates in mass spectrometry: On bead digestion, Rapigest digestion and the classic in gel digestion. The On bead digestion consists of trypsin being added directly to the beads after the DNA pull down, so that digestion happens while the proteins are still bound on their target DNA and the DNA is still bound on the magnetic streptavidin beads as previously described (Turriziani et al., 2014). Rapigest digestion refers to an in solution digestion in the presence of Rapigest (Waters), a reagent capable of enhancing enzymatic digestion of proteins (Unnikrishnan et al., 2012, Steuble et al., 2010).

To compare the three methods I used the same DNA pull down protocol with the only difference being that I opted for a 15min incubation of the bead-DNA conjugate in the *axolotl* oocyte lysate. The reason I wanted to test the 15min incubation primarily relied to that fact that Iurlaro et al (Iurlaro et al., 2013) had opted for the same incubation time in their cytosine modification interactome screen. Potentially, a lower incubation time in the oocyte lysate could result in more specific interactors and less background binders. To run the comparison between the different digestion protocols non-modified cytosine mNanog promoter was used.

After their preparation, the samples were analyzed with an ESI-QUAD-TOF mass spectrometry instrument as described in Chapter 2, section 2.12. The

obtained peaks where matched against the *axolotl* EST library through MASCOT.

As shown in Table 5.1, the in gel digestion proved to be the weakest in terms of protein identification as it only identified 37 proteins in contrast with Rapigest digestion, which resulted in 121 protein IDs from 60µl of DNA saturated magnetic beads. Unfortunately the on bead digestion sample from the same bead volume was lost during mass spectrometry analysis. For the case of 200µl mNanog promoter coated beads again the in gel digestion resulted in the least amount of identified proteins, Rapigest digestion came next with 206 protein IDs, while the on bead digestion protocol proved to be the most efficient with 564 proteins identified. Comparing the 15min incubation of bead-DNA conjugate in oocyte lysate with the 2hr incubation from previous experiments where the pull down eluate was in gel digested prior to mass spectrometry, it is shown that a longer incubation results in higher identification rates when it comes to in gel digestion. However, the on bead digestion of the 15min incubation resulted in approximately the same number of identified proteins as the in gel digestion of the 2hr incubation.

The numbers seen in Table 13 clearly show that on bead digestion is the most efficient protocol in terms of protein identification; however a 2hr incubation of the bead-DNA conjugate in *axolotl* oocyte lysate will potentially generate more data compared to the 15min incubation. Therefore I decided to combine the on bead digestion protocol with the 2hr incubation.

| **Number of identified proteins** | | | | | |
|---|---|---|---|---|---|
| *15min incubation* | | | | *2hr incubation* | |
| **Digestion protocol** | **In gel** | **On bead** | **Rapigest** | **In gel** | **Digestion protocol** |
| Volume: 60µl | 37 | - | 121 | - | - |
| Volume: 200µl | 142 | 564 | 206 | 523 | Volume: 200µl |

**Table 5.1. Comparison of the total number of proteins identified from the three different digestion protocols and two volumes of magnetic streptavidin beads, saturated with mNanog promoter.**

### 5.4.4 Non-modified C/5mC/5hmC/5fC interactome screen

The purified mNanog promoter bearing either one of the modifications was bound to magnetic streptavidin beads as described in Chapter 4. The beads were subsequently incubated in *axolotl* oocyte lysate and the DNA pull down eluate was prepared for mass spectrometry using the on bead digestion protocol. The samples were processed on a Sciex TripleTof 6600 mass spectrometry instrument. Peptides were matched against the *axolotl* EST library through a search with the PEAKS software. Technical replicates (n=6) were analyzed for each cytosine modification type. A label-free quantification was run with 5 replicates for each cytosine modification type, after the removal of the outlier replicate for each cytosine case. The results of the label-free quantification are shown in the heatmap of Fig. 5.3. The protein IDs of the heatmap, the search parameters and results statistics can be seen in the appendix (Fig. 7.1-7.4, Tables 7.1-7.3). The quantification is based on the relative intensities of peptide features that were detected in the samples. The features of the same peptide across samples are aligned together with the use of the retention alignment algorithm (Lin et al., 2013). The proteins are clustered together if they exhibit a similar expression trend across the samples. Highly expressed proteins are marked in red while a low protein expression is shown in green color. As seen in the heatmap protein enrichment is observed in all the cytosine counterparts assayed. The 5fC-mNanog promoter shows the highest number of binders, while also demonstrating 5fC exclusive binding factors. The 5hmC modification appears to attract the lowest number of binders. Apart from 5fC, the rest of the cytosine modifications do not demonstrate any significant exclusive binding (Fig. 5.3). Highlighted in the heatmap are the proteins CHAF1A and CHAF1B, subunits of the trimeric Chromatin Assembly Factor 1 (CAF-1) complex (CHAF1A, CHAF1B and p50), as well as the subunits of the dimeric FACT (Facilitates Chromatin Transcription) complex, SUPT16H and SSRP1. Additionally highlighted are the XPC complex, composing of the proteins XPC,

CETN2 and RAD23B and the Replication Protein A subunit RPA1 and RPA3. At the same time highlighted are UHRF1 and ERCC6L, which show a specific enrichment for 5mC and 5hmC in the screen. The respective IDs of the heatmap proteins are presented in the appendix.

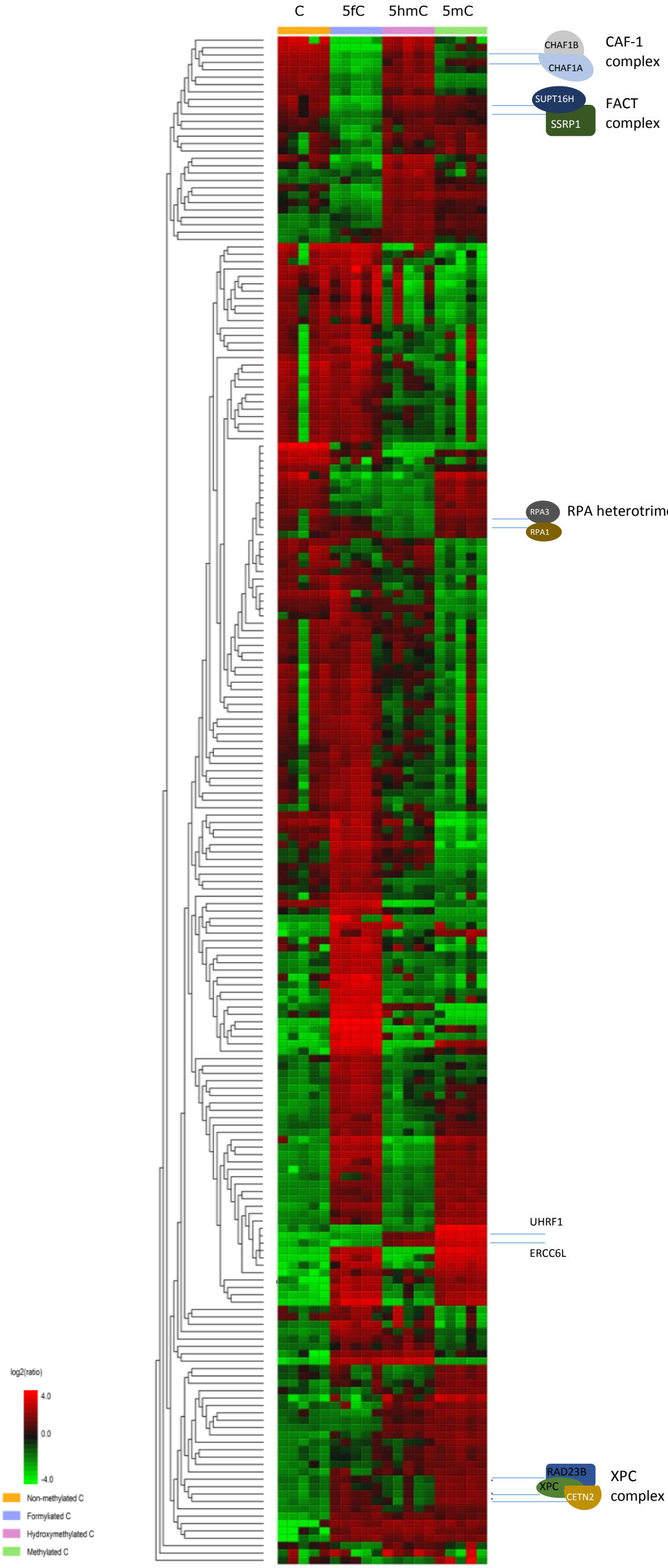



**Figure 5.3. Heatmap representing the protein enrichment for each cytosine modification on mNanog promoter.**

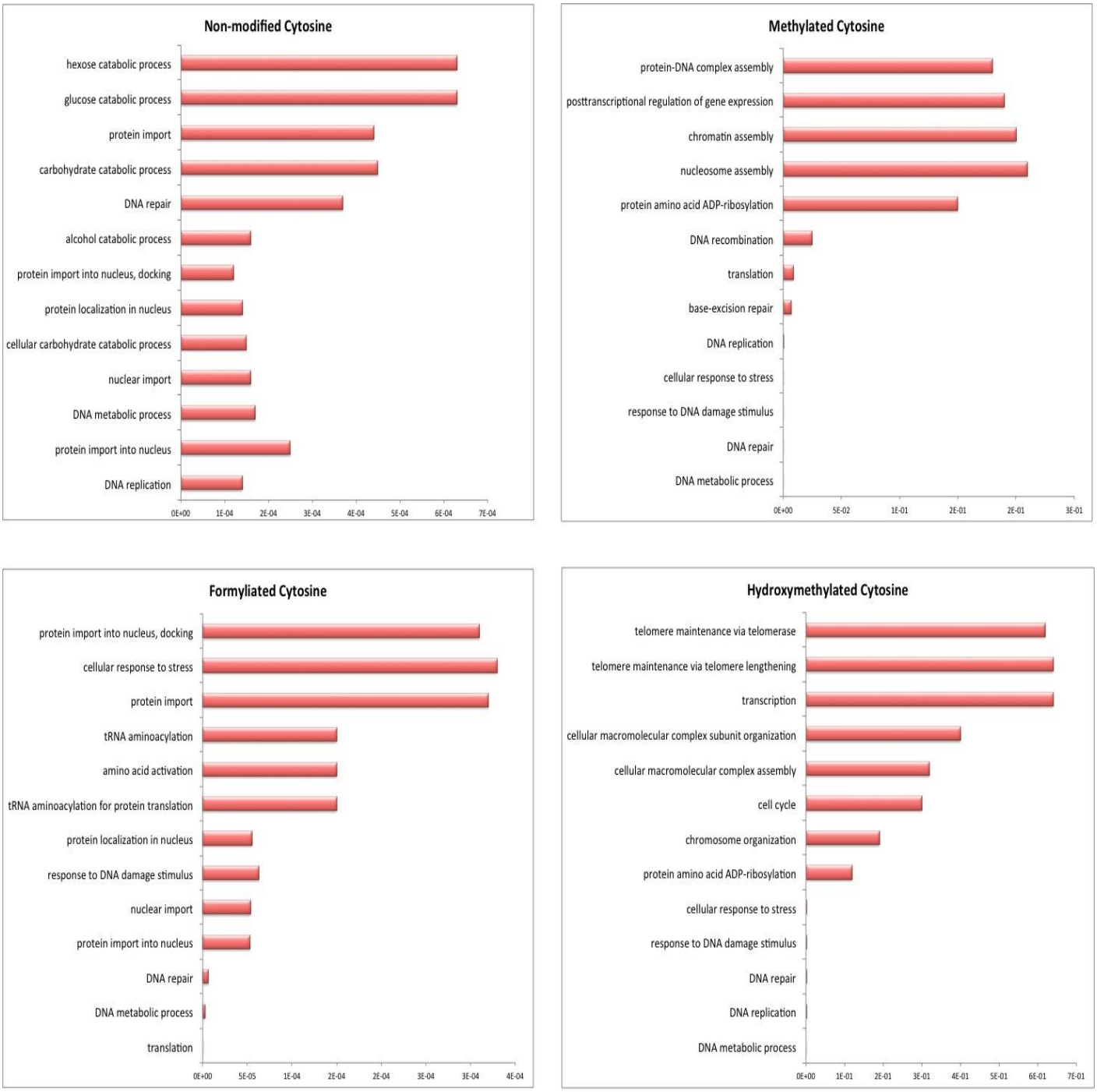



**Figure 5.4. Gene Ontology (GO) analysis of the binding factors showing relative enrichment for each cytosine modification (Benjamini corrected p-value).**

Gene Ontology analysis (GO DAVID) for the proteins enriched in each cytosine modification (Fig. 5.4), showed that the DNA replication pathway was significantly enriched in unmethylated C and 5hmC, while the DNA metabolic process pathway was highly enriched in all four cytosines. It should be noted that the GO category for DNA replication lists proteins such as the FACT complex, the CAF-1 complex subunits, RBBP4/7 and LIG1, which were identified in the screen, however are not strictly linked with DNA replication but operate in multiple cellular pathways as will be discussed below. The oocytes I used for the lysate are not capable of inducing DNA replication potentially because the ability of a female germ cell to induce DNA replication is linked with its maturity (Gurdon, 1968). The protein translation pathway was enriched in 5fC, indicative of high translational activity, potentially reflecting transcriptional activation under the 5fC modification. The genetic locus employed in the present research includes the TSS of the mNanog gene and a part of its first exon, supporting the idea that potential binders may include regulators of transcriptional activity. Additionally, the DNA repair pathway was significantly enriched in all three modified cytosine versions, indicating that 5mC, 5hmC and 5fC potentially trigger DNA repair pathways. The 5mC-mNanog promoter shows a significant enrichment for molecules associated with DNA damage response.

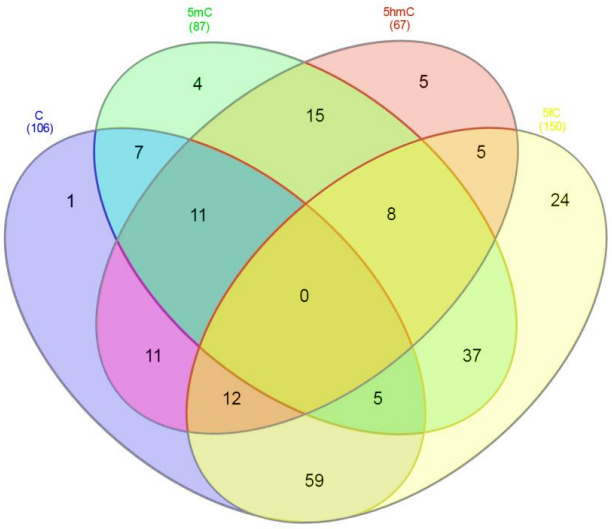



**Figure 5.5. Venn diagram comparing the overlapping and exclusive binders between the different cytosine modifications. The diagram was generated with the InteractiVenn tool.**

To obtain further insight into the data I looked into the overlapping and exclusive factors between each cytosine modification (Fig. 5.5). As discussed previously, the 5mC mark attracts methyl-binding proteins such as UHRF1, which was also identified in the present screen as a 5mC and 5hmC specific binder. Such binding preferences reflect potential biochemical relations between an epigenetic mark and a protein domain, and might also indicate the regulatory significance of each epigenetic mark. Additionally, common binders among different modifications might indicate biochemical or biological convergence between different modifications. Assessing modification specific and non-specific binders might hint potential demethylation pathways and also delineate how each cytosine is interpreted in the context of the oocyte.

From 87 5mC binding molecules, only 4 molecules TIA1, CKAP5, GM20390 and a histone of the H3 family, HIST3H2 were found to be 5mC exclusive binders in the screen. Only 23 were common between non-modified C and 5mC, with C-mNanog promoter binders amounting to a total of 106. The lower number of C:5mC common binders is a translation of the impact the methyl mark has on the surface of DNA. Any factors that would preferentially bind to non-modified cytosines in the context of a locus are repelled by the presence of 5mC. GO analysis of the 5mC:C common factors showed enrichment for proteins associated with DNA replication and DNA repair pathways (data not shown). The 5mC:5hmC common binders again showed high enrichment for DNA repair related factors, while UHRF1 was identified amongst the factors shared only between 5mC:5hmC.

Interestingly, a large number of binding molecules were shared between 5mC:5fC; a total of 50 common binders primarily linked with DNA repair and mRNA translation processes. Iurlaro et al (Iurlaro et al., 2013) reported potential association of the 5fC mark with transcriptional repression as the team observed transcriptional repression associated molecules to bind on 5fC much like 5mC in their screen. The high number of 5fC:5mC overlapping factors in the present screen is reminiscent of the observation of Iurlaro et al.

The entire XPC complex was found among the 5mC:5fC enriched molecules. The XPC complex comprises of XPC, CETN2 and Rad23b and has been associated with the NER repair pathway. Recently the XPC complex was characterized as a stem cell coactivator complex (SCC) because of its involvement with the transcriptional network of Nanog/Oct4/Sox2 (Cattoglio et al., 2015). Additionally, XPC was found to directly interact with TDG and stimulate its glycosylase activity, suggesting that XPC could potentially be involved in demethylation (Shimizu et al., 2010). Our screen did not identify TDG, but identified other glycosylases such as NTHL1 and MPG enriched in C:5mC:5fC and C:5fC respectively.

PARP1 was another molecule enriched in 5mC, 5hmC and 5fC that a recent report showed as important for the establishment of DNA methylation (Caiafa et al., 2009). More specifically PARP1 was found to inhibit DNMT1 dependent methylation when the latter physically interacted with the poly-ADP-ribose tail of PARP1 (Caiafa et al., 2009). An additional role of PARP1 in DNA demethylation was reported in PGCs, where the bulk of DNA demethylation occurs and PARP is highly expressed, while its inhibition impairs PGC demethylation (Ciccarone et al., 2012). More specific to the role PARP1 might be playing in the context of the oocyte it has been shown that paryliation is upregulated on paternal and maternal chromatin after fertilization, while the same was also observed in parthenogenetically activated oocyte (on the maternal chromatin), suggesting that maternal factors control the upregulation of paryliation (Imamura et al., 2004).

The number of 5fC:C common binders was 76. This group was enriched for protein localization, RNA binding and translation associated factors. The presence of proteins like PARS (Prolyl t-RNA synthetase) and SARS (Seryl-tRNA synthetase), two tRNA synthetases, RPs (Ribosomal Proteins) and Translation Elongation Factors such as EEF1B2 (Eukaryotic Translation Elongation Factor 1 beta 2), suggest that C and 5fC-mNanog promoter versions are transcriptionally active and produce mRNA.

While C, 5mC and 5hmC demonstrated a limited number of exclusive binders, 5fC showed a total number of 24 binders exclusive to 5fC in a total of 150 binders.

### 5.4.5 5mC binders: enriched for DNA repair factors, core histones and UHRF1

A high contribution to DNA repair pathways was observed in 5mC-mNanog promoter binders screen. More specifically, the Base Excision Repair (BER) pathway was enriched as seen on the GO analysis (Fig. 5.4), while proteins implicated in the Nucleotide Excision Repair (NER) pathway were also found among the 5mC binders.

NTHL1 (NTH like 1), a glycosylase catalyzing the first step of BER and PNKP (Polynucleotide Kinase Phosphatase 3'), which restores blocked 3' and 5' termini to 3'-OH and 5'-phosphate ends after AP cleavage, were found to bind 5mC-mNanog promoter. While the BER pathway has been proposed as a potential candidate pathway for active DNA demethylation NTHL1 and PNKP have not been linked to DNA demethylation yet. However, recently a DNA phosphatase 3' (ZDP) was reported to function in active DNA demethylation in the plant Arabidopsis (Martinez-Macias et al., 2012). DNA demethylation studies have reported 5mC DNA glycosylace activity in plants through the factor DEMETER, a similar active demethylation mechanism has not been reported in animals, interestingly, though ZDP and PNKP show sequence similarities (Martinez-Macias et al., 2012).

POLB (DNA Polymerase β/beta), responsible for gap filling in BER DNA repair and LIG1 (DNA ligase 1), which ligates remaining nicks at the final stage of BER were also identified amongst the 5mC binders. TDP1 (Tyrosyl-DNA phosphodiesterase 1), found to catalyse AP site cleavage and termini repair subsequently (Lebedeva et al., 2011), along with the helicases RECQL and RECQL5 participating in BER as secondary factors (Veith and Mangerich, 2015) were also identified in the 5mC screen. Additionally, the paryliation factors PARP1-3 were found enriched in the 5mC-mNanog promoter screen and

PARP1 has been reported to be closely linked to the BER pathway as its inhibition will compromise DNA repair by preventing the recruitment BER factors (Reynolds et al., 2015).

Although the BER pathway was fairly represented, factors belonging to the NER pathway were also identified. Characteristically, the entire XPC complex, composing of XPC, RAD23B and CETN2 was purified as a 5mC-mNanog promoter binder. The XPC complex, as discussed previously, is involved in lesion recognition during GG-NER. ERCC6L, a paralog of ERCC6 (CSB), was identified as a 5mC binder and could potentially be similarly involved in NER. RPA1 and RPA3, components of the RPA factor implicated in the extension and stabilisation of the nucleotide bubble formed during NER, were also identified as 5mC binding factors. It has to be noted here that DNA ligase LIG1 and POLB are common between BER and NER, and are therefore reported as contributors to both DNA repair pathways in the present thesis.

MSH3, implicated in DNA mismatch repair was also present in the screen.

The dimeric complex of FACT, SUPT16H and SSRP1, along with UHRF1 and three linker histones (H1E, H1B, H1C) were also identified in the 5mC screen. The connection between linker histones and methylated DNA has been demonstrated before and is potentially linked with chromatin compaction (Tazi and Bird, 1990, Ball et al., 1983).

SUB1/PC4 is a factor linked with transcriptional activation and repression as well as DNA repair. More specifically SUB1/PC4 was reported to assist in protein-DNA interactions of Pol II and Pol III transcription factor complexes, while it was also found to repress Pol II transcription (Conesa and Acker, 2010). SUB1/PC4 knock down HeLa cells resulted in opened chromatin and elevated gene transcription, an additional indication of its repressive activity (Das et al., 2006). The same factor was also reported to accumulate at DNA lesions, linking it with DNA repair, NER specifically judged by its interaction with TFIIH (Conesa and Acker, 2010).

SETX (Senataxin) has been reported to play a role in genetic integrity maintenance specifically in meiotic cells (Becherel et al., 2013). YBX2, on the other hand, appears to be germ cell specific and studies have linked it with

fertility. It is primarily associated with RNA binding, however it could potentially be acting as a transcription factor as well (Snyder et al., 2015).

HNRNPK is an RNA binding ribonucleoprotein associated with pre-mRNA. HNRNPK knockdown enhanced the efficiency of iPS reprogramming, while its overexpression had the opposite effect (Bao et al., 2015). PURA is a DNA and RNA binding factor and potential transcription factor participating in viral but cellular regulation as well. Recent findings suggest its implication in DNA repair (White et al., 2009).

EVC2, an additional 5mC binder identified in the screen, has been found to positively modulate the hedgehog pathway, important during embryo development (Blair et al., 2011).

CHAF1A, a subunit of the CAF-1 complex which has been linked with heterochromatin organization and epigenetic silencing (Houlard et al., 2006) was found amongst the 5mC binders. Recent findings have reported CHAF1A and CHAF1B repression to enhance iPS reprogramming and promote the activation of pluripotency associated genes such as Oct4 and Nanog (Cheloufi et al., 2015).

It has to be noted that additional factors such as histones and ribosomal proteins were also identified in the 5mC screen. Among the identified histones was Oocyte type linker histone B4, also known as H1oo.

The identified proteins formed a well-connected network as shown after STRING analysis (Fig. 5.6).

