Genetic and epigenetic mechanisms of the regulation of mouse embryonic stem cells self-renewal by the pluripotency transcription factor Nanog

Victor Heurtier

Genetic and epigenetic mechanisms of the regulation of mouse embryonic stem cells self-renewal by the pluripotency transcription factor Nanog

Victor HEURTIER

Thèse de doctorat de Biologie

Dirigée par Pablo NAVARRO GIL

Laboratoire Epigénétique des cellules souches – Institut Pasteur

Sorbonne Université

Ecole doctorale Complexité du Vivant (ED 515)

Présentée et soutenue publiquement le 18/09/2018

Devant un jury composé de :

Président : Pr MOUCHEL-VIELH Emmanuèle, Professeur

Rapporteur : Dr SAVATIER Pierre, Directeur de Recherche

Rapporteur : Dr RADA-IGLESIAS Alvaro, Chef d'équipe

Examinateur : Dr AZUARA Veronique, Maître de conférences

Examinateur : Dr PINSKAYA Marina, Maître de conférences

Directeur de thèse : Dr NAVARRO GIL Pablo, Chef d'équipe

Abstract

Mouse embryonic stem (ES) cells are derived from the pre-implantation blastocyst and are able to maintain their pluripotent state through virtually limitless cell divisions *in vitro*. The study of the mechanisms regulating the specific features of ES cells led to the discovery of the transcription factors (TFs) governing their unique transcriptome. Among those TFs, Nanog plays a central role in the gene regulatory network that supports ES cells self-renewal. However, the molecular mechanisms by which Nanog exerts its functions are not fully elucidated.

We adapted an inducible CRISPR activation (CRISPRa) system in mouse ES cells and demonstrated its functionality to stimulate transcription at endogenous loci. We then accurately determined the list of Nanog responsive genes using gain (CRISPRa cell line) and loss of function experiments and total RNA sequencing. Moreover, the DNA binding profiles of distinct pluripotency TFs were assessed genome-wide by chromatin immunoprecipitation (ChIP) and sequencing upon Nanog depletion and revealed the importance of the latter in the regulation of the pluripotency network activity at thousands of loci. We additionally showed that Nanog employs distinct mechanisms to activate and repress its targets through binding rather distal or proximal DNA regulatory elements respectively. We further investigated the action of Nanog overexpression in the absence of the LIF cytokine (otherwise mandatory to support ES cells self-renewal) and showed that Nanog both activates pluripotency sustaining factors and, to a greater extent, represses differentiation determinants. Furthermore, we found that Nanog-dependent gene repression is strongly correlated with the enrichment of the repressive histone mark H3K27me3 thus proposing a new mechanism through which Nanog could indirectly repress its targets such as Otx2, a major early differentiation factor.

We further explored the regulation of long non-coding RNAs (lncRNAs) expression by Nanog and selected non-coding candidates that could potentially be involved in the maintenance of mouse ES cell self-renewal. So far, only one of them seems to give promising results and will be further investigated.

Finally, a fortuitously observed off-target effect of a specific gRNA, together with CRISPRa, led to the induction of the 2-cell-like state transcriptomic profile spontaneously found in a small subset of ES cells. CRISPRa off-target binding sites were determined genome-wide by ChIP sequencing and candidate genes potentially responsible for the induction of the observed effect are now under investigation.

À mon père,

Qui m’a toujours encouragé et soutenu,

Sans qui cette thèse ne serait sans doute pas.

Acknowledgements

I would like to express my deep gratitude to my supervisor Pablo Navarro Gil for giving me the opportunity to work first as a Master and then a PhD student in his laboratory. Thank you for the great liberty you gave me in my work and the faith you had in me which allowed me to feed my curiosity and develop my scientific autonomy. Thank you for useful critiques and accurate analyses which pushed me to do always better and helped me building my scientific critical spirit.

I want to thank all the past and present members of the laboratory with whom I had a real pleasure to work, discuss, exchange and laugh every day during these five years and made my time in the lab enjoyable. Thank you Agnès for your kindness, advice and support as well as all you do for us every day and that we often don't even realize. Thank you Philippe for all our discussions, from science to politics, from music to fishing, without forgetting our silly jokes. I wish to thank Nicola for his excellent advice that I didn't follow enough. It was also a pleasure to work with Inma, always enthusiastic and encouraging, thank you for your help and kindness. I am particularly grateful for the help of Nick in analysing the massive amount of data we generate, for great discussions and with whom it was a pleasure to work with. Thank you Thaleia for your help, humour and rational scientific point of view. Thank you Laure for your patience and for the personal engagement you put in the project. Finally, thank you Laurence for your help with all our administrative work. I wish you all best of luck and success in your future professional and personal lives.

I wish to acknowledge the help provided by Didier Montarras by introducing me to the world of research and of Institut Pasteur from the beginning. I am very grateful for your tutoring and help. Un grand merci Didier.

I would like to thank Damien for his help on the gRNA design and our long discussions.

I want to thank my PhD committee (Alice Jouneau, Marina Pinskaya, Cosmin Saveanu) for following my work all along my thesis and for their useful critiques and suggestions.

I would like to thank the whole Department of Stem Cells and Developmental Biology, and more specifically Aurélien, Brendan, Sylvia, Sabela, Mathieu, Coralie, Aurélie, Clémire, Catherine, Edith, Marc, Sylvain, Glenda.

I want to thank Sara for welcoming me in the Department during my Master and for her support.

I would like to thank the organizers of the MD-PhD programs from the Sorbonne Université and the Ecole de l'INSERM for allowing me to do this PhD.

I want to thank Alexandra for her everyday life support, kindness and care as well as her constant and precious help in the lab.

Je voudrais finalement remercier ma famille pour leur soutien tout au long de mes études, leur affection et leur intérêt dans mon travail. Merci à mes deux frères Lucien et Sylvain, à ma grand-mère Mimine et à mes parents, Stéphanie et Jean-François.

Table of content

Introduction

I. Early development & embryonic stem cells

A. Early mouse embryo development

The fertilized oocyte as well as all the blastomeres resulting from its first cleavages are able to produce all embryonic as well as extraembryonic tissues originating from the embryo, they are totipotent (Boroviak and Nichols 2014). After fertilization, while the spermatozoid has been largely emptied for the benefit of high mobility and contains nearly only the paternal genome, the stock of maternal transcripts and proteins allows for the first steps of development to occur including the transcriptional activation of the zygotic genome around the 2-cell stage. The first five cleavages of the embryo occur at constant volume, without any cell growth, and consist of subsequent divisions of the initial egg cytosol **(Fig. 1.1)**. Apart from the initial zygote, only the two blastomeres resulting from its first cleavage are able to autonomously produce a full embryo, while blastomeres produced at later cleavages can only participate to all embryonic and extraembryonic tissues when aggregated with host blastomeres (Chazaud and Yamanaka 2016). Although 8-cell stage blastomeres were considered as identical and equipotent, accumulating evidences show that they actually differ on several aspects including DNA methylation status, histone modifications enrichment and transcription factors expression (Tabansky et al. 2013). Yet, the functional significance of those early differences is not totally deciphered.

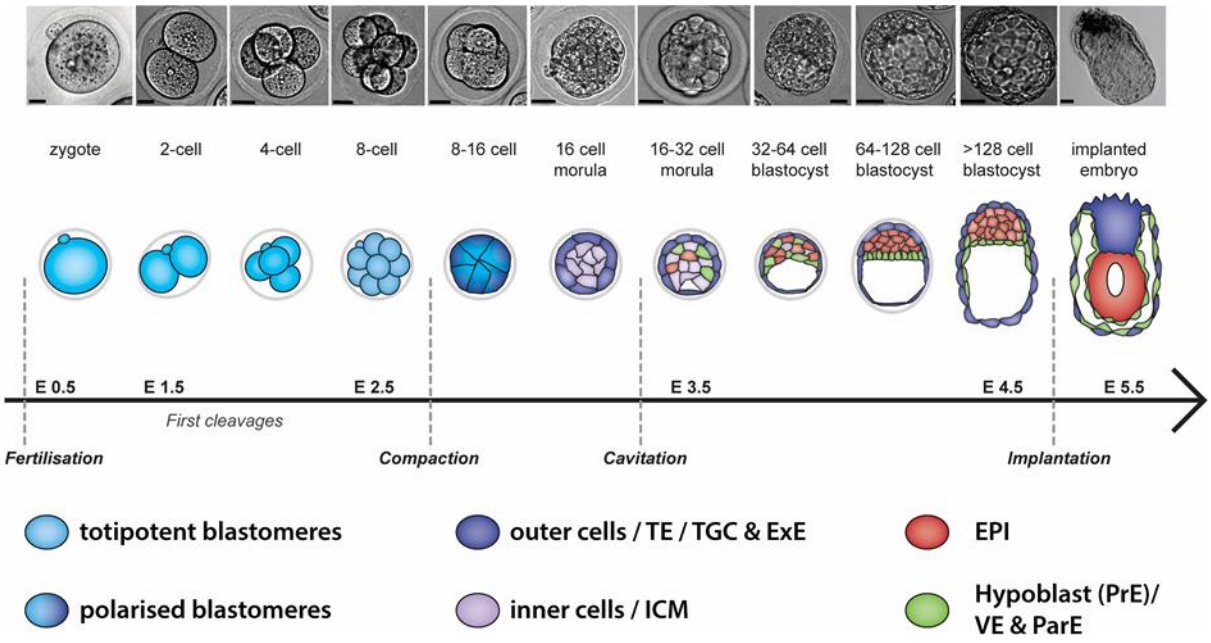


Fig. 1.1. Pictures and schematic representation of the early development of the mouse embryo form the fertilisation until its implantation (Piliszek et al. 2016).

Until the eight-cell stage, the dividing blastomeres, surrounded by the *zona pellucida* inherited form the egg, look like a loose aggregate of spherical cells. Starting from the eight-cell stage the embryo undergoes a morphologically noticeable event called compaction **(Fig. 1.1)**. This corresponds to the increase of cell-cell interactions and generates a cohesive sphere therefore establishing a layer of polarized cells outside the embryo and a small apolar population inside (White et al. 2016). It has been shown that blastomeres facing the exterior of the embryo at this stage are strongly biased to differentiate towards the extraembryonic trophectoderm (TE) lineage which will be necessary for implantation, placenta development and embryonic health (Rossant and Tam 2009). On the contrary, blastomeres confined at the centre of the compacted embryo tend to be part of the inner cell mass (ICM) of the blastocyst which will generate the whole foetus as well as additional extra-embryonic tissues. The Hippo pathway, known for its cellular role in integrating physical cues from the surrounding environment (cell-cell contact, adhesion strength, physical stress), through the inhibition of the transcription factor Yap1 is largely involved in this process (Nishioka et al. 2009). The next cleavages will therefore give rise to the first lineage choice in the embryo: TE outside and ICM inside **(Fig. 1.1)**. Master transcriptional regulators have been characterized in the control of this choice as Cdx2 and Gata3 for TE commitment and Oct4, Sox2 and Nanog for the ICM specification (Chazaud and Yamanaka 2016). A second cell fate choice will happen within the ICM population after it includes about 15 cells, separating the primitive endoderm (PrE) from the epiblast (Epi). The PrE will generate extra-embryonic tissues as the yolk sac whereas the Epi will generate the whole foetus. While Gata6 and Nanog, markers of the PrE and the Epi respectively are first co-expressed in the early ICM, they progressively become mutually exclusive in the late blastocyst stage showing a salt-and-pepper distribution. Importantly, the FGF pathway through its ligand/receptor dyad FGF4/FGFR2 plays a key role in this segregation as inhibition of the pathway through chemical agents completely abolishes PrE development (Chazaud et al. 2006; Kang et al. 2013). Practically, a cross-talk is organized in such a way that cells from the PrE express the receptor and Epi cells synthetize the ligand but are not responsive to it thus reinforcing the nascent differential identities (Ohnishi et al. 2014). It has not been clearly established whether initial positioning of the cells within the ICM determines or not their PrE/Epi choice. It was recently shown that apoptosis is clearly involved in the proper segregation of Epi/PrE cells, acting as a selective mechanism. As the Epi progenitors start to emerge from the ICM population they quickly proliferate and are

definitively engaged in their lineage while PrE cells are more plastic and have been shown to be able to revert to an Epi *state in vivo* (Xenopoulos et al. 2015).

After blastocyst implantation, the founder cells of the gametes, called Primordial Germ Cells (PGCs), will emerge from the ICM cells. In contrast to the rest of the embryo, they will undergo a complete epigenetic "resetting". Indeed a major regulator of gene expression, i.e. DNA methylation, will be completely erased from the PGCs before being slowly re-established in a sex-dependent manner. While the genetic information brought by the parental genomes is obviously different, a more drastic dissimilarity between the two gametes is found in the way those two genomes are pre-modified. This high asymmetry, arising from the divergent differentiation between male and female germ cells, will be partially maintained during the life of the new embryo and will widely contribute to its proper development (Weaver, Susiarjo, and Bartolomei 2009; Smallwood and Kelsey 2012; Z. D. Smith and Meissner 2013).

B. The establishment of pluripotency in vitro

The characterization of Embryonic Stem (ES) cells mostly results from the study of teratocarcinomas ("monster tumors") known for ages as tumors containing many types of differentiated tissues. Their exploration truly started after their transplantation was established, allowing their continuous maintenance through serial grafting (Leroy C. Stevens and Little 1954). Strikingly, single but not all cells arising from those tumors were able to regenerate a new multilineage cellular mass that was transplantable when injected into a recipient animal. This was demonstrating the existence in the tumor of a sub-population of cells, called Embryonic Carcinoma (EC) cells, able to both self-renew and generate multiple cell types (Kleinsmith and Pierce 1964; Martin 1981). Surprisingly, those cells morphologically resembled cells present in the early mouse embryo prior to gastrulation and were able to form three dimensional structures when aggregated together showing common characteristic of the early embryo (so called embryoid bodies) (Martin and Evans 1975). They also had markers of early embryos (Martin 1980). In addition, it was shown that the same kind of tumors could be generated when some of the cells originating from an early mouse embryo were grafted into extra-uterine sites of an adult mice (L. C. Stevens 1968; Solter, Skreb, and Damjanov 1970; Diwan and Stevens 1976). Blastocyst injection then opened new routes to explore the relationship between EC cells and embryo development. Many of them failed to incorporate or caused abnormal development and tumors probably due to their abnormal karyotypes. But few EC cells injected in the blastocyst gave rise to what is called "chimeric" embryos originating

from both recipient and grafted cells. The link between EC cells and early development of the embryo was clearly established (Bradley et al. 1984). It was shown later on that pluripotent cells could actually be derived from the aforementioned PGCs and were most likely at the origin of teratocarcinomas (Matsui, Zsebo, and Hogan 1992; Resnick et al. 1992; Leitch et al. 2013).

Stable EC cell lines were clonally derived and cultured on non-dividing feeder cells and their culture conditions were optimized to increase their proliferation efficiency and their differentiation potential (Kahan and Ephrussi 1970). Two independent teams therefore tried to derive in identical conditions cells coming from the early embryo before implantation. This was finally achieved in 1981 (Evans and Kaufman 1981; Martin 1981). These cells derived *in vitro* from early stage embryos were able to self-renew clonally, gave rise to teratocarcinomas when injected in adult compartments, and finally produced chimeric animals when injected back in pre-implantation embryos and could contribute to the germ line of the new born (Bradley et al. 1984). The so-called mouse Embryonic Stem Cells were born.

C. Origin and properties

Embryonic Stem cells are derived from early blastocyst between E3.5 and E4.5 originating from the ICM compartment. However, depending on the culture conditions, transcriptionally they rather resemble to the naïve epiblast cells (Boroviak et al. 2014). They are usually derived either on feeder cells (mouse embryonic fibroblasts) or gelatine coated surfaces. ES cells exert the two properties defining a stem cell: they maintain their ability to differentiate in all the tissues that will constitute an adult body (pluripotency) through virtually infinite divisions when kept in particular conditions in vitro (self-renewal).

The ability of ES cells to contribute to many tissues including the germline when injected into an embryo combined with the increasing easiness of their genetic manipulation, selection and clonal expansion radically improved the generation of genetically modified animals. Notably, homologous recombination allows the precise and specific targeting of a given genome sequence and opens the way to dissect gene function and regulatory genetic effects (Doetschman et al. 1987; Thomas and Capecchi 1987).

D. The spectrum of pluripotent cells

Besides ES cells, other *in vitro* derived cells form distinct origins are pluripotent and able to self-renew. First, the aforementioned EC cells of tumor origin show these two properties despite their relative genomic instability and their poor capacity to contribute to chimeric animals (Bradley et al. 1984). Second, Embryonal germ (EG) cells that can be derived from the primordial germ cells (PGCs) found in the post-implantation embryo strongly resemble ES cells despite showing clone-dependent erasure of genomic imprinting. This property is most likely at the origin of the development of teratocarcinomas *in vivo* through abnormal reprogramming possibly mediated by the ectopic activation of the LIF/Stat3 pathway (Leitch et al. 2013). Mouse pluripotent cell lines could as well be derived from post-implantation epiblasts (called Epiblast stem cells, EpiSCs). EpiSCs show dependency to different signalling pathways and a distinct expression profile of pluripotent markers compared to ES cells. They can form teratocarcinomas showing multiple lineage origins, a hallmark of pluripotent cells, but are unable to contribute to chimeras upon blastocyst injection (Brons et al. 2007; Tesar et al. 2007). A low E-cadherin expression has been suggested to be responsible of this inability, impeding EpiSCs to integrate into the pre-implantation embryo (Ohtsuka, Nishikawa-Torikai, and Niwa 2012). However, EpiSCs are still able to contribute to the development of various lineages, including the germ line, when grafted at specific locations in the post-implantation embryos (Y. Huang et al. 2012; Kojima et al. 2014). Based on the detailed characterization of the markers of the pluripotent state, Takahashi and Yamanaka (2006) managed to induce the reprograming of differentiated cells through the artificial expression of pluripotency-associated transcription factors (named reprogramming factors). The so called iPSCs (induced pluripotent stem cells) show identical properties to ES cells. The reprogramming process involves the combination of the silencing of the somatic program and the induction of the pluripotency-associated gene network. It was shown that the somatic cell type of origin as well as the reprogramming culture condition greatly impact the requirement and mechanisms of the reprogramming progression. These artificially reset cells have now been produced from a wide diversity of differentiated tissues in both mice an human but have also been shown to keep discreet traces of their tissue of origin due to incomplete epigenetic erasure (Roost et al. 2017; Shi et al. 2017; Apostolou and Stadtfeld 2018).

E. The maintenance of the pluripotent state

While pluripotency is a relatively brief and transient property of the early embryo, it was shown to be extraordinarily stable *in vitro*. Given their great potential for regenerative medicine as well as the excellent model that they constitute to investigate early cell fate choices, a massive effort was made to understand the major cues regulating the specific properties of mouse ES cells. We will therefore briefly review some of the key identified mechanisms leading to mouse ES cells self-renewal.

II. Signalling pathways regulating pluripotency

Cells constituting the early developing embryo are subjected to numerous external cues coming from adjacent or more distal cells, mechanical constrains or maternal tissues, among others. They have to properly integrate and respond to these signals in order to follow the proper developmental path of the embryo. Mouse ES cells nicely recapitulate these properties and show high responsiveness to their environment. However, we will mainly focus on three distinct signalling pathways regulating mouse ES cells self-renewal despite the implication of other extracellular signals: the LIF, FGF and Wnt pathways **(Fig. 1.2.)**.

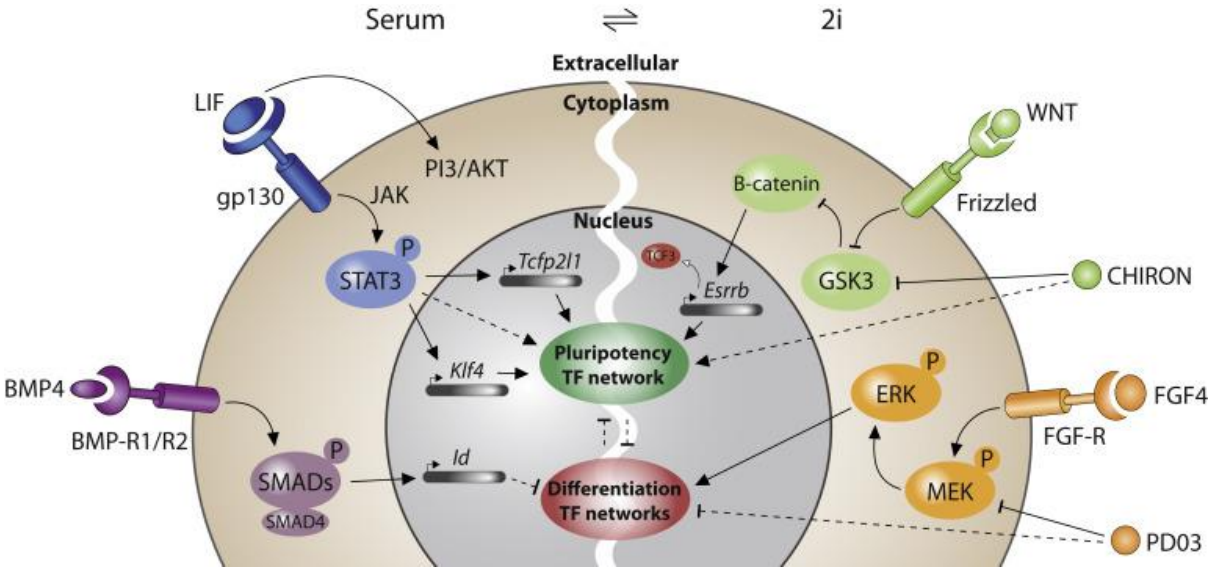


Fig. 1.2. Schematic representation of the main signalling pathways regulating the self-renewal of mouse ES cells, from (Hackett and Surani 2014).

A. LIF signalling

The Leukaemia Inhibitory Factor (LIF) belongs to the interleukin (IL)-6 family cytokines regrouping the ligands that require the receptor co-activator glycoprotein 130 (gp130) to transduce their signal. The LIF was first reported in 1984 (Koopman and Cotton 1984) and more precisely described later on (A. G. Smith and Hooper 1987) as a secreted factor present in Buffalo rat liver extracts with differentiation-inhibiting activity (thus named DIA) for ES cell and EC cells. The provision of this factor was shown to eliminate the requirement for feeder cells in pluripotent cells culture. Independently, the gene of a protein identified as a differentiation-promoting and proliferation-inhibiting factor for murine myeloid leukemic cells

in vitro, therefore called LIF, was cloned (Gearing et al. 1987). Soon after, the exact connexion between DIA and LIF was demonstrated. It was further showed that ES cell cultured in LIF-containing medium without feeder cells were able to keep their pluripotent status through numerous passages and give rise to a high degree of chimaerism after transplantation in the blastocyst (Austin G. Smith et al. 1988; Williams et al. 1988). It was thus remarkable that a unique cytokine has such opposite effects in two distinct cellular contexts. Feeder cells were later shown to produce the LIF in a form that is first associated to the cell surface before being released in the medium (Rathjen et al. 1990). Other members of the IL6 family have since been shown to be able to stimulate the LIF pathway and partially substitute for the need of LIF activation to promote ES cell self-renewal (Jennifer Nichols, Chambers, and Smith 1994; Rose et al. 1994).

The LIF acts through its specific transmembrane receptor, the LIF receptor (LIFR) (Gearing et al. 1991). The binding of LIF to the LIFR mediates the recruitment of gp130 and their heterodimerisation (Gearing et al. 1992). This contact activates the gp130-associated JAK kinase by a trans-phosphorylation process which in turn further phosphorylates the LIFR and gp130. Phosphorylated gp130 is bound by Signal transducer and activator of transcription 3 (Stat3) which can then be activated by JAK phosphorylation. JAK kinase can also activate the extracellular signal-related kinase (ERK) and phosphatidyl inositol-3 kinase (PI3K) pathways. However LIF-stimulation of these pathways doesn't promote mouse ES cells self-renewal. The activation of ERK is actually antagonistic to self-renewal and a mutated form of gp130 precluding ERK activation leads to enhanced self-renewal (Boeuf et al. 1997; Matsuda et al. 1999; Burdon et al. 1999; Hitoshi Niwa et al. 2009; Martello, Bertone, and Smith 2013). The activation of Stat3 triggers its dimerization and translocation into the nucleus where it transcriptionally regulates downstream targets (Boeuf et al. 1997; H. Niwa et al. 1998; Matsuda et al. 1999) The disruption of the LIF signalling by LIF deprivation, LIFR or Stat3 depletion or chemical inhibition of JAK lead to the differentiation of mouse ES cells cultivated in serum (Chambers et al. 2003; Hall et al. 2009a). Despite the critical role of the LIF signalling in the maintenance of pluripotency *in vitro* in FCS/LIF, no evidence for early development defects have been reported for LIF, LIFR, gp130, JAK1 null embryos (Stewart et al. 1992; M. Li, Sendtner, and Smith 1995; Yoshida et al. 1996; Rodig et al. 1998; Nakashima et al. 1999). However, it was reported that Stat3 knockout embryos show abnormal development of the epiblast at E6.5 stage (Takeda et al. 1997). Importantly, gp130 KO embryos have been shown not to be able to survive diapause, a process where mouse embryos are arrested at the late

blastocyst stage due to their impossible implantation induced either naturally in lactating females or artificially upon ovariectomy. Therefore, the apparent discrepancy between the requirement of LIF signalling for the *in vitro* culture of mouse ES cells in FCS/LIF and early embryo development seems to find an answer with the idea that LIF/gp130 signalling is important for the long-term maintenance and survival of pluripotent epiblast (J. Nichols et al. 2001). An idea that has to be however revisited since Stat3 expression and activation through the LIF signalling was more recently shown to be essential in the maintenance of pluripotency in the ICM but not in the blastocyst (Do et al. 2013). Interestingly, mouse ES cells states that would resemble the *in vivo* diapause embryo were shown to be induced by Myc depletion or inhibition of the mTor complex but the effect of LIF deprivation was not analysed in these contexts (Bulut-Karslioglu et al. 2016; Scognamiglio et al. 2016). Due to its crucial role in the maintenance of mouse ES cells self-renewal, an intensive effort has been deployed on the identification and characterization of the downstream targets of the JAK/STAT3 signalling pathway. Several factors, including c-Myc, Klf4, Tfcp2l1, Klf5, Tbx3, Gbx2 and Esrrb (Cartwright et al. 2005; X. Zhang et al. 2008; Bourillot et al. 2009; Hall et al. 2009a; Hitoshi Niwa et al. 2009; Martello, Bertone, and Smith 2013; Tai and Ying 2013; Ye et al. 2013a; D. Huang et al. 2018) have been identified as direct targets of the LIF pathway and shown to allow for diverse levels of compensation of LIF withdrawal effects. Furthermore, genome-wide studies indicate that Stat3 and its downstream targets are involved in the activation of Oct4, Sox2 and Nanog in ESCs, hence connecting the LIF pathway to the core of the pluripotency network (Xi Chen et al. 2008; Hitoshi Niwa et al. 2009; Do et al. 2013). Stat3 recruitment on its transcriptional targets was further reported to be facilitated by Brg1, a component of the esBAF chromatin remodelling complex. Brg1 was shown to be necessary for the targeting of Stat3 to many genomic places by antagonising the effect of the Polycomb PRC2 complex that triggers the organisation of facultative heterochromatin through the deposition of the repressive H3K27me3 histone mark (Ho et al. 2009, 2011; Novershtern and Hanna 2011; Singhal et al. 2014). One of the most responsive genes to LIF signalling is the Suppressor of cytokine signalling 3 (Socs3) (Duval et al. 2000; Krebs and Hilton 2009). Socs3 is actually involved in a negative feedback loop repressing Stat3 activation. It acts by intercalating its kinase inhibitory domain in the activation domain of JAK1. Its overexpression leads to differentiation of mouse ES cells in FCS/LIF due to the strong repression of the LIF stimulating effects. However, its loss, while triggering higher activation of Stat3 and JAK1 also tend to promote spontaneous differentiation towards endodermal lineages most likely due to the concomitant

hyper-activation of the MAPK pathway by JAK kinase (Duval et al. 2000; Nicholson et al. 2000; Schmitz et al. 2000; Forrai et al. 2006; Boyle et al. 2009).

However, the LIF signalling alone without serum is not sufficient to maintain self-renewal (Ying, Nichols, et al. 2003). In fact, the main messenger provided by serum addition was shown to be the Bone morphogenetic protein (BMP). Indeed, BMP provision can bypass the need for serum by inducing the Id (Inhibitor of DNA-binding/differentiation) proteins. Ids inhibit the activity of master differentiation factors as Pax, bHLH, myoD, Ets, mash1 by heterodimerisation with these helix-loop-helix transcription factors (Ying, Nichols, et al. 2003; Ying, Stavridis, et al. 2003; K. Zhang et al. 2010; Davies et al. 2013). Therefore, upon serum withdrawal, ES cells continue to proliferate but differentiate over five to six days, mostly into neural precursors. However, BMP pro self-renewal activity stricly depends on LIF co-stimulation as LIF withdrawal in serum containing medium leads to non-neural differentiation (Wiles and Johansson 1999; Ying, Nichols, et al. 2003; Malaguti et al. 2013).

