

Dissertation  
submitted to the  
Combined Faculties for the Natural Sciences and for Mathematics  
of the Ruperto-Carola University of Heidelberg, Germany  
for the degree of  
Doctor of Natural Sciences

presented by

MSc. in Genetics: Mahmoud-Reza Rafiee  
born in: Tehran, Iran  
Oral-examination: 25.November 2015



Understanding pluripotency by global and targeted  
quantification of  
chromatin-associated proteins

Referees: Dr. Lars Steinmetz  
Prof. Dr. Michael Knop



## Summary

The maintenance of pluripotency in embryonic stem cells (ESC) is regulated by a network of chromatin-associated proteins coordinated by three master transcription factors Oct4, Sox2 and Nanog. To understand how different states of pluripotency are established, I developed three methods for studying chromatin-associated proteins globally, protein-targeted and locus-targeted in mouse ESC.

Firstly, to study chromatin protein composition in a global manner, for the first time I developed *in-vitro* enzymatic labeling of chromatin by biotinylated nucleotides using Terminal deoxynucleotidyltransferase (TdT). As a result, more than 5000 proteins were significantly enriched in mouse ESCs in comparison to the negative control omitting the biotinylation step. In addition to the canonical chromatin-binding proteins, SICAP suggests chromatin association of some unexpected proteins such as Fgf4, which is a growth factor. This observation was further verified by immuno-staining.

Secondly, I combined SICAP with chromatin immuno-precipitation (ChIP-SICAP) to identify proteins that interact with a target protein specifically on chromatin. Using endogenous Oct4, Sox2 and Nanog (OSN) as the targets of ChIP-SICAP, I identified about 400 proteins, as the overlap of the three assays. These 400 proteins include a large number of established interaction partners of the target proteins known to participate in the core pluripotency network (e.g. Rex1, Prdm14, Tcf3, Sall4, Esrrb, Tbx3, Stat3 etc). To reveal the co-localization sites with OSN, I selected Trim24. Interestingly, using ChIP-seq it turned out that Trim24 co-localizes with OSN on many super-enhancers of pluripotency.

Thirdly, I developed a method to identify proteins bound to the Nanog promoter using biotinylated oligonucleotides. The specificity of the method, called targeted isolation of genomic regions (TIGR), was validated using qPCR and high-throughput sequencing. Hence, several proteins have been identified that are known to bind to, and regulate transcriptional activity of Nanog. Comparing the meta-stable and the ground-state of pluripotency, TIGR identified several nucleoporins that associate with the Nanog promoter preferentially in the ground-state of pluripotency. Using ChIP-qPCR I could validate the association of Nup98 to the Nanog promoter.

Taken together, the data generated by the aforementioned methods expand the circuitry of pluripotency, and shed a new light on the differences between the ground-state and meta-stable state of pluripotency. Additionally, the newly developed methods are highly generalizable and independent of cell culture or genetic engineering so that they can be used for studying diverse biological systems.

## Zusammenfassung

Die Pluripotenz in embryonalen Stammzellen (ESC) wird durch ein komplexes Netzwerk von Chromatin-assoziierten Proteinen aufrecht erhalten, das von den drei Master-Transkriptionsfaktoren Oct4, Sox2 und Nanog kontrolliert wird. Um zu verstehen, auf welche Weise verschiedene Zustände der Pluripotenz etabliert werden, habe ich drei unterschiedliche, einander komplementäre Methoden zur Untersuchung von Chromatin-assoziierten Proteinen entwickelt: 1) Analyse auf globaler Ebene, 2) in Bezug auf bestimmte Proteine und 3) in Bezug auf bestimmte Gen-Loci in Maus-ESCs.

Im ersten Ansatz zur Charakterisierung der Gesamtheit der Chromatin-assoziierten Proteine auf globaler Ebene, entwickelte ich die erste beschriebene Methode zur *in vitro* enzymatischen Markierung von Chromatin durch biotinylierte Nukleotide unter Verwendung des Enzyms Terminal Desoxynukleotidyltransferase (TdT) (SICAP). Als Ergebnis wurden mehr als 5,000 Proteine als signifikant in Maus ESCs angereichert identifiziert, im Vergleich zur Negativkontrolle ohne Biotinylierungsschritt. Zusätzlich zu den kanonischen Chromatin-bindenden Proteinen, ergab die SICAP Methode Chromatin-Assoziierung für einige unerwartete Proteine wie z.B. den Wachstumsfaktor Fgf4. Diese Beobachtung konnte durch Immunfärbung bestätigt werden.

Im zweiten Ansatz kombinierte ich die SICAP Methode mit Chromatin Immunpräzipitation (ChIP- SICAP), um Proteine spezifisch zu identifizieren, die mit einem Zielprotein an Chromatin wechselwirken. Die jeweils unabhängige Verwendung endogener Oct4, Sox2 und Nanog (OSN) Proteine als Zielproteine für ChIP-SICAP ergab die Identifizierung von etwa 400 Proteinen, die in allen drei Tests gemeinsam detektiert wurden. Diese 400 Proteine umfassten eine große Anzahl von etablierten Interaktionspartnern, die bekanntermaßen Teil des Pluripotenz-Netzwerks sind (z.B. Rex1, Prdm14, Tcf3, Sall4, Esrrb, Tbx3, Stat3). Am Beispiel Trim24 charakterisierte ich genauere Gen-Loci der Co-Lokalisierung mit den OSN Faktoren. Interessanterweise stellte sich in ChIP-seq Experimenten heraus, dass Trim24 mit OSN auf vielen Super-Enhancern co-lokalisiert.

Im dritten Ansatz entwickelte ich eine Methode, um spezifisch Proteine zu identifizieren, die an einen bestimmten Promotor, in diesem Fall den Nanog-Promotor gebunden sind. Die Spezifität der Methode, die biotinylierte Oligonukleotide verwendet und Gezielte Isolierungsmethode genomischer Regionen (targeted isolation of genomic regions, TIGR) genannt wird, wurde unter Verwendung von qPCR und Hochdurchsatz-Sequenzierung bestätigt. Mehrere Proteine, die bekanntermaßen den Nanog-Promotor binden und dessen Transkriptionsaktivität regulieren, wurden identifiziert. Vergleicht man den meta-stabilen mit dem Grundzustand der Pluripotenz, identifizierte die TIGR Methode mehrere Nukleoporine (z.B. Nup98), die an den Nanog Promotor bevorzugt im Grundzustand der Pluripotenz binden. Dies konnte durch Verwendung von ChIP-PCR validiert werden.

Zusammenfassend erlauben die durch die oben erwähnten Methoden generierten Datensätze ein erweitertes Verständnis der Regulation des Netzwerkes der Pluripotenz. Zudem können die neu entwickelten Methoden stark verallgemeinert und unabhängig von der Zellkultur oder Gentechnologie angewandt werden, und sind daher für die Untersuchung verschiedenster biologischer Systeme relevant.

Dedicated to my parents and my wife who have always been devoted to my success.

‘My heart was illuminated by a thousand suns; but could never ever touch the perfection of a single atom.’

Ibn Sina (980-1037 AD)



# Contents

<b>1</b>	<b>Introduction and Objectives</b>	<b>1</b>
<b>2</b>	<b>From Cell Fate Engineering to Chromatin-Associated Proteins</b>	<b>5</b>
2.1	Chromatin-associated proteins involved in cellular reprogramming . . . . .	6
2.2	Principles of the maintenance of pluripotency in different states . . . . .	10
2.3	Novel methods for studying chromatin-associated proteins	16
<b>3</b>	<b>SICAP: Global View of Chromatin Composition</b>	<b>21</b>
3.1	Designing a new approach for capturing chromatin . . .	23
3.2	Evaluating the specificity and efficiency of chromatin purification by SICAP . . . . .	25
3.3	Identification of chromatin composition in mouse embryonic stem cells . . . . .	25
3.4	Comparing chromatin protein composition between the ground-state and the metastable state of pluripotency .	31
<b>4</b>	<b>ChIP-SICAP: Regional View of Chromatin</b>	<b>37</b>
4.1	Experimental design and proof-of-principle . . . . .	39
4.2	Comparing Oct4, Sox2 and Nanog co-localized proteins on chromatin between 2i and serum media . . . . .	41
4.3	Studying the core-transcriptional circuitry of mouse embryonic stem cells . . . . .	44
4.4	Verification of Trim24 as a novel component of the core-circuitry . . . . .	47
4.5	Studying proteins co-localized with PRC2 on chromatin	50
4.6	ChIP-SICAP is the reciprocal validation of SICAP . . .	52
<b>5</b>	<b>Toward Single Locus Chromatin Composition by TIGR</b>	<b>55</b>
5.1	Experimental design of the assay . . . . .	56
5.2	Evaluating the specificity and efficiency of TIGR . . . .	57
5.3	Identification of proteins associated with the Nanog promoter . . . . .	58
5.4	Comparing the Nanog promoter-associated proteins in 2i and serum conditions . . . . .	60

5.5	Validation of Nup98 as a novel protein that binds to Nanog promoter . . . . .	60
<b>6</b>	<b>Studying the effect of Nup210 on Somatic Cell Reprogramming</b>	<b>65</b>
6.1	Design of the assay . . . . .	66
6.2	Nup210 is required at least for the initiation phase of reprogramming . . . . .	68
6.3	Nup210 is needed for normal proliferation of the MEF cells . . . . .	69
<b>7</b>	<b>Discussion</b>	<b>71</b>
7.1	Specific <i>in-vitro</i> labeling of <i>in-vivo</i> fixed chromatin . .	71
7.2	Dynamics of chromatin-associated proteins in the ground-state and metastable states of pluripotency . . . . .	73
7.3	Reciprocal validation of SICAP and the antibody using ChIP-SICAP . . . . .	74
7.4	Expanding the core-transcriptional circuitry of pluripotency . . . . .	75
7.5	Toward identification of locus-specific associated proteins	76
7.6	Conclusion . . . . .	77
<b>8</b>	<b>Material and Methods</b>	<b>79</b>
8.1	Cell culture and cell fixation . . . . .	79
8.2	SICAP . . . . .	80
8.3	ChIP-SICAP . . . . .	81
8.4	TIGR . . . . .	82
8.5	High pH fractionation and mass spectrometric analysis	84
8.6	ChIP-seq . . . . .	85
8.7	ChIP-seq data analysis . . . . .	86
8.8	List of the antibodies used for the IP . . . . .	87
8.9	GO analysis . . . . .	87
8.10	HPA antibodies and immuno-fluorescence . . . . .	88
8.11	Treating the ES cells with the HA-tagged Fgf4 . . . . .	88
8.12	Nup210 knockdown . . . . .	89
8.13	qRT-PCR and List of the qPCR primers . . . . .	90
	<b>References</b>	<b>93</b>
	<b>Appendices</b>	<b>117</b>
A.1	Abbreviations . . . . .	117
A.2	Supplementary Tables . . . . .	119
A.3	List of publications . . . . .	121
A.4	Acknowledgement . . . . .	122

# 1

## Introduction and Objectives

---

A cellular phenotype is organized by the network of transcription factors (TF) dictating the cell-type specific gene expression profile. Therefore, studying the TF networks and other chromatin-associated proteins are necessary in order to understand how the cellular identity is established. For example, during the development, TFs determine the fate of the cells by inducing the expression of a set of genes required for the differentiation pathway, and suppressing the alternative routes [1]. Additionally, in order to confine the cells in the correct track multiple epigenetic modifications are added to DNA and chromatin by cytosine methylation and histone marks, respectively [2, 3]. As a result, development is unidirectional, and the differentiated cells are phenotypically stable [4-6].

On the other hand, by applying different *in-vitro* or *in-vivo* treatments [7] progressing in developmental processes is reversible. For instance, by replacing the nucleus of an oocyte with that of a somatic cell, it is possible to re-establish pluripotency [8-10]; the ability of a cell to differentiate to all embryonic lineages. Moreover, fusing somatic cells with teratocarcinoma cells [11, 12], embryonic germ cells (EGC) and ESCs [13] generates pluripotent-like cells. Finally evolving these approaches has led to the induction of pluripotency in somatic cells by expressing a few TFs [14]. The latter approach is the easiest and the most viable way of producing patient-specific pluripotent stem cells. Even by expressing a few TFs it is possible to change the cell fate of somatic cells to another somatic cell type without pluripotent intermediates [15-17]. As a result, these TF-based strategies for cellular reprogramming and trans-differentiation inspired many studies on

the implicated mechanism in removing epigenetic properties in somatic cells, and gaining the new TF networks. Unfortunately, TF-based reprogramming procedures in almost all cases are incomplete, inefficient and time-consuming. Consequently, studying the involved mechanism among heterogeneous population of the cells is very difficult. Here I have summarized the current knowledge of reprogramming and cell-fate engineering in chapter 2. In contrast to TF-based reprogramming, induction of pluripotency by cell-fusion and nuclear transplantation is rapid, as the pluripotency markers of the donor nucleus are detectable within a few cell divisions [18, 19]. Therefore, the question arises: what factors in the pluripotent cells are able to erase the epigenetic memory of the somatic cells. In other words, some chromatin-associated factors in pluripotent cells are needed to assist the TFs to induce pluripotency. Finding an answer for this question inspired me to focus on the dynamics of chromatin composition.

Nevertheless, studying chromatin composition is challenging due to its complexity and the contaminants that may carry over during biochemical purification of chromatin. Although chromatin is precipitated easily by traditional centrifugation-based methods, many contaminants may bind to chromatin during cell lysis and chromatin precipitation. Additionally, there is no negative control for this approach, hence, it is hard to rule out artifacts. Furthermore, the nuclear membrane and the endoplasmic reticulum (ER) membrane are fused together. Therefore, even if the nuclei are separated purely from the cytoplasm, it is hard to distinguish between ER proteins and nuclear proteins. Moreover, nuclear proteins are not always bound to chromatin. In fact, it is important to distinguish between the nucleosolic and the chromatin-bound fraction of a given protein. Logically the latter one is more relevant to gene expression regulations.

To distinguish contaminants from true chromatin-binding proteins, Kustatscher et. al.[20] used a multi-classifier approach that assigns a probability of chromatin-binding to each protein, called interphase chromatin probability (ICP). Nevertheless, the fixed ICPs ignore the fact that the chromatin composition is highly dynamic, and that many

transcriptional regulators may bind to chromatin in a context dependent manner.

Therefore, I sought to develop new enabling methods for reliable isolation of chromatin-associated proteins using specific DNA-labeling reactions by terminal deoxytransferase (TdT). Indeed, TdT is a template-independent enzyme that adds biotinylated nucleotides specifically to DNA. In addition, biotin-streptavidin is the strongest non-covalent interaction, which allows very stringent washing steps. Taken together these features provide a platform for developing a robust method for selective isolation of chromatin-associated proteins (SICAP), which I have developed and used for the first time to study the chromatin protein composition (chromatome) in mouse ESCs in chapter 3.

I was also interested to look into the dynamics of chromatome changes, therefore, two cellular states of pluripotency in mouse ES cells were compared. Traditionally, mouse ES cells are cultured by serum, which is referred to as the meta-stable state of pluripotency. Alternatively, serum can be replaced by 2 protein kinase inhibitors (2i)[21], PD0325901 and CHIR99021, driving the ES cells into a condition resembling the preimplantation epiblast, referred to as the ground state of pluripotency [22]. In fact, PD0325901 inhibits mitogen activated protein kinase (Mek) to prevent Fgf4-mediated differentiation, while, CHIR99021 improves cell viability and proliferation by inhibiting glycogen synthase kinase-3 (Gsk3b) thus promoting Wnt signaling [21]. As a result, 2i medium provides a better environment to maintain pluripotency by eliminating fluctuations in the expression of pluripotency transcription factors [23] like Nanog and Rex1 (Zfp42). Comparing 2i and serum conditions provides an interesting model system for studying epigenetic reprogramming as it has been shown that in serum medium DNA is highly methylated, while, in 2i medium it becomes demethylated [24].

Although SICAP is a robust tool for studying chromatome in a global manner, protein composition of chromatin varies regionally according to the different sites of the genome. Therefore, I was looking for a more detailed approach to study the core transcriptional cir-

---

cuitry of the pluripotency. To achieve this aim, I combined chromatin immunoprecipitation (ChIP) and SICAP to identify chromatin-bound interactants of the master TFs of pluripotency including Oct4, Sox2 and Nanog. As will be discussed in chapter 4, comparing 2i and serum conditions using ChIP-SICAP reveals enhanced chromatin-binding of several pluripotency-associated factors, while, their total quantity is not regulated. This observation may give an answer to a question in the field why the 2i condition is a better medium for the maintenance of pluripotency. Furthermore, ChIP-SICAP validates SICAP results, which I have used to verify *Dazl* chromatin-binding in chapter 4.

Although ChIP-SICAP reveals co-localization of proteins on the genome, these do not necessarily occur in all genomic loci. Therefore, I decided to develop a method for identification of proteins interacting with any locus of interest. Although several researchers have tried to achieve this aim by different approaches [25-27], none of them has introduced a versatile method to the field. For targeted isolation of genomic regions (TIGR), I designed specific oligonucleotide probes for the *Nanog* promoter to identify the enriched proteins in comparison the to the non-targeting probe. Using TIGR for comparing proteins differentially associated with the *Nanog* promoter in the ground-state and meta-stable state, it turned out that several nucleoporin preferentially bind to the *Nanog* promoter in 2i medium. I validated this result by ChIP-qPCR using Nup98 antibody, which will be explained in chapter 5.

In chapter 6, I focused on somatic cell reprogramming to understand the role of Nup210 in the induction of pluripotency. Finally, in chapter 7 I discussed the advantages of the newly developed methods in comparison with previous methods.

# 2

## From Cell Fate Engineering to Chromatin-Associated Proteins

---

Cellular reprogramming essentially entails the removal of the epigenetic memory of a somatic cell, and establishing a new transcriptional network in order to achieve a new phenotype. The most extreme example of reprogramming is the induction of pluripotency in somatic mammalian cells by Oct4, Sox2, Klf4 and c-Myc [14]. After this initial discovery it was demonstrated that induction of pluripotency is feasible by a combination of different TFs [28-31], miRNAs [32], and even chemical compounds [33] (reviewed in [34]). Furthermore, it is possible to change the identity of one cell type to another by expressing the master TFs of the target lineage, a procedure which is called trans-differentiation [15-17]. Since the generation of induced pluripotent stem cells (iPSC), it has received a lot of attention in field due to its clinical potential. Therefore, during the last decade many groups have been trying to elucidate the implicated mechanism of reprogramming in order to improve it.

In principle, during the reprogramming procedure the epigenetic properties of the somatic cells are gradually removed, being replaced by a transcriptional circuitry that sustains pluripotency. Indeed, the pluripotency transcriptional network is mainly regulated by the same transcription factors that have been used to induce reprogramming. In other words, the three master transcription factors Oct4, Sox2 and Nanog not only are required for the maintenance of pluripotency in the ES cells, but also are able to impose the pluripotency transcriptional framework in the somatic cells. Therefore, studying chromatin biology of pluripotency has been very helpful for understanding and enhancing cellular reprogramming. Chromatin functions are carried

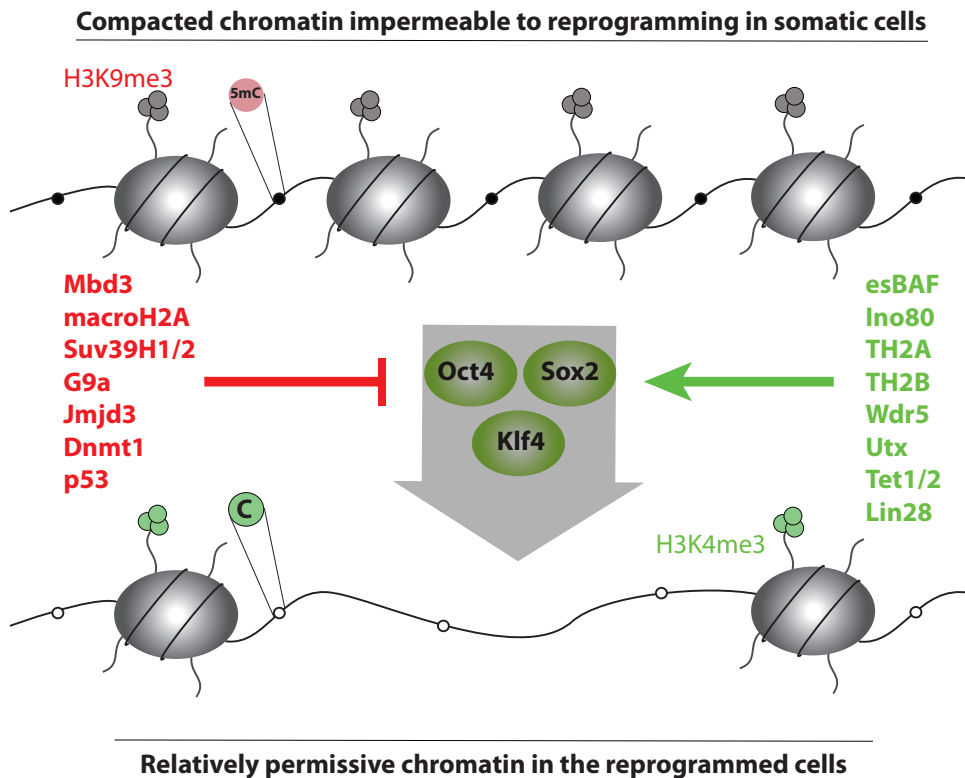
out by numerous proteins. From histones and their modifications to transcription factors and chromatin remodelers, many proteins are involved in the maintenance of pluripotency. Despite genome-wide RNAi-screening for the identification of regulators of pluripotency [35], many regulators of pluripotency have remained uncharacterized due to the inherent limitations of knock-down/out approaches (e.g. off-target knock down, lethal consequence of knock down, insufficient knock down, over-lapping functions of the proteins). On the other hand, with the advancement of mass spectrometry in the recent years, several proteomic approaches have emerged for identification of novel chromatin-associated proteins. Thus in this chapter, I have reviewed what we have learned so far from chromatin biology of pluripotency and reprogramming. Then, I have covered the recent methods for studying chromatin protein composition.

## **2.1 Chromatin-associated proteins involved in cellular reprogramming**

Upon expressing Oct4, Sox2, Klf4, Myc (hereafter as OSKM) the transcriptional network of the somatic cells is perturbed, and gradually the ES cells markers are expressed during a process that normally takes 2 weeks. Consequently, ES-like characteristics are established [36, 37]. It has been shown that OSK bind to various regions of the genome, and have a pioneering role by recruiting other TFs and chromatin remodelers [38-40]. In fact, in order to reach to the pluripotent state, somatic repressive heterochromatin should be converted to the ES-like permissive euchromatin. Therefore, chromatin remodelers required for the maintenance of the ES cells, such as esBAF (Brg-associated factors in ES cells) [41] and Ino80 (inositol-requiring 80) [42], also play important roles in reprogramming. As a result, expressing BAF or Ino80 components together with OSK facilitates binding of the TFs to their targets, and improves the efficiency of reprogramming [42, 43]. On the other hand, in the absence of Mbd3, a NuRD component required for silencing pluripotency genes during differentiation of the ES cells



[44], reprogramming is dramatically boosted [45], even without c-Myc or Sox2 [46]. Nevertheless, this phenomenon seems to be a context-dependent observation, as another study has claimed the necessity of Mbd3 for reprogramming [47].



**Figure 2.1:** Chromatin-associated proteins influencing the somatic cell reprogramming. In order to accomplish cellular reprogramming, compact chromatin has to be opened, demethylated, and suppressive histone marks such as H3K9me3 should be converted to H3K4me3. Chromatin-associated proteins with positive and negative effect in these procedures have been shown with green and red, respectively. 5mC: 5-methylcytosine. Adapted from [48].

In addition to the chromatin remodelers, histone variants also have impact on the induction of pluripotency. Interestingly, TH2A and TH2B, two histone variants highly present in oocytes, are able to improve cellular reprogramming [49]. In fact, the expression of TH2A and TH2B in somatic cells increases the DNase I sensitivity. There-

fore, their contributions in reprogramming is probably mediated by opening the chromatin structure [49].

In contrast to TH2A and TH2B, macroH2A is a barrier to reprogramming [50]. Interestingly, loss of macroH2A in ES cells interferes with the differentiation of the cells due to the partial inactivation of pluripotency genes during the differentiation [51].

Noteworthy chromatin-associated proteins related to the histone modifications play important roles during cellular reprogramming. Wdr5, a component of Trithorax complex, is required for efficient reprogramming of somatic cells by rapid di-methylation of H3K4 in the initial stage of reprogramming [52, 53]. Additionally, Wdr5 is involved in ES cell self-renewal by regulating H3K4 methylation.

On the other hand, Suv39H1/H2 hinders reprogramming [39] by trimethylation of H3K9, which is a marker of heterochromatin regions. In fact, OSKM do not bind to the genomic regions with this histone marks. Therefore, knocking down Suv39H1/H2 enhances the efficiency of reprogramming [39, 54]. Likewise, inhibiting G9a histone methyltransferase is beneficial for reprogramming due to the decreasing H3K9 di-methylation [55, 56].

Although decreasing H3K9 is helpful for reprogramming, the effect of H3K27 tri-methylation on cellular reprogramming depends on the enzymes and the loci. For instance, decreasing Utx, which is an enzyme with the ability to remove H3K27 tri-methylation [57], severely affects reprogramming [58]. In fact, Utx removes H3K27 methylation from the promoters of the pluripotency genes such as *Egf4*, *Sall4* and *Sall1*. In contrast, down-regulation of *Jmjd3*, a H3K27 demethylase, improves reprogramming by blocking *Ink4/Arf* expression and decreasing cell senescence [59]. Additionally, for efficient cellular reprogramming components of the PRC2 (polycomb repressive complex 2) are needed to maintain H3K27 at least in some of their targets [54, 60].

One of the key events during the cellular reprogramming is demethylation of the genome. In fact, DNA demethylation is critical for successful iPS reprogramming, and reactivation of gene expression. Therefore, aberrantly methylated promoters and enhancers may cause abnormali-

ties in gene expression in the iPS cells, and incomplete conversions [61]. As a result, targeting Dnmt1 (DNA methyltransferase 1) by knock-down or chemical compounds like Aza deoxy-Cytidine, significantly enhances iPS cell generation [62]. Nevertheless, knocking down Dnmt3a and Dnmt3b has almost no effect on reprogramming, suggesting that de novo DNA methylation does not have a major role in dedifferentiation [63].

Furthermore, Tet1 and Tet2 enzymes with the ability to oxidize methyl Cytosine have been reported to contribute in somatic cellular reprogramming [64, 65], as their knock downs affects the efficiency of iPS cell generation. In the case of Tet1 this effect seems to be vitamin C-dependent [66]. However, recently it has been shown that the contribution of Tet1 and Tet2 is only until mesenchymal to epithelial transition (MET), and knocking out all the Tet enzymes in epithelial cells has no effect on reprogramming [67].

In addition to the TFs, and proteins related to the epigenetic properties, cell cycle regulators also impact iPSC generation. In fact, ES cells have a relatively short cell cycle with a quick G1 phase, and long S phase. Upon differentiation, the cell cycle is longer and the G1 phase is dominated. Interestingly as an early event during cellular reprogramming, the cell cycle changes, and becomes similar to that of the ES cells [68]; as a result, the faster cell cycle the higher rate of cellular reprogramming, and vice versa. Hence, down-regulation of p53/p21 or over-expressing lin28 significantly accelerates the rate of reprogramming [69]. It seems that replication of DNA provides a window of opportunity for OSKM or other TFs to bind to their targets on the genome, before these regions are able to inherit the epigenetic characters of the parental DNA [48].

Taken together, in all the aforementioned cases a chromatin-associated protein with a critical role in the maintenance of the ES cells, contributes to the somatic cell reprogramming. Thus by studying pluripotency, and by applying new technologies for studying chromatin protein composition it should be possible to find potent candidates that facilitate or accelerate iPS cell generation.

## 2.2 Principles of the maintenance of pluripotency in different states

Pluripotency is a potential capability of a cell in generating all cells of an adult organism. Naturally pluripotent cells are generated during embryonic day 4 (E4). At this time, the embryo is called blastocyst, and consists of two cellular components: the inner-cell mass (ICM), and the trophoctoderm (TE). Before the implantation, the ICM cells are also divided into two groups, the epiblast cells that homogeneously express *Nanog*, and the primitive endoderm cells that express *Gata6* [70]. Developmentally the pre-implantation epiblast cells are at the ‘ground state’ of pluripotency. In other words, these cells are the ancestors of all future adult cells. Therefore, they are called ‘naive’ pluripotent stem cells, which means they have no developmental bias [71]. Functionally, these cells are able to produce a healthy and live embryo if they are engrafted into a tetraploid embryo (4N complementation assay) [72]. In fact, the tetraploid cells are not capable of completing the fetal development. At the later stage of the development (E5), when the embryo has implanted, the epiblast cells are functionally called ‘primed’ pluripotent cells, as they are primed to start the differentiation programs. As a result, these epiblast stem cells (EpiSC) are no longer able to pass the 4N complementation assay [22].

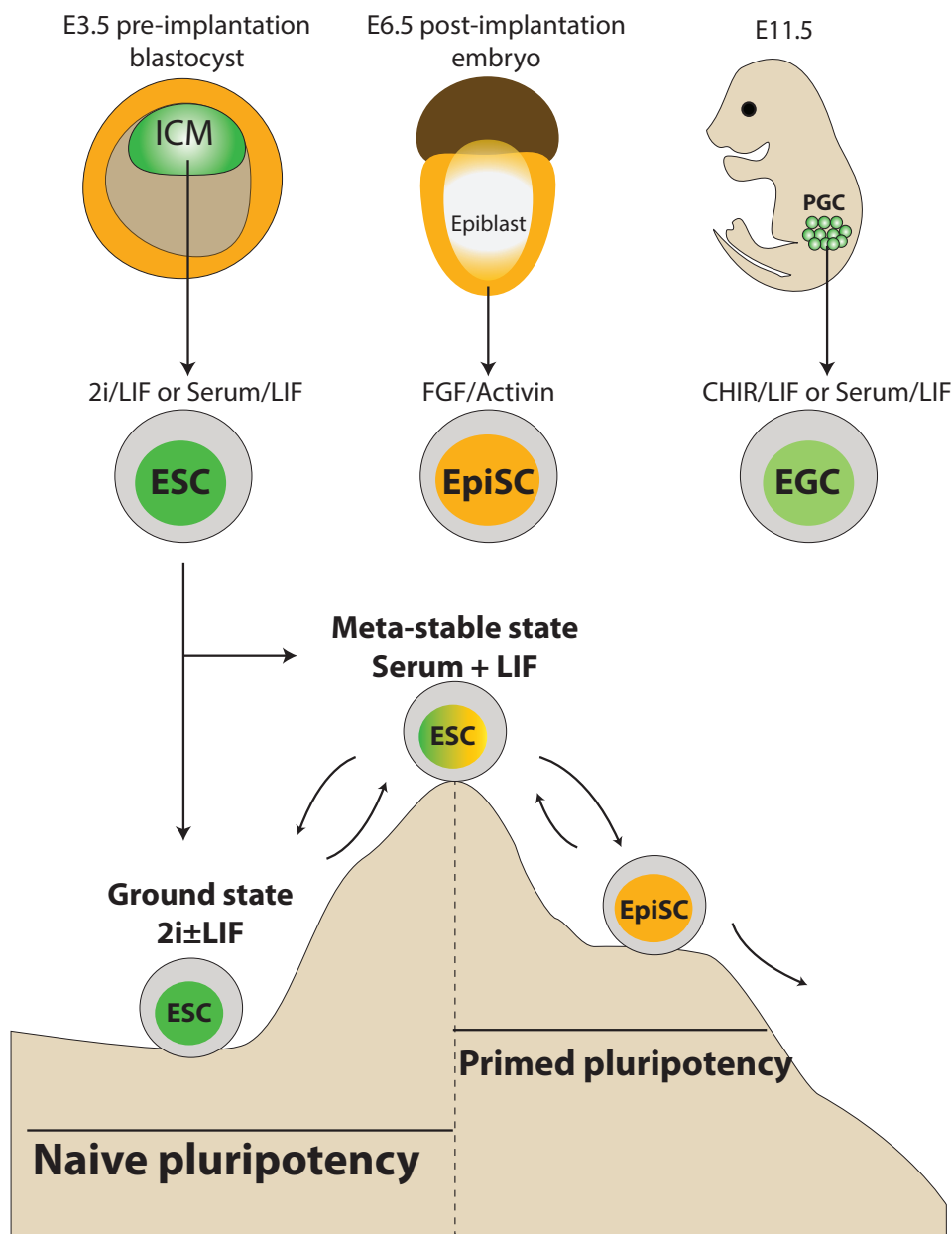
From the epigenetic point of view, in the naive state the genome is globally de-methylated, and the paternal X-chromosome is reactivated in the female embryos. Therefore, epigenetically there is no boundary. However, in the primed state the genome is methylated again, and the cells are ready to start the lineage commitment programs [73]. Interestingly by applying appropriate in vitro conditions it is possible to pause the cells in the pluripotent stage. Initially this was achieved by culturing ICM cells using serum-containing medium on a layer of the inactivated fibroblast cells, which is called the feeder layer. Later it turned out that the feeder layer secretes a factor (leukemia inhibitory factor, LIF) that supports the self-renewal of the ES cells [74, 75]. Therefore, the feeder cells can be omitted if LIF is added to the medium, and a

supportive extracellular matrix like gelatin applied to the culture plate for attaching the ES cells to the plastic. The ES cells grown using serum and LIF with or without feeder are called ‘serum’ ES cells.

Indeed, LIF supports self-renewal of the ES cells, by the activation of JAK kinase and the Stat3 pathway. Without Stat3, adding LIF to the medium does not contribute to the self-renewal of the ES cells. Conversely, overexpression of Stat3 is enough to keep the ES cells undifferentiated without adding LIF [76]. It seems that Stat3 induces the expression of Tfc2l1, which is also a TF. In fact, Tfc2l1 is necessary for LIF-responsiveness, and its forced expression supports LIF-independent self-renewal [77, 78].

It has been shown that Bmp4 is the critical component of serum, as in the absence of serum it allows the maintenance of the ES self-renewal together with LIF [79]. In fact, Bmp4 activates the expression of Inhibitor of differentiation (Id) genes via the Smad pathway [79].

Culturing the ES cells using serum and LIF supports the naive pluripotency, as the ES cells are able to contribute to the chimeric mouse, or tetraploid embryos. However due to the undefined composition of serum, contradicting signaling pathways are activated in the ES cells. As a result, there is quite some heterogeneity among the serum ES cells, as pluripotency markers such as Nanog and Rex1 fluctuate between high and low level of expression [80-82]. In other words, some cells have kept their naive pluripotent identity, and some others are primed to differentiate. Nevertheless, these subpopulations of the ES cells dynamically convert their identity [83]. Therefore, the serum ES cells are considered to be in a ‘metastable’ state of pluripotency [73].



**Figure 2.2:** Diversity of pluripotent stem cells. Functionally ICM-derived stem cells are called naive pluripotent stem cells. Culturing the ICM-derived cells in 2i medium, with or without LIF, stabilizes naive pluripotency. ES cells cultured in 2i medium are closely similar to the ICM cells (the ground state). Although, serum and LIF stabilizes naive pluripotency, the ES cells are heterogeneous from different aspects (the meta-stable state). Following implantation of the embryo, stem cells derived from epiblast (EpiSCs) are functionally called primed pluripotent stem cells, as they are primed to be differentiated. Furthermore, primordial germ cells (PGCs) can generate embryonic germ cells (EGCs) by in vitro culturing, which are highly comparable to the ES cells. Adapted from [73]

After years of studying on the signaling pathways in the mouse ES cells, it turned out that FGF/ERK pathway is responsible for triggering differentiation of the ES cells. Therefore, by adding a specific chemical inhibitor of the FGF pathway (PD0325901, or PD03) to the medium of the ES cells, they are able to self-renew without LIF or serum [21]. In order to improve the viability and self-renew of the cells, the chemical inhibitor of GSK3 (CHIRON) was supplemented, as well [21, 84, 85]. Therefore, the new serum-free medium is called ‘2i’ referring to the 2 chemical inhibitors.