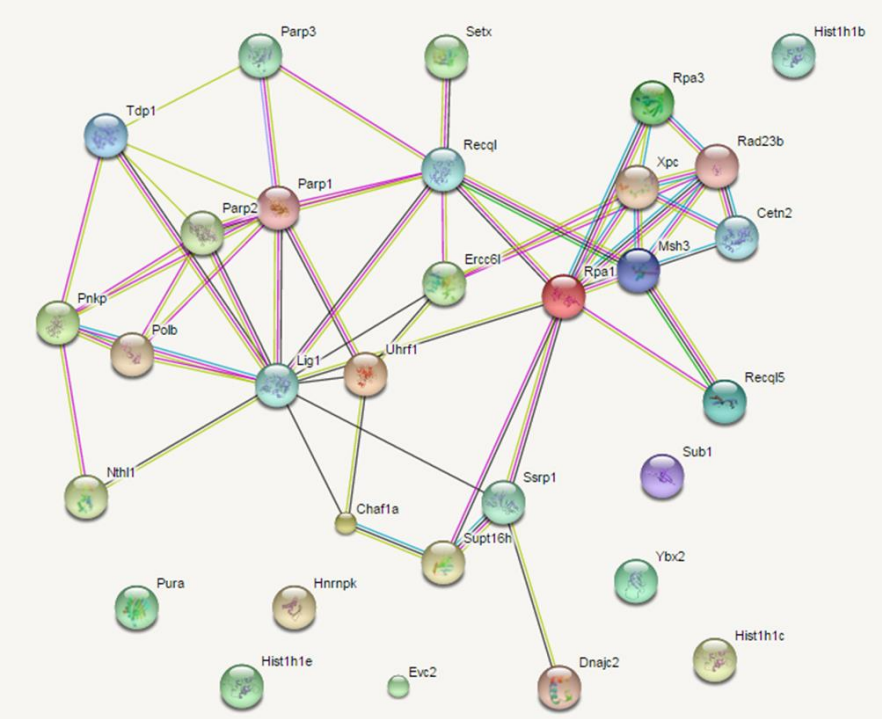



**Figure 5.6. Representative protein network of 5mC-mNanog promoter binders as identified in the screen.**

### 5.4.6 5hmC binders: chromatin remodelers and DNA repair enrichment

The FACT complex, SUPT16H and SSRP1, was enriched in the 5hmC-mNanog binders' screen as well. FACT, facilitates chromatin transcription, has been reported to promote transcriptional elongation by removing H2A-H2B dimers from chromatin hence positively impacting the progression of RNA pol II (Belotserkovskaya et al., 2003). More recently the FACT complex was found to specifically interact with H2A.X and facilitate its placement and displacement on chromatin (Heo et al., 2008). The team also showed that PARP1 paryliates SUPT16H, hence repressing H2AX exchange (Heo et al., 2008). Two subunits of the CAF-1 (Chromatin Assembly Factor 1) were also identified: CHAF1A and CHAF1B. CAF-1, which is comprised of three subunits in total CHAF1A, CHAF1B and p50, has been associated with heterochromatin maintenance, centromere organization and chromatin restoration after DNA repair as

reviewed in (Yu et al., 2015). CAF-1 has also been shown to play a role in the activation of Notch target genes by potentially associating with Suppressor of Hairless (SuH), which regulates the deposition of activating epigenetic marks in Drosophila (Yu et al., 2013). RBBP4 and RBBP7 were pulled down in the 5hmC screen. Both proteins are subunits of the Nucleosome Remodelling Deacetylase (NuRD) complex, which is one of the cellular co-repressor complexes and has been reported to recruit Histone Deacetylase enzymes to chromatin (Millard et al., 2016). Moreover, RBBP4/7 has been found to associate and contribute to additional complexes such as CAF-1 (Parthun et al., 1996, Verreault et al., 1996) and HAT1 , a Histone Acetyltransferase complex (Verreault et al., 1998). RBBP4/7 also directly interacts with H3 and H4 histones as part of their histone chaperone role and might potentially act as histone recruiters (Nowak et al., 2011). YBX1 and YBX2, also identified in the screen are Y-Box binding proteins. The first is associated with multiple cellular roles such as transcriptional regulation, DNA repair and mRNA regulation (splicing, packaging, translation) (Lyabin et al., 2014), while YBX2 is a potential transcription factor highly expressed in germ cells as discussed previously. PURA, a probable physical interactor of YBX1 as shown by protein network analysis, was also found in the 5hmC-mNanog promoter screen.

The DNA helicases RECQL and RECQL5 were also identified as 5hmC binders. Cathepsin J (Ctsj), a factor not well characterized but potentially associated with protease activity was also identified in the screen. Interestingly, the expression profile of Ctsj seems to be restricted to placental tissue and could potentially be associated with embryonic development (Tisljar et al., 1999). Karyopherin a7 (KPNA7), found to be important for porcine embryo development (Wang et al., 2012), was found among the 5hmC-mNanog promoter binders. NAP1L4, a member of the nucleosome assembly family and a histone chaperone was found in the screen.

Two clamp proteins were also identified in the 5hmC screen. RFC3 and PCNA physically interact and RFC3 potentially recruits PCNA to chromatin. The clamps associate with DNA replication, DNA repair and also serve as

interactors for different factors. UBTF, Upstream Binding Transcription Factor, associated with RNA pol I and rRNA regulation, as well as coordination of polymerase II gene transcription (Sanij et al., 2015), was also identified as an 5hmC binder. Additionally, UHRF1, also known to bind 5hmC, and ERCC6L were also found among the 5hmC binders, however both ERCC6L and UHRF1 bound 5mC more specifically.

Additionally, SETX, TDP1, PNKP and EVC2 were identified as 5hmC binding factors, much like in the case of the 5mC screen.

### 5.4.7 5fC binders: 5mC resemblance and the methionine cycle

The 5fC-mNanog promoter screen generated the highest number of binding factors compared to 5mC, 5hmC and non-modified cytosine mNanog promoter. As shown in the Venn diagram (Fig. 5.5) there was a significant overlap between 5fC and 5mC. Much like in the case of 5mC, the BER pathway is very well represented, recapitulating all the steps from damage recognition to DNA polymerase repair and nick ligation. The BER pathway factors identified in the screen were: NTHL1, MPG, APEX1, PNKP, PCNA, POLB, LIG1, RECQL, RECQL5, FEN1 and the paryliation protein PARP1. From the above BER associated factors MPG, APEX1 and FEN1 were not identified in the 5mC screen. MPG (N-Methylprunine DNA Glycosylase) is a glycosylase capable of initiating the BER pathway, while APEX1 (Apurinic/Apyrimidinic Endodeoxyribonuclease 1) is a protein that recognizes AP sites and by using them as substrate cleaves its 5' end. FEN1 (Flap structure specific Endonuclease 1) removes the flap DNA strand created during LP-BER.

Additionally and in resemblance to the 5mC binders, NER pathway associated molecules are present in the 5fC screen. The factors XPC, RAD23B, CETN2, RPA1, RPA3, RFC3 and LIG1 are the NER participants identified in the 5fC-mNanog promoter screen. RFC3 (Replication Factor C subunit 3) is a subunit of a DNA clamp complex participating in DNA repair and replication. Additional

factors shared among 5mC and 5fC screen were the core histones H1E, H1B, H1C, DNAJC2, EVC2 and SUB1 and the oocyte specific linker histone B4.

Interestingly, three of the identified binders of 5fC-mNanog promoter belong to the control of methylation cycle in the cell: AHCY, SHMT1 AND BHMT. AHCY (Adenosylhomocysteinase) is involved in the conversion of SAH (S-Adenosyl-Homocysteine) to HCY (Homocysteine) and adenosine by hydrolysis. SAH is a competitor of methyl-transferase reactions depending upon SAM (S-Adenosyl-Methionine), which is a major substrate for methylation by donating a methyl group. AHCY might therefore be involved in the regulation of methylation levels by regulating the concentration of SAH. Notably, depletion of AHCY causes lethality in mice (Miller et al., 1994), while its stable overexpression causes cell apoptosis in human cells (Hermes et al., 2008). AHCY deficiency in yeast will inhibit histone methylation through SAH (Tehlivets et al., 2013). SHMT1 (Serine Hydroxymethyltransferase 1) converts serine and tetrahydrofolate into 5,10-tetrahydrofolate in the folate cycle. This reaction provides one-carbon units for the synthesis of methionine, thymidylate and purines. BHMT (Betaine Homocysteine Methyltransferase) on the other hand, carries out the methylation of homocysteine by transferring a methyl group from betaine. The participation of each enzyme in the folate and methionine cycles and their relation to DNA methylation are depicted in Fig. 5.7.

Calcium Response Factor (CARF), identified in both 5mC and 5fC is a DNA binding factor that has been found to bind loosened chromatin on the promoter of the BDNF (Brain Derived Neurotrophic Factor) gene and positively regulate its transcription (Tao et al., 2002). NAP1L4 (Nucleosome Assembly Protein 1 like 4), also identified in 5hmC screen and SMC4 are involved in chromosome maintenance, with the latter belonging to the condensin complex which plays a role in mitosis, chromosome condensation and re-annealing of single stranded DNA (Sutani and Yanagida, 1997).

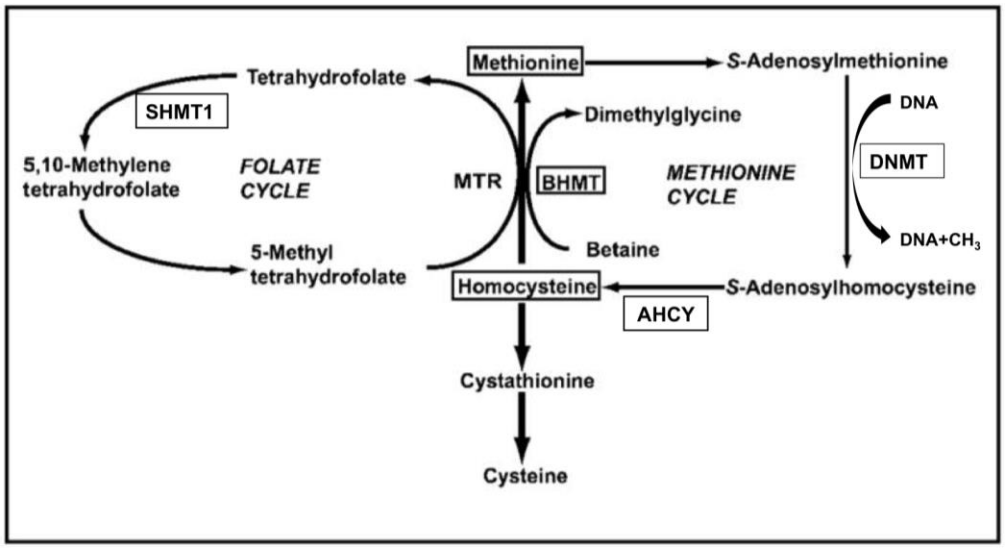



**Figure 5.7. Folate and methionine cycles and the participation of the identified SHMT1, BHMT and AHCY proteins. SHMT1 is catalyses the formation of 5,10-Methylene tetrahydrofolate from tetrahydroformate (THF), which is then converted to 5-Methyl tetrahydrofolate that can subsequently be used as a methyl donor for the methylation of homocysteine. BHMT catalyses the methylation of homocycteine by using betaine as a methyl donor. The universal methyl donor SAM (S-adenosylmethionine) is formed from methionine and is used from methyltransferases such as DNMTs to methylate DNA or other substrates. AHCY, lastly, is responsible for the conversion of SAH to homocysteine.**

Interestingly, ATIC (5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase) which is a bifunctional protein implicated in purine biosynthesis was identified in the 5fC screen. ATIC has formyltransferase activity and while its presence in the screen might also be related to tRNA synthetases or ribosomal proteins found as 5fC interactors, its association to the formyl group is important as it could also be potentially hinting alternative 5fC removal pathways prior to, or after, the excision of the 5fC from DNA.

### 5.4.8 Non-modified cytosine binders: low number of binders and absence of PARP

The non-modified cytosine mNanog promoter (C-mNanog promoter) shares a number of binders with the 5mC, 5hmC and 5fC screens. DNA repair associated factors belonging to the BER and NER pathways are present in the C-mNanog promoter screen, however enzymes that regulate the DNA repair pathway are not recapitulated in their entirety as is the cases for 5mC and 5fC. MPG, NTHL1 and APEX1 were the identified BER factors, RPA1 and RPA3 were the NER related factors while LIG1, also identified in the screen, is related to both NER and BER. The identification of a low number of DNA repair associated factors as compared to 5mC and 5fC likely reflects that extensive DNA repair does not take place on the non-modified version of the mNanog promoter. Additionally the high number of DNA repair associated binders in the cases of 5mC and 5fC demonstrates higher DNA specificity in regards to their binding. The DNA repair factors are not sequence or base specific, therefore potentially they might be involved in the screening of the DNA fragment in search for DNA lesions, hence explaining their presence in the C-mNanog promoter screen. Moreover, the DNA repair associated binders found enriched in the C-mNanog promoter screen represent only the initial steps of base and AP site cleavage for the case of BER, and the stage after the unwinding of the damaged DNA for the NER case. In contrast with 5mC and 5fC, the C-mNanog promoter screen demonstrates only partial reflection of the DNA repair mechanisms of BER and NER.

Notably, PARP proteins were not enriched in the C-mNanog promoter screen, indicating that potentially PARPs are associated with the recognition of the modified cytosines and potentially initiate the DNA damage response. As discussed previously, significant enrichment of paryliation has been reported on the surface of murine chromatin after fertilization as a result of oocyte activity (Imamura et al., 2004). The enrichment of PARP1 in 5mC, 5fC and 5hmc over C, indicates a specific interaction between the modified DNA and

PARP1 either through the recognition of the modification or the recruitment of PARP1. The CHAF1A and CHAF1B subunits of the CAF-1 (Chromatin Assembly Factor 1) complex, along with the core histones H2AO, H2BO and H2BB that were identified as 5hmC binders, were also enriched in the C-mNanog promoter screen indicating that the CAF-1 complex might be responsible for depositing the latter histones onto chromatin in oocytes. The NuRD complex subunits RBBP4 and RBBP7, also falling into the category of factors identified both in the 5hmC and non-modified C screens, might potentially be associated with histone deposition rather than transcriptional repression, in the present context.

Interestingly, the proteins involved in the folate and methionine cycles SHMT1, BHMT and AHCY were also identified as C-mNanog promoter binders.

| | C | 5mC | 5hmC | 5fC |
|---|---|---|---|---|
| H1E, H1B, H1C, B4 | | 🔴 | | 🔴 |
| FACT complex | 🟢 | 🟢 | 🟢 | |
| SUB1/PC4 | | 🔴 | | 🔴 |
| CARF | | 🟢 | | 🟢 |
| RBBP4/7 (NuRD) | 🔴 | | 🔴 | |
| CAF-1 | 🔴 | | 🔴 | |

**Table 5.2. The predicted transcriptional impact of representative binders identified in the screen, as predicted based on existing publications. The circles denote the presence of the respective factors in among the identified binders of each cytosine modification. A red circle denotes negative (repressive) influence on transcription, and the green circle denotes potential transcriptional activation**

### 5.4.9 BER or NER: Fold enrichment of NER and BER associated factors in C/5mC/5hmC/5fC

Since some DNA repair associated factors were also found in the non-modified cytosine, I assessed the relative fold enrichment (log2 ratio of abundance) of the BER and NER DNA repair molecules identified in the screen in all four cytosine counterparts (Table 5.3). NTHL1, MPG and APEX1 appear more abundant in the case of C and judging by their relative abundance in the 5mC-mNanog promoter, they seem to be repelled by the 5mC mark (Table 5.3). Interestingly, though, APEX1 is more than two-fold enriched in the case of 5fC. Alternatively, the rest of the BER associated factors are more abundant in the modified cytosine screens. For the NER pathway, the XPC complex responsible for the recognition of the lesion reaches almost 3-fold abundance in 5mC, while also being enriched in 5hmC and 5fC compared to C, although at lower levels. The same applies to POLB, one of the DNA polymerases that carries out gap filling in DNA lesion sites for both NER and BER. RPA1, RPA3 and RFC3 also appear to be more abundant in 5mC compared to C, 5hmC and 5fC. Strikingly, ERCC6L demonstrates ultra-specific binding for 5mC at an eleven-fold enrichment compared to C. ERCC6L is also clustered together with UHRF1, another highly specific binder of 5mC, in the heatmap of Fig. 5.3 because of their similar expression pattern. Judging by the difference in the fold-enrichment of the binders, the NER related factors, and specifically the XPC-RAD23B-CETN2 DNA damage recognition complex, appears more strongly related to 5mC. The binding of BER related molecules (except for NTHL1 and MPG) appears to be more targeted to either 5hmC or 5fC, with the exceptions of TDP1, which appears 5mC specific and PARP1, which is equally abundant in 5mC and 5fC.

Interestingly, APEX1 and FEN1 show a high enrichment for 5fC, while RECQL5 shows a five-fold enrichment for the modified cytosines in comparison to C, thus being interesting candidates for the study of their specific relation to each modification.

| | | C | 5mC | 5hmC | 5fC |
|---|---|---|---|---|---|
| BER | NTHL1 | 1 | 0.33 | 0.05 | 0.36 |
| | MPG | 1 | 0.03 | 0.02 | 0.09 |
| | APEX1 | 1 | 0.35 | 0.09 | 2.35 |
| | PNKP | 1 | 3.03 | 3.24 | 3.7 |
| | PCNA | 1 | 1.25 | 2.18 | 1.67 |
| | TDP1 | 1 | 2.24 | 1.92 | 1.35 |
| | RECQL | 1 | 1.94 | 2.68 | 1.82 |
| | RECQL5 | 1 | 5.18 | 4.59 | 5.51 |
| | FEN1 | 1 | 1.56 | 2.18 | 3.97 |
| | PARP1 | 1 | 2.3 | 1.74 | 2.33 |
| NER | XPC | 1 | 2.86 | 1.41 | 2.09 |
| | RAD23B | 1 | 2.44 | 1.23 | 1.66 |
| | CETN2 | 1 | 3.12 | 1.48 | 2.15 |
| | RPA1 | 1 | 1.17 | 0.58 | 0.93 |
| | RPA3 | 1 | 1.22 | 0.56 | 0.97 |
| | RFC3 | 1 | 1.11 | 1.87 | 2.21 |
| | ERCC6L | 1 | 11.26 | 4.2 | 0.81 |
| BER/NER | LIG1 | 1 | 1.75 | 0.86 | 1.7 |
| | POLB | 1 | 3.76 | 0.89 | 2.37 |

**Table 5.3. Relative fold enrichment of BER and NER contributing factors identified in the screen, based on the log2 ration of abundance of each protein in the respective average abundance between the replicates of each sample.**

## 5.5 Conclusion

The screen for different cytosine modification binders identified a multitude of binding factors. Among the different cytosine modifications assayed, 5fC and 5hmC demonstrated the highest and lowest number of binders, respectively. The latter was also observed by Spuijt et al (Spruijt et al., 2013) in the context of mESC indicating that 5fC potentially is an epigenetic mark attracting multiple factors and impacting the respective biological pathways. Additionally, the screen revealed that many more proteins bound exclusively to 5fC than any other cytosine counterpart. The same observation has also been reported previously in mESC (Iurlaro et al., 2013) highlighting the epigenetic significance of 5fC in a pluripotent cell context.

Interestingly, the screen revealed only a small number of potential transcription factors, contrary to expectations. This could potentially be attributed to the immature stage of the oocytes used in the experiment. The

presence of transcription factors at such an early oocyte stage might be in limited amounts as previously reported (Wang et al., 2010), and experimental handling, which inevitably causes some protein loss, might be further hindering their detection. It would be interesting to attempt the same experiment with MII (Metaphase II) oocytes to see whether transcription factor binding levels improve.

While other studies carrying out DNA pull down experiments with the different cytosine modification counterparts have reported TET protein binding to 5mC and 5hmC, no TET binders were identified in the present study. Active demethylation through TET enzymatic activity has been the most prominent concept in the field of methylation reprogramming, while DNA repair pathways have been proposed as a secondary possibility. Interestingly, the screen presented in this chapter identified multiple DNA repair associated factors enriched in all the modified cytosine counterparts of the screen. BER repair contributors are enriched in 5mC, 5hmC and 5fC, with FEN1, PCNA and TDP1 being more abundant in 5fC, 5hmC and 5mC respectively. Whether the binding of the latter three molecules has a specific biological impact needs to be researched. It would be interesting to assess whether their disruption would impact the levels of the respective cytosine modifications in the oocyte. NER associated factors, on the other hand, although appearing enriched in all modified cytosines compared to C, demonstrated a preference for 5mC (Table 5.3). The XPC complex, which is primarily involved in the recognition of DNA lesions during GG-NER, was identified in the screen and all its subunits (XPC, CETN2, RAD23B) were found almost three-fold enriched in 5mC compared to C-mNanog promoter. 5fC demonstrated a two-fold enrichment of the XPC complex factors over non-modified cytosine, while 5hmC also showed XPC enrichment although not as high as 5mC or 5fC.

The function of the NER repair pathway in active DNA demethylation has been proposed when NER related factors were found directly implicated in the demethylation of the Oct4 promoter in Xenopus oocytes (Barreto et al., 2007). The emergence of TETs and the respective active demethylation pathway challenged such findings. While drawing sound conclusions requires

additional experimentation, the findings in the present chapter highlight the association of NER and the XPC complexes specifically with the 5mC mark. Findings have suggested that NER functions mainly through recognizing chromatin structure. Bearing that in mind, it is difficult to discern at this stage whether the XPC complex recognizes the 5mC mark itself or its impact on chromatin conformation. 5mC has been recently reported to reduce DNA flexibility (Ngo et al., 2016), while linker histones that were found to be enriched 5mC binders in the present screen, also have an impact on chromatin conformation (Tazi and Bird, 1990, Ball et al., 1983). Judging by the enrichment of BER DNA repair contributors in the case of 5mC/5hmC/5fC, one cannot exclude the potential role they might be playing in the repair of the respective modification. What is interesting, though, is that the oocyte at this stage recognises all DNA modifications as potential lesions and DNA repair factors are recruited. It is therefore likely that the demethylation wave observed after fertilization might be triggered not by the epigenetic mark as such, but by its impact on DNA and chromatin.

While the screen was successful in identifying binding factors one can only hypothesize on the impact they had on the potential removal of the cytosine mark. Whether all modified cytosines are removed and replaced by a non-modified cytosine, or additional processing occurs prior to complete demethylation during the incubation of mNanog promoter in the oocyte lysate needs to be further defined. It would be very interesting to see how each cytosine modification is altered (or not) by carrying out the same experiment of DNA pull down and obtaining the mNanog promoter for nucleotide analysis.

It also has to be noted that the modified cytosines were uniformly incorporated in the mNanog promoter during End-Point PCR. Therefore all the cytosines (except for the primer) carried the mark and there was no CpG specific modification. Therefore that should also have its consequences on both DNA conformation and binding factors. Again, it would be of high interest to conduct the same experiment, however this time introducing each modification only in the CpG context. Interestingly, UHRF1 which is a

methylated (and hydroxymethylated) CpG binder reported to preferentially bind hemimethylated DNA and be repelled by full methylation (Bianchi and Zangi, 2013), was found highly enriched in the 5mC screen. Recent findings have associated UHRF1 with DNA repair as a DNA damage recognition factor (Tian et al., 2015). UHRF1 is unlikely to be recruited for DNA methylation maintenance at the prophase oocyte stage; therefore its presence in the screen indicates participation in alternative pathway and DNA repair is a likelihood. Of note should also be that the hyper-methylated status of the DNA mimics that of sperm. Therefore the 5mC screen ideally recapitulates what commences during sperm DNA remodelling.

Interestingly the XPC complex was enriched in both 5mC and 5fC, with 5fC enrichment being lower that 5mC. The XPC complex comprises of XPC, RAD23B and CETN2 and all three are key components of NER. The XPC subunit recognises DNA damage and initiates DNA repair while RAD23B and CETN2 mostly operate to stabilise XPC binding to DNA. Possible implications of XPC in demethylation were reported recently when XPC knockdown lead to reduced promoter activation and loss of DNA demethylation (Le May et al., 2010). In the same study XPC was also linked with promoter activation as opposed to generic DNA demethylation. Additionally, the XPC complex has been characterised as having stem cell coactivator qualities (SCC) by participating in the activation of pluripotency associated genes (Cattoglio et al., 2015). Identifying the XPC complex as a 5mC:5fC binder potentially indicates that 5mC and 5fC might be removed from genetic sequences by the same molecular mechanism in the context of the oocyte.

ERCC6L, a paralog of ERCC6 was also highly enriched in the 5mC-mNanog promoter screen. Protein network analysis showed potential associations between UHRF1 and ERCC6L, while UHRF1 and ERCC6L showed similar expression profiles and were clustered together in the heatmap (Fig. 5.3). ERCC6, also known as CSB, is a factor that contributes to the NER pathway. Notably, ERCC6 has been proposed to participate in DNA repair, where damage prevents RNA polymerase from elongating the transcription of a gene (Selby and Sancar, 2003). However, it is not known whether ERCC6L functions

in the same manner as ERCC6. According to Uniprot, ERCC6L can act as a chromatin tension censor and whether this means that ERCC6L is mostly related to 5mC induced conformational changes needs to be delineated. However, its eleven-fold enrichment in the 5mC screen as compared to C renders the factor an interesting candidate for study in relation to the methyl mark.

The oxidative demethylation pathway (TET dependent) is widely accepted as the prevailing mode of demethylation. There has been no proof in any cellular system that 5hmC, 5fC and 5caC could be deposited on cytosine as individual chemical groups, much like 5mC is by methyltransferases. However, 5fC and 5caC can both be removed by TDG and their gaps subsequently filled by C. The relative enrichment of DNA repair factors across all modified cytosines could serve as an indication that potentially the oocyte is capable of removing all modified cytosine by DNA repair. In the same manner that TDG has been reported to recognise and remove 5fC and 5caC, 5mC and 5hmC might also be recognised by additional glycosylases or base/nucleotide excision factors. Interestingly, however the glycosylases identified in the screen, NTHL1 and MPG showed enrichment for the non-modified cytosine counterpart and were repelled by modifications. At the same time APEX1, an AP lyase of the BER pathway was found to be relatively enriched in C and in over two-fold enrichment in 5fC binding. The presence of APEX1 on non-modified cytosine is not well justified. Maybe APEX1 participates in alternative pathways independent of cytosine modifications. Previously, however, APEX1 inhibition has been reported to lead to diminished demethylation in murine PGCs (Hajkova et al., 2010). Additionally, RECQL5, a DNA helicase showed high specificity for modified cytosines, potentially reflecting a more universal treatment of DNA modification during fertilization.