B. FGF signalling

As previously mentioned, *in vivo*, the FGF signalling activity is detected during the early development of the embryo and is a key pathway regulating the segregation of the PrE and the Epi (G. Guo and Smith 2010; Frankenberg et al. 2011; Kang et al. 2013; Bessonnard et al. 2014). *In vitro*, the inactivation of this pathway blocks the spontaneous differentiation and enhances the self-renewal of mouse ES cells (Jennifer Nichols et al. 1998; Burdon et al. 1999; Kunath et al. 2007; Stavridis et al. 2007; Ying et al. 2008; Joo et al. 2014). Indeed, ES cells produce significant amounts of Fibroblast growth factor 4 (Fgf4) and express its associated receptor, Fgfr2. Although initially supposed to be an autocrine self-renewal stimulator, the deletion of Fgf4 in mouse ES cells clearly demonstrated its role in differentiation (Wilder et al. 1997). Further, the autocrine secretion of Fgf4 prompts ES cells towards differentiation (Kunath et al. 2007; Stavridis et al. 2007). This secreted factor lays upstream of the MEK/ERK kinases pathway in mouse ES cells. Accordingly, Erk2 KO cells have been shown to be strongly resistant to differentiation (Kunath et al. 2007). The mechanisms through which the FGF signal is driving differentiation are diverse. First, Erk has been shown to directly operate on the pluripotency factors by phosphorylation triggering their ubiquitination and degradation, as shown for Klf4 and Klf2 (M. O. Kim et al. 2012; Yeo et al. 2014; Dhaliwal et al. 2018). Second, it was shown that Erk is also directly recruited to chromatin and strikingly, towards Polycomb complex targeted regions, strongly enriched for developmental regulators in ES

cells. The Polycomb complex is known for its role in the formation of facultative heterochromatin and deposition of its associated repressive histone mark, H3K27me3 (see below). It was further shown that Erk has a positive impact on the recruitment of this complex and the formation of repressive domains thus participating in the inhibition of its target genes. However, the tethering of Erk at these genomic loci is also associated with a significant enrichment in the Serine-5 phosphorylated form of RNA-Pol II, involved in the initiation of transcription. A phosphorylation mark that was additionally shown to be triggered, among others, by Erk. Differentiation accompanied by activation of these developmental genes was further associated with a reduction of Erk positioning at these sites. Therefore, it seems that Erk plays a dual role on key developmental regulators activation as Eomes, Otx2 or the Hox genes. It simultaneously promotes transcription initiation, fitting with the conception of Erk as a differentiation stimulator, but also prevents further expression through Polycomb complex cooperation, therefore poising these genes for rapid activation upon differentiation cues (Tee et al. 2014). As most signalling pathways, Fgf signal holds a negative feedback loop mechanism. A CRISPR/Cas9 screening identified the Rsk1 kinase, a downstream phosphorylated target of the Fgf pathway, as a strong repressor of Erk activation. Genetic or chemical invalidation of Rsk1 resulted in elevated Erk phosphorylation leading to abnormal kinetics of differentiation characterized by the faster extinction of naïve pluripotency factors and activation of epiblast and lineage specific markers (Nett et al. 2018). The chemical inhibition of Erk (by PD0325901, one of the two components of the "2i" medium) also sustains robust ES cell self-renewal in the presence of LIF or upon simultaneous inhibition of a repressor of the Wnt pathway (Ying et al. 2008).

C. Wnt signalling

The main effector of the canonical Wnt pathway is β-catenin. In the absence of Wnt stimulation, β-catenin is phosphorylated by GSK3β with the support of additional scaffold proteins leading to its degradation. When the Frizzled (Fzd) receptors of the Wnt pathway are stimulated, β-catenin is stabilized and interacts with the Lef/Tcf factors which normally act as transcriptional repressors (Clevers 2006). In mouse ES cells, the interaction between β-catenin and Tcf3 leads to the abolition of its repressing effects by dissociating Tcf3 from its DNA-binding sites (C.-I. Wu et al. 2012; Shy et al. 2013). Importantly, Tcf3 deletion in ES cells phenocopies the chemical inhibition of GSK3β by increasing resistance to differentiation (Pereira, Yi, and Merrill 2006; G. Guo et al. 2011; Wray et al. 2011) suggesting that the main

target mediating β-catenin effect is Tcf3. In addition, Tcf3 has been shown to colocalize with many pluripotency factors on key regulatory regions to directly repress pluripotency factors as Nanog and is clearly established as a crucial factor to instruct early differentiation of mouse ES cells (Pereira, Yi, and Merrill 2006; Cole et al. 2008; G. Guo et al. 2011; Martello et al. 2012; Leeb et al. 2014). Another important role for β-catenin to sustain self-renewal has been suggested in cell adhesion (Lyashenko et al. 2011). β-catenin pathway has also been shown to repress neuronal differentiation (Kielman et al. 2002; Haegele et al. 2003), this was further supported by a recent study showing that methylation or acetylation of the lysine 49 of β-catenin protein orchestrates a balance between repression of neuronal differentiation factors and activation of mesodermal markers (Hoffmeyer et al. 2017). The chemical inhibition of GSK3β (by CHIR99021) thus mainly prevents β-catenin degradation and is achieved by the second of the two component of the "2i" medium (Ying et al. 2008). In this condition, by inhibition of MEK/ERK and GSK3β kinases, spontaneous differentiation of mouse ES cells is abolished, LIF and serum stimulation become facultative, and the pluripotent state is considered as naïve.

III. Transcription Factors-based regulation of pluripotency

A. Pluripotency factors

"Pluripotency factors" qualify internal regulators of pluripotent cells that specifically orchestrate the maintenance of the ES cell state. Therefore, genes that are involved in general functions such as metabolism, cytoskeleton or transcription, are not included in this category even though they might be necessary for ES cell self-renewal. Many of these factors have been identified with diverse roles and regulatory skills, converging towards the same outcome, the propagation of mouse ES cells identity throughout cell division. We will first focus on what has been proposed to be the three major pluripotency factors: Oct4, Sox2 and Nanog. Then, we will concentrate on a functional property shared by several pluripotency factors: the induction of LIF-independent self-renewal. Finally, the role of Otx2 in the early establishment of differentiation will be examined.

B. Oct4 (Pou5f1)

The octamer-binding transcription factor 4 (Oct4) is considered to be the central pluripotency factor. It was historically the first transcription factor to be identified as strongly associated with early development and pluripotency (H. R. Schöler et al. 1989; Hans R. Schöler et al. 1990; Okamoto et al. 1990; Palmieri et al. 1994). Oct4 is detected in oocytes, during the first cleavages of the embryo, then restricted to the inner cell mass of the blastocyst and later detected exclusively in the germ cell lineage. *In vitro*, Oct4 was found only in EC, EG, and ES cells and was lost upon differentiation of these different cell types. Oct4 KO embryos fail to develop a pluripotent ICM despite developing until the blastocyst stage. Instead, ICM cells differentiate into extraembryonic trophoblast lineage and the resulting absence of pluripotent cells, in turn, impairs the proper development of their surrounding trophoblastic cells inducing the premature death of the embryo (Jennifer Nichols et al. 1998). No pluripotent cell lines could be established *in vitro* in the same study showing the essential role of Oct4 in maintaining the pluripotent state. Moreover, induced deletion of Oct4 *in vitro* leads to loss of self-renewal and differentiation (Hitoshi Niwa, Miyazaki, and Smith 2000). Surprisingly, this differentiation is strongly biased towards the trophoblast lineage, which is not a common fate for ES cells. All these results established the vital role of Oct4 in the propagation of pluripotent cells. However, the overexpression of Oct4 does not reinforce ES cell self-renewal. On the contrary, its forced

expression, even at a low extent, also drives differentiation (Hitoshi Niwa, Miyazaki, and Smith 2000). This finding suggested that Oct4 plays a dual role in self-renewal and differentiation and that its level should be tightly regulated to maintain pluripotency. An hypothesis for that fact, is that Oct4 is still expressed in the post implantation epiblast thus also suggesting a role in the progress of lineage commitment *in vivo* (Thomson et al. 2011; Osorno et al. 2012; DeVeale et al. 2013). Nevertheless, interestingly, heterozygous expression of Oct4 was shown to sustain self-renewal however impairing differentiation (Karwacki-Neisius et al. 2013).

C. Sox2

The SRY-box transcription factor 2 (Sox2) is also essential for ES cell self-renewal. Sox2 depletion in mouse ES cells results in trophoblast differentiation, thus phenocopying the loss of Oct4 (Masui et al. 2007). The consequences of its inactivation *in vivo* are less clear due to the persistence of the factor from maternal origin (Avilion et al. 2003) but was shown to be required for trophectoderm formation (Keramari et al. 2010). However, overexpression of Sox2 predisposes mouse ES cells to differentiation in neuroectoderm, mesoderm, and trophectoderm but not endoderm (S. Zhao et al. 2004; J. L. Kopp et al. 2008), showing that Sox2 expression levels should be tightly regulated for efficient self-renewal, as shown previously for Oct4 levels. Sox2 is known as an Oct4 partner with which it binds DNA on a chimeric Oct/Sox DNA element. Oct4 and Sox2 motifs were initially found to be in close contiguity at some regulatory elements found close to pluripotency and developmentally related genes. It was further shown that the juxtaposition of the two motifs was important for such elements to display their activatory functions. Finally, a chimeric motif was proposed to mediate the binding of the two factors in close interaction with each other (D. C. Ambrosetti, Basilico, and Dailey 1997; D.-C. Ambrosetti et al. 2000; J.-L. Chew et al. 2005; Kuroda et al. 2005). Thanks to genome-wide studies, it was further attested that Oct4 and Sox2 share a broad proportion of common targets among the genome of mouse ES cells and that shared binding sites mostly harbour the aforementioned chimeric motif (Loh et al. 2006; Xi Chen et al. 2008). However, in contrast with this common view, the role of Sox2 in mouse ES cells was shown to be mainly mediated through the indirect activation of Oct4, as Sox2 depletion can be fairly compensated by the forced expression of oct4 (Masui et al. 2007), thus questioning the functional relevance of their highly correlated genome-wide binding profiles. Interestingly, as reported for the recruitment of Stat3 at many genomic sites, the chromatin remodeller Brg1 strongly facilitates Oct4 and, to a lower extent, Sox2 binding at a vast number of regions in pluripotent cells as well as during

the reprogramming process of somatic cells (Kidder, Palmer, and Knott 2009; King and Klose 2017) therefore placing this chromatin modifier factor at a central role in the support of pluripotency factors function. Together with Oct4 and Sox2, a third factor was shown to share a great overlap of targets throughout the genome, Nanog, setting the triumvirate of the core pluripotency factors.

D. Nanog

Nanog was first identified in an ES-specific set of genes (Ramalho-Santos et al. 2002) and was later described as a homeobox-containing gene preferentially expressed in pre-implantation embryo and ES cell (S.-H. Wang et al. 2003). It was then further characterized by two independent functional screenings (Chambers et al. 2003; Mitsui et al. 2003) where Nanog was the only retrieved factor whose forced expression in ES cell could circumvent LIF dependency for efficient self-renewal. Given its ability to lock ES cells in a self-renewing state, it was consequently named, inspired from Irish mythology, "Tir na nog" the "land of the ever young". Nanog sequence homologies led the authors of these two studies to link it with the Nk2 homeoprotein family. However, apart from its homeodomain the rest of Nanog protein didn't resemble any previously characterized protein. In addition, Nanog DNA recognition domain seems to be different from the Nk2 consensus binding motif. *In vitro*, Nanog was shown to be expressed in ES and EC cells and rapidly decreasing upon differentiation. Moreover, Nanog was shown to be regulated by Oct4 and Sox2 which directly bind its promoter through a canonical Oct/Sox motif and activate its expression (D. Y. Wu and Yao 2005; Kuroda et al. 2005; Rodda et al. 2005). Nanog expression was shown to start in the early embryo from the 8 cell stage (Chambers et al. 2003; Hart et al. 2004; Dietrich and Hiiragi 2007) and is homogenously expressed in all the blastomeres of the ICM until E3.5, when a reciprocal expression between Nanog and Gata6 appears (Chazaud et al. 2006; Plusa et al. 2008). The subsequent activation of FGF signalling in PrE restricts Nanog expression to the epiblast (Messerschmidt and Kemler 2010; Frankenberg et al. 2011). Therefore, Nanog presented all the characteristics of a key pluripotency factor.

In human ES cells also, NANOG overexpression enhances self-renewal, allowing for culture at lower cell density and in absence of feeder cells (Darr, Mayshar, and Benvenisty 2006). Nanog overexpression further showed additional abilities in supporting pluripotency. It was shown to strongly increase the efficiency of neural stem (NS) cells nuclear reprogramming when fused with Nanog overexpressing ES cells compared to WT. However, Nanog alone was

not able to reprogram NS cell when ectopically expressed alone. This suggests a role for Nanog in orchestrating the instatement of a pluripotent genome/transcriptome in conjunction with other ES cell partners rather than by autonomous activation (José Silva et al. 2006). Nanog was additionally shown to reprogram murine F9 EC cells in ES-like cells and EpiSCs into Epi-iPSCs (Yanmei Chen, Zhongwei Du, and Zhen Yao 2006; Jose Silva et al. 2009). On the contrary, when NS cells nuclear reprogramming was performed by fusion with Nanog KO ES cells, no naïve pluripotent cells could be obtained. It was thus postulated that Nanog expression was strictly required for the acquisition of naïve pluripotency assessed by the ability of the cells to survive in 2i medium (Ying et al. 2008; Jose Silva et al. 2009). This results was later contradicted by showing that Nanog was dispensable for the generation of iPSCs if, for instance, ascorbic acid was added to the culture medium (Schwarz et al. 2014; A. C. Carter et al. 2014). Ascorbic acid (Vitamin C) has been reported to activate DNA (Tet) and histones (Kdm) demethylation enzymes in mouse pluripotent cells (Blaschke et al. 2013; Ebata et al. 2017). However, even though Nanog was not included in the official quartet of the "Yamanaka factors" for generating mouse iPS cells (Takahashi and Yamanaka 2006), when combined with Oct4, Sox2 and Lin28, Nanog could reprogram human somatic cells to pluripotency (Yu et al. 2007). These results suggest an important role of Nanog in the late step of the reprogramming process, which can however be substituted in specific contexts.

While forced expression of Nanog noticeably impaired differentiation, on the contrary, Nanog KO ES cell differentiated massively with parietal endoderm-like morphology while still being able to be kept for serial passaging at low efficiency (Chambers et al. 2003). However, Nanog expression was shown to be dispensable for the maintenance of pluripotency in mouse ES cells (Chambers et al. 2007). Indeed, stable knock-out cells could be propagated for many passages despite their strong tendency to differentiate. Antibiotics selection cassette knocked-in endogenous loci of Nanog allowed those cells to be artificially "purified" when maintained under selection. This most likely indirectly selects for the presence of an upstream activator of Nanog or the absence of a Nanog repressor but is so far unknown. Hence, despite this clever trick, these factual Nanog KO ES cells were able to properly self-renew, were karyotypically stable, expressed the markers of the naïve state of pluripotency (Rex1, Stella) and were still capable of differentiation *in vitro* and *in vivo* upon selection release (Chambers et al. 2007). This clearly assessed the dispensability of Nanog for *in vitro* maintenance of pluripotency.

Remarkably, Nanog was shown to be heterogeneously expressed in ES cell cultured in serum-containing medium with both positive and negative states being reversible, showing dynamic fluctuations of Nanog expression (Chambers et al. 2007). Moreover, all these cells

were still Oct4 positive confirming the contingency of Nanog expression for ES cell self-renewal. However, as Nanog KO cells, Nanog negative cells were associated to a higher probability to differentiate and a lower proliferation or higher death rate. An important number of reports then focused on Nanog heterogeneous expression which has been associated with random monoallelic expression, stochastic burst of transcription or Nanog auto-repression negative feedback loop (Martinez Arias and Brickman 2011; Navarro et al. 2012; Fidalgo et al. 2012; MacArthur et al. 2012; Filipczyk et al. 2013; Faddah et al. 2013; Abranches et al. 2014). These fluctuations, also reported but to a much lower degree in 2i medium, support a model where Nanog different levels allow mouse ES cells for exploring their differentiation potential and eventually respond to external cues without being definitely committed to differentiate. A recent study analysed the mechanisms of 2i medium inhibitors on the suppression of this heterogeneity and reported both the death of Nanog null cells upon addition of the two chemical compounds as well as a rapid activation of Nanog expression (Hastreiter et al. 2018).

Therefore, an important point was to determine whether the dispensability of Nanog to sustain self-renewal of mouse pluripotent cells *in vitro* also existed *in vivo*. It was first reported that Nanog null ES cells could be grafted in recipient embryos and participate in the development of chimeric animals excluding the PGCs pool (Chambers et al. 2007). This stated the notion of a non-essential role of Nanog in embryo development. However, Nanog null embryos were later shown to be unable to give rise to the pluripotent cells of the ICM rather leading to trophoblastic differentiation and a consequent absence of development of the PrE. The same result was observed when these embryos were derived in culture (Jose Silva et al. 2009; Messerschmidt and Kemler 2010; Frankenberg et al. 2011). It was therefore demonstrated that Nanog is strictly necessary for the establishment of pluripotency *in vivo*. Inversely, the role of Nanog in the development of germ cells that was reported to be crucial was recently reconsidered showing that Nanog null cells can actually complete germ line development despite a broad depletion of PGCs number (M. Zhang et al. 2018).

Therefore, in order to properly interpret the preceding reports and unravel the functional relevance of Nanog in these various contexts, a precise understanding of the mechanisms through which Nanog instates or supports pluripotency and self-renewal of pluripotent cells has to be settled. So far, many studies point out the role of the activation or the repression of a given factor consequently not providing a very clear picture of Nanog regulatory function genome-wide, some of these studies will be reviewed thereafter. The redundancy of the

pluripotency network certainly participates in obscuring our understanding of these mechanisms.

It was first reported that Nanog was not intervening in the LIF pathway and was neither a canonical target of Stat3 transcription factor (Chambers et al. 2003). More recently, Nanog was reported to act in synergy with the LIF cascade on the phosphorylated form of Stat3 (pSTAT3) (Stuart et al. 2014). Nanog was proposed to directly repress Socs3 expression to potentiate the effect of the LIF stimulus. Nanog and LIF pathway were also shown to synergistically act on Klf4 induction, one of the most responsive pluripotency factor to LIF stimulation (Stuart et al. 2014). Nanog actually shows an intricate relationship with Klf factors 2, 4 and 5 and the LIF pathway. Indeed, Klf4 and 5 have been shown to be responsive targets of the LIF pathway, they also activate Nanog and reciprocally. Instead, Klf2 is not activated by LIF stimulation but also positively regulates Nanog. All of them share a multitude of common genomic targets with Nanog, thus strongly suggesting a broad redundancy of all these factors when co-expressed, also supported by their common propensity to support mouse ES cells self-renewal. Given this tight and redundant association, it is not surprising that Nanog and the Klf factors are all dispensable for mouse ES cells self-renewal. (Hitoshi Niwa, Miyazaki, and Smith 2000; J. Jiang et al. 2008; Parisi et al. 2008; Hall et al. 2009a; Bourillot et al. 2009; P. Zhang et al. 2010). The LIF Pathway was also proposed to have an effect on Nanog protein stability. It was reported that the deubiquitinase USP21, a transcriptional target of the LIF/STAT3 pathway, deubiquitinates and stabilizes Nanog protein (Jin et al. 2016). In addition, USP21 is also a phosphorylation target of ERK signalling leading to its dissociation from Nanog and the increased degradation of the latter. This study therefore showed how a balance between the LIF and ERK pathways can directly act on the expression level of a pluripotency factor at the post-transcriptional level. The ERK pathway was also proposed to negatively impact Nanog effect by two additional mechanisms. First, ERK1 was shown to directly phosphorylate Nanog protein decreasing its stability (S.-H. Kim et al. 2014, 1). Second, by activating Brf1 that is responsible for the destabilization of pluripotency factors messenger RNAs, in particular Nanog, and promoting mesendoderm commitment (Tan and Elowitz 2014). In return, Nanog was shown to inhibit some of the canonical positive ERK targets as Sox17, and reciprocally activate some of its negative targets as Klf2 (Festuccia et al. 2012; Yeo et al. 2014; Hamilton and Brickman 2014).

Several studies revealed the genome-wide binding of Nanog in mouse ES cells and showed that Nanog, Oct4 and Sox2 share a substantial number of common genomic targets

(Loh et al. 2006; Xi Chen et al. 2008; Whyte et al. 2013). However, Nanog is not able to promote self-renewal in absence of Oct4 and both factors show relatively different features concerning the way they support self-renewal (Hitoshi Niwa, Miyazaki, and Smith 2000; Chambers et al. 2003, 2007). Therefore, a still open question is to understand the relevance of this extensive overlap of binding for rather functionally distinguishable factors. The three core pluripotency factors remarkably bind most of the genes playing a role in the regulation of mouse ES cells self-renewal or down regulated upon differentiation, starting form themselves. RNAi experiments showed that Nanog and Oct4 can both activate or repress gene expression with a bias towards the activation ability (Loh et al. 2006) Remarkably, many genes showing binding of Nanog in their vicinity didn't show any transcriptional response upon Nanog loss suggesting that the redundancy of binding of the factors may mask the effect upon loss of a single one. Surprisingly, Nanog overexpression was additionally shown to trigger a rather partial and variable rescue of the expression of pluripotency factors upon retinoic acid directed differentiation for 2 days (Loh et al. 2006), suggesting that not all of the pluripotency factors expression is necessary for Nanog-mediated blocking of lineage commitment. Strikingly, the best binding motifs predicted for Nanog from two independent studies *in vitro* (Mitsui et al. 2003; Jauch et al. 2008) and identified from two studies based on Nanog binding regions in mouse ES cells (Loh et al. 2006; Xi Chen et al. 2008) showed an astonishing discrepancy, raising the question of Nanog DNA-binding specificity and recruitment to chromatin.

Furthermore, Nanog has been demonstrated to physically interact with many partners (J. Wang et al. 2006; Costa et al. 2013; Gagliardi et al. 2013). Many pluripotency factors are found in the list (Oct4, Sox2,, Esrrb, Klf5) as well as more generic factors as chromatin structure regulators (NuRD, Polycomb, LSD1). Yet, the relevance of all these contacts is still to be determined. However, for instance, Nanog interaction with Tet1 and Tet2 DNA demethylases importance was demonstrated in the activation of gene expression in mouse ES cells and during the reprogramming process (Costa et al. 2013).

Another approach aiming at unravelling the key downstream targets of Nanog function consisted in identifying the best transcriptionally responsive genes upon Nanog induction. One of those appeared to be Esrrb (Festuccia et al. 2012). Esrrb can variably substitute for Nanog function in LIF-independent self-renewal and reprogramming of EpiSCs as well as PGCs development (Festuccia et al. 2012; Martello et al. 2012; M. Zhang et al. 2018). Similarly to Nanog, Esrrb is dispensable for ESC self-renewal and quickly down-regulated during exit of naïve pluripotency. Ncoa3 was shown to be an important cofactor of Esrrb transcriptional activity and both bind at many genomic positions together with Oct4/Sox2/Nanog (Percharde

et al. 2012). However, Esrrb could barely substitute for Nanog function in providing LIF-independent self-renewal suggesting that additional mechanisms are employed by Nanog to mediate this function.

Conversely, the study of the repression of Nanog target genes seems particularly relevant when considering the effect of Nanog in preventing differentiation. Indeed, several Nanog repressed genes have been reported to prime ESCs towards differentiation: Otx2, Tcf15, Pou3f1, Sox17 (Niakan et al. 2010; Festuccia et al. 2012; D. Acampora, Di Giovannantonio, and Simeone 2013; Davies et al. 2013; Zhu et al. 2014; Buecker et al. 2014; Dario Acampora et al. 2017). Therefore, it is possible that the repression of these factors also plays an essential role in preserving the equilibrium between self-renewal and differentiation of mouse ES cells orchestrated by Nanog. In support of this idea, several repressive complexes, including PRC1 and PRC2 complexes were identified as Nanog partner proteins (Gagliardi et al. 2013) while many of these developmental factors are targets of these repressive complexes (Azuara et al. 2006; Bernstein et al. 2006; Leeb and Wutz 2007). However, despite the reported tendency of pluripotency and polycomb factors to colocalize on repressed loci (Boyer et al. 2006), the mechanisms of polycomb complexes recruitment to chromatin are still elusive (Holoch and Margueron 2017). Moreover, Nanog and PRC2 have been reported to have largely non-overlapping binding profiles genome-wide (Xi Chen et al. 2008), it's thus unlikely that Nanog would be responsible for the direct recruitment of this complex to its genomic targets.

Therefore, several major aspects of Nanog supportive activity to self-renewal still need to be explored and leave open questions: How does functional redundancy of pluripotency factors account for the differentiation block imposed by Nanog? How does the simultaneous binding of Nanog to common binding regions influence the behaviour of its local partners and what is the functional relevance of this redundant recruitment? How does Nanog induce the repression of its targets?

E. LIF independent self-renewal

The overexpression of some of the pluripotency factors allows self-renewal of mouse ES cells in the absence of functional LIF signalling. Since mouse ES cells secrete some LIF which then acts as a paracrine factor, strict LIF independence is usually validated through the inactivation of one of the key LIF signalling components by the use of a JAK inhibitor, a LIFR antagonist (hLIF-05), a dominant-negative form of Stat3 or is assessed in Lifr KO ES cells (Chambers et al. 2003). Among the three core pluripotency transcription factors, only Nanog

has been shown to be able to induce LIF-independent self-renewal of mouse ES cells. This property was reported to happen in the absence of Klf4 and Tbx3 activity (Hitoshi Niwa et al. 2009). Moreover, as aforementioned , expression of many pluripotency factors didn't seem to be rescued by Nanog overexpression during early differentiation upon retinoic acid treatment (Loh et al. 2006) despite the known ability of Nanog to sustain self-renewal in this context (Chambers et al. 2003). From the Klf factors involved in pluripotency (Klf2, 4 and5), only Klf2 has been shown to robustly sustain self-renewal independently of LIF, even though it is the only one of the three not being a target of the LIF pathway (Hall et al. 2009a). However, Klf4 and 5 also showed the capacity to support self-renewal in the absence of LIF but to a lower extent. In addition, even though Klf2 is not as efficient as Nanog in promoting LIF-independence, its effect was shown to be Nanog-independent. However, it is not known whether Nanog can sustain self-renewal in the absence of Klf2 and LIF (Hall et al. 2009a). The overexpression of Esrrb was also reported to provide LIF independence (Festuccia et al. 2012). Here again, while performing much less efficiently than Nanog in that task, this effect was shown to be Nanog-independent. Moreover, in the same study, it was reported that Nanog was not able to confer LIF independency in Esrrb KO cells. However, it's worth mentioning that Esrrb depleted cells have a greatly impaired self-renewal, even in presence of LIF, and it would thus be interesting to validate this result in a different context, starting from a *bona fide* mouse ES cells culture. Another factor has been shown to provide cytokine independent self-renewal, Tfcp2l1 (Martello, Bertone, and Smith 2013; Ye et al. 2013b, 1). Tfcp2l1 has been identified as a major factor of the LIF/Stat3 pathway able to substitute for its upstream regulators. The comparison to the gold standard factor, Nanog, revealed again a lower efficiency. Finally, overexpression of a stabilized mutant of Myc, a downstream target of the LIF pathway, also maintains self-renewal of mouse ES cells in the absence of exogenous provision of LIF (Cartwright et al. 2005). It was shown that upon LIF withdrawal, Thr58 of c-Myc is phosphorylated by GSK3β, resulting in its degradation. However, if a Myc variant where Thr58 is replaced with an alanine residue, is overexpressed, mouse ES cells can self-renew in the absence of LIF and Stat3. Whereas few distinct mechanisms leading to this common phenotype have been proposed, a comprehensive comparison of the self-renewing cells resulting from the overexpression of these different factors in the absence of LIF might elucidate the essential processes required to confer LIF-independence.

F. Otx2

As presented above, signalling pathways related to self-renewal and pluripotency transcription factors are engaged in a self-sustaining crosstalk characterized by a large redundancy of intracellular effectors. However, the naïve pluripotent state is established in the E4.5 epiblast *in vivo* and is rapidly disassembled around E5.5 (Boroviak et al. 2014). Therefore, how is this multi-layered pluripotency network rapidly dismantled in response to external differentiation cues? It has been proposed that the pluripotency network and the differentiation regulators are finely wired to ensure an efficient and coordinated hand over (Hackett and Surani 2014). In addition, as aforementioned, Nanog is heterogeneously expressed in mouse ES cells cultured in Serum/LIF condition and such variations have been proposed to allow ES cells to randomly explore differentiation directions that can be reverted or confirmed according to the environment instructions (Chambers et al. 2007; Martinez Arias and Brickman 2011). Interestingly, Otx2 shows a comparable expression profile, with Otx2 high cells showing reduced Nanog protein level and vice versa (D. Acampora, Di Giovannantonio, and Simeone 2013). Furthermore, Otx2 has been shown to be repressed by many pluripotency transcription factors, including Nanog (Parisi et al. 2010; Festuccia et al. 2012; Grabole et al. 2013; Okashita et al. 2016; Dario Acampora et al. 2017) and to be strongly downregulated in 2i medium (Marks et al. 2012). Additionally, Otx2 was reported to be strongly induced during the conversion of ES cells into EpiSC and to interact with and reposition Oct4 at novel regulatory regions thus allowing the proper activation of EpiSC enhancers (Buecker et al. 2014; Yang et al. 2014).