While PD03 prevents differentiation of the ES cells, inhibition of Gsk3 by CHIRON stabilizes B-Catenin. As a result, B-Catenin moves into the nucleus, and prevents the repressive effect of Tcf3 on Oct4 and Sox2 [86]. In fact, Tcf3 depletion prevents differentiation of the ES cells [87], and substitutes Gsk3 inhibition [88] for improving self-renewal. Remarkably, the 2i ES cells homogeneously express the pluripotency markers such as Nanog and Prdm14. In addition, the genome of the ES cells is globally de-methylated [89, 90]. Furthermore, 2i ES cells contribute to the chimera embryos with a higher rate in comparison to serum ES cells. Importantly, using 2i medium germ line-competent ES cells can be generated from impermeable strains like non-obese diabetic (NOD) or FVB [91-93]. Therefore, the ES cell grown in 2i medium satisfy the definition of naive pluripotency better than the serum ES cells, nevertheless, both of them are functionally considered as naive pluripotent stem cells. However, in terms of transcriptional signature and epigenetic profile, the serum ES cells are different with the pre-implantation epiblast cells, while, the 2i (with or without LIF) ES cells are closely similar to the ground-state epiblast cells. In fact, single-cell transcriptome analysis indicates that 2i ES cells are clustered closely with E4.5 epiblast cells [94, 95]. Therefore, serum ES cells are not characterized as ground-state, whereas, the 2i (+LIF) ES cells are the most similar representative of the ground-state epiblast cells in *in-vitro* conditions [73].

As was mentioned previously, the identity of the ES cells primarily depends on three master transcription factors, Oct4, Sox2 and Nanog

(OSN), which together constitute the core-transcriptional circuitry of pluripotency. Additionally, they recruit other pluripotency auxiliary TFs such as Tbx3, Esrrb, Klf4, Klf2, Sall4, Stat3, Tfcp2l1 and Prdm14 etc. to specific enhancers and promoters in order to maintain the transcriptional network of pluripotency [73]. Although the auxiliary TFs are markers of pluripotent stem cells, depending on 2i or serum conditions they may become essential or dispensable for the maintenance of pluripotency. For instance, in serum condition Stat3 and Tfcp2l1 are indispensable [77], while, Esrrb and Klf2 are not required for the maintenance of pluripotency. Conversely, in 2i medium the latter two TFs are indeed essential [96], whereas, Stat3 and Tfcp2l1 are not imperative. Although 2i medium is a better condition for persevering pluripotency, OSN and most of the pluripotency auxiliary TFs do not dramatically change in expression level between 2i and serum conditions at least at the level of mRNA. For example, Oct4, Sox2, Nanog, Tbx3, Esrrb, Klf4 and Klf2 do not show significant changes, while, Prdm14, Dppa3 (Stella), Tfcp2l1 and Tcl1 are up-regulated 3-fold [95]. On the other hand, Utf1, Lin28b and Id genes are down-regulated in 2i medium (between 2 to 5-fold). Therefore, one remaining question in the field is how 2i medium provides a better tuned condition for the ES cells, while there is no such a significant change in the expression of pluripotency-associated genes [97]. Furthermore, Myc is dramatically down-regulated in 2i medium (32-fold). In fact, such an intensive down-regulation of Myc may bring about cell cycle arrest, which is not tolerated in the ES cells, and may lead to their differentiation. Thus another unanswered question is how 2i ES cells are able to compensate the loss of the Myc network. One of the remarkable differences between 2i and serum conditions, is the reduction of global DNA methylation in the 2i ES cells [24, 89, 98]. Basically, ICM cells are globally hypomethylated just before the implantation [99, 100], whereas, after the implantation the genome of the epiblast cells is methylated again. Interestingly, the hypomethylation occurs rapidly in 2i medium, and it can be acquired on the correct genomic sites in the serum medium [6]. Initially it was thought that TET-mediated oxida-



tion of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) is needed for the demethylation of DNA. However, Tet1 and Tet2 knock out ES cells are still able to demethylate their genome in 2i medium. Surprisingly, knocking out Prdm14 in the ES cells prevents their DNA hypomethylation in the 2i condition [90, 101]. It seems that Prdm14 contributes in DNA demethylation by suppressing the expression of de novo DNMTs [90], and by promoting the activity of the TET enzymes [102].

In contrast to male ES cells, the genome of the female ES cells is hypomethylated in serum condition. It has been shown that the hypomethylation of the female genome is due to the activation of both X chromosomes. As a result, MAPK and Gsk3 pathways are inhibited, and Akt pathway is stimulated [103].

In addition to DNA methylation, histone marks are different between 2i and serum conditions. For example, H3K27me3, a histone mark generated by PRC2, is depleted in the 2i condition [104]. This could be due to the inhibition of ERK pathway, as it is required for the activity of Eed [104]. Interestingly, PRC2 is not required for the maintenance of pluripotency, however, it is needed for appropriate differentiation of the ES cells. Therefore, removing H3K27me3 mark from developmental genes does not activate them, indicating that there are other mechanisms for repressing the differentiation genes in the ES cell [95]. In 2i condition RNA polymerase 2 (Pol II) is paused at the proximal promoters more frequently than the serum condition especially at the promoter of developmental genes. Thus PolII pausing could be an important transcriptional regulatory mechanism in the ground state of pluripotency.

Recent studies claim that H3K9me2 and H3K9me3, the histone marks of heterochromatic regions, are globally lower in the 2i condition [89]. On the other hand, H3K4me3, an activating histone mark, increases in 2i medium. At the same time, Padi4 is up-regulated, which is able to convert arginine to citrulline on histone H1. As a result, chromatin is decondensed [105].

In contrast to the 2i and serum conditions for preserving naive

---

pluripotency, primed pluripotency is stable by adding bFgf and Activin A to the medium of the EpiSCs [106]. EpiS cells are able to differentiate *in-vitro*, and to generate teratoma *in-vivo*, however, they are not able to contribute in chimeric mouse. In comparison to the ES cells, EpiSCs express *Esrrb*, *Klf4*, *Klf5*, *Nanog*, *Prdm14* and *Zfp42* (*Rex1*) at lower levels [107]. In addition, the female X-chromosome is deactivated in the EpiS cells, and the promoters of pluripotency-associated genes are methylated. Furthermore, pluripotent stem cells can be derived from primordial germ cells (PGC). Hence, they are called embryonic germ cells (EGC), which are almost indistinguishable from the ES cells [89].

Altogether, plasticity of pluripotent stem cells in transforming from one state to the other state of pluripotency provides valuable model systems for studying cellular reprogramming. Indeed, chromatin-associated proteins are at the core of maintaining and abolishing pluripotency. Hence by developing new enabling tools for studying chromatin biology we should be able to tackle the unanswered questions, gain new insight in the mechanism of gain and loss of pluripotency, and design new approaches for improving somatic cell reprogramming.

### **2.3 Novel methods for studying chromatin-associated proteins**

In order to understand the functions of the genome, it is essential to study the organization of proteins that occupy the genome and that regulate its function. Therefore, in the Encyclopedia of DNA Elements (ENCODE) project, many research groups have contributed to profile the binding sites of TFs and other chromatin-associated proteins on the genome. In addition to the TF-occupancies, characterization of histone modifications and their localizations on the genome were the other goals of the ENCODE. To achieve these goals, Chromatin immuno-precipitation (ChIP) coupled to the high-throughput sequencing (ChIP-seq) has been the method of choice. Furthermore, to shed light on the 3D structure of the genome they have studied long-range DNA-DNA interactions by methods such as Hi-C and ChIA-PET.

Although ChIP-seq has been a very successful method, by definition it targets proteins one at a time and thus does not uncover the protein composition of chromatin in an unbiased manner. Indeed, due to the artifacts associated with the isolation of chromatin, identification of novel chromatin-associated proteins by mass spectrometry has been challenging. Nevertheless, recently in parallel with advances in mass spectrometry, several new methods have been developed to reveal the complex array of chromatin-associated proteins. Here I have described some of these novel methods developed since 2009, to discuss their advantages and limitations.

One of the problems in adapting the ChIP method for proteomics studies is the large amount of the background proteins. In fact, ChIP-seq takes the advantage of DNA sequencing as the read-out, which provides two dimensions for the signal: 1- location of the signal on the genome, and 2- the intensity of the signal. As result, the differences with the negative control is determined easily when the intensity of the signal is mapped throughout the genomic loci. However, using mass spectrometry only the intensity of the signal can be used as the read-out. In other words, the enriched proteins are identified using a ratio in comparison to the negative control. Using the gentle washing steps that are applied to the ChIP procedure, it is hard to remove the background proteins efficiently, which is sufficient when targeting only a single protein as in ChIP, but which becomes problematic when using an unbiased approach such as mass spectrometry. Therefore, some true positives are rejected because of lack of sufficient enrichment, and on the other hand some false positives are included among the true positives due to the inadequate removal of the background proteins. To solve this issue, two methods have been developed, modified ChIP (mChIP) [108] and ChIP-MS (mass spectrometry)[109], both of which somehow used tandem affinity purification (TAP)-tagging for increasing the signal to noise ratio.

In mChIP, no formaldehyde cross-linking is used. Instead, chromatin is fragmented by sonication, or using micrococcal nuclease (MNase) digestion, followed by tandem-affinity-purification (TAP) tagging. Then

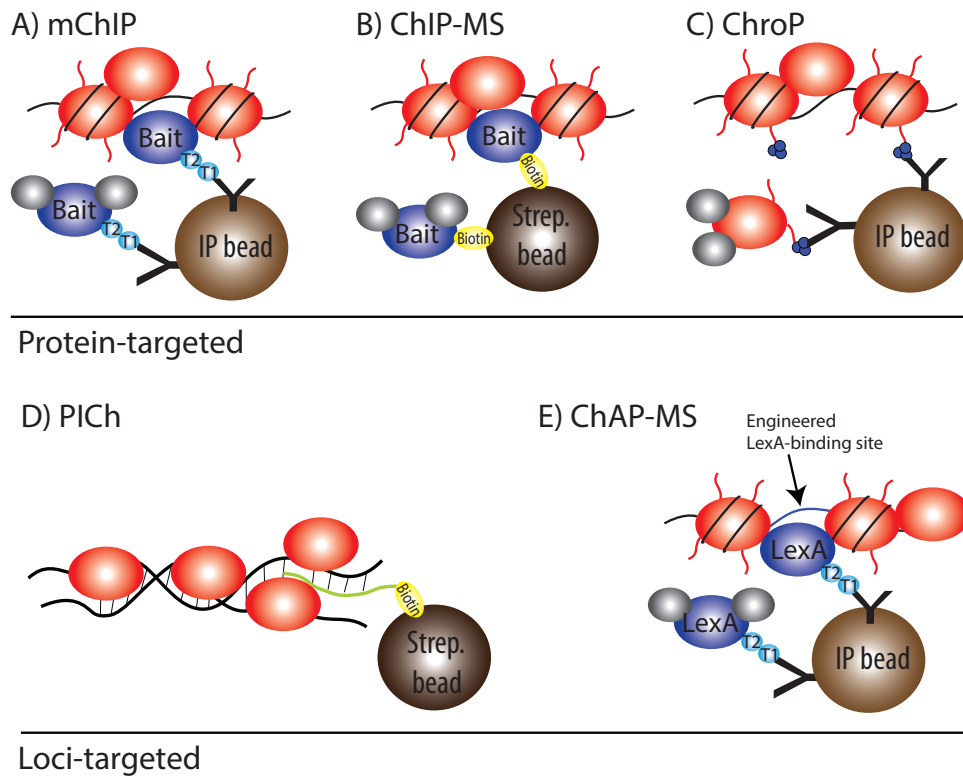
proteins are resolved using the SDS-PAGE to remove the antibody bands from the other proteins, and finally the proteins are digested to be injected to the mass spectrometer for protein identification. Indeed, TAP-tagging makes the IP very specific, and makes the method feasible for the proteins without proper antibodies. However, the efficiency of IP is quite low, and when it is carried out two times sequentially it becomes even less efficient. As a result, a massive number of input cells is needed (10 billion cells), which is normally only achievable for yeast cell culture but not mammalian cells.

In ChIP-MS, however, robust formaldehyde cross-linking is applied (3%, 30 min). In addition, the target protein has a His-tag, and a bacterial peptide sequence to be biotinylated using the endogenous biotin-ligases (HTB-tag). Therefore, no antibody is needed, and the target protein can be captured using streptavidin beads. As a result, stringent washing can be applied to remove the background proteins. This method was used for the first time to study the dosage compensation in drosophila using 1 billion cells as the input. Although the amount of required cells is less than the previous method, still it is not routinely achievable. In addition, using mChIP or ChIP-MS tagging the proteins may affect the function of the bait, or may bring about artifacts due the expression level of the bait.

In line with the aforementioned methods, chromatin proteomics (ChroP) [110] is a method for identification of chromatin-associated proteins using histone modifications as the bait. There is no need to tag the bait histones, probably because histones are very abundant, which may help to improve the signal to noise ratio. Additionally, ChroP has been applied to mammalian cells using 100 million cells.

The above mentioned methods focus on a target protein to identify physical interactions in a global fashion. To identify proteins associated with the locus of interest, Dejardin and Kingstone have developed a method called proteomics of isolated chromatin fragments (PICh) [111]. In this method they have used a biotinylated oligonucleotide as the probe to target the telomeres in human and drosophila cells. Additionally, they have successfully targeted pericentromeric regions

in mouse ES cells [112]. The main advantage of this method is that it is independent of antibodies. In fact, designing and synthesizing an oligonucleotide probe is much simpler than producing a ChIP-grade antibody. However, so far it has only been useful for the target loci with tens of copies per genome.



**Figure 2.3:** The recent methods for studying chromatin can be divided into two groups: Protein-targeted methods (including mChIP, ChIP-MS and ChroP), and Loci-targeted methods (PICh and ChAP-MS). A) In mChIP the target protein is purified by TAP-tagging. B) In ChIP-MS the target protein is expressed in conjunction with a bacterial peptide to be biotinylated by endogenous biotin-ligases. C) In ChroP, histone modifications are targeted by specific antibodies, to identify the co-localized proteins. D) In PICh the target loci are captured using a biotinylated oligonucleotide probe. E) in ChAP-MS LexA-binding site is engineered in the target locus, and subsequently it is captured by expression and immuno-precipitation of LexA. T1: Tag1, T2: Tag2.

Another method to identify proteins that bind to a given genomic locus is called chromatin affinity purification with mass spectrometry (ChAP-MS)[25]. In this method, LexA-binding site is integrated adjacent to the locus of interest. Then LexA is expressed in the cells to bind to the engineered locus. Using this method, proteins-associated with Gal1 locus have been studied in yeast in the presence of Glucose and Galactose to turn the target locus transcriptionally off and on, respectively. Although some of the expected proteins have been identified using this method, the amount of required input cells is enormously high (100 billion cells). In addition, cloning the LexA-binding site is a barrier for expanding this method to other cellular systems, and may interfere with the expression of the target loci. To circumvent this, instead of LexA-binding site transcription activator-like (TAL) was used for targeting the locus of interest [26]. However, it seems that TAL does not bind to the Gal1 locus when the expression is turned off. In other words, TAL binds to the transcriptionally active targets. Altogether, identification of chromatin-associated proteins certainly contributes to understanding the mechanism of transcriptional regulation, but is technically challenging. Therefore, many investigator shave attempted to develop novel methods for achieving this aim. It is clear that there are some gaps in the methods to be filled to circumvent the shortcomings of current methods. For example, mChIP, ChIP-MS and ChroP fail to answer a question: are the identified proteins located on the chromatin, or are they soluble interactants of the bait proteins? Beyond multi-copy loci such as telomeres, are we able to identify proteins that associate with a single locus? Answering these questions demands developing new methods, which will be discussed in the following chapters.

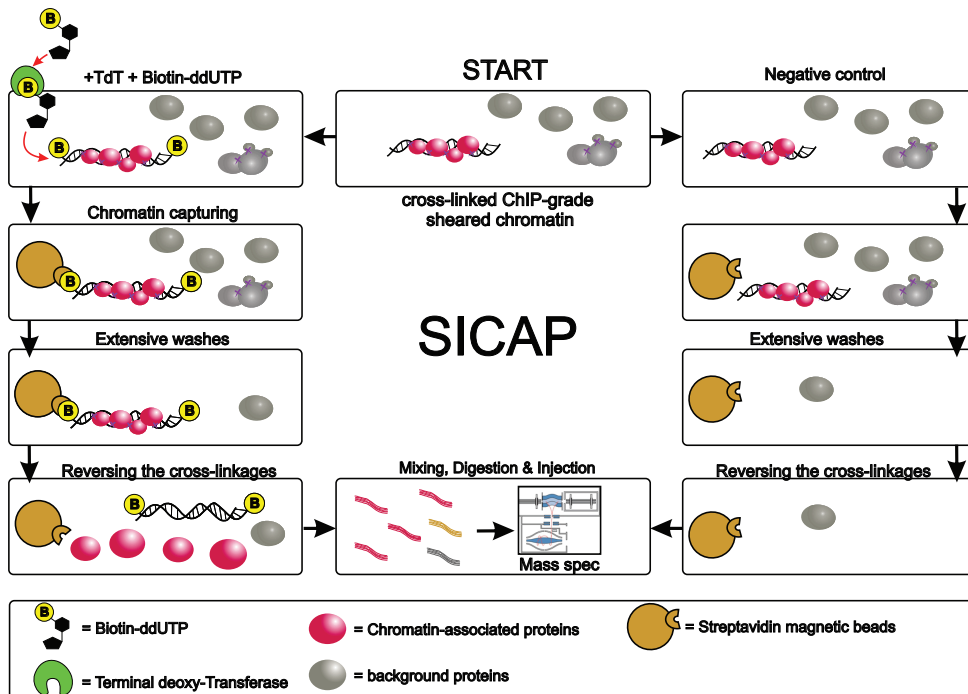
# 3

## SICAP: Global View of Chromatin Composition

---

Chromatin is an organized cellular complex that contains genomic DNA together with proteins and RNAs [113], which are required for the repair, duplication and transcription of the genome. A wide array of proteins is involved in each of these biological processes, and therefore the identification of the implicated proteins is absolutely essential to gain better mechanistic insights into genome regulation. Constructing the global view of chromatin has remained challenging due to the background proteins associated with chromatin purification. Chromatin is a very large complex with the propensity to precipitate easily after cell lysis. In addition, DNA is a highly charged molecule; therefore, many non-specific chromatin proteins may bind to chromatin during cell lysis. The traditional method developed for the purification of chromatin [114] does not have a proper negative control to rule out the contaminants. For instance, mitotic chromosomes, which are quite dense and distinguishable, could be isolated using sucrose and Percoll gradients [115]. Although this approach was used to identify new substrates of Aurora B Kinase [116], there are many non-canonical proteins. Considering that there is no negative control for density centrifugation, the final list of proteins may look unreliable. To solve this problem recently Ohta et. al. [20] have used machine learning approaches with different classifiers called multiclassifier combinatorial proteomics (MCCP) to set up a random forest analysis. As a result, chromosomal proteins were distinguished with high specificity. Nevertheless, the classifiers used in this study are feasible for mitotic chromosomes. Consequently, the method can not be extended to interphase chromatin. Many important transcription factors and chromatin remodelers detach from

the condensed chromosome during mitosis. Therefore, more recently Kustatscher et. al. [20] used a similar bioinformatic approach to define interphase chromatin probability (ICP) for proteins. Although this approach could be useful in specific cases [117], it does not reflect the dynamic essence of chromatin binding due to the fixed probabilities that are assigned to the proteins. In other words, for some transcription factors such as STATs or SMADs always low ICPs are considered, while depending on the model system of interest these factors may bind more or less to chromatin.



**Figure 3.1:** Schematic presentation of SICAP. For further details, please refer to the text (section 3.1).

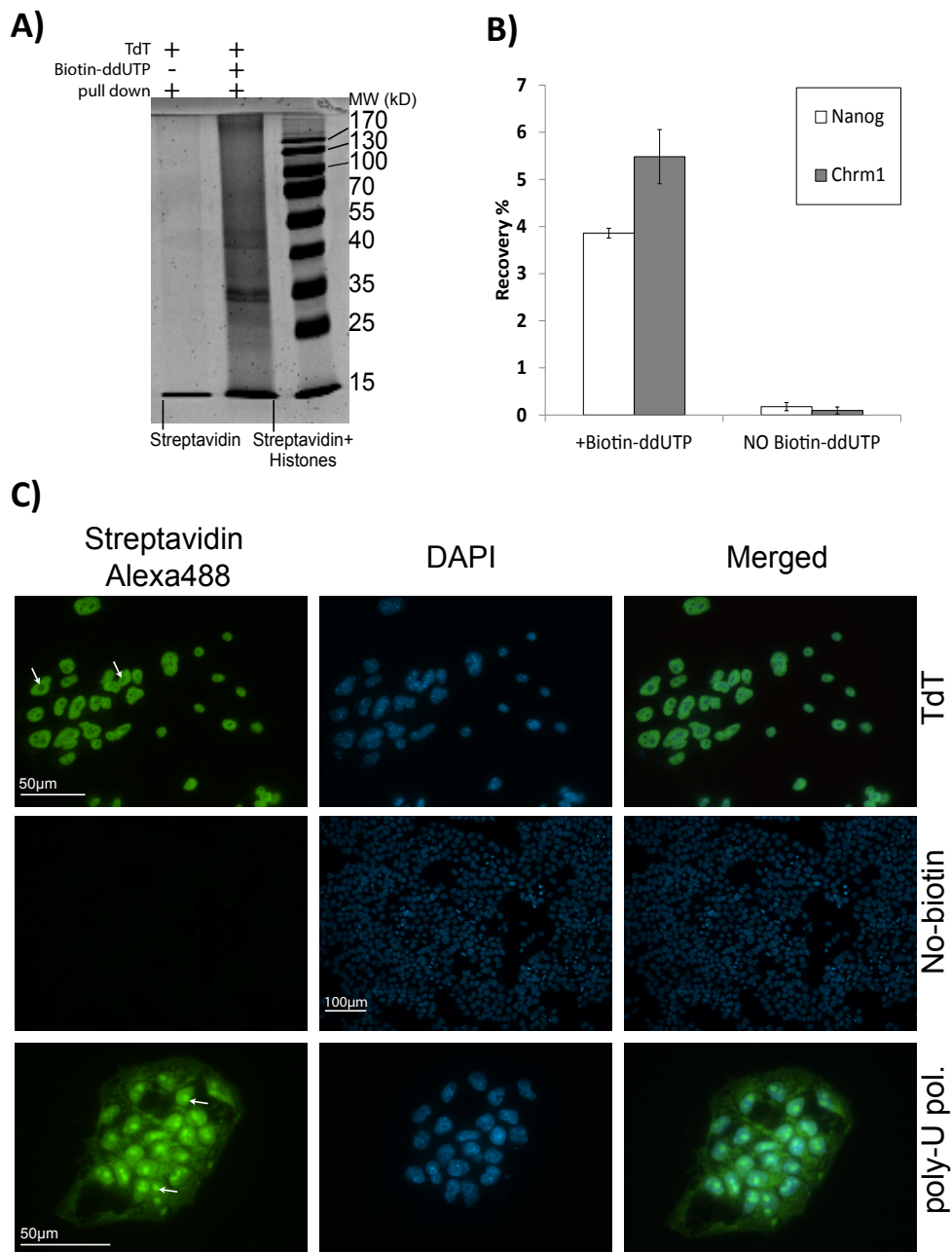
In this study to avoid the problems of traditional chromatin-prep methods with the background proteins, and to take the dynamics of chromatin into consideration practically, I developed a new method called Selective Isolation of Chromatin Associated Proteins (SICAP, fig. 3.1). Using SICAP we are able to specifically label DNA, and



purify the *in-vivo* fixed chromatin fragments from many other protein-complexes.

### 3.1 Designing a new approach for capturing chromatin

To identify the global protein composition of chromatin in an unbiased way, I developed SICAP. To do that, first of all I treated mouse ES cells with formaldehyde 1.5% (v/v) for 15min to cross-link proteins to DNA essentially like in the established ChIP protocols [118]. To exclude proteins cross-linked via RNA, I added RNase A to the samples. Next I treated the sample with terminal deoxynucleotidyltransferase (TdT) in the presence of biotin-ddUTP in order to specifically label free 3'-ends of DNA. Chromatin fragments were captured using streptavidin beads. Then the beads were washed extensively by SDS 1%, NaCl 2M, iso-propanol 20% (v/v) and Acetonitrile 50% (v/v) in water. Finally, to break the crosslinks the samples were boiled (fig. 3.1). As the negative control, essentially the same procedure was carried out just without adding biotin-ddUTP, hence, chromatin fragments were not labeled. Finally, the proteins were loaded on SDS-PAGE to be compared with the negative control. As a result, it turned out that SICAP enriches for proteins with very low background from non-specific binding of the proteins to the beads (fig. 3.2A). This was further investigated by quantitative mass spectrometry to identify proteins enriched relative to the negative control using SICAP.



**Figure 3.2:** Comparing SICAP with the negative control. A) SDS-PAGE shows the proteins enriched by SICAP procedure with and without biotin-ddUTP. B) The efficiency of recovering DNA for three independent replicates relative to the input was calculated using qPCR. C) The fluorescent microscopy of the cells using streptavidin-Alexa488 indicates. In the first row: using TdT and biotin-ddUTP the nuclei are labeled; but neither cytoplasm nor nucleoli. In the second row: as a negative control, biotin-nucleotides were not added to the cells. In the third row: using Poly-U polymerase and biotin-UTP in addition to the nuclei, cytoplasm and nucleoli are stained. The white arrows indicate nucleoli, which shows a strong signal using RNA-labeling.

## 3.2 Evaluating the specificity and efficiency of chromatin purification by SICAP

Although TdT adds nucleotide specifically to DNA (not RNA), to be more confident about the specificity and efficiency of DNA-labeling I utilized a microscopy assay. Briefly the cells were fixed, and then they were treated with DNase I to produce nicks in DNA. Subsequently the cells were treated with TdT and biotin-ddUTP. Then the cells were exposed to Streptavidin Alexa-fluor 488.

Fluorescent microscopy clearly demonstrated efficient and specific staining of nuclei, but not cytoplasm nor nucleoli (fig. 3.2C, top row). Conversely to see the difference of RNA-labeling with DNA-labeling, I replaced TdT with Ecoli polyU polymerase; in addition, I replaced biotin-ddUTP with biotin-UTP. As a result, this time I observed cytoplasmic signal, and the intensive staining of nucleoli (fig. 3.2C, bottom row). Although in this experiment I used DNase I to produce nicks in DNA, in SICAP I omitted this step to avoid biases due to DNase I hypersensitive sites. In SICAP, chromatin is chopped into small fragments by sonication, hence, chromatin fragments are labeled by TdT. Therefore, there is no need to make nicks by DNase I. In addition, I checked the efficiency of DNA recovery, which has turned out to be about 3-5% relative to the input DNA fig. 3.2B. Thus using SICAP the *in-vivo* fixed DNA-protein complexes are efficiently and specifically enriched such that only with 4 million cells I identified many chromatin proteins; while, the most recent study has used 100 million cells [20] for chromatin enrichment for proteomics (ChEP).

## 3.3 Identification of chromatin composition in mouse embryonic stem cells

Although I loaded the SICAP samples on SDS-PAGE to visually compare it with the negative control (fig. 3.2A), proteins purified by SICAP

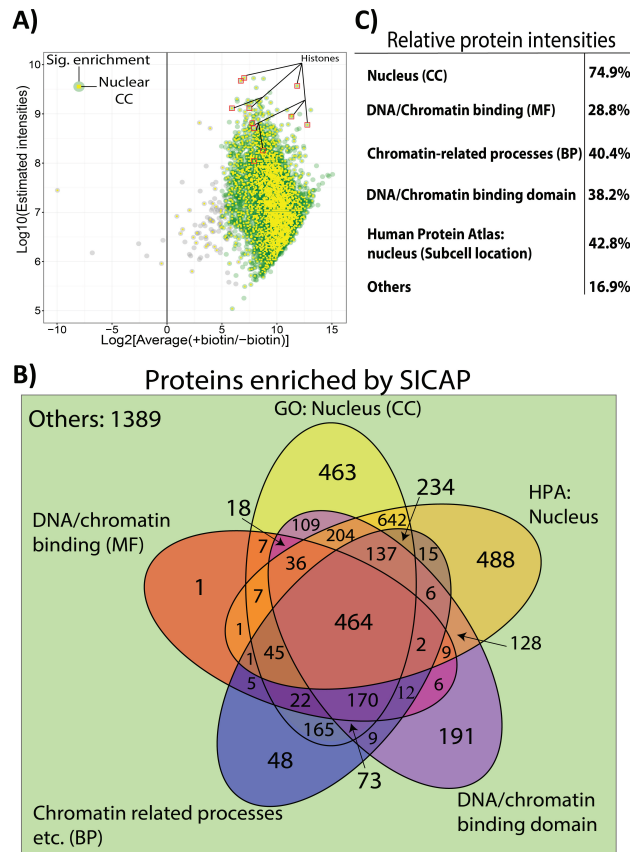
are free of detergents. Therefore, after tryptic digestion and desalting, the peptides can be injected to the mass spectrometer. In order to decrease the complexity of the sample, I divided the peptides into 10 fractions using high pH reverse phase liquid chromatography (pH = 10). As a result, after protein digestion and mass spectrometric analysis 5231 proteins were identified using two independent replicates, 5106 of which were significantly (adj. p-value  $\leq 0.05$ ) enriched over the negative control (fig. 3.3A and Supplementary Table 3.1) as determined by a moderated t-test using the limma package [119]. Once again this result confirms very low non-specific background proteins by SICAP.

To assess whether prior evidence may link these 5106 proteins to chromatin, I referred to Gene Ontology (GO) as well as the presence of DNA or chromatin binding domains. The GO database provides evidence for nuclear localization of 2794 proteins (fig. 3.3B) while 806 proteins have been annotated as chromatin or DNA-binders. These include many transcription factors known to be involved in stem cell maintenance such as Oct4, Sox2, Nanog, Klf4, Tcf3, Vps72, Stat3, Esrrb, Sall4, Lin28, Rbpj, Tbx3, Notch1, B-Catenin and Nkap.

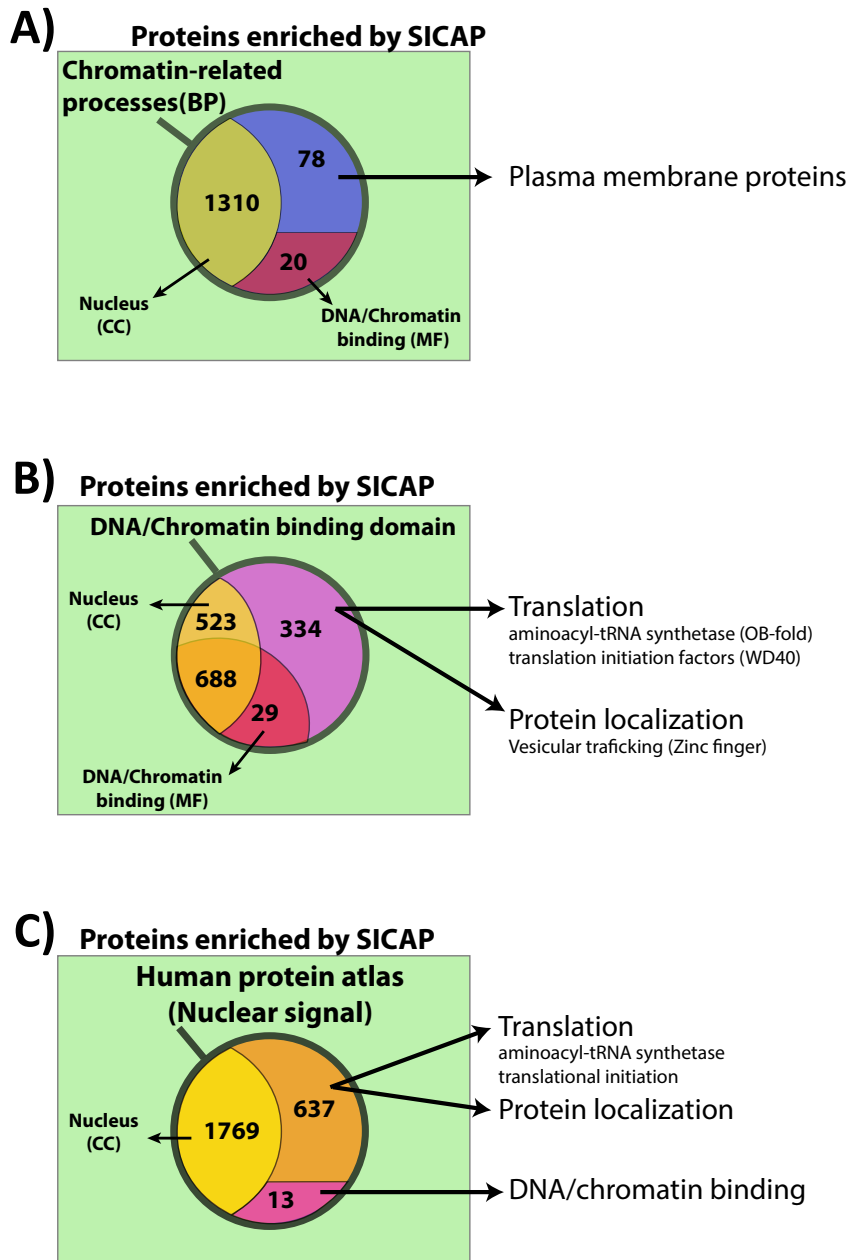
Moreover 1407 proteins are involved in biological processes related to chromatin functionality, including transcriptional regulators such as RNA polymerase II subunits as well as many components of the mediator complex. Some plasma membrane proteins would not be expected to be present in this group (i.e. Arf4, Atp2b4, Lims1, Rab23 etc. fig. 3.4A and Supplementary Table 3.1). Interestingly some of these membrane-proteins (i.e. E-cadherin [120], Dab2ip [121], Mavs [122], Rab23 [123] and Tlr2 [124]) have been shown to be involved in nuclear import and might thus interact with chromatin in the process.

Additionally, I found 1574 proteins with DNA or chromatin binding domains, most of which have indeed been annotated with GO-terms related to nuclear localization or as DNA or chromatin-binder proteins, with the exception of 334 proteins (fig. 3.4B). Gene enrichment analysis of these 334 proteins indicated that 25 proteins are involved in translation (i.e. proteins with canonical aminoacyl-tRNA ligase activity or translation initiation factor activity). Interestingly, aminoacyl-tRNA

synthetases have been reported in the nuclei as a high molecular weight complex [125]. Furthermore 28 proteins are involved in protein transport with DNA/chromatin-binding domains (i.e. proteins involved in vesicle-mediated transport proteins such as Sec23a, Sec23b, Sec24a, Sec24c, Sec24d, Sec31a with zinc finger domains).



**Figure 3.3:** Identification of the enriched proteins by SICAP. A) Significantly enriched proteins by SICAP (adj. p-value  $\leq 0.05$ ) and the background proteins were stained by green and gray dots, respectively. The yellow dot shows nuclear localization based on CC of GO. Histones have been shown by an additional red rectangle B) the enriched proteins were categorized based on CC = Nucleus (Yellow), MF = DNA or Chromatin binding (Red), BP = Transcriptional-regulation, DNA repair, DNA replication and chromatin modification (Chromatin-related processes, Blue), DNA or chromatin binding domain (Purple) and HPA sub-cellular localization = Nucleus (Orange). C) the relative intensities of the proteins have been shown for each of the aforementioned categories.