An additional factor identified in the screen as a 5mC:5fC binder is PARP1, an enzyme catalysing paryliation, a post-translational modification attaching a poly-ADP tail to amino acids. PARP1 has been suggested to play the role of the XRCC1 recruiter on DNA lesions and also participate in PGC demethylation (Ciccarone et al., 2012, Gavin et al., 2013). Additionally, PARP1 was found to

be recruited in the paternal pronucleus, and when inhibited demethylation of the paternal DNA was reduced (Hajkova et al., 2010).

The modified cytosine binding screen did not identify any of the usual suspects of demethylation: the TETs. However, recent findings have suggested that oocyte demethylation is potentially carried out through TET independent pathways. Based on my findings in the present chapter, I propose that the NER DNA repair pathway factors appear to most prominently bind the methylated DNA compared to the BER pathway that does not show 5mC specific enrichment. However, the enrichment of BER factors also hints potential cooperation of the NER and BER pathway in the context of oocyte demethylation. From the NER factors, it is the XPC complex that recognises the 5mC mark as shown by its significant 5mC enrichment and by the known function of the XPC as a lesion recognising factor. However, it is likely that the XPC complex does not recognise and bind the 5mC mark per se, but the impact the 5mC modification has on DNA conformation. I also propose that in the oocyte context all cytosine modifications trigger the DNA repair response and might potentially be removed via the NER or BER repair pathway. Additionally, judging by the relative enrichment of the BER associated factors in the 5mC screen, the two pathways might be cooperating in the context of oocyte methylation recognition and eventual demethylation. The participation of each identified factor in the BER and NER repair pathways is depicted in Fig. 5.8.

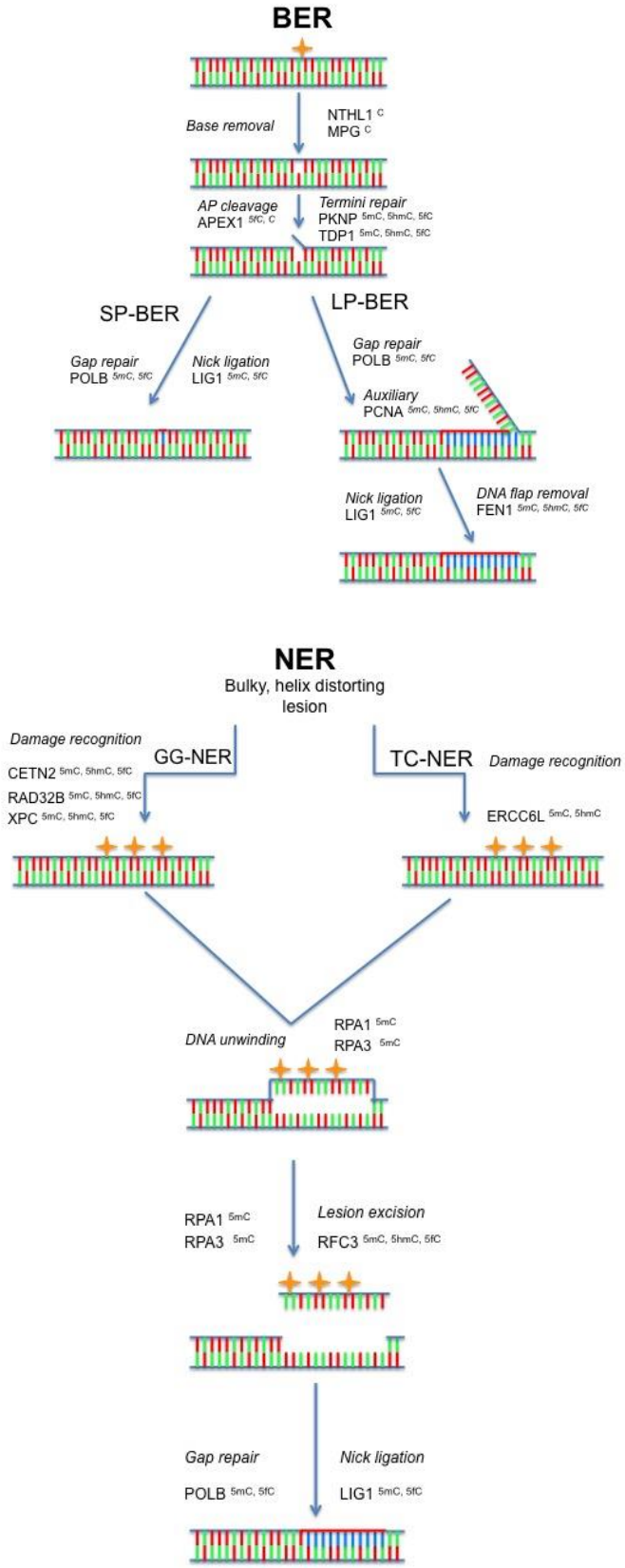



**Figure 5.8. The BER and NER pathways and which step the DNA repair factors identified in the screen contribute to. Note that contribution of additional factors is needed for DNA repair completion, specifically in the case of NER. However, in this schematic only the participation of identified factors is highlighted.**

# Chapter 6: Discussion

## *6.1 Protein immunoprecipitation and antibody optimization*

Protein immunoprecipitation is a standard tool for the identification of physical interactors of a protein of interest. However, when employed to delineate AxNanog co-factors in *axolotl* oocytes, it indicated and led to the discovery of an antibody malfunction. The AxNanog antibody that was generated by previous lab members, while being capable of recognizing in vitro translated and recombinant Nanog failed to identify native AxNanog. The antibody was primarily recognizing the GroEL equivalent in the oocyte, resulting in a prominent that was mistakenly associated with AxNanog. In my attempt to solve the problem by purifying the antibody from the GroEL antibody contaminant, I also discovered that a significant percentage of recombinant AxNanog was trapped in inclusion bodies over the course of its overexpression in *E. Coli*. Eventually, through various applications, I was able to purify an improved version of the AxNanog antibody from immunized rabbit serum. However, the new batch of AxNanog antibody was not able to recognize AxNanog natively, while being able to detect overexpressed AxNanog in pre-iPS cell via immunocytochemistry. As a result there was no immunoprecipitation presented in Chapter 3. However, while the work carried out was not fruitful in terms of protein network identification, it was particularly useful for the development of troubleshooting and result interpretation skills.

## 6.2 Optimization and application of DNA pull down in axolotl oocyte lysates

In the context of the present thesis, the work presented in Chapter 4 represents the preparatory steps that eventually led to the differential cytosine modification binders screen of Chapter 5. The initial design of the project aimed at identifying mNanog promoter binding transcription factors in the context of the *axolotl* oocyte, and also comparing binding factors between embryo and oocyte. The optimization experiments as well as mass spectrometry protein identification trials were carried out; however the final large scale experiment was not carried out. The halt was not a result of failure, but of plan alteration based on publications (Spruijt et al., 2013, Iurlaro et al., 2013) that added a new insight to the initial project. Nonetheless, with the embryo and oocyte comparison I was still able to identify factors binding to the mNanog promoter, while some of them were identified in the C/5mC/5hmC/5fC-mNanog promoter screen presented in Chapter 5. More specifically, SSRP1 the subunit of the FACT complex, and UHRF1, that were identified as embryonic binders of the mNanog promoter. Additionally, PARP1 and the subunits of the nucleosome assembly complex CAF1, CHAF1A and CHAF1B that appeared to be split between embryo and oocyte, are both present in the oocyte, judging by the data presented in Chapter 5.

## 6.3 DNA repair and oocyte demethylation

The oocyte is capable of reprogramming chromatin through loosening up chromatin with key reprogramming factors as previously demonstrated by J. Gurdon and I. Wilmut (Gurdon and Wilmut, 2011). Notably, extensive demethylation takes place in the oocyte at the onset of fertilization to prepare the highly methylated paternal chromatin for zygote formation. Paternal DNA demethylation occurs actively, through a replication independent mechanism (Hajkova et al., 2010, Oswald et al., 2000, Wu and Zhang, 2010). The demethylating capacity of the oocyte has been linked with its elevated efficiency in cell fate reprogramming as compared to iPS (Ma et al, 2014).

Extensive research has been carried out to define the demethylation pathway and the TET protein family has been reported as primarily responsible for DNA demethylation through the successive oxidation of 5mC. TET3 is highly expressed in mature oocytes, while 5hmC has also been detected on paternal DNA following DNA demethylation (Wossidlo et al., 2011, Gu et al., 2011). Additionally, the presence of 5fC and 5caC was also reported in zygotes (Inoue et al., 2011b). Such findings led to the suggestion that demethylation in oocytes occurs through the successive oxidation of 5mC by the TET pathway. However, recent findings have added a new perspective to this view. Notably, the publications of Amouroux et al (Amouroux et al., 2016) and Santos et al (Santos et al., 2013) reported that zygote DNA demethylation occurs in two distinct phases. Phase I demethylation commences during the zygotic stages PN0-PN2, while Phase II demethylation takes place during PN3-PN5 (Santos et al., 2013). Both teams report TET3 dependent 5hmC accumulation during Phase II (PN3-PN4 onwards), however the loss of demethylation in Phase I is not linked with TET activity or 5hmC increase. Additionally, Amouroux et al (Amouroux et al., 2016) reported that 5hmC formation during Phase II is dependent not only on TET3 activity, but on de novo DNA methyltransferase activity as well. The findings of Amouroux et al (Amouroux et al., 2016), introduce a new perspective into the demethylating capacity of the oocyte by

suggesting that paternal chromatin loses the 5mC mark during the early zygote stages to potentially be remethylated and successively oxidized by the oocyte in the stages of PN3-PN4 and onwards. The independence of Phase I oocyte demethylation from TET activity suggests that alternative demethylation pathways operate, especially during early zygote stages where paternal sperm DNA is subjected to demethylation. The NER and BER DNA repair pathways have been proposed as alternative demethylation pathways (Franchini et al., 2012, Gehring et al., 2009). Recently, both NER and BER associated factors have been linked with oocyte demethylation. More specifically, Long Patch-BER (LP-BER) was reported to function through AID deamination concomitantly with hydroxymethylation in later zygote demethylation (Phase II) (Santos et al., 2013). Hajkova et al (Hajkova et al., 2010) has also reported BER related activity in murine zygotes as core contributors of the BER pathway, APEX1 and PARP1 were found to play a significant role in paternal DNA demethylation. The NER DNA repair pathway has been associated with active DNA demethylation and silenced promoter activation in Xenopus oocytes with the mediation of Gadd45a (Schafer et al., 2010, Barreto et al., 2007).

In Chapter 5 I focused on identifying cytosine modification specific binders on the genetic sequence of the mNanog promoter and in the context of oocyte. More specifically, mNanog promoter that was amplified with the incorporation of either non modified cytosine (C) or one of the 5mC/5hmC/5fC was incubated in *axolotl* oocyte lysate to recapitulate the events taking place during the early stages of zygote formation in relation to DNA demethylation and modified cytosine interpretation. Genetic loci associated with the regulation of pluripotency such as Oct4 and Nanog promoters are methylated during male gametogenesis and undergo demethylation in the oocyte after fertilization (Gu et al., 2011). Therefore factors binding to the different cytosine counterparts would reflect the demethylation pathway employed by the oocyte to demethylate and activate genes during fertilization. The results obtained from the differential

modification screen would generate data corroborating either one of the already proposed mechanisms of demethylation (TET oxidation, BER, NER etc.) or suggest new perspectives in oocyte demethylation. Additionally, multiple studies have suggested that the oxidation products of 5mC, aside from their role in demethylation, also come with their own epigenetic significance. Iurlaro et al (Iurlaro et al., 2013) showed that 5fC attracts factors linked with transcriptional repression resembling the attributes of 5mC. Different cell lineages appear to interpret cytosine modifications in different ways, judging by the different factors that bind each cytosine modification, as reported by Spruijt et al (Spruijt et al., 2013). Therefore with a screen of modification specific binders information concerning how the oocyte interprets such marks would also surface. Factors associated with gene repression or activation would suggest whether the respective cytosine modification is interpreted as a repressor or activator chromatin mark, respectively.

The molecules binding mNanog promoter in our screen were identified through mass spectrometry and binders showing enrichment for each cytosine modification were revealed with label free quantification.

Surprisingly, the screen did not identify TET family proteins binding to any of the cytosine modifications, although Tet2 and 3 mRNA is found in *axolotl* oocytes (transcriptome analysis, Matt Loose). Interestingly a high number of DNA repair pathway associated molecules were identified to bind 5mC, 5fC and 5hmC-mNanog promoters. The BER associated factors identified were: APEX1, PNKP, POLB, LIG1, TDP1, RECQL, RECQL5, FEN1, PCNA and PARP1. The BER pathway is recapitulated almost in its entirety in the modified cytosines, i.e. all the steps required for its completion are represented with an identified factor, except for the initial DNA glycosylase activity, responsible for the base removal step (Table 5.3). Interestingly two DNA glycosylases, NTHL1 and MPG, were mostly enriched in non-modified cytosine mNanog promoter, potentially indicating that modified cytosines repel DNA glycosylases. Alternatively, the process of damaged/modified base recognition and removal could also be

faster in the case of modified cytosines resulting in transient DNA glycosylase binding and base removal, and a swift progression to the successive BER repair steps. The latter suggestion is also based on the fact that the BER associated factors that were enriched in the modified cytosine screens were downregulated in the case of non-modified cytosine. As discussed above, APEX1 and PARP1 have been implicated with oocyte demethylation in previous reports (Hajkova et al., 2010). Additionally, the presence of FEN1 and PCNA in the screen is more indicative of LP-BER (Long Patch) in the recognition and treatment of modified DNA as previously reported by Santos et al (Santos et al., 2013). While there is significant enrichment of BER factors binding to the modified cytosine containing mNanog promoter, it is difficult to discern whether the BER is more specific to either one of the cytosine modifications. There appear to be preferences in terms of some BER factors being more enriched in 5mC, 5hmC or 5fC, like FEN1 and TDP1 that are enriched in 5fC and 5mC, respectively. However, in average the BER pathway appears robustly represented in all three cytosine modifications, potentially reflecting a universal recognition and treatment of modified cytosines in the context of the oocyte.

The NER associated factors showing enrichment in the modified cytosine screen were XPC, CETN2, RAD23B, ERCC6L, RPA1, RPA3, LIG1 and POLB. Because LIG1 and POLB are implicated in both NER and BER DNA repair, they are classified as contributors of both pathways in the context of the present screen. While NER repair involves a large number of participating molecules, a significant recapitulation of the steps involved in NER is still observed, much like in the case of BER. In this case, as well, the NER pathway factors showed enrichment across the modified cytosines while being significantly downregulated in non-modified cytosine-mNanog promoter. However, in this case the DNA repair contributors were more enriched in 5mC, as compared to 5hmC and 5fC. Notably though, the NER factors were more enriched in 5fC than 5hmC. The XPC complex, composing of XPC, RAD23B and CETN2, was found three-fold enriched in 5mC compared to non-modified cytosine

showing binding specificity towards methylated DNA. The XPC complex carries out global genome damage recognition in NER, while its binding affinity to a lesion site is analogous to the degree of helical distortion (Sugasawa, 2008). Interestingly, 5mC was found to decrease DNA flexibility (Ngo et al., 2016), an attribute that potentially impacts XPC recognition. The ERCC6L factor, a paralog of ERCC6, which is implicated in NER repair, was found as a highly enriched 5mC binder at a surprising 11-fold enrichment. While it is not known yet whether ERCC6L functions in the same manner as ERCC6, its 5mC specificity makes it an interesting molecule for further study. ERCC6L potentially plays a role in murine embryo development, judging by its reported upregulated and downregulated expression in embryos and adults, respectively (Chen Xianggui 2004). Notably, the expression profile of ERCC6L was found very similar to that of UHRF1, resulting in the two clustering together in the heatmap of Chapter 5 (Fig. 5.3). The clustering indicates that the two factors are binding 5mC with high specificity, while there might also be a direct interaction between the two; however confirmatory experiments are needed to support the latter proposition. Additionally, 5mC and 5fC-mNanog promoter were also enriched for core histone molecules that are associated with compact and transcriptionally inactive DNA, thus suggesting a cross-talk between histones, methylation and DNA repair in the context of oocyte demethylation. Oocyte demethylation through NER has been previously proposed (Schafer et al., 2010, Barreto et al., 2007), while the oocyte reportedly contains the entire NER machinery and is very efficient at repairing bulky DNA adducts (Ackerman and Iakoucheva, 2000).

Based on the above findings I suggest that DNA modifications trigger DNA repair mechanism recognition in the context of the oocyte. While the BER pathway shows drastic enrichment in all modified cytosines, it does not appear to preferentially target either 5mC, 5hmC or 5fC. An interesting question to ponder upon based on this findings would be if sperm DNA was hyper-hydroxymethylated or hyper-formyliated, rather than hypermethylated, would it still be subjected to de-hydroxymethylation and

de-formyliation, respectively. So is the oocyte at this stage “programmed” to recognise and correct through DNA repair cytosine modifications in general, or is it only able to demethylate, rather than de-modify in general. To find that the BER pathway is enriched across 5mC, 5hmC and 5fC hints that potentially not only 5mC but 5hmC and 5fC are removed from DNA via DNA repair. It is possible that different factors that specifically recognise each modification might be acting as recruiters of BER, however that remains to be studied.

The enrichment profile of the NER factors across modified cytosines also shows that potentially the oocyte is capable of recognizing all cytosine modifications as DNA lesions in need of repair. However, NER appears to have higher enrichment in 5mC, compared to the other cytosine counterparts, thus demonstrating a 5mC preference. The XPC complex appears to be a specific recognizer of the 5mC mark, although its specificity might not be related to direct 5mC recognition. The XPC complex is primarily involved in the recognition of DNA helix distorting lesions in the course of global genome NER (GG-NER). So potentially in the case of oocyte demethylation, it is the impact of the 5mC mark that is being recognized from the XPC complex. Interestingly the XPC complex has been recently found to selectively activate Nanog, Oct4 and Sox2 in stem cells and the complex was henceforth termed Stem Cell Coactivator (SCC) (Fong et al., 2011). This finding along with the reports of Xenopus oocyte lysate successfully activating Oct4 promoter by demethylation potentially suggest that NER, aside from being a generic DNA repair mechanism might also have the ability to specifically target pluripotency associated genetic regions. However, that remains to be further delineated. The fact that 5fC is the next modification more enriched in NER factors after 5mC indicates that 5mC and 5fC might share similar biochemical attributes in the cell context.

Interestingly my data also suggest that the two DNA repair pathways of NER and BER might act in concert during oocyte mediated DNA demethylation. This has been previously suggested as a possibility when Gadd45a was found to be implicated in BER and NER repair and in both cases affecting

demethylation (Niehrs and Schafer, 2012). More specifically the model suggested by Niehrs and Schafer (Niehrs and Schafer, 2012) called for NER DNA repair commencing through the mediation of Gadd45a on one DNA strand, and BER DNA repair occurring on the other DNA strand. According to the model, the hemimethylated DNA resulting after NER demethylation would act as substrate for the recruitment of BER factors and the commencement of DNA repair. While the present study did not identify Gadd45a, the concomitant identification of a substantial number of BER and NER factors, significantly recapitulating both DNA repair mechanisms reflects their potential parallel activity in oocyte demethylation. More interestingly my screen identified UHRF1, a known binder of methylated cytosines in a CpG context, which has been implicated in methylation maintenance during DNA replication by preferentially binding hemimethylated DNA and recruiting DNMT1 (Liu et al., 2013). Notably, it was recently reported that UHRF1 binds specifically to hemimethylated DNA and repels fully methylated CpGs (Bianchi and Zangi, 2013). While as discussed in Chapter 5 the presence of UHRF1 at this oocyte stage is unlikely to be related to DNA methylation maintenance, its presence might be linked to the recognition of hemimethylation and subsequent recruitment of BER factors. UHRF1 has been recently implicated in DNA repair though the recognition of DNA lesions (Tian et al., 2015) or the interaction with the DNA glycosylase MPG (Liang et al., 2013). Additionally UHRF1 has been reported to associate with PARP1 (De Vos et al., 2014), which was also identified as a binder enriched in modified cytosines, 5mC and 5fC in particular. PARP1 catalyzes paryliation, hence adding a tail of ADP ribose units to its targets including itself. PARP1 has been involved with the BER repair pathway through its ability to parylate and recruit BER factors at the site of a DNA lesion (Wei and Yu, 2016, Horton et al., 2014). What unfolds with the aforementioned reports and the factors identified in my screen is a model of UHRF1 and PARP1 mediated recognition and paryliation of a hemimethylated DNA region, and the subsequent recruitment of BER factors. In the case where the DNA repair pathways of BER and NER cooperate in oocyte demethylation, GG-NER could commence without the need of a specific recruiter as the XPC

complex recognizes DNA lesions presenting helix distortion. However, in the case of BER a recruiter would be needed to target the BER mechanism to the remaining methylated cytosines. I hereby suggest that UHRF1 that is capable of robustly recognizing hemimethylated CpG and was identified ultra-specifically enriched as a 5mC binder (35-fold enrichment compared to C-mNanog promoter), is the factor that targets the hemimethylated DNA that is left after NER activity commences. PARP1 that has been found to associate with UHRF1 then acts by parylating itself and other potential surrounding proteins, thus attracting additional BER factors to complete BER repair (Fig. 6.1).

While additional experimentation is needed to corroborate the model hereby proposed, the present screen provides significant indications that a BER/NER collaborative activity might be responsible for the demethylation wave occurring in oocyte demethylation of the previously discussed Phase I, i.e. zygotic stages of PN0-PN2. The screen presented in Chapter 5 is the first attempt to identify modified cytosine binders by utilizing a differential DNA pull down screen in oocytes. My study has successfully established a protocol for efficient pull down and identification of modified and non-modified cytosine binders in the context of the oocyte.

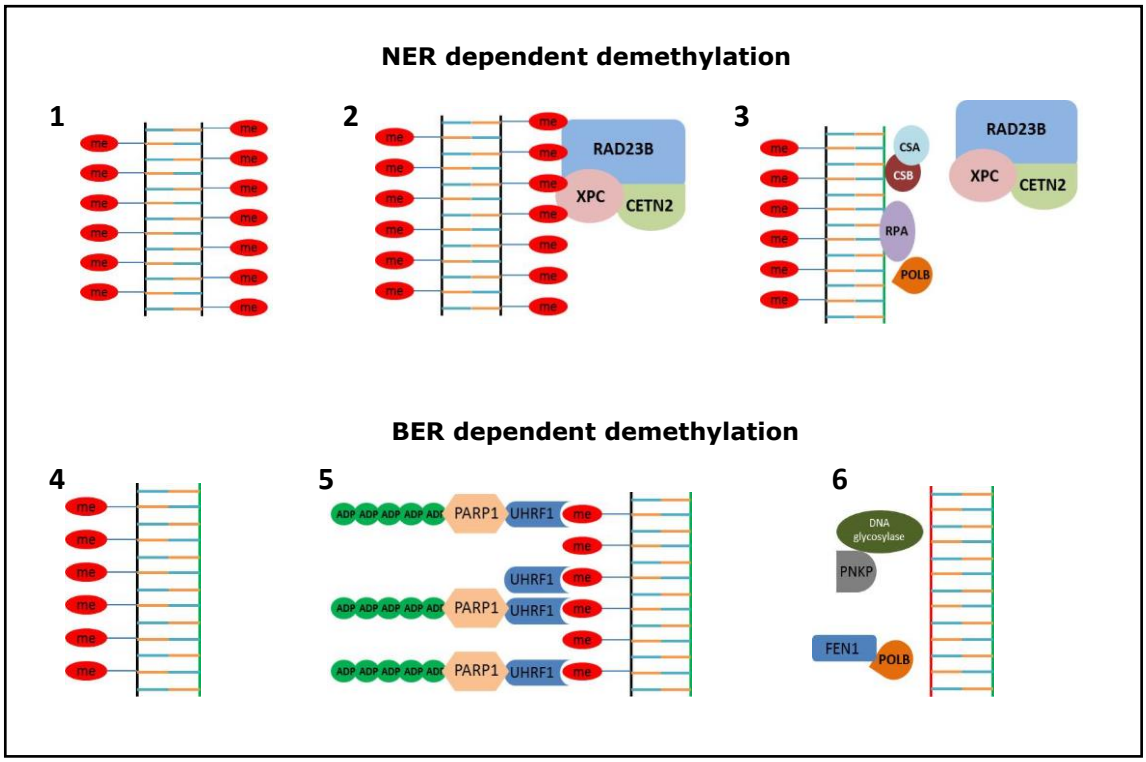



**Figure 6.1. Model proposing NER/BER dependent demethylation in the context of the oocyte. Fully methylated DNA (1) is recognised by the XPC complex, through its impact on DNA conformation (2). The XPC complex then recruits additional NER factors that complete demethylation through DNA repair on one DNA strand (3). The completion of NER results in DNA hemimethylation (4). The hemimethylated DNA is substrate for UHRF1 binding which subsequently interacts with PARP1 (5). PARP1 acts through paryliating itself and potentially other proteins, thus creating poly-ADP-ribose tails that are recognized by BER associated proteins (5). The completion of BER repair results in non-methylated (non-modified) DNA (6).**

## Future perspectives

The significant enrichment of BER and NER factors in our modified cytosines screen is indicative of their parallel activity in the context of the oocyte. Additionally, the presence of UHRF1 and PARP1 as enriched 5mC binders hints a mechanism of BER recruitment to the hemimethylated DNA resulting after the completion of demethylation through NER repair.