Otx2 was initially shown to play a key role in the development of neuroectodermal tissues (D. Acampora et al. 1995). However, Otx2 expression was also detected in E3.5 blastocyst and ubiquitously expressed in the epiblast of E5.5 blastocyst (D. Acampora, Di Giovannantonio, and Simeone 2013) suggesting an unanticipated role for Otx2 during peri-implantation development. *In vitro*, Otx2 KO mouse ES cells show profound defects to respond to Fgf stimulation and undergo EpiSC conversion (D. Acampora, Di Giovannantonio, and Simeone 2013). In addition, Otx2 overexpression drives mouse ES cells to differentiate even when cultured in 2i/LIF condition, indicating that Otx2 can actively dissolve the naïve pluripotency network (Buecker et al. 2014). On the basis of few markers, Otx2 overexpressing cells were shown to be comparable to Nanog KO ES cells whereas Otx2 KO ES cells strongly resemble Nanog overexpressing ones (Dario Acampora et al. 2017). The double knock-out of Otx2 and Nanog led also to a primed state with a reduced capacity to integrate pre-implantation

blastocyst. All these results showed a marked antagonism between Nanog and Otx2 orchestrating the balance between primed (EpiSC) and naïve (ES) pluripotent states. Considering the convergent view of all these results, the report showing that the deletion of Otx2 binding site in Nanog promoter led to a decrease of Nanog expression and a conversion of ES cells towards Epi-like cells state was quite surprising, suggesting a positive effect for local Otx2 binding on Nanog expression (Dario Acampora et al. 2016). *In vivo,* the effect of this deletion was consistent with the *in vitro* consequences, showing a down regulation of Nanog in epiblast cells with an expansion of the PrE population, suggesting a clear, unanticipated, role for Otx2 as a transcriptional activator of Nanog expression. This result suggests an ambiguous effect of Otx2 on the regulation of Nanog where low level of Otx2 would promote a direct, positive, effect on Nanog expression whereas higher level would lead to a negative, indirect, effect most likely mediated by ES cells commitment towards Epiblast lineage. However, we cannot exclude that the modification of the 8bp centred on Otx2 binding site within Nanog promoter affected the binding of other activating factors.

IV. Polycomb regulation and bivalent domains in pluripotency

A. Polycomb complexes and functions

Facultative heterochromatin corresponds to genomic regions that are kept silent in a given cell type but retain the potential to be activated by a specific stimulus and therefore corresponds to different regions for every cell type. Such loci usually correspond to developmental genes that require to be expressed under specific temporal and/or spatial contexts to specify the particular features of a given tissue. H3K27me3 is the characteristic mark of facultative heterochromatin and is deposited by Polycomb. Two different Polycomb repressive complexes exist called PRC1 and PRC2. Both polycomb complexes are involved in the deposition of different modifications on distinct histone tails but the resulting effect generally leads to transcriptional repression (Schuettengruber et al. 2017).

PRC2 is responsible for the mono-, di-, and tri-methylation of H3K27 (R. Cao et al. 2002). This histone modification is deposited by the histone methyltransferase Enhancer of zeste homolog 2 (Ezh2) or its homologue Ezh1. Ezh2 is accompanied by additional partners, Eed, Suz12, and Rbap46/48 which help stabilizing the PRC2 complex on its histone substrate and enhance Ezh2 catalytic activity. These four proteins are considered to compose the core of the PRC2 complex but additional components such as Jarid2, Aebp2, and polycomb-like proteins (Pcl1-3) help in regulating PRC2 activity (Margueron and Reinberg 2011). Among the three different methylation states of H3K27, H3K27me3 has been associated to gene expression repression while H3K27me2 has also been found enriched in intergenic regions to block H3K27 acetylation and maintain enhancers inactive. Finally, mono-methylation of H3K27 co-localizes with intragenic active transcriptional units and H3K36me3 (Ferrari et al. 2014).

PRC1 complex is responsible for the ubiquitination of histone H2A on lysine 119 (H2AK119ub) which participates in chromatin compaction and triggers gene silencing. PRC1 is more heterogeneous than PRC2 in its organisation and can be formed by different sets of partners grouped around an essential catalytic core composed by Ring1A/B E3 ubiquitin ligases. The canonical PRC1 complex includes the proteins Bmi1 and a polycomb-like chromobox homolog (Cbx), which recognizes H3K27me3 mark deposited by PRC2 (Morey and Helin 2010). A non-canonical PRC1 complex has also been identified where the Rybp

protein replaces both Cbx and Bmi1 proteins. In this form, PRC1 doesn't recognize H3K27me3, and the mechanism of its recruitment on chromatin is not clearly understood (Gil and O'Loghlen 2014).

B. Bivalent domains and Polycomb regulation in pluripotent cells

In mouse ES cells, H3K27me3 and PRCs can be found at large genomic domains covering repressed developmental regulators (Boyer et al. 2006; Mikkelsen et al. 2007; Endoh et al. 2008; Ku et al. 2008). Inversely, the trithorax system is associated with trimethylation of histone H3 lysine 4 (H3K4me3) and is associated with active transcription site (Kingston and Tamkun 2014). However, a subset of mouse ES cells CpG islands (CGIs) contain nucleosomes marked by both H3K27me3 and H3K4me3 histone marks **(Fig. 1.3)**. The genomic sites containing this antagonistic combination of histone modifications have been referred as "bivalent" (Azuara et al. 2006; Bernstein et al. 2006). Promoters showing the simultaneous presence of these two chromatin marks are also characterized by the occupancy of the aforementioned Serine-5 phosphorylated form of RNA-Pol II associated with transcription initiation and has been shown to be mediated by ERK phosphorylation (Brookes et al. 2012; Tee et al. 2014). The presence of this poised state of RNA-Pol II has been proposed to contribute to the robust activation of developmental genes during the exit from pluripotency and the initiation of differentiation (Voigt et al. 2012). Regarding their equivocal chromatin features, upon differentiation, bivalent domains typically resolve in monovalent H3K4me3 or H3K27me3 at genes that are activated or silenced, respectively, according to their novel cell type (Mikkelsen et al. 2007; Voigt et al. 2012; Ferrari et al. 2014) **(Fig. 1.3)**.

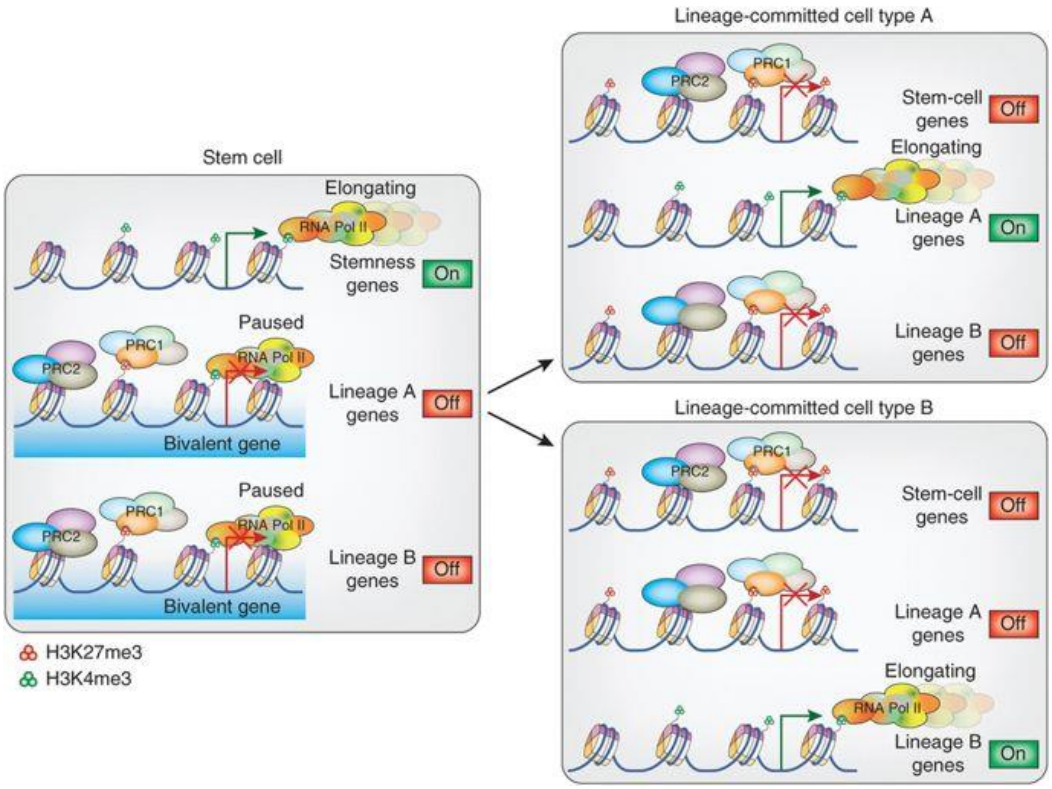


Fig. 1.3. Schematic representation of bivalent domains showing H3K27me3, H3K4me3 and poised RNA Pol II enrichment and their resolution towards differentiation (Di Croce and Helin 2013).

Despite an apparent simple model proposing bivalency as a way to poise key developmental factors for rapid activation, the functional evidences supporting that view are remains relatively limited. The study of KO ES cells for key PRC2 complex components, as Suz12, Eed or Jarid2 indicate that PRC2 complex has a limited impact on self-renewal. Although many bivalent genes tend to get upregulated upon single or combinatorial depletion of PRC2 members, mouse ES cells are not subjected to massive differentiation. However, their differentiation is greatly impaired, with inappropriate silencing of pluripotency factors and abnormal expression of developmental markers in agreement with the severe phenotypes observed in the developing embryo (O'Carroll et al. 2001; Pasini et al. 2007; Shen et al. 2008; Chamberlain, Yee, and Magnuson 2008; G. Li et al. 2010; Landeira and Fisher 2011; Riising et al. 2014). From these results, few studies proposed that Polycomb occupancy could only be the reflection of an absence of transcriptional activity upon a given locus (Chamberlain, Yee, and Magnuson 2008; Riising et al. 2014) with relatively few functional consequences. Remarkably, the phenotype of PRC1 inactivation in mouse ES cells shows more dramatic effects. Ring1b depletion is quickly followed by cell death, morphological changes and

upregulation of PRC1-repressed genes (Leeb and Wutz 2007; van der Stoop et al. 2008) with double Ring1 KO showing even more severe phenotype (Stock et al. 2007; Endoh et al. 2008). Therefore, a possible hypothesis is that the loss of PRC2 and H3K27me3 at bivalent domains is compensated by the continuing presence of non-canonical PRC1 and the maintenance of H2AK119ub preventing a dramatic upregulation of PRC2 targeted genes.

Interestingly, 2i medium culture of mouse ES cells lead to a genome-wide reorganisation of H3K27me3 coinciding with a decrease of the mark at bivalent promoters with no apparent transcriptional change of these genes (Marks et al. 2012). However the loss of H3K27me3 at bivalent CGIs has been shown to come with an increased enrichment in the body of their associated genes (Illingworth et al. 2016). PRC2 complex was further shown not to be responsible nor necessary for gene silencing in 2i medium (Galonska et al. 2015) compared to serum/LIF condition where mouse ES cells appeared to be more unstable upon Eed depletion. One possibility for the absence of transcriptional response in 2i medium, despite an acute loss of H3K27me3, is that Erk inhibition impedes Ser5 RNA-polII phosphorylation at bivalent promoters therefore precluding transcriptional initiation (Tee et al. 2014).

Finally, the mechanism of PRC2 recruitment onto chromatin is still an open question. Polycomb-like proteins (Pcl) have been reported to show high affinity for H3K36me3 histone mark through their Tudor domain as well as for methylated CPGs and have been proposed to be key factors in the recruitment of PRC2 complex to chromatin. Moreover, their activity in mouse ES cells has been associated with both repression of bivalent genes and repression of pluripotency genes upon early differentiation (Ballaré et al. 2012; Brien et al. 2012, 19; Hunkapiller et al. 2012; Cai et al. 2013). Besides, PRC2 complex have been shown to bind RNA with high affinity and many discoveries made in the field of long non-coding RNAs (lncRNAs) suggest that these transcripts may act as scaffolds to guide histone modifying complexes to their genomic targets thus proposing a possible model of RNA-mediated recruitment of PRC2 throughout the genome (J. Zhao et al. 2010; Brockdorff 2013; Kaneko et al. 2013; Kaneko, Son, et al. 2014; Kaneko, Bonasio, et al. 2014a; Brockdorff 2017).

V. Long non-coding RNAs & mouse ES cells

A. Description and characteristics

The improvement of high-throughput sequencing technology has greatly improved our knowledge about the transcriptional activity in coding and non-coding regions of mammalian genomes. For instance, it was estimated that more than 70% of the human genome might be actually transcribed in some conditions while only 1 to 2% of it is supposed to code for protein-related genes (Dinger et al. 2009; K. C. Wang and Chang 2011). An important portion of these non-coding transcription units have been shown to fall under the category of lncRNAs (The FANTOM Consortium et al. 2014).

As explicitly revealed in their denomination, lncRNAs are defined by two common properties: their length must be greater than 200nt to differentiate them from shorter non-coding RNAs (miRNAS, piRNAs) and their coding potential must not indicate the possible production of any or any functional protein. The latter criteria is a matter of debate since many lncRNAs have been shown to interact with ribosomal units, although this is not a real evidence for effective translation (G.-L. Chew et al. 2013; Ji et al. 2015; Carlevaro-Fita et al. 2016; Zeng, Fukunaga, and Hamada 2018), and that some, however predicted as non-coding, were responsible for the synthesis of small functional peptides (Nelson et al. 2016). Indeed, in the case where an open reading frame (ORF) is present within such a transcript, the determination of its non-coding potential is based on the short size of the ORF or its absence of conservation among a large panel of related species (M. F. Lin, Jungreis, and Kellis 2011; L. Wang et al. 2013). An inconvenient drawback of this arbitrary (size) and negative (non-coding) definition of lncRNAs is that they harbour a large diversity of molecules with distinct functional and mechanistic properties that is, however, inherent to their recent discovery (Ulitsky and Bartel 2013).

Numerous sub-categories of lncRNAs have been proposed based on different features such as their location relative to surrounding coding genes (intergenic, sense/antisense overlapping, divergent) or the chromatin context embedding their transcription start site (TSS) (promoter of a coding gene, enhancer, CTCF binding) but will probably need to be reconsidered based on functional or intrinsic properties of the lncRNAs themselves (Mattick and Rinn 2015). While most stable lncRNAs are transcribed by RNA-pol II, possess a 5' cap, are polyadenylated and often spliced, some use different ways for 3' maturation or are

otherwise usually unstable and rapidly degraded transcripts (Wilusz, Freier, and Spector 2008; M. Guttman 2009; Lloret-Llinares et al. 2015; Zong et al. 2016).

The sequence conservation across species of these RNAs is on average much lower than for their coding counterparts (Ulitsky and Bartel 2013). It is still not clearly understood whether this suggests that the selection pressure applied on lncRNAs is based on different criteria than the one applied on mRNAs or if it simply reflects the insignificant functionality of a vast majority of them. Indeed, if the structures of lncRNAs are more important than their actual sequences, which might be suggested by the fact that *in silico* predictions of secondary structures for lncRNAs lead to complex and organised folding (Pegueroles and Gabaldón 2016), it is conceivable that their DNA sequence evolved rapidly under lower constrains.

The average level and tissue-specificity of lncRNAs expression also shows a distinguishable behaviour compared to coding genes. Indeed, the vast majority of lncRNAs are expressed at a lower level than mRNAs, possibly suggesting their specific expression in some sub-populations of cells within a given tissue or cell culture (Cabili et al. 2011; Pauli et al. 2012; Derrien et al. 2012). In agreement with this idea, the expression of lncRNAs is more strongly restricted to a specific tissue or cell type than it is for coding genes. Such specific features argue for a role of lncRNAs in the maintenance of cell identity or in highly specified functional biological tasks (S. J. Liu et al. 2017).

B. LncRNAs function in mouse ES cells

First of all, it's worth mentioning that new ways of action are constantly proposed for lncRNAs. This is not surprising given the incredibly high diversity of transcripts grouped together in the lncRNA family. We will therefore focus on a limited subset of transcripts, chosen to be representative of the diverse functions fulfilled by the lncRNAs, and that have more specifically been shown to play a functional role in pluripotent cells. In particular, the three main classes of regulatory lncRNAs will be mentioned: regulating neighbouring gene(s) in *cis*, regulating in *trans* the transcription of distant genes in the nucleus, acting on regulation of gene expression in the cytoplasm.

Despite the very large number of lncRNAs found to be specifically expressed in mouse pluripotent cells, very few of them have been functionally characterized so far (M. Guttman 2009; Mitchell Guttman et al. 2010, 2011; Lv et al. 2015; Bergmann et al. 2015; Bogu et al. 2016).

Gm15055 is a typical *cis*-regulatory lncRNA (G.-Y. Liu et al. 2016). It is highly expressed in mouse ES cells where it was shown to be positively regulated by Oct4 through a cis-regulatory element. Gm15055 is located 50kb upstream of the Hoxa gene cluster towards which it was shown to recruit the PRC2 complex resulting in the maintenance of H3K27me3 deposition on Hoxa gene promoters and induce transcriptional repression. It was further shown by chromosome conformation capture experiments that Gm15055 locus directly contacts multiple sites of the Hoxa gene cluster in mouse ES cells thus facilitating the cis targeting of Gm15055 RNA to the Hoxa genes.

The lncRNA TUNA (for Tcl1 upstream neuron-associated lincRNA) was first identified in an RNAi screen in mouse ES cells (N. Lin et al. 2014). TUNA was shown to be critical for mouse ES cells self-renewal as well as neural differentiation. It was further reported to have a positive impact on reprogramming efficiency when overexpressed. Remarkably, TUNA was shown to contain a 200-nt long RNA sequence displaying a strong evolutionary conservation across vertebrates allowing for its interaction with three RNA-binding proteins: Ptbp1, hnRNP-K, and Ncl. TUNA RNA and its three partners were further demonstrated to colocalize at the promoters of Nanog, Sox2 and Fgf4, whereas the precise mechanism through which this complex would be recruited at such locus was not assessed. However, Sox2 RNAi experiment revealed that it shares with TUNA a large portion of misregulated genes upon their depletion in mouse ES cells. Given the fact that both genes are involved in neurogenesis and show a highly similar expression pattern along neurodevelopment, the authors suggested a close relationship between Sox2 and TUNA regulatory functions.

Panct1, as TUNA RNA, was first identified in a RNAi screen (Chakraborty et al. 2012) looking for lncRNAs whose expression would be necessary for preserving mouse ES cells self-renewal. It is a sense overlapping lncRNA included in the protein coding gene Tobf1. Interestingly, the depletion of Tobf1 in mouse ES cells also leads to the loss of pluripotency, but more surprisingly Panct1 and Tobf1 were shown to colocalize in the nuclear space of mouse ES cells (Chakraborty et al. 2017) and to specifically form discreet foci in early G1 phase. The DNA binding of Tobf1, shown to be dependent of Panct1, was mapped genome-wide and revealed to overlap significantly with Oct4 binding. Strikingly, mutating an octamer-like motif in Panct1 RNA strongly diminishes the strength of Tobf1, and to a lower extent of Oct4, localization and recruitment to their common targets proposing a regulatory role of Panct1/Tobf1 complex on the recruitment of Oct4 as specific regions in cell cycle-dependent manner.

Linc-RoR (regulator of reprogramming), despite acting at distance from its transcriptional site, represents a radically different class of lncRNAs compared to Panct1 and TUNA RNAs. Linc-RoR is a human-specific cytosolic lncRNA which was first identified as enhancing the reprogramming efficiency of iPSCs when overexpressed (Loewer et al. 2010). However, linc-RoR was later shown to act as a microRNA sponge, buffering the effect of miR-145 and repressing its negative impact on OCT4, NANOG and SOX2 levels (Y. Wang et al. 2013). Its expression was shown to be directly controlled by the pluripotency factors and necessary for human ES cells self-renewal thus creating a self-sustaining feedback loop of the pluripotency network.

Finally, lincU was very recently studied on the basis of its regulation by the pluripotency factor Nanog in mouse ES cells (Jiapaer et al. 2018). LincU was shown to be localized in the cytoplasm where it stabilizes Dusp9 protein, an ERK-specific phosphatase, preventing its ubiquitination and degradation. This effect logically results in the repression of the ERK1/2 signalling pathway activity. Hence, upon depletion of lincU, mouse ES cells self-renewal is severely impaired while its overexpression induces the ground state of pluripotency. Remarkably, lincU is evolutionary conserved in human genome, and its effect on self-renewal is also conserved in human ES cells.

Therefore, it has been clearly established that lncRNAs can display regulatory functions on self-renewal of pluripotent cells through diverse functions and mechanisms involving all the layers of gene expression regulation (transcriptional, post-transcriptional and at the protein level). Given the extremely low percentage of them that have been characterized so far, we can expect that many more of them will find a place in the complex network regulating pluripotency.

VI. From CRISPR discovery to CRISPR activators

A. Historical background

Although genome engineering with CRISPR/Cas9 system is now a routinely used technique to insert, delete or modify DNA sequences in living organisms, it took more than 20 years after its first discovery (Ishino et al. 1987) for this tool to be used as a genome engineering tool in eukaryotic cells. While studying the IAP enzyme in *Escherichia coli*, Ishino et al. cloned and sequenced a 1.7 kb chromosomal fragment containing the iap gene and noticed "An unusual structure […] in the 3'-end flanking region of iap […]. Five highly homologous sequences of 29 nucleotides were arranged as direct repeats with 32 nucleotides as spacing." With the increase of sequenced prokaryotic genomes in the 90s, it appeared that a great portion of bacteria and archaea actually possess the same kind of short repeated sequences. They were frequently organized in clusters but always regularly separated by unique sequences of constant length. They were first called Short Regularly Spaced Repeats (SRSRs) (Mojica et al. 2000) before getting their actual name CRISPR for Clustered Regularly Interspaced Short Palindromic Repeats (Jansen et al. 2002). It was therefore shown that in most species those clusters were flanked on one side by a common "leader" sequence. The repeats and their leader sequences were shown to be conserved within a species, but different between species. Four unique CRISPR-associated (Cas) genes were identified always adjoining a CRISPR locus indicating that Cas genes and CRISPR loci might be functionally linked. It was also shown that those CRISPR arrays were transcribed (T.-H. Tang et al. 2002). A deep bioinformatics study (Haft et al. 2005) dug into uncharacterized genes in the neighbourhood of CRISPR loci and found many additional protein families strictly linked to CRISPR loci across multiple prokaryotic species. They showed that Cas genes number can be larger than previously expected with up to 20 different ones and can also be located between two clusters of repeat. CRISPR loci were later classified on the basis of the Cas proteins they contained and were grouped in three classes, with the class I and III harbouring many Cas proteins acting in complex in opposite to the Class II having few effector Cas proteins. A big step-forward was taken when close analyses focused on the unique spacer sequences separating the clustered repeats and revealed their extrachromosomal, phage or plasmid-associated origins (Pourcel, Salvignol, and Vergnaud 2005; Mojica et al. 2005; Bolotin et al. 2005). It was additionally shown that the presence of exogenous sequences from a given virus within the CRISPR loci was positively correlated with the ability of the prokaryote to resist the viral infection. It was

consequently postulated that CRISPR arrays serve as stable immune memory platforms leading to functional defence against pathogens. In 2007 the link between viral infection, viral spacer sequences insertion in CRISPR arrays, Cas proteins effector functions and viral resistance was clearly established (Barrangou et al. 2007). It was later evidenced that CRISPR transcripts are processed in small RNAs (CRISPR RNAs, crRNAs) containing single spacers to guide by a base-pairing mechanism the Cas nuclease activity (Brouns et al. 2008). A type III CRISPR was shown to act on DNA rather than RNA as it was preventing plasmid conjugation in bacteria (Marraffini and Sontheimer 2008). The same year, the importance of the presence in the target sequence of the short protospacer-adjacent motifs (PAMs) for Cas9-mediated cleavage of DNA was demonstrated (Deveau et al. 2008). However, the proper demonstration of Cas9 as being the only enzyme within its Cas genes cluster able to cleave DNA was done in 2010 (Garneau et al. 2010). A new component of the type II CRISPR systems was characterized the next year: the non-coding tracrRNA (trans-activating crRNA) was shown to be complementary of the repeated sequence of the CRISPR array and to hybridize with the crRNA to allow its maturation by endogenous RNase III and Cas9 protein. It was therefore demonstrated that Cas9 and two short non-coding transcripts only were required for targeted DNA cleavage. The route for targeted genome editing was opened.

It was first shown that the type II CRISPR system and subsequent interference function was transferrable from one bacterial strain to another (Sapranauskas et al. 2011). Soon after, purified Cas9 guided by crRNAs was shown to be able to cleave target DNA in vitro (Gasiunas et al. 2012; Jinek et al. 2012). A chimeric non-coding RNA resulting from a pseudo-fusion of the crRNA and the tracrRNA (called single guide RNA, sgRNA) was then engineered and shown to reproduce the function of the cr and tracrRNA duo in vitro. Finally, two studies reported simultaneously for the first time targeted genome editing in mammalian cells with type II CRISPR Cas9 (Cong et al. 2013; Mali, Yang, et al. 2013). Non Homologous End Joining (NHEJ) or Homology Directed Repair (HDR) mediated genome modification was obtained by heterologous expression of Cas9 and a sgRNA or a crRNA/tracrRNA hybrid allowing modification of a single gene as well as multiple genes at once in different human cell lines. After these pioneer studies, the use of CRISPR/Cas9 system as well as new Cas proteins exploded rapidly to become a widely used technology in numerous different organisms.

B. Summarized way of action

As type II CRISPR/Cas9 has been the most used and characterized CRISPR system so far, we will only focus on its particular properties. The full response of CRISPR/Cas9 system after phage or plasmid invasion is commonly divided in three steps. First, Cas1/2 complex is involved in the cleavage of the exogenous DNA in short sequences and their insertion within the CRISPR array between the crRNA-associated repeats. Those small fragments are not randomly picked from the foreign DNA but selected to be followed by a few nucleotides-long motif (Protospacer Adjacent Motif or PAM) that is necessary for further cleavage by Cas9 protein. Second, the newly modified CRISPR array is transcribed in a long non-functional fusion of crRNA and further matured by cleavage in mature crRNAs by endogenous RNase III. Repeated portion of the crRNA allows for hybridization with the trans-activating CRISPR RNA (tracrRNA) and allows the resulting pair of transcripts to interact with Cas9 protein. Third, Cas9 protein recognizes its specific target through base pairing with the spacer sequence of the crRNA and only if followed by the required PAM sequence subsequently cleaves the invading DNA molecule (Hille and Charpentier 2016; F. Jiang and Doudna 2017).

C. Genome engineering

First, as mentioned before, a fusion of the crRNA and the tracrRNA called single guide RNA (sgRNA) have been shown to efficiently replace the need of separated transcripts. This chimeric RNA reproduces the repeat/anti-repeat hybridisation of the two original RNAs by internal looping and incorporates the three stem loops of the tracrRNA necessary for Cas9 interaction and activation. Cas9 is a bilobed protein including a REC lobe involved in the match detection of the sgRNA and its target sequence and a second lobe including two nuclease domains (called RuvC and HNH based on their biochemical structure homology with known nucleases) and a PAM-interacting domain (PI) responsible for PAM detection (Nishimasu et al. 2014). HNH nuclease domain is responsible for cleavage of the DNA strand complementary to the crRNA while RuvC domain cuts its opposite strand. Cas9 protein has been shown to dramatically change its three-dimensional conformation upon sgRNA binding in such a way that it seems to serve as a real scaffold for the protein allowing nuclease sites to become open and thus potentially active (Jinek et al. 2014; Lim et al. 2016). Once interacting with a sgRNA Cas9 can start scanning the genome. It has been shown that this scanning is first depending on PAM recognition followed by DNA strands separation and RNA-DNA duplex formation

(Anders et al. 2014; Sternberg et al. 2014). Only an extensive pairing between the sgRNA and its target will allow for efficient double strand cleavage by Cas9 three base pairs upstream of the PAM. However, it was shown that a short homology was sufficient for transient binding of the sgRNA/Cas9 complex but that only a high overlap between the sgRNA and the target sequences leads to efficient cleavage (X. Wu et al. 2014a). More precisely, Cas9 specificity has been extensively characterized and showed that Cas9 allows for few off-targets throughout the sgRNA sequence with more sensitivity for mismatches close to the PAM sequence (X.-H. Zhang et al. 2015; Tycko, Myer, and Hsu 2016). Off-targets effects led scientists to derive a semi active form of Cas9 original protein by inactivation of one of its two nuclease domains (Gasiunas et al. 2012; Mali, Aach, et al. 2013; Ran et al. 2013). Thus, two sgRNAs targeting opposite strands were necessary for double DNA strand break greatly increasing the specificity of cleavage genome-wide.

The easiness of sgRNA design and expression vector generation made of CRISPR/Cas9 system a very quickly expanding tool. It has been extensively used for genome engineering in cultured cells, embryos and live organisms as well as a versatile tool for large scale screening for gene inactivation and even started to be used for therapeutics goals.