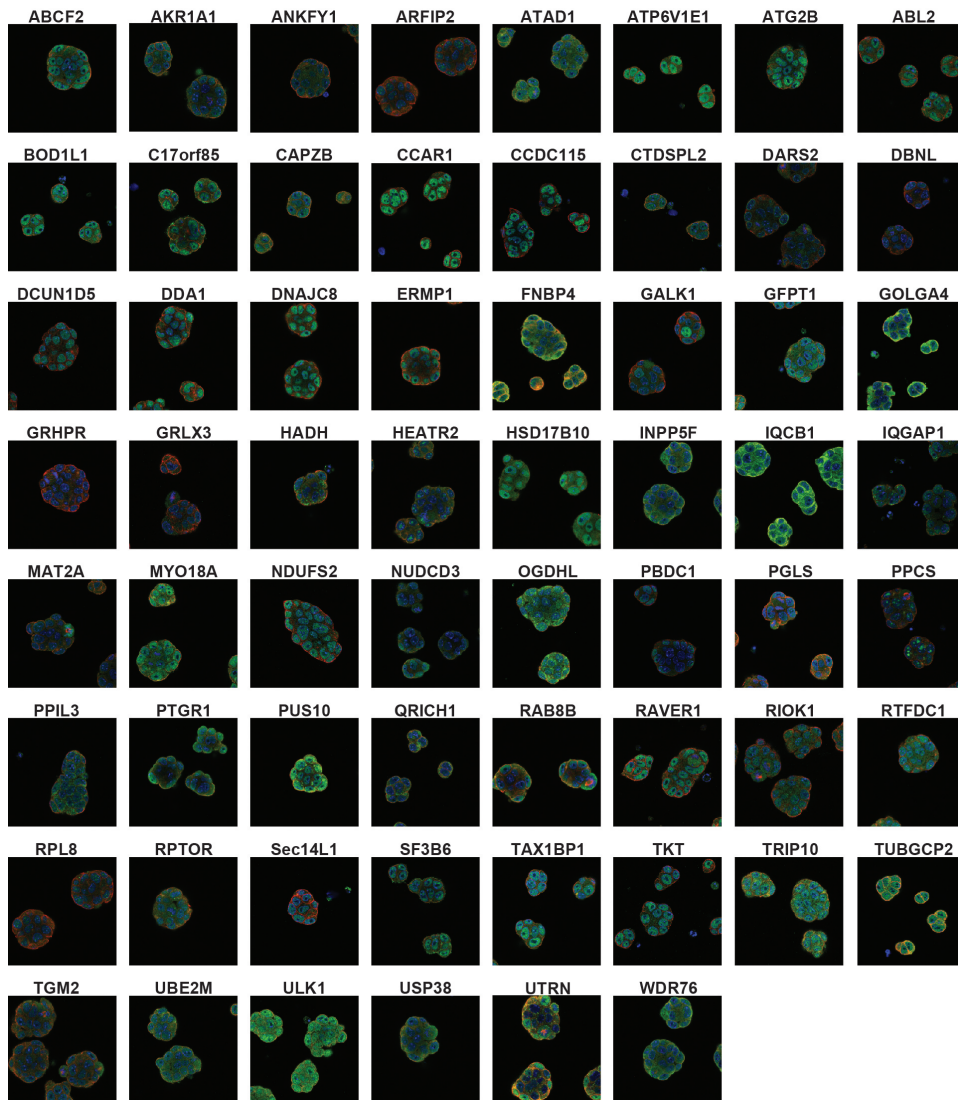


**Figure 3.4:** Unexpected proteins enriched by SICAP. A) The Venn diagram shows proteins annotated by GO as transcriptional-regulation, most of which have nuclear localization, except some plasma membrane proteins. B) The Venn diagram shows proteins with DNA or chromatin binding domains, most of which have nuclear localization or they are in fact chromatin-binders. Some proteins involved in translation or protein-localization do not have nuclear localization based on GO. C) The Venn diagram shows proteins with nuclear signal based on HPA database, most of which overlap with GO. However, based on HPA some proteins involved in translation or protein-localization have nuclear localization.

In order to check the localization of these proteins I consulted the Human Protein Atlas (HPA [126]). Mapping all 5106 SICAP-enriched proteins to HPA revealed nuclear localization for at least 2400 proteins in our list, among which 650 proteins that had not been annotated as a nuclear protein based on GO database (fig. 3.4C). Interestingly, this included Sec23a, Sec24b and Copa (involved in vesicular-trafficking), Cars2, Ears2, Lars2 (amino-acyl tRNA synthetases) and Eif2b2, Eif2b4, Eif4h (translation initiation factors) thus supporting our SICAP data. It should be noted that protein annotation based on staining in the human cancer cell lines used in HPA may not be taken as a direct proof for nuclear localization in the mouse ES cells used in our study. Conversely, absence of such localization in HPA does not disprove our SICAP results.

Importantly, by considering the relative intensities of the enriched proteins about 75% of them are localized in the nucleus, while, 17% of the enriched proteins have no known relevance to the chromatin (fig. 3.3C). To more directly correlate SICAP-derived chromatin-interaction to nuclear localization in mouse ES cells I selected 132 proteins from our SICAP data without prior GO annotations related to nuclear localization or functionality. Immuno-fluorescence microscopy using HPA-validated antibodies revealed nuclear localization of 62 proteins (fig. 3.5, and Supplementary Table 3.2) including proteins involved in vesicular trafficking (i.e. Sec14L1, Rab8a, Ankfy1), as well as translation (i.e. Rpl8).

It is noteworthy that detecting nuclear localization by immuno-fluorescent and chromatin-binding by SICAP are not equivalent; but complementary evidence. In fact, a nuclear protein may or may not bind to chromatin. For example, p53 is a nuclear protein, however, it binds to DNA only once it is activated. On the other hand, using immuno-fluorescent microscopy we may see a strong plasma membrane signal for E-cadherin. Nevertheless, it does not disprove the chromatin binding of the protein [127].



**Figure 3.5:** Immuno-fluorescent staining of some proteins enriched by SICAP with previously unknown nuclear localization using HPA antibodies. Blue shows nuclei (DAPI staining), Red shows microtubules, and green shows the target proteins. 62 out of 134 proteins clearly show nuclear localization. For further details, please refer to the Supplementary Table 3.2. These immuno-staining experiments were carried out by Dr. Peter Thul, KTH Royal Institute of Technology.

Altogether, SICAP has indicated that chromatin composition is much more diverse than previously expected. Thanks to the high specificity of TdT in labeling DNA, the ability to eliminate contaminating proteins by extensive washing, and comparing with a non-biotinylated negative control, SICAP is a reliable approach.

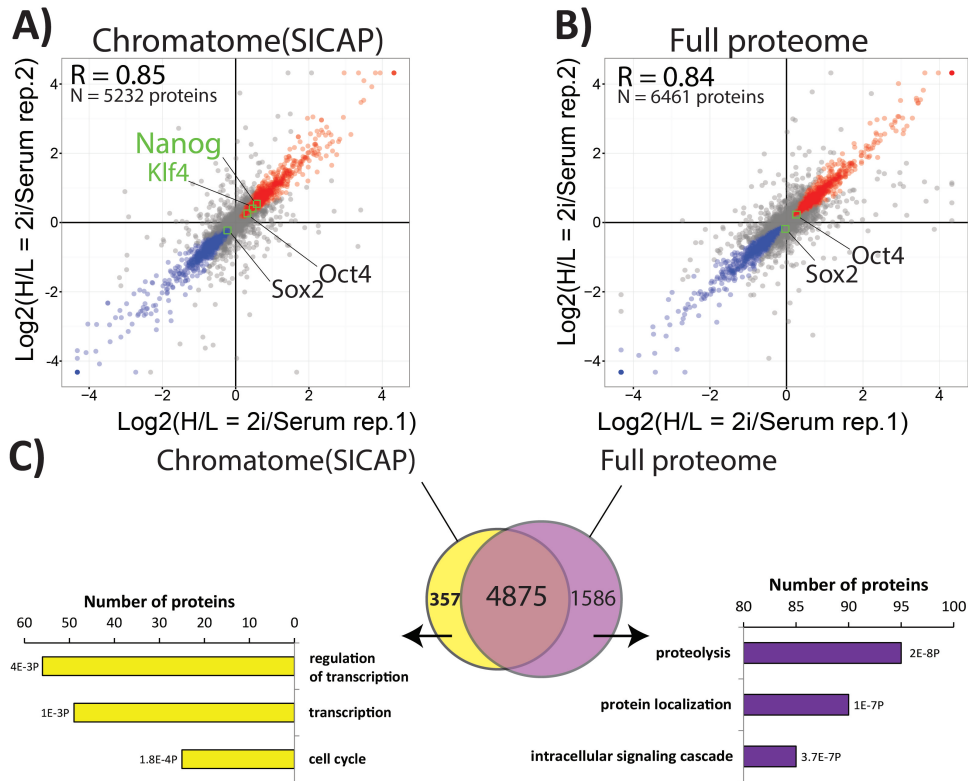


### 3.4 Comparing chromatin protein composition between the ground-state and the metastable state of pluripotency

To achieve a new cellular state, chromatin should recruit some proteins in order to regulate the expression of a specific set of genes. To identify the recruited/dismissed chromatin proteins in a high-throughput manner I used SICAP to quantitatively compare the chromatin composition of ES cells in the ground-state (2i conditions) and the meta-stable state of pluripotency (serum conditions) after labeling cells with heavy and light SILAC amino acids, respectively. In addition, I obtained full proteome data from the same cells to investigate the correlation between changes in protein levels and chromatin localization. The SICAP and total proteome experiments identified 5232 and 6461 proteins, respectively, each quantified with a Pearson correlation of 0.85 between biological replicates indicating high reproducibility (fig. 3.6A, B and Supplementary Table 3.3). Qualitative comparison of SICAP and proteome datasets indicate that 357 proteins were only identified using SICAP (fig. 3.6C) such as Nanog and Klf4 (fig. 3.6A, B). This may be readily explained by the higher complexity of the full proteome and the enrichment of otherwise low abundance chromatin-bound factors via SICAP, rendering them accessible for mass spectrometric detection. These proteins mainly represent processes related to transcription (fig. 3.6C) as may be expected for a chromatin-enriched pool of proteins.

Comparing the chromatome of the two cell states using SICAP indicates about 1518 and 1521 proteins differentially associate with chromatin in 2i and serum media, respectively (adj.  $p \leq 0.05$ , fig. 3.6A). Comparing the full proteomes of the two cellular conditions reveals about 1500 differentially expressed proteins in each of the media (adj.  $p \leq 0.05$ , fig. 3.6B); however, GO analysis of proteins higher in 2i medium indicated that the SICAP data is more enriched for chromatin-related biological processes in comparison to the full proteome and the transcriptome data by Marks et. al. [95] (fig. 3.7A, and Supplementary

Table 3.3).



**Figure 3.6:** Comparing 2i and serum condition of the ES cells by SICAP and full proteome. A and B) proteins significantly higher in 2i and serum were stained in red and blue, respectively ( $p$ -value  $\leq 0.05$ ). Proteins without significant change were stained in gray. C) The Venn diagram shows proteins identified only by SICAP and the full proteome, which are significantly enriched in transcriptional-regulation and proteolysis, respectively. P: p-value of the enrichments, R: Pearson Regression

In addition, SICAP data were enriched in processes related to (m)RNA processing, suggesting the capture of proteins involved in co-transcriptional regulation at chromatin. In serum medium the most significant enriched genes using SICAP are involved in non-coding RNA metabolism and translation (fig. 3.7B, and supplementary table 3.3). In fact, rRNA transcription and maturation occur co-transcriptionally in the nucleolus [128]; in addition, tRNA aminoacylation [125] and translation [129, 130] have been reported in the nucleus. Given the fact that translation-related genes are not up-regulated according to

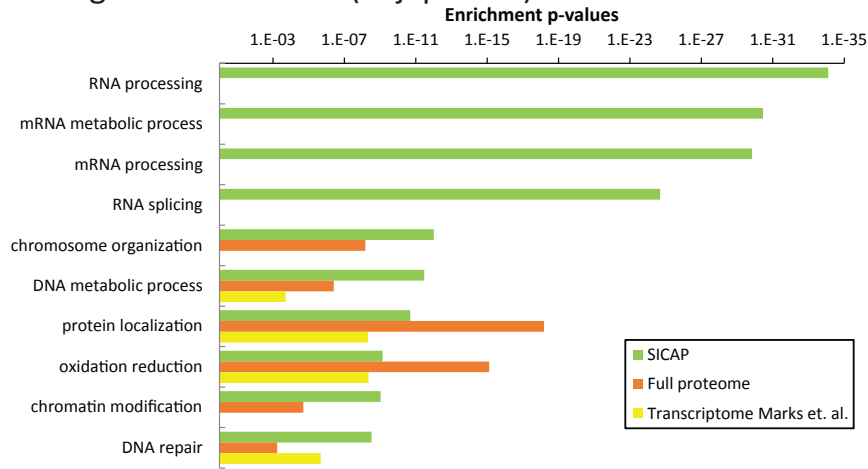
the full proteome comparison, their differential binding to chromatin in serum condition suggests other non-canonical functions that remain to be determined.

When directly comparing full proteome to SICAP data, changes in protein expression levels between 2i and serum conditions showed a reasonably high correlation with differences in chromatin association ( $R=0.58$ , fig. 3.8A). For instance, Dnmt3a and Dnmt3b were among the strongest enriched proteins in serum condition in both data sets, suggesting global DNA methylation in the meta-stable state. Additionally, only SICAP data indicate that Uhrf1, required for the maintenance of 5mC significantly binds to chromatin in serum condition. This is in line with bisulfite sequencing results by Habibi et. al [24] and with the current model of decreased DNA methylation and enhanced transcription in 2i cells.

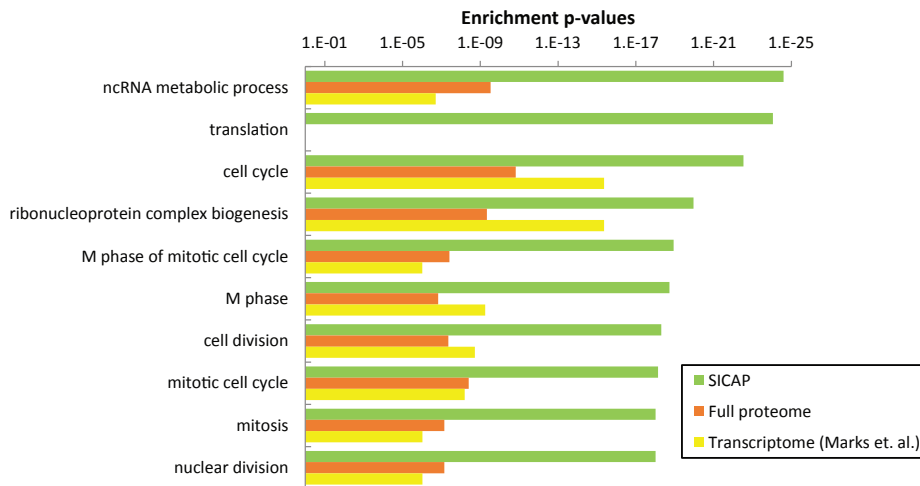
Among the proteins with the strongest increase in chromatin-binding and overall expression in 2i cells we observed an uncharacterized protein Gm13128/Pramef17. Interestingly Gm13128/Pramef17 is highly similar (70% identical) to Pramel7, which has been identified for its important role in the maintenance of pluripotency in the ES cells acting downstream of LIF-Stat3 signaling [131]. PRAME was first identified as an antigen expressed in human melanomas [132]. Interestingly members of PRAME family, which contain Leucine-rich repeats, are expressed preferentially in gametogenic tissues and in tumors [133]. Our observation that Pramef17 is highly expressed and binds to chromatin in 2i medium may indicate an unknown chromatin-related function of the PRAME family in pluripotency, as well.

Surprisingly we observed Fgf4 among the prominent up-regulated chromatin binders in the ground-state of pluripotency (fig. 3.8A). This is highly unexpected since Fgf4 is a growth factor not known to associate with chromatin. Interestingly, immuno-fluorescent staining using a monoclonal antibody indeed confirmed strong nuclear localization of Fgf4 in addition to staining in the plasma membrane as expected for a secretory protein (fig. 3.8B).

**A) Genes higher in 2iL-media (adj.  $p < 0.05$ )**



**B) Genes higher in serum-media (adj.  $p < 0.05$ )**

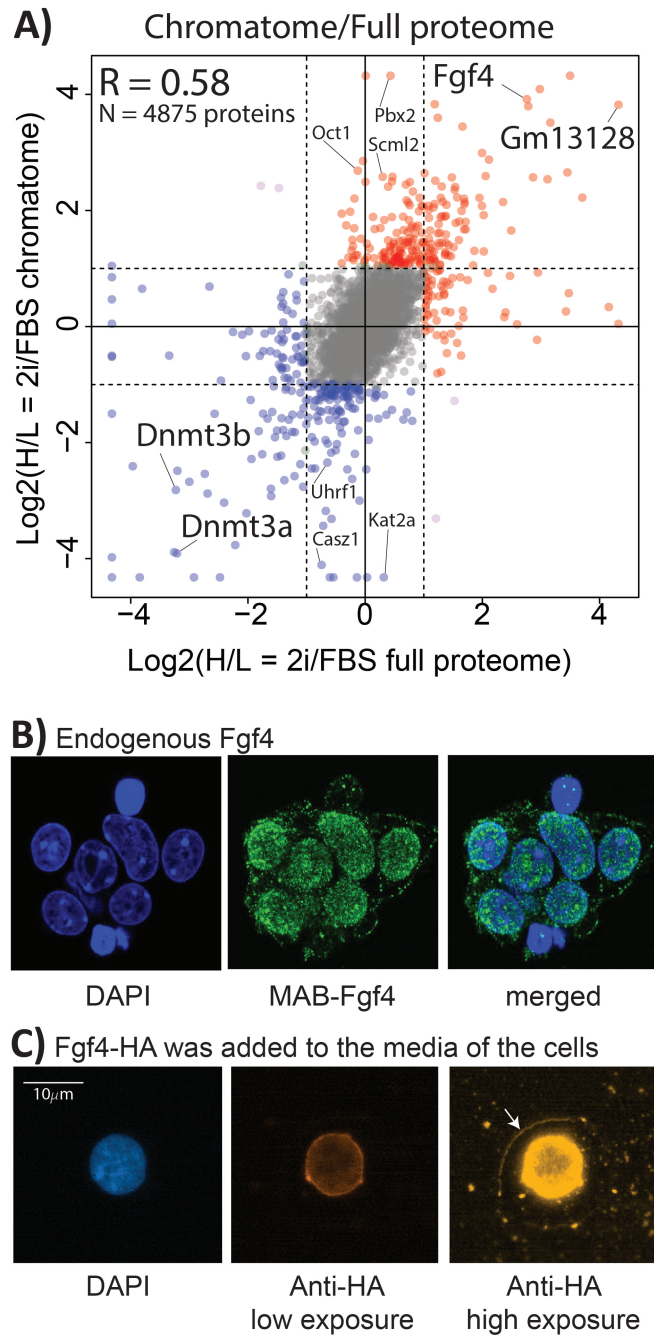


**Figure 3.7:** Enrichment analysis of the biological processes for the significant genes in A) 2i medium, and B) serum medium using SICAP (green), full proteome (orange) and transcriptome (yellow, data produced by Marks et. al. [95])

To further investigate whether ES cells possess a mechanism to transport Fgf4 to the nucleus we generated a stable HEK cell line expressing HA-tagged Fgf4. Treating ES cells with the filtered conditioned medium followed by immuno-fluorescent staining using an HA-antibody revealed nuclear localization of Fgf4 mainly in cells outside the ES cell colony, again with additional staining in the plasma membrane (fig. 3.8C). This indicates that secreted Fgf4 is able to move

into the nucleus in a paracrine fashion. Interestingly, Fgf4 is known to induce differentiation by a pathway that is inhibited by the combined effects of the Mek2 and Gsk3b (i.e. 2i) inhibitors. It remains to be determined if nuclear localization provides a mechanism to silence the differentiation-promoting function of Fgf4.

Although changes in chromatin interaction in 2i versus serum conditions closely follow overall expression for most proteins, about 235 proteins showed a significant difference in chromatin binding without an appreciable change in their total abundance (fig. 3.8A). Such a pattern indicates dynamic association with chromatin controlled by other mechanisms than translational regulation. For instance, the transcriptional regulators Oct1 and Pbx2 show an 8 to 16-fold increase in chromatin-binding in 2i condition, without a change in their total quantity (Supplementary Table 3.3). Another interesting example is Scml2, a polycomb repressive complex 1 (PRC1) associated protein, that binds to chromatin globally in 2i condition. It seems that Scml2-binding to chromatin occurs independently of PRC1 and PRC2, since the other members of these complexes do not show significant changes in expression or chromatin binding. In serum conditions, the transcription factor Casz1 and histone acetyltransferase Kat2A are highly enriched in chromatin. The fact that these prominent examples all concern bona fide chromatin components with established roles in transcriptional regulation suggest that they may fulfill important functions in sustaining ground or meta-stable states of pluripotency. Shuttling to chromatin independent of altering their expression level may provide a mechanism to do so in a fast manner.



**Figure 3.8:** Quantitative comparison between the SICAP and the full proteome dynamics. A) The plot shows full proteome and SICAP ratios between 2i to serum on the x and y-axis, respectively. There are 235 proteins without significant change in their total quantity while their chromatin-binding is significant. B) Immuno-staining by monoclonal antibody (MAB) against Fgf4 indicates nuclear localization of Fgf4 in the ESCs. C) Fgf4 fused with the HA-tag was added to the medium of the ES cells, and then immuno-staining was carried out using anti-HA antibody. As a result, nuclear localization of the tagged protein was observed. The white arrow indicates the ligand probably in the cell membrane.

# 4

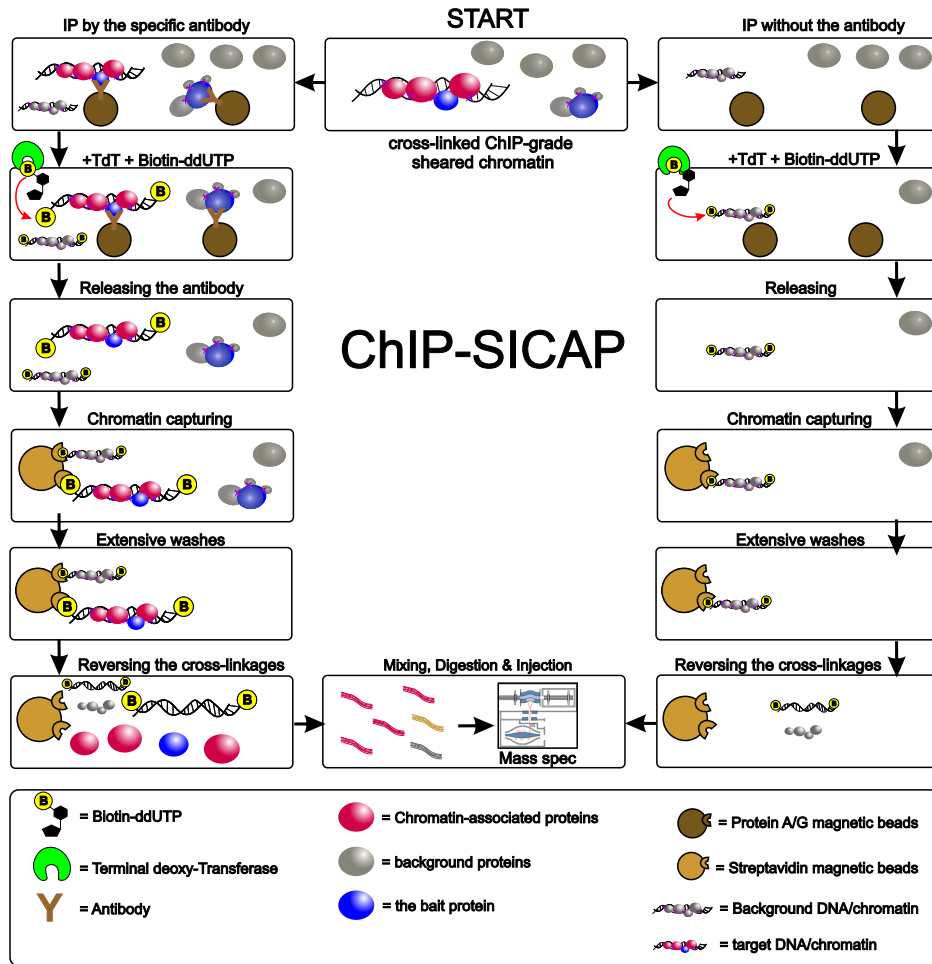
## ChIP-SICAP: Regional View of Chromatin

---

Chromatin immuno-precipitation (ChIP, [134, 135]) is a powerful method for finding the genome-wide identification of binding sites of a given transcription factor. At the end of this procedure, the enriched DNA fragments are identified by qPCR, microarray chips [136], or high-throughput sequencing [137]. Therefore, the antibody, contaminant proteins or non-chromatin-bound proteins do not interfere with the final read-out, as they are removed during the procedure of DNA purification. However, injecting the ChIP samples into the mass spectrometer usually does not produce reliable results due to contaminating proteins. In fact, many non-specific proteins such as cytoplasmic proteins may stick to the beads, the plastic tubes or even DNA. Using gentle washing steps applied to the ChIP procedure it is hard to remove the contaminant proteins, and as a result the final sample is not clean enough to demonstrate reproducible and reliable results. In addition, the target protein of the ChIP is not always attached to the chromatin, hence, it may carry over many non-chromatin-bound interactants.

Hence I developed a solution for double purification and excluding the non-chromatin-bound proteins by combining ChIP and SICAP (ChIP-SICAP, fig. 4.1). Using this method, I studied proteins in the vicinity of Oct4, Sox2 and Nanog (hereafter OSN) on chromatin in order to identify novel proteins connected to the core-circuitry of mouse ESCs in 2i and serum media. In other words, using ChIP-SICAP we can study proteins on small fragments of DNA with the target protein, which I call co-localized proteins on chromatin. It is noteworthy that co-localized proteins (identified by ChIP-SICAP) do not necessarily interact directly with the bait protein, but they are in close vicinity to

the bait protein on the same DNA fragment (in average about 200-bp). Therefore, it is conceivable that they have functional cross talk.



**Figure 4.1:** Schematic representation of ChIP-SICAP. For further details, please refer to the text (section 4.1).

Here I identified 408 proteins, as the result of OSN ChIP-SICAP. These 408 proteins include a large number of established interaction partners of the target proteins known to participate in the core pluripotency network (e.g. Rex1, Prdm14, Tcf3, Sall4, Esrrb, Tbx3, Stat3 etc). Interestingly, it turned out that many of these regulators of pluripotency do not show such a significant change in their total quantity, however, their chromatin bindings are significantly different in 2i and



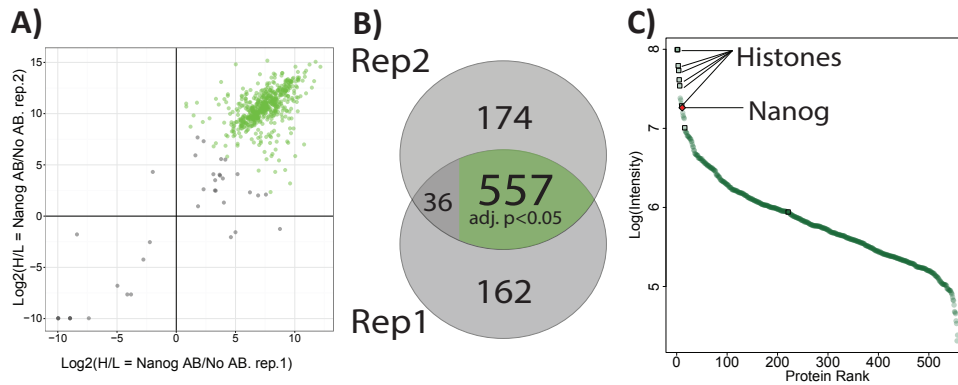
serum media. In addition, I identified many proteins as novel components of the core-circuitry. Therefore, I selected Trim24, as a protein remained undetected by the previous affinity capture methods, to verify the ChIP-SICAP result. Interestingly it turned out that Trim24 binds to OSN on many critical genomic loci such as *Nanog*, *Prdm14*, *Tbx3* and *Sox2* enhancers preferentially in 2i medium.

Then I focused on Suz12 and the co-localized proteins on chromatin. In fact, polycomb group (PcG) proteins have important roles during development. Specifically, polycomb repressive group 2 (PRC2), which Suz12 is a member of, represses transcription via di- and trimethylation of lysine K27 on histone H3 (H3K27me<sub>2/3</sub>). As a result of applying Suz12 ChIP-SICAP, the entire PRC2 plus many known interactors of Suz12 were identified.

## 4.1 Experimental design and proof-of-principle

I first carried out ChIP-SICAP using a *Nanog* antibody in ES cells cultured with heavy-SILAC 2i medium compared to a negative control cultured with light-SILAC 2i medium. In the negative control I excluded the *Nanog* antibody, however, I included the DNA biotinylation step. As a result, 640 proteins were identified, out of which 557 proteins were significantly higher in the *Nanog* antibody-treated sample over the negative control (adj. p-value  $\leq 0.05$ , fig. 4.2A, B and Supplementary Table 4.1).

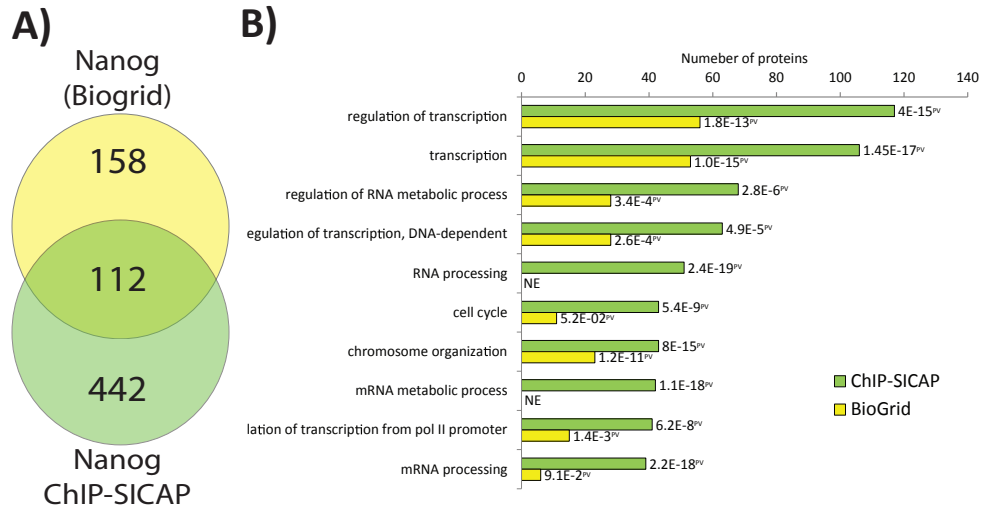
Ranking the identified proteins based on their estimated intensities indicates that histones are the most abundant proteins, followed by *Nanog* and its (potential) interaction partners (fig. 4.2C). This observation alone leads to two important conclusions, namely i) that the antibody has pulled-down the expected target, thereby independently validating the antibody, and ii) that *Nanog* co-purifies all nucleosome histones proving that it binds to chromatin. This is in contrast to previous *Nanog* affinity-capture assays (BioGrid), which were unable



**Figure 4.2:** Nanog ChIP-SICAP in comparison to the no-antibody control. A) The scatterplot shows proteins non-significant and significantly enriched proteins (adj.  $p$ -value  $\leq 0.05$ ) over the no-antibody control by gray and green color, respectively. B) The Venn diagram shows the overlap of the two replicates. C) The plot shows the estimated intensities of the significant proteins and their ranking from top to down.

to identify the histones. In fact, using co-immunoprecipitation chromatin is sedimented to avoid its binding to the beads, and then soluble interactants are pulled down.

Of the 557 identified proteins, 112 overlapped with the 270 proteins previously identified to interact with Nanog (as deposited in BioGrid) (fig. 4.3A). Comparing the outersects of the datasets (442 and 158 proteins from ChIP-SICAP and the previous affinity captures, respectively) indicates that Regulation of transcription is the most significant biological process (BP) that has been identified in both protein groups (fig. 4.3B). Nevertheless, using ChIP-SICAP the number of proteins associated with the BP term is higher, and the  $p$ -value of the enrichment is also less. The relatively small intersect (20%) indicates that the scope of chromatin-associated proteins identified by ChIP-SICAP is quite different to that of the soluble interactants found by other methods. In fact, without DNA-labeling it is hard to distinguish between these two sets of networks.



**Figure 4.3:** Comparing ChIP-SICAP with the previous affinity purification methods for Nanog. A) The Venn diagram shows the overlap of Nanog ChIP-SICAP (green) with the previous affinity purification methods (yellow, reference: BioGrid). B) Comparing the enriched biological processes between the outersects. NE = not enriched, PV = p-value of the enrichment.

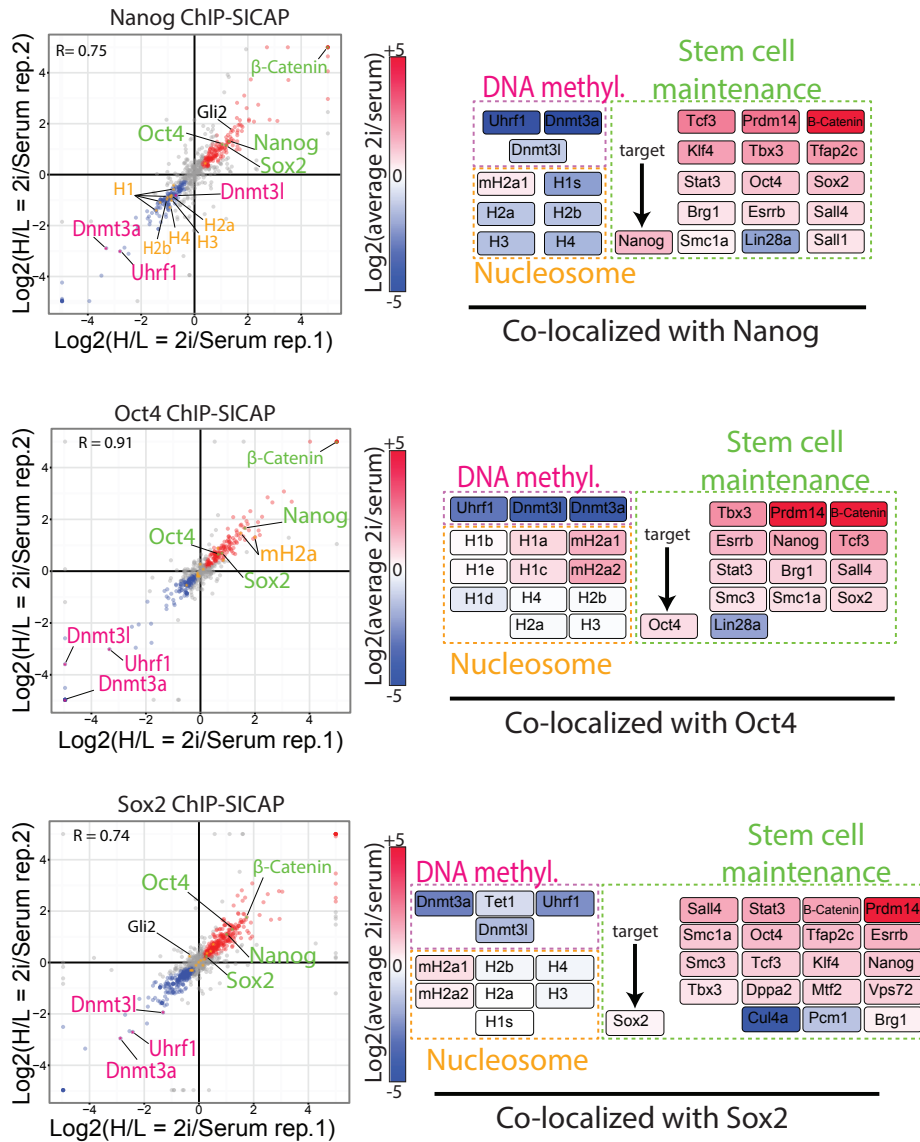
## 4.2 Comparing Oct4, Sox2 and Nanog co-localized proteins on chromatin between 2i and serum media

Being confident of detecting only a few background proteins in the no-antibody control, I next compared proteins co-localized with Nanog, Oct4 and Sox2 using ESCs grown in serum (light medium) and 2i medium (heavy SILAC). In these experiments heavy and light cells were combined from the beginning of the assay, therefore chromatin fragments of both samples were simultaneously labeled by biotin-ddUTP.