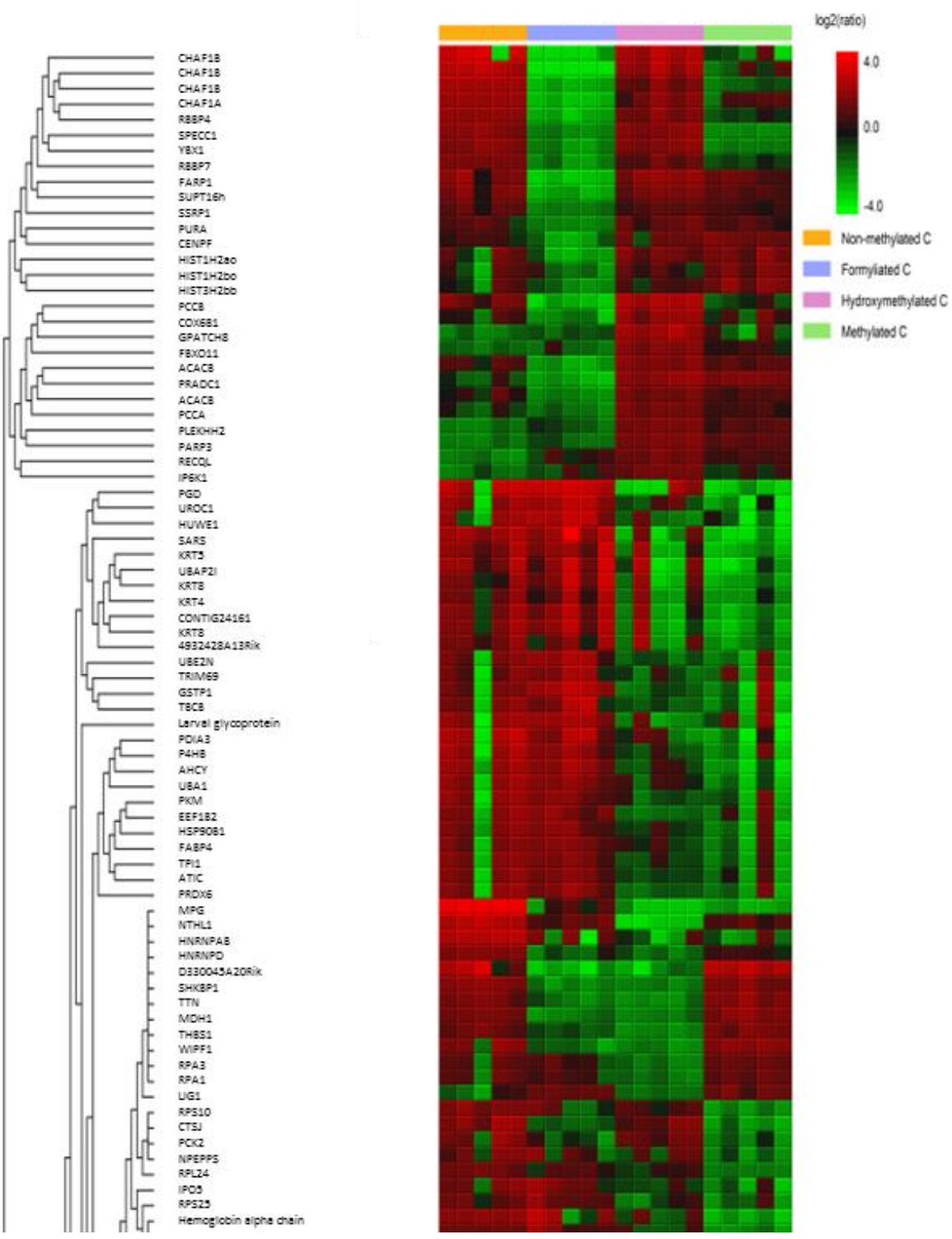
However, additional experiments are required to confirm the proposed model. While our screen successfully identified a multitude of binders, we do not know what the impact of the oocyte treatment was on each cytosine modification. The question of whether the modified cytosines and 5mC in specific were repaired and replaced by non-modified cytosines remains. To obtain such an insight the same DNA pull down experiment could be conducted, without however concentrating on protein identification. The incubated DNA could be retrieved and subjected to nucleotide LS/MS to delineate its cytosine modification profile. If successfully obtained, such data would prove that oocyte lysate treated DNA is subjected to demethylation in the oocyte.

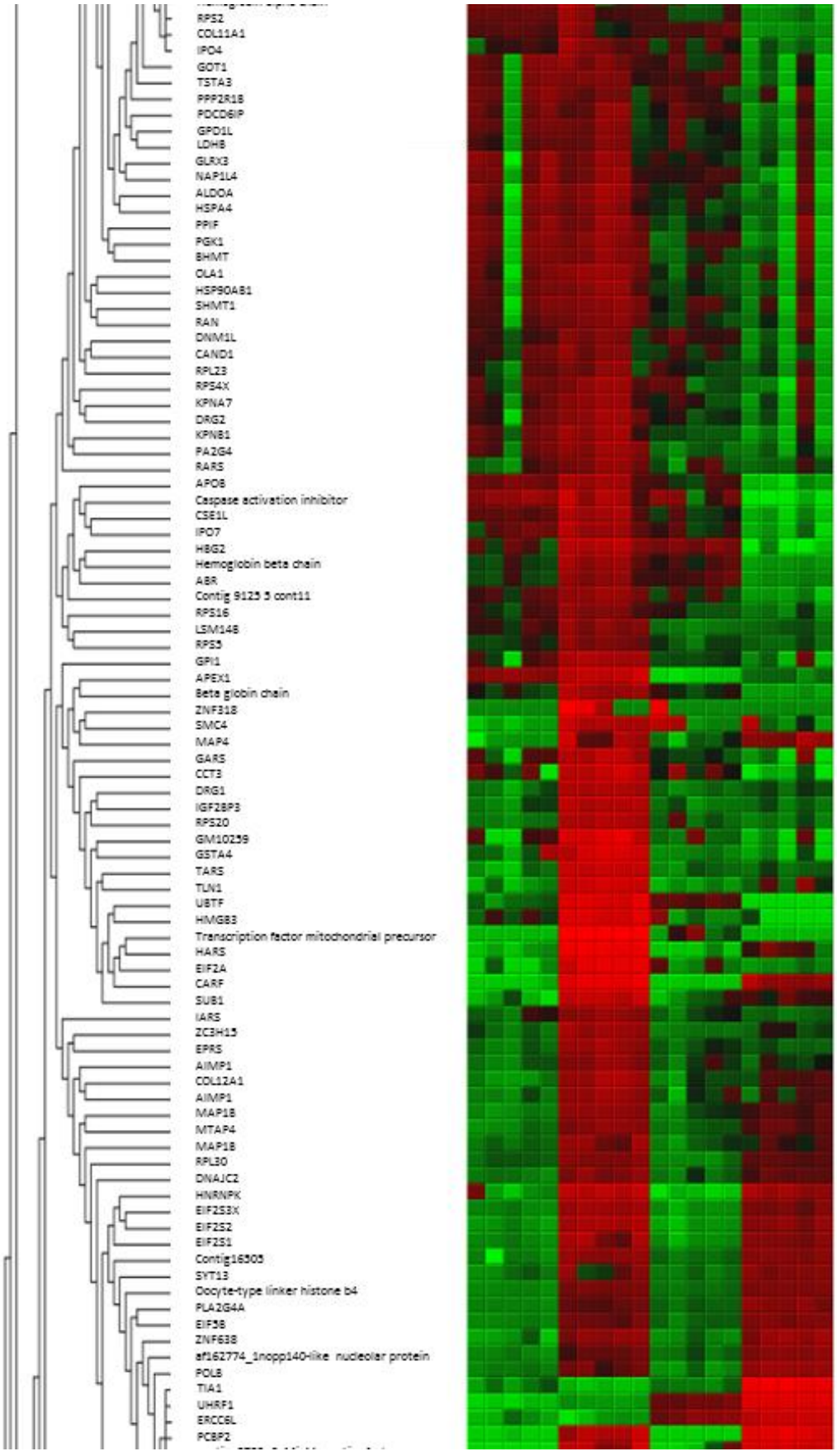
To further clarify the role of NER and BER pathways in DNA demethylation, factors of each pathway could be inhibited in the oocyte either with the use of morpholinos or specific antibodies (or agents) to block their activity. With the factor(s) blocked I would attempt to carry out similar DNA pull down experiments aiming to identify protein binders through mass spectrometry as well as study the treated DNA for cytosine modifications through nucleotide LC/MS. I would be able to see how inhibiting given factors impacts DNA modifications as well see how inhibition impacts DNA binding. To study whether the observed BER/NER activity is sequence specific or specifically related to mNanog promoter an additional DNA sequence could be added to the experiment. Moreover, the BER and NER pathway contribution to DNA demethylation could be studied in the context of ESC. Differential cytosine modification screen with the mNanog promoter could be carried out with ESC

lysates to delineate whether the same or different/additional molecules would bind and show similar activity in the context of ESC.

# Appendix

## Heatmap and identified proteins in the 5mC/5hmC/5fC and non-modified cytosine screen




RPS2
COL11A1
IPO4
GOT1
TSTA3
PPP2R1B
PDCD6IP
GPD1L
LDHB
GLRX3
NAP1L4
ALDOA
HSPA4
PPIF
PGK1
BHMT
OLA1
HSP90AB1
SHMT1
RAN
DNM1L
CAND1
RPL23
RPS4X
KPNA7
DRG2
KPNB1
PA2G4
RARS
APOB
Caspase activation inhibitor
CSE1L
IPO7
HBG2
Hemoglobin beta chain
ABR
Contig 9125 3 cont11
RPS16
LSM14B
RPS3
GPI1
APEX1
Beta globin chain
ZNF318
SMC4
MAP4
GARS
CCT3
DRG1
IGF2BP3
RPS20
GM10259
GSTA4
TARS
TLN1
UBTF
HMGB3
Transcription factor mitochondrial precursor
HARS
EIF2A
CARF
SUB1
IARS
ZC3H15
EPRS
AIMP1
COL12A1
AIMP1
MAP1B
MTAP4
MAP1B
RPL30
DNAJC2
HNRNPK
EIF2S3X
EIF2S2
EIF2S1
Contig16505
SYT13
Oocyte-type linker histone b4
PLA2G4A
EIF3B
ZNF638
af162774_1nopp140-like nucleolar protein
POLB
TIA1
UHRF1
ERCC6L
PCBP2

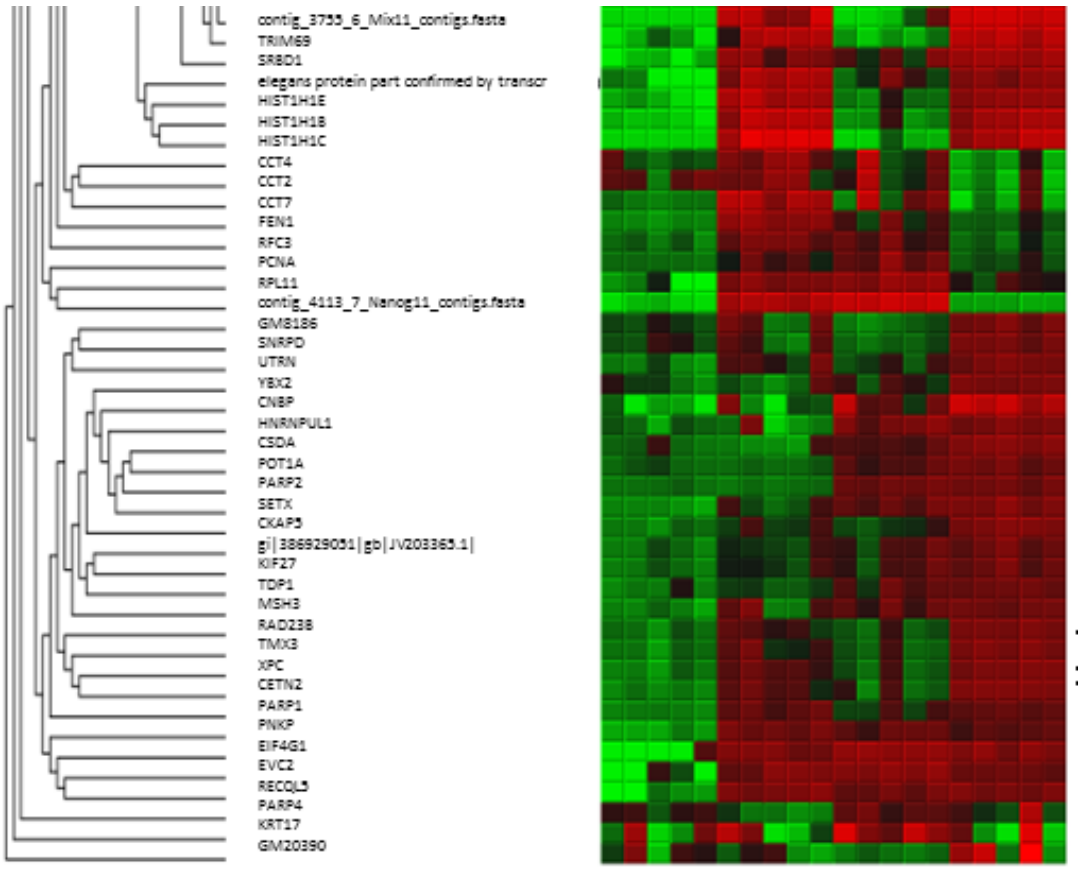



**Figure 7.1. (Fig. 5.3 annotated) Heat map showing the relative protein enrichment of each cytosine modification. Here the respective proteins appear annotated.**

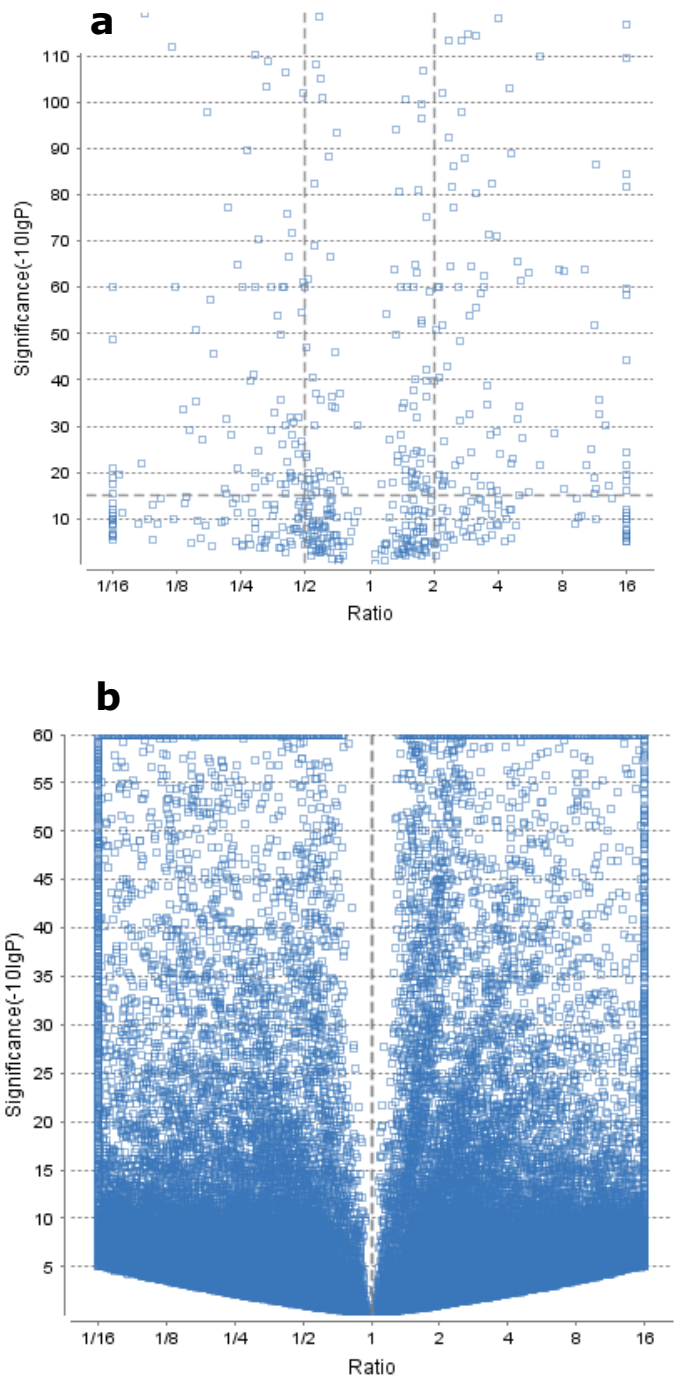



**Figure 7.2. Volcano plot of the analysed proteins and peptides. a) For proteins. The vertical dotted grey line represents the threshold set for the label free quantification. b) For peptides.**

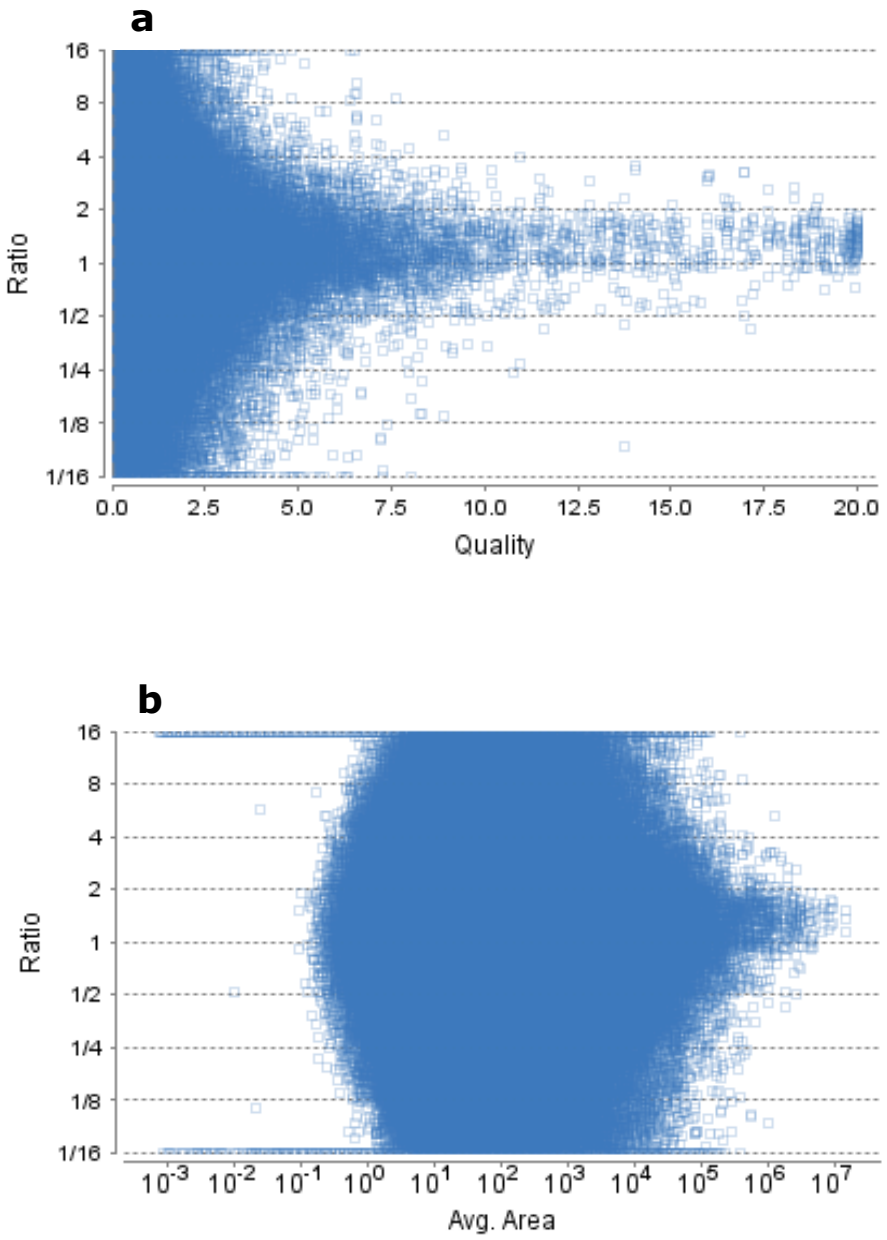



**Figure 7.3.The distribution of feature vector (peptides) ratio. a) By quality. The higher quality of a peptide indicates that the peptide is more quantifiable. b) By intensity. Directly related to the MS signal intensity Higher intensity peptide features are more reproducible across samples.**

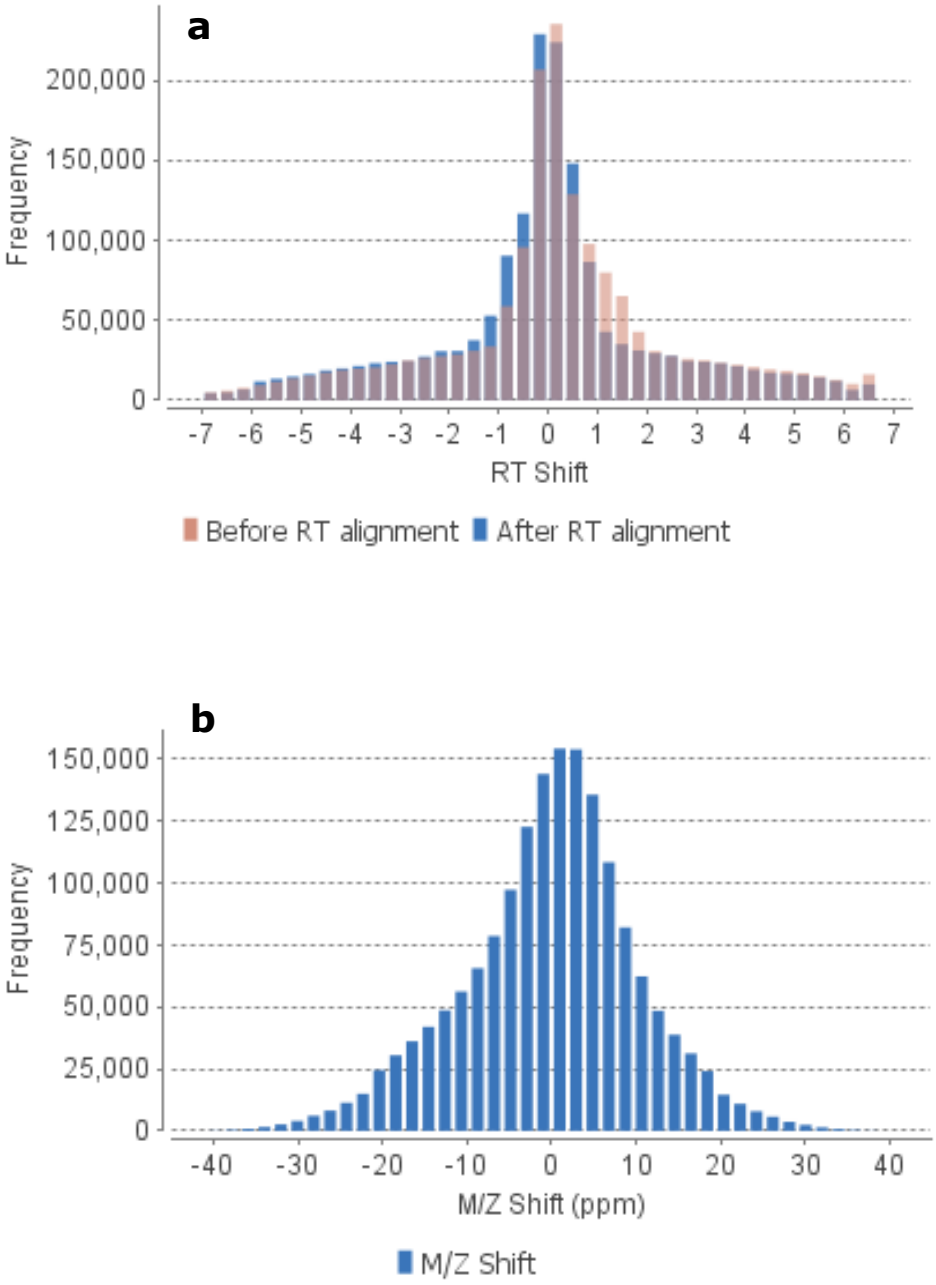



**Figure 7.4. RT and m/z Shift Distribution. a) Retention time shift distribution to the base sample. b) Distribution of the m/z shift (in ppm) of the data with respect to the base sample.**

## Summary Tables

| | |
|---|---|
| Retention time | ≥0 |
| Retention time | ≤120 |
| Feature fold change | ≥1 |
| Quality | ≥0 |
| Avg. Area | ≥0E0 |
| Charge | ≥1 |
| Charge | ≤10 |
| Confident sample number | ≥1 |
| With peptide ID | false |
| Normalization | Use TIC |
| Protein significance | ≥15 |
| Protein fold change | ≥2 |
| Confident unique supports | ≥1 |

**Table 7.1. Result filtration parameters.**

| | |
|---|---|
| Features | 4324340 |
| Features with ID | 33987 |
| Feature vectors | 1034540 |
| Feature vectors with ID | 5316 |
| Protein groups | 207 |

**Table 7.2. Feature vector and protein result filtration parameters.**

Quantification type: Label free quantification
Mass Error Tolerance: 20.0 ppm
Retention Time Shift Tolerance: 6.0 min
Dependent on PID: 3,9,44,49,79,104,14,39,109,114,119,124,19,34,
64,69,84,89,24,29,54,59,94,99
FDR Threshold: 1%
Samples: 20 samples in 4 groups
: Sample 2 Sample 3 Sample 4 Sample 5
: Sample 6 Sample 7 Sample 9 Sample 10
: Sample 11 Sample 12 Sample 13 Sample 14
: Sample 16 Sample 17 Sample 18 Sample 19
: Sample 20 Sample 21 Sample 22 Sample 24
Reference Sample: Sample 2 (auto detected)
Training Samples: Sample 2, Sample 3 (auto detected)

**Table 7.3. Search parameters for the heatmap in Fig. 5.3.**

# References


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

ABOUSSEKHRA, A., BIGGERSTAFF, M., SHIVJI, M. K., VILPO, J. A., MONCOLLIN, V., PODUST, V. N., PROTIC, M., HUBSCHER, U., EGLY, J. M. & WOOD, R. D. 1995. Mammalian DNA nucleotide excision repair reconstituted with purified protein components. *Cell,* 80**,** 859-68.