D. Other kind of versatile DNA-binding protein

Gene targeting pre-existed long before the "CRISPR era". In 1988 it was shown that DNA double strand breaks at specific site greatly increased the efficiency of gene targeting in yeast (Rudin and Haber 1988). A comparable effect was later shown in mammalian cells through the use of the yeast meganuclease I-SceI and its large recognition sequence (Rouet, Smih, and Jasin 1994) producing both HDR and NHEJ. Few modifications of this and other meganucleases were further tested (Epinat et al. 2003) with the goal to build custom endonucleases but its limited plasticity and the apparition of better alternatives rapidly stopped their development.

FokI, a bacterial restriction enzyme working only upon dimerization on dual DNA targets, was shown to be composed of two separated nuclease and DNA-binding domains. This discovery opened the route to the customized design of site-specific DNA targeting enzyme through the use of FokI cutting domain fused to modular DNA-binding protein. In that aim, Zinc Finger proteins were the first kind of DNA-binding protein to be used (Y. G. Kim, Cha, and Chandrasegaran 1996). However, despite their relatively easy design (a single zinc finger

domain interacts with a 3 bp-sequence, (Pavletich and Pabo 1991), they had a major drawback: most of them were highly cytotoxic asking for laborious optimization. A new class of DNA-binding protein was later characterized. Transcription-activator-like effector (TALE) originates from the plant pathogen *Xanthomonas* bacteria. Each repeat domain recognizes a single base thus allowing for the creation of truly flexible site-specific DNA binding domain. When fused to DNA cleavage domain of FokI, TALENs (N for nuclease) had very low cytotoxicity (Guilinger et al. 2014). Thus, a large effort was made to simplify their design and many expression vectors were generated but the development of CRISPR rapidly decreased their use.

E. Transcriptional modulation

Thanks to its flexible DNA binding capacity, many groups rapidly envisaged Cas9 as a new cargo protein for delivery to a specific genomic location of any functional effectors, as done previously with ZF and TALE proteins. The first step on that way was to turn Cas9 into an inert DNA-binding protein. This was done by introducing one inactivating mutation in both RuvC and HNH enzymatic sites (Jinek et al. 2012). Recruitment of the newly created inactive form of Cas9 (dCas9 for enzymatically dead Cas9) was first published (Qi et al. 2013) in the bacteria *E. coli* where its recruitment was shown to strongly impede transcription initiation and elongation at promoter or gene body respectively in a strand-specific manner. This result was confirmed by nascent RNA sequencing and showed a clear block of transcription about 20bp upstream of the gRNA targeted site. This effect was shown to be highly specific by total RNA sequencing. Multiple genes could also be affected at once by multiple gRNAs delivery. Efficiency was clearly higher when targeting the close proximity of the transcription start of the gene (TSS). The system was further applied in human cells to chromosomally-integrated viral fluorescent reporters. In that case, however, gene expression decrease was much lower, reaching in the best case 50% decrease with most of tested gRNAs not producing any effect. This might reflect an intrinsic property of Cas9, originally existing to target DNA in prokaryotic organisms where genome structure is largely different than eukaryotic chromatin. This CRISPR-mediated repression of gene expression by steric competition with RNA-polymerase or DNA-binding activator of transcription was called CRISPR interference (CRISPRi). Another study using CRISPRi in bacteria (Bikard et al. 2013) demonstrated its strong potency in gene repression and further developed an opposite system for transcription activation by fusing to dCas9 the ω subunit of RNA polymerase, previously shown to be

sufficient for transcription induction. When dCas9-ω was recruited close to the TSS (-35bp), repression of transcription was reproduced but when moved further a strong induction could be obtained at specific positions (around -90bp).

CRISPRi was concomitantly developed and optimized in eukaryotic cells by fusing separately to the inactive dCas9 cargo (Gilbert et al. 2013) three different transcriptional repressive domains: the KRAB (Krüppel associated box) domain of Kox1, the CS (Chromo Shadow) domain of HP1α, or the WRPW domain of Hes1. In human cells (HEK293T), KRAB domain was showing the best repressive effect when recruited on the promoter or the coding part of a stably integrated artificial fluorescent reporter, outperforming dCas9 recruitment alone. This effect was shown to be stable over time (up to 5 days), induced with a single gRNA and highly specific of the targeted gene. This was reported on transgenic reporters as well as endogenous genes where dCas9-KRAB frequently performs better than dCas9 resulting to about 50% decrease of protein expression as assessed by FACS analysis. Here again, the effect was shown to work properly for multiplexing by targeting several genes at the same time. The same system was shown to be functional in yeast when KRAB domain was replaced by Mxi1 a yeast transcriptional repressor. An important result was that dCas9-KRAB also works on regulatory elements such as enhancers where dCas9 alone doesn't. However, dCas9 recruitment alone on a Tetracycline responsive element (TRE) can prevent TET-ON activation of a reporter thus suggesting steric competition between dCas9 and other programmable TFs. The efficiency of dCas9-KRAB fusion was assessed in other organisms (Gao et al. 2014; Farzadfard, Perli, and Lu 2013) and later used for genome-wide phenotypical screenings upon gene inactivation (Gilbert et al. 2014). A later study (Kearns et al. 2015) compared LSD1 (an H3K4 and K9me demethylase) and KRAB fusion with dCas9 recruitment on regulatory regions of Oct4 and Tbx3 genes in mouse ES cells. Their results suggested that LSD1 was more efficient at triggering repression when recruited at distal regulatory elements than KRAB domain which was shown to better perform at proximal enhancers or promoters. Their inhibition effect was shown to be achieved without histone occupancy modification or three dimensional re-organization of the targeted locus but rather by post-translational modification of histones. Instead of using histone modifying enzymes, another group (X. S. Liu et al. 2016) used a DNA-methyltransferase fused to dCas9 to induce gene repression. They showed that fusion of Dnmt3a with dCas9 was able to induce methylation of endogenous loci and subsequent efficient gene repression. They could also show that this system was able to block CTCF binding at a given site in a Dnmt3a activity-dependent manner. A 3D modification of

the targeted region was further assessed by 3C assay. The binding specificity of the system was evaluated by ChIPseq experiment and revealed few off-target sites with low binding signal and marginal methylation effects. This latter result is worth being compared with a more recent study (Galonska et al. 2018) showing that dCas9 fusion to the catalytic domain of Dnmt3a (cat3a) was able to non-specifically methylate CpGs at a genome-wide scale in the absence of gRNA. However, this effect was not mediated by dCas9 binding to the genome but rather to an increased concentration of cat3a in the nucleus.

CRISPR activation (CRISPRa) was first developed (Gilbert et al. 2013) by fusing dCas9 to VP64 (four copies of the viral protein VP16 C-terminal domain) or p65 activation domain (a member of the NF-κB family) both known for their transcriptional activation properties. Strong activation was obtained in human cells when targeting the promoter of a stably integrated transgenic fluorescent reporter preceded by a minimal CMV promoter and three repeated targetable sites for a single gRNA. This result was rapidly confirmed with a smaller number of VP16 domains fusion (VP48) (Cheng et al. 2013). But when applied to endogenous loci with a single binding site, absolutely no effect was observed. A higher number of VP16 domain was therefore fused to dCas9. A new dCas9-VP160 was thus tested with a single gRNA, the level of activation was still very poor but a synergistic effect was shown when using a combination of gRNAs targeting the same promoter. These results were further confirmed by following studies using the same experimental set up (Maeder et al. 2013; Perez-Pinera, Kocak, et al. 2013; Farzadfard, Perli, and Lu 2013).

A new generation of CRISPRa systems followed these pioneer tools including three potent technologies: SunTag, SAM and VPR. The SunTag system (Tanenbaum et al., 2014) was initially developed for live tracking of protein dynamics and localization. This was achieved by fusion to the protein of interest of a repeated array of small epitopes (coming from the yeast GCN4 protein) which were specifically recognized by diffusible antibodies (single chain fragments variable) conjugated to fluorescent tags, therefore allowing for great amplification of the fluorescent signal (**Fig. 1.4)**. An elegant adaptation of this tool was set up for CRISPR activation purpose. The repeated epitopes array was fused to dCas9 and the diffusible antibody to VP64 allowing for the local accumulation of multiple copies of VP64. The system was shown to be impressively more powerful than all previously reported CRISPRa systems in mammalian cells generating high induction of endogenous loci with a single gRNA. This system demonstrates its robustness and specificity in a genome-wide screening (Gilbert

et al. 2014) using a broad gRNAs library. Based on the Cas9-gRNA complex crystal structure (Nishimasu et al. 2014) and identified free-of-interaction loops within the gRNA molecule, another group (Konermann et al. 2015) imagined the use of the guide RNA itself as a recruiting component for transcriptional activation. The rationale for this idea was based on the fact that the gRNA is in very close proximity to the chromatin thus potentially producing a stronger effect than when the effector is fused to the large Cas9 protein and possibly far from the targeted DNA sequence. They therefore inserted RNA aptamers recognized by the MS2 protein (Peabody 1993) within the gRNA loops. The system included a dCas9-VP64 fusion, a gRNA with two aptamer loops and a MS2 protein fused with p65 and HSF1, two transactivation domains (**Fig. 1.4)**. This novel CRISPPRa complex was referred as SAM for Synergistic Activation Mediator. SAM performed tremendously better than dCas9-VP64. The authors identified the best window to be targeted for a high efficiency of induction being from -200 to +1bp relative to the TSS of the target gene. As previously reported, it was found that basal level of expression of a gene highly influences the degree of induction, lowest expressed genes being the best induced ones. Multiplexing with up to 10 gRNAs targeting 10 different genes was performed, however producing lower induction than when each was targeted separately. RNA sequencing experiments confirmed the high specificity of the system and validation with a functional screening was performed. Finally, The VPR system (Chavez et al. 2015) came from a screening testing more than twenty fusions of activation domains from Mediator/Transcription Initiation Complex subunits or transcription factors to dCas9 for activation of a fluorescent reporter. VP64, p65 and Rta were identified as the three best transcription inducer proteins (**Fig. 1.4)**. Tested on endogenous genes the dCas9-VP64-p65-Rta outperformed VP64 alone in human cells. Multiple genes could be induced at once with high efficiency and specificity. VPR was shown to work as well in drosophila and mouse cell lines, showing the versatility of the designed system.

A different system was based on the histone acetylase protein p300 ability to promote transcriptional activation (Hilton et al. 2015). The enzymatically active domain of p300 (called p300 core) only was fused to dCas9 (**Fig. 1.4)**. Induction with this new fusion was now as good as or even a little bit better compared to VP64 while both fusions were expressed at the same level. The effect was shown to be fully dependent on the enzymatic power of p300 core. The most noteworthy result of this study is that while VP64 was only able to induce downstream transcriptional response when targeted to promoters, p300 core was able to produce not only better induction when recruited to promoters but also to induce a nice response through

proximal and distal enhancers. However, it is important to notice that good levels of transcriptional induction were always obtained with pooled gRNAs.

A good comparison (Chavez et al. 2016) of most of the CRISPRa systems (VP64, VPR, SAM, Scaffold, SunTag, p300 core, VP160, dual VP-64) was performed and reported SunTag, VPR and SAM systems to be the most efficient ones (**Fig. 1.4)**.

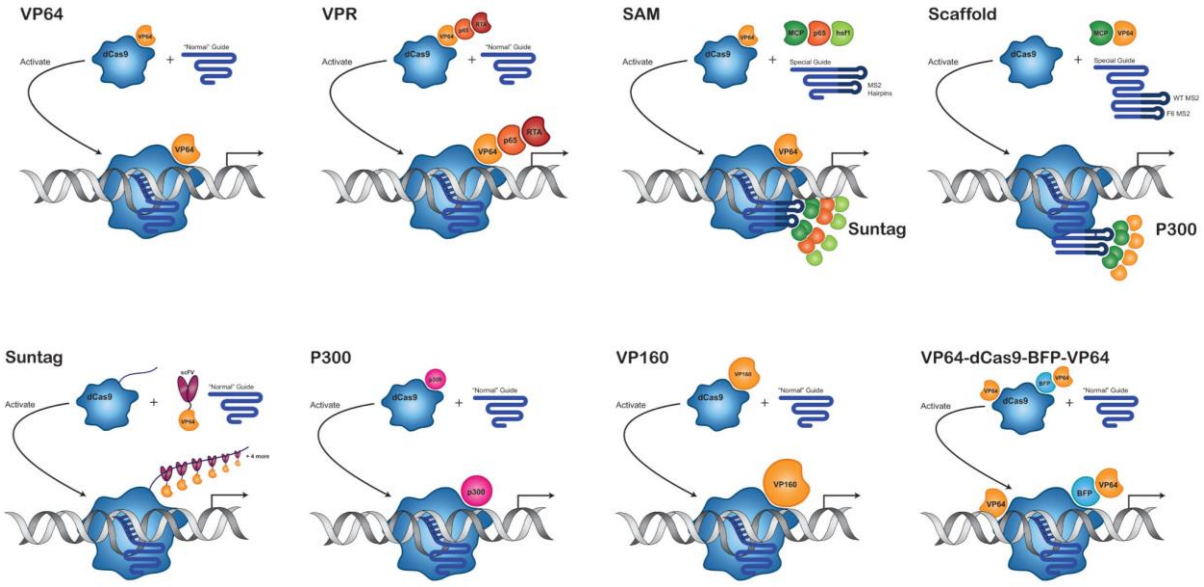


Fig. 1.4. Schematic representations of most of the aforementioned CRISPRa systems, from Chavez et al. 2016.

F. Examples of CRISPRa studies in stem cell biology and cell reprogramming

The targeted differentiation of pluripotent cells in specific lineages has been efficiently performed with CRISPRa strategies. For instance, Wei et al. (2016) obtained a high induction of Cdx2 in mESCs leading to, when cells were placed in Trophectoderm Stem Cell (TSC) culture media, a good induction of TSC-like cells (Cdx2+) obtained at a higher efficiency compared with a Cdx2 exogenous transgene transfection. The same study was performed with Gata6 and Extra Embryonic-like (XEN-like) cells with comparable results. Chavez et al. (2015) used VPR to differentiate iPS cells in neuronal lineages through the transfection of a pool of 30 gRNAs targeting NGN2 or NEUROD1 and showed that this was sufficient to trigger strong neuronal differentiation of pluripotent cells cultured in self-renewing condition.

An interesting advantage of CRISPRa compared to conventional cDNA-mediated techniques in transient overexpression of factors was reported by Black et al. (2016). Indeed, they could show that, while both kind of non-integrated transgenes were eliminated with the same kinetics over time, CRISPRa lead to a more durable overexpression due to the epigenetic remodelling of the targeted locus inducing a prolonged effect compared to exogenous expression. The authors used both ways to activate Brn2, Ascl1, and Mytl1 transcription factors in MEFs in order to induce neuronal differentiation. They showed that endogenous expression of the factors was strongly sustained after CRISPRa transgenes disappearance compared to cDNA vectors where endogenous expression rapidly decreased after transgenes elimination. Therefore, a largely more efficient reprogramming was obtained through CRISPRa vectors transient transfection in MEFs.

Finally, it was recently shown for the first time that endogenous induction of Sox2 locus alone could lead to the reprogramming of mouse fibroblasts into induced pluripotent cells. The authors (P. Liu et al. 2018) activated seven reprogramming candidate genes (Oct4, Sox2, Klf4, c-Myc, Nr5a2, Glis1, Cebpa) with the SunTag system through combinatorial and multiplexed inductions. By this way, they could efficiently obtain reprogramming of MEFs and found that endogenous activation of Sox2 alone along with a chemical cocktail was sufficient to induce MEFs reprogramming into iPSCs. When re-differentiated in MEFs, those cells could be efficiently reprogrammed again by SunTag activation. The same effect was obtained when recruiting the SunTag system to Oct4 promoter and proximal enhancer. This study underlined the fact that endogenous locus activation can result in different effects than transgenic expression methods probably due to local effects or recapitulation of “natural” pluripotency associated genome architecture therefore enhancing reprogramming efficiency.

Methods

Heurtier & Owens et al

The molecular logic of Nanog-induced self-renewal

Methods

Cell culture

Regular cultures:

ES cells (E14Tg2a and derivatives) were cultured at 37°C in 7% CO2 on 0.1% gelatine (SIGMA, G1890-100G) in DMEM+GlutaMax-I (Gibco 31966-021), 10% FCS (Sigma F7524), 100 µM 2-mercaptoethanol (Gibco 31350-010), 1× MEM non-essential amino acids (Gibco 1140-035) and 10 ng/ml recombinant LIF (MILTENYI BIOTEC, 130-099-895). Cells were passaged 1:10 every 2–3 days with 1X trypsin-EDTA 0.05% (Thermo 25300062). Only when mentioned cells were cultured in 2i medium containing 0.5X DMEM/F12 (Gibco 31331093), 0.5X Neurobasal (Gibco 21103049), 0.5% N2 supplement 100X (Gibco 17502048), 1% B27 supplement 50X (Gibco 17504044), 10µg/mL Insulin (Sigma I1882-100MG), 2 mM L-Glutamine (Invitrogen 91139), 0.05% BSA (Sigma A3311-10G), 100 µM 2-mercaptoethanol (Gibco, 31350-010), 10 ng/ml recombinant LIF (MILTENYI BIOTEC, 130-099-895), 1 µM PD0325901 (Axon Medchem Bv Axon-1408), 3 µM CHIR99021 (Axon Medchem Bv Axon-1386). When specified cells were treated with 1 µg/mL of Doxycycline (Sigma D3072-1ML), 1 µg/mL Puromycin (Sigma P9620-10ML), 400 µg/mL Hygromycin B (Sigma H3274-50MG), 0.2 mg/mL G418 (Sigma G8168-10ML), 1µM Tamoxifen (Sigma H7904-5MG).

Nanog loss- and gain-of-function:

For all SunTag induction experiments, 30.000 cells per cm^2 were plated in presence or absence of Doxycycline/LIF. Medium was changed every day after one wash with the same medium. At day 3, cells were either lysed on the plate for RNA extraction or harvested for Western blot lysates, microscopy slides and chromatin preparation. For induction kinetics over 6 days, cells were passaged at day 3 at the same density. 44iN cells (Festuccia et al., 2012) were kept in culture with Doxycycline for at least 3 passages, except when explicitly stated. Subsequently, 30.000 cells per cm^2 were plated in presence or absence of Doxycycline for the indicated

times. To culture 44iN cells in the absence of Nanog long-term, the cells were maintained under G418 selection, as previously described (Festuccia et al., 2012). Finally, 44NERT (Navarro et al., 2012) cells were cultured under G418 selection and plated at 30.000 cells per cm^2 to perform Tamoxifen induction kinetics for the indicated times.

Clonal assays:

Clonal assays were performed by plating 600 cells/P6 well in +/- LIF and +/- Dox, in parallel. Medium was changed every day after one wash with the same medium. After 6 days, colonies were fixed (25% Citrate solution, Sigma 854: 67% Acetone, Sigma 270725; 8% Formaldehyde, Sigma F8775) for 1min and stained for 20min with Alkaline Phosphatase staining kit (Sigma, 86R). Number of undifferentiated, mixt and differentiated colonies was then assessed on a stereo-microscope (NIKON-SMZ1500). For serial cloning assay, all cells from -LIF +Dox condition were harvested at day 6, counted and plated again at clonal density for 6 additional days, as indicated, and processed as above. Finally, to assess the pluripotent state of Nanog SunTag cells cultured in -LIF +Dox, all cells were harvested at day 6 and passaged 1:4 in FCS/LIF for 1 day to ensure correct cell adhesion. The next day medium was replaced by 2i/LIF.

Generation of SunTag ES cells

Cloning of the SunTag and gRNA expressing vectors:

The SunTag vectors were obtained from Addgene (#60903, #60904). The PiggyBac vectors containing the rtTA trans-activator (PB-CAG-rtTA) and a TRE-driven expression cassette (A-ND2) as well as the PBase vector were kindly provided by Dr. Pentao Liu (Gao et al., 2013). To generate the PB-TRE-SunTag-dual-Hygro vector (see Sup. Fig1) expressing the two moieties of the SunTag system, we first modified #60903 as follows. The TRE promoter was PCR-amplified and inserted upstream of dCas9. Next, we excised the lentiviral part of the vector downstream of the WPRE element and ensured its integrity by ligating a short WPRE amplicon. Both 5' and 3' LTRs were PCR-amplified from PB-CAG-rtTA and sequentially inserted on both sides of the TRE-dCas9-GCN4-BFP cassette, and a bGH polyA signal was PCR amplified and inserted downstream of the BFP sequence. Finally, an IRES-Hph (Hygromycin resistance) cassette was amplified by PCR and inserted in frame with the dCas9-GCN4-BFP cassette upstream of the bGH polyA signal, generating the PB-TRE-dCas9-GCN4-P2A-BFP-IRES-Hph-pA vector. To

modify #60904 we first PCR-amplified the scFv-GCN4-sfGFP-VP64-GB1-NLS cassette and inserted it in A-ND2 vector downstream of the TRE promoter. After a subsequent PCR amplification of the resulting cassette, it was finally inserted in the modified #60903 vector, generating our final Dox-inducible SunTag construct. To generate the gRNA expression vector, we used the #51133 plasmid (Addgene) to introduce the U6-gRNA-PGK-PuroR-pA cassette in the PiggyBac backbone of the PB-CAG-rtTA. The resulting plasmid is referred to as PB-gRNA-Puro. This vector was used to clone annealed oligos corresponding to the 20 bp of the sgRNA sequence preceded by specific overhangs (5'-CACC and 5'TTTG).

Establishment of Parental SunTag ES cells:

Subconfluent E14Tg2a cells were transfected with 5 µL of Lipofectamine 2000 (ThermoFisher), 0.8 µg of PB-CAG-rtTA, 0.8 µg of dual SunTag PiggyBac vector and 0.4 µg of the PBase vector. Upon Doxycycline treatment (1 µg/mL) the cells were selected with Hygromycin B for 10 days. Single clones were manually picked and expanded in absence of Doxycycline and Hygromycin B. After expansion and stock freezing, induction of GFP/BFP expression was analysed for 6 clones on a LSR II Fortessa (Becton-Dickinson). Data were analysed using the FlowJo software suite (Tree Star). The 2 clones showing the best percentages of GFP/BFP positive cells under Doxycycline treatment and an absence of fluorescent signals in absence of Doxycycline were kept for further experiment (C1 and C2). The karyotype of C1 and C2 cells were established using colcemid arrest (4 h; 100 ng/ml−1; Gibco, 15212-012), hypotonic shock (NaCitrate 0.017 M, KCl 0.03 M) and cold acetic acid–methanol (1:3) fixation at 4°C. Fixed cells were spread by dropping on pre-heated glass slides, mounted (Vectashield; VectorLab, H1200) and imaged using a Nikon Eclipse X microscope equipped with: 63× oil immersion objective (N.A1.4); LUMENCOR excitation diodes; Hamamatsu ORCA-Flash 4.0LT camera; NIS Elements 4.3 software. Chromosomes number was then counted manually with NIS Elements 4.3 software for a minimum of 20 randomly chosen cells per clone.

gRNA design and selection:

To design gRNAs targeting the SunTag system to the Nanog and Otx2 promoters, we first identified all the potential targets of 20 nucleotides preceding a 'NGG' protospacer adjacent motif (PAM) on both strands. Those having a GC content between 35% and 85%, and not containing a stretch of 4 or more repeated nucleotides were kept. A potential efficiency score

was calculated for each candidate guide given the sequence and using a predictive model, as described (Doench et al., 2014). To control off-target effects and ensure DNA targeting specificity, the remaining list of sgRNA candidates was mapped on the complete mm9 mouse genome using the EMBOSS Fuzznuc tool, allowing various ambiguities and complex search patterns. As CRISPR-CAS9 efficiency depends on sgRNA-target similarity pattern (Hsu et al., 2013), candidates having off-targets with only 0, 1 or 2 mismatches were excluded, while off-targets with 5 mismatches or more were ignored, as well as off-targets not followed by a PAM. Off-targets with 3 or 4 mismatches were then sorted by the lowest number of off-targets with 3 mismatches in the 5' end, then by the highest proportion of off-targets with 2 mismatches or more in the seed. The efficiency of promoter induction with candidate gRNAs was then experimentally tested by transient transfection and the best gRNA was kept for further analyses.

Generation of Nanog-, Otx2- and Nanog/Otx2-SunTag ES cells:

C1 and C2 were lipofected as described above with 1 µg of PB-gRNA-Puro and 0.5 µg of PBase-expression vector, selected with Puromycin for 4 days and plated clonally under selection. Twelve clones (six originating from C1 and C2) were selected for RT-qPCR induction of either Nanog or Otx2. One clone from C1 and another one from C2 showing good induction levels were finally used for all experiments. For Nanog-Otx2 dual SunTag cells, Nanog-SunTag C1 and C2 cells were lipofected as above with Otx2 gRNA. Ten clones were then picked for each C1 and C2, expanded, and analysed for Otx2 induction. One clone originating from each transfection showing good induction levels of both Nanog and Otx2 were used for all experiments.

Microscopy:

Bright field microscopy:

Cell culture dishes pictures were taken on a Nikon Eclipse Ti-S inverted microscope equipped with: CFI S Plan Fluor ELWD 20X objective; 89 North PhotoFluor LM-75; Hamamatsu ORCA-Flash 4.0LT camera; NIS Elements 4.3 software.

Single-molecule RNA Fluorescent In Situ Hybridisation (smFISH):

Cells were washed in 1X PBS, trypsinized, pelleted, washed again in 1X PBS and resuspended in 2mL of DMEM/FCS medium. Cells were fixed with 1% Formaldehyde (Sigma F8775) with slow agitation. Fixation reaction was stopped by addition of 300µL of 1M glycine (SIGMA G7126-500G) for 5min. Cells were then pelleted at 4°C, washed in cold 1X PBS, and pelleted again. Cells were resuspended in cold 1%BSA 1X PBS at 1 million cells/mL and cytospun at 400 rpm (Low acceleration) for 5 min on SuperFrost slides (Thermo J1800AMNT). Slides were air dried and stored in 70° EtOH at 4°C. Each spot was incubated at 37°C for 3hrs with hybridization cocktail (10% Formamide, 2X SSC buffer, 1µg/mL BSA, 1µL of E.Coli RNAs at 1µg/mL, 1µL of Nanog probe at 20 pmol/µL). The slides were washed 3 times in 2X SSC 10% Formamide for 30min at 37°C and mounted in Vectashield medium with DAPI (Vector-abcys H-1200). Nanog pre-messenger probe was designed using Stellaris Probe Designer version 4.2 on Biosearch Technologies website with the maximum masking level (5) and was synthetized by the same company. Image stacks (0.5 µm gap) were acquired using a Nikon Eclipse X microscope equipped with: 63× oil immersion objective (N.A1.4); LUMENCOR excitation diodes; Hamamatsu ORCA-Flash 4.0LT camera; NIS Elements 4.3 software. The number of active transcription sites per cell was automatically counted with FISH-quant program (Mueller et al. 2013). We used ImageJ software to measure the intensity of the signal at active transcription sites. A line of 16 pixels, with constant orientation being kept over all images, was centred on each analysed spot. The function Analyse>Measure was used to get pixel intensities along the line. The intensity of each pixel was then normalised to the mean of the first and last pixel of each individual analysed spot. Spots were randomly selected until at least 40 had been quantified for each of the Nanog SunTag clones in both -Dox and +Dox conditions.

RNA, protein and chromatin analyses:

Western Blot:

One million cells were lysed in 100 µL of Laemmli Sample Buffer (Biorad, 1610737) and 10 µL loaded in wells of 10X Mini Protean TGX gels 10 % (Biorad 456-1034). Migration was performed at 20 mA for 1h30 and transfer to nitrocellulose membranes (GE Healthcare RPN303D) was performed at 300 mA for 1h at 4°C. Ponceau (Sigma P7170-1L) staining was used to check loading homogeneity and a picture was taken with a ChemiDoc MP Imaging System (Biorad 1705062). Membrane was blocked in PBS-0.1% Tween20 (PBST) 5%BSA for 1hr

at room temperature. Membrane was incubated with primary antibodies at 4°C overnight in blocking buffer, washed 3 times for 5min in PBST at room temperature and incubated in blocking buffer with secondary antibodies for 1ht at RT. Membranes were washed three times in PBST and revealed with Clarity Max Western ECL Substrate (Biorad 1705062). Chemoluminescence was imaged on a ChemiDoc MP Imaging System (Biorad 1705062). Membrane stripping was done with mild stripping buffer (water 0.15% glycine, 0.1% SDS, 0.1% Tween 20, adjusted to pH 2.2) for three washes of 5 min. Membrane was blocked again after stripping and processed as above. The antibodies used can be found in Table S4.

RNA extraction and Reverse Transcription:

Cells were lysed with 1mL TRIzol (ThermoFisher) and RNAs extracted according to manufacturer's protocol. To eliminate any genomic DNA contamination, this was followed by an additional DNAse I treatment (Qiagen 79254) for 20min at 37°C followed by phenol:chloroform purification. RNAs were resuspended in Ultrapure DNAse/Rnase Free Distilled Water (Thermo 10977035). Reverse Transcription was performed with 1µg of total RNAs with random hexamers or specific primers for strand specific RT-qPCR (Table S4; Roche 04379012001) following manufacturer's protocol on a TM 100 Thermal Cycler (Biorad).