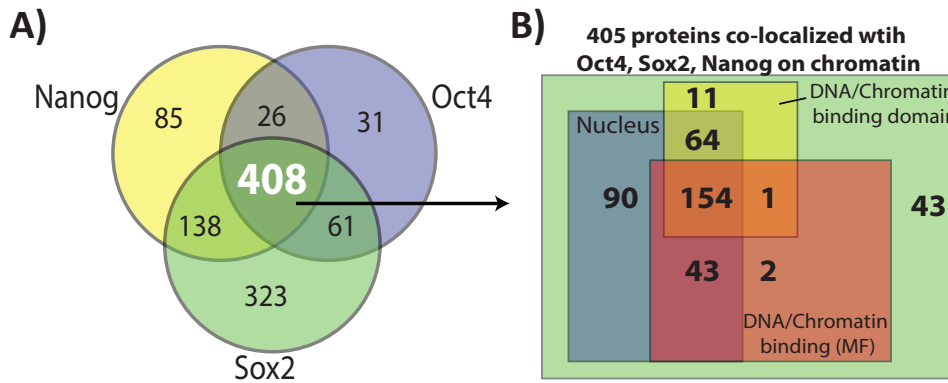
Firstly, I performed ChIP-SICAP using Nanog. I detected 650 proteins, out of which about 300 proteins were significantly different between 2i and FBS conditions (adj. p-value  $\leq 0.1$ , fig. 4.4A, and Supplementary Table 4.2). Interestingly using Nanog ChIP-SICAP, B-catenin has the highest fold-change of chromatin-binding in 2i compared to serum condition ( $>20$ -fold difference). This is expected from the fact

that one of the kinase inhibitors in 2i medium is Gsk3b inhibitor (CHIR99021), resulting in activation of Wnt signaling and translocation of B-catenin to the nucleus co-localizing with Nanog on chromatin. We also detected Oct4 and Sox2 in association with Nanog binding sites, which is consistent with the tight interconnection of the three factors in the proposed model for the core transcriptional circuitry of the ES cells [138]. Additionally, we observed several factors related to stem cell maintenance that differentially associate with Nanog binding sites in 2i medium, including Prdm14, Tfc2l1, Rex1(Zfp42), Tcf3 (Tcf7l1), Sall4, Esrrb, Tbx3, Stat3, Klf4, Tfap2c, Smarca4 (Brg1) (Figure 4.4A). Indeed, prior studies have shown the important roles of these proteins in either stem maintenance or reprogramming to pluripotent stem cells.

Remarkably, in serum condition we observed all nucleosome components including histones H2A, H2B, H3, H4 and multiple histone H1 variants with significantly higher ratios compared to 2i. Therefore, it seems that DNA is more accessible to Nanog in 2i medium, while DNA is relatively more packed around Nanog-binding sites in serum condition. This indicates that ChIP-SICAP can inform on global chromatin structure. Finally, Nanog preferentially interacts with DNA methylases (Dnmt3a, Dnmt3l, Uhrf1) under serum conditions. Extending these experiments to Oct4 and Sox2 as entry-points for differential ChIP-SICAP between 2i and serum conditions essentially reproduced the results for Nanog, with subtle differences (fig. 4.4B, C). Importantly, each experiment identified all three master TFs, confirming the close connection between OSN in the pluripotency network. Like Nanog, ChIP-SICAP using Oct4 and Sox2 also shows a high ratio of B-Catenin in 2i medium, although with a less extreme ratio for Sox2. Interestingly most of the nucleosome components did not show a significant change using Oct4 and Sox2 ChIP-SICAP, however, macroH2A1 and macroH2A2 are surprisingly higher for Oct4 in 2i condition. Indeed, macroH2As are transcriptionally-suppressive variants of H2A [139-141] therefore, it seems that in 2i condition some of the Oct4 targets are repressed by recruiting macroH2A.



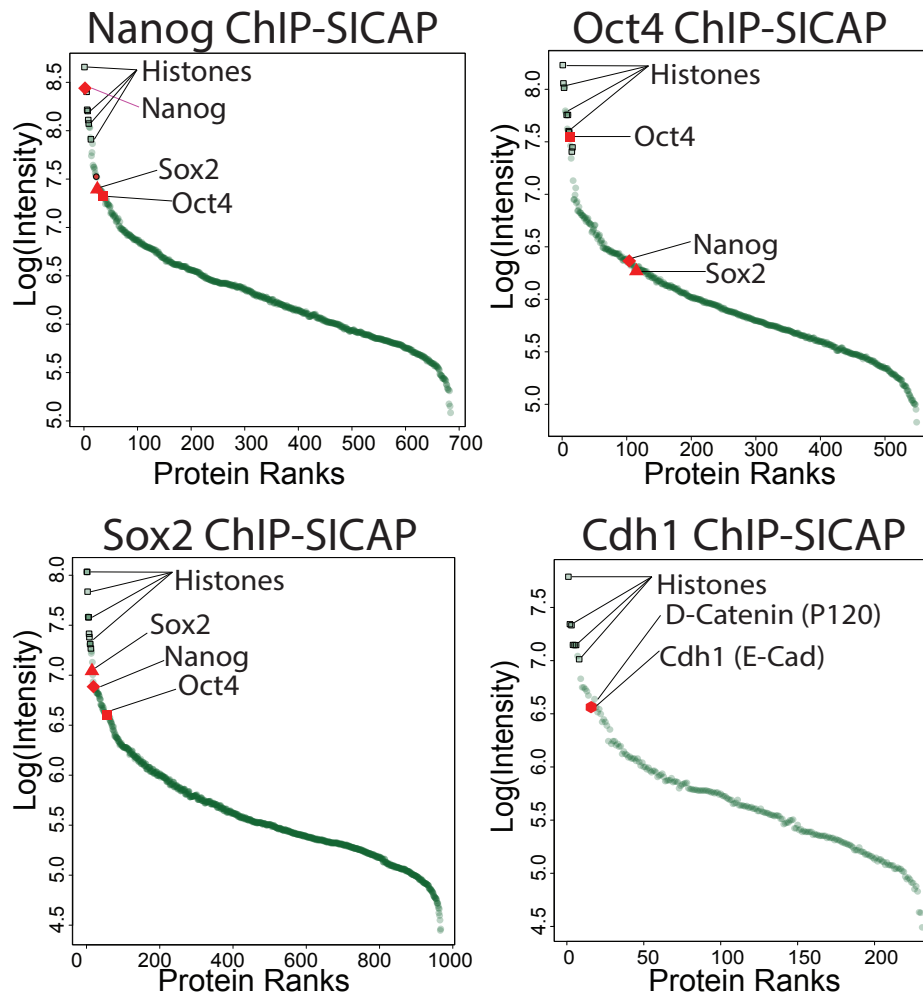
**Figure 4.4:** Comparative ChIP-SICAP between 2i and serum conditions using Nanog, Oct4 and Sox2 as the target (bait). Proteins significantly higher in 2i, serum and non-significant proteins are red, blue and gray, respectively (adj. p-value  $\leq 0.1$ ).



**Figure 4.5:** Proteins co-localized with Oct4, Sox2 and Nanog on chromatin. A) The Venn diagram shows the overlap of Oct4, Sox2 and Nanog ChIP-SICAP. B) Categorization of the 408 proteins based on GO terms and protein domains.

### 4.3 Studying the core-transcriptional circuitry of mouse embryonic stem cells

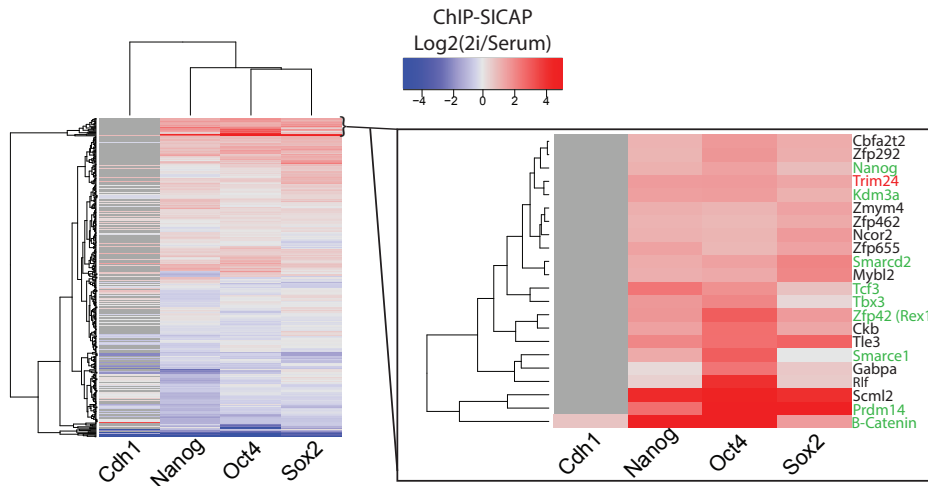
Here I reasoned that proteins identified repeatedly using Oct4, Sox2 and Nanog ChIP-SICAP are most likely connected to the core-circuitry. Therefore, after subtracting the potential false-positives identified in a no-antibody ChIP-SICAP in ES cells in 2i and serum media (primarily endogenously biotinylated proteins such as pyruvate carboxylase, Mccc1, Mccc2 and so on), I identified 408 proteins in the overlap among the three ChIP-SICAP experiments (fig. 4.5A). This represents the large majority of the proteins in the individual experiments (77% of Oct4, 44% of Sox2 and 62% of Nanog 70-80%) indicating the tight interaction within the network. GO annotation of these 408 proteins indicate that 86% (351 proteins) have a nuclear localization while 49% (200 proteins) have a DNA-binding or chromatin-binding function, indicating the enrichment of proteins with the expected functionality. Furthermore, about 10% of the proteins (43 proteins) have no annotation relating them to nucleus, DNA-binding, chromatin-binding or the relevant domains (fig. 4.5B).



**Figure 4.6:** Ranking the proteins based on their estimated intensities. Each ChIP-SICAP experiment indicates the target protein on top of the intensities, after the histones. Using Cdh1(E-Cadherin) neither Oct4, Sox2, Nanog nor many other proteins were identified, and vice versa.

Because of the very high overlap between the OSN ChIP-SICAP experiments we aimed to verify that results were not accidental by carrying out a ChIP-SICAP experiment using E-Cadherin (Cdh1) as a protein foreign to the pluripotency network. In fact, E-cadherin is a plasma membrane protein rather than a canonical chromatin-binder, however, it can be cleaved by gamma-secretase, gamma-secretase and caspase-3 releasing specific C-terminal fragments (E-Cad/CTFs) 1, 2 and 3, respectively, which are able to translocate into the nucleus and

bind to chromatin [142]. As expected, ranking of identified proteins in the Cdh1 ChIP-SICAP assay indicated that Cdh1 is one of the most intense proteins after the histones (fig. 4.6). Interestingly, Cdh1 was identified mostly by C-terminal peptides. In addition, its known interactors Beta-Catenin and Delta-Catenin (p120) were identified, of which the latter is required for nuclear localization and DNA-binding of (E-cad/CTF2) [142]. While p120 and several other proteins were specifically identified using Cdh1, neither Oct4, Sox2, Nanog nor many other stem cell maintenance factors were detected in the Cdh1 ChIP-SICAP experiment. Thus protein-DNA interactions revealed by ChIP-SICAP are specific, and depend on the protein used as a bait.



**Figure 4.7:** The heatmap of the overlapping proteins based on their 2i/Serum ratios by ChIP-SICAP. Cdh1 (E-Cadherin) is clustered differently with Nanog, Oct4 and Sox2. The first vertical cluster was enlarged in the right panel. The green genes are known components of the core-circuitry of pluripotency.

Specificity was further confirmed by clustering of the OSN and Cdh1 ChIP-SICAP results, where Oct4, Sox2 and then Nanog clustered together while Cdh1 was separated as an out-group (fig. 4.7). In addition, many stem cell regulators tightly grouped together, including Nanog, beta-catenin, Rex1, Tcf3, Tbx3, Kdm3a (Jmjd1a) and Prdm14 Remarkably, Prdm14 is one the most prominent factors whose association with the OSN network is strongly induced in 2i conditions.

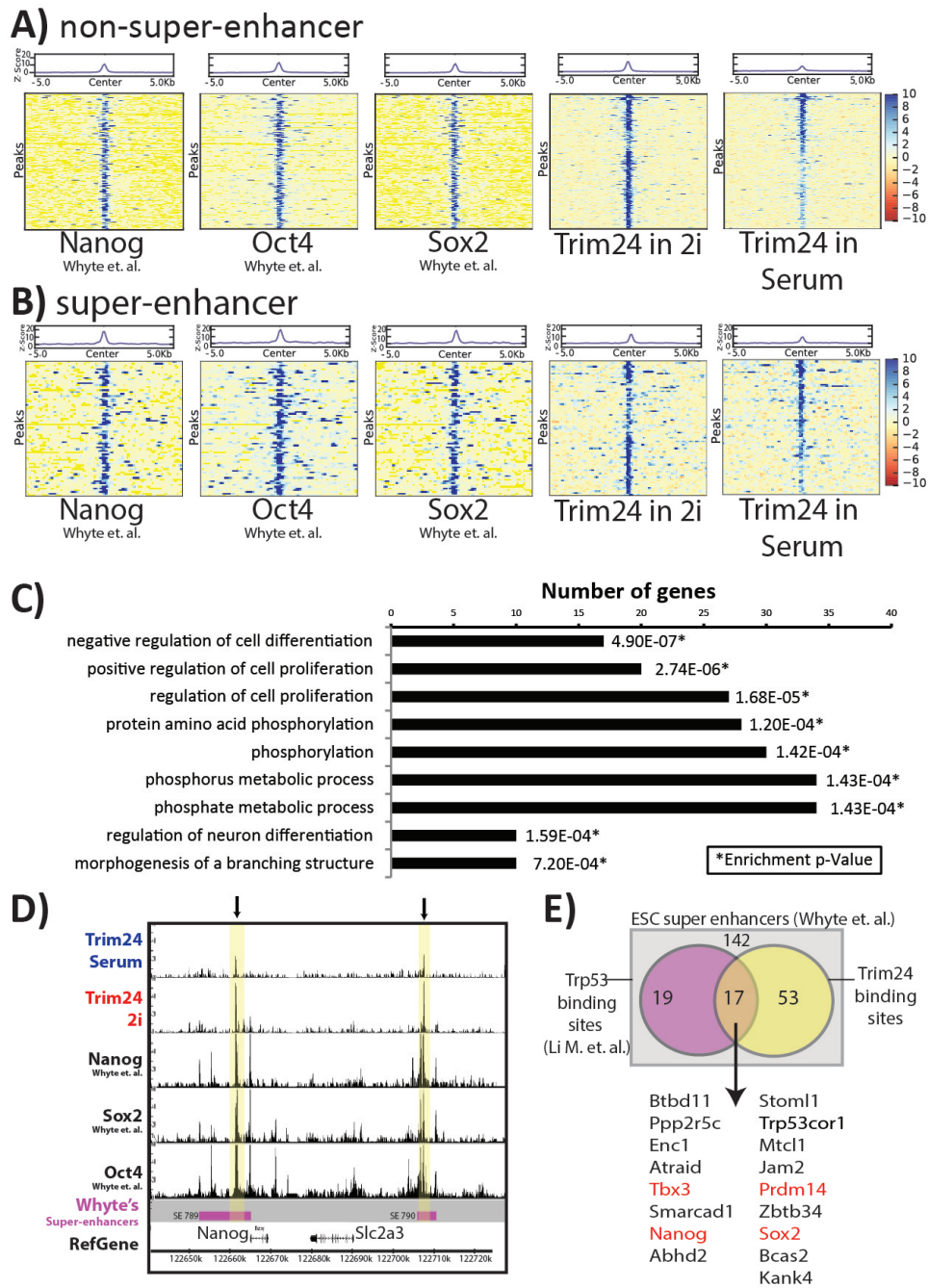


In fact, Yamaji et al [143] have previously shown that Prdm14 has an important role in the generation of primordial germ cells (PGC) by epigenetic reprogramming and re-establishing pluripotency. Furthermore, we found that several components of the BAF complex (Smarcd2, Smarcb1, Smarcc1, Smarcd1 and Smarca4) as well as histone demethylases Kdm3a (Jmjd1a) and Kdm3b (Jmjd1b) are significantly higher in 2i medium (all adj. p-value  $\leq 0.1$ , fig. 4.8). This is in line with their demonstrated role in promoting reprogramming or regulating pluripotency (e.g. Smarcc1 (BAF155) and Smarca4 [144]; Smarcd1 (Baf60a, [145]), Kdm3b and Kdm3a [146, 147]).

#### 4.4 Verification of Trim24 as a novel component of the core-circuitry

Among the novel candidates in the first cluster in fig. 4.7, I observed Trim24, an E3-ubiquitin ligase that subjects p53 to degradation [148]. Since Trim24 has not been associated with the pluripotency network before, we decided to verify the co-localization of Trim24 with OSN by ChIP-seq. Interestingly the ChIP-seq result revealed that Trim24 co-localized with OSN in about 800 enhancers (fig. 4.9A, B), including 70 super-enhancers defined by Whyte et. al. [149]. Additionally, Trim24 preferentially binds to 443 enhancers in 2i-condition compared to serum medium, which is in line with our proteomic data. Interestingly these enhancers are close to genes involved in negative regulation of cell differentiation (fig. 4.9C, D i.e. Nanog, Lif, Nr0b1 and Gli2, gene enrichment p-value  $3.08E-5$ , Supplementary Table 4.3). Additionally, based on the genome-wide occupancy of p53 in mouse ES cells [150] and our ChIP-seq data, Trim24 and p53 co-localize on several of the ESC super-enhancers (fig. 4.9E), remarkably including the super-enhancers of Nanog, Prdm14, Sox2 and Tbx3. Thus it is conceivable that the effect of p53, a potent inhibitor of pluripotency, on these super-enhancers is moderated by Trim24 in 2i medium, possibly by its targeting for degradation. Alton et al have shown that knocking down Trim24 in mouse ESCs leads to p53-dependept apoptosis [148].





**Figure 4.9:** The co-localization of Oct4, Sox2, Nanog and Trim24 on the ES enhancers A) The Trim24 ChIP-seq profile of the signals on the non-super-enhancers of the mouse ESCs. The non-super-enhancers have been defined by Whyte et. al. [149] B) The ChIP-seq profile of the signals on super-enhancers defined by of the mouse ESCs. The super-enhancers have been defined by Whyte et. al. [149] C) The enriched biological processes associated with the genes around the up-bound enhancers in 2i condition. D) The profile of the ChIP-seq signal on Nanog super-enhancers. E) The Venn diagram shows where Trim24 and Trp53 are co-localized on the super-enhancers. The Trp53 binding sites were obtained from [150].

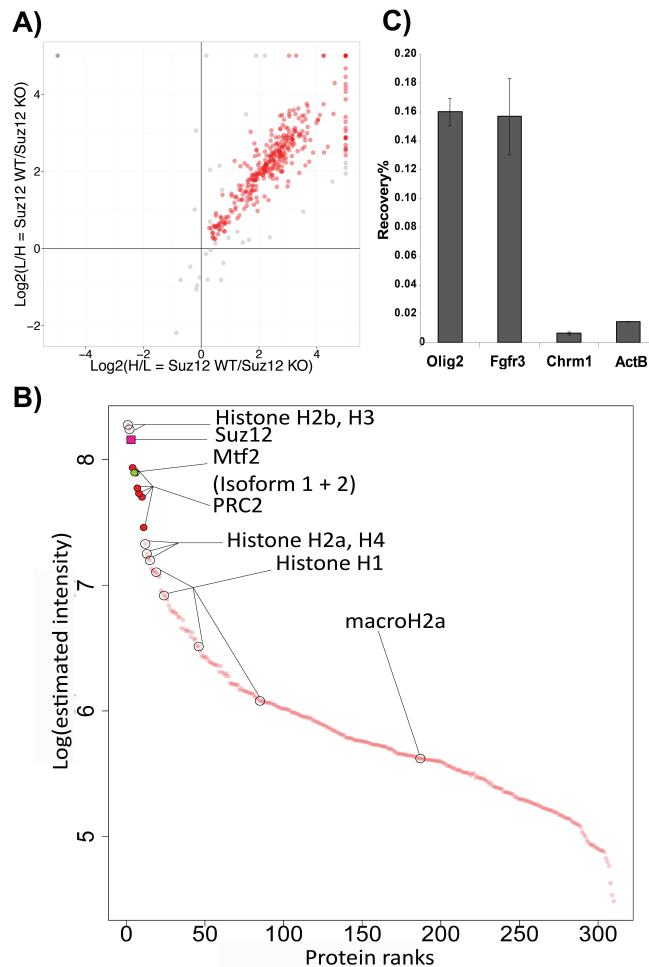
## 4.5 Studying proteins co-localized with PRC2 on chromatin

After applying ChIP-SICAP to the core-transcriptional regulatory circuitry of the mouse ES cells, I used this method to study proteins co-localized with Suz12, and hence PRC2 proteins in these cells. In mammalian cells PRC2 consists of 6 core components: Suz12, Ezh2 or Ezh1, Eed, Rbbp4 and Rbbp7, which are involved in silencing gene transcription. Remarkably, PRC2 proteins have important roles in expressing developmental genes; such that by knocking out Suz12, Ezh2 or Eed mouse embryos die on day E7.5 to E8.5 [151-153]. Identification of proteins co-localized with PRC2 proteins on chromatin may help us to understand how they are recruited to the target genes to be silenced.

To achieve this goal, I carried out Suz12 ChIP-SICAP in the wild type ES cells, and as the negative control I used Suz12 knock-out (SKO) ES cells. Additionally, I reversed the SILAC labeling between the replicates as an independent biological replicate. In other words, in the first replicate the wild type ES cells are labeled with heavy SILAC, and the SKO ES cells are labeled with light SILAC; and vice versa in the second replicate. Hence I mixed the wild type and the SKO ES cells from the beginning of the assay thereby minimizing experimental bias. As a result, 383 proteins were identified as the intersect of two replicates (fig. 4.10A, B and Supplementary Table 4.4), out of which 348 proteins were significantly enriched compared to the negative control (adj. p-value  $\leq 0.1$ ). A few background proteins were identified using both replicates in the knock out samples. In fact, in this case knock out control is the best negative control, and evaluates the off-target binding of the antibody. This indicates the benefit of double purification by the target protein and by DNA.

As expected I observed all 6 core components of PRC2. Furthermore, ranking the identified proteins based on the estimated intensities shows histones and Suz12 and the other PcG proteins as the most abundant proteins. Interestingly I observed two isoforms of Mtf2, which are required for binding of PRC2 to the repressed Hox genes, and the inac-

tive X-chromosome [154, 155], as well as the global level of H3K27me3 in mouse ES cells [156]. In fact, the first 102 aminoacids of the isoform 1 are missing in the isoform 2, consequently the Tudor domain is lost. It has been shown that the Tudor domain binds to methylated H3K36, and it is required for the recruitment of PRC2 by Mtf2 [156]. However, it is not yet known what the function of the isoform 2 is.



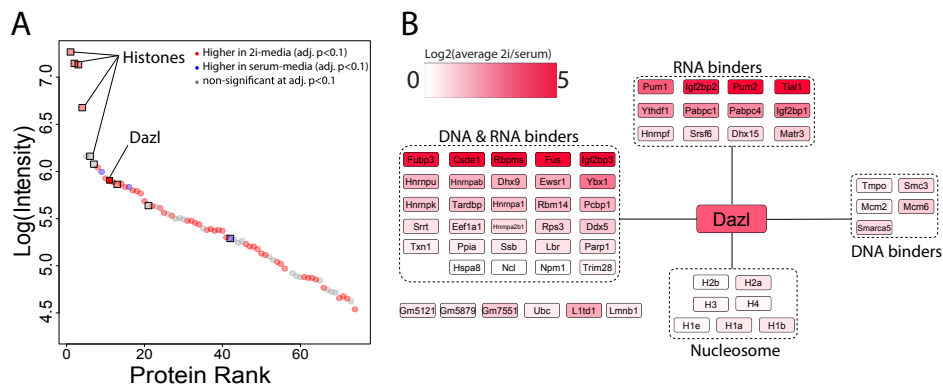
**Figure 4.10:** Suz12 ChIP-SICAP: A) The scatterplot indicates the enriched proteins by targeting Suz12 in comparison to the Suz12 knockout control with red and gray dot, respectively (adj.  $p$ -value $<0.1$ ). B) The proteins were ranked based on their estimated intensities. C) qPCR demonstrates enrichment of Suz12 targets Olig2 and Fgfr3 in comparison to the background Chrm1 and ActB. The DNA was retrieved from a ChIP-SICAP sample prep.

Moreover, I observed Jarid2 co-localized with Suz12. Indeed, it has been demonstrated that Jarid2 binds to many PcG target genes. In addition, it is essential and sufficient for recruiting PcG proteins, as well as differentiation of the ES cells and proper development [157]. Furthermore, using Suz12 ChIP-SICAP I identified several uncharacterized proteins such as AU022751 and Gm340, which are interesting candidates to determine their functional relevance with PcG proteins. All combined, these results demonstrate the power of ChIP-SICAP identifying all known core and peripheral PRC2 proteins in a single experiment with high confidence, in addition to several new candidates. Additionally, I wanted to retrieve DNA at the end of the ChIP-SICAP procedure. Therefore, one can carry out qPCR, sequencing and mass spectrometry in parallel with one sample prep. To achieve this aim, I used SP3 peptide purification after the protein digestion [158]. In fact, using this protocol the peptides bind to paramagnetic beads coated with carboxyl groups by acetonitril as an organic solvent. I assumed that DNA fragments do not bind to the beads, and therefore I kept the acetonitril supernatant. Then I evaporated the liquid, and I reconstituted the solution by Tris-HCl 10mM. After that I subjected the solution to phenol-chloroform-isoamylalcohol procedure to purify DNA. Finally, qPCR results demonstrated a successful enrichment of Suz12 binding sites such as Fgfr3 and Olig2 in comparison to ActB and Chrm1 (fig. 4.10C). Thus, using ChIP-SICAP, not only the protein co-localizations are detected, but also DNA can be retrieved to be used for qPCR or sequencing.

## 4.6 ChIP-SICAP is the reciprocal validation of SICAP

In chapter 3 using the SICAP procedure I have shown that many known chromatin-binding proteins together with the novel ones are identified making use of the fact that SICAP captures DNA, followed by identification of DNA-interacting proteins by mass spectrometry. As a reciprocal validation for the novel chromatin-binding proteins one may

utilize ChIP-SICAP, in which a protein is immuno-precipitated and then DNA is labeled to isolate the chromatin-bound proteins from the other interactants. In particular, ChIP-SICAP will provide stronger evidence of chromatin-binding compared to immuno-fluorescence microscopy, which at best can show nuclear localization, while adding the advantage of directly validating the specificity of the used antibody.



**Figure 4.11:** Proof of chromatin-binding for Dazl by ChIP SICAP. A) Sorting the identified proteins by ChIP-SICAP based on their intensities indicates histones on top following Dazl. B) Proteins with  $\log_2$  average  $>0$  were considered as true positives, as Dazl is highly up-regulated in 2i medium. Looking at the GO terms associated with the enriched proteins indicates several proteins with DNA/RNA-binding functions.

Here I selected Dazl from the SICAP data as a non-canonical chromatin bound protein, yet showing a strongly induced association to chromatin in 2i condition (Supplementary Table 3.3). Dazl is a cytoplasmic mRNA-binding protein that binds to the 3'-UTRs [159] and inhibits the translation of key transcripts during the development of primordial germ cells including Sox2, Rex1(Zfp42) and Sall4, as well as pro-apoptotic genes such as Caspase 7 [160]. In addition, Dazl enhances translation of Tet1 resulting in a global increase in DNA-hydroxymethylation in mouse ES cells grown in 2i conditions [161]. Dazl was identified in SICAP despite the treatment of the samples with RNase A and the use of TdT which specifically reacts with DNA but not RNA, thus lending strong support to the notion that Dazl is a genuine DNA-interacting protein. In addition, another 900 RNA-

binding proteins were uncovered, suggesting that proteins interacting with both DNA and RNA may be a common phenomenon. To directly show that Dazl associates with chromatin and to identify its interaction partners we used an antibody against Dazl in a ChIP-SICAP approach comparing ES cells in 2i and serum conditions following differential SILAC labeling. Of note, Dazl is much higher expressed in 2i cells (Supplementary Table 3.3), thus we expected to see that the true co-localizers would be higher in 2i medium.

Interestingly we identified histones as the most abundant proteins, then Dazl, and then several other proteins that are capable of binding to RNA and DNA such as Pum2, Fus, HNRNPs, Sfpq, Ddx3x, Srrt, Tardbp, most of which are higher in 2i medium (Figure 4.11 and Supplementary Table 4.5). As a result, we can confirm the chromatin-association of Dazl, which may indicate Dazl binding to its targets during transcription.



# 5

## Toward Single Locus Chromatin Composition by TIGR

---

Although ChIP-seq is a powerful method, it determines the binding sites for one protein of interest in each experiment. If we would like to identify all the proteins that associate with a given locus, it is hard to carry out ChIP-seq for hundreds of proteins. The traditional method for this aim is an in-vitro assay, in which short oligonucleotides encompassing the given sequence are immobilized on beads, and then treated with the nuclear lysate [162-164]. Finally, the enriched proteins in comparison to the negative control are identified by mass spectrometry or western blotting. Using this method normally hundreds of proteins are enriched in comparison to the negative controls, which is a clear indication of non-specific proteins bound to short oligonucleotides (45-bp)[165]. In addition, depending on how stringently the samples are washed, the final result might be different and non-reproducible. Therefore, the best solution is capturing the in-vivo fixed chromatin to avoid such artifacts. To achieve this aim Dejardin et al. have developed a method called proteomics of isolated chromatin fragments (PICh)[166]. In PICh after cross-linking using formaldehyde, chromatin is fragmented by sonication and then the loci of interest are isolated using biotinylated oligonucleotide probes. Specifically, he used locked nucleic acid (LNA)-based probes to enhance the hybridization of the probe with the target, which is cross-linked to the proteins. Using this method Dejardin could detect the proteins associated with the telomeres and pericentromeric regions [166-168] benefiting from the fact that each has tens of copies per cell. Unfortunately, they stated in his article [166] that PICh is not sensitive enough to detect single locus-associated proteins. Here I tried to improve this strategy by a

self-invented protocol for targeted isolation of genomic regions (TIGR). In fact, by changing the probe-hybridization conditions TIGR method is more convenient and efficient. In this study I applied TIGR to the Nanog promoter, and then I confirmed the specificity of the targeting by high-throughput sequencing. Following injecting the sample into the mass spec about 800 proteins were reproducibly identified, among which there are tens of known Nanog promoter binding proteins. Finally using ChIP-qPCR, I validated Nup98 as a novel protein that binds to the Nanog promoter.

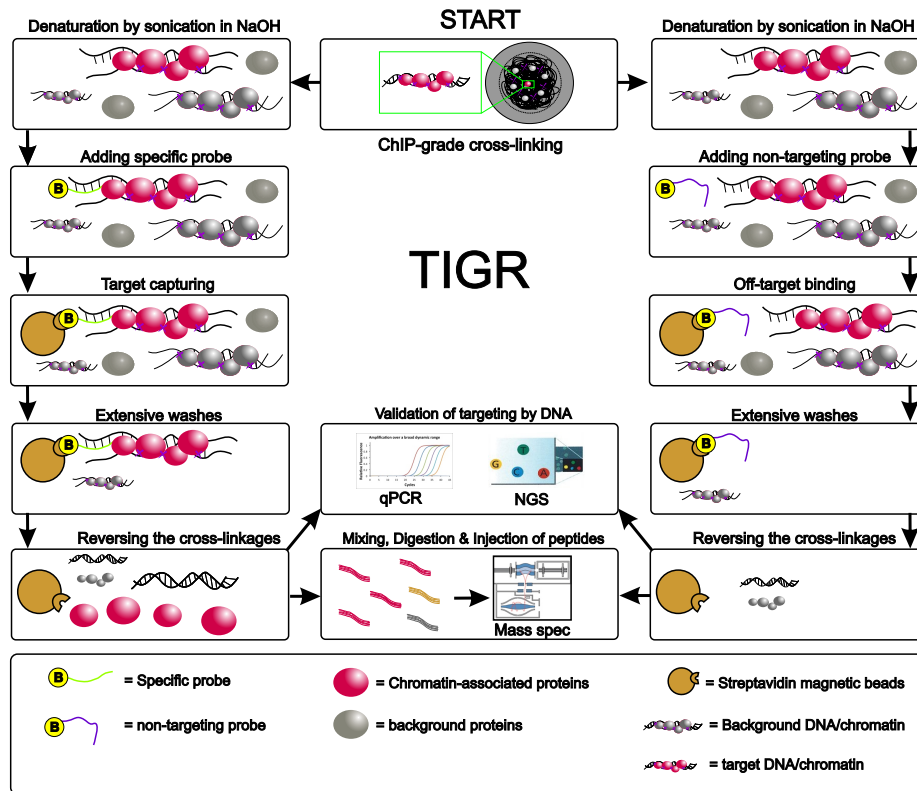
## 5.1 Experimental design of the assay

To lyse the cells, and partially purify chromatin I used DNAzol. DNAzol is a mixture of guanidinium and other denaturing reagents for extraction of genomic DNA. In the original protocol of DNAzol, ethanol is used to precipitate DNA. However, I found out that cross-linked chromatin precipitates just by spinning at 5000g without adding ethanol. Therefore, by removing ethanol less contaminant proteins precipitate together with the chromatin. Furthermore, in order to improve the efficiency of hybridization I used NaOH, which is used in Southern blotting for DNA denaturation and probe hybridization. After the sonication to fragment chromatin into shorter pieces, I added the oligonucleotide probes. Then I changed the buffer to bring the pH back to the neutral condition. As a result, DNA is renatured, and the probe may bind to its target. Moreover, heating at 55 ° C to 65 ° C for 5 min is also necessary to hybridize the probe efficiently. Since formaldehyde cross-linking is sensitive to heating, I did not heat above 65 ° C for 5 min, while, in PICCh protocol essentially hybridization is carried out by heating to 71 ° C for 7 min, and without adding NaOH.

In this study I designed two adjacent probes for the Nanog promoter and one non-targeting probe. The probes work efficiently when the lengths are 50 bp. In fact, comparing the ratios of the Nanog probe to the scramble probe provides a means to rule out background proteins. Additionally, comparing the result of the two adjacent probes indicates

the reproducibility of the results.