ACKERMAN, E. J. & IAKOUCHEVA, L. M. 2000. Nucleotide excision repair in oocyte nuclear extracts from Xenopus laevis. *Methods,* 22**,** 188-93.

ALBERIO, R., JOHNSON, A. D., STICK, R. & CAMPBELL, K. H. 2005. Differential nuclear remodeling of mammalian somatic cells by Xenopus laevis oocyte and egg cytoplasm. *Exp Cell Res,* 307**,** 131-41.

ALBERIO, R., JOHNSON, A. D. 2012. *Epigenetic reprogramming with oocyte molecules* [Online]. [Accessed].

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

AMOUROUX, R., NASHUN, B., SHIRANE, K., NAKAGAWA, S., HILL, P. W., D'SOUZA, Z., NAKAYAMA, M., MATSUDA, M., TURP, A., NDJETEHE, E., ENCHEVA, V., KUDO, N. R., KOSEKI, H., SASAKI, H. & HAJKOVA, P. 2016. De novo DNA methylation drives 5hmC accumulation in mouse zygotes. *Nat Cell Biol,* 18**,** 225-33.

ANDREU-VIEYRA, C. V., CHEN, R., AGNO, J. E., GLASER, S., ANASTASSIADIS, K., STEWART, A. F. & MATZUK, M. M. 2010. MLL2 is required in oocytes for bulk histone 3 lysine 4 trimethylation and transcriptional silencing. *PLoS Biol,* 8.

ANGELOV, D., BONDARENKO, V. A., ALMAGRO, S., MENONI, H., MONGELARD, F., HANS, F., MIETTON, F., STUDITSKY, V. M., HAMICHE, A., DIMITROV, S. & BOUVET, P. 2006. Nucleolin is a histone chaperone with FACT-like activity and assists remodeling of nucleosomes. *EMBO J,* 25**,** 1669-79.

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

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

ARIOKA, Y., WATANABE, A., SAITO, K. & YAMADA, Y. 2012. Activation-induced cytidine deaminase alters the subcellular localization of Tet family proteins. *PLoS One,* 7**,** e45031.

ARTUS, J., PILISZEK, A. & HADJANTONAKIS, A. K. 2011. The primitive endoderm lineage of the mouse blastocyst: sequential transcription factor activation and regulation of differentiation by Sox17. *Dev Biol,* 350**,** 393-404.

AUMAN, H. J., NOTTOLI, T., LAKIZA, O., WINGER, Q., DONALDSON, S. & WILLIAMS, T. 2002. Transcription factor AP-2 gamma is essential in the extra-embryonic lineages for early postimplantation development (vol 129, pg 2733 2001). *Development,* 129**,** 3280-3280.

AVILION, A. A., NICOLIS, S. K., PEVNY, L. H., PEREZ, L., VIVIAN, N. & LOVELL-BADGE, R. 2003. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev,* 17**,** 126-40.

AVVAKUMOV, G. V., WALKER, J. R., XUE, S., LI, Y., DUAN, S., BRONNER, C., ARROWSMITH, C. H. & DHE-PAGANON, S. 2008. Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature,* 455**,** 822-5.

AWE, J. P. & BYRNE, J. A. 2013. Identifying Candidate Oocyte Reprogramming Factors Using Cross-Species Global Transcriptional Analysis. *Cell Reprogram*.

BACHMAN, M., URIBE-LEWIS, S., YANG, X., BURGESS, H. E., IURLARO, M., REIK, W., MURRELL, A. & BALASUBRAMANIAN, S. 2015. 5-Formylcytosine can be a stable DNA modification in mammals. *Nat Chem Biol,* 11**,** 555-7.

BACHMAN, M., URIBE-LEWIS, S., YANG, X., WILLIAMS, M., MURRELL, A. & BALASUBRAMANIAN, S. 2014. 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat Chem,* 6**,** 1049-55.

BACHVAROVA, R. F., CROTHER, B. I. & JOHNSON, A. D. 2009. Evolution of germ cell development in tetrapods: comparison of urodeles and amniotes. *Evol Dev,* 11**,** 603-9.

BALBOULA, A. Z., STEIN, P., SCHULTZ, R. M. & SCHINDLER, K. 2015. RBBP4 regulates histone deacetylation and bipolar spindle assembly during oocyte maturation in the mouse. *Biol Reprod,* 92**,** 105.

BALL, D. J., GROSS, D. S. & GARRARD, W. T. 1983. 5-methylcytosine is localized in nucleosomes that contain histone H1. *Proc Natl Acad Sci U S A,* 80**,** 5490-4.

BAO, X., WU, H., ZHU, X., GUO, X., HUTCHINS, A. P., LUO, Z., SONG, H., CHEN, Y., LAI, K., YIN, M., XU, L., ZHOU, L., CHEN, J., WANG, D., QIN, B., FRAMPTON, J., TSE, H. F., PEI, D., WANG, H., ZHANG, B. & ESTEBAN, M. A. 2015. The p53-induced lincRNA-p21 derails somatic cell reprogramming by sustaining H3K9me3 and CpG methylation at pluripotency gene promoters. *Cell Res,* 25**,** 80-92.

BARRETO, G., SCHAFER, A., MARHOLD, J., STACH, D., SWAMINATHAN, S. K., HANDA, V., DODERLEIN, G., MALTRY, N., WU, W., LYKO, F. & NIEHRS, C. 2007. Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. *Nature,* 445**,** 671-5.

BARTKE, T., VERMEULEN, M., XHEMALCE, B., ROBSON, S. C., MANN, M. & KOUZARIDES, T. 2010. Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell,* 143**,** 470-84.

BEAUJEAN, N., HARTSHORNE, G., CAVILLA, J., TAYLOR, J., GARDNER, J., WILMUT, I., MEEHAN, R. & YOUNG, L. 2004. Non-conservation of mammalian preimplantation methylation dynamics. *Curr Biol,* 14**,** R266-7.

BECHEREL, O. J., YEO, A. J., STELLATI, A., HENG, E. Y., LUFF, J., SURAWEERA, A. M., WOODS, R., FLEMING, J., CARRIE, D., MCKINNEY, K., XU, X., DENG, C. & LAVIN, M. F. 2013. Senataxin plays an essential role with DNA damage response proteins in

meiotic recombination and gene silencing. *PLoS Genet,* 9**,** e1003435.
BECK, C., ROBERT, I., REINA-SAN-MARTIN, B., SCHREIBER, V. & DANTZER, F. 2014. Poly(ADP-ribose) polymerases in double-strand break repair: focus on PARP1, PARP2 and PARP3. *Exp Cell Res,* 329**,** 18-25.
BELOTSERKOVSKAYA, R., OH, S., BONDARENKO, V. A., ORPHANIDES, G., STUDITSKY, V. M. & REINBERG, D. 2003. FACT facilitates transcription-dependent nucleosome alteration. *Science,* 301**,** 1090-3.
BEYER, T. A., WEISS, A., KHOMCHUK, Y., HUANG, K., OGUNJIMI, A. A., VARELAS, X. & WRANA, J. L. 2013. Switch enhancers interpret TGF-beta and Hippo signaling to control cell fate in human embryonic stem cells. *Cell Rep,* 5**,** 1611-24.
BHATTACHARYA, S. K., RAMCHANDANI, S., CERVONI, N. & SZYF, M. 1999. A mammalian protein with specific demethylase activity for mCpG DNA. *Nature,* 397**,** 579-83.
BHOUMIK, A. & RONAI, Z. 2008. ATF2: a transcription factor that elicits oncogenic or tumor suppressor activities. *Cell Cycle,* 7**,** 2341-5.
BIAN, Y., ALBERIO, R., ALLEGRUCCI, C., CAMPBELL, K. H. & JOHNSON, A. D. 2009. Epigenetic marks in somatic chromatin are remodelled to resemble pluripotent nuclei by amphibian oocyte extracts. *Epigenetics,* 4**,** 194-202.
BIANCHI, C. & ZANGI, R. 2013. UHRF1 discriminates against binding to fully-methylated CpG-Sites by steric repulsion. *Biophys Chem,* 171**,** 38-45.
BIDDLE, A., SIMEONI, I. & GURDON, J. B. 2009. Xenopus oocytes reactivate muscle gene transcription in transplanted somatic nuclei independently of myogenic factors. *Development,* 136**,** 2695-703.
BILODEAU, S., KAGEY, M. H., FRAMPTON, G. M., RAHL, P. B. & YOUNG, R. A. 2009. SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes Dev,* 23**,** 2484-9.
BLAIR, H. J., TOMPSON, S., LIU, Y. N., CAMPBELL, J., MACARTHUR, K., PONTING, C. P., RUIZ-PEREZ, V. L. & GOODSHIP, J. A. 2011. Evc2 is a positive modulator of Hedgehog signalling that interacts with Evc at the cilia membrane and is also found in the nucleus. *BMC Biol,* 9**,** 14.
BOROVIAK, T., LOOS, R., BERTONE, P., SMITH, A. & NICHOLS, J. 2014. The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. *Nat Cell Biol,* 16**,** 516-28.
BOSTICK, M., KIM, J. K., ESTEVE, P. O., CLARK, A., PRADHAN, S. & JACOBSEN, S. E. 2007. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science,* 317**,** 1760-4.
BOURC'HIS, D. & BESTOR, T. H. 2004. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature,* 431**,** 96-9.
BOURC'HIS, D., XU, G. L., LIN, C. S., BOLLMAN, B. & BESTOR, T. H. 2001. Dnmt3L and the establishment of maternal genomic imprints. *Science,* 294**,** 2536-9.
BOYER, L. A., LEE, T. I., COLE, M. F., JOHNSTONE, S. E., LEVINE, S. S., ZUCKER, J. P., GUENTHER, M. G., KUMAR, R. M., MURRAY, H. L., JENNER, R. G., GIFFORD, D. K., MELTON, D. A., JAENISCH, R. & YOUNG, R. A. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell,* 122**,** 947-56.

BRADLEY, A., EVANS, M., KAUFMAN, M. H. & ROBERTSON, E. 1984. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature,* 309**,** 255-6.
BRAMBRINK, T., HOCHEDLINGER, K., BELL, G. & JAENISCH, R. 2006. ES cells derived from cloned and fertilized blastocysts are transcriptionally and functionally indistinguishable. *Proc Natl Acad Sci U S A,* 103**,** 933-8.
BRIGGS, R. & KING, T. J. 1952. Transplantation of Living Nuclei From Blastula Cells into Enucleated Frogs' Eggs. *Proc Natl Acad Sci U S A,* 38**,** 455-63.
BRONS, I. G., SMITHERS, L. E., TROTTER, M. W., RUGG-GUNN, P., SUN, B., CHUVA DE SOUSA LOPES, S. M., HOWLETT, S. K., CLARKSON, A., AHRLUND-RICHTER, L., PEDERSEN, R. A. & VALLIER, L. 2007. Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature,* 448**,** 191-5.
BROOK, F. A. & GARDNER, R. L. 1997. The origin and efficient derivation of embryonic stem cells in the mouse. *Proc Natl Acad Sci U S A,* 94**,** 5709-12.
BRUNIQUEL, D. & SCHWARTZ, R. H. 2003. Selective, stable demethylation of the interleukin-2 gene enhances transcription by an active process. *Nat Immunol,* 4**,** 235-40.
BUEHR, M., MEEK, S., BLAIR, K., YANG, J., URE, J., SILVA, J., MCLAY, R., HALL, J., YING, Q. L. & SMITH, A. 2008. Capture of authentic embryonic stem cells from rat blastocysts. *Cell,* 135**,** 1287-98.
BUI, H. T., WAKAYAMA, S., KISHIGAMI, S., KIM, J. H., VAN THUAN, N. & WAKAYAMA, T. 2008. The cytoplasm of mouse germinal vesicle stage oocytes can enhance somatic cell nuclear reprogramming. *Development,* 135**,** 3935-45.
BURDON, T., CHAMBERS, I., STRACEY, C., NIWA, H. & SMITH, A. 1999. Signaling mechanisms regulating self-renewal and differentiation of pluripotent embryonic stem cells. *Cells Tissues Organs,* 165**,** 131-43.
BYRNE, J. A., SIMONSSON, S., WESTERN, P. S. & GURDON, J. B. 2003. Nuclei of adult mammalian somatic cells are directly reprogrammed to oct-4 stem cell gene expression by amphibian oocytes. *Curr Biol,* 13**,** 1206-13.
CAIAFA, P., GUASTAFIERRO, T. & ZAMPIERI, M. 2009. Epigenetics: poly(ADP-ribosyl)ation of PARP-1 regulates genomic methylation patterns. *FASEB J,* 23**,** 672-8.
CAO, S., BENDALL, H., HICKS, G. G., NASHABI, A., SAKANO, H., SHINKAI, Y., GARIGLIO, M., OLTZ, E. M. & RULEY, H. E. 2003. The high-mobility-group box protein SSRP1/T160 is essential for cell viability in day 3.5 mouse embryos. *Mol Cell Biol,* 23**,** 5301-7.
CATTOGLIO, C., ZHANG, E. T., GRUBISIC, I., CHIBA, K., FONG, Y. W. & TJIAN, R. 2015. Functional and mechanistic studies of XPC DNA-repair complex as transcriptional coactivator in embryonic stem cells. *Proc Natl Acad Sci U S A,* 112**,** E2317-26.
CHAMBERS, I., COLBY, D., ROBERTSON, M., NICHOLS, J., LEE, S., TWEEDIE, S. & SMITH, A. 2003. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell,* 113**,** 643-55.
CHAMBERS, I., SILVA, J., COLBY, D., NICHOLS, J., NIJMEIJER, B., ROBERTSON, M., VRANA, J., JONES, K., GROTEWOLD, L. & SMITH, A. 2007. Nanog safeguards pluripotency and mediates germline development. *Nature,* 450**,** 1230-4.
CHANG, L., LORANGER, S. S., MIZZEN, C., ERNST, S. G., ALLIS, C. D. & ANNUNZIATO, A. T. 1997. Histones in transit: cytosolic histone

complexes and diacetylation of H4 during nucleosome assembly in human cells. *Biochemistry,* 36**,** 469-80.
CHATFIELD, J., O'REILLY, M. A., BACHVAROVA, R. F., FERJENTSIK, Z., REDWOOD, C., WALMSLEY, M., PATIENT, R., LOOSE, M. & JOHNSON, A. D. 2014. Stochastic specification of primordial germ cells from mesoderm precursors in axolotl embryos. *Development,* 141**,** 2429-40.
CHAZAUD, C., YAMANAKA, Y., PAWSON, T. & ROSSANT, J. 2006. Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. *Dev Cell,* 10**,** 615-24.
CHELOUFI, S., ELLING, U., HOPFGARTNER, B., JUNG, Y. L., MURN, J., NINOVA, M., HUBMANN, M., BADEAUX, A. I., EUONG ANG, C., TENEN, D., WESCHE, D. J., ABAZOVA, N., HOGUE, M., TASDEMIR, N., BRUMBAUGH, J., RATHERT, P., JUDE, J., FERRARI, F., BLANCO, A., FELLNER, M., WENZEL, D., ZINNER, M., VIDAL, S. E., BELL, O., STADTFELD, M., CHANG, H. Y., ALMOUZNI, G., LOWE, S. W., RINN, J., WERNIG, M., ARAVIN, A., SHI, Y., PARK, P. J., PENNINGER, J. M., ZUBER, J. & HOCHEDLINGER, K. 2015. The histone chaperone CAF-1 safeguards somatic cell identity. *Nature,* 528**,** 218-24.
CHEN XIANGGUI , L. Y., ZANG MINGXI , PEI XINRONG , XU YAJUN , GAO LIFANG 2004. cDNA cloning and expression analysis of mouse gene encoding the protein Ercc6l which is a novel member of SNF2 family. *Prog Biochem Biophy,* 31**,** 443–448.
CHEW, J. L., LOH, Y. H., ZHANG, W., CHEN, X., TAM, W. L., YEAP, L. S., LI, P., ANG, Y. S., LIM, B., ROBSON, P. & NG, H. H. 2005. Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. *Mol Cell Biol,* 25**,** 6031-46.
CHIA, N. Y., CHAN, Y. S., FENG, B., LU, X., ORLOV, Y. L., MOREAU, D., KUMAR, P., YANG, L., JIANG, J., LAU, M. S., HUSS, M., SOH, B. S., KRAUS, P., LI, P., LUFKIN, T., LIM, B., CLARKE, N. D., BARD, F. & NG, H. H. 2010. A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature,* 468**,** 316-20.
CHIN, M. H., MASON, M. J., XIE, W., VOLINIA, S., SINGER, M., PETERSON, C., AMBARTSUMYAN, G., AIMIUWU, O., RICHTER, L., ZHANG, J., KHVOROSTOV, I., OTT, V., GRUNSTEIN, M., LAVON, N., BENVENISTY, N., CROCE, C. M., CLARK, A. T., BAXTER, T., PYLE, A. D., TEITELL, M. A., PELEGRINI, M., PLATH, K. & LOWRY, W. E. 2009. Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell,* 5**,** 111-23.
CHOU, D. M., ADAMSON, B., DEPHOURE, N. E., TAN, X., NOTTKE, A. C., HUROV, K. E., GYGI, S. P., COLAIACOVO, M. P. & ELLEDGE, S. J. 2010. A chromatin localization screen reveals poly (ADP ribose)-regulated recruitment of the repressive polycomb and NuRD complexes to sites of DNA damage. *Proc Natl Acad Sci U S A,* 107**,** 18475-80.
CICCARONE, F., KLINGER, F. G., CATIZONE, A., CALABRESE, R., ZAMPIERI, M., BACALINI, M. G., DE FELICI, M. & CAIAFA, P. 2012. Poly(ADP-ribosyl)ation acts in the DNA demethylation of mouse primordial germ cells also with DNA damage-independent roles. *PLoS One,* 7**,** e46927.
CONESA, C. & ACKER, J. 2010. Sub1/PC4 a chromatin associated protein with multiple functions in transcription. *RNA Biol,* 7**,** 287-90.
COSTA, Y., DING, J., THEUNISSEN, T. W., FAIOLA, F., HORE, T. A., SHLIAHA, P. V., FIDALGO, M., SAUNDERS, A., LAWRENCE, M., DIETMANN, S., DAS, S., LEVASSEUR, D. N., LI, Z., XU, M., REIK,

W., SILVA, J. C. & WANG, J. 2013. NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature,* 495**,** 370-4.

COUCOUVANIS, E. & MARTIN, G. R. 1999. BMP signaling plays a role in visceral endoderm differentiation and cavitation in the early mouse embryo. *Development,* 126**,** 535-46.

CRAWFORD, D. J., LIU, M. Y., NABEL, C. S., CAO, X. J., GARCIA, B. A. & KOHLI, R. M. 2016. Tet2 Catalyzes Stepwise 5-Methylcytosine Oxidation by an Iterative and de novo Mechanism. *J Am Chem Soc,* 138**,** 730-3.

DAS, C., HIZUME, K., BATTA, K., KUMAR, B. R., GADAD, S. S., GANGULY, S., LORAIN, S., VERREAULT, A., SADHALE, P. P., TAKEYASU, K. & KUNDU, T. K. 2006. Transcriptional coactivator PC4, a chromatin-associated protein, induces chromatin condensation. *Mol Cell Biol,* 26**,** 8303-15.

DAVIS, T. L., YANG, G. J., MCCARREY, J. R. & BARTOLOMEI, M. S. 2000. The H19 methylation imprint is erased and re-established differentially on the parental alleles during male germ cell development. *Hum Mol Genet,* 9**,** 2885-94.

DAWLATY, M. M., BREILING, A., LE, T., BARRASA, M. I., RADDATZ, G., GAO, Q., POWELL, B. E., CHENG, A. W., FAULL, K. F., LYKO, F. & JAENISCH, R. 2014. Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. *Dev Cell,* 29**,** 102-11.

DE VOS, M., EL RAMY, R., QUENET, D., WOLF, P., SPADA, F., MAGROUN, N., BABBIO, F., SCHREIBER, V., LEONHARDT, H., BONAPACE, I. M. & DANTZER, F. 2014. Poly(ADP-ribose) polymerase 1 (PARP1) associates with E3 ubiquitin-protein ligase UHRF1 and modulates UHRF1 biological functions. *J Biol Chem,* 289**,** 16223-38.

DE VOS, M., SCHREIBER, V. & DANTZER, F. 2012. The diverse roles and clinical relevance of PARPs in DNA damage repair: current state of the art. *Biochem Pharmacol,* 84**,** 137-46.

DEB-RINKER, P., LY, D., JEZIERSKI, A., SIKORSKA, M. & WALKER, P. R. 2005. Sequential DNA methylation of the Nanog and Oct-4 upstream regions in human NT2 cells during neuronal differentiation. *J Biol Chem,* 280**,** 6257-60.

DEMPSEY, L. A., LI, M. J., DEPACE, A., BRAY-WARD, P. & MAIZELS, N. 1998. The human HNRPD locus maps to 4q21 and encodes a highly conserved protein. *Genomics,* 49**,** 378-84.

DIETRICH, J. E. & HIIRAGI, T. 2007. Stochastic patterning in the mouse pre-implantation embryo. *Development,* 134**,** 4219-31.

DING, J., XU, H., FAIOLA, F., MA'AYAN, A. & WANG, J. 2012. Oct4 links multiple epigenetic pathways to the pluripotency network. *Cell Res,* 22**,** 155-67.

DING, L., PASZKOWSKI-ROGACZ, M., NITZSCHE, A., SLABICKI, M. M., HENINGER, A. K., DE VRIES, I., KITTLER, R., JUNQUEIRA, M., SHEVCHENKO, A., SCHULZ, H., HUBNER, N., DOSS, M. X., SACHINIDIS, A., HESCHELER, J., IACONE, R., ANASTASSIADIS, K., STEWART, A. F., PISABARRO, M. T., CALDARELLI, A., POSER, I., THEIS, M. & BUCHHOLZ, F. 2009. A genome-scale RNAi screen for Oct4 modulators defines a role of the Paf1 complex for embryonic stem cell identity. *Cell Stem Cell,* 4**,** 403-15.

DIXON, J. E., ALLEGRUCCI, C., REDWOOD, C., KUMP, K., BIAN, Y., CHATFIELD, J., CHEN, Y. H., SOTTILE, V., VOSS, S. R., ALBERIO, R. & JOHNSON, A. D. 2010. Axolotl Nanog activity in mouse embryonic stem cells demonstrates that ground state pluripotency is conserved from urodele amphibians to mammals. *Development,* 137**,** 2973-80.

DOEGE, C. A., INOUE, K., YAMASHITA, T., RHEE, D. B., TRAVIS, S., FUJITA, R., GUARNIERI, P., BHAGAT, G., VANTI, W. B., SHIH, A., LEVINE, R. L., NIK, S., CHEN, E. I. & ABELIOVICH, A. 2012. Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. *Nature,* 488**,** 652-5.