Quantitative real-time PCR:

Real-time PCR reactions were performed in duplicates in 384-well plates with a LightCycler 480 (Roche) using 4.5µL of LightCycler 480 SYBR Green I Master (Roche, 04707516001), 5µL of sample and 0.25µL of each primer at 20µM in a final reaction volume of 10µL. Standard and melting curves were generated to verify the amplification efficiency (>85%) and the production of single DNA species. PCR primer sequences are listed in Table S4. The 2dCt method was used both for ChIP and RT-PCR analysis. For the former, all values were corrected to the input; for the latter, Tbp was used to normalise the data.

Chromatin preparation:

Nanog SunTag cells (3.10^7) used for H3K27me3 analysis, were resuspended in 3 ml PBS and crosslinked for 10 min at room temperature with 1% formaldehyde (Sigma F8775). Crosslinking was stopped with 0.125 mM glycine for 5 min at room temperature. 44iN cells (3.10^7) used for TF binding profiling were crosslinked in 3 ml of freshly prepared PBS-DSG

2 mM at pH 7.0 (Sigma, 80424-5 mg) for 50 min at room temperature with occasional shaking. After pelleting and washing in PBS, cells were incubated for 10 min in 3 ml PBS 1% formaldehyde (Sigma F8775), quenched with 0.125 MM glycine. After fixation, cells were pelleted, washed twice with ice-cold PBS and resuspended in 6 ml of swelling buffer (25 mM Hepes pH 7.95, 10 mM KCl, 10 mM EDTA) freshly supplemented with 1X protease inhibitor cocktail (PIC-Roche, 04 693 116 001) and 0.5% NP-40. After 30 min on ice with occasional shaking, the suspension was centrifuged and resuspended in 450 µl of TSE150 (0.1% SDS, 1% Triton, 2 mM EDTA, 20 mM Tris-HCl pH8, 150 mM NaCl) buffer, freshly supplemented with 1X PIC. Samples were split in 3 (150µL) and sonicated in 1.5 mL tubes (Diagenode) using a Bioruptor Pico (Diagenode) for 7 cycles divided into 30 s ON - 30 s OFF sub-cycles at maximum power, in circulating ice-cold water. After centrifugation (10 min, full speed, 4 °C), the supernatant was stored at −80 °C. Five microlitres was used to quantify the chromatin concentration and check DNA size (typically 200–500 bp).

Chromatin immunoprecipitation (ChIP):

The chromatin was pre-cleared for 90 min on a rotating wheel at 4 °C in 300 µl of TSE150 containing 50 µl of pG Sepharose beads (Sigma, P3296-5 ML) 50% slurry, previously blocked with BSA (500 µg ml−1; Roche, 5931665103) and yeast tRNA (1 µg ml−1; Invitrogen, AM7119). Immunoprecipitations were performed overnight rotating on-wheel at 4 °C in 500 µl of TSE150. 20 µL was set apart for input DNA extraction and precipitation. 50 µL of blocked pG beads 50% slurry was added for 4 h rotating on-wheel at 4 °C. Beads were pelleted and washed for 5 min rotating on-wheel at room temperature with 1 ml of buffer in the following order: 3 × TSE150, 1 × TSE500 (as TSE150 but 500 mM NaCl), 1× washing buffer (10 mM Tris-HCl pH8, 0.25M LiCl, 0.5% NP-40, 0.5% Na-deoxycholate, 1 mM EDTA), and 2 × TE (10 mM Tris-HCl pH8, 1 mM EDTA). Elution was performed in 100 µl of elution buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8) for 15 min at 65 °C after vigorous vortexing. Eluates were collected after centrifugation and beads rinsed in 150 µl of TE-SDS1%. After centrifugation, the supernatant was pooled with the corresponding first eluate. For both immunoprecipitated and input chromatin, the crosslinking was reversed overnight at 65 °C, followed by proteinase K treatment, phenol/chloroform extraction and ethanol precipitation. The antibodies and the amount of chromatin used for PCR or sequencing analyses is indicated in Table S4.

Chromatin accessibility (ATAC):

50,000 viable cells were washed with cold 1X PBS, pelleted by centrifugation for 5 min at 500g at 4°C, resuspended in 50 µl of transposition reaction mix (25 µl of Tagmentation DNA buffer, 2.5 µl Tagment DNA enzyme (Illumina) and 22.5 µl nuclease-free H2O) and incubated for 30 min at 37 °C. DNA was purified with a MinElute PCR purification kit (Qiagen),

Libraries preparation and sequencing

RNA-seq:

We used 0.5 µg of total RNA to purify polyadenylated mRNAs and to build an RNA library, using TruSeq Stranded mRNA Sample Prep Kit (Illumina, #RS-122-9004DOC), as recommended by the manufacturer. Directional libraries were checked for concentration and quality on DNA chips with the Bioanalyser (Agilent). More precise and accurate quantification were performed with sensitive fluorescent-based quantitation assays ("Quant-It" assays kit and QuBit fluorometer, Invitrogen). Samples were normalized to 2nM and multiplexed. After denaturation using 0.1N NaOH (5' at room temperature), the samples were diluted to 9pM and loaded on the flowcell. Sequencing was performed on the HiSeq 2500 sequencer (Illumina) in 65 bases V4 single-end mode.

ChIP-seq:

Chiped DNA was end repair in a total volume of 50µL (sample 37.5 µL, 10mM dNTPs 2 µL, NEB T4 ligase buffer 5 µL, NEB T4 polymerase 2.5 µL, NEB Klenow polymerase 0.5 µL, NEB T4 PNK 2.5 µL) and incubated for 30min at 20°C. After DNA purification (see below), A-tailing was subsequently performed in a total volume of 25 µL (sample 20 µL, NEB Buffer 2 2.5 µL, 5mM dATP 1 µL, NEB Klenow 3'-5' exo minus 1.5 µL) at 37°C for 30min. Illumina TruSeq adapters were used for libraries indexing, ordered from IDT Company with 5' phosphate modification. Illumina adapters' compatibility was checked with the online tool checkmyindex (checkmyindex.pasteur.fr) for multiplexing. Truseq adapters were annealed with Illumina Universal adapter at 20µM each in 1X NEB Buffer 2 on a TM 100 Thermal Cycler (Biorad). Adapters ligation was performed in a total volume of 25 µL (sample 19.25 µL, NEB 10x T4 ligase buffer 2.5 µL, 0.2µM adapter 1.25 µL, NEB T4 ligase 2 µL) at 16°C overnight. After DNA purification, DNA was amplified in a total volume of 50 µL (sample 19.5 µL, Pico green 1 µL –

1:10 in water; Life Technologies P7589 –, 25 µL of KAPA HiFi HOTSTART Ready mix – NC0295239 –, PCR 1.0 10µM 1 µL, PCR 2.0, 10µM 1 µL) on a LightCycler 480 (Roche). PCR primers are listed in Table S4. Any sample reaching an absolute fluorescence value of 6 was taken out from the plate at the end of the last extension. Any library requiring more than 16 cycles of amplification to reach this level was discarded and reprepared. DNA was finally purified and the concentration was measured with a Qubit 4 Fluorometer (Thermo Q33226). Libraries quality check and size estimation were then performed on an Agilent 2200 Tape Station with High Sensitivity D5000 ScreenTape (Agilent Technologies, 5067-5592) and High Sensitivity D1000 Reagents (Agilent Technologies, 5067-5585) using 1-2ng of material. Libraries were subsequently adjusted to equimolar concentration of 2nM according to fragments average size and concentration prior mixing them for subsequent sequencing on the HiSeq 2500 sequencer (Illumina) in 65 bases V4 single-end mode.

ATAC-seq:

Libraries were prepared as described (Buenrostro et al., 2015) but replacing NEB Next High-Fidelity 2X PCR Master Mix, by KAPA HiFi HotStart (KapaBiosystems KM2602) for PCR amplification, after determining the number of cycles needed by qPCR. The concentration and quality of the libraries were assessed as described above. Libraries were sequenced on the HiSeq 2500 sequencer (Illumina) in 65 bases V4 paired-end mode.

SPRI beads preparation and DNA purification:

SPRI beads for DNA purification were prepared as follows: 50% (w/v) PEG 8000 stock was prepared by progressively pouring nuclease-free water on 12.5g of PEG 8000 (Promega V3011) until reaching a total volume of 25mL. 1mL of vigorously resuspended Sera-Mag Magnetic Speed beads (Ge Healthcare 65152105050250) were washed three times with washing buffer (10 mM Tris HCl pH 8, 1 mM EDTA, 0.05% Tween 20) by vortexing (15sec) and supernatant removal with an PureProteom Protein G Magnetic Bead System 11740343 (EMD Millipore). Then, incomplete storage buffer was prepared (25mL NaCl 5 M, 6.25mL Nuclease-free water, 0.5mL Tris base 1 M, 0.5mL Disodium EDTA 0.1 M) and 1mL was used to resuspend the beads after last washing step by vortexing 15sec. Resuspended beads were added to the rest of the incomplete storage buffer and the mix was vortexed for 30sec. 17.5mL of 50% (w/v) PEG 8000 stock was then slowly added to the mix. Finally, 250µL of Tween 20 were added and

the mix was slowly inverted until solution homogenised. Beads were further aliquoted in 2mL Eppendorf tubes and kept at 4°C. Optimal Beads:sample ratios for precise DNA size recovery were finally assessed using a 50bp DNA step ladder (Sigma S7025-50UG) and an Agilent 2200 Tape Station using High Sensitivity D5000 ScreenTape (Agilent Technologies, 5067-5592) and High Sensitivity D1000 Reagents (Agilent Technologies, 5067-5585). DNA purification steps were performed in round bottom 96 well plates (Thermo 611U96) using an Agencourt SPRI Plate Super Magnet Plate (Beckman Coulter A32782) at room temperature. DNA sample and beads solution were mixed (0.8 to 2:1 ratio) in the P96 well and left for 5min and washed twice with 200µL 70% EtOH. DNase-free water was added to the beads for DNA elution, with successive rounds of pipetting. After 5min, the samples were repositioned on the magnetic plate for 5min and the supernatant was collected.

Informatic analyses

Data and availability:

A summary of data collected and sequenced for RNA-seq, ChIP-seq along with public data used to identify Nanog binding regions is available in Table S5. Briefly, for *Nanog* induction RNA-seq between 2-4 replicates were sequenced per condition; for 44iN ChIP-sequencing we collected two replicates per factor for plus and minus Dox, and correspondingly 2 replicates per +/- Dox for ATAC-seq. For H3K27me3 ChIP-seq we collected and sequenced 4 replicates per Lif/Dox condition, some of which were excluded due to quality issues (see Identification of H3K27me3 Domains). To identify Nanog binding regions we combined Nanog ChIP-seq from four independent studies: GSE56312, (Galonska et al., 2015); GSE11724, (Marson et al., 2008); GSE44288, (Whyte et al., 2013); GSE55404, (Lee et al., 2015). To compare to previously published Nanog targets, microarray data was taken from: Nishiyama et al., 2009 and Festuccia et al., 2012.

ChIP-seq Data Processing:

For all ChIP-seq samples (Nanog public datasets, 44iN +/-Dox, H3K27me3) reads were aligned with bowtie2 (Langmead and Salzberg, 2012) in the mm9 genome, with options “-k 10” for all samples and additionally “-I 0 -X 1000 --no-discordant --no-mixed” for paired-end samples from Lee et al. Reads were additionally filtered for those with a single alignment (mapping

quality: 255) and an edit distance less than 4 (mean edit distance for paired reads). For Nanog and 44iN datasets distinct reads aligning with identical coordinates were treated as duplicates and collapsed to one.

RNA-seq Data Processing:

Stranded RNA-seq reads were aligned to the mm9 genome using STAR (Dobin and Gingeras, 2015) and quantified by RSEM (Li and Dewey, 2011) using the RSEM-STAR pipeline, with additional options "--seed 1618 --calc-pme --calc-ci --estimate-rspd --forward-prob 0.0".

ATAC-seq Data Processing:

Paired end 65bp ATAC-seq reads were trimmed by aligning read pairs to discover regions of reverse complementarity surrounded by Nextera sequencing adapters. Similarly to ChIP-seq processing, reads were aligned to mm9 genome using bowtie2 (Langmead and Salzberg, 2012) with options "-k 10 -I 0 -X 1000 --no-discordant --no-mixed", and filtered for reads with a single alignment mean edit distance less than 4 between read pairs. Heatmaps and meta plots were generated by marking ATAC-seq cut sites, left-most and right-most coordinates of each read with shifted inwards by 4bp, as recommended (Buenrostro et al., 2015), and base pair of the cut site and the two surrounding were marked. Total cut-site signal was normalised for sequencing depth and averaged over replicates.

Identifying Nanog binding regions from Public datasets

Six independent samples from four studies were combined, as specified above. Peaks were called against relevant inputs/controls for all samples using macs2 (Feng et al., 2012) with "callpeak -q 0.2 -g mm", with the exception of (Lee et al., 2015) in which controls were unavailable, in this case macs2 was run without controls and peaks are called against a local background model. Peaks intersecting with the mm9 blacklist (ENCODE Project Consortium 2012) were excluded along with those on chrM and chrY. This resulted in 69,088 & 25,047 peaks for (Galonska et al., 2015), 17,950 & 10,347 peaks for (Marson et al., 2008), 31,062 peaks for (Whyte et al., 2013), and 27,888 peaks for (Lee et al., 2015). Combined, this resulted in 85,697 candidate Nanog binding regions, which were further filtered to those occurring in at least two independent samples resulting in 39,164 peaks. Reflecting that these regions are representative of Nanog binding, we found the fraction of reads in these peaks to be high over

all samples: 17.1%, 23.9% (Galonska et al. 2015); 16.6% & 16.6% (Marson et al., 2008); 12.8% (Whyte et al., 2013); 25.8% (Lee et al., 2015). Finally, to focus on those regions with clear Nanog binding we additionally filtered peaks to those with a minimum height (averaged over all samples) of 1 read per million resulting in 27,782 peaks.

Clustering of 44iN ChIP-seq

ChIP-seq signal for Esrrb, Oct4, Sox2 and Brg1 was quantified over 1kb (+/-500bp) centred on Nanog peaks. Our objective was to identify the set of regions with strong co-binding between at least one of the factors and Nanog. Employing preliminary clustering on the sequencing depth normalised signals from the four factors, we found regions segregated into those displaying nanog co-binding and nanog solo behaviour. We found that a peak height threshold of ~30 reads normalised to the mean sequencing depth (~2.7 reads per million) for at least one factor in one condition captured the co-binding versus solo distinction. Applying this threshold resulted in 13,515 nanog solo regions, and 14,259 nanog co-binding regions which were subject to further comprehensive clustering. To identify patterns of co-binding that do not depend of differences between occupancy between factors or globally between sites (i.e. to group together Nanog sites which have the same binding pattern at potentially different occupancy levels), we normalise each factor to the same mean occupancy, and then normalise by the maximum peak height over all factors at each region. We apply k-means clustering on the combined normalised signal from +/- 250 bp surrounding the summit of each nanog peak for each factor. Formally, if h^+_{ijk} is the normalised read depth in +Dox for factor i at region j at position k, and h^-_{ijk} correspondingly for -Dox, then the trace over the entire region of length is $t^+_{ij} = (h^+_{ij1}, \ldots, h^+_{ijm})$ and similarly t^-_{ij} for -Dox. The combined trace is denoted $t_i j = (t^+_{ij}, t^-_{ij})$

We calculate a mean +Dox occupancy score for each factor: $\sigma_i = \sum_{j,k} h^+_{ijk}$

and then denote the normalised trace over n factors $\tau_j = (t_{1j}/\sigma_1, \ldots, t_{nj}/\sigma_n)$, finally denoting its max-normalisation as $\bar{\tau} = \tau_j / \max_k \{\tau_{jk}\}$.

We then apply a k-means clustering with a Euclidean distance metric on the $\bar{\tau_j}$. K-means clustering was applied using the Clustering package in Julia. We selected k to provide a clear

group of regions were Nanog is required for other factors to bind or not. We found k = 8 to provide a good balance between summarising broad Nanog dependence and identifying complex co-binding relationships: when evaluating the cluster assignments for k and (k+1), in the range 2≤k≤20, we found that the Rand Index (Rand, 1971), defined as the number of pairs of regions assigned to the equivalent cluster over the total numbers of pairs exceeded 0.9 for k≥7. Importantly, only at k=8 the Oct4 only dependent cluster which retains accessibility (Fig 3) was resolved, and for k = 9 and k = 10 the overall cluster identity in terms of dominant factors and dependent vs independent assignments no longer changed.

Identification of H3K27me3 Domains

Candidate H3K27me3 domains were identified by MACS2 (Feng et al., 2012), run in broad peak mode with options "--broad -q 0.2 -g mm" against relevant inputs. A high level on concordance in peaks between the samples was observed. We merged all peak regions within 3kb, taking those present in at least 8 samples (i.e. all four replicates of at least two conditions), to focus on broad domains. This resulted in 6240 H3K27me3 domains. We noticed that certain replicates were outlying, and excluded those samples that had the maximum reads per peak over all replicates of a condition in at least 60% of the peaks. This excluded a sample from each of SunTag +LIF-Dox, SunTag, +LIF+Dox and SunTag -LIF+Dox. We considered a gene to be embedded within a H3K27me3 domain, if its loci intersected with the domain by at least 1bp or if a domain lay within 4kb of the TSS and that domain did not intersect with another gene.

Clustering of H3K27me3

To identity a broad set of H3K27 domains with a dynamic response to removal of LIF and Nanog induction, we clustered the total number of reads per peak max normalised over all SunTag conditions by k-means clustering with k = 3.

RNA-seq Differential Expression Analysis in the presence of LIF

RSEM estimated read counts per sample were rounded for use with DESeq2 (Love et al., 2014). Genes with at least 10 normalised counts in all replicates of at least one condition were considered for differential expression analysis. For all differential expression tests DESeq2 was run without independent filtering and without any fold change shrinkage, genes with FDR < 0.05 are considered differentially expressed. For +LIF samples, Nanog responsive genes in

SunTag were identified by a Wald Test with the formula ~Dox on SunTag samples, and correspondingly the same was applied to 44iN samples. Since the overlap between the two systems is good (contingency table accounting for up, down, and non-significantly regulated χ^2=2126.97, df=4, p≈0.0 and a large statistical effect Cramér's V = 0.267), and non-significant genes in either SunTag or 44iN had the correct fold change when the gene was significantly misregulated in the other setup (Fig S2), we tested to find those genes consistently mis-regulated by induction of Nanog in SunTag or 44iN by a Likelihood ratio test. More specifically, we tested the alternative hypothesis ~Cell + Dox + Cell:Dox over the null model ~Cell, where Cell is a factor indicating SunTag or 44iN and Dox indicating Dox treatment. The likelihood ratio test identified 419 Nanog responsive genes (FDR < 0.05), including 152/164 genes identified by SunTag alone, and 115/141 genes identified by 44iN alone. To select a list of Nanog responsive genes in +Lif we required that a gene with differentially expression with FDR < 0.05 in any of SunTag alone, 44iN alone or the likelihood ratio test, resulting in 457 genes.

RNA-seq Differential Expression Analysis in the absence of LIF

For SunTag Nanog RNA-seq in +/- LIF and +/- Dox we opted for Wald tests on contrasts on the formula ~LIF + Dox + LIF:Dox. This allowed us to identify genes that responded to LIF or Dox in independent manner along with genes whose response to Dox was dependent on LIF. We tested three variables: LIF, Dox and the sum of Dox and LIF:Dox interaction term, resulting in a fold change due to the loss of LIF, a fold change due to the addition of Dox in +LIF and a fold change due to the addition of Dox in -LIF. Classifying genes as either activated, repressed or not significant for each variable results in the assignment of genes to one of 24 different patterns of response. We tested 15,301 genes and found 7999 genes (52.2%) had no significant change in expression in either +/-LIF or +/- Dox; 2790 (18.2%) and 2648 (17.3%) genes were activated and repressed upon loss of LIF with no Dox response; 684 (4.7%) genes were repressed on loss of LIF and activated by Nanog in -LIF; 528 (3.4%). The remaining ~5% of genes either had a Nanog response in +LIF only, or a Nanog response in -LIF in which the gene did not respond to the removal of LIF. To assess the potential Otx2-driven compensation of Nanog effects, we applied an identical analysis to Nanog alone, noting those cases in which a gene previously assigned as a Nanog target in -LIF, is now either not significant or is

significant in the opposing direction. To detect those genes where the Otx2 only partially compensates for Nanog rescue (i.e. expression is not returned to -LIF -Dox levels), we combined Nanog and Nanog/Otx2 SunTag samples and tested ~LIF + DoxGuide + LIF:DoxGuide, where DoxGuide is a factor representing +/-Dox and either Nanog alone or Nanog/Otx2 guides. We tested for the difference between the two guides in +Dox -LIF.

Gene Peak Proximity Enrichments

To determine whether a set of Nanog responsive genes were enriched in proximity to a set of Nanog peaks. We calculated the distance between the TSS of each gene that had been tested for differential expression and the set of Nanog peaks. We then performed fisher exact tests between the association of genes in the responsive set to all expressed genes within *x*bp of a peak, for *x* in the range [1, 1e+8] bp.

References for Methods

Buenrostro, J., Wu, B., Chang, H., and Greenleaf, W. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol *109*, 21.29.1-21.29.9.

Dobin, A., and Gingeras, T.R. (2015). Mapping RNA-seq Reads with STAR. Curr Protoc Bioinformatics *51*, 11.14.1-19.

Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B.L., Xavier, R.J., and Root, D.E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. Nat. Biotechnol. *32*, 1262–1267.

Feng, J., Liu, T., Qin, B., Zhang, Y., and Liu, X.S. (2012). Identifying ChIP-seq enrichment using MACS. Nat Protoc *7*, 1728–1740.

Festuccia, N., Osorno, R., Halbritter, F., Karwacki-Neisius, V., Navarro, P., Colby, D., Wong, F., Yates, A., Tomlinson, S.R., and Chambers, I. (2012). Esrrb Is a Direct Nanog Target Gene that Can Substitute for Nanog Function in Pluripotent Cells. Cell Stem Cell *11*, 477–490.

Galonska, C., Ziller, M.J., Karnik, R., and Meissner, A. (2015). Ground State Conditions Induce Rapid Reorganization of Core Pluripotency Factor Binding before Global Epigenetic Reprogramming. Cell Stem Cell *17*, 462–470.

Gao, X., Yang, J., Tsang, J.C.H., Ooi, J., Wu, D., and Liu, P. (2013). Reprogramming to Pluripotency Using Designer TALE Transcription Factors Targeting Enhancers. Stem Cell Reports *1*, 183–197.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nature Biotechnology 31, 827–832.Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lee, B.-K., Shen, W., Lee, J., Rhee, C., Chung, H., Kim, K.-Y., Park, I.-H., and Kim, J. (2015). Tgif1 Counterbalances the Activity of Core Pluripotency Factors in Mouse Embryonic Stem Cells. Cell Reports *13*, 52–60.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology *15*, 550.

Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., et al. (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521–533.

Navarro, P., Festuccia, N., Colby, D., Gagliardi, A., Mullin, N.P., Zhang, W., Karwacki-Neisius, V., Osorno, R., Kelly, D., Robertson, M., et al. (2012). OCT4/SOX2-independent Nanog autorepression modulates heterogeneous Nanog gene expression in mouse ES cells. The EMBO Journal *31*, 4547–4562.

Nishiyama, A., Xin, L., Sharov, A.A., Thomas, M., Mowrer, G., Meyers, E., Piao, Y., Mehta, S., Yee, S., Nakatake, Y., et al. (2009). Uncovering Early Response of Gene Regulatory Networks in ESCs by Systematic Induction of Transcription Factors. Cell Stem Cell *5*, 420–433.

Rand, W.M. (1971). Objective Criteria for the Evaluation of Clustering Methods. Journal of the American Statistical Association *66*, 846–850.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. Cell *153*, 307–319.

As most of the methods used in the thesis are described in the manuscript methods only the one which are not included or not sufficiently detailed within it will be included in that section. All experiments have been done in the same conditions than in the manuscript if not precised.

A. sgRNA cloning

Two oligonucleotides corresponding to the 20 bp of the sgRNA sequence preceded by the following overhangs were synthesized:
5' – CACCNNNNNNNNNNNNNNNNNN – 3'
3' – NNNNNNNNNNNNNNNNNNNNCAAA – 5'

1ug of Pb-gRNA-Puro plasmid was digested with BbsI for 1hr at 37°C (1 µg Plasmid, 1 µl FastDigest BbsI, 2 µl 10X FastDigest Buffer, X µl H2O for 20 µl total). In the meanwhile the pair of oligos were annealed (5 µl of each oligo at 100 µM) by heating up at 95°C for 5 min and cooling at RT on the bench for 45 min. 3 µL of cooled oligo mix was diluted in 750 µL of water. Ligation was performed for 30 min at RT (2 µL of BbsI-digested plasmid with no need for purification, 2 µl of diluted oligo mix, 2 µl 10X T4 ligase Buffer (NEB), 13 µl H2O, 1 µl T4 ligase (NEB)). To avoid high background, ligation reaction was followed by additional BbsI restriction for 10 min at 37°C (add 3µL 10X FastDigest Buffer, 6µL H2O, 1µL BbsI to ligation mix). 2 µL of the mix was then transformed in competent bacteria. The next day, 2 colonies were picked and after miniprep and plasmid purification submitted to Sanger sequencing with the following primer : ACTATCATATGCTTACCGTAAC.

B. Bio-informatics analysis with Seqmonk program (LASER selection)

RNAseq results validation with Festuccia and Marks datasets: For Marks et al. 2012, tables were directly taken from the publication. From Festuccia et al. 2012: from the microarrays results table provided in the article, all genes with absolute value of log2FC(12h/0h) greater than 0.2 were selected as Nanog induced/repressed according to expression change. Genes' lists were further quantified through RPKM quantification on our four RNAseq samples.

LASER selection with Seqmonk program: Reads were aligned with bowtie 2 in the platform for Bioinformatics at the Institut Pasteur facility (Caroline Proux).

A running window of 400 bp with a step of 150 bp was performed to bin the genome. Reads were counted on both strands separately. RPM quantification was done on every bins. Bins with log2(RPM)<-5 in all samples were discarded from the analyses. A percentile normalization was applied on the retained bins values. Bins were filtered for log2(2i/FCS)>2, then sequentially log(dox/iN)>2, and log2(fcs/iN) > 2. Bins overlapping CDS on the same strand were further discarded. Only probes with at least Log2(RPM)>-4.5 in two samples were finally kept.

Contig probe generator: by default parameters. A percentile normalization was performed on the obtained contigs. Only contigs with at least one value of Log2(RPM)>-4.5 were kept. Selection by expression level: log2(2i/FCS)>1, log(dox/iN)>1, and log2(fcs/iN)>1. Contigs overlapping CDS on the same strand were discarded.

C. Cell fractionation

ES cells were lysed in hypotonic lysis buffer (10 mM Hepes pH 7.9, 10 mM KCl, 1.5 mM MgCl2, 0.1% Triton X-100, 0.34 M sucrose, 10% glycerol, 1 mM DTT, 1× protease inhibitor cocktail (Roche) and 300 U/ml RNasIN (Promega)) for 6 min on ice; ten million cells in 200µl of lysis buffer. The nuclei were isolated from the cytoplasmic fraction by centrifugation at 1300 g for 5 min. The supernatant was then re-centrifuged 5 min at 20 000 g, and the purified cytoplasmic fraction was taken apart to a new tube where three volumes of TRIzol® were added in order to extract cytoplasmic RNA. The nuclei pellets were washed in lysis buffer and RNA was extracted by TRIzol® addition following manufacturer's protocol. (Adapted from X. Q. D. Wang and Dostie 2017)

D. Single cell sorting

ES cells transfected with the PiggyBac vectors (see Results section VII. B) and treated with Dox for 48 hours were trypsinized and resuspended in FCS/LIF medium, filtered through a 40µm cell strainer and kept on ice. The highest GFP-positive ES cells were sorted as single cell per well in a gelatinized 96-well plate (containing FCS/LIF medium) using a FacsARIA III cell sorter (Becton-Dickinson), while keeping samples on ice.

E. Generation of LASER 23 KO ES cells

1 million E14Tg2a WT cells were plated in a 6-well plate at D0. Meanwhile, 1 µg of a plasmid containing the 5'-deletion guide and a Cas9-mCherry cassette was pre-mixed with 3 µg of a plasmid containing the 3'-deletion guide and a Puromycin cassette in 250 µL of DMEM without serum. 5 µL of Lipofectamin 2000 (Invitrogen, 11668-019) was added to 250 µL of DMEM Medium without serum in a separated tube. After 5 min, both tubes were mixed (final volume 500µL) and left for 30 min at room temperature to allow complexes formation. Finally, the mix was added to the medium. Transient Puromycin (1 µg/mL) selection was performed from D1 to D4 to select. At D4 cells were plated at clonal density, and clones were picked 10 days later. Genomic DNA was isolated with NucleoSpin Tissue DNA extraction Kit (Macherey-Nagel, 740952.50), and screened by qPCR with genomic primers matching inside and outside the deleted region. Finally, deletion was confirmed by PCR and migration on agarose gel. The resulting DNA amplicon originating form correctly deleted clones was sent for sequencing for validation.