In order to check the enrichment of the Nanog promoter, I carried out qPCR for the Nanog promoter and a few other genomic loci including *Slc2a3*, which is the nearest gene to the Nanog locus. Furthermore, the isolate DNA fragments were subjected to next-generation sequencing. Then I injected the proteins obtained by the protocol to the mass spectrometer (fig. 5.1).



**Figure 5.1:** Schematic presentation of TIGR procedure. For more information, please refer to the text (section 5.1).

## 5.2 Evaluating the specificity and efficiency of TIGR

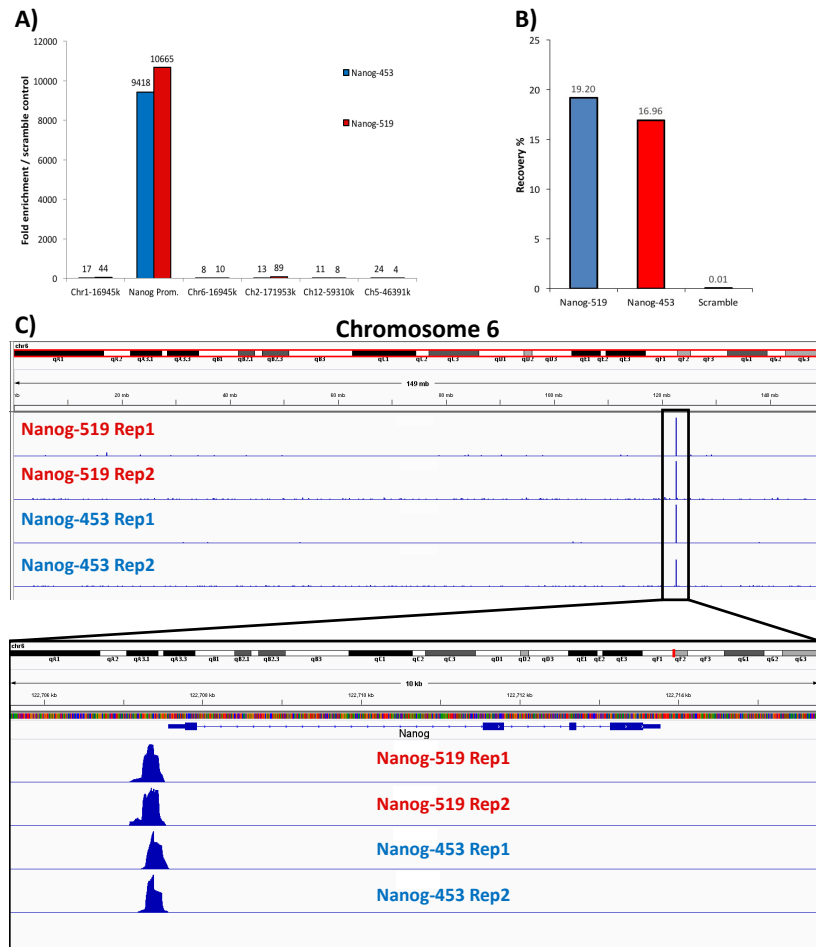
To check the specificity of targeting, the isolated chromatin fragments were subjected to DNA extraction and qPCR. It turned out that Nanog promoter is enriched 9000 to 10000-fold in comparison to the scrambled controls (fig. 5.2A). Additionally, the efficiency of recovering the

target was calculated by comparing with an input control, which is not subjected to pull down. The result indicated that between 17% to 19% of the chromatin fragments encompassing the Nanog promoter are recovered (fig. 5.2B). Moreover, the enrichment was calculated by comparing with the scramble probe. Furthermore, the pulled down samples were ligated to the Illumina linkers and submitted to high-throughput sequencing. As a result, it turned out that the Nanog promoter is the major peak, which encompasses a region <1 kb around the binding sites of the probes (fig. 5.2C). and the non-specific peaks are much less intense and non-reproducible. Altogether, TIGR efficiently and specifically targets the locus of interest.

### 5.3 Identification of proteins associated with the Nanog promoter

After confirming the specificity of the TIGR pull down, the samples were prepared for injecting to the mass spec. In this assay, the two Nanog specific probes Nanog-519 and Nanog-453, positioned just 10 bp apart in the Nanog-locus, were applied to the cells already grown with heavy and intermediate SILAC labels, respectively. In addition, the scramble probe was applied to the light-SILAC cells. After hybridizing the probes, the three samples were mixed and subjected to streptavidin pull down. Following stringent washing steps and reversing the cross-linking, the samples were injected at once into the mass spectrometer. The ratios of the proteins identified by Nanog specific probes over the scramble probe were calculated, and significantly enriched proteins were determined. Interestingly the proteins identified by the independent probes are very well correlated, and their overlap is very high (fig. 5.3A, B and Supplementary Table 5.1). There are 841 proteins at the overlap of the two Nanog probes, which make sense for a 1kb fragment of the chromatin. Looking at the biological processes enriched among these 841 proteins, 'regulation of transcription' is significantly enriched (fig. 5.3C). Additionally, many Nanog binding proteins were identified such as Oct4, Smarca4, p53, Sall4 and

B-Catenin (fig. 5.3D). Moreover, some general transcription factors and pol2 members were identified among the enriched proteins. Taken together, the enriched proteins are relevant to the target locus, therefore, it is worthwhile validating the unknown interactants.



**Figure 5.2:** TIGR efficiently and specifically enriches the Nanog promoter. A) The bar chart indicates the enrichment of the Nanog promoter, and the most intense off-target binding sites. The enrichment was calculated over the scramble (non-targeting) control. The locations of the off-target sites were determined by high throughput sequencing. B) The percentage of recovering the Nanog promoter relative to the input control. C) The tracks indicate the enrichment of the sequencing reads throughout the chromosome 6. Nanog-519 and Nanog-435: probes targeting the Nanog promoter at 519 and 453bp upstream of translation start site, respectively.

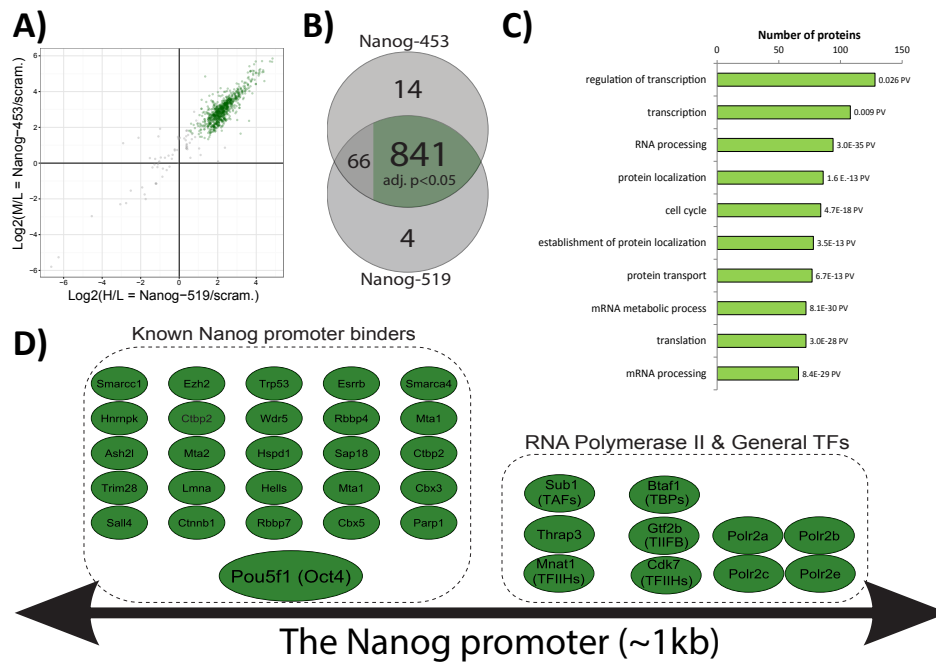
## 5.4 Comparing the Nanog promoter-associated proteins in 2i and serum conditions

In order to find the proteins related to the regulation of the Nanog promoter among the novel candidates identified in the previous experiment, I compared the two conditions of pluripotency by culturing the ES cells in 2i+LIF (2iL) and serum conditions. In fact, in 2iL medium Nanog has a more homogenous expression pattern among the cells. Therefore, comparing 2iL and serum condition is an interesting model system for finding Nanog promoter regulators. Thus I applied a Nanog probe using the established TIGR protocol, to the ES cells grown in 2i and serum media with heavy and light SILAC labeling, respectively. In this experiment the identified proteins were quantitatively compared between 2iL and serum (fig. 5.4A, Supplementary Table 5.2). In order to remove the non-specific proteins I looked at the overlap of this experiment with the previous one (fig. 5.4B). Once again many known interactors of the Nanog promoter were identified, most of which do not show such a significant change between 2i and serum in term of binding to the Nanog promoter, except Lamin A. Looking into the potentially novel proteins that differentially bind to the Nanog promoter, many of them are nuclear proteins and some of them are cytoplasmic (fig. 5.4C). Therefore, this experiment provided additional information about the dynamics of the proteins recruited to the promoter. Although several known Nanog promoter binding proteins were identified in this experiment, to be confident about the novel interactors I sought a reciprocal validation using ChIP-qPCR.

## 5.5 Validation of Nup98 as a novel protein that binds to Nanog promoter

Interestingly among the nuclear proteins that differentially bind to the Nanog-locus in TIGR, there are 6 nuclear pore proteins, which are expected to be in the nuclear membrane. Although nuclear membrane in

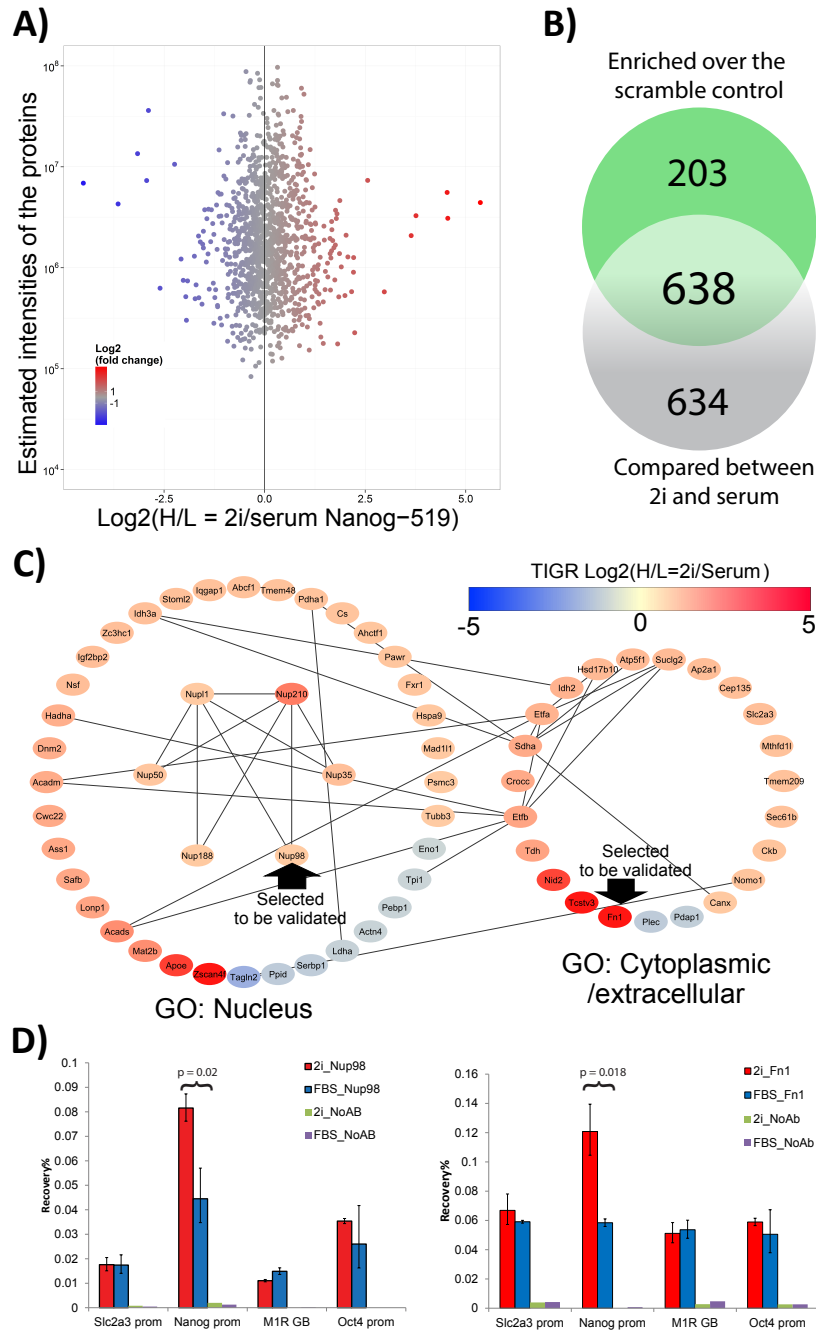
some points is connected to the chromatin, Nanog locus is not located at the periphery of the nucleus. To verify that Nup98 may also bind to the Nanog promoter in the ES cells I used ChIP-qPCR. Interestingly this experiment not only demonstrated the binding of Nup98 to the Nanog promoter but also indicates its preferential binding in 2iL medium (fig. 5.4D), which is in line with the TIGR result.



**Figure 5.3:** Proteins identified by targeting the Nanog promoter using TIGR. A) Comparing proteins isolated by the Nanog probes with the negative control. The heavy- and medium- SILAC cells were subjected to TIGR procedure using Nanog-519 and Nanog-453, respectively; and the light-SILAC cells was treated with the scramble probe. Proteins enriched significantly (adj. p-value  $\leq 0.05$ ) in comparison to the scramble control are shown in green. B) Comparing the TIGR results between the two Nanog probes using the Venn diagram. C) Biological processes related to the enriched proteins (shown in green in A and B). D) Some of the known interactors of the Nanog promoter, detected among the enriched proteins using the TIGR procedure.

Furthermore, I checked fibronectin (Fn1), as a cytoplasmic candidate using ChIP-qPCR. Interestingly Fn1 also shows enriched on the Nanog promoter preferentially in 2i medium. Comparing the immunoenrichment pattern of Nup98 with Fn1 indicates that Nup98 more specifically binds to the Nanog promoter, while, Fn1 generally binds to chromatin. Previously it has been shown that Nup98 binds to the promoter of genes involved in development, cell signaling and cell cycle-related processes [169, 170]. Although Nup98 does not have a known DNA-binding domain it binds to chromatin, and via the GLFG (glycine-leucine-phenylalanine-glycine) domain it is able to recruit other transcription factors like CBP/P300 and histone deacetylases (HDACs [171, 172]). Moreover, the GLFG bodies also promote the binding of intra-nuclear Nup98 to each other as foci, which are called GLFG bodies [173, 174]. These studies suggest that Nup98 has a primary role as an activator of transcription [175], which according to our data may count Nanog among its target genes. Altogether using comparative TIGR I found several known and novel Nanog-promoter binding proteins in addition to their dynamics on the promoter. Among the novel candidates I selected Nup98, and for the first time using ChIP-qPCR I could verify that it binds to the Nanog promoter preferentially in a cellular condition that Nanog expression has less fluctuation (2iL medium).





**Figure 5.4:** Comparing proteins associated with the Nanog promoter using the TIGR procedure. A) Proteins identified using the probe Nanog-519 are shown in a blue to red spectrum representing 2i over serum ratios. B) The Venn diagram shows the overlap between this experiment and the previous experiment (shown in fig. 5.3) C) Proteins differentially bound to the Nanog promoter are divided into two groups: nuclear and cytosolic D) ChIP-qPCR indicates Nup98 and Fn1 bind to the Nanog promoter, preferentially in 2i condition. prom. = promoter, GB = gene body.



# 6

## Studying the effect of Nup210 on Somatic Cell Reprogramming

---

As was mentioned previously, somatic cells can be reprogrammed to pluripotent stem cells by expressing the master or auxiliary TFs of pluripotency such as OSKM. However, this procedure is inefficient, and as a result of that many heterogeneous intermediate cell populations are generated. Nevertheless, some studies have tried to define a few markers as the cornerstones of this procedure [176, 177]. These markers are expressed sequentially among the cells that are successfully reprogrammed to the iPS cells during a time course, which normally takes 2 weeks. Therefore, focusing on the cells expressing the cornerstone markers may help us to understand the mechanism of cellular reprogramming. To this end, Hansson et. al. compared dynamics of protein expression during the reprogramming at 6 time points based on Thy1, SSEA1 and Oct4-eGFP expression [178]. Among the interesting information obtained by this study, remarkably it turned out that Nup210 is the only nucleoporin up-regulated during cellular reprogramming (fig. 6.1A).

As was mentioned in chapter 5, in addition to the classical roles of nucleoporins in transportation, they have important other functions including gene expression regulation [179]. Furthermore, recent studies indicate cell-type specific expression of nucleoporins [180, 181]. Interestingly, Nup210 is preferentially expressed in the epithelial cells during the embryonic development [182]. Strikingly, it has been reported that Nup210 is not detected in myoblast and mouse ESCs, however, becomes expressed during the differentiation [183]. Furthermore, suppressing Nup210 blocks differentiation by increasing apoptosis, and decreasing the genes required for differentiation [183]. Recently it has

been suggested that Nup210 has anti-apoptotic roles during myoblast differentiation [184]. Nevertheless how Nup210 performs its function has not been determined, yet.

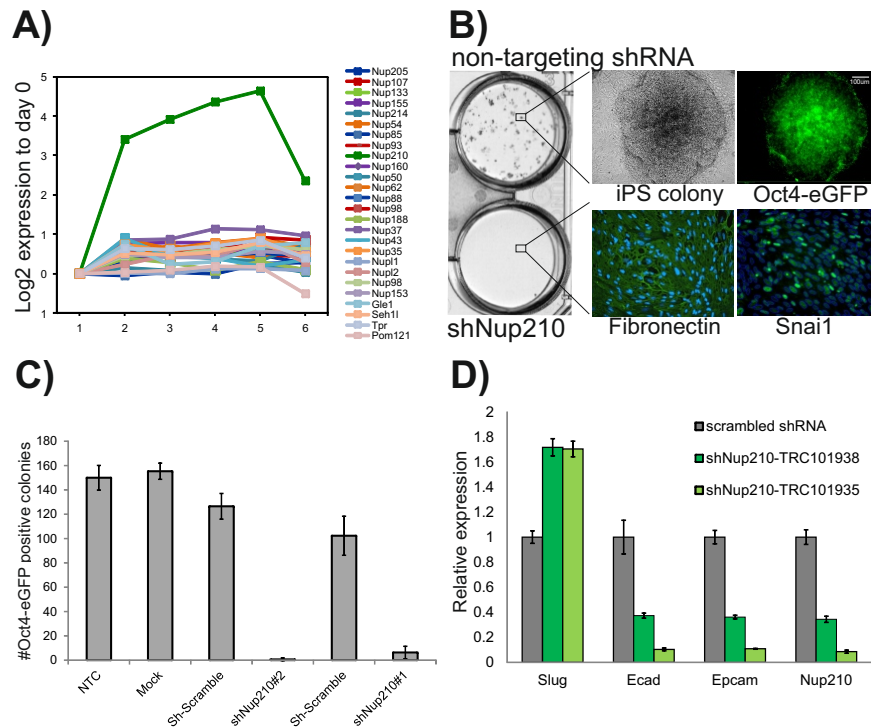
Up-regulation of Nup210 during cellular reprogramming is another striking observation suggesting Nup210 fulfills an important role during dedifferentiation, which I have investigated in this chapter.

## 6.1 Design of the assay

For this study, I used reprogrammable mouse embryonic fibroblast (MEF) generated by Stadtfeld et. al. [185], containing a cassette for expressing Oct4, Sox2, Klf4 and Myc (OSKM). The cassette was engineered in *Colla1* locus, downstream of a dox-inducible promoter. In addition, the MEF cells are able to express reverse tetracycline-dependent transactivator (rtTA) via another knock-in on the *Rosa26* locus. Therefore, upon adding doxycycline to the medium of the MEF cells they are able to start reprogramming without further transduction of OSKM. The advantage of this system in comparison to the dox-inducible lentiviral system is avoiding the epigenetic silencing of the OSKM. As a result, transgene expression is more efficient, and homogenous among the MEF cells. Furthermore, in these MEF cells eGFP coding sequence has been inserted into the endogenous Oct4 locus. Thus, after complete reprogramming and establishing the pluripotency the endogenous Oct4 is derepressed, and eGFP is expressed as well. As a result, the iPS colonies are eGFP positive, which is easily distinguishable from the intermediates of reprogramming.

In order to investigate the role of Nup210 in reprogramming, I knocked down Nup210 by two different shRNAs during the generation of the iPS cells. Thus if Nup210 has an important role, abolishing it should affect cellular reprogramming. To do that, shRNAs are transferred via lentiviruses produced by human embryonic kidney (HEK) cells. Initially, HEK cells are transfected by the plasmids encoding the shRNAs via a lentiviral vector. In addition, the other necessary genes for the production of the viruses are transfected to the HEK cells. As

a result, viral genome is encapsulated and secreted to the medium of the HEK cells. The medium is collected, filtered and concentrated. To determine the titer of the viruses, some regular MEF cells are infected. Then the genomic DNA is isolated, and the number of integrated viral genome is determined by qPCR. Hence, the Nup210 shRNAs and the scramble shRNA are transduced equimolar to the reprogrammable MEF cells. Then doxycycline is added to the medium of the cells. After 2 weeks as the read-out I counted the number of Oct4-eGFP positive colonies, and I compared it with the scramble control.



**Figure 6.1:** Studying the role of Nup210 in reprogramming. A) Expression profiling of nucleoporins during reprogramming was carried out by Hansson et. al [178]. Nup210 is exclusively up-regulated. B) In contrast to the non-targeting shRNA, using shNup210s generation of Oct4-eGFP positive (iPS) colonies is hindered; fibroblast cells keep their identity, as they express Snail and extracellular matrix of fibronectin. The antibody signal is green, and DAPI is blue. C) The plot indicates the number of Oct4-eGFP positive colonies after 2 weeks. shNup210#1: TRC101938, shNup210#2:TRC101935. D) Comparing the mRNA expression level of epithelial and mesenchymal markers between the scramble control, and the shNup210s.

## 6.2 Nup210 is required at least for the initiation phase of reprogramming

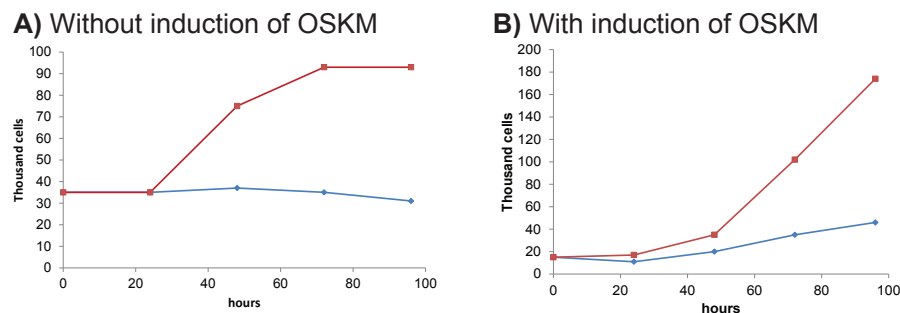
Strikingly, knocking down Nup210 in the reprogrammable MEF cells dramatically reduces the number of Oct4-eGFP positive colonies generated at the end of two weeks (fig. 6.1B, C). Indeed, knocking down Nup210 by two independent shRNAs rules out the possibility of an accidental result. Therefore, Nup210 has a critical role in successful induction of pluripotency.

Previously, cellular reprogramming has been divided into 3 steps: Initiation, maturation and stabilization [186]. For each step, some markers or events have been defined. The hall mark of the initiation phase is indeed mesenchymal to epithelial transition (MET) that can be visualized by attenuation of mesenchymal markers such as Snail and expression of epithelial markers such as E-cadherin. During the maturation phase, Nanog is derepressed, and the cells are committed to irreversible reprogramming. Finally, during the stabilization phase pluripotency-associated markers are expressed [185].

Morphology of the Nup210 knock-down cells suggests that the cells have not accomplished MET. To consolidate this observation, I used immuno-staining by Snail as a TF specifically expressed in mesenchymal cells. In addition, I used fibronectin as a marker highly expressed by the fibroblast cells. As a result, immuno-staining indicates that Nup210 knock down cells still express Snail and fibronectin. Furthermore, comparing epithelial markers such as E-cadherin and EpCAM between the Nup210 knock down cells and the scramble control by qPCR indicates that Nup210 knock down cells have failed to express epithelial markers. Conversely, a mesenchymal marker like Slug has remained higher in the Nup210 knock down cells. Altogether, knocking down Nup210 clearly hinders cellular reprogramming, and the MEF cells are not able to pass the first phase of reprogramming, quite likely they retain their fibroblast identity despite expressing OSKM for 2 weeks.

### 6.3 Nup210 is needed for normal proliferation of the MEF cells

To investigate further why Nup210 is required for MET, I knocked it down in the MEF cells with and without the induction of OSKM. In both cases comparing the growth curves with the scramble control clearly indicates that the cells do not proliferate normally as a result of Nup210 suppression. Without expressing OSKM, the cells stop proliferating, whereas, expression of OSKM up-regulates Nup210, and to some extent alleviates cellular senescence. Therefore, Nup210 is needed for normal proliferation of the MEF cells. Interestingly, it has been previously shown that rapid proliferation is essential for the induction of pluripotency, conversely, cell cycle arrest inhibits successful reprogramming [187]. Thus knocking down Nup210 inhibits normal proliferation of the cells, and consequently OSKM are unable to push the cells toward an epithelial identity, hence, cellular reprogramming fails in the initiation phase.



**Figure 6.2:** The growth curve of the MEF cells after the suppression of Nup210. A) The growth curve of the MEF cells transduced with shNup210 and the scramble shRNA are shown in blue and red, respectively. Doxycycline was not added to the medium of the cells. B) Doxycycline was added to the medium of the MEF cells to induce the expression of OSKM. Dramatic up-regulation of Nup210 (based on fig. 6.1A), alleviates the effect of shNup210.

Interestingly, TIGR results (chapter 5) suggest Nup210 is a Nanog promoter-associated protein. Additionally, SICAP (chapter 3) using the mouse ES cells indicates that Nup210 is a chromatin-associated

protein. Although nucleoporins traditionally have been considered as nuclear transporters, recent studies have revealed their roles in the maintenance of genome integrity, and DNA repair. For instance, Nup84 is involved in relocation of Double-strand DNA break to the nuclear periphery, which probably assists the repair process [188]. Furthermore, telomeric chromatin is protected by binding to the Mlp1 and Mlp2 nucleoporins via Yku70/Yku80 complex [189, 190]. Unprotected telomeres eventually arrest cell cycle [191, 192]. Additionally, ribosomal DNA (rDNA) repeats are associated to the nuclear envelope via CLIP proteins, which are critical for the stability of the chromatin loops at these loci [193, 194].

Altogether, Nup210 is a chromatin-associated protein that is essential for cell proliferation, and iPS generation. The mechanism by which Nup210 affects cell cycle, whether by DNA replication or other pathways, remains to be elucidated further.



# 7

## Discussion

---

By comparing chromatin-associated proteins between two cellular states it should be possible to understand how the cells are able to change their transcriptional programs in order to establish a new phenotype. However, due to the complexity of chromatin, studying the DNA-bound proteins has been challenging. Although ChIP-seq has provided valuable information about the genome-wide occupancy of one protein in each experiment, it is blind to other chromatin-associated proteins. To identify chromatin-bound proteins in an unbiased manner, the available methods are not solid enough to rule out the artifacts and contaminants associated with the isolation of chromatin. For example, nuclei and the chromatin fraction are commonly isolated based on the large size and insolubility, respectively. However, using these procedures many unexpected proteins such as ribosomal proteins or growth factors are identified as well. Therefore, it is hard to exclude that they are contaminants from the endoplasmic reticulum (ER), as the membrane of the nuclei and the ER are fused together. Even for a canonical chromatin-binding protein it is hard to distinguish between the chromatin-bound fraction and the soluble pool in the nucleosol.

### 7.1 Specific *in-vitro* labeling of *in-vivo* fixed chromatin

Here I have aimed to develop novel methods for studying dynamics of chromatin-associated proteins globally, and in a protein-targeted fashion by SICAP and ChIP-SICAP, respectively. I have demonstrated

the power of these methods by studying chromatin composition in the ES cells, and by comparing the ground-state and the metastable state of pluripotency to identify multiple known as well as novel proteins involved in the maintenance of pluripotency.

Firstly, I used formaldehyde cross-linking and sonication as it is conventionally used in ChIP, therefore, it is unlikely that too many proteins are cross-linked to the chromatin, as the validity of ChIP results with formaldehyde cross-linking has already been demonstrated [195]. Furthermore, after the sonication I spun the sheared chromatin to sediment the large molecular complexes and cell debris. As a result, chromatin and other cellular complexes are fragmented into small fragments.

The key benefit of SICAP and ChIP-SICAP methods resides in the TdT-mediated biotinylation of sheared DNA, thereby specifically labeling DNA. Capture on streptavidin beads facilitates stringent washing, providing distinct advantages over centrifugation-based and contamination-prone chromatin-precipitation. The benefit of ChIP-SICAP over mChIP [108], ChIP-MS[109], ChroP[110] and RIME[196] resides in the subsequent enrichment of the bait protein and DNA (essentially constituting a DNA-based pull-down) formally proving that the target protein and its partners bind to chromatin. Additionally, by removing the background proteins, including the antibody and protein A/G, the ratio of signal to noise is enhanced, and false negatives/positives are dramatically reduced. As a result, ChIP-SICAP works with a substantially reduced number of cells (20 million) in comparison to mChIP (10 billion), ChIP-MS (1 billion) and ChroP (100 million).

Importantly, ChIP-SICAP revealed that proteins co-localized on the chromatin are quite different to the soluble interactants. Indeed, this observation has not been feasible previously, as the previous affinity-capture methods are not able to distinguish between the chromatin-bound proteins and soluble interactants. Thus using ChIP-SICAP we are able to identify novel partners for a protein that are functionally more relevant to the chromatin-linked roles of the target protein.

## 7.2 Dynamics of chromatin-associated proteins in the ground-state and metastable states of pluripotency

The ability of SICAP to determine changes in chromatin composition revealed that chromatin is highly dynamic, evidenced by the large number of proteins preferentially associating to DNA in either 2iL or serum conditions (fig. 3.3A). Furthermore, dynamics of chromatin-binding does not necessarily follow the changes at the expression level of a given protein. For instance, Uhrf1, required for the maintenance of 5mC, is only significant based on SICAP data in serum condition, which is supported by the fact that DNA methylation is higher in the same condition [24, 89]. Moreover, Scml2, a polycomb-associated protein, does not show significant changes in its total level of expression, while the chromatin-binding in 2i condition is indeed significant. Considering that DNA is hypomethylated in 2i condition, the SICAP result is in line with a recent study that has shown Scml2 association to chromatin upon hypomethylation of DNA[112]. Taken together, dynamics of chromatin-binding revealed by SICAP is more influential in comparison to the full proteomics in terms of understanding the mechanism of transcriptional regulation.

Furthermore, using ChIP-SICAP it turned out that several key pluripotency factors such as Tbx3, Stat3, Sall4, Esrrb and Zfp42 (Rex1) show enhanced chromatin interaction in the ground state of pluripotency, however, without a change in their overall expression level. This observation may provide an explanation for how 2i medium provide a better condition for culturing ES cells. Additionally, based on our result Tcf3 binds to the chromatin upon the activation of the Wnt pathway. Therefore, our result supports the model that  $\beta$ -Catenin converts Tcf3 from an inhibitor complex to an activator of OSN [87], rather than wiping it out [86].

Strikingly SICAP revealed that the growth factor Fgf4 binds to chromatin especially in 2iL condition. Interestingly we showed that

HA-tagged Fgf4 secreted from HEK293 cells can translocate to the nucleus of ES cells. Although this is an unknown property of Fgf4, studies on the mechanism of nuclear import of Fgf1[197] and Fgf2[198] have identified cytosolic interactors and karyopherins to mediate transport through the nuclear pore. Although we can only speculate on the function of chromatin-bound Fgf4, it is at least striking that PD0325901, one of the 2iL inhibitors, inhibits Fgf4 signaling to prevent differentiation. As a result, we observe massive up-regulation of Fgf4 in 2i medium (>20-fold). One explanation is that excessive Fgf4 may be confined in the nucleus to avoid over activation of the signaling pathway. Clearly, further studies will be required to investigate this unexpected trait of Fgf4, which should include ChIP-seq to determine genome-wide occupancy. Experiments in this direction failed in our hands because of the lack of a ChIP-grade antibody. Moreover, I have tried to generate a stable cell line for expressing Fgf4 fused with the HA-tag, however, over-expressing Fgf4 leads to significant differentiation of the ES cells even in 2iL medium.

### **7.3 Reciprocal validation of SICAP and the antibody using ChIP-SICAP**

By design ChIP-SICAP targets the chromatin-bound fraction of the bait protein and proteins that bind in its vicinity. After mass spectrometry, for all tested targets (Oct4, Sox2, Nanog, Suz12 and Dazl) this resulted in a typical pattern ranking core histones as the most abundant proteins, followed by the bait protein. Histones are expected from the fact that ChIP-SICAP investigates native chromatin, however, histones have normally not been detected by affinity purification of OSN. Identification of the bait protein provides an independent validation of the antibody and its ability to pull down the bait under ChIP conditions. Thereby ChIP-SICAP follows the recommendations that were recently proposed for the quality-control of antibodies in affinity purification strategies[199]. Finally, typically dozens to hundreds of proteins follow at lower intensities, which we interpret as proteins

that co-localize with the bait at a decreasing number of genomic loci. These proteins do not necessarily physically interact with the bait protein, but are co-isolated on the same DNA fragment. Considering that chromatin was sheared to an average size of 200 bp, this still implies close protein spacing to facilitate functional cross-talk.

Additionally, ChIP-SICAP procedure is a proof of chromatin-binding for a protein of interest. In fact, observing the histones on top of the protein intensities followed by the target protein, verifies chromatin-binding of the target protein. Therefore, the unexpected proteins identified by SICAP can be further consolidated by ChIP-SICAP. As an example, using this approach I could demonstrate that *Dazl* is a chromatin-binding protein, which co-localizes with several other proteins with the ability to bind to DNA and RNA. This observation suggests that post-transcriptional regulations are started right away from the chromatin.

It is noteworthy that the result of ChIP-SICAP indicates proteins co-localized with the bait protein on many genomic loci, however, these co-localizations should be repeated frequently enough to provide the mass spectrometer with enough material to identify the peptides. In other words, the identified proteins are among the typical co-localizers of the bait protein over many genomic loci. Thus proteins identified by the ChIP-SICAP procedure are potentially good candidates to be verified for binding to the locus of interest.

## **7.4 Expanding the core-transcriptional circuitry of pluripotency**

Using ChIP-SICAP we identified about 400 proteins that co-localized with all 3 core factors *Nanog*, *Oct4* and *Sox2* (OSN) reflecting high inter-connectivity most likely at multiple genomic loci. The tight clustering of novel proteins among these well-established components of the core-circuitry prompted us to validate *Trim24*, an ubiquitin ligase targeting p53 for degradation[200]. ChIP-seq not only confirmed association of *Trim24* to the OSN network, but also showed that it fre-

quently did so at super-enhancers co-occupied with p53 near Nanog, Sox2, Tbx3 and Prdm14 loci. This may explain how p53 is restrained within the core-circuitry to safeguard pluripotency in 2iL medium. Indeed, the important role of Trim24 in the development of blastocyst, and activation of embryonic transcription has been previously demonstrated[201].