DU, Q., LUU, P. L., STIRZAKER, C. & CLARK, S. J. 2015. Methyl-CpG-binding domain proteins: readers of the epigenome. *Epigenomics,* 7**,** 1051-73.

DURBIN, J. E., HACKENMILLER, R., SIMON, M. C. & LEVY, D. E. 1996. Targeted disruption of the mouse Stat1 gene results in compromised innate immunity to viral disease. *Cell,* 84**,** 443-50.

EDELBROCK, M. A., KALIYAPERUMAL, S. & WILLIAMS, K. J. 2013. Structural, molecular and cellular functions of MSH2 and MSH6 during DNA mismatch repair, damage signaling and other noncanonical activities. *Mutat Res,* 743-744**,** 53-66.

EISTETTER, H. R., MILLS, A. & ARKINSTALL, S. J. 1993. Signal transduction mechanisms of recombinant bovine neurokinin-2 receptor stably expressed in baby hamster kidney cells. *J Cell Biochem,* 52**,** 84-91.

EMELYANOV, A. V., RABBANI, J., MEHTA, M., VERSHILOVA, E., KEOGH, M. C. & FYODOROV, D. V. 2014. Drosophila TAP/p32 is a core histone chaperone that cooperates with NAP-1, NLP, and nucleophosmin in sperm chromatin remodeling during fertilization. *Genes Dev,* 28**,** 2027-40.

ENGEL, N., TRONT, J. S., ERINLE, T., NGUYEN, N., LATHAM, K. E., SAPIENZA, C., HOFFMAN, B. & LIEBERMANN, D. A. 2009. Conserved DNA methylation in Gadd45a(-/-) mice. *Epigenetics,* 4**,** 98-9.

EVANS, M. J., KAUFMAN, M.H. 1981. Establishment in culture of pluripotential cells from mouse embryos. *Nature,* 292**,** 154 - 156.

FALNES, P. O., JOHANSEN, R. F. & SEEBERG, E. 2002. AlkB-mediated oxidative demethylation reverses DNA damage in Escherichia coli. *Nature,* 419**,** 178-82.

FARTHING, C. R., FICZ, G., NG, R. K., CHAN, C. F., ANDREWS, S., DEAN, W., HEMBERGER, M. & REIK, W. 2008. Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. *PLoS Genet,* 4**,** e1000116.

FAZZIO, T. G., HUFF, J. T. & PANNING, B. 2008. An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. *Cell,* 134**,** 162-74.

FONG, Y. W., INOUYE, C., YAMAGUCHI, T., CATTOGLIO, C., GRUBISIC, I. & TJIAN, R. 2011. A DNA repair complex functions as an Oct4/Sox2 coactivator in embryonic stem cells. *Cell,* 147**,** 120-31.

FRANCHINI, D. M., SCHMITZ, K. M. & PETERSEN-MAHRT, S. K. 2012. 5-Methylcytosine DNA demethylation: more than losing a methyl group. *Annu Rev Genet,* 46**,** 419-41.

FRANKENBERG, S. & RENFREE, M. B. 2013. On the origin of POU5F1. *BMC Biol,* 11**,** 56.

FUJIKURA, J., YAMATO, E., YONEMURA, S., HOSODA, K., MASUI, S., NAKAO, K., MIYAZAKI JI, J. & NIWA, H. 2002. Differentiation of embryonic stem cells is induced by GATA factors. *Genes Dev,* 16**,** 784-9.

FULKA, J., JR. & FULKA, H. 2007. Somatic cell nuclear transfer (SCNT) in mammals: the cytoplast and its reprogramming activities. *Adv Exp Med Biol,* 591**,** 93-102.

GAGLIARDI, A., MULLIN, N. P., YING TAN, Z., COLBY, D., KOUSA, A. I., HALBRITTER, F., WEISS, J. T., FELKER, A., BEZSTAROSTI, K., FAVARO, R., DEMMERS, J., NICOLIS, S. K., TOMLINSON, S. R.,

POOT, R. A. & CHAMBERS, I. 2013. A direct physical interaction between Nanog and Sox2 regulates embryonic stem cell self-renewal. *EMBO J,* 32**,** 2231-47.
GAO, S., CHUNG, Y. G., PARSEGHIAN, M. H., KING, G. J., ADASHI, E. Y. & LATHAM, K. E. 2004. Rapid H1 linker histone transitions following fertilization or somatic cell nuclear transfer: evidence for a uniform developmental program in mice. *Dev Biol,* 266**,** 62-75.
GAO, T., ZHENG, J., XING, F., FANG, H., SUN, F., YAN, A., GONG, X., DING, H., TANG, F. & SHENG, H. Z. 2007. Nuclear reprogramming: the strategy used in normal development is also used in somatic cell nuclear transfer and parthenogenesis. *Cell Res,* 17**,** 135-50.
GAO, Z., COX, J. L., GILMORE, J. M., ORMSBEE, B. D., MALLANNA, S. K., WASHBURN, M. P. & RIZZINO, A. 2012. Determination of protein interactome of transcription factor Sox2 in embryonic stem cells engineered for inducible expression of four reprogramming factors. *J Biol Chem,* 287**,** 11384-97.
GAVIN, D. P., CHASE, K. A. & SHARMA, R. P. 2013. Active DNA demethylation in post-mitotic neurons: a reason for optimism. *Neuropharmacology,* 75**,** 233-45.
GEHRING, M., REIK, W. & HENIKOFF, S. 2009. DNA demethylation by DNA repair. *Trends Genet,* 25**,** 82-90.
GOLL, M. G., KIRPEKAR, F., MAGGERT, K. A., YODER, J. A., HSIEH, C. L., ZHANG, X., GOLIC, K. G., JACOBSEN, S. E. & BESTOR, T. H. 2006. Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Science,* 311**,** 395-8.
GORE, A., LI, Z., FUNG, H. L., YOUNG, J. E., AGARWAL, S., ANTOSIEWICZ-BOURGET, J., CANTO, I., GIORGETTI, A., ISRAEL, M. A., KISKINIS, E., LEE, J. H., LOH, Y. H., MANOS, P. D., MONTSERRAT, N., PANOPOULOS, A. D., RUIZ, S., WILBERT, M. L., YU, J., KIRKNESS, E. F., IZPISUA BELMONTE, J. C., ROSSI, D. J., THOMSON, J. A., EGGAN, K., DALEY, G. Q., GOLDSTEIN, L. S. & ZHANG, K. 2011. Somatic coding mutations in human induced pluripotent stem cells. *Nature,* 471**,** 63-7.
GOTTSCHALK, A. J., TIMINSZKY, G., KONG, S. E., JIN, J., CAI, Y., SWANSON, S. K., WASHBURN, M. P., FLORENS, L., LADURNER, A. G., CONAWAY, J. W. & CONAWAY, R. C. 2009. Poly(ADP-ribosyl)ation directs recruitment and activation of an ATP-dependent chromatin remodeler. *Proc Natl Acad Sci U S A,* 106**,** 13770-4.
GU, T. P., GUO, F., YANG, H., WU, H. P., XU, G. F., LIU, W., XIE, Z. G., SHI, L., HE, X., JIN, S. G., IQBAL, K., SHI, Y. G., DENG, Z., SZABO, P. E., PFEIFER, G. P., LI, J. & XU, G. L. 2011. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature,* 477**,** 606-10.
GURDON, J. B. 1962. The transplantation of nuclei between two species of Xenopus. *Dev Biol,* 5**,** 68-83.
GURDON, J. B. 1968. Changes in somatic cell nuclei inserted into growing and maturing amphibian oocytes. *J Embryol Exp Morphol,* 20**,** 401-14.
GURDON, J. B., ELSDALE, T. R. & FISCHBERG, M. 1958. Sexually mature individuals of Xenopus laevis from the transplantation of single somatic nuclei. *Nature,* 182**,** 64-5.
GURDON, J. B., PARTINGTON, G. A. & DE ROBERTIS, E. M. 1976. Injected nuclei in frog oocytes:RNA synthesis and protein exchange. *J Embryol Exp Morphol,* 36**,** 541-53.
GURDON, J. B. & WILMUT, I. 2011. Nuclear transfer to eggs and oocytes. *Cold Spring Harb Perspect Biol,* 3.

HAASE, A., OLMER, R., SCHWANKE, K., WUNDERLICH, S., MERKERT, S., HESS, C., ZWEIGERDT, R., GRUH, I., MEYER, J., WAGNER, S., MAIER, L. S., HAN, D. W., GLAGE, S., MILLER, K., FISCHER, P., SCHOLER, H. R. & MARTIN, U. 2009. Generation of induced pluripotent stem cells from human cord blood. *Cell Stem Cell,* 5**,** 434-41.

HAJKOVA, P., JEFFRIES, S. J., LEE, C., MILLER, N., JACKSON, S. P. & SURANI, M. A. 2010. Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science,* 329**,** 78-82.

HALLEY-STOTT, R. P., PASQUE, V., ASTRAND, C., MIYAMOTO, K., SIMEONI, I., JULLIEN, J. & GURDON, J. B. 2010. Mammalian nuclear transplantation to Germinal Vesicle stage Xenopus oocytes - a method for quantitative transcriptional reprogramming. *Methods,* 51**,** 56-65.

HAN, D. W., DO, J. T., GENTILE, L., STEHLING, M., LEE, H. T. & SCHOLER, H. R. 2008. Pluripotential reprogramming of the somatic genome in hybrid cells occurs with the first cell cycle. *Stem Cells,* 26**,** 445-54.

HANSIS, C., BARRETO, G., MALTRY, N. & NIEHRS, C. 2004. Nuclear reprogramming of human somatic cells by xenopus egg extract requires BRG1. *Curr Biol,* 14**,** 1475-80.

HASHIMOTO, H., HORTON, J. R., ZHANG, X. & CHENG, X. 2009. UHRF1, a modular multi-domain protein, regulates replication-coupled crosstalk between DNA methylation and histone modifications. *Epigenetics,* 4**,** 8-14.

HAYASHI, K., OHTA, H., KURIMOTO, K., ARAMAKI, S. & SAITOU, M. 2011. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell,* 146**,** 519-32.

HE, L., LIU, H. & TANG, L. 2012. SWI/SNF chromatin remodeling complex: a new cofactor in reprogramming. *Stem Cell Rev,* 8**,** 128-36.

HE, Y. F., LI, B. Z., LI, Z., LIU, P., WANG, Y., TANG, Q., DING, J., JIA, Y., CHEN, Z., LI, L., SUN, Y., LI, X., DAI, Q., SONG, C. X., ZHANG, K., HE, C. & XU, G. L. 2011. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science,* 333**,** 1303-7.

HENDRICH, B. & BIRD, A. 1998. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol,* 18**,** 6538-47.

HEO, K., KIM, H., CHOI, S. H., CHOI, J., KIM, K., GU, J., LIEBER, M. R., YANG, A. S. & AN, W. 2008. FACT-mediated exchange of histone variant H2AX regulated by phosphorylation of H2AX and ADP-ribosylation of Spt16. *Mol Cell,* 30**,** 86-97.

HERMANN, A., GOYAL, R. & JELTSCH, A. 2004. The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J Biol Chem,* 279**,** 48350-9.

HERMES, M., OSSWALD, H., RIEHLE, R., PIESCH, C. & KLOOR, D. 2008. S-Adenosylhomocysteine hydrolase overexpression in HEK-293 cells: effect on intracellular adenosine levels, cell viability, and DNA methylation. *Cell Physiol Biochem,* 22**,** 223-36.

HOGENBIRK, M. A., HEIDEMAN, M. R., VELDS, A., VAN DEN BERK, P. C., KERKHOVEN, R. M., VAN STEENSEL, B. & JACOBS, H. 2013. Differential programming of B cells in AID deficient mice. *PLoS One,* 8**,** e69815.

HOLLIDAY, R. & PUGH, J. E. 1975. DNA modification mechanisms and gene activity during development. *Science,* 187**,** 226-32.

HORTON, J. K., STEFANICK, D. F., PRASAD, R., GASSMAN, N. R., KEDAR, P. S. & WILSON, S. H. 2014. Base excision repair defects invoke hypersensitivity to PARP inhibition. *Mol Cancer Res,* 12**,** 1128-39.
HOULARD, M., BERLIVET, S., PROBST, A. V., QUIVY, J. P., HERY, P., ALMOUZNI, G. & GERARD, M. 2006. CAF-1 is essential for heterochromatin organization in pluripotent embryonic cells. *PLoS Genet,* 2**,** e181.
HOUSTON, D. W. & KING, M. L. 2000. Germ plasm and molecular determinants of germ cell fate. *Curr Top Dev Biol,* 50**,** 155-81.
HOUSTON, D. W., ZHANG, J., MAINES, J. Z., WASSERMAN, S. A. & KING, M. L. 1998. A Xenopus DAZ-like gene encodes an RNA component of germ plasm and is a functional homologue of Drosophila boule. *Development,* 125**,** 171-80.
HSIEH, C. L. 1994. Dependence of transcriptional repression on CpG methylation density. *Mol Cell Biol,* 14**,** 5487-94.
HU, G., KIM, J., XU, Q., LENG, Y., ORKIN, S. H. & ELLEDGE, S. J. 2009. A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. *Genes Dev,* 23**,** 837-48.
HUSSEIN, S. M., BATADA, N. N., VUORISTO, S., CHING, R. W., AUTIO, R., NARVA, E., NG, S., SOUROUR, M., HAMALAINEN, R., OLSSON, C., LUNDIN, K., MIKKOLA, M., TROKOVIC, R., PEITZ, M., BRUSTLE, O., BAZETT-JONES, D. P., ALITALO, K., LAHESMAA, R., NAGY, A. & OTONKOSKI, T. 2011. Copy number variation and selection during reprogramming to pluripotency. *Nature,* 471**,** 58-62.
IMAMURA, T., NEILDEZ, T. M., THENEVIN, C. & PALDI, A. 2004. Essential role for poly (ADP-ribosyl)ation in mouse preimplantation development. *BMC Mol Biol,* 5**,** 4.
INOUE, A., OGUSHI, S., SAITOU, M., SUZUKI, M. G. & AOKI, F. 2011a. Involvement of mouse nucleoplasmin 2 in the decondensation of sperm chromatin after fertilization. *Biol Reprod,* 85**,** 70-7.
INOUE, A., SHEN, L., DAI, Q., HE, C. & ZHANG, Y. 2011b. Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. *Cell Res,* 21**,** 1670-6.
INOUE, A. & ZHANG, Y. 2011. Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science,* 334**,** 194.
IQBAL, K., JIN, S. G., PFEIFER, G. P. & SZABO, P. E. 2011. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc Natl Acad Sci U S A,* 108**,** 3642-7.
ITO, S., D'ALESSIO, A. C., TARANOVA, O. V., HONG, K., SOWERS, L. C. & ZHANG, Y. 2010. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature,* 466**,** 1129-33.
ITO, S., SHEN, L., DAI, Q., WU, S. C., COLLINS, L. B., SWENBERG, J. A., HE, C. & ZHANG, Y. 2011. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science,* 333**,** 1300-3.
ITO, T., IKEHARA, T., NAKAGAWA, T., KRAUS, W. L. & MURAMATSU, M. 2000. p300-mediated acetylation facilitates the transfer of histone H2A-H2B dimers from nucleosomes to a histone chaperone. *Genes Dev,* 14**,** 1899-907.
IURLARO, M., FICZ, G., OXLEY, D., RAIBER, E. A., BACHMAN, M., BOOTH, M. J., ANDREWS, S., BALASUBRAMANIAN, S. & REIK, W. 2013. A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol,* 14**,** R119.

IVANOVA, N., DOBRIN, R., LU, R., KOTENKO, I., LEVORSE, J., DECOSTE, C., SCHAFER, X., LUN, Y. & LEMISCHKA, I. R. 2006. Dissecting self-renewal in stem cells with RNA interference. *Nature,* 442**,** 533-8.

IWAFUCHI-DOI, M., MATSUDA, K., MURAKAMI, K., NIWA, H., TESAR, P. J., ARUGA, J., MATSUO, I. & KONDOH, H. 2012. Transcriptional regulatory networks in epiblast cells and during anterior neural plate development as modeled in epiblast stem cells. *Development,* 139**,** 3926-37.

JHA, S., SHIBATA, E. & DUTTA, A. 2008. Human Rvb1/Tip49 is required for the histone acetyltransferase activity of Tip60/NuA4 and for the downregulation of phosphorylation on H2AX after DNA damage. *Mol Cell Biol,* 28**,** 2690-700.

JIANG, B. H., TSENG, W. L., LI, H. Y., WANG, M. L., CHANG, Y. L., SUNG, Y. J. & CHIOU, S. H. 2015. Poly(ADP-Ribose) Polymerase 1: Cellular Pluripotency, Reprogramming, and Tumorogenesis. *Int J Mol Sci,* 16**,** 15531-45.

JIANG, J., CHAN, Y. S., LOH, Y. H., CAI, J., TONG, G. Q., LIM, C. A., ROBSON, P., ZHONG, S. & NG, H. H. 2008. A core Klf circuitry regulates self-renewal of embryonic stem cells. *Nat Cell Biol,* 10**,** 353-60.

JIANG, L., ZHANG, J., WANG, J. J., WANG, L., ZHANG, L., LI, G., YANG, X., MA, X., SUN, X., CAI, J., HUANG, X., YU, M., WANG, X., LIU, F., WU, C. I., HE, C., ZHANG, B., CI, W. & LIU, J. 2013. Sperm, but not oocyte, DNA methylome is inherited by zebrafish early embryos. *Cell,* 153**,** 773-84.

JIN, S. G., GUO, C. & PFEIFER, G. P. 2008. GADD45A does not promote DNA demethylation. *PLoS Genet,* 4**,** e1000013.

JOHNSON, A. D., BACHVAROVA, R. F., DRUM, M. & MASI, T. 2001. Expression of axolotl DAZL RNA, a marker of germ plasm: widespread maternal RNA and onset of expression in germ cells approaching the gonad. *Dev Biol,* 234**,** 402-15.

JOHNSON, A. D., CROTHER, B., WHITE, M. E., PATIENT, R., BACHVAROVA, R. F., DRUM, M. & MASI, T. 2003. Regulative germ cell specification in axolotl embryos: a primitive trait conserved in the mammalian lineage. *Philos Trans R Soc Lond B Biol Sci,* 358**,** 1371-9.

JOST, J. P. 1993. Nuclear extracts of chicken embryos promote an active demethylation of DNA by excision repair of 5-methyldeoxycytidine. *Proc Natl Acad Sci U S A,* 90**,** 4684-8.

JOST, J. P., SIEGMANN, M., SUN, L. & LEUNG, R. 1995. Mechanisms of DNA demethylation in chicken embryos. Purification and properties of a 5-methylcytosine-DNA glycosylase. *J Biol Chem,* 270**,** 9734-9.

JULLIEN, J., ASTRAND, C., HALLEY-STOTT, R. P., GARRETT, N. & GURDON, J. B. 2010. Characterization of somatic cell nuclear reprogramming by oocytes in which a linker histone is required for pluripotency gene reactivation. *Proc Natl Acad Sci U S A,* 107**,** 5483-8.

JULLIEN, J., PASQUE, V., HALLEY-STOTT, R. P., MIYAMOTO, K. & GURDON, J. B. 2011. Mechanisms of nuclear reprogramming by eggs and oocytes: a deterministic process? *Nat Rev Mol Cell Biol,* 12**,** 453-9.

JUTRAS, B. L., VERMA, A. & STEVENSON, B. 2012. Identification of novel DNA-binding proteins using DNA-affinity chromatography/pull down. *Curr Protoc Microbiol,* Chapter 1**,** Unit1F 1.

KAGEY, M. H., NEWMAN, J. J., BILODEAU, S., ZHAN, Y., ORLANDO, D. A., VAN BERKUM, N. L., EBMEIER, C. C., GOOSSENS, J., RAHL, P. B., LEVINE, S. S., TAATJES, D. J., DEKKER, J. & YOUNG, R. A. 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature,* 467**,** 430-5.

KALKAN, T. & SMITH, A. 2014. Mapping the route from naive pluripotency to lineage specification. *Philos Trans R Soc Lond B Biol Sci,* 369.
KEE, K., ANGELES, V. T., FLORES, M., NGUYEN, H. N. & REIJO PERA, R. A. 2009. Human DAZL, DAZ and BOULE genes modulate primordial germ-cell and haploid gamete formation. *Nature,* 462**,** 222-5.
KIDDER, B. L., PALMER, S. & KNOTT, J. G. 2009. SWI/SNF-Brg1 regulates self-renewal and occupies core pluripotency-related genes in embryonic stem cells. *Stem Cells,* 27**,** 317-28.
KIM, D., KIM, C. H., MOON, J. I., CHUNG, Y. G., CHANG, M. Y., HAN, B. S., KO, S., YANG, E., CHA, K. Y., LANZA, R. & KIM, K. S. 2009. Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins. *Cell Stem Cell,* 4**,** 472-6.
KIM, K., DOI, A., WEN, B., NG, K., ZHAO, R., CAHAN, P., KIM, J., ARYEE, M. J., JI, H., EHRLICH, L. I., YABUUCHI, A., TAKEUCHI, A., CUNNIFF, K. C., HONGGUANG, H., MCKINNEY-FREEMAN, S., NAVEIRAS, O., YOON, T. J., IRIZARRY, R. A., JUNG, N., SEITA, J., HANNA, J., MURAKAMI, P., JAENISCH, R., WEISSLEDER, R., ORKIN, S. H., WEISSMAN, I. L., FEINBERG, A. P. & DALEY, G. Q. 2010. Epigenetic memory in induced pluripotent stem cells. *Nature,* 467**,** 285-90.
KIYONARI, H., KANEKO, M., ABE, S. & AIZAWA, S. 2010. Three inhibitors of FGF receptor, ERK, and GSK3 establishes germline-competent embryonic stem cells of C57BL/6N mouse strain with high efficiency and stability. *Genesis,* 48**,** 317-27.
KLOET, S. L., BAYMAZ, H. I., MAKOWSKI, M., GROENEWOLD, V., JANSEN, P. W., BERENDSEN, M., NIAZI, H., KOPS, G. J. & VERMEULEN, M. 2015. Towards elucidating the stability, dynamics and architecture of the nucleosome remodeling and deacetylase complex by using quantitative interaction proteomics. *FEBS J,* 282**,** 1774-85.
KLOSE, R. J., KALLIN, E. M. & ZHANG, Y. 2006. JmjC-domain-containing proteins and histone demethylation. *Nat Rev Genet,* 7**,** 715-27.
KLOSE, R. J. & ZHANG, Y. 2007. Regulation of histone methylation by demethylimination and demethylation. *Nat Rev Mol Cell Biol,* 8**,** 307-18.
KONG, P. C., ZHU, Y., WANG, M. S., LI, H. P., CHEN, X. J. & JIANG, M. X. 2013. Reprogramming of round spermatids by the germinal vesicle cytoplasm in mice. *PLoS One,* 8**,** e78437.
KONG, Q., XIE, B., LI, J., HUAN, Y., HUANG, T., WEI, R., LV, J., LIU, S. & LIU, Z. 2014. Identification and characterization of an oocyte factor required for porcine nuclear reprogramming. *J Biol Chem,* 289**,** 6960-8.
KOOPMAN, P. & COTTON, R. G. 1984. A factor produced by feeder cells which inhibits embryonal carcinoma cell differentiation. Characterization and partial purification. *Exp Cell Res,* 154**,** 233-42.
KOVACS, K. A., STEINMANN, M., MAGISTRETTI, P. J., HALFON, O. & CARDINAUX, J. R. 2003. CCAAT/enhancer-binding protein family members recruit the coactivator CREB-binding protein and trigger its phosphorylation. *J Biol Chem,* 278**,** 36959-65.
KRIAUCIONIS, S. & HEINTZ, N. 2009. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science,* 324**,** 929-30.
KRISHNAKUMAR, R. & KRAUS, W. L. 2010. The PARP side of the nucleus: molecular actions, physiological outcomes, and clinical targets. *Mol Cell,* 39**,** 8-24.
KUNATH, T., SABA-EL-LEIL, M. K., ALMOUSAILLEAKH, M., WRAY, J., MELOCHE, S. & SMITH, A. 2007. FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem

cells from self-renewal to lineage commitment. *Development,* 134**,** 2895-902.
KURODA, T., TADA, M., KUBOTA, H., KIMURA, H., HATANO, S. Y., SUEMORI, H., NAKATSUJI, N. & TADA, T. 2005. Octamer and Sox elements are required for transcriptional cis regulation of Nanog gene expression. *Mol Cell Biol,* 25**,** 2475-85.
LA SALLE, S., MERTINEIT, C., TAKETO, T., MOENS, P. B., BESTOR, T. H. & TRASLER, J. M. 2004. Windows for sex-specific methylation marked by DNA methyltransferase expression profiles in mouse germ cells. *Dev Biol,* 268**,** 403-15.
LAI, M., LIANG, L., CHEN, J., QIU, N., GE, S., JI, S., SHI, T., ZHEN, B., LIU, M., DING, C., WANG, Y. & QIN, J. 2016. Multidimensional Proteomics Reveals a Role of UHRF2 in the Regulation of Epithelial-Mesenchymal Transition (EMT). *Mol Cell Proteomics,* 15**,** 2263-78.
LAW, J. A. & JACOBSEN, S. E. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet,* 11**,** 204-20.
LAWSON, K. A., DUNN, N. R., ROELEN, B. A., ZEINSTRA, L. M., DAVIS, A. M., WRIGHT, C. V., KORVING, J. P. & HOGAN, B. L. 1999. Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes Dev,* 13**,** 424-36.
LAWSON, K. A. & HAGE, W. J. 1994. Clonal analysis of the origin of primordial germ cells in the mouse. *Ciba Found Symp,* 182**,** 68-84; discussion 84-91.
LE MAY, N., MOTA-FERNANDES, D., VELEZ-CRUZ, R., ILTIS, I., BIARD, D. & EGLY, J. M. 2010. NER factors are recruited to active promoters and facilitate chromatin modification for transcription in the absence of exogenous genotoxic attack. *Mol Cell,* 38**,** 54-66.
LEBEDEVA, N. A., RECHKUNOVA, N. I. & LAVRIK, O. I. 2011. AP-site cleavage activity of tyrosyl-DNA phosphodiesterase 1. *FEBS Lett,* 585**,** 683-6.
LEIDGENS, S., BULLOUGH, K. Z., SHI, H., LI, F., SHAKOURY-ELIZEH, M., YABE, T., SUBRAMANIAN, P., HSU, E., NATARAJAN, N., NANDAL, A., STEMMLER, T. L. & PHILPOTT, C. C. 2013. Each member of the poly-r(C)-binding protein 1 (PCBP) family exhibits iron chaperone activity toward ferritin. *J Biol Chem,* 288**,** 17791-802.
LEUNG, L., KWONG, M., HOU, S., LEE, C. & CHAN, J. Y. 2003. Deficiency of the Nrf1 and Nrf2 transcription factors results in early embryonic lethality and severe oxidative stress. *J Biol Chem,* 278**,** 48021-9.
LI, E., BESTOR, T. H. & JAENISCH, R. 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell,* 69**,** 915-26.
LI, P., TONG, C., MEHRIAN-SHAI, R., JIA, L., WU, N., YAN, Y., MAXSON, R. E., SCHULZE, E. N., SONG, H., HSIEH, C. L., PERA, M. F. & YING, Q. L. 2008. Germline competent embryonic stem cells derived from rat blastocysts. *Cell,* 135**,** 1299-310.
LI, Y., ZHANG, Q., YIN, X., YANG, W., DU, Y., HOU, P., GE, J., LIU, C., ZHANG, W., ZHANG, X., WU, Y., LI, H., LIU, K., WU, C., SONG, Z., ZHAO, Y., SHI, Y. & DENG, H. 2011. Generation of iPSCs from mouse fibroblasts with a single gene, Oct4, and small molecules. *Cell Res,* 21**,** 196-204.
LIANG, C., ZHANG, X., SONG, S., TIAN, C., YIN, Y., XING, G., HE, F. & ZHANG, L. 2013. Identification of UHRF1/2 as new N-methylpurine DNA glycosylase-interacting proteins. *Biochem Biophys Res Commun,* 433**,** 415-9.
LIANG, J., WAN, M., ZHANG, Y., GU, P., XIN, H., JUNG, S. Y., QIN, J., WONG, J., COONEY, A. J., LIU, D. & SONGYANG, Z. 2008. Nanog

and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. *Nat Cell Biol,* 10**,** 731-9.
LIN, C. J., KOH, F. M., WONG, P., CONTI, M. & RAMALHO-SANTOS, M. 2014. Hira-mediated H3.3 incorporation is required for DNA replication and ribosomal RNA transcription in the mouse zygote. *Dev Cell,* 30**,** 268-79.
LIN, H., HE, L. & MA, B. 2013. A combinatorial approach to the peptide feature matching problem for label-free quantification. *Bioinformatics,* 29**,** 1768-75.
LIN, Y. L., SHIVJI, M. K., CHEN, C., KOLODNER, R., WOOD, R. D. & DUTTA, A. 1998. The evolutionarily conserved zinc finger motif in the largest subunit of human replication protein A is required for DNA replication and mismatch repair but not for nucleotide excision repair. *J Biol Chem,* 273**,** 1453-61.
LISTER, R., PELIZZOLA, M., KIDA, Y. S., HAWKINS, R. D., NERY, J. R., HON, G., ANTOSIEWICZ-BOURGET, J., O'MALLEY, R., CASTANON, R., KLUGMAN, S., DOWNES, M., YU, R., STEWART, R., REN, B., THOMSON, J. A., EVANS, R. M. & ECKER, J. R. 2011. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature,* 471**,** 68-73.
LIU, X., GAO, Q., LI, P., ZHAO, Q., ZHANG, J., LI, J., KOSEKI, H. & WONG, J. 2013. UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. *Nat Commun,* 4**,** 1563.
LIU, Y., ZHANG, B., KUANG, H., KORAKAVI, G., LU, L. Y. & YU, X. 2016. Zinc finger protein 618 regulates UHRF2 (ubiquitin like with PHD and ring finger domains 2) function as a specific 5-hydroxymethylcytosine reader. *J Biol Chem*.
LIVIGNI, A., PERADZIRYI, H., SHAROV, A. A., CHIA, G., HAMMACHI, F., MIGUELES, R. P., SUKPARANGSI, W., PERNAGALLO, S., BRADLEY, M., NICHOLS, J., KO, M. S. & BRICKMAN, J. M. 2013. A conserved Oct4/POUV-dependent network links adhesion and migration to progenitor maintenance. *Curr Biol,* 23**,** 2233-44.
LOH, Y. H., WU, Q., CHEW, J. L., VEGA, V. B., ZHANG, W., CHEN, X., BOURQUE, G., GEORGE, J., LEONG, B., LIU, J., WONG, K. Y., SUNG, K. W., LEE, C. W., ZHAO, X. D., CHIU, K. P., LIPOVICH, L., KUZNETSOV, V. A., ROBSON, P., STANTON, L. W., WEI, C. L., RUAN, Y., LIM, B. & NG, H. H. 2006. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet,* 38**,** 431-40.
LYABIN, D. N., ELISEEVA, I. A. & OVCHINNIKOV, L. P. 2014. YB-1 protein: functions and regulation. *Wiley Interdiscip Rev RNA,* 5**,** 95-110.
MA, H., MOREY, R., O'NEIL, R. C., HE, Y., DAUGHTRY, B., SCHULTZ, M. D., HARIHARAN, M., NERY, J. R., CASTANON, R., SABATINI, K., THIAGARAJAN, R. D., TACHIBANA, M., KANG, E., TIPPNER-HEDGES, R., AHMED, R., GUTIERREZ, N. M., VAN DYKEN, C., POLAT, A., SUGAWARA, A., SPARMAN, M., GOKHALE, S., AMATO, P., WOLF, D. P., ECKER, J. R., LAURENT, L. C. & MITALIPOV, S. 2014. Abnormalities in human pluripotent cells due to reprogramming mechanisms. *Nature,* 511**,** 177-83.
MA, M., GUO, X., WANG, F., ZHAO, C., LIU, Z., SHI, Z., WANG, Y., ZHANG, P., ZHANG, K., WANG, N., LIN, M., ZHOU, Z., LIU, J., LI, Q., WANG, L., HUO, R., SHA, J. & ZHOU, Q. 2008. Protein expression profile of the mouse metaphase-II oocyte. *J Proteome Res,* 7**,** 4821-30.
MACHON, O., MASEK, J., MACHONOVA, O., KRAUSS, S. & KOZMIK, Z. 2015. Meis2 is essential for cranial and cardiac neural crest development. *BMC Dev Biol,* 15**,** 40.

MAITI, A. & DROHAT, A. C. 2011. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J Biol Chem,* 286**,** 35334-8.
MAITI, A., HWANG, B. J., ARMWOOD, C. J., LU-CHANG, A. L. & DROHAT, A. C. 2013. TDG excision of fC may be a predominant element of pathways for active DNA demethylation. *Faseb Journal,* 27.
MALLANNA, S. K., ORMSBEE, B. D., IACOVINO, M., GILMORE, J. M., COX, J. L., KYBA, M., WASHBURN, M. P. & RIZZINO, A. 2010. Proteomic analysis of Sox2-associated proteins during early stages of mouse embryonic stem cell differentiation identifies Sox21 as a novel regulator of stem cell fate. *Stem Cells,* 28**,** 1715-27.
MARTIN, G. R. 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A,* 78**,** 7634-7638.
MARTIN, G. R. & EVANS, M. J. 1975. Differentiation of clonal lines of teratocarcinoma cells: formation of embryoid bodies in vitro. *Proc Natl Acad Sci U S A,* 72**,** 1441-5.
MARTINEZ-MACIAS, M. I., QIAN, W., MIKI, D., PONTES, O., LIU, Y., TANG, K., LIU, R., MORALES-RUIZ, T., ARIZA, R. R., ROLDAN-ARJONA, T. & ZHU, J. K. 2012. A DNA 3' phosphatase functions in active DNA demethylation in Arabidopsis. *Mol Cell,* 45**,** 357-70.
MARTINOWICH, K., HATTORI, D., WU, H., FOUSE, S., HE, F., HU, Y., FAN, G. & SUN, Y. E. 2003. DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. *Science,* 302**,** 890-3.
MASUI, S., NAKATAKE, Y., TOYOOKA, Y., SHIMOSATO, D., YAGI, R., TAKAHASHI, K., OKOCHI, H., OKUDA, A., MATOBA, R., SHAROV, A. A., KO, M. S. & NIWA, H. 2007. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol,* 9**,** 625-35.
MATHEMA, V. B. & KOH, Y. S. 2012. Inhibitor of growth-4 mediates chromatin modification and has a suppressive effect on tumorigenesis and innate immunity. *Tumour Biol,* 33**,** 1-7.
MATSUDA, T., NAKAMURA, T., NAKAO, K., ARAI, T., KATSUKI, M., HEIKE, T. & YOKOTA, T. 1999. STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. *EMBO J,* 18**,** 4261-9.
MATSUMURA, H. & TADA, T. 2008. Cell fusion-mediated nuclear reprogramming of somatic cells. *Reprod Biomed Online,* 16**,** 51-6.
MAYER, W., NIVELEAU, A., WALTER, J., FUNDELE, R. & HAAF, T. 2000. Demethylation of the zygotic paternal genome. *Nature,* 403**,** 501-2.
MCLAY, D. W. & CLARKE, H. J. 2003. Remodelling the paternal chromatin at fertilization in mammals. *Reproduction,* 125**,** 625-33.
MENG, Z., MOROISHI, T. & GUAN, K. L. 2016. Mechanisms of Hippo pathway regulation. *Genes Dev,* 30**,** 1-17.
MESSERSCHMIDT, D. M. & KEMLER, R. 2010. Nanog is required for primitive endoderm formation through a non-cell autonomous mechanism. *Dev Biol,* 344**,** 129-37.
MICHAEL, P. 1984. Are the primordial germ cells (PGCs) in urodela formed by the inductive action of the vegetative yolk mass? *Dev Biol,* 103**,** 109-16.
MILLARD, C. J., VARMA, N., SALEH, A., MORRIS, K., WATSON, P. J., BOTTRILL, A. R., FAIRALL, L., SMITH, C. J. & SCHWABE, J. W. 2016. The structure of the core NuRD repression complex provides insights into its interaction with chromatin. *Elife,* 5**,** e13941.

MILLER, M. W., DUHL, D. M., WINKES, B. M., ARREDONDO-VEGA, F., SAXON, P. J., WOLFF, G. L., EPSTEIN, C. J., HERSHFIELD, M. S. & BARSH, G. S. 1994. The mouse lethal nonagouti (a(x)) mutation deletes the S-adenosylhomocysteine hydrolase (Ahcy) gene. *EMBO J,* 13**,** 1806-16.

MITSUI, K., TOKUZAWA, Y., ITOH, H., SEGAWA, K., MURAKAMI, M., TAKAHASHI, K., MARUYAMA, M., MAEDA, M. & YAMANAKA, S. 2003. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell,* 113**,** 631-42.

MIYAMOTO, K., PASQUE, V., JULLIEN, J. & GURDON, J. B. 2011. Nuclear actin polymerization is required for transcriptional reprogramming of Oct4 by oocytes. *Genes Dev,* 25**,** 946-58.

MIZUGUCHI, G., SHEN, X., LANDRY, J., WU, W. H., SEN, S. & WU, C. 2004. ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science,* 303**,** 343-8.

MORGAN, H. D., DEAN, W., COKER, H. A., REIK, W. & PETERSEN-MAHRT, S. K. 2004. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *J Biol Chem,* 279**,** 52353-60.

MUNZEL, M., GLOBISCH, D., BRUCKL, T., WAGNER, M., WELZMILLER, V., MICHALAKIS, S., MULLER, M., BIEL, M. & CARELL, T. 2010. Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew Chem Int Ed Engl,* 49**,** 5375-7.

MURATA, K., KOUZARIDES, T., BANNISTER, A. J. & GURDON, J. B. 2010. Histone H3 lysine 4 methylation is associated with the transcriptional reprogramming efficiency of somatic nuclei by oocytes. *Epigenetics Chromatin,* 3**,** 4.

NAKAGAWA, M., KOYANAGI, M., TANABE, K., TAKAHASHI, K., ICHISAKA, T., AOI, T., OKITA, K., MOCHIDUKI, Y., TAKIZAWA, N. & YAMANAKA, S. 2008. Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat Biotechnol,* 26**,** 101-6.

NAKATAKE, Y., FUKUI, N., IWAMATSU, Y., MASUI, S., TAKAHASHI, K., YAGI, R., YAGI, K., MIYAZAKI, J., MATOBA, R., KO, M. S. & NIWA, H. 2006. Klf4 cooperates with Oct3/4 and Sox2 to activate the Lefty1 core promoter in embryonic stem cells. *Mol Cell Biol,* 26**,** 7772-82.

NG, R. K., DEAN, W., DAWSON, C., LUCIFERO, D., MADEJA, Z., REIK, W. & HEMBERGER, M. 2008. Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5. *Nat Cell Biol,* 10**,** 1280-90.

NGO, T. T., YOO, J., DAI, Q., ZHANG, Q., HE, C., AKSIMENTIEV, A. & HA, T. 2016. Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nat Commun,* 7**,** 10813.

NICHOLS, J. & SMITH, A. 2009. Naive and primed pluripotent states. *Cell Stem Cell,* 4**,** 487-92.

NICHOLS, J. & SMITH, A. 2012. Pluripotency in the embryo and in culture. *Cold Spring Harb Perspect Biol,* 4**,** a008128.

NICHOLS, J., ZEVNIK, B., ANASTASSIADIS, K., NIWA, H., KLEWE-NEBENIUS, D., CHAMBERS, I., SCHOLER, H. & SMITH, A. 1998. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell,* 95**,** 379-91.

NIEHRS, C. & SCHAFER, A. 2012. Active DNA demethylation by Gadd45 and DNA repair. *Trends Cell Biol,* 22**,** 220-7.

NIEUWKOOP, P. D. 1969. The formation of the mesoderm in urodelean amphibians. *Roux's Arch. Entw. Org.***,** 341-373.

NISHIOKA, N., INOUE, K., ADACHI, K., KIYONARI, H., OTA, M., RALSTON, A., YABUTA, N., HIRAHARA, S., STEPHENSON, R. O., OGONUKI, N., MAKITA, R., KURIHARA, H., MORIN-KENSICKI, E. M., NOJIMA, H., ROSSANT, J., NAKAO, K., NIWA, H. & SASAKI, H. 2009. The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass. *Dev Cell,* 16**,** 398-410.

NIWA, H., BURDON, T., CHAMBERS, I. & SMITH, A. 1998. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev,* 12**,** 2048-60.

NIWA, H., MIYAZAKI, J. & SMITH, A. G. 2000. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat Genet,* 24**,** 372-6.

NIWA, H., TOYOOKA, Y., SHIMOSATO, D., STRUMPF, D., TAKAHASHI, K., YAGI, R. & ROSSANT, J. 2005. Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell,* 123**,** 917-29.

NOWAK, A. J., ALFIERI, C., STIRNIMANN, C. U., RYBIN, V., BAUDIN, F., LY-HARTIG, N., LINDNER, D. & MULLER, C. W. 2011. Chromatin-modifying complex component Nurf55/p55 associates with histones H3 and H4 and polycomb repressive complex 2 subunit Su(z)12 through partially overlapping binding sites. *J Biol Chem,* 286**,** 23388-96.

OKANO, M., BELL, D. W., HABER, D. A. & LI, E. 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell,* 99**,** 247-57.

OKANO, M., XIE, S. & LI, E. 1998. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat Genet,* 19**,** 219-20.

OKUMURA-NAKANISHI, S., SAITO, M., NIWA, H. & ISHIKAWA, F. 2005. Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells. *J Biol Chem,* 280**,** 5307-17.

OSWALD, J., ENGEMANN, S., LANE, N., MAYER, W., OLEK, A., FUNDELE, R., DEAN, W., REIK, W. & WALTER, J. 2000. Active demethylation of the paternal genome in the mouse zygote. *Curr Biol,* 10**,** 475-8.

OZTURK, N., SINGH, I., MEHTA, A., BRAUN, T. & BARRETO, G. 2014. HMGA proteins as modulators of chromatin structure during transcriptional activation. *Front Cell Dev Biol,* 2**,** 5.

PAPADOPOULOU, T., KAYMAK, A., SAYOLS, S. & RICHLY, H. 2016. Dual role of Med12 in PRC1-dependent gene repression and ncRNA-mediated transcriptional activation. *Cell Cycle***,** 1-15.

PAPP, B. & PLATH, K. 2011. Reprogramming to pluripotency: stepwise resetting of the epigenetic landscape. *Cell Res,* 21**,** 486-501.

PARDO, M., LANG, B., YU, L., PROSSER, H., BRADLEY, A., BABU, M. M. & CHOUDHARY, J. 2010. An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. *Cell Stem Cell,* 6**,** 382-95.

PARK, I. H., ZHAO, R., WEST, J. A., YABUUCHI, A., HUO, H., INCE, T. A., LEROU, P. H., LENSCH, M. W. & DALEY, G. Q. 2008. Reprogramming of human somatic cells to pluripotency with defined factors. *Nature,* 451**,** 141-6.

PARTHUN, M. R., WIDOM, J. & GOTTSCHLING, D. E. 1996. The major cytoplasmic histone acetyltransferase in yeast: links to chromatin replication and histone metabolism. *Cell,* 87**,** 85-94.

PAUN, A. & PITHA, P. M. 2007. The IRF family, revisited. *Biochimie,* 89**,** 744-53.

PELTON, T. A., SHARMA, S., SCHULZ, T. C., RATHJEN, J. & RATHJEN, P. D. 2002. Transient pluripotent cell populations during primitive

ectoderm formation: correlation of in vivo and in vitro pluripotent cell development. *J Cell Sci,* 115**,** 329-39.
PEREIRA, L., YI, F. & MERRILL, B. J. 2006. Repression of Nanog gene transcription by Tcf3 limits embryonic stem cell self-renewal. *Mol Cell Biol,* 26**,** 7479-91.
PFAFFENEDER, T., HACKNER, B., TRUSS, M., MUNZEL, M., MULLER, M., DEIML, C. A., HAGEMEIER, C. & CARELL, T. 2011. The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew Chem Int Ed Engl,* 50**,** 7008-12.
PFEIFFER, M. J., SIATKOWSKI, M., PAUDEL, Y., BALBACH, S. T., BAEUMER, N., CROSETTO, N., DREXLER, H. C., FUELLEN, G. & BOIANI, M. 2011. Proteomic analysis of mouse oocytes reveals 28 candidate factors of the "reprogrammome". *J Proteome Res,* 10**,** 2140-53.
PHILPOTT, A. & LENO, G. H. 1992. Nucleoplasmin remodels sperm chromatin in Xenopus egg extracts. *Cell,* 69**,** 759-67.
PICHLER, G., WOLF, P., SCHMIDT, C. S., MEILINGER, D., SCHNEIDER, K., FRAUER, C., FELLINGER, K., ROTTACH, A. & LEONHARDT, H. 2011. Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways. *J Cell Biochem,* 112**,** 2585-93.
PLUSA, B., PILISZEK, A., FRANKENBERG, S., ARTUS, J. & HADJANTONAKIS, A. K. 2008. Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst. *Development,* 135**,** 3081-91.
POLO, S. E., KAIDI, A., BASKCOMB, L., GALANTY, Y. & JACKSON, S. P. 2010. Regulation of DNA-damage responses and cell-cycle progression by the chromatin remodelling factor CHD4. *EMBO J,* 29**,** 3130-9.
RAMOS, I., MARTIN-BENITO, J., FINN, R., BRETANA, L., ALORIA, K., ARIZMENDI, J. M., AUSIO, J., MUGA, A., VALPUESTA, J. M. & PRADO, A. 2010. Nucleoplasmin binds histone H2A-H2B dimers through its distal face. *J Biol Chem,* 285**,** 33771-8.
RATHBONE, A. J., FISHER, P. A., LEE, J. H., CRAIGON, J. & CAMPBELL, K. H. 2010. Reprogramming of ovine somatic cells with Xenopus laevis oocyte extract prior to SCNT improves live birth rate. *Cell Reprogram,* 12**,** 609-16.
REALE, A., MATTEIS, G. D., GALLEAZZI, G., ZAMPIERI, M. & CAIAFA, P. 2005. Modulation of DNMT1 activity by ADP-ribose polymers. *Oncogene,* 24**,** 13-9.
REIK, W. & SURANI, M. A. 2015. Germline and Pluripotent Stem Cells. *Cold Spring Harb Perspect Biol,* 7.
REYNOLDS, P., COOPER, S., LOMAX, M. & O'NEILL, P. 2015. Disruption of PARP1 function inhibits base excision repair of a sub-set of DNA lesions. *Nucleic Acids Res,* 43**,** 4028-38.
RIGGS, A. D. 1975. X inactivation, differentiation, and DNA methylation. *Cytogenet Cell Genet,* 14**,** 9-25.
RODDA, D. J., CHEW, J. L., LIM, L. H., LOH, Y. H., WANG, B., NG, H. H. & ROBSON, P. 2005. Transcriptional regulation of nanog by OCT4 and SOX2. *J Biol Chem,* 280**,** 24731-7.
ROHMAN, M. & HARRISON-LAVOIE, K. J. 2000. Separation of copurifying GroEL from glutathione-S-transferase fusion proteins. *Protein Expr Purif,* 20**,** 45-7.
ROSNER, M. H., VIGANO, M. A., OZATO, K., TIMMONS, P. M., POIRIER, F., RIGBY, P. W. & STAUDT, L. M. 1990. A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. *Nature,* 345**,** 686-92.
ROSSANT, J. 2008. Stem cells and early lineage development. *Cell,* 132**,** 527-31.

SAKAI, Y., SUETAKE, I., SHINOZAKI, F., YAMASHINA, S. & TAJIMA, S. 2004. Co-expression of de novo DNA methyltransferases Dnmt3a2 and Dnmt3L in gonocytes of mouse embryos. *Gene Expr Patterns,* 5**,** 231-7.

SANCAR, A., LINDSEY-BOLTZ, L. A., UNSAL-KACMAZ, K. & LINN, S. 2004. Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu Rev Biochem,* 73**,** 39-85.

SANIJ, E., DIESCH, J., LESMANA, A., POORTINGA, G., HEIN, N., LIDGERWOOD, G., CAMERON, D. P., ELLUL, J., GOODALL, G. J., WONG, L. H., DHILLON, A. S., HAMDANE, N., ROTHBLUM, L. I., PEARSON, R. B., HAVIV, I., MOSS, T. & HANNAN, R. D. 2015. A novel role for the Pol I transcription factor UBTF in maintaining genome stability through the regulation of highly transcribed Pol II genes. *Genome Res,* 25**,** 201-12.

SANTOS, F., PEAT, J., BURGESS, H., RADA, C., REIK, W. & DEAN, W. 2013. Active demethylation in mouse zygotes involves cytosine deamination and base excision repair. *Epigenetics Chromatin,* 6**,** 39.

SCHAFER, A., SCHOMACHER, L., BARRETO, G., DODERLEIN, G. & NIEHRS, C. 2010. Gemcitabine functions epigenetically by inhibiting repair mediated DNA demethylation. *PLoS One,* 5**,** e14060.

SCHANIEL, C., ANG, Y. S., RATNAKUMAR, K., CORMIER, C., JAMES, T., BERNSTEIN, E., LEMISCHKA, I. R. & PADDISON, P. J. 2009. Smarcc1/Baf155 couples self-renewal gene repression with changes in chromatin structure in mouse embryonic stem cells. *Stem Cells,* 27**,** 2979-91.

SCHOLER, H. R., RUPPERT, S., SUZUKI, N., CHOWDHURY, K. & GRUSS, P. 1990. New type of POU domain in germ line-specific protein Oct-4. *Nature,* 344**,** 435-9.