Results

VII. Adaptation of the CRISPRa SunTag system in mouse ES cells

A. Why developing CRISPR-activation?

Gain and loss of function experiments have been extensively carried out to assess the functional relevance of a given gene. Loss of function experiment proceeds through the deprivation or the alteration of a gene product and assess its molecular or phenotypical consequences. If any resulting trait is observed, the interpretation of the result might be entangled by indirect processes but the causality between the alteration of the gene product and its downstream effect is usually straightforward. In the case of gain of function assay, where a hyperactivec form of a protein is synthetized or where a gene product is overexpressed, the interpretation of the result is way more complicated. Indeed, it is impossible to predict whether the resulting increased activity will only enhance the natural activity of the gene or also provide it novel functions (partners, targets…). However, pluripotent stem cell biology greatly benefited from such kind of approaches allowing for artificially enforced cellular fate. Notably, but not exhaustively, the characterization of the famous reprogramming factors and the generation of iPS cells (Takahashi and Yamanaka 2006), the discovery of the regulatory role of Nanog in mouse ES cells self-renewal (Chambers et al. 2003) as well as the targeted differentiation of pluripotent stem cells towards multiple lineages (Fujikura et al. 2002; Hitoshi Niwa et al. 2005; Theodorou et al. 2009; Magnúsdóttir et al. 2013; Yamamizu et al. 2013, 2016).

Nevertheless, while near transcriptome-wide loss of function screenings have been efficiently performed in mammalian cells for years, mostly thanks to small interfering RNA libraries as well as random mutagenesis or genomic insertion (Paddison et al. 2004; Carpenter and Sabatini 2004; J. M. Silva et al. 2005), large scale gain of function studies were strongly hindered by technological reasons: the need of massive full length cDNAs production and cloning despite their large sequence size. This led either to blind and low-efficiency - but successful - methods (Harrington et al. 2001; Chambers et al. 2003) or moderate scale, gene candidate-based, time-consuming approaches (Takahashi and Yamanaka 2006; Yamamizu et al. 2016).

The development of CRISPRa, guided by a very short RNA molecule, opened a new route to flexible transcriptome-wide gain of function screenings mirroring the efficiency of

knock-down strategies with high specificity (Gilbert et al. 2014) but also greatly simplified lower scale studies (P. Liu et al. 2018).

Furthermore, CRISPRa directly enhances the transcription of a gene at its natural locus, which thus allows for recapitulating the role of a locally acting RNA molecule. Indeed, some lncRNAs have been shown to act in the vicinity of their transcription site or on the same chromosome with poor locus specificity (Lee and Jaenisch 1997; Y. A. Tang et al. 2010). Remarkably, the RNA molecule was even shown to be dispensable for mediating some lncRNAs function for which rather transcription or splicing itself was shown to be important (Engreitz et al. 2016). Consequently, their overexpression from a cDNA containing vector randomly inserted in the genome would precisely not reproduce their endogenous role, strongly precluding the use of such a method for functional screening and characterization of this kind of molecule (K. C. Wang et al. 2011; Maamar et al. 2013). Moreover, proceeding through induction of endogenous loci, CRISPRa requires minimal knowledge about the structure of the targeted gene apart from a loose idea about its transcription initiation site. Thus, this method is fairly appropriate for functional investigation of poorly characterized transcription units as many lncRNAs are to date.

CRISPRa thus appears as a real opportunity for the stem cell biology field which tremendously benefited from overexpression studies and represents a great perspective for the exploration of the poorly analysed lncRNAs family. In our case, Nanog endogenous activation was used to refine the understanding of its molecular mechanism orchestrating mouse ES cells self-renewal. Then, in a second approach, a small number of lncRNAs potentially involved in mouse pluripotent cells biology were functionally interrogated using CRISPR-based overexpression. Finally, an unpredictable effect of CRISPRa brought us through embryo early cleavages field.

B. Construction of the first SunTag cell line generation

As previously mentioned, the SunTag system is one of the most powerful CRISPR activation system reported in the literature to efficiently enhance transcription in eukaryotic cells (Tanenbaum et al. 2014). We thus decided to introduce it in WT mouse ES cells. The transgenes coding for the two components of the SunTag complex were publicly available as lentiviral vectors which can therefore easily be transduced in mammalian cells. However, lentiviral vectors are known to be progressively silenced in pluripotent cells (Yao et al. 2004).

This effect is particularly famous during the process of iPSCs reprogramming (Takahashi and Yamanaka 2006; Hotta and Ellis 2008), where retroviral exogenous vectors are progressively shut off until the end of the conversion and is considered as a hallmark of fully reprogrammed pluripotent cells (Stadtfeld, Maherali, and Hochedlinger 2008). We consequently looked for another integration method. It appeared that the PiggyBac transposon system was an excellent candidate for this aim.

This natural transposon was discovered in the "cabbage looper" butterfly and has been shown to allow for transposition of small as well as very large cargos with a high efficiency in mouse and human cells (Ding et al. 2005; M. A. Li et al. 2011; Rostovskaya et al. 2012; R. Li et al. 2013; X. Li et al. 2013; Chantzoura et al. 2015). PiggyBac vectors have been used in many studies on mammalian pluripotent stem cells for random gene activation or trapping screenings, reprogramming assays or cell line construction with great success (W. Wang et al. 2008; G. Guo et al. 2009; G. Guo and Smith 2010; X. Wu et al. 2014b). Importantly, no active silencing of this type of transposon in mammalian cells has been reported so far. Thus, this natural hijacked system appears to be proficient for stable, random insertion of DNA sequences in mouse cell genome. Practically, such insertion is achieved by co-transfection of a plasmid containing a sequence of interest (cargo) flanked by PiggyBac 5' and 3' LTRs along with a PiggyBac transposase coding vector. The transiently expressed transposase specifically recognizes the PiggyBac LTRs and, through a "cut and paste" mechanism, proceed to the integration of the cargo together with its flanking LTRs within the genome of the transfected cell. It was shown that PiggyBac transposition events are biased through open and AT rich genomic regions and that PiggyBac transposase generally integrate the transposon at the centre of an AATT motif **(Fig. 2.1)** (Malcolm J. Fraser et al. 1995; M. J. Fraser et al. 1996)**.** From a technical point of view, very little amount of DNA is necessary to be transfected for genomic integration compared to more "passive" methods like electroporation or nucleofection and prior linearization of the vector is not required.

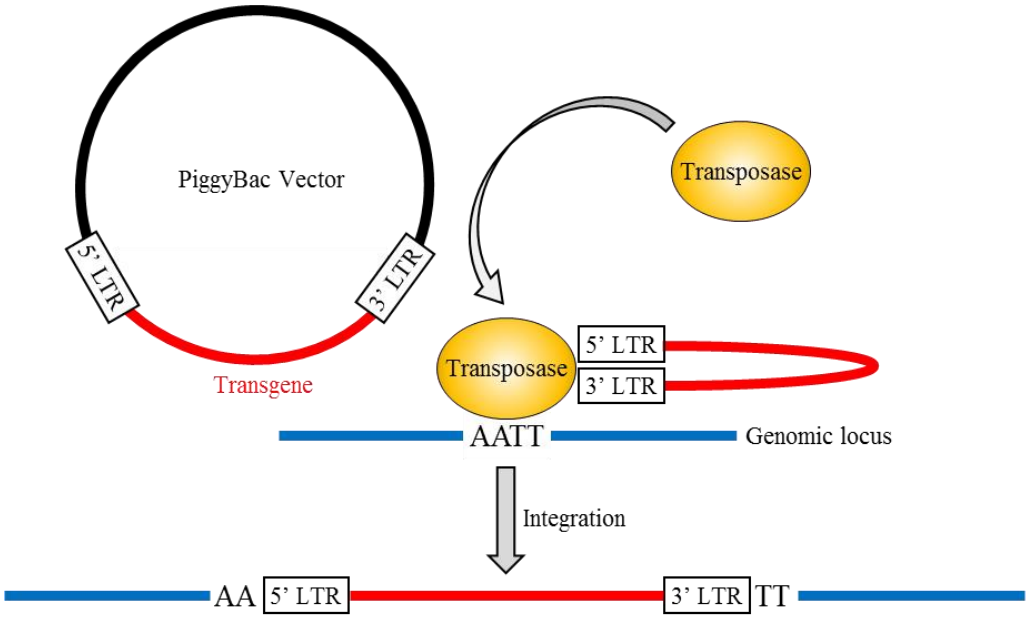


Fig. 2.1. Schematic representation of the PiggyBac transposition from a PiggyBac vector mediated by its specific transposase within an AATT genomic motif.

We therefore decided to insert the SunTag transgenes in PiggyBac vectors. However, in order to generate a tightly controllable form of this transcriptional induction system we introduced Tet-ON promoters upstream of the two SunTag modules, making them inducible under Doxycycline (Dox) treatment (Gossen and Bujard 1992). This was progressively achieved and led, together with an unanticipated issue (see below), to the construction of two generations of SunTag mouse ES cells lines.

The SunTag system is composed of two fusion proteins: - 1. the inactivated form of Cas9 enzyme (dCas9) (Qi et al. 2013) directly fused to ten repeats of a short yeast epitope (from the GCN4 protein) - 2. a single-chain variable fragment (scFv), corresponding to the fusion of the variable regions of an immunoglobulin (Huston et al. 1988), specifically recognizing the GCN4 yeast epitope (Hanes et al. 1998). This diffusible antigen binding protein carries four copies of the transactivation domain of the viral protein VP16 (VP64) (Chris I. Ace et al. 1988; Cousens et al. 1989) **(Fig. 2.2)**. Recruited to a specific promoter through any gRNA sequence, this sophisticated complex has been shown to potently activate expression of the downstream transcriptional unit (Tanenbaum et al. 2014).

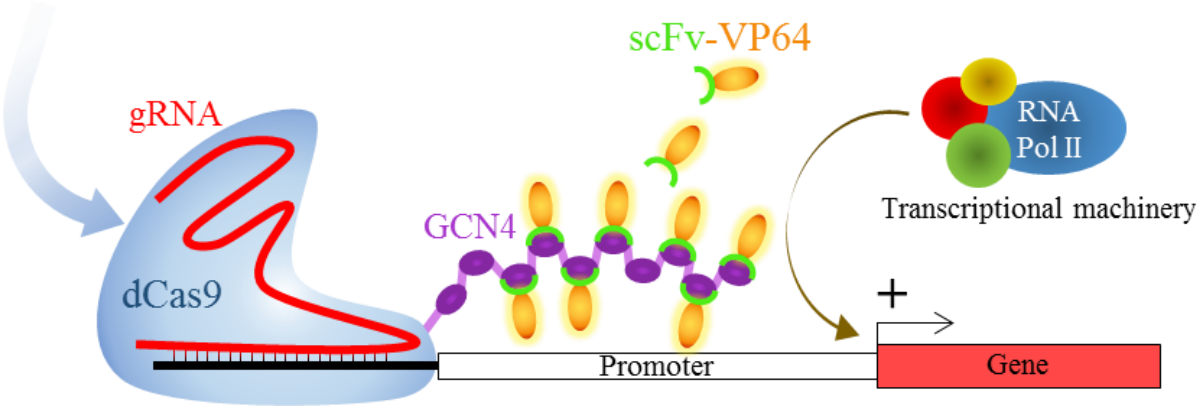


Fig. 2.2. Schematic representation of the SunTag system for endogenous transcriptional activation. The dCas9 protein in complex with the gRNA scans the genome looking for genomic sequence match. The scfv-VP64 chimeric fusion recognizes and binds the GCN4 epitopes presented by dCas9 and activates transcription around the target binding site through recruitment of the transcriptional machinery and chromatin modifying enzymes.

Our first attempt to create a stable SunTag cell line was done by co-transfecting two PiggyBac vectors harbouring the Dox-dependent rtTA transactivator and the scFv-GCN4-GFP-VP64 fusion downstream of an rtTA-responsive promoter (TRE) along with the PiggyBase expressing vector in WT male ES cells (E14Tg2a) **(Fig. 2.3 A)**. Due to the GFP molecule present in the scFv fusion, stable integration of the two transgenes were expected to provide GFP expression under Dox treatment. Therefore, after transfection, cells were treated with Dox for two days and single GFP positive cells were sorted by FACS in separated wells. Growing clones were expanded and evaluated for GFP expression level and homogeneity under Dox treatment by fluorescent microscope observation. Based on this evaluation, six clones were sub-selected and further analysed by FACS. One of them (clone 5) showed together a high level of GFP induction, no detectable GFP expression in the absence of induction as well as a relatively homogenous GFP expression **(Fig. 2.3 B)**.

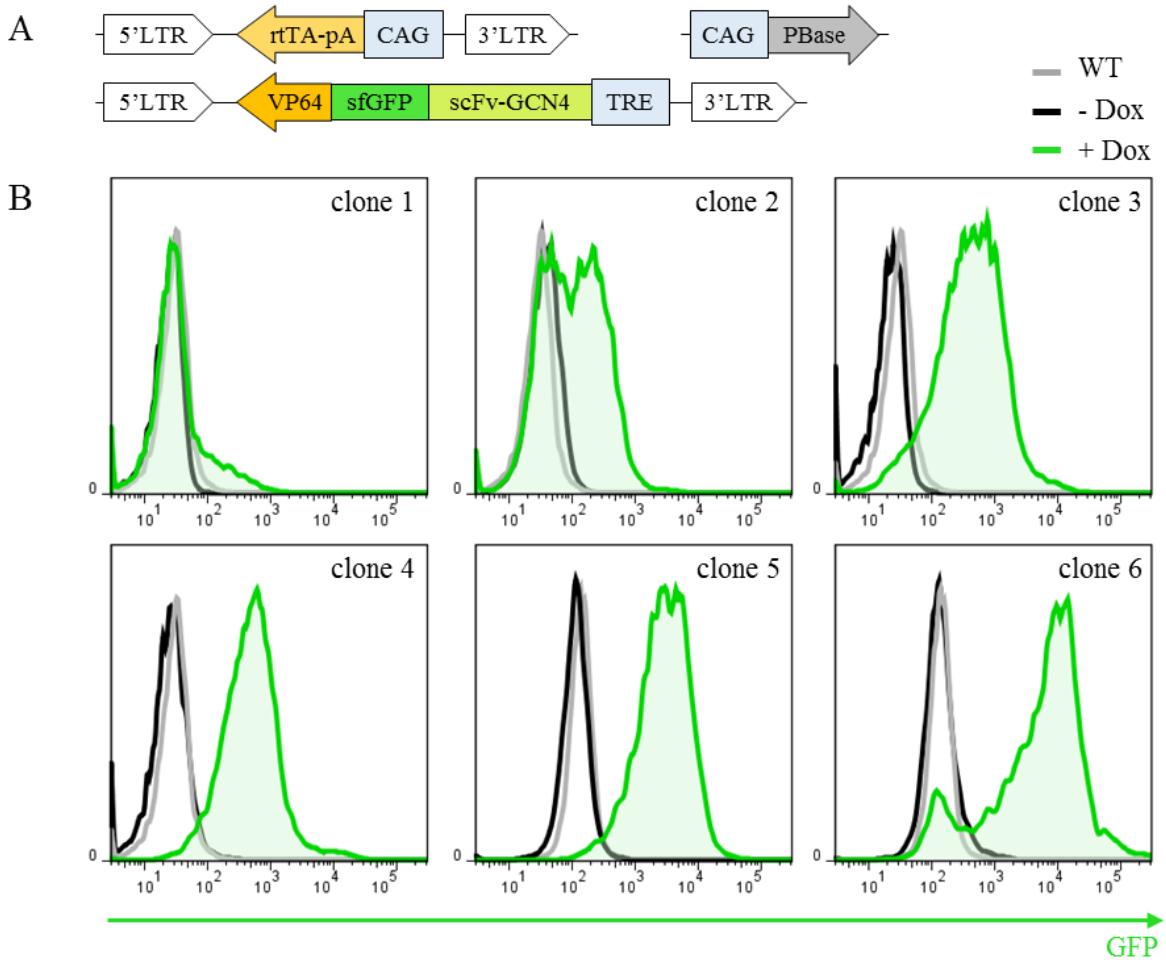


Fig. 2.3. A. Set of the three co-transfected vectors. **B.** FACS analysis of the six selected clones after 48h of treatment. Clone 5 shows the best proportion of GFP positive cells under induction with high level of GFP expression and no leakiness of expression in the absence of Dox treatment.

This clone was further transfected with a linearized vector harbouring an rtTA-responsive transgene composed of the dCas9-GCN4x10 fusion of the SunTag system linked to a blue fluorescent reporter (TagBFP) by a self-cleavage peptide (P2A) together with a constitutively expressed Neomycin resistance **(Fig. 2.4 A)**. Transfected cells were plated at clonal density and, after a week of antibiotic treatment, eighty-eight clones were manually picked. After expansion, all clones were put under Dox treatment and evaluated for GFP and BFP induction. Twenty-four of them were thus selected for further examination. The main purpose of this cell line being to functionally appreciate the effect of the overexpression of any gene on ES cell self-renewal, we decided to assess the ability of these clones to either reinforce or impair self-renewal when placed in corresponding conditions.

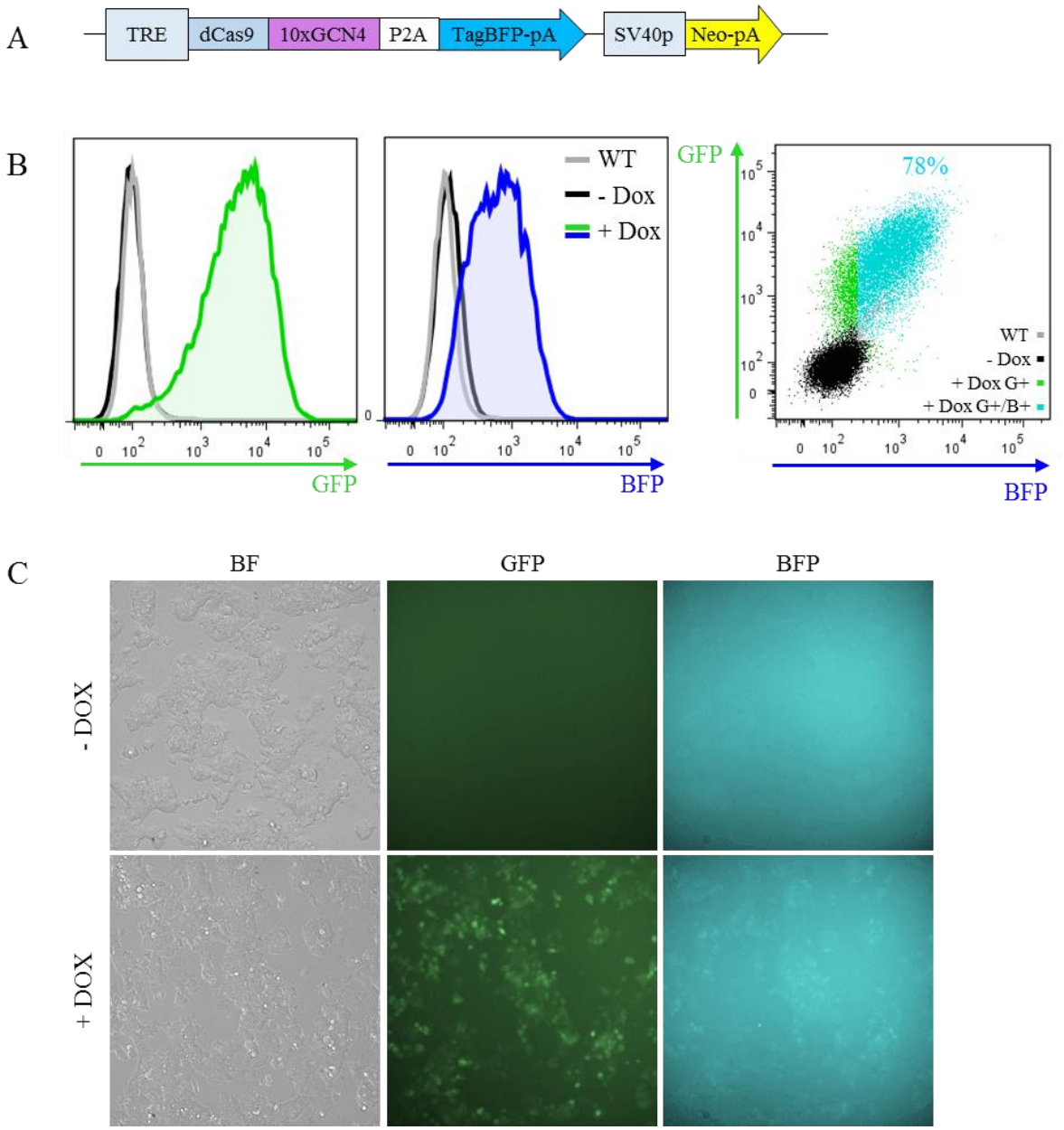


Fig. 2.4. A. Second integrated vector. **B.** FACS analysis of the selected clones at 24 hours post-treatment. Double GFP/BFP positive (G/B+) cells represent 78% of the whole population in the Dox treated sample (+Dox). 20% of the Dox treated cells were GFP positive only (G+). WT and non treated cells were analysed for comparison. **C.** Microscopy images of the first SunTag clone after 24 hours of Dox treatment.

Thereby, we evaluated their capacity to grow in 2i medium and differentiate upon LIF withdrawal or N2B27 medium culture in the presence or not of Dox treatment. Most of the clones formed nice dome-shaped colonies in 2i condition and undergo massive differentiation in adequate media while some barely differentiated or showed extensive death during differentiation and were hence discarded from further analysis.

Remarkably, fluorescent reporters signal was progressively lessening over both 2i adaptation and differentiation protocol most probably reflecting decreased expression of the transgenes due to their random integration in genomic regions essentially active in undifferentiated ES cell (not shown).

Finally, one of the clone passing the latter selection and showing wide induction of the two reporters was analysed by FACS **(Fig. 2.4 B).** This revealed simultaneous expression of the transgenes in about 78% of the cells (BFP/GFP positive) under Dox treatment. The Dox-dependent and widespread induction of the reporters could also be objectified by fluorescent microscopy **(Fig. 2.4 C)** despite the relatively low brightness of the TagBFP reporter.

However, following more cautious observation it appeared that this final clone showed abnormal morphological changes at low density in serum/LIF medium during early induction of the transgenes. Colonies appeared less compacted with many individualizing cells, resembling early differentiation **(Fig. 2.5)**. Yet, none of the tested pluripotency markers showed significant transcriptional change by RT-qPCR (not shown) upon Dox stimulation. Nevertheless, we decided to only use this clone for preliminary tests of the SunTag system and generate in the meanwhile a new generation of SunTag ES cell line.

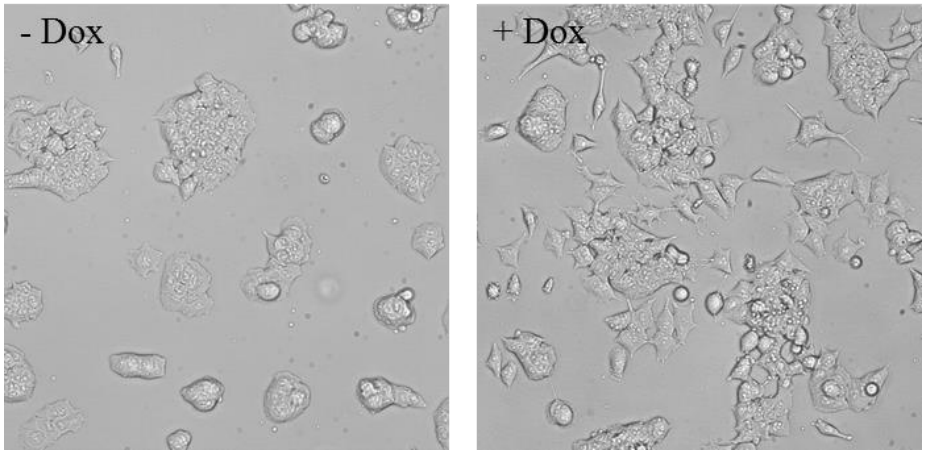


Fig. 2.5. Bright field pictures of the SunTag clone with (+Dox) or without (-Dox) Dox treatment for 48 hours. Decompaction of the colonies can be observed in +Dox condition.

C. Construction of the second SunTag ES cell line generation

As mentioned before, PiggyBac transposase has been shown to preferentially integrate the “donor” PiggyBac transposon towards open genomic regions. The advantage of this feature is that the resulting integrated vector will tend to be decently expressed in the transfected cell

line. But, the risk of integration in transcribed genes is also significantly increased compared with a truly random insertion and can thus lead to endogenous gene inactivation as well as the expression of chimeric proteins from endogenous and exogenous origins. In addition, even if PiggyBac transfection parameters can be optimized to promote single copy integration (W. Wang et al. 2008), the number of incorporated PiggyBac transposons is impossible to predict. Therefore, the number of transgenes to randomly integrate should be minimized as much as possible. In order to reduce the chance of such events we decided to reduce the number of vectors necessary to construct our SunTag cell line. This would also simplify the integration of the SunTag vectors in any given cell line. We finally inserted both components of the SunTag complex in a single PiggyBac vector downstream of two, head-to-head positioned, Dox responsive promoters **(Fig. 2.6 A)**.

This new vector was transfected along with the rtTA coding transgene and the PiggyBac transposase in early passage E14 WT cells **(Fig. 2.6 A)**. Cells were plated at clonal density in presence of Dox and Hygromycin as resistance was conferred by dCas9 transgene expression. Forty-eight single colonies were manually picked and expression of GFP and BFP reporters in presence of Dox was assessed under a fluorescent microscope. Based on this first examination with, this time, more careful attention to their morphological behaviour under induction, six clones were sub-selected for further evaluation. Reversibility of the reporters' expression after Dox stimulation was tested: none of the 6 clones showed residual GFP/BFP signal when put back in the absence of Dox for one day (not shown). All clones differentiated properly in the absence of LIF and in N2B27 medium and formed round shape colonies in 2i medium. Dox treatment didn't show any substantial effect on cell fate in any of these culture conditions. For better evaluation of the expression of the transgenes, the six clones were analysed by FACS in the presence and absence of Dox treatment **(Fig. 2.6 B)**. Three of them (clones 3, 4 and 5) showed leakiness of expression of both fluorescent reporters in the absence of stimulation when compared to WT cells. Clone 2 presented a significant proportion (40%) of Dox treated cells with no GFP or BFP signal. Clones 1 and 6 showed no or minor leakiness of expression in the absence of induction and revealed a large percentage of BFP/GFP positive cells with a reasonably sharp distribution of fluorescence under Dox treatment.

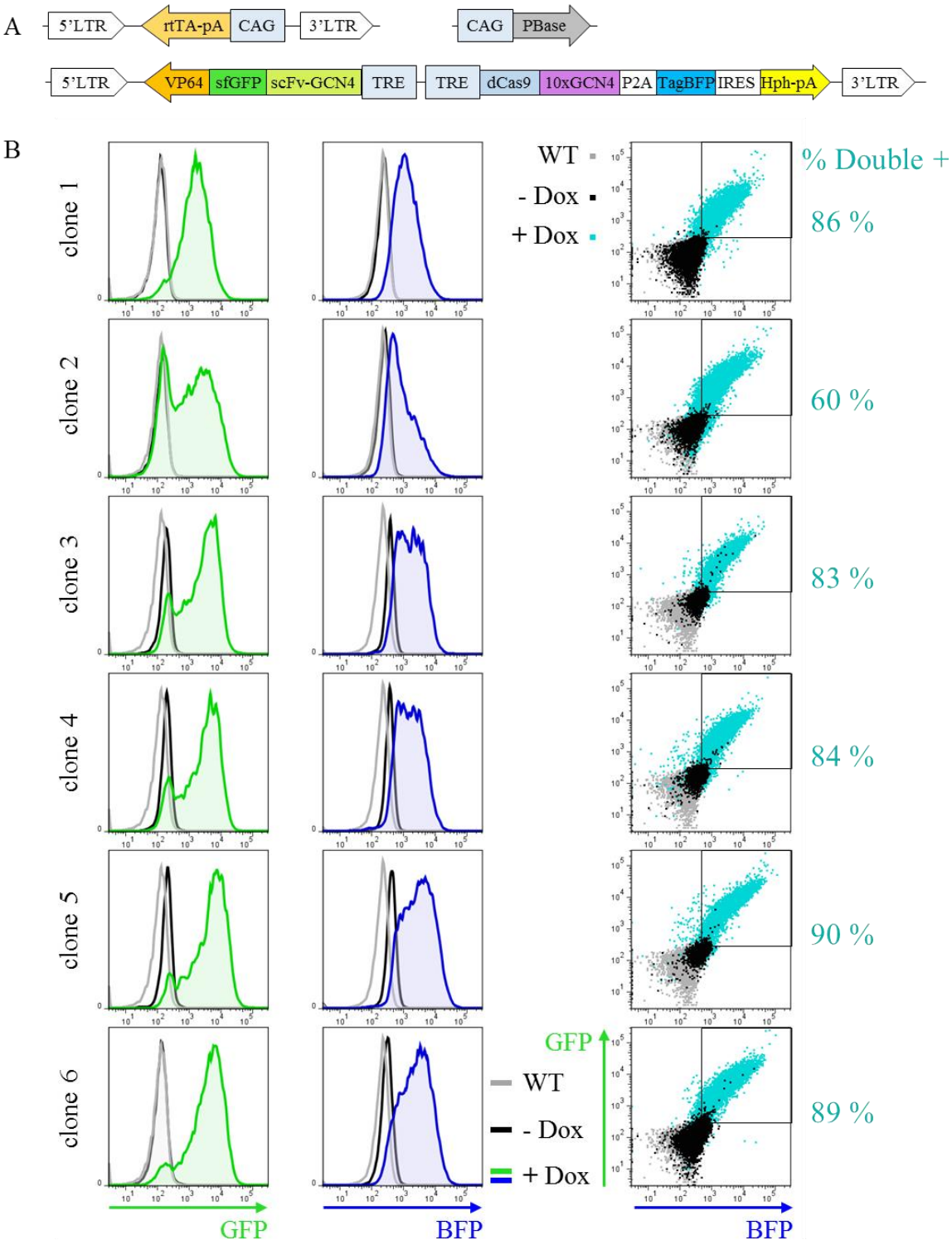


Fig. 2.6. A. Second set of vectors for the SunTag cell line construction. **B.** FACS analysis of the six selected clones after 48h of treatment. GFP and BFP levels were evaluated. Only clones 1 and 6 showed no leak of expression of the two transgenes (- Dox) as well as a good proportion of activated cells under Dox induction (+ Dox). Sharp and diagonal distributions of the Dox treated cells on the right panel show very high correlation between GFP and BFP expression.