Indeed, one of the questions in the field is that how 2i ES cells can survive while Myc is dramatically down-regulated in 2i condition ( 30-fold). Here we observe that Trim24 co-localizes with OSN on the enhancers of the genes related to cell proliferation such as Kit, MycN, FoxP1, Mdm4, Gli2 and Kras. Therefore, it is conceivable that in the absence of Myc, Trim24 positively regulates the cell proliferation, as it is a potent oncogene in breast cancer[202], [203].

## 7.5 Toward identification of locus-specific associated proteins

Although ChIP-SICAP is a robust method, it needs a good antibody for the target protein. Indeed, identification of proteins associated with a locus of interest would be more straight-forward if we could capture the given locus by oligonucleotide probes. In other words, in this case the aim of the method is exactly complementary to that of ChIP. Previously Dejardin and Kingston have tried to achieve this aim by LNA probes using a method called PICh[111], however, this method has so far remained limited to the repetitive regions with many copies per cells [111, 112, 204]. Likewise, Kennedy-Darling used the same strategy to identify proteins associated with rDNA, telomeric X-element, and Gal 1-10 in yeast[205]. However, the efficiency of recovering the target is only about 1%. Here I have aimed to improve this strategy by modifying the hybridization of the probe using alkaline conditions instead of neutral pH. Starting from a self-invented protocol, I called this procedure as TIGR. Evaluated by qPCR and high-throughput sequencing, the hybridization is quite efficient and specific (about 15% to 20% in the case of Nanog promoter). Remarkably several TFs and chromatin

proteins known to associate with the Nanog promoter have been identified using TIGR. Being verified independently by ChIP-qPCR, for the first time I could demonstrate that Nup98 binds to the Nanog promoter. These results all promising that TIGR method is a powerful and generalizable tool to identify proteins that interact with individual genomic loci whose functionality may have been suggested by genetic or computational approaches, but whose biochemical composition may now be unraveled by TIGR.

## 7.6 Conclusion

Throughout this thesis, three novel methods were invented for studying chromatin protein composition. Firstly, SICAP was presented for studying global composition and dynamics of chromatin-associated proteins. Using this method DNA-protein complexes are specifically labeled in an enzymatic reaction by TdT to be purified. This approach was used for the first time in the field. This opens up the way for unbiased identification of unknown chromatin complexes, many of which have been ignored due to the lack of a specific method in the field. Additionally, proteins with subtle changes in their expression levels are commonly ignored in studying biological systems, however, thanks to SICAP I have shown that some of these proteins bind to chromatin significantly, which is quite important from transcriptional and epigenetic point of view.

Secondly, I combined ChIP and SICAP to focus on the chromatin-bound partners of a protein of interest. This combination provides a solution for shortcomings in previous methods such as mChIP, ChIP-MS, ChroP and RIME. Indeed, thanks to the molecular mechanism of ChIP-SICAP we can confirm that the identified proteins are bound to chromatin. Furthermore, ChIP-SICAP provides the most specific validation of SICAP. Considering that DNA-labeling is carried out *in vitro*, SICAP and ChIP-SICAP can be generically applied to any cell

culture system as well as tissues.

Thirdly, I developed and applied TIGR for targeting the Nanog promoter by regular oligonucleotides probes. At the present time, I'm in the process of expanding this method to other loci and different species in order to evaluate the accuracy of the results, and possibly to improve the method.



# 8

## Material and Methods

---

### 8.1 Cell culture and cell fixation

46c mouse ES cells were grown feeder-free on 0.2% gelatinized cell culture plates in either traditional ES medium with serum or 2i medium. The serum medium contained DMEM high glucose (Life technologies, 11965-092) supplemented with 15% fetal bovine serum (Life technologies, 10270-106), 100  $\mu$  M MEM non-essential aminoacids (Life technologies, 11140-050), 1xGlutamax (Life technologies, 35050-061), 1x penicillin and streptomycin (Life technologies, 15140-122), 100 $\mu$ M of 2-mercaptoethanol (Sigma, M7522), and 200ng/ml of LIF (EMBL, protein expression core facility). The 2i+LIF (2iL)-medium contained DMEM/F12 medium for SILAC (Pierce, 88215), 100 $\mu$ MMEM non-essential aminoacids (Life technologies, 11140-050), 1xGlutamax (Life technologies, 35050-061), 1x penicillin and streptomycin (Life technologies, 15140-122), 100 $\mu$ M of 2-mercaptoethanol (Sigma, M7522), 0.5mg/ml of BSA (Sigma, A3059), 200ng/ml of LIF (EMBL, protein expression core facility), 1 $\mu$ M of PD0325901(Reagents Direct, 39-C68), 3 $\mu$ M of CHIR99021 (Reagents Direct, 27-H76). In addition, for light SILAC, 100 mg/ml of Lysine (L8662), 100 mg/ml of Arginine (Sigma, A6969) and 100 mg/ml of Proline (Sigma, P5607) were added to the 2i medium. For Heavy SILAC 100 mg/ml of 13C6,15N2-L-Lysine HCl (Silantes, 211604102), 100 mg/ml of 13C6,15N4-L-Arginine HCl (Silantes, 201604102) and 100 mg/ml of Proline (Sigma, P5607) were

added to the 2iL-medium. For cell fixation, the cells were harvested by Stempro Accutase (Life technologies, A11105-01), and spun 5min at 200g to remove the medium. Then the cell pellet was resuspended in 1.5% formaldehyde (Pierce, 28906) in PBS. After 15 min incubation at room temperature with occasional rotations, 125mM Glycine (Merck, 56-40-6) was added to the solution to quench cross-linking. Then the cells were washed twice with PBS, counted and aliquoted, and stored at  $-80^{\circ}\text{C}$  as dry cell pellets.

## 8.2 SICAP

About 4 million fixed ES cells were used per condition and replicate. The cells were resuspended in 900  $\mu\text{l}$  TE buffer plus Complete protease inhibitor (Roche, 11873580001), and kept on ice for 5min. Then 100  $\mu\text{l}$  Triton X-100 10% was added, and the cells were incubated another 5min on ice. Cells were washed twice by resuspending them in 10mM Tris-HCl (pH = 7.5), spinning afterward at 500g for 5min, and removing the supernatant. Cells were sonicated in 130  $\mu\text{l}$  of 10mM Tris-HCl using CovarisS220 for 430 seconds by Duty cycle: 10%, Intensity: 5 and Cycle/Burst: 200 to shear chromatin to an average size of 200 nucleotides, followed by centrifugation at 12000g for 10min to sediment cell debris, and collection of the supernatant. Then 30  $\mu\text{g}$  of RNase A (Life technologies, EN0531) was added to the clear supernatant, and incubated at  $37^{\circ}\text{C}$  for 30min. Then the sheared chromatin of the heavy SILAC sample was mixed with 1xTdT buffer, 50  $\mu\text{M}$  of Biotin-11-ddUTP (Jenabioscience, NU-1619-BIOX) and 60U of TdT (Life technologies, EP0161) to biotinylate DNA, while the sheared chromatin of the light SILAC (negative control) sample was mixed with 1xTdT buffer and 60U of TdT leaving out Biotin-11-ddUTP. The samples were incubated at  $37^{\circ}\text{C}$  for 30min with 1000 RPM agitation. To solubilize the precipitated sheared chromatin, 252  $\mu\text{l}$  Ampure XP (Beckman, A63880) was added to the suspension. After mixing and incubation at room temperature for 5min, the beads were adsorbed by a magnet, and the supernatant was removed. Then 700 $\mu\text{l}$  ethanol 70% (v/v) was

added to the beads on the magnet without resuspending them. The ethanol was removed, and this step was repeated once again. After that Tris-HCl 10mM was added to the beads to eluted DNA, and they were resuspended by a short sonication. The beads were then removed, and the supernatant was collected. Then heavy and light SILAC samples were mixed, and 1% SDS was added to the sheared chromatin. The volume was adjusted to 1.5ml with IP buffer (Tris-HCl 50mM pH = 7.5, Triton X100 1%, NP40 0.5%, EDTA pH= 8 5mM and NaCl 150mM). Then 50  $\mu$ l Streptavidin beads (NEB, S1420S) were added to the solution. After 30min rotation head-to-tail, the beads were successively washed 3times with 1% SDS, once with 2MNaCl, twice with 20% iso-propanol, and 5 times with 50% acetonitril. The beads were then resuspended in 0.1% RapiGest (Waters, 186001861) and 5mMDTT. To reverse the cross-links, the beads were boiled at 95°C for 20min. Finally, the beads were removed, and 10mM iodoacetamide was added to the solution. After 30min incubation, again 5mM DTT was added to the solution. Then 100ng Lys-C (Wako, 125-05061) and 500ng Trypsin (Promega, V5280) were added to the samples to digest the proteins.

To compare chromatin protein composition, 4 million ES cells grown in heavy SILAC 2iL medium were mixed with 4 million ES cells grown in serum medium and processed as described above.

### 8.3 ChIP-SICAP

Sheared chromatin was prepared as described above, however from 24 million ES cells. After sedimentation of the cell debris by spinning, the antibody was added to the sheared chromatin obtained from the heavy SILAC sample, while to the light SILAC sample no antibody was added to serve as a negative control. After an overnight agitation at 4°C, the samples were spun at 12000g for 10min. Then 90% of the solution was collected and the volume was adjusted to 1ml with the IP buffer. Then 30 $\mu$ l Dynabeads protein A or protein G (Life technologies, 10001D or 10003D, respectively depending on the antibody) were added to the samples. Following 2 hours of rotation, the beads were washed once

with 10mM Tris-HCl, and once with 1xTdT buffer. Then the beads were resuspended in 1xTdT buffer, 60 U of TdT and 50 $\mu$ M of Biotin-11-ddUTP, and incubated at 37°C for 30min. Subsequently the beads were washed 6 times with ice-cold IP buffer, resuspended in 200mM of DTT and 7.5% of SDS, and incubated at 37°C for 30min. Then the beads removed on the magnet, and the solution was collected. At this point the heavy and light labeled samples were mixed, and the volume was adjusted to 1.5ml by IP buffer. Then 50  $\mu$ l of Streptavidin beads were added, and the samples were rotated 30min at room temperature. Then the beads were washed, and proteins were digested as described in the SICAP procedure. To compare proteins co-localized with the bait protein in 2iL and serum media, 24 million ES cells grown in heavy SILAC 2iL medium were mixed with 24 million ES cells grown in serum medium with natural aminoacids. The rest of the procedure was performed as described above.

## 8.4 TIGR

In this experiment the Nanog promoter was targeted by two specific probes with the following sequences:

Nanog-519: CTTGACCTGAAACTTCCCCTAGAGATCGCCAGGG  
TCTGGAGGTGCA

Nanog-453: AGGTGCAGCCGTGGTTAAAAGATGAATAAAGTG  
AAATGAGGTAAAGCCTCTT

In addition, a non-targeting probe with the following sequence was used as a negative control:

Scram: AGGTGCAGCCGTGGTTAAAAGATGAATAAAGTGAAAT-  
GAGGTAAAGCCTCTT

All probes were biotinylated at the 5'-end via a linker called triethyleneglycol (TEG). About 24 million cells per probe. Probe Nanog-519 was applied to the heavy SILAC cells, Nanog-453 was applied to the intermediate SILAC cells, and Scram was applied to the light SILAC cells. The cells were split into 6 microtubes (4 million cells per tube). The cells in each tube were resuspended in 300  $\mu$ l of IP buffer, and

5 $\mu$ l RNase A (10 mg/ml, Fermentas EN0531) was added to each tube. Then the tubes were incubated at 37 °C for 15-30 min. After that 500  $\mu$ l of DNAzol (Life technologies, 10503-027) was added to the tubes. Then the tubes were vortexed vigorously for a few seconds, and the spun at 5000g for 2min. The supernatant was discarded, and the pellet was resuspended again in 500  $\mu$ l of DNAzol. Once again, the tubes were vortexed vigorously for a few seconds, and the spun at 5000g for 2min. The supernatant was discarded, and each tube was resuspended gently in 300 $\mu$ l of NaOH 25mM. Then the samples were incubated at 37 °C for 30 min, without agitation. After that, the tubes were sonicated using Bioruptor Plus (Diagenode), 14 cycles in the high mode, 30 seconds on, 30 seconds off. Then 1ul of the diluted the relevant probe (1 pmol/ $\mu$ l) was added to each tube, and the content of the tube was applied to an Amicon ultrafiltration column (30kD, Millipore, UFC503096). The columns were spun 5 min at 12000g, and then 300  $\mu$ l of BW buffer (Tris-HCl 10mM pH=7.5, EDTA 1mM, NaCl 1M) was added to the column. Again columns were spun 10 min at 12000g to reduce the volume to about 100ul. After that the liquid was collected, and poured into a 2-ml Eppendorf tube with 300  $\mu$ l of BW buffer. Then the tubes were incubated at 65 °C for 5 min, 600  $\mu$ l of BW buffer was added to each tube. After that the content of 3 tubes with heavy SILAC samples, 3 tubes with medium SILAC samples and 3 tubes with light SILAC samples were mixed in a 15 ml-Falcone tube (2 Falcone tubes, 9ml in each tube). Then 50ul of Streptavidin magnetic beads (NEB, S1420S) was added to each tube, and the tubes were rotated for 1-hour head to tail. After that the beads were washed 2 times with 10 ml of SDS wash buffer-1 (Tris-HCl 10mM pH=7.5, EDTA 1mM, NaCl 350 mM, SDS 0.5%); once with 10 ml of SDS wash buffer-2 (Tris-HCl 10mM pH=7.5, EDTA 1mM, NaCl 200 mM, SDS 1%); two times with 10 ml of iso-propanol wash buffer (2-propanol 20% in water (v/v), NaCl 500mM); four times with 10 ml of acetonitrile wash buffer (acetonitrile 40% in water (v/v), NaCl 500mM). Then beads were resuspended in 0.8 ml of the acetonitrile wash buffer, and transferred to a 2-ml Eppendorf tube. Finally the acetonitrile wash

buffer was discarded and the beads were digested. The beads were then resuspended in 0.1% RapiGest (Waters, 186001861) and 5mM DTT. To reverse the cross-links, the beads were boiled at 95°C for 20min. Finally, the beads were removed, and 10mMiodoacetamide was added to the solution. After 30min incubation, again 5mM DTT was added to the solution. Then 100ng Lys-C (Wako, 125-05061) and 500ng Trypsin (Promega, V5280) were added to the samples to digest the proteins.

## 8.5 High pH fractionation and mass spectrometric analysis

Following digestion of the proteins, the peptides were cleaned up using stage-tipping procedure [206]. In order to confine streptavidin peptides into discrete fractions in order to minimize interference with MS detection of genuine prey peptides, the samples were subjected to fractionation using high pH reversed-phase chromatography. Peptides were fractionated on an Agilent 1200 Infinity HPLC system with a Gemini C18 column (3  $\mu\text{m}$ , 110 Å, 100 1.0  $\times$  mm, Phenomenex) using a linear 60 min gradient from 0% to 35% (v/v) acetonitrile in 20 mM ammonium formate (pH 10) at a flow rate of 0.1 ml/min. Elution of peptides was detected with a variable wavelength UV detector set to 254 nm. Thirty two 1-min fractions were collected that were subsequently pooled into ten fractions using a post-concatenation strategy as previously described. Each fraction was then analyzed using LC-MS on a Orbitrap Velos Pro (Thermo Fisher Scientific) connected to a nanoAcquity UPLC via a nanoelectrospray ion source (Waters). Peptides were separated with a BEH300C18 (75  $\mu\text{m}$   $\times$  250 mm, 1.7  $\mu\text{m}$ ) UPLC column (Waters) using a stepwise 60-min or 240-min gradients, respectively depending on ChIP-SICAP or SICAP from 3% to 85% (v/v) acetonitrile in 0.1% (v/v) formic acid at a flow rate of 300 nl/min. The LTQ-Orbitrap Velos Pro mass spectrometer was operated in data-dependent mode, acquiring one survey MS scan in the orbitrap followed by up to 15 fragmentation scans (TOP15) of the most abundant ions analyzed in the LTQ by CID fragmentation. Only charge

states of two and higher were allowed for fragmentation. Essential MS settings were: full MS: AGC = 106, maximum ion time = 500 ms, m/z range = 375–1600, resolution = 30 000 FWHM; MS2: AGC = 30 000, maximum ion time = 50 ms, minimum signal threshold = 1500, dynamic exclusion time = 30 s, isolation width = 2 Da, normalized collision energy = 40, activation Q = 0.25. The MS spectra were analyzed using Proteome Discoverer 1.4 (Thermo Fisher Scientific), and the proteins were identified using MASCOT search engine (Matrix Science) against the *Mus musculus* proteome of the Uniprot database. The Percolator algorithm [207] was used to limit FDR rates to a q-value < 0.01

## 8.6 ChIP-seq

Shearing chromatin and IP were performed as described above for ChIP-SICAP, however omitting the DNA-labeling. In addition, an aliquot was taken from the clear sheared chromatin to be used as an input control. After the IP and six rounds of washing steps with the ice-cold IP buffer, the beads were resuspended in TE buffer plus 1% of SDS. Then the samples were boiled at 95°C for 20min, and 40µg proteinase K was added for protein digestion at 55°C for 30min. Next DNA was purified using phenol/chloroform isoamylalcohol and precipitated using glycogen and ethanol. Finally, DNA was resuspended in 30µl of Tris-HCl 10mM. To prepare the library for Illumina sequencing, purified ChIP DNA was end-repaired by Klenow, T4 DNA polymerase and T4 polynucleotide kinase. Then DNA fragments were subjected to A-tailing, and NEBNext adapter ligation (NEB Index Primers Set 1, E7335S). Following PCR for 12 cycles, the amplicons were size-selected by mixing 50µl PCR products with 30µl of Ampure XP beads. The supernatant was collected, and again 45 µl of AmpureXP beads was added. After 2 rounds of washing with 70% ethanol, the DNA was eluted in 50µl of Tris-HCl 10mM. Once again the eluted DNA was mixed with 48µl of AmpureXP beads, and after the washing, they were eluted by 15 µl of Tris-HCl 10mM. Sequencing was carried out by

Illumina HiSeq 2000 according to the manufacturer’s protocols.

## 8.7 ChIP-seq data analysis

Unless stated otherwise, analysis was performed in a local installation of Galaxy [208-210] maintained by the EMBL Genome Biology Computational Support. The 50-bp single-end reads were aligned to build version NCBI37/MM9 of the mouse genome using Bowtie version 2.0 [211] using standard options (Galaxy Tool version 0.2, sensitive preset). Reads failing to be mapped or mapping at several locations (as identified by the XS tag set by bowtie2) were removed using the “Filter SAM” tool and the “Select” tool, respectively. Read duplicates were identified and removed using Picard’s Mark Duplicates (<http://broadinstitute.github.io/picard>). Sequencing data quality was assessed using FastQC and the Deeptools package[212]. ChIP quality was estimated by cross-correlation using the “SPP” tool as suggested by ENCODE ChIP-seq guidelines[213]. Finally, reproducibility of ChIP replicates and final peak selection was achieved using the IDR pipeline of Landt et. al.[213], and implemented following instructions from <https://sites.google.com/site/anshulkundaje/projects/idr>.

We used MACS version 2 [214] as the underlying peak caller with recommended options (i.e. setting p-value cutoff to 1e-3). The final list of peaks used in the study corresponds to the “optimal” list (i.e. using peak called on merged replicates) at a 2% IDR cutoff. Heatmaps were produced using the Deeptools computeMatrix and heatmapmer tools on input-subtracted coverage files (see below). Coverage files (bigwig format) were generated using the Deeptools bam Coverage tool using the “Normalize coverage to 1x” option. Input subtraction (from ChIP signal) was performed using the Deeptools bigwig Compare tool. An average fragment size of 200 bp and bin sizes of 50 bp was systematically used. The annotation of the peaks and the nearby genes were carried out using the R package ChIPseeker [215].



## 8.8 List of the antibodies used for the IP

Nanog (D2A3) XP Rabbit mAb (Cell Signaling Tech, 8822 S) was applied 1:100 (v/v) to the sheared chromatin obtained from 24 million cells for the ChIP-SICAP assay. Oct-4A (C30A3C1) Rabbit mAb (Cell Signaling Tech, 5677 S) was applied 1:50 (v/v) to the sheared chromatin obtained from 24 million cells for the ChIP-SICAP assay. Human SOX2 Affinity Purified Polyclonal Ab (R&D Systems, AF2018) 25  $\mu$ g was applied to the sheared chromatin obtained from 24 million cells for the ChIP-SICAP assay. E-Cadherin (24E10) Rabbit mAb (Cell Signaling Tech, 3195s) was applied 1:50 (v/v) to the sheared chromatin obtained from 24 million cells for the ChIP-SICAP assay. Dazl antibody (Cell Signaling Tech, 8042 S) was applied 1: 50 (v/v) to the sheared chromatin obtained from 24 million cells for the ChIP-SICAP assay. Trim24/TIF1a antibody (Bethyl lab, A300-815A) 2.5  $\mu$ g was applied to the sheared chromatin obtained from 24 million cells for the ChIP-Seq assay. Mouse Monoclonal Fgf4 antibody (R&D systems, MAB58461) was used for immuno-fluorescent microscopy; diluted 1:200 in BSA 1% in PBS-T. HA-tag (C29F4) Rabbit monoclonal antibody (Cell Signaling Tech, 3724S) was used for immuno-fluorescent microscopy; diluted 1:500 in BSA 1% in PBS-T. Anti-rabbit IgG conjugated with Alexa Fluor 555 (Cell Signaling Tech, 4413S) was used as the secondary antibody for immuno-fluorescent microscopy; diluted 1:500 in BSA 1% in PBS with 0.1% Tween20. Anti-rat IgG conjugated with Alexa Fluor 488 (A-11006) was used as the secondary antibody for immuno-fluorescent microscopy; diluted 1:500 in BSA 1% in PBS with 0.1% Tween20 .

## 8.9 GO analysis

Annotations of the genes were determined either using Perseus software[216], or were downloaded directly from Uniprot database. Additionally sub-cellular localizations based on HPA antibodies were downloaded from the website: <http://www.proteinatlas.org/>; GO biological processes and their enrichments were determined using DAVID bioin-

formatics resources[217]

## 8.10 HPA antibodies and immuno-fluorescence

In this study, 159 primary antibodies validated within the Human Protein Atlas (HPA) project [126] were selected that detect proteins encoded by 134 different genes in Human cells. The antibodies were selected either by a high sequence similarity (>49%) in the antigenic region of the human and murine orthologs or show a high similarity in the staining of human and murine cell lines, respectively.

Immuno-fluorescent stainings were prepared in 96-well glass bottom plates (Greiner Bio-One, 655891) coated with gelatin. 30000 mES cells cultivated in 2i medium were seeded per well and incubated for 24h. Cell fixation, permeabilization, and immuno-fluorescent staining were performed according to the standard protocol at the HPA [218]. Briefly, cells were washed with PBS prior fixation for 15 min with 4% ice-cold paraformaldehyde in growth medium supplemented with 10% FBS or in PBS, and permeabilized with Triton X-100 in PBS for 15 min. The cells were then incubated with the primary antibodies, and/or anti-alpha-Tubulin (Abcam, ab7291) diluted according to the provider's recommendation overnight at 4 °. After 3-5 washing with PBS, the cells were incubated with the secondary antibodies for 1-1.5h. Following a nuclei-staining with DAPI for a few min. After 5 washing, the wells were filled with PBS and/or glycerol.

## 8.11 Treating the ES cells with the HA-tagged Fgf4

Fgf4 cDNA was amplified by PCR using the Forward primer:

5'-TTTGGATCCGCCACCATGGCGAAACGCGGGCCGA-3'

and the reverse primer:

5'-TTTGTCGACTCAGGCATAGTCAGGCACGTCATAAGGATATC  
CCAGTCTAGGAAGGAAGTG-3'.

In fact the reverse primer contained the HA-tag coding sequence. The

PCR product was digested by BamHI and Sall, and then it was cloned into pLV-mCherry (pLV-mCherry was a gift from Pantelis Tsoulfas, Addgene plasmid # 36084) using the same enzymes instead of mCherry. The coding sequence was then confirmed by Sanger sequencing. Subsequently HEK293T cells were transfected by pLV-Fgf4, psPAX2, and VSV-G. To do that, 8 million cells were seeded in a T75 flask, after an overnight incubation the cells were transfected using FuGENE HD. After 8 hours the medium of the cells were changed with the fresh one. About 2 days after the transfection, the virus-containing medium was collected, filtered, and concentrated using Amicon 100kD ultrafiltration tubes. The concentrated medium was used to infect a new batch of HEK293T cells. The transduced HEK293T cells were expanded to a 10cm dish. When the cells became confluent, the medium of the cells was replaced with the medium containing 1% FBS for 24 hours. Then the medium was collected, filtered, and concentrated by Amicon ultrafiltration tube 3kD from 10ml to 0.5ml. The ES cells were seeded in a 96-well plate using 2i medium. During the seeding 20ul of the concentrated HA-Fgf4 was added to the medium of the cells. After 24 hours, again 20ul of the concentrated HA-Fgf4 was added to the medium of the cells. Finally, after 2 hours, the cells were fixed using formaldehyde 4%, and subjected to immuno-fluorescent microscopy using anti-HA antibody as described previously.

## 8.12 Nup210 knockdown

As it has been previously published [178], HEK293T cells were transfected by psPAX2, VSV-G and a lentiviral vector: Nup210 shRNAs (TRCN0000101935 or TRCN0000101938), pLKO.1 scrambled shRNA or the mock vector. One overnight after the transfection, the medium of the cells were changed with the fresh medium with 1% FBS. About 2 days after the transfection, the medium of the cells were collected, filtered, and concentrated by Amicon ultra-15 centrifugal filter unit 100kD. Then regular MEF cells were infected with different volumes of the concentrated viruses to measure the integration copy number

of the viruses into the genome of the host cells by quantitative PCR, hence, the titer of viruses was calculated for the next step. The primer sequences were as follows:

primer name	sequence (5'-3')
Rpph1-F	AGTCTGAATTGGGTTATGAGG
Rpph1-R	AGGTGAGTTCCCAGAGAGCA
pLKO-F	GAATAGTAGACATAATAGCAACAGAC
pLKO-R	CCAAAGTGGATCTCTGCTGTC.

About  $10^6$  reprogrammable MEF cells [185] in passage 2 were split into a 6-well plate. The day after seeding, the medium of the cells were changed with the fresh one plus 4 mg/ml polybrene, and the cells were infected with equal titer of viruses. One day later, the viral medium were replaced with fresh medium. Two days after the infection the secondary MEF cells were seeded on 3wells with feeder layer, and over-expression of the reprogramming factors were induced by adding  $1\mu\text{g/ml}$  of Doxycycline to the medium. After 15 days, the number of iPS colonies was compared.

### 8.13 qRT-PCR and List of the qPCR primers

As it has been previously published [178], to do qRT-PCR RNA was extracted by QIAGEN RNeasy mini kit. After measuring the concentrations, RNA was treated with 2U turbo DNase Ambion for about 30 min. Then DNase was inactivated according to the manual, and cDNA was synthesized by Takara PrimeScript RT Master Mix (Perfect Real Time) in  $20\mu\text{l}$ . Real-time PCR was carried out by ABI7500 using ABI SYBR green master mix. The thermal condition of the PCR was as follows: 95 °C 10 min, 95 °C 15 sec, 60 °C 30 sec, 72 °C 30 sec. Melt curve analysis was then used as a test of specificity. The primer sequences for the amplification of the targets were as follows:

<b>primer name</b>	<b>sequence (5'-3')</b>
Nup210-F	ATGTGGTTGAAGCCGGATAC
Nup210-R	AAAACCTCAGCAGGAAGCACAG
Slug-F	ATGCCCAGTCTAGGAAATCG
Slug-R	AGTGTGAGTTCTAATGTGTCCTTG
Snail-F	CGGATGTGAAGAGATAACCAGTG
Snail-R	GAAGATGCCAGCGAGGATG
Ecad-F	CAGTGGTCTTTCAGCTCCTTC
Ecad-R	AAGGCACAGTTTATATCTCAGCA
Epcam-F	TCTGGCACTGCTAACTGGTAC
Epcam-R	CTTTCAGATAGGGTGTGGATAG
ActB-F	ACAGGATGCAGAAGGAGATTAC
ActB-R	TAAAACGCAGCTCAGTAACAGTC



## References

- [1] T. Graf, and T. Enver, “Forcing cells to change lineages,” *Nature*, vol. 462, no. 7273, pp. 587-94, Dec 3, 2009.
- [2] A. D. Goldberg, C. D. Allis, and E. Bernstein, “Epigenetics: a landscape takes shape,” *Cell*, vol. 128, no. 4, pp. 635-8, Feb 23, 2007.
- [3] W. Xie, M. D. Schultz, R. Lister, Z. Hou, N. Rajagopal, P. Ray, J. W. Whitaker, S. Tian, R. D. Hawkins, D. Leung, H. Yang, T. Wang, A. Y. Lee, S. A. Swanson, J. Zhang, Y. Zhu, A. Kim, J. R. Nery, M. A. Urich, S. Kuan, C. A. Yen, S. Klugman, P. Yu, K. Suknuntha, N. E. Propson, H. Chen, L. E. Edsall, U. Wagner, Y. Li, Z. Ye, A. Kulkarni, Z. Xuan, W. Y. Chung, N. C. Chi, J. E. Antosiewicz-Bourget, I. Slukvin, R. Stewart, M. Q. Zhang, W. Wang, J. A. Thomson, J. R. Ecker, and B. Ren, “Epigenomic analysis of multilineage differentiation of human embryonic stem cells,” *Cell*, vol. 153, no. 5, pp. 1134-48, May 23, 2013.
- [4] J. Zhu, M. Adli, J. Y. Zou, G. Verstappen, M. Coyne, X. Zhang, T. Durham, M. Miri, V. Deshpande, P. L. De Jager, D. A. Bennett, J. A. Houmard, D. M. Muoio, T. T. Onder, R. Camahort, C. A. Cowan, A. Meissner, C. B. Epstein, N. Shoresh, and B. E. Bernstein, “Genome-wide chromatin state transitions associated with developmental and environmental cues,” *Cell*, vol. 152, no. 3, pp. 642-54, Jan 31, 2013.
- [5] D. Jost, “Bifurcation in epigenetics: implications in development, proliferation, and diseases,” *Phys Rev E Stat Nonlin Soft Matter Phys*, vol. 89, no. 1, pp. 010701, Jan, 2014.
- [6] Z. Shipony, Z. Mukamel, N. M. Cohen, G. Landan, E. Chomsky, S. R. Zeligler, Y. C. Fried, E. Ainfunder, N. Friedman, and A. Tanay, “Dynamic and static maintenance of epigenetic memory in pluripotent and somatic cells,” *Nature*, vol. 513, no. 7516, pp. 115-9, Sep 4, 2014.
- [7] M. Abad, L. Mosteiro, C. Pantoja, M. Canamero, T. Rayon, I. Ors, O. Grana, D. Megias, O. Dominguez, D. Martinez, M. Manzanares, S. Ortega, and M. Serrano, “Reprogramming in vivo produces teratomas and iPS cells with totipotency features,” *Nature*, vol. 502, no. 7471, pp. 340-5, Oct 17, 2013.

- 
- [8] J. B. Gurdon, T. R. Elsdale, and M. Fischberg, "Sexually mature individuals of *Xenopus laevis* from the transplantation of single somatic nuclei," *Nature*, vol. 182, no. 4627, pp. 64-5, Jul 5, 1958.
- [9] J. B. Gurdon, "The developmental capacity of nuclei taken from differentiating endoderm cells of *Xenopus laevis*," *J Embryol Exp Morphol*, vol. 8, pp. 505-26, Dec, 1960.
- [10] J. B. Gurdon, "The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles," *J Embryol Exp Morphol*, vol. 10, pp. 622-40, Dec, 1962.
- [11] R. A. Miller, and F. H. Ruddle, "Pluripotent teratocarcinoma-thymus somatic cell hybrids," *Cell*, vol. 9, no. 1, pp. 45-55, Sep, 1976.
- [12] D. Solter, "From teratocarcinomas to embryonic stem cells and beyond: a history of embryonic stem cell research," *Nat Rev Genet*, vol. 7, no. 4, pp. 319-27, Apr, 2006.
- [13] T. P. Zwaka, and J. A. Thomson, "A germ cell origin of embryonic stem cells?," *Development*, vol. 132, no. 2, pp. 227-33, Jan, 2005.
- [14] K. Takahashi, and S. Yamanaka, "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors," *Cell*, vol. 126, no. 4, pp. 663-76, Aug 25, 2006.
- [15] M. Ieda, J. D. Fu, P. Delgado-Olguin, V. Vedantham, Y. Hayashi, B. G. Bruneau, and D. Srivastava, "Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors," *Cell*, vol. 142, no. 3, pp. 375-86, Aug 6, 2010.
- [16] T. Vierbuchen, A. Ostermeier, Z. P. Pang, Y. Kokubu, T. C. Sudhof, and M. Wernig, "Direct conversion of fibroblasts to functional neurons by defined factors," *Nature*, vol. 463, no. 7284, pp. 1035-41, Feb 25, 2010.
- [17] P. Huang, Z. He, S. Ji, H. Sun, D. Xiang, C. Liu, Y. Hu, X. Wang, and L. Hui, "Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors," *Nature*, vol. 475, no. 7356, pp. 386-9, Jul 21, 2011.
- [18] R. Jaenisch, and R. Young, "Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming," *Cell*, vol. 132, no. 4, pp. 567-82, Feb 22, 2008.
- [19] M. Boiani, S. Eckardt, H. R. Scholer, and K. J. McLaughlin, "Oct4 distribution and level in mouse clones: consequences for pluripotency," *Genes Dev*, vol. 16, no. 10, pp. 1209-19, May 15, 2002.
- [20] G. Kustatscher, N. Hegarat, K. L. Wills, C. Furlan, J. C. Bukowski-Wills, H. Hochegger, and J. Rappsilber, "Proteomics of a fuzzy organelle:



- 
- interphase chromatin,” *EMBO J*, vol. 33, no. 6, pp. 648-64, Mar 18, 2014.
- [21] Q. L. Ying, J. Wray, J. Nichols, L. Batlle-Morera, B. Doble, J. Woodgett, P. Cohen, and A. Smith, “The ground state of embryonic stem cell self-renewal,” *Nature*, vol. 453, no. 7194, pp. 519-23, May 22, 2008.
- [22] J. Nichols, and A. Smith, “Naive and primed pluripotent states,” *Cell Stem Cell*, vol. 4, no. 6, pp. 487-92, Jun 5, 2009.
- [23] J. Wray, T. Kalkan, and A. G. Smith, “The ground state of pluripotency,” *Biochem Soc Trans*, vol. 38, no. 4, pp. 1027-32, Aug, 2010.
- [24] E. Habibi, A. B. Brinkman, J. Arand, L. I. Kroeze, H. H. Kerstens, F. Matarese, K. Lepikhov, M. Gut, I. Brun-Heath, N. C. Hubner, R. Benedetti, L. Altucci, J. H. Jansen, J. Walter, I. G. Gut, H. Marks, and H. G. Stunnenberg, “Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells,” *Cell Stem Cell*, vol. 13, no. 3, pp. 360-9, Sep 5, 2013.
- [25] S. D. Byrum, A. Raman, S. D. Taverna, and A. J. Tackett, “ChAP-MS: a method for identification of proteins and histone posttranslational modifications at a single genomic locus,” *Cell Rep*, vol. 2, no. 1, pp. 198-205, Jul 26, 2012.
- [26] S. D. Byrum, S. D. Taverna, and A. J. Tackett, “Purification of a specific native genomic locus for proteomic analysis,” *Nucleic Acids Res*, vol. 41, no. 20, pp. e195, Nov, 2013.
- [27] Z. J. Waldrip, S. D. Byrum, A. J. Storey, J. Gao, A. K. Byrd, S. G. Mackintosh, W. P. Wahls, S. D. Taverna, K. D. Raney, and A. J. Tackett, “A CRISPR-based approach for proteomic analysis of a single genomic locus,” *Epigenetics*, vol. 9, no. 9, pp. 1207-11, Sep, 2014.
- [28] J. Yu, M. A. Vodyanik, K. Smuga-Otto, J. Antosiewicz-Bourget, J. L. Frane, S. Tian, J. Nie, G. A. Jonsdottir, V. Ruotti, R. Stewart, Slukvin, II, and J. A. Thomson, “Induced pluripotent stem cell lines derived from human somatic cells,” *Science*, vol. 318, no. 5858, pp. 1917-20, Dec 21, 2007.
- [29] B. Feng, J. Jiang, P. Kraus, J. H. Ng, J. C. Heng, Y. S. Chan, L. P. Yaw, W. Zhang, Y. H. Loh, J. Han, V. B. Vega, V. Cacheux-Rataboul, B. Lim, T. Lufkin, and H. H. Ng, “Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb,” *Nat Cell Biol*, vol. 11, no. 2, pp. 197-203, Feb, 2009.
- [30] J. C. Heng, B. Feng, J. Han, J. Jiang, P. Kraus, J. H. Ng, Y. L. Orlov, M. Huss, L. Yang, T. Lufkin, B. Lim, and H. H. Ng, “The nuclear receptor
-

---

Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells,” *Cell Stem Cell*, vol. 6, no. 2, pp. 167-74, Feb 5, 2010.