SCHUBELER, D. 2015. Function and information content of DNA methylation. *Nature,* 517**,** 321-6.

SEISENBERGER, S., ANDREWS, S., KRUEGER, F., ARAND, J., WALTER, J., SANTOS, F., POPP, C., THIENPONT, B., DEAN, W. & REIK, W. 2012. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol Cell,* 48**,** 849-62.

SELBY, C. P. & SANCAR, A. 2003. Characterization of transcription-repair coupling factors in E. coli and humans. *Methods Enzymol,* 371**,** 300-24.

SHEN, L., KONDO, Y., GUO, Y., ZHANG, J., ZHANG, L., AHMED, S., SHU, J., CHEN, X., WATERLAND, R. A. & ISSA, J. P. 2007. Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. *PLoS Genet,* 3**,** 2023-36.

SHEN, L., WU, H., DIEP, D., YAMAGUCHI, S., D'ALESSIO, A. C., FUNG, H. L., ZHANG, K. & ZHANG, Y. 2013. Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell,* 153**,** 692-706.

SHIMIZU, Y., UCHIMURA, Y., DOHMAE, N., SAITOH, H., HANAOKA, F. & SUGASAWA, K. 2010. Stimulation of DNA Glycosylase Activities by XPC Protein Complex: Roles of Protein-Protein Interactions. *J Nucleic Acids,* 2010.

SHUCK, S. C., SHORT, E. A. & TURCHI, J. J. 2008. Eukaryotic nucleotide excision repair: from understanding mechanisms to influencing biology. *Cell Res,* 18**,** 64-72.

SILVA, J., NICHOLS, J., THEUNISSEN, T. W., GUO, G., VAN OOSTEN, A. L., BARRANDON, O., WRAY, J., YAMANAKA, S., CHAMBERS, I. & SMITH, A. 2009. Nanog is the gateway to the pluripotent ground state. *Cell,* 138**,** 722-37.

SIMONSSON, S. & GURDON, J. 2004. DNA demethylation is necessary for the epigenetic reprogramming of somatic cell nuclei. *Nat Cell Biol,* 6**,** 984-90.
SINNAMON, J. R., WADDELL, C. B., NIK, S., CHEN, E. I. & CZAPLINSKI, K. 2012. Hnrpab regulates neural development and neuron cell survival after glutamate stimulation. *RNA,* 18**,** 704-19.
SLEETH, K. M., SORENSEN, C. S., ISSAEVA, N., DZIEGIELEWSKI, J., BARTEK, J. & HELLEDAY, T. 2007. RPA mediates recombination repair during replication stress and is displaced from DNA by checkpoint signalling in human cells. *J Mol Biol,* 373**,** 38-47.
SMILEY, J. A., KUNDRACIK, M., LANDFRIED, D. A., BARNES, V. R., SR. & AXHEMI, A. A. 2005. Genes of the thymidine salvage pathway: thymine-7-hydroxylase from a Rhodotorula glutinis cDNA library and iso-orotate decarboxylase from Neurospora crassa. *Biochim Biophys Acta,* 1723**,** 256-64.
SMITH, A. G., HEATH, J. K., DONALDSON, D. D., WONG, G. G., MOREAU, J., STAHL, M. & ROGERS, D. 1988. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature,* 336**,** 688-90.
SMITH, A. G. & HOOPER, M. L. 1987. Buffalo rat liver cells produce a diffusible activity which inhibits the differentiation of murine embryonal carcinoma and embryonic stem cells. *Dev Biol,* 121**,** 1-9.
SMITH, T. A. & HOOPER, M. L. 1983. Medium conditioned by feeder cells inhibits the differentiation of embryonal carcinoma cultures. *Exp Cell Res,* 145**,** 458-62.
SMITH, Z. D., SINDHU, C. & MEISSNER, A. 2016. Molecular features of cellular reprogramming and development. *Nat Rev Mol Cell Biol,* 17**,** 139-54.
SNYDER, E., SOUNDARARAJAN, R., SHARMA, M., DEARTH, A., SMITH, B. & BRAUN, R. E. 2015. Compound Heterozygosity for Y Box Proteins Causes Sterility Due to Loss of Translational Repression. *PLoS Genet,* 11**,** e1005690.
SONG, C. X., SZULWACH, K. E., DAI, Q., FU, Y., MAO, S. Q., LIN, L., STREET, C., LI, Y., POIDEVIN, M., WU, H., GAO, J., LIU, P., LI, L., XU, G. L., JIN, P. & HE, C. 2013. Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell,* 153**,** 678-91.
SPRUIJT, C. G., GNERLICH, F., SMITS, A. H., PFAFFENEDER, T., JANSEN, P. W., BAUER, C., MUNZEL, M., WAGNER, M., MULLER, M., KHAN, F., EBERL, H. C., MENSINGA, A., BRINKMAN, A. B., LEPHIKOV, K., MULLER, U., WALTER, J., BOELENS, R., VAN INGEN, H., LEONHARDT, H., CARELL, T. & VERMEULEN, M. 2013. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell,* 152**,** 1146-59.
STADLER, M. B., MURR, R., BURGER, L., IVANEK, R., LIENERT, F., SCHOLER, A., VAN NIMWEGEN, E., WIRBELAUER, C., OAKELEY, E. J., GAIDATZIS, D., TIWARI, V. K. & SCHUBELER, D. 2011. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature,* 480**,** 490-5.
STAVRIDIS, M. P., LUNN, J. S., COLLINS, B. J. & STOREY, K. G. 2007. A discrete period of FGF-induced Erk1/2 signalling is required for vertebrate neural specification. *Development,* 134**,** 2889-94.
STEIN, R., RAZIN, A. & CEDAR, H. 1982. In vitro methylation of the hamster adenine phosphoribosyltransferase gene inhibits its expression in mouse L cells. *Proc Natl Acad Sci U S A,* 79**,** 3418-22.
STEUBLE, M., GERRITS, B., LUDWIG, A., MATEOS, J. M., DIEP, T. M., TAGAYA, M., STEPHAN, A., SCHATZLE, P., KUNZ, B., STREIT, P. &

SONDEREGGER, P. 2010. Molecular characterization of a trafficking organelle: dissecting the axonal paths of calsyntenin-1 transport vesicles. *Proteomics,* 10**,** 3775-88.
SUDA, Y., SUZUKI, M., IKAWA, Y. & AIZAWA, S. 1987. Mouse embryonic stem cells exhibit indefinite proliferative potential. *J Cell Physiol,* 133**,** 197-201.
SUETAKE, I., SHINOZAKI, F., MIYAGAWA, J., TAKESHIMA, H. & TAJIMA, S. 2004. DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. *J Biol Chem,* 279**,** 27816-23.
SUGASAWA, K. 2008. XPC: its product and biological roles. *Adv Exp Med Biol,* 637**,** 47-56.
SUTANI, T. & YANAGIDA, M. 1997. DNA renaturation activity of the SMC complex implicated in chromosome condensation. *Nature,* 388**,** 798-801.
SWIERS, G., CHEN, Y. H., JOHNSON, A. D. & LOOSE, M. 2010. A conserved mechanism for vertebrate mesoderm specification in urodele amphibians and mammals. *Dev Biol,* 343**,** 138-52.
TADA, M., TADA, T., LEFEBVRE, L., BARTON, S. C. & SURANI, M. A. 1997. Embryonic germ cells induce epigenetic reprogramming of somatic nucleus in hybrid cells. *EMBO J,* 16**,** 6510-20.
TAHILIANI, M., KOH, K. P., SHEN, Y., PASTOR, W. A., BANDUKWALA, H., BRUDNO, Y., AGARWAL, S., IYER, L. M., LIU, D. R., ARAVIND, L. & RAO, A. 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science,* 324**,** 930-5.
TAJRISHI, M. M., TUTEJA, R. & TUTEJA, N. 2011. Nucleolin: The most abundant multifunctional phosphoprotein of nucleolus. *Commun Integr Biol,* 4**,** 267-75.
TAKAHASHI, K., TANABE, K., OHNUKI, M., NARITA, M., ICHISAKA, T., TOMODA, K. & YAMANAKA, S. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell,* 131**,** 861-72.
TAKAHASHI, K. & YAMANAKA, S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell,* 126**,** 663-76.
TAM, P. P. & ZHOU, S. X. 1996. The allocation of epiblast cells to ectodermal and germ-line lineages is influenced by the position of the cells in the gastrulating mouse embryo. *Dev Biol,* 178**,** 124-32.
TAMANAHA, E., GUAN, S., MARKS, K. & SALEH, L. 2016. Distributive Processing by the Iron(II)/alpha-Ketoglutarate-Dependent Catalytic Domains of the TET Enzymes Is Consistent with Epigenetic Roles for Oxidized 5-Methylcytosine Bases. *J Am Chem Soc*.
TAO, H., LIU, W., SIMMONS, B. N., HARRIS, H. K., COX, T. C. & MASSIAH, M. A. 2010. Purifying natively folded proteins from inclusion bodies using sarkosyl, Triton X-100, and CHAPS. *Biotechniques,* 48**,** 61-4.
TAO, X., WEST, A. E., CHEN, W. G., CORFAS, G. & GREENBERG, M. E. 2002. A calcium-responsive transcription factor, CaRF, that regulates neuronal activity-dependent expression of BDNF. *Neuron,* 33**,** 383-95.
TARANGER, C. K., NOER, A., SORENSEN, A. L., HAKELIEN, A. M., BOQUEST, A. C. & COLLAS, P. 2005. Induction of dedifferentiation, genomewide transcriptional programming, and epigenetic reprogramming by extracts of carcinoma and embryonic stem cells. *Mol Biol Cell,* 16**,** 5719-35.
TAZI, J. & BIRD, A. 1990. Alternative chromatin structure at CpG islands. *Cell,* 60**,** 909-20.

TEHLIVETS, O., MALANOVIC, N., VISRAM, M., PAVKOV-KELLER, T. & KELLER, W. 2013. S-adenosyl-L-homocysteine hydrolase and methylation disorders: yeast as a model system. *Biochim Biophys Acta,* 1832**,** 204-15.
TEPEREK, M. & MIYAMOTO, K. 2013. Nuclear reprogramming of sperm and somatic nuclei in eggs and oocytes. *Reprod Med Biol,* 12**,** 133-149.
TERANISHI, T., TANAKA, M., KIMOTO, S., ONO, Y., MIYAKOSHI, K., KONO, T. & YOSHIMURA, Y. 2004. Rapid replacement of somatic linker histones with the oocyte-specific linker histone H1foo in nuclear transfer. *Dev Biol,* 266**,** 76-86.
TESAR, P. J. 2005. Derivation of germ-line-competent embryonic stem cell lines from preblastocyst mouse embryos. *Proc Natl Acad Sci U S A,* 102**,** 8239-44.
TESAR, P. J., CHENOWETH, J. G., BROOK, F. A., DAVIES, T. J., EVANS, E. P., MACK, D. L., GARDNER, R. L. & MCKAY, R. D. 2007. New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature,* 448**,** 196-9.
THOMSON, J. A., ITSKOVITZ-ELDOR, J., SHAPIRO, S. S., WAKNITZ, M. A., SWIERGIEL, J. J., MARSHALL, V. S. & JONES, J. M. 1998. Embryonic stem cell lines derived from human blastocysts. *Science,* 282**,** 1145-7.
TIAN, Y., PARAMASIVAM, M., GHOSAL, G., CHEN, D., SHEN, X., HUANG, Y., AKHTER, S., LEGERSKI, R., CHEN, J., SEIDMAN, M. M., QIN, J. & LI, L. 2015. UHRF1 contributes to DNA damage repair as a lesion recognition factor and nuclease scaffold. *Cell Rep,* 10**,** 1957-66.
TISLJAR, K., DEUSSING, J. & PETERS, C. 1999. Cathepsin J, a novel murine cysteine protease of the papain family with a placenta-restricted expression. *FEBS Lett,* 459**,** 299-304.
TOMIOKA, M., NISHIMOTO, M., MIYAGI, S., KATAYANAGI, T., FUKUI, N., NIWA, H., MURAMATSU, M. & OKUDA, A. 2002. Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex. *Nucleic Acids Res,* 30**,** 3202-13.
TREWICK, S. C., HENSHAW, T. F., HAUSINGER, R. P., LINDAHL, T. & SEDGWICK, B. 2002. Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. *Nature,* 419**,** 174-8.
TUREK-PLEWA, J. & JAGODZINSKI, P. P. 2005. The role of mammalian DNA methyltransferases in the regulation of gene expression. *Cell Mol Biol Lett,* 10**,** 631-47.
TURRIZIANI, B., GARCIA-MUNOZ, A., PILKINGTON, R., RASO, C., KOLCH, W. & VON KRIEGSHEIM, A. 2014. On-beads digestion in conjunction with data-dependent mass spectrometry: a shortcut to quantitative and dynamic interaction proteomics. *Biology (Basel),* 3**,** 320-32.
UNNIKRISHNAN, A., AKIYOSHI, B., BIGGINS, S. & TSUKIYAMA, T. 2012. An efficient purification system for native minichromosome from Saccharomyces cerevisiae. *Methods Mol Biol,* 833**,** 115-23.
VAN DEN BERG, D. L., SNOEK, T., MULLIN, N. P., YATES, A., BEZSTAROSTI, K., DEMMERS, J., CHAMBERS, I. & POOT, R. A. 2010. An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell,* 6**,** 369-81.
VARDIMON, L., KRESSMANN, A., CEDAR, H., MAECHLER, M. & DOERFLER, W. 1982. Expression of a cloned adenovirus gene is inhibited by in vitro methylation. *Proc Natl Acad Sci U S A,* 79**,** 1073-7.
VEITH, S. & MANGERICH, A. 2015. RecQ helicases and PARP1 team up in maintaining genome integrity. *Ageing Res Rev,* 23**,** 12-28.
VERREAULT, A., KAUFMAN, P. D., KOBAYASHI, R. & STILLMAN, B. 1996. Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4. *Cell,* 87**,** 95-104.

VERREAULT, A., KAUFMAN, P. D., KOBAYASHI, R. & STILLMAN, B. 1998. Nucleosomal DNA regulates the core-histone-binding subunit of the human Hat1 acetyltransferase. *Curr Biol,* 8**,** 96-108.
WADA, K., ITOGA, K., OKANO, T., YONEMURA, S. & SASAKI, H. 2011. Hippo pathway regulation by cell morphology and stress fibers. *Development,* 138**,** 3907-14.
WANG, J., RAO, S., CHU, J., SHEN, X., LEVASSEUR, D. N., THEUNISSEN, T. W. & ORKIN, S. H. 2006. A protein interaction network for pluripotency of embryonic stem cells. *Nature,* 444**,** 364-8.
WANG, S., KOU, Z., JING, Z., ZHANG, Y., GUO, X., DONG, M., WILMUT, I. & GAO, S. 2010. Proteome of mouse oocytes at different developmental stages. *Proc Natl Acad Sci U S A,* 107**,** 17639-44.
WANG, X., PARK, K. E., KOSER, S., LIU, S., MAGNANI, L. & CABOT, R. A. 2012. KPNA7, an oocyte- and embryo-specific karyopherin alpha subtype, is required for porcine embryo development. *Reprod Fertil Dev,* 24**,** 382-91.
WARD, W. S. 2010. Function of sperm chromatin structural elements in fertilization and development. *Mol Hum Reprod,* 16**,** 30-6.
WARREN, L., MANOS, P. D., AHFELDT, T., LOH, Y. H., LI, H., LAU, F., EBINA, W., MANDAL, P. K., SMITH, Z. D., MEISSNER, A., DALEY, G. Q., BRACK, A. S., COLLINS, J. J., COWAN, C., SCHLAEGER, T. M. & ROSSI, D. J. 2010. Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell,* 7**,** 618-30.
WATANABE, D., SUETAKE, I., TADA, T. & TAJIMA, S. 2002. Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis. *Mech Dev,* 118**,** 187-90.
WEBER, U., RODRIGUEZ, E., MARTIGNETTI, J. & MLODZIK, M. 2014. Luna, a Drosophila KLF6/KLF7, is maternally required for synchronized nuclear and centrosome cycles in the preblastoderm embryo. *PLoS One,* 9**,** e96933.
WEI, H. & YU, X. 2016. Functions of PARylation in DNA Damage Repair Pathways. *Genomics Proteomics Bioinformatics,* 14**,** 131-9.
WHITE, M. K., JOHNSON, E. M. & KHALILI, K. 2009. Multiple roles for Puralpha in cellular and viral regulation. *Cell Cycle,* 8**,** 1-7.
WHYTE, W. A., ORLANDO, D. A., HNISZ, D., ABRAHAM, B. J., LIN, C. Y., KAGEY, M. H., RAHL, P. B., LEE, T. I. & YOUNG, R. A. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell,* 153**,** 307-19.
WILLIAMS, A. B. & SCHUMACHER, B. 2016. p53 in the DNA-Damage-Repair Process. *Cold Spring Harb Perspect Med,* 6.
WILLIAMS, R. L., HILTON, D. J., PEASE, S., WILLSON, T. A., STEWART, C. L., GEARING, D. P., WAGNER, E. F., METCALF, D., NICOLA, N. A. & GOUGH, N. M. 1988. Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature,* 336**,** 684-7.
WOLFFE, A. P. & MATZKE, M. A. 1999. Epigenetics: regulation through repression. *Science,* 286**,** 481-6.
WOSSIDLO, M., NAKAMURA, T., LEPIKHOV, K., MARQUES, C. J., ZAKHARTCHENKO, V., BOIANI, M., ARAND, J., NAKANO, T., REIK, W. & WALTER, J. 2011. 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat Commun,* 2**,** 241.
WRAY, J., KALKAN, T. & SMITH, A. G. 2010. The ground state of pluripotency. *Biochem Soc Trans,* 38**,** 1027-32.
WU, D. Y. & YAO, Z. 2005. Isolation and characterization of the murine Nanog gene promoter. *Cell Research,* 15**,** 317-324.

WU, D. Y. & YAO, Z. 2006. Functional analysis of two Sp1/Sp3 binding sites in murine Nanog gene promoter. *Cell Research,* 16**,** 319-322.
WU, H., WU, X., SHEN, L. & ZHANG, Y. 2014. Single-base resolution analysis of active DNA demethylation using methylase-assisted bisulfite sequencing. *Nat Biotechnol,* 32**,** 1231-40.
WU, H. & ZHANG, Y. 2014. Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell,* 156**,** 45-68.
WU, Q., CHEN, X., ZHANG, J., LOH, Y. H., LOW, T. Y., ZHANG, W., ZHANG, W., SZE, S. K., LIM, B. & NG, H. H. 2006. Sall4 interacts with Nanog and co-occupies Nanog genomic sites in embryonic stem cells. *J Biol Chem,* 281**,** 24090-4.
WU, S. C. & ZHANG, Y. 2010. Active DNA demethylation: many roads lead to Rome. *Nat Rev Mol Cell Biol,* 11**,** 607-20.
WUENSCH, A., HABERMANN, F. A., KUROSAKA, S., KLOSE, R., ZAKHARTCHENKO, V., REICHENBACH, H. D., SINOWATZ, F., MCLAUGHLIN, K. J. & WOLF, E. 2007. Quantitative monitoring of pluripotency gene activation after somatic cloning in cattle. *Biol Reprod,* 76**,** 983-91.
WYLIE, C. 1999. Germ cells. *Cell,* 96**,** 165-74.
XU, G. L. & WONG, J. M. 2015. Oxidative DNA demethylation mediated by Tet enzymes. *National Science Review,* 2**,** 318-328.
XU, L., CHEN, Y. C., CHONG, J., FIN, A., MCCOY, L. S., XU, J., ZHANG, C. & WANG, D. 2014. Pyrene-based quantitative detection of the 5-formylcytosine loci symmetry in the CpG duplex content during TET-dependent demethylation. *Angew Chem Int Ed Engl,* 53**,** 11223-7.
XU, R. H., SAMPSELL-BARRON, T. L., GU, F., ROOT, S., PECK, R. M., PAN, G., YU, J., ANTOSIEWICZ-BOURGET, J., TIAN, S., STEWART, R. & THOMSON, J. A. 2008. NANOG is a direct target of TGFbeta/activin-mediated SMAD signaling in human ESCs. *Cell Stem Cell,* 3**,** 196-206.
YAMAGATA, K., UEDA, J., MIZUTANI, E., SAITOU, M. & WAKAYAMA, T. 2010. Survival and death of epiblast cells during embryonic stem cell derivation revealed by long-term live-cell imaging with an Oct4 reporter system. *Dev Biol,* 346**,** 90-101.
YAMAGUCHI, H. & MIYAZAKI, M. 2014. Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. *Biomolecules,* 4**,** 235-51.
YAN, J. X., WAIT, R., BERKELMAN, T., HARRY, R. A., WESTBROOK, J. A., WHEELER, C. H. & DUNN, M. J. 2000. A modified silver staining protocol for visualization of proteins compatible with matrix-assisted laser desorption/ionization and electrospray ionization-mass spectrometry. *Electrophoresis,* 21**,** 3666-72.
YING, Q. L., WRAY, J., NICHOLS, J., BATLLE-MORERA, L., DOBLE, B., WOODGETT, J., COHEN, P. & SMITH, A. 2008. The ground state of embryonic stem cell self-renewal. *Nature,* 453**,** 519-23.
YING, Y., LIU, X. M., MARBLE, A., LAWSON, K. A. & ZHAO, G. Q. 2000. Requirement of Bmp8b for the generation of primordial germ cells in the mouse. *Mol Endocrinol,* 14**,** 1053-63.
YING, Y. & ZHAO, G. Q. 2001. Cooperation of endoderm-derived BMP2 and extraembryonic ectoderm-derived BMP4 in primordial germ cell generation in the mouse. *Dev Biol,* 232**,** 484-92.
YOUNG, R. A. 2011. Control of the embryonic stem cell state. *Cell,* 144**,** 940-54.
YU, J., VODYANIK, M. A., SMUGA-OTTO, K., ANTOSIEWICZ-BOURGET, J., FRANE, J. L., TIAN, S., NIE, J., JONSDOTTIR, G. A., RUOTTI, V., STEWART, R., SLUKVIN, II & THOMSON, J. A. 2007. Induced

pluripotent stem cell lines derived from human somatic cells. *Science,* 318**,** 1917-20.
YU, Z., LIU, J., DENG, W. M. & JIAO, R. 2015. Histone chaperone CAF-1: essential roles in multi-cellular organism development. *Cell Mol Life Sci,* 72**,** 327-37.
YU, Z., WU, H., CHEN, H., WANG, R., LIANG, X., LIU, J., LI, C., DENG, W. M. & JIAO, R. 2013. CAF-1 promotes Notch signaling through epigenetic control of target gene expression during Drosophila development. *Development,* 140**,** 3635-44.
YURTTAS, P., MORENCY, E. & COONROD, S. A. 2010. Use of proteomics to identify highly abundant maternal factors that drive the egg-to-embryo transition. *Reproduction,* 139**,** 809-23.
ZHANG, H. M., LI, L., PAPADOPOULOU, N., HODGSON, G., EVANS, E., GALBRAITH, M., DEAR, M., VOUGIER, S., SAXTON, J. & SHAW, P. E. 2008. Mitogen-induced recruitment of ERK and MSK to SRE promoter complexes by ternary complex factor Elk-1. *Nucleic Acids Res,* 36**,** 2594-607.
ZHANG, P., NI, X., GUO, Y., GUO, X., WANG, Y., ZHOU, Z., HUO, R. & SHA, J. 2009. Proteomic-based identification of maternal proteins in mature mouse oocytes. *BMC Genomics,* 10**,** 348.
ZHANG, P., SUN, Y. & MA, L. 2015. ZEB1: at the crossroads of epithelial-mesenchymal transition, metastasis and therapy resistance. *Cell Cycle,* 14**,** 481-7.
ZHARKOV, D. O. 2008. Base excision DNA repair. *Cellular and Molecular Life Sciences,* 65**,** 1544-1565.
ZHENG, P. & DEAN, J. 2007. Oocyte-specific genes affect folliculogenesis, fertilization, and early development. *Semin Reprod Med,* 25**,** 243-51.
ZHOU, T., BENDA, C., DUZINGER, S., HUANG, Y., LI, X., LI, Y., GUO, X., CAO, G., CHEN, S., HAO, L., CHAN, Y. C., NG, K. M., HO, J. C., WIESER, M., WU, J., REDL, H., TSE, H. F., GRILLARI, J., GRILLARI-VOGLAUER, R., PEI, D. & ESTEBAN, M. A. 2011. Generation of induced pluripotent stem cells from urine. *J Am Soc Nephrol,* 22**,** 1221-8.
ZHU, B., ZHENG, Y., HESS, D., ANGLIKER, H., SCHWARZ, S., SIEGMANN, M., THIRY, S. & JOST, J. P. 2000. 5-methylcytosine-DNA glycosylase activity is present in a cloned G/T mismatch DNA glycosylase associated with the chicken embryo DNA demethylation complex. *Proc Natl Acad Sci U S A,* 97**,** 5135-9.