Interestingly, with this new, "all-in-one" SunTag vector, BFP and GFP levels in single cell now clearly showed a linear relationship **(Fig.2.6 B)** thus ensuring simultaneous expression of the two parts of the SunTag complex in the same cell. This obviously increases the probability of homogenous and efficient induction among the cell population. The karyotype of clone 1 and 6 (thereafter called SunTag clones 1 and 2 for the rest of this manuscript) was examined and showed regular modal number of forty chromosomes. **(Fig. 2.7)**. Both clones were able to induce expression of the transgenes in differentiation conditions (-LIF and N2B27) for two days under Dox treatment although to a lesser extent (not shown) thus demonstrating their potential to induce gene expression in self-renewing ES cells as well as early differentiating cells.

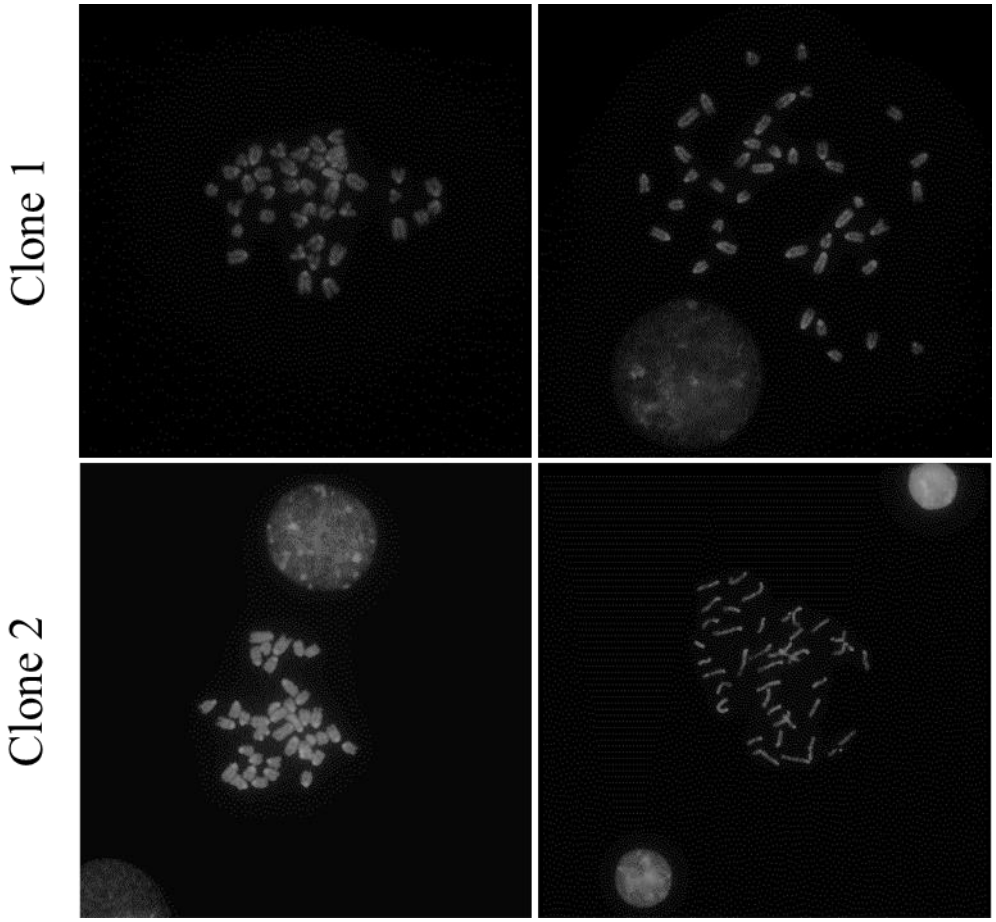


Fig. 2.7. Chromosome spread of clone 1 (top) and 2 (bottom) showing 40 chromosomes karyotypes. Cells were treated for 4 hours with Colcemid before fixation. More than 20 karyotypes were examined for each clone showing a modal number of 40 chromosomes.

D. D. Transcriptional Induction tests

We first decided to assess the efficiency of our slightly modified SunTag system on four endogenous promoters targeted by multiple gRNAs. We picked two pluripotency factors expressed at reasonably high level in ES cell and two lncRNAs which expression is relatively low and have been previously linked with ES cell biology (Mitchell Guttman et al. 2011). Six gRNAs were designed upstream of Nanog promoter, two upstream of Esrrb and linc1242 promoters and three upstream of AK031828 lncRNA **(Fig. 2.8 A)** (sequences listed in Table 1). The gRNAs were preferentially designed, when possible, in the region laying between fifty to four hundred base pairs upstream of the targeted gene TSS as being reported to be the best region for CRISPRa activity (Gilbert et al. 2014). All these gRNAs were cloned into a PiggyBac vector harbouring a gRNA cassette and driving Puromycin resistance. These vectors were transfected separately along with the PiggyBac transposase **(Fig. 2.8 B)** for stable integration in the SunTag clone resulting from the first generation (section I.A.). After Puromycin selection, cells were plated in the presence or not of Dox and RNAs were extracted two days later. Induction of the targeted genes was tested by RT-qPCR **(Fig. 2.9 A)**.

Nanog mature and pre-messenger RNA (Nanog pre) expressions were induced with five out of the six tested gRNAs. This induction raises from two to four fold. The only gRNA designed downstream of Nanog TSS appeared to be inactive, most likely due to the reported interference of dCas9 binding with transcription initiation when recruited on or closely downstream of a TSS (Qi et al. 2013; Gilbert et al. 2013) **(Fig. 2.9 A)**. The same experiment was done with four out of the six gRNAs in parallel where LIF was withdrawn at the beginning of the induction **(Fig. 2.9 B).** While Nanog expression dropped in that condition as expected, the SunTag induction fold change under Dox treatment was here much higher. However, the "absolute" levels of Nanog reached in the absence and the presence of LIF were actually really close to each other, suggesting that a maximal level of Nanog induction was obtained in both conditions with our experimental set up.

The overexpression of lncRNA AK0301828 was achieved with all three designed gRNAs targeting its promoter. A minimum of threefold induction was obtained despite the fact that two gRNAs were located further than five hundred base pairs away from its TSS. However, no induction was observed for Esrrb neither linc1242 despite the *a priori*, rationally well-designed and located double gRNAs. **(Fig. 2.8 A and 2.9 A)**. These results quickly illustrate the gRNA and gene-dependent activity of CRISPR activation method.

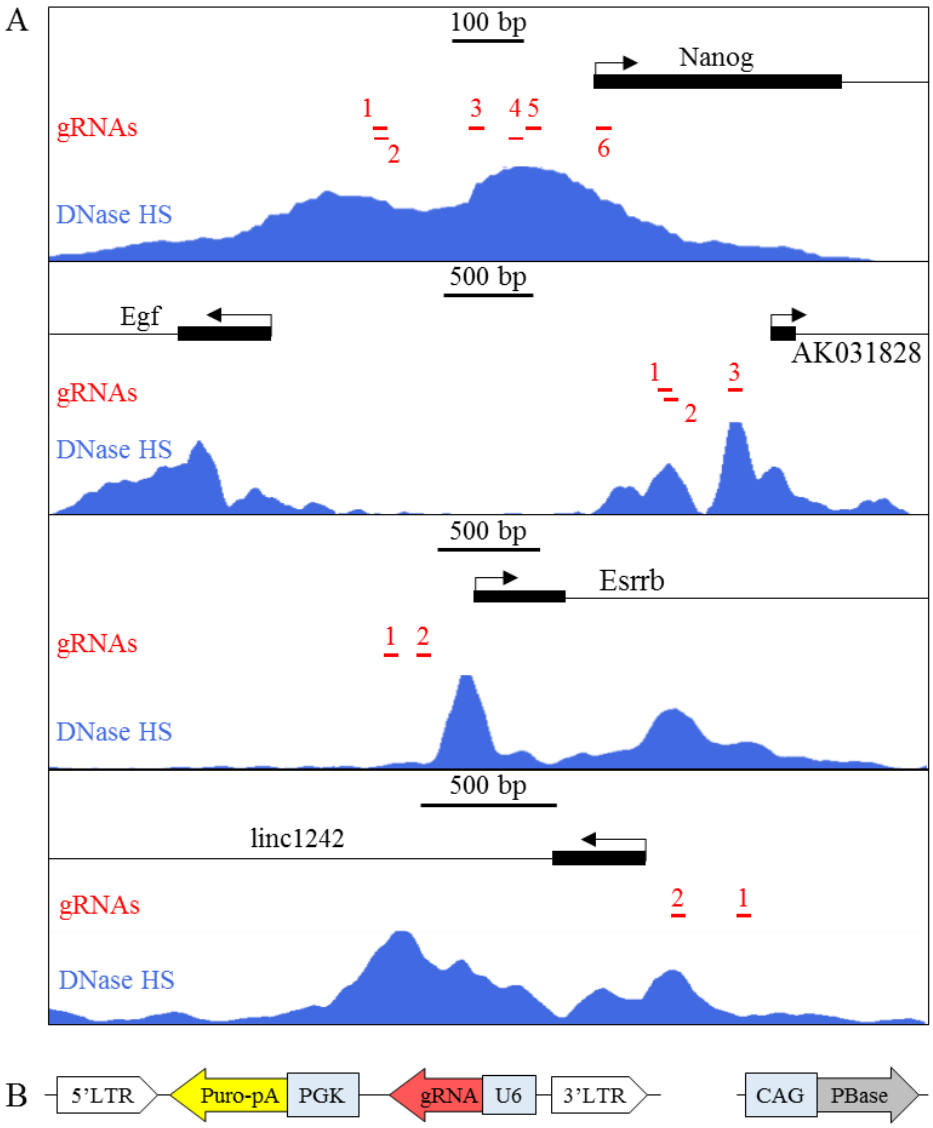


Fig. 2.8 A. Schematic illustration of the gRNAs location on the promoter of each targeted gene. Genomic scale is represented above each track. DNase Hyper sensibility signal is showed to illustrate chromatin openness and was obtained from the UCSC genome browser (ENCODE track) **B.** PiggyBac vector for gRNA expression: gRNA transcription is driven by a RNA Pol-III U6 promoter. A Puromycin resistance cassette is transcribed from an independent PGK promoter. Its transfection along with the PBase coding vector allows for stable genomic integration.

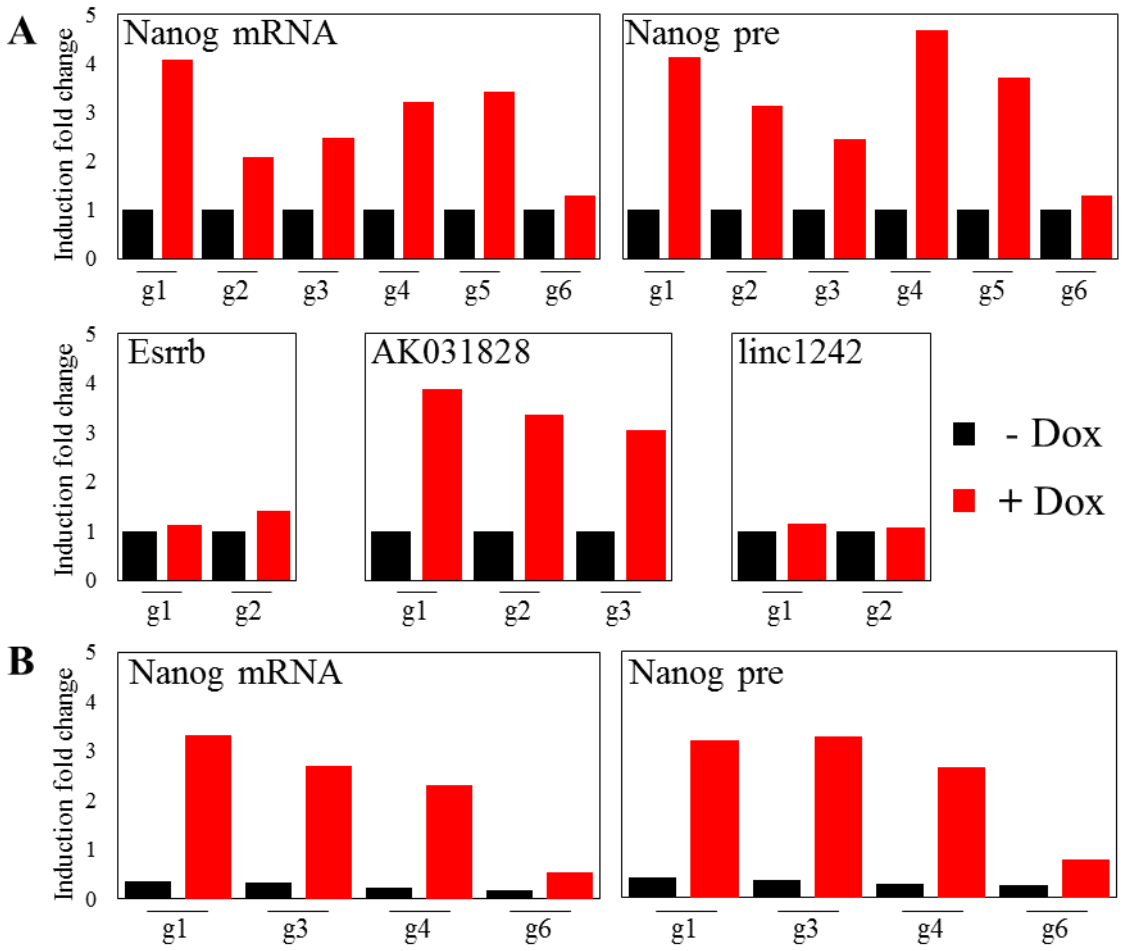


Fig. 2.9 A. RT-qPCR measure of the induction of endogenous genes with the first generation SunTag clone in serum/LIF condition. Cells were treated with Dox for 48hrs. Values are normalized to Tbp level and - Dox condition is set to 1 for each gRNA. **B.** Same experiment as **A.** where LIF was withdrawn at the beginning of Dox treatment. Since both experiments were done in parallel, values are normalized to +LIF level without Dox from **A**. such that induction fold changes are comparable.

We validated the induction of Nanog expression with the six clones of the second SunTag cell line generation (section I.B.) with one of the best inducing gRNA (gRNA 1). As shown before, the level of Nanog reached upon induction in presence or absence of LIF was comparable and all clones appeared to show equivalent induction strength in both conditions **(Fig. 2.10 A)**. In addition, the two definitive SunTag clones (1 and 2) harbouring the same gRNA were cultured in 2i medium for three days and plated again for three more days in 2i with Dox treatment or not. While Nanog basal level was already higher than compared to serum containing culture, a mild increase (lower than two fold) seems to be still achieved, though not

reaching the same level obtained in serum culture. This effect appeared to be consistent between mature as well as pre messenger Nanog RNA **(Fig. 2.10 B)**.

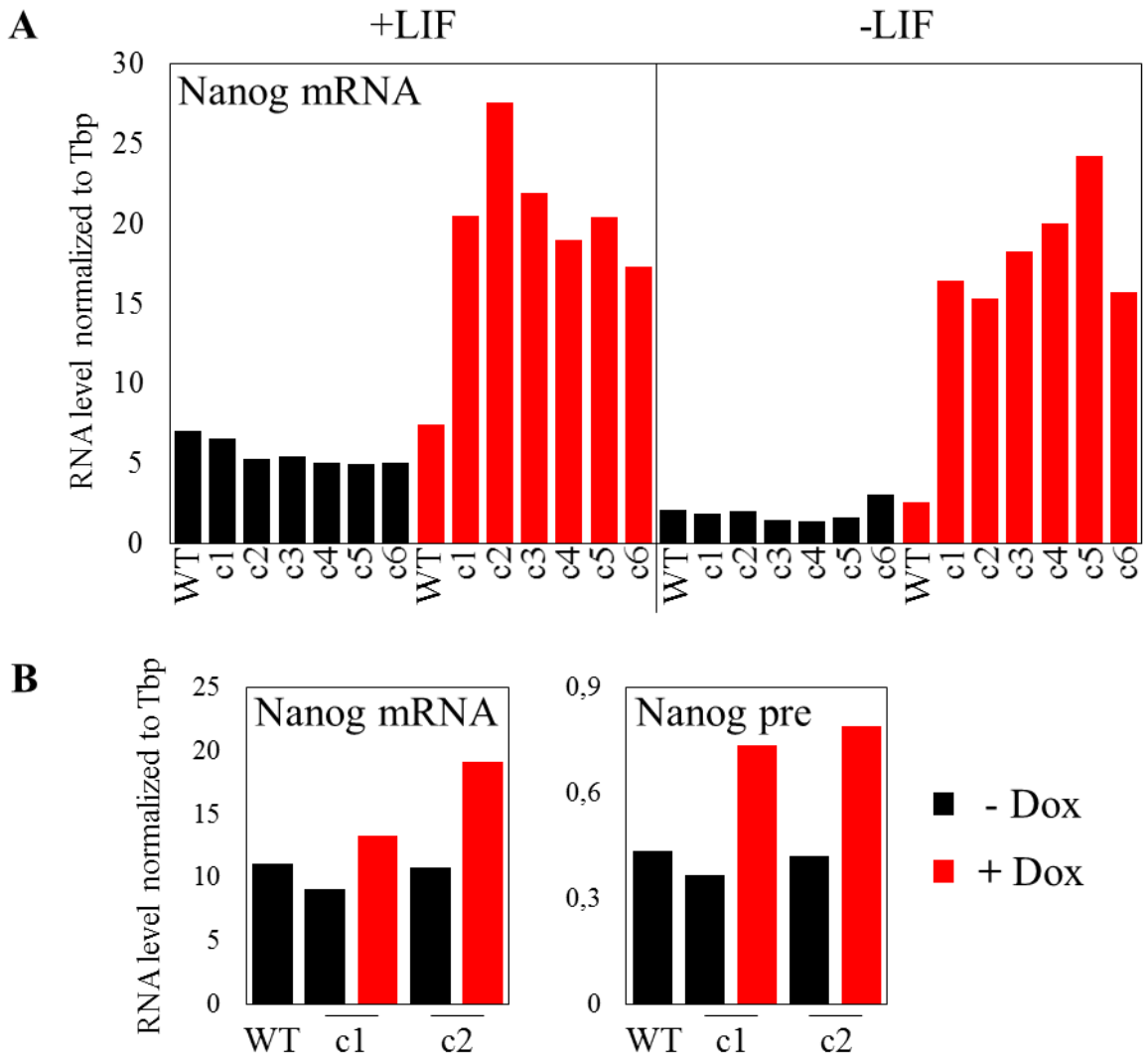


Fig. 2.10 A. RT-qPCR measure of induction of Nanog gene with the SunTag cell line in serum/LIF (left) or upon LIF withdrawal (right) in the six SunTag clones (second generation) compared to WT cells. Values are normalized to Tbp level. **B.** RT-qPCR measure of induction of Nanog gene with the two SunTag clones. Cells were cultured for 3 days in 2i/LIF and kept in 2i/LIF for 3 additional days with or without Dox treatment in parallel with WT cells. Values are normalized to Tbp level.

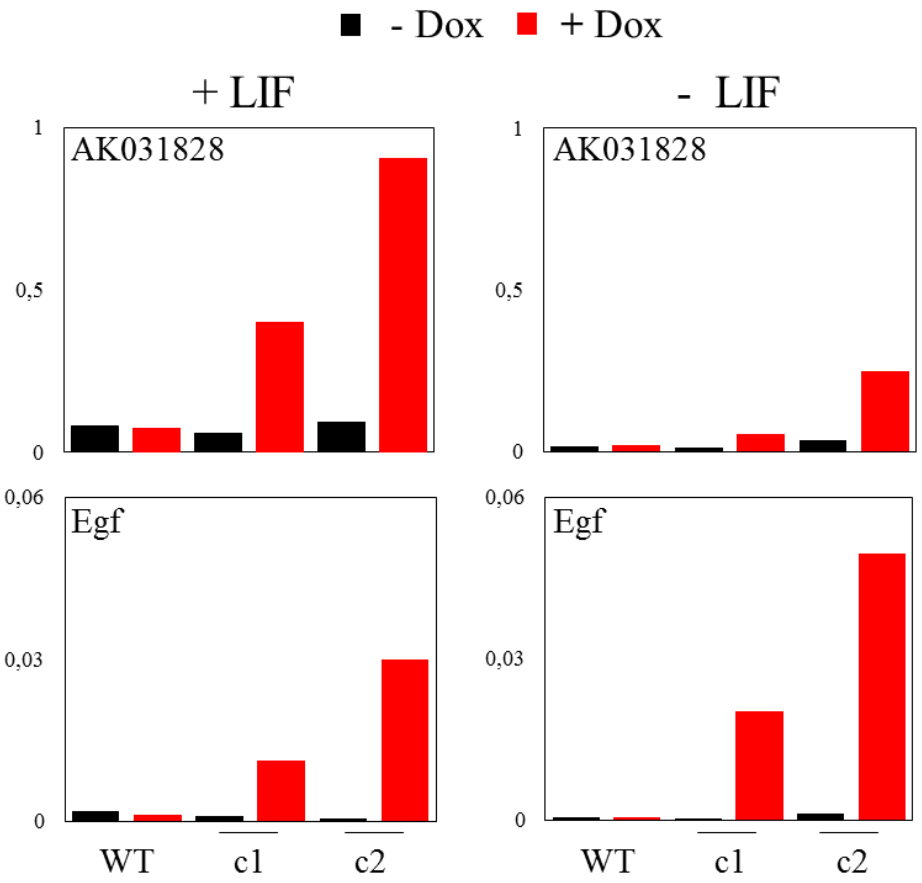


Fig. 2.11. RT-qPCR measure of AK031828 lncRNA induction and its closest neighbouring gene Egf with gRNA g1 (see Fig. 2.8. A). Experiment was performed with both SunTag clones in FCS/LIF or upon LIF withdrawal for 2 days in parallel with WT cells. Values are normalized to Tbp level.

Interestingly, it appeared that Egf, the closest gene to AK031828 initiating 3kb upstream in the opposite direction **(Fig. 2.8 A)**, was strongly upregulated by the overexpression of its non-coding neighbour **(Fig. 2.11)**. But surprisingly, while the overexpression of AK031828 was severely reduced, if not completely abolished, in the absence of LIF, likely due to its natural extinction upon LIF withdrawal, the simultaneous induction of Egf showed an opposite tendency, although its basal expression didn't show any significant change. This result subtly suggests that activation of both transcripts results from two uncoupled mechanisms, therefore unveiling a proximal enhancer-like function for SunTag complex binding.

Indeed, such a function unmistakably reminds VP16 transactivation domain ability to recruit histone acetyl-transferases (Choy and Green 1993; Utley et al. 1998; Vignali et al. 2000; Kundu et al. 2000). This potential effect prompted us to explore the capacity of SunTag to activate gene expression from a long-distance regulatory element. We therefore designed two gRNAs targeting a putative long distance (about 150kb) enhancer of Pax6 gene identified by a

collaborating group (Noordermeer D. laboratory, Gif-sur-Yvette, France) as physically interacting with Pax6 promoter in mouse ES cells by Hi-C/4C experiments (unpublished) **(Fig. 2.12).** Pax6 is a transcription factor known as a master developmental regulator of ectodermal tissues (Osumi et al. 2008; Blake and Ziman 2014). Its CpG island promoter is one of the so-called bivalent domains of mouse ES cells and is therefore, as well as its gene body, targeted by the PRC2 polycomb complex and covered by a large H3K27me3 repressive domain (Bernstein et al. 2006; Sachs et al. 2013). Therefore, we designed two additional gRNAs on Pax6 promoter in order to assess whether this gene was refractory to transcriptional activation by CRISPR activation as Cas9 binding has been shown to be affected by heterochromatin structures (Xiaoyu Chen et al. 2016; Jensen et al. 2017). All four gRNAs were integrated separately in one of the two SunTag clones (clone 1). Cells were treated or not with Dox for two days and transcriptional effects were assessed by RT-qPCR **(Fig. 2.13)**.

Recruitment of the SunTag complex to Pax6 promoter lead to a three and seven fold increase of expression with gRNA 1 and 2 respectively showing a good responsiveness to Pax6 locus for CRISPRa stimulation despite polycomb-mediated repression. Likewise, both gRNAs targeting the aforementioned distal regulatory region resulted in the increased transcription of local enhancer RNAs (eRNAs). However, in that case, no effects were detected on Pax6 transcription as shown by intronic primers pair **(Fig. 2.13)**. This negative result then raised the possibility that this enhancer could actually be involved in the regulation of another surrounding gene. Subsequently, the genes lying in the same topologically associated domain (TAD) (Nora et al. 2012; Dixon et al. 2012; Sexton et al. 2012) were delineated according to public Hi-C data (3D genome browser, Yue Lab). All of them appeared to be unaffected by the stimulation of Pax6 promoter or its putative enhancer **(Fig. 2.13)**. Thus, we were unable to decipher the regulatory function of this region within the scrutinized genomic region and to assess whether the transcriptional induction of this distal enhancer was sufficient for gene activation.

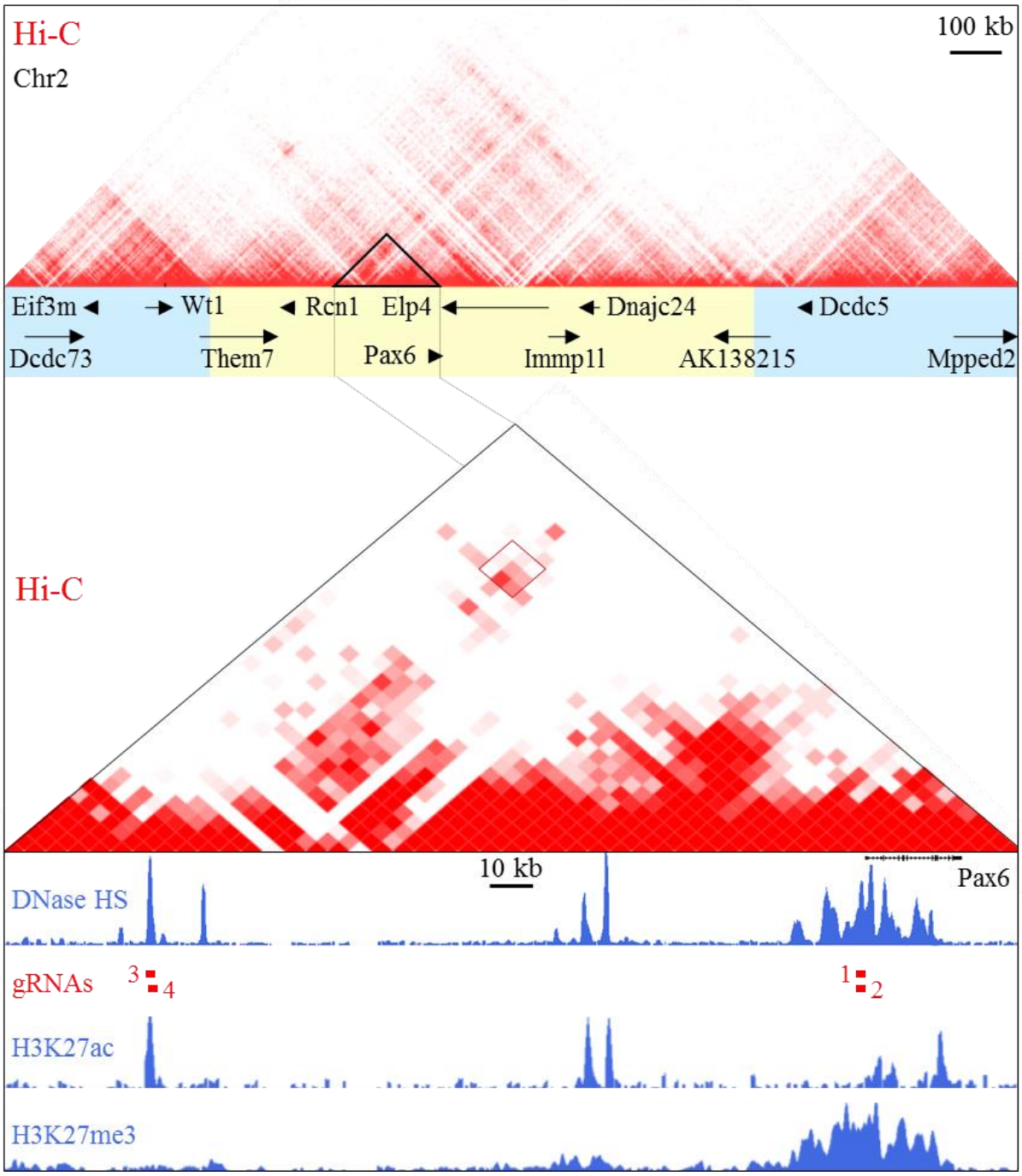


Fig. 2.12. Pax6 neighbourhood and gRNAs design. Pax6 TAD (yellow) and the two surrounding TADs (blue) are shown in the upper picture with Hi-C heatmap obtained from 3D genome browser - Yue Lab website - with mouse ES cells dataset from (Bonev et al. 2017). Black triangle corresponds to the zoomed in region below. gRNAs 3 and 4 are designed over the targeted putative enhancer of Pax6 while gRNAs 1 and 2 directly recruit the SunTag system to its promoter. DNase hypersensitivity (HS) signal is shown to illustrate chromatin openness (UCSC browser – ENCODE), H3K27ac and H3K27me3 enrichments are shown to attest for enhancer regions and polycomb-mediated Pax6 repression respectively.