[31] Y. Buganim, D. A. Faddah, A. W. Cheng, E. Itskovich, S. Markoulaki, K. Ganz, S. L. Klemm, A. van Oudenaarden, and R. Jaenisch, “Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase,” *Cell*, vol. 150, no. 6, pp. 1209-22, Sep 14, 2012.

[32] F. Anokye-Danso, C. M. Trivedi, D. Juhr, M. Gupta, Z. Cui, Y. Tian, Y. Zhang, W. Yang, P. J. Gruber, J. A. Epstein, and E. E. Morrisey, “Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency,” *Cell Stem Cell*, vol. 8, no. 4, pp. 376-88, Apr 8, 2011.

[33] P. Hou, Y. Li, X. Zhang, C. Liu, J. Guan, H. Li, T. Zhao, J. Ye, W. Yang, K. Liu, J. Ge, J. Xu, Q. Zhang, Y. Zhao, and H. Deng, “Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds,” *Science*, vol. 341, no. 6146, pp. 651-4, Aug 9, 2013.

[34] T. W. Theunissen, and R. Jaenisch, “Molecular control of induced pluripotency,” *Cell Stem Cell*, vol. 14, no. 6, pp. 720-34, Jun 5, 2014.

[35] N. Y. Chia, Y. S. Chan, B. Feng, X. Lu, Y. L. Orlov, D. Moreau, P. Kumar, L. Yang, J. Jiang, M. S. Lau, M. Huss, B. S. Soh, P. Kraus, P. Li, T. Lufkin, B. Lim, N. D. Clarke, F. Bard, and H. H. Ng, “A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity,” *Nature*, vol. 468, no. 7321, pp. 316-20, Nov 11, 2010.

[36] K. Adachi, and H. R. Scholer, “Directing reprogramming to pluripotency by transcription factors,” *Curr Opin Genet Dev*, vol. 22, no. 5, pp. 416-22, Oct, 2012.

[37] H. Niwa, “The pluripotency transcription factor network at work in reprogramming,” *Curr Opin Genet Dev*, vol. 28, pp. 25-31, Oct, 2014.

[38] L. Magnani, J. Eeckhoutte, and M. Lupien, “Pioneer factors: directing transcriptional regulators within the chromatin environment,” *Trends Genet*, vol. 27, no. 11, pp. 465-74, Nov, 2011.

[39] A. Soufi, G. Donahue, and K. S. Zaret, “Facilitators and impediments of the pluripotency reprogramming factors’ initial engagement with the genome,” *Cell*, vol. 151, no. 5, pp. 994-1004, Nov 21, 2012.

[40] K. S. Zaret, and J. S. Carroll, “Pioneer transcription factors: establishing competence for gene expression,” *Genes Dev*, vol. 25, no. 21, pp. 2227-41, Nov 1, 2011.

[41] L. Ho, J. L. Ronan, J. Wu, B. T. Staahl, L. Chen, A. Kuo, J. Lessard,

- 
- A. I. Nesvizhskii, J. Ranish, and G. R. Crabtree, “An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency,” *Proc Natl Acad Sci U S A*, vol. 106, no. 13, pp. 5181-6, Mar 31, 2009.
- [42] L. Wang, Y. Du, J. M. Ward, T. Shimbo, B. Lackford, X. Zheng, Y. L. Miao, B. Zhou, L. Han, D. C. Fargo, R. Jothi, C. J. Williams, P. A. Wade, and G. Hu, “INO80 facilitates pluripotency gene activation in embryonic stem cell self-renewal, reprogramming, and blastocyst development,” *Cell Stem Cell*, vol. 14, no. 5, pp. 575-91, May 1, 2014.
- [43] N. Singhal, J. Graumann, G. Wu, M. J. Arauzo-Bravo, D. W. Han, B. Greber, L. Gentile, M. Mann, and H. R. Scholer, “Chromatin-Remodeling Components of the BAF Complex Facilitate Reprogramming,” *Cell*, vol. 141, no. 6, pp. 943-55, Jun 11, 2010.
- [44] K. Kaji, I. M. Caballero, R. MacLeod, J. Nichols, V. A. Wilson, and B. Hendrich, “The NuRD component Mbd3 is required for pluripotency of embryonic stem cells,” *Nat Cell Biol*, vol. 8, no. 3, pp. 285-92, Mar, 2006.
- [45] Y. Rais, A. Zviran, S. Geula, O. Gafni, E. Chomsky, S. Viukov, A. A. Mansour, I. Caspi, V. Krupalnik, M. Zerbib, I. Maza, N. Mor, D. Baran, L. Weinberger, D. A. Jaitin, D. Lara-Astiaso, R. Blecher-Gonen, Z. Shipony, Z. Mukamel, T. Hagai, S. Gilad, D. Amann-Zalcenstein, A. Tanay, I. Amit, N. Novershtern, and J. H. Hanna, “Deterministic direct reprogramming of somatic cells to pluripotency,” *Nature*, vol. 502, no. 7469, pp. 65-70, Oct 3, 2013.
- [46] M. Luo, T. Ling, W. Xie, H. Sun, Y. Zhou, Q. Zhu, M. Shen, L. Zong, G. Lyu, Y. Zhao, T. Ye, J. Gu, W. Tao, Z. Lu, and I. Grummt, “NuRD blocks reprogramming of mouse somatic cells into pluripotent stem cells,” *Stem Cells*, vol. 31, no. 7, pp. 1278-86, Jul, 2013.
- [47] R. L. dos Santos, L. Tosti, A. Radziskeuskaya, I. M. Caballero, K. Kaji, B. Hendrich, and J. C. Silva, “MBD3/NuRD facilitates induction of pluripotency in a context-dependent manner,” *Cell Stem Cell*, vol. 15, no. 1, pp. 102-10, Jul 3, 2014.
- [48] B. Nashun, P. W. Hill, and P. Hajkova, “Reprogramming of cell fate: epigenetic memory and the erasure of memories past,” *EMBO J*, vol. 34, no. 10, pp. 1296-308, May 12, 2015.
- [49] T. Shinagawa, T. Takagi, D. Tsukamoto, C. Tomaru, L. M. Huynh, P. Sivaraman, T. Kumarevel, K. Inoue, R. Nakato, Y. Katou, T. Sado, S. Takahashi, A. Ogura, K. Shirahige, and S. Ishii, “Histone variants enriched
-

---

in oocytes enhance reprogramming to induced pluripotent stem cells,” *Cell Stem Cell*, vol. 14, no. 2, pp. 217-27, Feb 6, 2014.

[50] A. Gaspar-Maia, Z. A. Qadeer, D. Hasson, K. Ratnakumar, N. A. Leu, G. Leroy, S. Liu, C. Costanzi, D. Valle-Garcia, C. Schaniel, I. Lemischka, B. Garcia, J. R. Pehrson, and E. Bernstein, “MacroH2A histone variants act as a barrier upon reprogramming towards pluripotency,” *Nat Commun*, vol. 4, pp. 1565, 2013.

[51] C. Creppe, P. Janich, N. Cantarino, M. Noguera, V. Valero, E. Musulen, J. Douet, M. Posavec, J. Martin-Caballero, L. Sumoy, L. Di Croce, S. A. Benitah, and M. Buschbeck, “MacroH2A1 regulates the balance between self-renewal and differentiation commitment in embryonic and adult stem cells,” *Mol Cell Biol*, vol. 32, no. 8, pp. 1442-52, Apr, 2012.

[52] Y. S. Ang, S. Y. Tsai, D. F. Lee, J. Monk, J. Su, K. Ratnakumar, J. Ding, Y. Ge, H. Darr, B. Chang, J. Wang, M. Rendl, E. Bernstein, C. Schaniel, and I. R. Lemischka, “Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network,” *Cell*, vol. 145, no. 2, pp. 183-97, Apr 15, 2011.

[53] R. P. Koche, Z. D. Smith, M. Adli, H. Gu, M. Ku, A. Gnirke, B. E. Bernstein, and A. Meissner, “Reprogramming factor expression initiates widespread targeted chromatin remodeling,” *Cell Stem Cell*, vol. 8, no. 1, pp. 96-105, Jan 7, 2011.

[54] T. T. Onder, N. Kara, A. Cherry, A. U. Sinha, N. Zhu, K. M. Bernt, P. Cahan, B. O. Marcarci, J. Unternaehrer, P. B. Gupta, E. S. Lander, S. A. Armstrong, and G. Q. Daley, “Chromatin-modifying enzymes as modulators of reprogramming,” *Nature*, vol. 483, no. 7391, pp. 598-602, Mar 29, 2012.

[55] Y. Shi, J. T. Do, C. Desponts, H. S. Hahm, H. R. Scholer, and S. Ding, “A combined chemical and genetic approach for the generation of induced pluripotent stem cells,” *Cell Stem Cell*, vol. 2, no. 6, pp. 525-8, Jun 5, 2008.

[56] J. Chen, H. Liu, J. Liu, J. Qi, B. Wei, J. Yang, H. Liang, Y. Chen, J. Chen, Y. Wu, L. Guo, J. Zhu, X. Zhao, T. Peng, Y. Zhang, S. Chen, X. Li, D. Li, T. Wang, and D. Pei, “H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs,” *Nat Genet*, vol. 45, no. 1, pp. 34-42, Jan, 2013.

[57] R. J. Klose, E. M. Kallin, and Y. Zhang, “JmjC-domain-containing proteins and histone demethylation,” *Nat Rev Genet*, vol. 7, no. 9, pp. 715-27, Sep, 2006.

- 
- [58] A. A. Mansour, O. Gafni, L. Weinberger, A. Zviran, M. Ayyash, Y. Rais, V. Krupalnik, M. Zerbib, D. Amann-Zalcenstein, I. Maza, S. Geula, S. Viukov, L. Holtzman, A. Pribluda, E. Canaani, S. Horn-Saban, I. Amit, N. Novershtern, and J. H. Hanna, "The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming," *Nature*, vol. 488, no. 7411, pp. 409-13, Aug 16, 2012.
- [59] W. Zhao, Q. Li, S. Ayers, Y. Gu, Z. Shi, Q. Zhu, Y. Chen, H. Y. Wang, and R. F. Wang, "Jmjd3 inhibits reprogramming by upregulating expression of INK4a/Arf and targeting PHF20 for ubiquitination," *Cell*, vol. 152, no. 5, pp. 1037-50, Feb 28, 2013.
- [60] G. Frigola, P. L. Germain, P. Laise, A. Cuomo, A. Blasimme, F. Gross, E. Signaroldi, G. Bucci, C. Sommer, G. Pruneri, G. Mazzarol, T. Bonaldi, G. Mostoslavsky, S. Casola, and G. Testa, "Cell reprogramming requires silencing of a core subset of polycomb targets," *PLoS Genet*, vol. 9, no. 2, pp. e1003292, 2013.
- [61] H. Ma, R. Morey, R. C. O'Neil, Y. He, B. Daughtry, M. D. Schultz, M. Hariharan, J. R. Nery, R. Castanon, K. Sabatini, R. D. Thiagarajan, M. Tachibana, E. Kang, R. Tippner-Hedges, R. Ahmed, N. M. Gutierrez, C. Van Dyken, A. Polat, A. Sugawara, M. Sparman, S. Gokhale, P. Amato, D. P. Wolf, J. R. Ecker, L. C. Laurent, and S. Mitalipov, "Abnormalities in human pluripotent cells due to reprogramming mechanisms," *Nature*, vol. 511, no. 7508, pp. 177-83, Jul 10, 2014.
- [62] T. S. Mikkelsen, J. Hanna, X. Zhang, M. Ku, M. Wernig, P. Schorderet, B. E. Bernstein, R. Jaenisch, E. S. Lander, and A. Meissner, "Dissecting direct reprogramming through integrative genomic analysis," *Nature*, vol. 454, no. 7200, pp. 49-55, Jul 3, 2008.
- [63] M. Pawlak, and R. Jaenisch, "De novo DNA methylation by Dnmt3a and Dnmt3b is dispensable for nuclear reprogramming of somatic cells to a pluripotent state," *Genes Dev*, vol. 25, no. 10, pp. 1035-40, May 15, 2011.
- [64] C. A. Doege, K. Inoue, T. Yamashita, D. B. Rhee, S. Travis, R. Fujita, P. Guarnieri, G. Bhagat, W. B. Vanti, A. Shih, R. L. Levine, S. Nik, E. I. Chen, and A. Abeliovich, "Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2," *Nature*, vol. 488, no. 7413, pp. 652-5, Aug 30, 2012.
- [65] Y. Costa, J. Ding, T. W. Theunissen, F. Faiola, T. A. Hore, P. V. Shliha, M. Fidalgo, A. Saunders, M. Lawrence, S. Dietmann, S. Das, D. N. Levasseur, Z. Li, M. Xu, W. Reik, J. C. Silva, and J. Wang, "NANOG-
-

---

dependent function of TET1 and TET2 in establishment of pluripotency,” *Nature*, vol. 495, no. 7441, pp. 370-4, Mar 21, 2013.

[66] J. Chen, L. Guo, L. Zhang, H. Wu, J. Yang, H. Liu, X. Wang, X. Hu, T. Gu, Z. Zhou, J. Liu, J. Liu, H. Wu, S. Q. Mao, K. Mo, Y. Li, K. Lai, J. Qi, H. Yao, G. Pan, G. L. Xu, and D. Pei, “Vitamin C modulates TET1 function during somatic cell reprogramming,” *Nat Genet*, vol. 45, no. 12, pp. 1504-9, Dec, 2013.

[67] X. Hu, L. Zhang, S. Q. Mao, Z. Li, J. Chen, R. R. Zhang, H. P. Wu, J. Gao, F. Guo, W. Liu, G. F. Xu, H. Q. Dai, Y. G. Shi, X. Li, B. Hu, F. Tang, D. Pei, and G. L. Xu, “Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming,” *Cell Stem Cell*, vol. 14, no. 4, pp. 512-22, Apr 3, 2014.

[68] S. Ruiz, A. D. Panopoulos, A. Herrerias, K. D. Bissig, M. Lutz, W. T. Berggren, I. M. Verma, and J. C. Izpisua Belmonte, “A high proliferation rate is required for cell reprogramming and maintenance of human embryonic stem cell identity,” *Curr Biol*, vol. 21, no. 1, pp. 45-52, Jan 11, 2011.

[69] J. Hanna, K. Saha, B. Pando, J. van Zon, C. J. Lengner, M. P. Creighton, A. van Oudenaarden, and R. Jaenisch, “Direct cell reprogramming is a stochastic process amenable to acceleration,” *Nature*, vol. 462, no. 7273, pp. 595-601, Dec 3, 2009.

[70] J. Silva, J. Nichols, T. W. Theunissen, G. Guo, A. L. van Oosten, O. Barrandon, J. Wray, S. Yamanaka, I. Chambers, and A. Smith, “Nanog is the gateway to the pluripotent ground state,” *Cell*, vol. 138, no. 4, pp. 722-37, Aug 21, 2009.

[71] A. Bradley, M. Evans, M. H. Kaufman, and E. Robertson, “Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines,” *Nature*, vol. 309, no. 5965, pp. 255-6, May 17-23, 1984.

[72] A. Nagy, J. Rossant, R. Nagy, W. Abramow-Newerly, and J. C. Roder, “Derivation of completely cell culture-derived mice from early-passage embryonic stem cells,” *Proc Natl Acad Sci U S A*, vol. 90, no. 18, pp. 8424-8, Sep 15, 1993.

[73] J. A. Hackett, and M. A. Surani, “Regulatory principles of pluripotency: from the ground state up,” *Cell Stem Cell*, vol. 15, no. 4, pp. 416-30, Oct 2, 2014.

[74] A. G. Smith, J. K. Heath, D. D. Donaldson, G. G. Wong, J. Moreau, M. Stahl, and D. Rogers, “Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides,” *Nature*, vol. 336, no.

---

6200, pp. 688-90, Dec 15, 1988.

[75] R. L. Williams, D. J. Hilton, S. Pease, T. A. Willson, C. L. Stewart, D. P. Gearing, E. F. Wagner, D. Metcalf, N. A. Nicola, and N. M. Gough, "Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells," *Nature*, vol. 336, no. 6200, pp. 684-7, Dec 15, 1988.

[76] H. Niwa, "How is pluripotency determined and maintained?," *Development*, vol. 134, no. 4, pp. 635-46, Feb, 2007.

[77] G. Martello, P. Bertone, and A. Smith, "Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor," *EMBO J*, vol. 32, no. 19, pp. 2561-74, Oct 2, 2013.

[78] S. Ye, P. Li, C. Tong, and Q. L. Ying, "Embryonic stem cell self-renewal pathways converge on the transcription factor Tfcp2l1," *EMBO J*, vol. 32, no. 19, pp. 2548-60, Oct 2, 2013.

[79] Q. L. Ying, J. Nichols, I. Chambers, and A. Smith, "BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3," *Cell*, vol. 115, no. 3, pp. 281-92, Oct 31, 2003.

[80] I. Chambers, J. Silva, D. Colby, J. Nichols, B. Nijmeijer, M. Robertson, J. Vrana, K. Jones, L. Grotewold, and A. Smith, "Nanog safeguards pluripotency and mediates germline development," *Nature*, vol. 450, no. 7173, pp. 1230-4, Dec 20, 2007.

[81] K. Hayashi, S. M. Lopes, F. Tang, and M. A. Surani, "Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states," *Cell Stem Cell*, vol. 3, no. 4, pp. 391-401, Oct 9, 2008.

[82] Y. Toyooka, D. Shimosato, K. Murakami, K. Takahashi, and H. Niwa, "Identification and characterization of subpopulations in undifferentiated ES cell culture," *Development*, vol. 135, no. 5, pp. 909-18, Mar, 2008.

[83] E. Abranches, E. Bekman, and D. Henrique, "Generation and characterization of a novel mouse embryonic stem cell line with a dynamic reporter of Nanog expression," *PLoS One*, vol. 8, no. 3, pp. e59928, 2013.

[84] K. Ogawa, R. Nishinakamura, Y. Iwamatsu, D. Shimosato, and H. Niwa, "Synergistic action of Wnt and LIF in maintaining pluripotency of mouse ES cells," *Biochem Biophys Res Commun*, vol. 343, no. 1, pp. 159-66, Apr 28, 2006.

[85] N. Sato, L. Meijer, L. Skaltsounis, P. Greengard, and A. H. Brivanlou,

---

“Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor,” *Nat Med*, vol. 10, no. 1, pp. 55-63, Jan, 2004.

[86] J. Wray, T. Kalkan, S. Gomez-Lopez, D. Eckardt, A. Cook, R. Kemler, and A. Smith, “Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation,” *Nat Cell Biol*, vol. 13, no. 7, pp. 838-45, Jul, 2011.

[87] M. F. Cole, S. E. Johnstone, J. J. Newman, M. H. Kagey, and R. A. Young, “Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells,” *Genes Dev*, vol. 22, no. 6, pp. 746-55, Mar 15, 2008.

[88] F. Yi, L. Pereira, J. A. Hoffman, B. R. Shy, C. M. Yuen, D. R. Liu, and B. J. Merrill, “Opposing effects of Tcf3 and Tcf1 control Wnt stimulation of embryonic stem cell self-renewal,” *Nat Cell Biol*, vol. 13, no. 7, pp. 762-70, Jul, 2011.

[89] H. G. Leitch, K. R. McEwen, A. Turp, V. Encheva, T. Carroll, N. Grabole, W. Mansfield, B. Nashun, J. G. Knezovich, A. Smith, M. A. Surani, and P. Hajkova, “Naive pluripotency is associated with global DNA hypomethylation,” *Nat Struct Mol Biol*, vol. 20, no. 3, pp. 311-6, Mar, 2013.

[90] M. Yamaji, J. Ueda, K. Hayashi, H. Ohta, Y. Yabuta, K. Kurimoto, R. Nakato, Y. Yamada, K. Shirahige, and M. Saitou, “PRDM14 ensures naive pluripotency through dual regulation of signaling and epigenetic pathways in mouse embryonic stem cells,” *Cell Stem Cell*, vol. 12, no. 3, pp. 368-82, Mar 7, 2013.

[91] A. Kanda, Y. Sotomaru, S. Shiozawa, and E. Hiyama, “Establishment of ES cells from inbred strain mice by dual inhibition (2i),” *J Reprod Dev*, vol. 58, no. 1, pp. 77-83, 2012.

[92] J. Nichols, K. Jones, J. M. Phillips, S. A. Newland, M. Roode, W. Mansfield, A. Smith, and A. Cooke, “Validated germline-competent embryonic stem cell lines from nonobese diabetic mice,” *Nat Med*, vol. 15, no. 7, pp. 814-8, Jul, 2009.

[93] D. ten Berge, D. Kurek, T. Blauwkamp, W. Koole, A. Maas, E. Eroglu, R. K. Siu, and R. Nusse, “Embryonic stem cells require Wnt proteins to prevent differentiation to epiblast stem cells,” *Nat Cell Biol*, vol. 13, no. 9, pp. 1070-5, Sep, 2011.

[94] T. Boroviak, R. Loos, P. Bertone, A. Smith, and J. Nichols, “The ability



---

of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification,” *Nat Cell Biol*, vol. 16, no. 6, pp. 516-28, Jun, 2014.

[95] H. Marks, T. Kalkan, R. Menafrá, S. Denissov, K. Jones, H. Hofemeister, J. Nichols, A. Kranz, A. F. Stewart, A. Smith, and H. G. Stunnenberg, “The transcriptional and epigenomic foundations of ground state pluripotency,” *Cell*, vol. 149, no. 3, pp. 590-604, Apr 27, 2012.

[96] G. Martello, T. Sugimoto, E. Diamanti, A. Joshi, R. Hannah, S. Ohtsuka, B. Gottgens, H. Niwa, and A. Smith, “Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal,” *Cell Stem Cell*, vol. 11, no. 4, pp. 491-504, Oct 5, 2012.

[97] J. C. Yeo, and H. H. Ng, “The transcriptional regulation of pluripotency,” *Cell Res*, vol. 23, no. 1, pp. 20-32, Jan, 2013.

[98] G. Ficz, T. A. Hore, F. Santos, H. J. Lee, W. Dean, J. Arand, F. Krueger, D. Oxley, Y. L. Paul, J. Walter, S. J. Cook, S. Andrews, M. R. Branco, and W. Reik, “FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency,” *Cell Stem Cell*, vol. 13, no. 3, pp. 351-9, Sep 5, 2013.

[99] S. A. Smallwood, S. Tomizawa, F. Krueger, N. Ruf, N. Carli, A. Segonds-Pichon, S. Sato, K. Hata, S. R. Andrews, and G. Kelsey, “Dynamic CpG island methylation landscape in oocytes and preimplantation embryos,” *Nat Genet*, vol. 43, no. 8, pp. 811-4, Aug, 2011.

[100] Z. D. Smith, M. M. Chan, T. S. Mikkelsen, H. Gu, A. Gnirke, A. Regev, and A. Meissner, “A unique regulatory phase of DNA methylation in the early mammalian embryo,” *Nature*, vol. 484, no. 7394, pp. 339-44, Apr 19, 2012.

[101] N. Grabole, J. Tischler, J. A. Hackett, S. Kim, F. Tang, H. G. Leitch, E. Magnusdottir, and M. A. Surani, “Prdm14 promotes germline fate and naive pluripotency by repressing FGF signalling and DNA methylation,” *EMBO Rep*, vol. 14, no. 7, pp. 629-37, Jul, 2013.

[102] N. Okashita, Y. Kumaki, K. Ebi, M. Nishi, Y. Okamoto, M. Nakayama, S. Hashimoto, T. Nakamura, K. Sugawara, N. Kojima, T. Takada, M. Okano, and Y. Seki, “PRDM14 promotes active DNA demethylation through the ten-eleven translocation (TET)-mediated base excision repair pathway in embryonic stem cells,” *Development*, vol. 141, no. 2, pp. 269-80, Jan, 2014.

[103] E. G. Schulz, J. Meisig, T. Nakamura, I. Okamoto, A. Sieber, C.

- 
- Picard, M. Borensztein, M. Saitou, N. Bluthgen, and E. Heard, "The two active X chromosomes in female ESCs block exit from the pluripotent state by modulating the ESC signaling network," *Cell Stem Cell*, vol. 14, no. 2, pp. 203-16, Feb 6, 2014.
- [104] W. W. Tee, S. S. Shen, O. Oksuz, V. Narendra, and D. Reinberg, "Erk1/2 activity promotes chromatin features and RNAPII phosphorylation at developmental promoters in mouse ESCs," *Cell*, vol. 156, no. 4, pp. 678-90, Feb 13, 2014.
- [105] M. A. Christophorou, G. Castelo-Branco, R. P. Halley-Stott, C. S. Oliveira, R. Loos, A. Radziszheuskaya, K. A. Mowen, P. Bertone, J. C. Silva, M. Zernicka-Goetz, M. L. Nielsen, J. B. Gurdon, and T. Kouzarides, "Cit-rullination regulates pluripotency and histone H1 binding to chromatin," *Nature*, vol. 507, no. 7490, pp. 104-8, Mar 6, 2014.
- [106] I. G. Brons, L. E. Smithers, M. W. Trotter, P. Rugg-Gunn, B. Sun, S. M. Chuva de Sousa Lopes, S. K. Howlett, A. Clarkson, L. Ahrlund-Richter, R. A. Pedersen, and L. Vallier, "Derivation of pluripotent epiblast stem cells from mammalian embryos," *Nature*, vol. 448, no. 7150, pp. 191-5, Jul 12, 2007.
- [107] N. Festuccia, R. Osorno, F. Halbritter, V. Karwacki-Neisius, P. Navarro, D. Colby, F. Wong, A. Yates, S. R. Tomlinson, and I. Chambers, "Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells," *Cell Stem Cell*, vol. 11, no. 4, pp. 477-90, Oct 5, 2012.
- [108] J. P. Lambert, L. Mitchell, A. Rudner, K. Baetz, and D. Figeys, "A novel proteomics approach for the discovery of chromatin-associated protein networks," *Mol Cell Proteomics*, vol. 8, no. 4, pp. 870-82, Apr, 2009.
- [109] C. I. Wang, A. A. Alekseyenko, G. LeRoy, A. E. Elia, A. A. Gorchakov, L. M. Britton, S. J. Elledge, P. V. Kharchenko, B. A. Garcia, and M. I. Kuroda, "Chromatin proteins captured by ChIP-mass spectrometry are linked to dosage compensation in *Drosophila*," *Nat Struct Mol Biol*, vol. 20, no. 2, pp. 202-9, Feb, 2013.
- [110] M. Soldi, and T. Bonaldi, "The proteomic investigation of chromatin functional domains reveals novel synergisms among distinct heterochromatin components," *Mol Cell Proteomics*, vol. 12, no. 3, pp. 764-80, Mar, 2013.
- [111] J. Dejardin, and R. E. Kingston, "Purification of proteins associated with specific genomic Loci," *Cell*, vol. 136, no. 1, pp. 175-86, Jan 9, 2009.
- [112] N. Saksouk, T. K. Barth, C. Ziegler-Birling, N. Olova, A. Nowak,

---

E. Rey, J. Mateos-Langerak, S. Urbach, W. Reik, M. E. Torres-Padilla, A. Imhof, J. Dejardin, and E. Simboeck, "Redundant mechanisms to form silent chromatin at pericentromeric regions rely on BEND3 and DNA methylation," *Mol Cell*, vol. 56, no. 4, pp. 580-94, Nov 20, 2014.

[113] T. Mondal, M. Rasmussen, G. K. Pandey, A. Isaksson, and C. Kanduri, "Characterization of the RNA content of chromatin," *Genome Res*, vol. 20, no. 7, pp. 899-907, Jul, 2010.

[114] J. Bonner, M. E. Dahmus, D. Fambrough, R. C. Huang, K. Marushige, and D. Y. Tuan, "The Biology of Isolated Chromatin: Chromosomes, biologically active in the test tube, provide a powerful tool for the study of gene action," *Science*, vol. 159, no. 3810, pp. 47-56, Jan 5, 1968.

[115] C. D. Lewis, and U. K. Laemmli, "Higher order metaphase chromosome structure: evidence for metalloprotein interactions," *Cell*, vol. 29, no. 1, pp. 171-81, May, 1982.

[116] C. Morrison, A. J. Henzing, O. N. Jensen, N. Osheroff, H. Dodson, S. E. Kandels-Lewis, R. R. Adams, and W. C. Earnshaw, "Proteomic analysis of human metaphase chromosomes reveals topoisomerase II alpha as an Aurora B substrate," *Nucleic Acids Res*, vol. 30, no. 23, pp. 5318-27, Dec 1, 2002.

[117] C. Alabert, J. C. Bukowski-Wills, S. B. Lee, G. Kustatscher, K. Nakamura, F. de Lima Alves, P. Menard, J. Mejlvang, J. Rappsilber, and A. Groth, "Nascent chromatin capture proteomics determines chromatin dynamics during DNA replication and identifies unknown fork components," *Nat Cell Biol*, vol. 16, no. 3, pp. 281-93, Mar, 2014.

[118] J. D. Nelson, O. Denisenko, and K. Bomsztyk, "Protocol for the fast chromatin immunoprecipitation (ChIP) method," *Nat Protoc*, vol. 1, no. 1, pp. 179-85, 2006.

[119] M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, and G. K. Smyth, "limma powers differential expression analyses for RNA-sequencing and microarray studies," *Nucleic Acids Res*, vol. 43, no. 7, pp. e47, Apr 20, 2015.

[120] G. Peignon, S. Thenet, C. Schreider, S. Fouquet, A. Ribeiro, E. Dusaulx, J. Chambaz, P. Cardot, M. Pincon-Raymond, and J. Le Beyec, "E-cadherin-dependent transcriptional control of apolipoprotein A-IV gene expression in intestinal epithelial cells: a role for the hepatic nuclear factor 4," *J Biol Chem*, vol. 281, no. 6, pp. 3560-8, Feb 10, 2006.

[121] D. Xie, C. Gore, J. Liu, R. C. Pong, R. Mason, G. Hao, M. Long, W.

- 
- Kabbani, L. Yu, H. Zhang, H. Chen, X. Sun, D. A. Boothman, W. Min, and J. T. Hsieh, "Role of DAB2IP in modulating epithelial-to-mesenchymal transition and prostate cancer metastasis," *Proc Natl Acad Sci U S A*, vol. 107, no. 6, pp. 2485-90, Feb 9, 2010.
- [122] B. Zhong, Y. Yang, S. Li, Y. Y. Wang, Y. Li, F. Diao, C. Lei, X. He, L. Zhang, P. Tien, and H. B. Shu, "The adaptor protein MITA links virus-sensing receptors to IRF3 transcription factor activation," *Immunity*, vol. 29, no. 4, pp. 538-50, Oct 17, 2008.
- [123] S. Chi, G. Xie, H. Liu, K. Chen, X. Zhang, C. Li, and J. Xie, "Rab23 negatively regulates Gli1 transcriptional factor in a Su(Fu)-dependent manner," *Cell Signal*, vol. 24, no. 6, pp. 1222-8, Jun, 2012.
- [124] N. Yu, S. Zhang, F. Zuo, K. Kang, M. Guan, and L. Xiang, "Cultured human melanocytes express functional toll-like receptors 2-4, 7 and 9," *J Dermatol Sci*, vol. 56, no. 2, pp. 113-20, Nov, 2009.
- [125] L. Nathanson, and M. P. Deutscher, "Active aminoacyl-tRNA synthetases are present in nuclei as a high molecular weight multienzyme complex," *J Biol Chem*, vol. 275, no. 41, pp. 31559-62, Oct 13, 2000.
- [126] M. Uhlen, L. Fagerberg, B. M. Hallstrom, C. Lindskog, P. Oksvold, A. Mardinoglu, A. Sivertsson, C. Kampf, E. Sjostedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A. Szigartyo, J. Odeberg, D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P. H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, and F. Ponten, "Proteomics. Tissue-based map of the human proteome," *Science*, vol. 347, no. 6220, pp. 1260419, Jan 23, 2015.
- [127] E. C. Ferber, M. Kajita, A. Wadlow, L. Tobiansky, C. Niessen, H. Ariga, J. Daniel, and Y. Fujita, "A role for the cleaved cytoplasmic domain of E-cadherin in the nucleus," *J Biol Chem*, vol. 283, no. 19, pp. 12691-700, May 9, 2008.
- [128] M. Kos, and D. Tollervey, "Yeast pre-rRNA processing and modification occur cotranscriptionally," *Mol Cell*, vol. 37, no. 6, pp. 809-20, Mar 26, 2010.
- [129] F. J. Iborra, D. A. Jackson, and P. R. Cook, "Coupled transcription and translation within nuclei of mammalian cells," *Science*, vol. 293, no. 5532, pp. 1139-42, Aug 10, 2001.
- [130] A. David, B. P. Dolan, H. D. Hickman, J. J. Knowlton, G. Clavarino, P. Pierre, J. R. Bennink, and J. W. Yewdell, "Nuclear translation visualized

---

by ribosome-bound nascent chain puromycylation,” *J Cell Biol*, vol. 197, no. 1, pp. 45-57, Apr 2, 2012.

[131] E. A. Casanova, O. Shakhova, S. S. Patel, I. N. Asner, P. Pelczar, F. A. Weber, U. Graf, L. Sommer, K. Burki, and P. Cinelli, “Pramel7 mediates LIF/STAT3-dependent self-renewal in embryonic stem cells,” *Stem Cells*, vol. 29, no. 3, pp. 474-85, Mar, 2011.

[132] N. van Baren, H. Chambost, A. Ferrant, L. Michaux, H. Ikeda, I. Millard, D. Olive, T. Boon, and P. G. Coulie, “PRAME, a gene encoding an antigen recognized on a human melanoma by cytolytic T cells, is expressed in acute leukaemia cells,” *Br J Haematol*, vol. 102, no. 5, pp. 1376-9, Sep, 1998.

[133] B. V. Mistry, Y. Zhao, T. C. Chang, H. Yasue, M. Chiba, J. Oatley, F. Diaz, and W. S. Liu, “Differential expression of PRAMEL1, a cancer/testis antigen, during spermatogenesis in the mouse,” *PLoS One*, vol. 8, no. 4, pp. e60611, 2013.

[134] M. J. Solomon, P. L. Larsen, and A. Varshavsky, “Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene,” *Cell*, vol. 53, no. 6, pp. 937-47, Jun 17, 1988.

[135] T. R. Hebbes, A. W. Thorne, and C. Crane-Robinson, “A direct link between core histone acetylation and transcriptionally active chromatin,” *EMBO J*, vol. 7, no. 5, pp. 1395-402, May, 1988.

[136] B. Ren, F. Robert, J. J. Wyrick, O. Aparicio, E. G. Jennings, I. Simon, J. Zeitlinger, J. Schreiber, N. Hannett, E. Kanin, T. L. Volkert, C. J. Wilson, S. P. Bell, and R. A. Young, “Genome-wide location and function of DNA binding proteins,” *Science*, vol. 290, no. 5500, pp. 2306-9, Dec 22, 2000.

[137] A. Barski, S. Cuddapah, K. R. Cui, T. Y. Roh, D. E. Schones, Z. B. Wang, G. Wei, I. Chepelev, and K. J. Zhao, “High-resolution profiling of histone methylations in the human genome,” *Cell*, vol. 129, no. 4, pp. 823-837, May 18, 2007.

[138] L. A. Boyer, T. I. Lee, M. F. Cole, S. E. Johnstone, S. S. Levine, J. R. Zucker, M. G. Guenther, R. M. Kumar, H. L. Murray, R. G. Jenner, D. K. Gifford, D. A. Melton, R. Jaenisch, and R. A. Young, “Core transcriptional regulatory circuitry in human embryonic stem cells,” *Cell*, vol. 122, no. 6, pp. 947-956, Sep 23, 2005.

[139] M. Buschbeck, I. Uribealgo, I. Wibowo, P. Rue, D. Martin, A. Gutierrez, L. Morey, R. Guigo, H. Lopez-Schier, and L. Di Croce, “The histone

---

variant macroH2A is an epigenetic regulator of key developmental genes,” *Nature Structural & Molecular Biology*, vol. 16, no. 10, pp. 1074-U95, Oct, 2009.

[140] M. J. Gamble, K. M. Frizzell, C. Yang, R. Krishnakumar, and W. L. Kraus, “The histone variant macroH2A1 marks repressed autosomal chromatin, but protects a subset of its target genes from silencing,” *Genes & Development*, vol. 24, no. 1, pp. 21-32, Jan 1, 2010.

[141] K. Ratnakumar, L. F. Duarte, G. LeRoy, D. Hasson, D. Smeets, C. Vardabasso, C. Bonisch, T. Y. Zeng, B. Xiang, D. Y. Zhang, H. T. Li, X. W. Wang, S. B. Hake, L. Schermelleh, B. A. Garcia, and E. Bernstein, “ATRX-mediated chromatin association of histone variant macroH2A1 regulates alpha-globin expression,” *Genes & Development*, vol. 26, no. 5, pp. 433-438, Mar 1, 2012.

[142] E. C. Ferber, M. Kajita, A. Wadlow, L. Tobiansky, C. Niessen, H. Ariga, J. Daniel, and Y. Fujita, “A role for the cleaved cytoplasmic domain of E-cadherin in the nucleus,” *Journal of Biological Chemistry*, vol. 283, no. 19, pp. 12691-12700, May 9, 2008.

[143] M. Yamaji, Y. Seki, K. Kurimoto, Y. Yabuta, M. Yuasa, M. Shigeta, K. Yamanaka, Y. Ohinata, and M. Saitou, “Critical function of Prdm14 for the establishment of the germ cell lineage in mice,” *Nature Genetics*, vol. 40, no. 8, pp. 1016-1022, Aug, 2008.

[144] N. Singhal, J. Graumann, G. M. Wu, M. J. Arauzo-Bravo, D. W. Han, B. Greber, L. Gentile, M. Mann, and H. R. Scholer, “Chromatin-Remodeling Components of the BAF Complex Facilitate Reprogramming,” *Cell*, vol. 141, no. 6, pp. 943-955, Jun 11, 2010.

[145] A. Alajem, A. Biran, A. Harikumar, B. S. Sailaja, Y. Aaronson, I. Livyatan, M. Nissim-Rafinia, A. G. Sommer, G. Mostoslavsky, V. R. Gerbasi, D. E. Golden, A. Datta, S. K. Sze, and E. Meshorer, “Differential Association of Chromatin Proteins Identifies BAF60a/SMARCD1 as a Regulator of Embryonic Stem Cell Differentiation,” *Cell Reports*, vol. 10, no. 12, pp. 2019-2031, Mar 31, 2014.

[146] Y. H. Loh, W. W. Zhang, X. Chen, J. George, and H. H. Ng, “Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells,” *Genes & Development*, vol. 21, no. 20, pp. 2545-2557, Oct 15, 2007.

[147] J. K. Chen, H. Liu, J. Liu, J. Qi, B. Wei, J. Q. Yang, H. Q. Liang, Y. Chen, J. Chen, Y. R. Wu, L. Guo, J. Y. Zhu, X. J. Zhao, T. R. Peng,

- 
- Y. X. Zhang, S. Chen, X. J. Li, D. W. Li, T. Wang, and D. Q. Pei, "H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs," *Nature Genetics*, vol. 45, no. 1, pp. 34-U62, Jan, 2013.
- [148] K. Allton, A. K. Jain, H. M. Herz, W. W. Tsai, S. Y. Jung, J. Qin, A. Bergmann, R. L. Johnson, and M. C. Barton, "Trim24 targets endogenous p53 for degradation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 28, pp. 11612-11616, Jul 14, 2009.
- [149] W. A. Whyte, D. A. Orlando, D. Hnisz, B. J. Abraham, C. Y. Lin, M. H. Kagey, P. B. Rahl, T. I. Lee, and R. A. Young, "Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes," *Cell*, vol. 153, no. 2, pp. 307-319, Apr 11, 2013.
- [150] M. M. Li, Y. L. He, W. Dubois, X. L. Wu, J. X. Shi, and J. Huang, "Distinct Regulatory Mechanisms and Functions for p53-Activated and p53-Repressed DNA Damage Response Genes in Embryonic Stem Cells," *Molecular Cell*, vol. 46, no. 1, pp. 30-42, Apr 13, 2012.
- [151] C. Faust, K. A. Lawson, N. J. Schork, B. Thiel, and T. Magnuson, "The Polycomb-group gene *eed* is required for normal morphogenetic movements during gastrulation in the mouse embryo," *Development*, vol. 125, no. 22, pp. 4495-4506, Nov, 1998.
- [152] D. O'Carroll, S. Erhardt, M. Pagani, S. C. Barton, M. A. Surani, and T. Jenuwein, "The Polycomb-group gene *Ezh2* is required for early mouse development," *Molecular and Cellular Biology*, vol. 21, no. 13, pp. 4330-4336, Jul, 2001.
- [153] D. Pasini, A. P. Bracken, M. R. Jensen, E. L. Denchi, and K. Helin, "Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity," *Embo Journal*, vol. 23, no. 20, pp. 4061-4071, Oct 13, 2004.
- [154] X. Z. Li, K. Isono, D. Yamada, T. A. Endo, M. Endoh, J. Shinga, Y. Mizutani-Koseki, A. P. Otte, M. Casanova, H. Kitamura, T. Kamijo, J. Sharif, O. Ohara, T. Toyada, B. E. Bernstein, N. Brockdorff, and H. Koseki, "Mammalian Polycomb-Like Pcl2/Mtf2 Is a Novel Regulatory Component of PRC2 That Can Differentially Modulate Polycomb Activity both at the Hox Gene Cluster and at Cdkn2a Genes (vol 31, pg 351, 2011)," *Molecular and Cellular Biology*, vol. 34, no. 14, pp. 2773-2773, Jul, 2014.
- [155] M. Casanova, T. Preissner, A. Cerase, R. Poot, D. Yamada, X. Z. Li, R. Appanah, K. Bezstarosti, J. Demmers, H. Koseki, and N. Brockdorff,
-

- 
- “Polycomblike 2 facilitates the recruitment of PRC2 Polycomb group complexes to the inactive X chromosome and to target loci in embryonic stem cells,” *Development*, vol. 138, no. 8, pp. 1471-1482, Apr 15, 2011.
- [156] C. Ballare, M. Lange, A. Lapinaite, G. M. Martin, L. Morey, G. Pascual, R. Liefke, B. Simon, Y. Shi, O. Gozani, T. Carlomagno, S. A. Benitah, and L. Di Croce, “Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity,” *Nature Structural & Molecular Biology*, vol. 19, no. 12, pp. 1257-+, Dec, 2012.
- [157] D. Pasini, P. A. C. Cloos, J. Walfridsson, L. Olsson, J. P. Bukowski, J. V. Johansen, M. Bak, N. Tommerup, J. Rappsilber, and K. Helin, “JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells,” *Nature*, vol. 464, no. 7286, pp. 306-U193, Mar 11, 2010.
- [158] C. S. Hughes, S. Foehr, D. A. Garfield, E. E. Furlong, L. M. Steinmetz, and J. Krijgsveld, “Ultrasensitive proteome analysis using paramagnetic bead technology,” *Molecular Systems Biology*, vol. 10, no. 10, Oct, 2014.
- [159] M. Ruggiu, R. Speed, M. Taggart, S. J. McKay, F. Kilanowski, P. Saunders, J. Dorin, and H. J. Cooke, “The mouse Dazla gene encodes a cytoplasmic protein essential for gametogenesis,” *Nature*, vol. 389, no. 6646, pp. 73-77, Sep 4, 1997.
- [160] H. H. Chen, M. Welling, D. B. Bloch, J. Munoz, E. Mientjes, X. Chen, C. Tramp, J. Wu, A. Yabuuchi, Y. F. Chou, C. Buecker, A. Krainer, R. Willemsen, A. J. Heck, and N. Geijsen, “DAZL Limits Pluripotency, Differentiation, and Apoptosis in Developing Primordial Germ Cells,” *Stem Cell Reports*, vol. 3, no. 5, pp. 892-904, Nov 11, 2014.
- [161] M. Welling, H. H. Chen, J. Munoz, M. U. Musheev, L. Kester, J. P. Junker, N. Mischerikow, M. Arbab, E. Kuijk, L. Silberstein, P. V. Kharchenko, M. Geens, C. Niehrs, H. van de Velde, A. van Oudenaarden, A. J. R. Heck, and N. Geijsen, “DAZL regulates Tet1 translation in murine embryonic stem cells,” *Embo Reports*, vol. 16, no. 7, pp. 791-802, Jul, 2015.
- [162] T. Burckstummer, C. Baumann, S. Bluml, E. Dixit, G. Durnberger, H. Jahn, M. Planyavsky, M. Bilban, J. Colinge, K. L. Bennett, and G. Superti-Furga, “An orthogonal proteomic-genomic screen identifies AIM2 as a cytoplasmic DNA sensor for the inflammasome,” *Nature Immunology*, vol. 10, no. 3, pp. 266-272, Mar, 2009.
- [163] G. Mittler, F. Butter, and M. Mann, “A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional



- 
- DNA elements,” *Genome Research*, vol. 19, no. 2, pp. 284-293, Feb, 2009.
- [164] E. D. Rubio, D. J. Reiss, P. L. Weicsh, C. M. Disteche, G. N. Filippova, N. S. Baliga, R. Aebersold, J. A. Ranish, and A. Krumm, “CTCF physically links cohesin to chromatin,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 24, pp. 8309-8314, Jun 17, 2008.
- [165] J. P. Lambert, T. Pawson, and A. C. Gingras, “Mapping physical interactions within chromatin by proteomic approaches,” *Proteomics*, vol. 12, no. 10, pp. 1609-1622, May, 2012.
- [166] J. Dejardin, and R. E. Kingston, “Purification of Proteins Associated with Specific Genomic Loci,” *Cell*, vol. 136, no. 1, pp. 175-186, Jan 9, 2009.
- [167] J. M. Antao, J. M. Mason, J. Dejardin, and R. E. Kingston, “Protein Landscape at *Drosophila melanogaster* Telomere-Associated Sequence Repeats,” *Molecular and Cellular Biology*, vol. 32, no. 12, pp. 2170-2182, Jun, 2012.
- [168] N. Saksouk, T. K. Barth, C. Ziegler-Birling, N. Olova, A. Nowak, E. Rey, E. Simboeck, J. Mateos-Langerak, S. Urbach, W. Reik, M. E. Torres-Padilla, A. Imhof, and J. Dejardin, “Redundant Mechanisms to Form Silent Chromatin at Pericentromeric Regions Rely on BEND3 and DNA Methylation (vol 56, pg 580, 2014),” *Molecular Cell*, vol. 57, no. 1, pp. 202-202, Jan 8, 2015.
- [169] M. Capelson, Y. Liang, R. Schulte, W. Mair, U. Wagner, and M. W. Hetzer, “Chromatin-Bound Nuclear Pore Components Regulate Gene Expression in Higher Eukaryotes,” *Cell*, vol. 140, no. 3, pp. 372-U100, Feb 5, 2010.
- [170] B. Kalverda, H. Pickersgill, V. V. Shloma, and M. Fornerod, “Nucleoporins Directly Stimulate Expression of Developmental and Cell-Cycle Genes Inside the Nucleoplasm,” *Cell*, vol. 140, no. 3, pp. 360-U81, Feb 5, 2010.
- [171] L. H. Kasper, P. K. Brindle, C. A. Schnabel, C. E. J. Pritchard, M. L. Cleary, and J. M. A. van Deursen, “CREB binding protein interacts with nucleoporin-specific FG repeats that activate transcription and mediate NUP98-HOXA9 oncogenicity,” *Molecular and Cellular Biology*, vol. 19, no. 1, pp. 764-776, Jan, 1999.
- [172] X. T. Bai, B. W. Gu, T. Yin, C. Niu, X. D. Xi, J. Zhang, Z. Chen, and S. J. Chen, “Trans-repressive effect of NUP98-PMX1 on PMX1-Regulated c-FOS gene through recruitment of histone deacetylase 1 by FG repeats,”
-

- 
- Cancer Research, vol. 66, no. 9, pp. 4584-4590, May 1, 2006.
- [173] E. R. Griffis, S. L. Xu, and M. A. Powers, "Nup98 localizes to both nuclear and cytoplasmic sides of the nuclear pore and binds to two distinct nucleoporin subcomplexes," *Molecular Biology of the Cell*, vol. 14, no. 2, pp. 600-610, Feb, 2003.
- [174] E. R. Griffis, N. Altan, J. Lippincott-Schwartz, and M. A. Powers, "Nup98 is a mobile nucleoporin with transcription-dependent dynamics," *Molecular Biology of the Cell*, vol. 13, no. 4, pp. 1282-1297, Apr, 2002.
- [175] T. M. Franks, and M. W. Hetzer, "The role of Nup98 in transcription regulation in healthy and diseased cells," *Trends in Cell Biology*, vol. 23, no. 3, pp. 112-117, Mar, 2013. [176] M. Stadtfeld, N. Maherali, D. T. Breault, and K. Hochedlinger, "Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse," *Cell Stem Cell*, vol. 2, no. 3, pp. 230-240, Mar, 2008.
- [177] T. Brambrink, R. Foreman, G. G. Welstead, C. J. Lengner, M. Wernig, H. Suh, and R. Jaenisch, "Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells," *Cell Stem Cell*, vol. 2, no. 2, pp. 151-159, Feb, 2008.
- [178] J. Hansson, M. R. Rafiee, S. Reiland, J. M. Polo, J. Gehring, S. Okawa, W. Huber, K. Hochedlinger, and J. Krijgsveld, "Highly Coordinated Proteome Dynamics during Reprogramming of Somatic Cells to Pluripotency," *Cell Reports*, vol. 2, no. 6, pp. 1579-1592, Dec, 2012.
- [179] Y. Liang, and M. W. Hetzer, "Functional interactions between nucleoporins and chromatin," *Curr Opin Cell Biol*, vol. 23, no. 1, pp. 65-70, Feb, 2011.
- [180] A. Ori, N. Banterle, M. Iskar, A. Andres-Pons, C. Escher, H. Khanh Bui, L. Sparks, V. Solis-Mezarino, O. Rinner, P. Bork, E. A. Lemke, and M. Beck, "Cell type-specific nuclear pores: a case in point for context-dependent stoichiometry of molecular machines," *Mol Syst Biol*, vol. 9, pp. 648, 2013.
- [181] M. Raices, and M. A. D'Angelo, "Nuclear pore complex composition: a new regulator of tissue-specific and developmental functions," *Nat Rev Mol Cell Biol*, vol. 13, no. 11, pp. 687-99, Nov, 2012.
- [182] M. Olsson, M. Ekblom, L. Fecker, M. Kurkinen, and P. Ekblom, "cDNA cloning and embryonic expression of mouse nuclear pore membrane glycoprotein 210 mRNA," *Kidney Int*, vol. 56, no. 3, pp. 827-38, Sep, 1999.
- [183] M. A. D'Angelo, J. S. Gomez-Cavazos, A. Mei, D. H. Lackner, and M. W. Hetzer, "A change in nuclear pore complex composition regulates cell

- 
- differentiation,” *Dev Cell*, vol. 22, no. 2, pp. 446-58, Feb 14, 2012.
- [184] J. S. Gomez-Cavazos, and M. W. Hetzer, “The nucleoporin gp210/Nup210 controls muscle differentiation by regulating nuclear envelope/ER homeostasis,” *J Cell Biol*, vol. 208, no. 6, pp. 671-81, Mar 16, 2015.
- [185] M. Stadtfeld, N. Maherali, M. Borkent, and K. Hochedlinger, “A reprogrammable mouse strain from gene-targeted embryonic stem cells,” *Nature Methods*, vol. 7, no. 1, pp. 53-U10, Jan, 2010.
- [186] P. Samavarchi-Tehrani, A. Golipour, L. David, H. K. Sung, T. A. Beyer, A. Datti, K. Woltjen, A. Nagy, and J. L. Wrana, “Functional Genomics Reveals a BMP-Driven Mesenchymal-to-Epithelial Transition in the Initiation of Somatic Cell Reprogramming,” *Cell Stem Cell*, vol. 7, no. 1, pp. 64-77, Jul 2, 2010.
- [187] S. Ruiz, A. D. Panopoulos, A. Herrerias, K. D. Bissig, M. Lutz, W. T. Berggren, I. M. Verma, and J. C. I. Belmonte, “A High Proliferation Rate Is Required for Cell Reprogramming and Maintenance of Human Embryonic Stem Cell Identity,” *Current Biology*, vol. 21, no. 1, pp. 45-52, Jan 11, 2011.
- [188] S. Nagai, K. Dubrana, M. Tsai-Pflugfelder, M. B. Davidson, T. M. Roberts, G. W. Brown, E. Varela, F. Hediger, S. M. Gasser, and N. J. Krogan, “Functional targeting of DNA damage to a nuclear pore-associated SUMO-dependent ubiquitin ligase,” *Science*, vol. 322, no. 5901, pp. 597-602, Oct 24, 2008.
- [189] V. Galy, J. C. Olivo-Marin, H. Scherthan, V. Doye, N. Rascalou, and U. Nehrbass, “Nuclear pore complexes in the organization of silent telomeric chromatin,” *Nature*, vol. 403, no. 6765, pp. 108-112, Jan 6, 2000.
- [190] F. Feuerbach, V. Galy, E. Trelles-Sticken, M. Fromont-Racine, A. Jacquier, E. Gilson, J. C. Olivo-Marin, H. Scherthan, and U. Nehrbass, “Nuclear architecture and spatial positioning help establish transcriptional states of telomeres in yeast,” *Nature Cell Biology*, vol. 4, no. 3, pp. 214-221, Mar, 2002.
- [191] Y. B. Deng, S. S. Chan, and S. Chang, “Telomere dysfunction and tumour suppression: the senescence connection,” *Nature Reviews Cancer*, vol. 8, no. 6, pp. 450-458, Jun, 2008.
- [192] B. Khadaroo, M. T. Teixeira, P. Luciano, N. Eckert-Boulet, S. M. Germann, M. N. Simon, I. Gallina, P. Abdallah, E. Gilson, V. Geli, and M. Lisby, “The DNA damage response at eroded telomeres and tethering to the nuclear pore complex,” *Nature Cell Biology*, vol. 11, no. 8, pp. 980-U161,
-

---

Aug, 2009.

[193] K. Mekhail, J. Seebacher, S. P. Gygi, and D. Moazed, "Role for perinuclear chromosome tethering in maintenance of genome stability," *Nature*, vol. 456, no. 7222, pp. 667-70, Dec 4, 2008.

[194] R. Bermejo, A. Kumar, and M. Foiani, "Preserving the genome by regulating chromatin association with the nuclear envelope," *Trends Cell Biol*, vol. 22, no. 9, pp. 465-73, Sep, 2012.

[195] J. Toth, and M. D. Biggin, "The specificity of protein-DNA crosslinking by formaldehyde: in vitro and in drosophila embryos," *Nucleic Acids Res*, vol. 28, no. 2, pp. e4, Jan 15, 2000.

[196] H. Mohammed, C. D'Santos, A. A. Serandour, H. R. Ali, G. D. Brown, A. Atkins, O. M. Rueda, K. A. Holmes, V. Theodorou, J. L. Robinson, W. Zwart, A. Saadi, C. S. Ross-Innes, S. F. Chin, S. Menon, J. Stingl, C. Palmieri, C. Caldas, and J. S. Carroll, "Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor," *Cell Rep*, vol. 3, no. 2, pp. 342-9, Feb 21, 2013.

[197] Y. Zhen, V. Sorensen, C. S. Skjerpen, E. M. Haugsten, Y. Jin, S. Walchli, S. Olsnes, and A. Wiedlocha, "Nuclear import of exogenous FGF1 requires the ER-protein LRRC59 and the importins Kpnalpha1 and Kpn-beta1," *Traffic*, vol. 13, no. 5, pp. 650-64, May, 2012.

[198] S. Meunier, M. G. Navarro, C. Bossard, H. Laurell, C. Touriol, E. Lacazette, and H. Prats, "Pivotal role of translokin/CEP57 in the unconventional secretion versus nuclear translocation of FGF2," *Traffic*, vol. 10, no. 12, pp. 1765-72, Dec, 2009.

[199] E. Marcon, H. Jain, A. Bhattacharya, H. Guo, S. Phanse, S. Pu, G. Byram, B. C. Collins, E. Dowdell, M. Fenner, X. Guo, A. Hutchinson, J. J. Kennedy, B. Krastins, B. Larsen, Z. Y. Lin, M. F. Lopez, P. Loppnau, S. Miersch, T. Nguyen, J. B. Olsen, M. Paduch, M. Ravichandran, A. Seitova, G. Vadali, M. S. Vogelsang, J. R. Whiteaker, G. Zhong, N. Zhong, L. Zhao, R. Aebersold, C. H. Arrowsmith, A. Emili, L. Frappier, A. C. Gingras, M. Gstaiger, A. G. Paulovich, S. Koide, A. A. Kossiakoff, S. S. Sidhu, S. J. Wodak, S. Graslund, J. F. Greenblatt, and A. M. Edwards, "Assessment of a method to characterize antibody selectivity and specificity for use in immunoprecipitation," *Nat Methods*, vol. 12, no. 8, pp. 725-31, Aug, 2015.

[200] K. Allton, A. K. Jain, H. M. Herz, W. W. Tsai, S. Y. Jung, J. Qin, A. Bergmann, R. L. Johnson, and M. C. Barton, "Trim24 targets endogenous p53 for degradation," *Proc Natl Acad Sci U S A*, vol. 106, no. 28, pp.

---

11612-6, Jul 14, 2009.

[201] M. E. Torres-Padilla, and M. Zernicka-Goetz, "Role of TIF1alpha as a modulator of embryonic transcription in the mouse zygote," *J Cell Biol*, vol. 174, no. 3, pp. 329-38, Jul 31, 2006.

[202] W. W. Tsai, Z. Wang, T. T. Yiu, K. C. Akdemir, W. Xia, S. Winter, C. Y. Tsai, X. Shi, D. Schwarzer, W. Plunkett, B. Aronow, O. Gozani, W. Fischle, M. C. Hung, D. J. Patel, and M. C. Barton, "TRIM24 links a non-canonical histone signature to breast cancer," *Nature*, vol. 468, no. 7326, pp. 927-32, Dec 16, 2010.

[203] S. Hatakeyama, "TRIM proteins and cancer," *Nat Rev Cancer*, vol. 11, no. 11, pp. 792-804, Nov, 2011.

[204] J. M. Antao, J. M. Mason, J. Dejardin, and R. E. Kingston, "Protein landscape at *Drosophila melanogaster* telomere-associated sequence repeats," *Mol Cell Biol*, vol. 32, no. 12, pp. 2170-82, Jun, 2012.

[205] J. Kennedy-Darling, H. Guillen-Ahlers, M. R. Shortreed, M. Scalf, B. L. Frey, C. Kendzierski, M. Olivier, A. P. Gasch, and L. M. Smith, "Discovery of Chromatin-Associated Proteins via Sequence-Specific Capture and Mass Spectrometric Protein Identification in *Saccharomyces cerevisiae*," *J Proteome Res*, vol. 13, no. 8, pp. 3810-25, Aug 1, 2014.

[206] J. Rappsilber, M. Mann, and Y. Ishihama, "Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips," *Nat Protoc*, vol. 2, no. 8, pp. 1896-906, 2007.

[207] L. Kall, J. D. Canterbury, J. Weston, W. S. Noble, and M. J. MacCoss, "Semi-supervised learning for peptide identification from shotgun proteomics datasets," *Nat Methods*, vol. 4, no. 11, pp. 923-5, Nov, 2007.

[208] D. Blankenberg, G. Von Kuster, N. Coraor, G. Ananda, R. Lazarus, M. Mangan, A. Nekrutenko, and J. Taylor, "Galaxy: a web-based genome analysis tool for experimentalists," *Curr Protoc Mol Biol*, vol. Chapter 19, pp. Unit 19 10 1-21, Jan, 2010.

[209] B. Giardine, C. Riemer, R. C. Hardison, R. Burhans, L. Elnitski, P. Shah, Y. Zhang, D. Blankenberg, I. Albert, J. Taylor, W. Miller, W. J. Kent, and A. Nekrutenko, "Galaxy: a platform for interactive large-scale genome analysis," *Genome Res*, vol. 15, no. 10, pp. 1451-5, Oct, 2005.

[210] J. Goecks, A. Nekrutenko, J. Taylor, and T. Galaxy, "Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences," *Genome Biol*, vol. 11, no. 8, pp. R86, 2010.

- 
- [211] B. Langmead, C. Trapnell, M. Pop, and S. L. Salzberg, "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome," *Genome Biol*, vol. 10, no. 3, pp. R25, 2009.
- [212] F. Ramirez, F. Dundar, S. Diehl, B. A. Gruning, and T. Manke, "deepTools: a flexible platform for exploring deep-sequencing data," *Nucleic Acids Res*, vol. 42, no. Web Server issue, pp. W187-91, Jul, 2014.
- [213] S. G. Landt, G. K. Marinov, A. Kundaje, P. Kheradpour, F. Pauli, S. Batzoglou, B. E. Bernstein, P. Bickel, J. B. Brown, P. Cayting, Y. Chen, G. DeSalvo, C. Epstein, K. I. Fisher-Aylor, G. Euskirchen, M. Gerstein, J. Gertz, A. J. Hartemink, M. M. Hoffman, V. R. Iyer, Y. L. Jung, S. Karmakar, M. Kellis, P. V. Kharchenko, Q. Li, T. Liu, X. S. Liu, L. Ma, A. Milosavljevic, R. M. Myers, P. J. Park, M. J. Pazin, M. D. Perry, D. Raha, T. E. Reddy, J. Rozowsky, N. Shores, A. Sidow, M. Slattery, J. A. Stamatoyannopoulos, M. Y. Tolstorukov, K. P. White, S. Xi, P. J. Farnham, J. D. Lieb, B. J. Wold, and M. Snyder, "ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia," *Genome Res*, vol. 22, no. 9, pp. 1813-31, Sep, 2012.
- [214] Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, and X. S. Liu, "Model-based analysis of ChIP-Seq (MACS)," *Genome Biol*, vol. 9, no. 9, pp. R137, 2008.
- [215] G. Yu, L. G. Wang, and Q. Y. He, "ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization," *Bioinformatics*, vol. 31, no. 14, pp. 2382-3, Jul 15, 2015.
- [216] J. Cox, and M. Mann, "MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification," *Nat Biotechnol*, vol. 26, no. 12, pp. 1367-72, Dec, 2008.
- [217] W. Huang da, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources," *Nat Protoc*, vol. 4, no. 1, pp. 44-57, 2009.
- [218] C. Stadler, M. Skogs, H. Brismar, M. Uhlen, and E. Lundberg, "A single fixation protocol for proteome-wide immunofluorescence localization studies," *J Proteomics*, vol. 73, no. 6, pp. 1067-78, Apr 18, 2010.

# Appendices

## A.1 Abbreviations

Abbreviation	Description
2iL	2i medium plus LIF
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
BP	Biological Process
BAF	Brg-Associated Factors
CC	Cellular Compartment
ChIP-MS	Chromatin Immuno-Precipitation with Mass Spectrometry
ChAP-MS	Chromatin Affinity purification with Mass Spectrometry
ChEP	Chromatin Enrichment for Proteomics
ChIP	Chromatin Immuno-Precipitation
ChroP	Chromatin Proteomics
EGC	Embryonic Germ Cells
ESC	Eembryonic Stem Cells
ER	Endoplasmic Reticulum
GO	Gene Ontology
GLFG	glycine-leucine-phenylalanine-glycine domain
HDAC	Histone Deacetylases
HPA	Human Proteim Atlas
iPSC	Induced Pluripotent Stem Cells
ICM	Inner-Cell Mass
ICP	Interphase Chromatin Probability
MET	Mesenchymal to Epithelial Transition
MNase	Micrococcal Nuclease
mChIP	modified Chromatin Immno-Precipitation
MF	Molecular Function
MAB	Monoclonal Antibody
MCCP	MultiClassifier Combinatorial Proteomics

*Continued on next page*

*Continued from previous page*

<b>Abbreviation</b>	<b>Description</b>
NSE	Non-Super-Enhancer
OSN	Oct4, Sox2 and Nanog
OSKM	Oct4, Sox2, Klf4, Myc
PRC1	Polycomb Repressive Complex 1
PRC2	Polycomb Repressive Complex 2
PGC	Primordial Germ Cells
PICh	Proteomics of Isolated Chromatin fragments
Pol II	RNA polymerase 2
SICAP	Selective Isolation of Chromatin-Associated Proteins
SE	Super-Enhancer
TAP	Tandem Affinity Purification
TIGR	Targeted Isolation of Genomic Regions
TdT	Terminal deoxynucleotidyl transferase
TAL	Transcription Activator-Like
TF	Transcription Factors
TE	Trophectoderm



## A.2 Supplementary Tables

Please find the enclosed CD for the following supplementary tables:

- **Supplementary Table 3.1:**  
Chromatin composition of the mouse ESCs
- **Supplementary Table 3.2:**  
Immuno-staining of some proteins enriched by SICAP
- **Supplementary Table 3.3:**  
Comparing chromatin protein composition, full proteome and transcriptome between 2i and serum conditions
- **Supplementary Table 4.1:**  
Nanog ChIP-SICAP in comparison with no-antibody ChIP-SICAP
- **Supplementary Table 4.2:**  
Comparing Oct4, Sox2, Nanog and E-cadherin ChIP-SICAP between 2i and serum conditions
- **Supplementary Table 4.3:**  
The overlap of Trim24 binding sites with OSN on non-super-enhancers, and super-enhancers
- **Supplementary Table 4.4:**  
Suz12 ChIP-SICAP

- **Supplementary Table 4.5:**  
**Dazl ChIP-SICAP**
  
- **Supplementary Table 5.1:**  
**TIGR using Nanog-519, Nanog-453 and the scramble probe**
  
- **Supplementary Table 5.2 :**  
**TIGR using Nanog-519 in 2i and serum conditions**

### A.3 List of publications

1. **Rafiee MR**, Girardot C, Thul P, Lundberg E, Krijgsveld J. Expanding the circuitry of pluripotency by selective isolation of chromatin-associated proteins. Manuscript in preparation.
2. Shahryari A, **Rafiee MR**, Fouani Y, Olliae NA, Samaei NM, Shafiee M, Semnani S, Vasei M, Mowla SJ. Two Novel Splice Variants of SOX2OT, SOX2OT-S1, and SOX2OT-S2 are co-upregulated with SOX2 and OCT4 in Esophageal Squamous Cell Carcinoma. *Stem Cells*. 32:126-34, 2013.
3. Hansson J, **Rafiee MR**, Reiland S, Polo JM, Gehring J, Okawa S, Huber W, Hochedlinger K, Krijgsveld J. Highly coordinated proteome dynamics during reprogramming of somatic cells to pluripotency. *Cell Rep*. 2: 1579-92, 2012.
4. Asadzadeh J, Asadi MH, Shakhssalim N, **Rafiee MR**, Kalhor HR, Tavallaei M, Mowla SJ. A Plausible Anti-Apoptotic Role of Up-Regulated OCT4B1 in Bladder Tumors. *Urol J*. 9:574-80, 2012.
5. Farashahi Yazd E, **Rafiee MR**, Soleimani M, Tavallaei M, Salmani MK, Mowla SJ. OCT4B1, a novel spliced variant of OCT4, generates a stable truncated protein with a potential role in stress response. *Cancer Lett*. 309:170-5, 2011.
6. **Rafiee MR**, Sokhansanj A, Naghizadeh MA, Farazmand A. Analysis of Y-chromosomal short tandem repeat (STR) polymorphism in an Iranian Sadat population. *Genetika*. 8:1105-9, 2009.
7. **Rafiee MR**, Sokhansanj A, Yoosefi M and Naghizadeh MA. Identification of Salt-Inducible Peptide with Putative Kinase Activity in Halophilic Bacterium *Virgibacillus halodenitrificans*. *Journal of Bioscience and Bioengineering*, 104: 178-181, 2007.

## A.4 Acknowledgement

First of all, I would like to greatly appreciate my supervisor Dr. Jeroen Krijgsveld, for trusting me and giving me the opportunity to do my PhD at EMBL. Without he, I do not know where I would be. In addition, I would like to acknowledge the Darwin Trust of Edinburgh for providing me with their generous fellowship.

Next, I would like to appreciate my TAC committee members: Dr. Christian Haering for many useful discussions and his supportive manner, Dr. Lars Steinmetz, Dr. Peter Neveu and Prof. Michael Knop for their supports and helpful comments in the meetings.

Furthermore, I would like to appreciate Genome Biology Computational Support; specially Dr. Charles Girardo for instructing me ChIP-Seq data analysis, and contributing in these studies. Additionally, I would like to thank Dr. Emma Lundberg's contribution from Human Protein Atlas for providing us with some antibodies for this research. In addition, I would like to express my thanks to Prof. Konrad Hochedlinger from Harvard medical school for providing us with the reprogrammable MEF cells.

These methods would have not been developed without the great core facilities of EMBL, specially: Proteomics Core facility, GeneCore, Advanced Light Microscopy, Protein Expression and Purification Core facility and Flow Cytometry Core facility. I would like to thank all staffs of these facilities.

I would to sincerely appreciate all present and past members of the lab, specially Dr. Jenny Hansson, Dr. Daniel Klimmeck and Dr. Gertjan Kramer for helpful discussions. From the other labs at EMBL, I am really thankful to Nirupama Ramanathan, Massimo Petretich, Christian Hoerner and Volker Lauschke for providing me with the cell lines and the plasmids I needed.

In addition, I would like to acknowledge the fellows who have been working on my methods to apply them to their studies: Dr. Annalisa Nicastrì, Olga Mikhaylichenko, Jakob Trendel and Dr. Charlotte Stadler.

I am also really thankful to my former teachers Dr. Hamid Kalhor and Dr. Javad Mowla, former colleagues and everlasting friends Hamid Khayatzadeh and Julian Gill.

The last but not the least, I would like to show my special thanks of gratitude to my parents and my dear wife who have been great treasures of my life.

THANKS AGAIN TO ALL WHO HELPED ME.