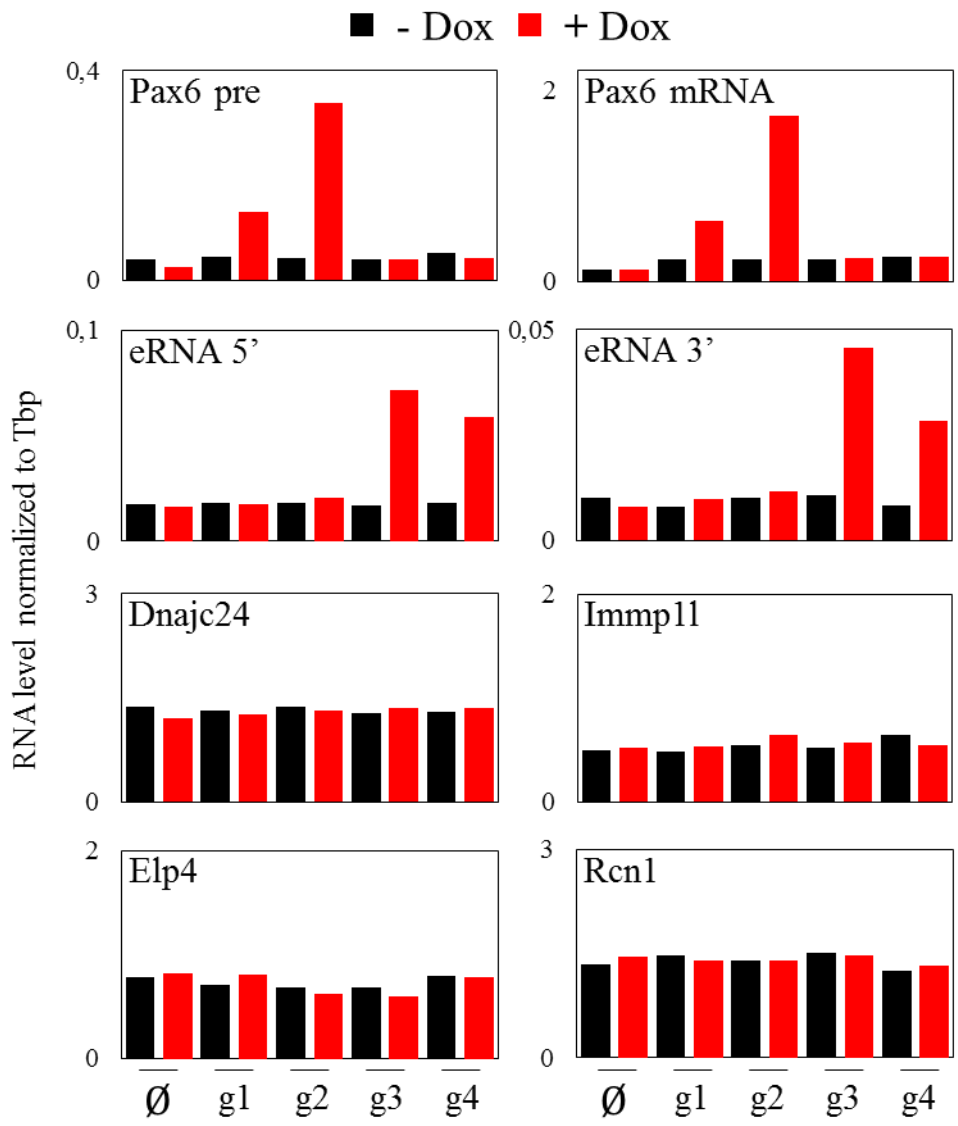


Fig. 2.13. RT-qPCR measure of Pax6 mature (mRNA) and pre-messenger (pre) RNA, enhancer RNAs upstream (eRNA 5') or downstream (eRNA 3') of the targeting gRNAs, and of the Pax6 TAD associated genes expression. This experiment was performed in one of the two SunTag clones in in serum/LIF medium. Cells were cultured for two days with Dox treatment or not. Values are normalized to Tbp level.

E. Discussion

PiggyBac transposon as an active genomic integration method is a powerful and convenient tool for stable integration of exogenous DNA. Moreover, its ease of cell delivery and insertion is a great resource when combined with the versatility of CRISPR technology based on gRNA's simple design and cloning (Xu et al. 2017). However, random integration of several vectors for the construction of a stable cell line revealed not to be an optimal choice. Indeed, the unexpected behaviour of the "first generation" SunTag clone **(Fig. 2.5)** can be most likely explained by the integration of one of the inducible SunTag vectors in a region where its activation led to the modified expression of a gene impacting ES cell fate or morphology. Genomic mapping of the insertion of the vectors was attempted by splinkerette PCR but remained unsuccessful (Horn et al. 2007). Moreover, the fact that Nanog induction in 2i medium was decreased relatively to what was obtained in serum containing medium **(Fig. 2.10 B)** might be the direct consequence of a lower expression of the SunTag components as fluorescent reporters signal was usually reduced in this medium under Dox treatment (not shown). Consequently, homologous recombination in a ubiquitously expressed locus like the well-known Rosa26 region (Friedrich and Soriano 1991) should perhaps have been favoured for the integration of the SunTag and rtTA vectors thus reducing the risk of endogenous gene expression alteration and transgene silencing upon varying cell fate. Nevertheless, the complete set of SunTag, rtTA and gRNA PiggyBac vectors constitutes a powerful toolkit for rapid gene overexpression in any mammalian cell type.

The fold changes order of magnitude reported in few CRISPRa papers is astronomical (up to five log10) (Perez-Pinera, Kocak, et al. 2013; Chavez et al. 2015; Polstein and Gersbach 2015; Chavez et al. 2016). One could be surprised by the moderate efficiency of one of these systems in our hands. First, in all the cases showing such efficiencies, gRNAs pools were used to activate a single gene as a synergistic effect has been shown for artificial transcriptional activators (Perez-Pinera, Ousterout, et al. 2013; Cheng et al. 2013). Second, such fold changes are only obtained for genes that are very lowly expressed or even not detectable in the cell line in which they were induced, thus producing amazingly high values of difference. In addition, a negative correlation between gene expression level and power of induction was clearly established (Konermann et al. 2015; Chavez et al. 2015, 2016). One of those studies in fact precised that, after CRISPRa induction, the resulting expression level of these genes remained lower or comparable to the level of expression in the tissue where they are naturally expressed

(Chavez et al. 2015). For comparison, in our case, the induction fold change of Egf, which is nearly not expressed in mouse ES cells, through targeting of AK031828, was indeed ranging from twelve to eighty-six **(Fig. 2.11)**. But in our hands, for active genes, the fold change of induction rarely exceed ten, closer to what has been reported before with a single gRNA targeting an already expressed gene (Cheng et al. 2013; Tanenbaum et al. 2014; Gao et al. 2014). However, in agreement with the literature, our results show that CRISPRa effect is highly gene-dependent. Two of our tested genes didn't respond at all to SunTag activation **(Fig. 2.9 A)**. Of note, Esrrb promoter gives rise to bidirectional transcription with synthesis of a non-annotated lncRNA transcribed in the opposite direction to Esrrb gene (**Fig. 2.14 A**, identified in section III). The two gRNAs designed and tested on this region are actually located in the first exon of this non-coding transcript. Since divergent lncRNAs expression has been shown to be important for adjacent gene regulation in mouse ES cells (Sigova et al. 2013; Luo et al. 2016), it would be interesting to see whether dCas9 recruitment at these positions disrupts the non-coding RNA expression possibly counteracting the otherwise potential induction of Esrrb.

Bidirectional induction of transcription seems to be a feature shared by CRISPRa as illustrated by the symmetrical induction of AK031828 and its divergent neighbouring gene Egf **(Fig. 2.8 A)**. In that case, SunTag recruitment about two kilobases upstream of Egf TSS likely acts as a proximal enhancer, despite the absence or the very low enrichment for enhancer characteristic features (H3K4me1, H3K27ac, p300) at this site **(Fig. 2.14 B)**. The lncRNA AK031828 is noticeably repressed upon LIF withdrawal, possibly related to the high enrichment in ERVK repeat elements found in its sequence whose activation has been previously linked with undifferentiated state of pluripotent stem cell (The FANTOM Consortium et al. 2014) **(Fig. 2.14 B)**. But while AK031828 induction is strongly reduced, the activation of Egf in the absence of LIF seems surprisingly higher than in + LIF condition, an effect that looks consistent between the two SunTag clones **(Fig. 2.11)**. This effect might be driven by local chromatin reorganization of the locus upon LIF withdrawal favouring physical contacts between the gRNA-targeted region and the promoter of Egf or by the redirection of the transcriptional machinery from AK031828 towards Egf promoter. This suggests that SunTag activity is partially dependent of and cannot completely override endogenous regulation of gene expression. Of note, a reverse but comparable response was detected upon SunTag activation of Nanog promoter on the transcriptional activity of its -5 kb enhancer (see section II) where bidirectional transcription was shown to be induced.

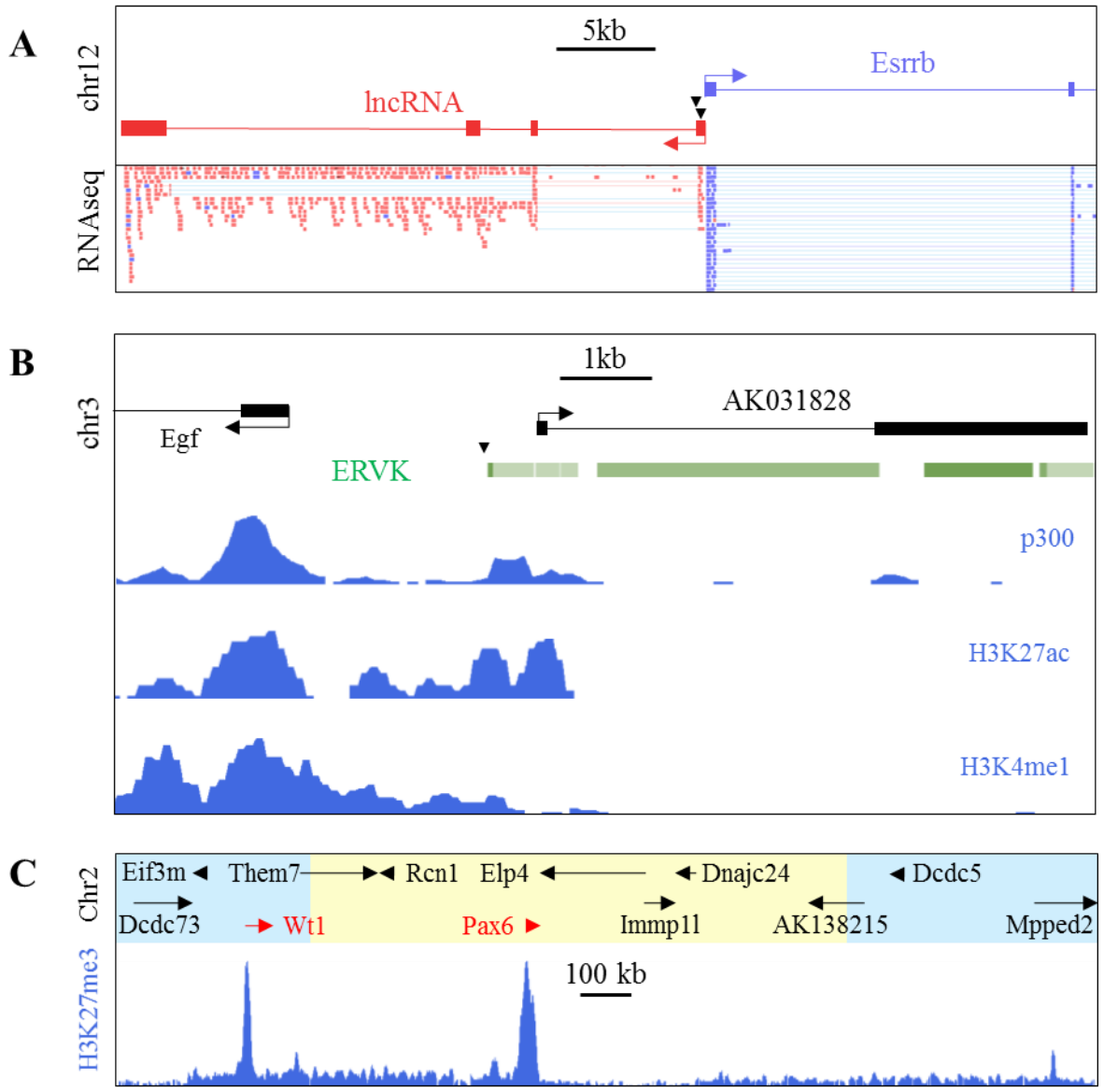


Fig. 2.14. A. Schematic representation of Esrrb and its divergent lncRNA. Small black triangles symbolized gRNAs targeting sites laying in the lncRNA first exon. RNA-seq reads from WT ES cells are shown below to attest of both divergent transcription. **B.** Schematic representation of AK031828 and Egf locus showing ERVK elements enrichment and characteristic enhancer features along the region (UCSC browser – ENCODE data) **C.** H3K27me3 enrichment in the region showed in **Fig. 2.12** showing the presence of two strong domains towards Pax6 and WT1 loci.

It was surprising for us to see that SunTag tethering to a given promoter can independently show proximal enhancer function for the closest neighbouring gene. Indeed, it was shown that dCas9-VP64 recruitment was truly inefficient in activating proximal enhancers, especially if compared to dCas9-p300 fusions (Hilton et al. 2015). However, the maximum number of VP64 copies theoretically recruited by the SunTag complex is ten times higher than

a direct dCas9-VP64 fusion thus possibly increasing its ability to activate enhancer or generate similar effect. Indeed, as mentioned before, VP16 transactivation domain has been clearly shown to recruit different histone acetylases (Choy and Green 1993; Utley et al. 1998; Vignali et al. 2000; Kundu et al. 2000). We were not able to demonstrate such an effect on the putative long-distance (150 kb) enhancer of Pax6. Despite transcriptional activation of local eRNAs no impact could be objectified on all tested surrounding genes **(Fig. 2.13)**. It is worth mentioning that the most distal activation reached through dCas9-p300core targeting to an enhancer was obtained for a forty-six kilobases distance but failed for a longer one (Hilton et al. 2015). This result was obtained on the well-characterized human β-globin locus where a single enhancer in known to activate four haemoglobin coding genes (D. Carter et al. 2002; W. Deng et al. 2014). Another regulatory region showing clear enhancer features (H3K27ac, H3K4me1, p300, visualized on UCSC browser, ENCODE data) is present between Pax6 gene and our targeted enhancer **(Fig. 2.12)**. The interaction between this region and the Pax6 locus seems relatively strong according to the Hi-C data shown previously. Thus, it is most likely that other regulatory regions also take part in orchestrating Pax6 expression and might possibly hold prevalent functions. This question could actually be addressed by activation of all regulatory elements surrounding Pax6 genes through single or multiplexed targeting with CRISPRa: an elegant approach used to dissect enhancers regulatory functions of a given gene by systematic and combinatorial repression or activation (Fulco et al. 2016; Carleton, Berrett, and Gertz 2017; Thormann et al. 2018). Another approach, to assess the role of this potential enhancer in early activation of Pax6, could be to repeat the same experiment at the onset of mouse ES cells differentiation towards neuronal lineage by retinoic acid treatment, a set up where Pax6 gets properly activated (Gajović et al. 1998). Pax6 promoter contains a high density CpG island and is largely covered by a repressive H3K27me3 domain. It was very recently shown that PRC2 complex in mouse ES cells is recruited in the first place to "nucleation sites" consisting of dense CpG island promoters and further spread out to other repressed regions by *cis* long range contacts (Oksuz et al. 2018). Interestingly, Pax6 was reported to be one of those specific regions. Hi-C heatmap of **Fig. 2.12** shows a long distance inter-TAD interaction between Pax6 and Wt1 loci (isolated red area in the top left of the Fig.). Moreover, it appears that Wt1 gene is also highly enriched in H3K27me3 histone mark **(Fig. 2.14 C)**. It is thus possible that Wt1 locus is one of the region where Polycomb mediated repression is spreading out from Pax6 promoter, therefore arguing for Pax6 locus as the core of a physically interacting, repressed, multiple regions hub. It is thus comprehensible that single long-distance enhancer activation doesn't succeed in inducing this central repressive region. Direct recruitment of SunTag

complex towards Pax6 promoter instead appeared to be strong enough to override this repressed state. Nonetheless, the fact that activation of Pax6 gene as well as local eRNAs didn't lead to any transcriptional effect among their four neighbouring genes also appears as a good indication of SunTag local specificity.

Finally, we built a set of PiggyBac vectors allowing for rapid generation of inducible SunTag cell lines. We further constructed a mouse ES cells line stably harbouring these transgenes and showed that we were able to, under Dox treatment control, activate endogenous genes expression and non-coding RNAs transcription. It also appeared that SunTag recruitment seems to confer proximal enhancer function but that such effect stays locally specific.

VIII. The molecular logic of Nanog-induced self-renewal

The molecular logic of Nanog-induced self-renewal

Victor Heurtier[1,2,a], Nick Owens[1,a], Inma Gonzalez[1], Florian Mueller[3], Caroline Proux[4], Damien Mornico[5], Philippe Clerc[1], Agnès Dubois[1], and Pablo Navarro[1]

[1]Epigenetics of Stem Cells, Department of Developmental and Stem Cell Biology, Institut Pasteur, CNRS UMR3738, 25 rue du Docteur Roux, 75015 Paris, France.
[2]Sorbonne Université, Collège Doctoral, F-75005 Paris, France.
[3]Imaging and Modelling, Department of Cell Biology and Infections, Institut Pasteur, CNRS UMR 3691, 25 rue du docteur Roux, Paris 75015, France.
[4]Transcriptome and EpiGenome, BioMics, Center for Innovation and Technological Research, Institut Pasteur, 28 rue du docteur Roux, 75015 Paris, France.
[5]Bioinformatics and Biostatistics Hub – C3BI, Institut Pasteur, CNRS USR 3756, 28 rue du docteur Roux, Paris 75015.
[a]Equal contribution

Transcription factor networks, together with histone modifications and signalling pathways, underlie the establishment and maintenance of gene regulatory architectures associated with the molecular identity of each cell type. However, how master transcription factors individually impact the epigenomic landscape and orchestrate the behaviour of regulatory networks under different environmental constraints is only very partially understood. Here, we show that the transcription factor Nanog deploys multiple distinct mechanisms to enhance embryonic stem cell self-renewal. In the presence of LIF, which fosters self-renewal, Nanog rewires the pluripotency network by promoting chromatin accessibility and binding of other pluripotency factors to thousands of enhancers. In the absence of LIF, Nanog blocks differentiation by sustaining H3K27me3, a repressive histone mark, at developmental regulators. Among those, we show that the repression of Otx2 plays a preponderant role. Our results underscore the versatility of master transcription factors, such as Nanog, to globally influence gene regulation during developmental processes.

Keywords: Nanog, Pluripotency Network, H3K27me3, Otx2, ES cells, Self-renewal, Differentiation

Correspondence: ***pnavarro@pasteur.fr***

Introduction

Gene regulatory networks driven by master transcription factors (TFs) play pivotal roles over a large spectrum of biological processes, from adaptive cell responses **(Shinozaki et al., 2003)** to cell fate specification during development **(Davidson et al., 2002)**. The key properties of TF networks, shared among cell types, developmental contexts and organisms **(Huang et al., 2005)**, are exemplified by the pluripotency network, which plays a dominant role during early mammalian embryogenesis **(Frum and Ralston, 2015)**. The robustness of this network allows the capture *ex vivo* of the transient biological identity of the pluripotent epiblast through the derivation of self-renewing Embryonic Stem (ES) cells **(Parfitt and Shen, 2014)**, which have enabled identification of key TFs (e.g. Oct4, Sox2, Nanog and Esrrb). The study of processes driving the balance between ES cell self-renewal and differentiation has provided us with a canonical picture of how TF networks operate, establishing self-sustaining regulatory loops and acting together through multiple promoters and enhancers **(Boyer et al., 2005; Loh et al., 2006; Chen et al., 2008; Kim et al., 2008)**. For instance, Oct4, without which pluripotent cells cannot be maintained **(Niwa et al., 2000)**, acts with the TF Sox2 to recognise and bind chimeric motifs **(Chew et al., 2005)** found at a large number of regulatory elements driving ES cell-specific transcription. Oct4 and Sox2 tend also to bind with other TFs, including Nanog and Esrrb, at multiple enhancers across the genome, to combinatorially coregulate a large number of targets. This simultaneous and concerted action over hundreds of common targets ensures extensive redundancy, and, therefore, robust genome-wide responses. How these TFs synergise at or compete for common regulatory elements, and how by these means they individually contribute to the network's activity, is however not well understood. Moreover, several TFs of the pluripotency network are directly connected to cell signalling, enabling ES cells to establish appropriate responses that are instructed extrinsically. A prominent example is provided by the LIF cytokine, which promotes self-renewal by activating several pluripotency TFs such as Esrrb **(Niwa et al., 2009; Huang et al., 2018)**. Hence, a key function of the pluripotency network is to integrate signalling cues to appropriately respond to changes in the environment, conferring the responsiveness of ES cells and their capacity to readily differentiate. In this regard, it is noteworthy that *Nanog* was first identified as a factor capable of bypassing the requirements for LIF: in the presence of ectopic Nanog expression, ES cell self-renewal is strongly enhanced and completely independent of LIF **(Chambers et al., 2003)**. In the current model, Nanog achieves LIF-independent self-renewal by activating LIF-responsive genes, in particular *Esrrb*. Hence, the Nanog-Esrrb axis and its intersection with LIF signalling represents a major mechanism by which intrinsic and extrinsic cues fine-tune self-renewal and avoid differentiation **(Festuccia et al., 2012)**. Yet, the precise mechanisms by which Nanog, and more generally the pluripotency network, controls differentiation genes are not fully understood. It is known, however, that differentiation genes adopt a particular chromatin state known as "bivalent" **(Azuara et al., 2006; Bernstein et al., 2006)**: while their promoters are enriched for H3K4me3, a mark of gene activity, they are simultaneously embedded within larger domains of H3K27me3, a repressive mark. During differentiation, this state is resolved in either H3K27 or K4me3 in a lineage-specific manner **(Mikkelsen et al., 2007)**. In agreement,

Polycomb Group proteins triggering H3K27me3 ensure appropriate cell fate changes **(Boyer et al., 2006; Pasini et al., 2007; Leeb et al., 2010)**. This underscores the importance of H3K27me3 as cells dismantle the pluripotency network, inhibit self-renewal and exit from pluripotency. Whether bivalent chromatin marks are governed by pluripotency TFs remains to be thoroughly addressed.

In this study, we explore the function of Nanog in mouse ES cells using inducible approaches of gain- and loss-of-function. We show that Nanog drives the recruitment of Oct4, Sox2 and Esrrb at thousands of regulatory regions, from where it mainly activates transcription. At these sites, Nanog also recruits Brg1 and promotes chromatin accessibility. On the contrary, to repress transcription Nanog does not recruit these TFs; rather, it frequently inhibits Oct4 or Sox2 binding. Nanog also binds at other enhancers where it acts redundantly with other TFs. However, in the absence of LIF the action of Nanog over these regulatory elements becomes dominant, particularly to promote transcription. This results in Nanog having an expanded action in the absence of LIF. Yet, its expanded repressive activity is not associated with ES cell enhancers. Rather, Nanog is required to maintain H3K27me3 at differentiation-associated genes. This is the case of the TF Otx2, whose downregulation by Nanog leads to LIF-independent self-renewal even when Esrrb is not expressed. Hence, Nanog deploys distinct molecular means to promote self-renewal and counteract differentiation: when the network is fully operative (in the presence of LIF), Nanog rewires its activity; when it is partially dismantled (in the absence of LIF), Nanog represses differentiation genes via H3K27me3. Overall, we reveal different modes and the varied logic employed by Nanog to orchestrate the three main features associated with self-renewal: the inter-dependencies between pluripotency TFs, LIF signalling, and bivalent chromatin domains.

Results

Inducible CRISPR-ON ES cells efficiently activate Nanog transcription. The SunTag system was developed as a versatile tool to either visualise specific molecules in live cells or to perform epigenome editing of endogenous loci when coupled to an enzymatically inert dCas9 (**Tanenbaum et al., 2014**). It involves the expression of diffusible antibodies (scFv) that interact with high affinity with 10 copies of the GCN4 epitope linked to an enzymatically inert Cas9 (dCas9). These scFv antibodies are fused to GFP and the potent activator VP64, such that upon expression of a gRNA targeting a given genomic region, several VP64 molecules are brought about with high efficiency and specificity. To provide increased flexibility to the system, and facilitate the generation of cell lines carrying an inducible CRISPR-ON system, we engineered a single vector expressing the two SunTag moieties under the control of a Tetracycline Responsive Element. Moreover, dCas9 is linked to BFP and HpH through P2A and IRES sequences, respectively (Fig. S1A). Hence, upon induction of the system with Doxycycline (Dox), the cells are expected to become green, blue, and Hygromycin-resistant, providing a high tractability. This vector was introduced in ES cells together with the rtTA activator: two clones (C1 and C2) showing a high percentage of green/blue cells upon Dox treatment and a strong induction of dCas9 and VP64 (Fig. S1B, C), were selected. They both self-renew normally and differentiate in the absence of LIF; their karyotypes are also normal (Fig. S1D).

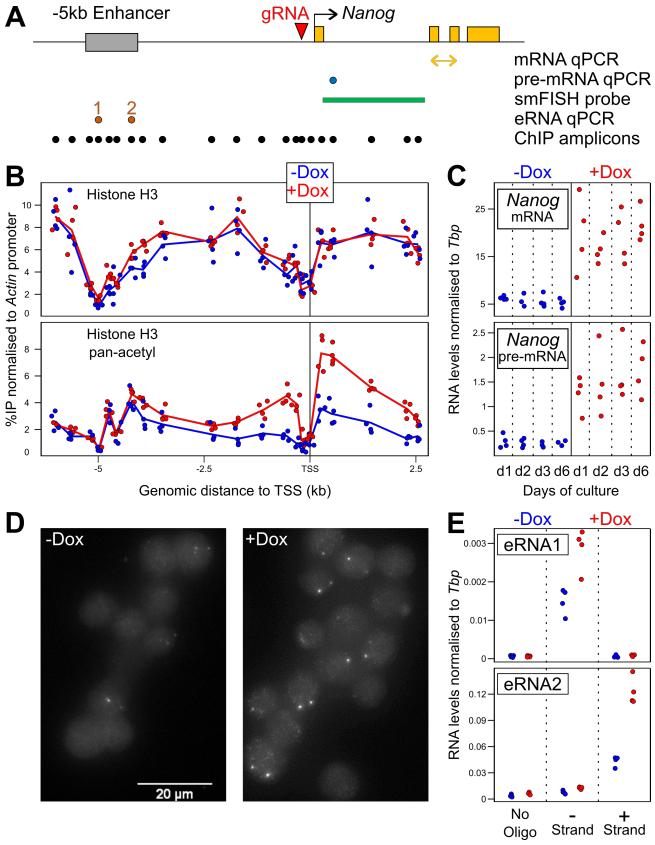


Fig. 1. CRISPR-ON ES cells for Dox-inducible activation of endogenous *Nanog*. **(A)** Schematic representation of the *Nanog* locus (black arrow: promoter; yellow boxes: exons; grey box: *Nanog* enhancer; red arrowhead: gRNA). Below, the position of the amplicons and probe used for the assays indicated on the right is shown. **(B)** ChIP across the *Nanog* locus monitoring total histone H3 (top) and pan-acetyl H3 (bottom) in the absence (blue) and after 72h of Dox treatment (red). Each dot represents normalised %IP measured in individual replicates and lines the averages. **(C)** Normalised levels of Nanog mRNA (top) and pre-mRNA (bottom) after the indicated number of days in the absence (blue) and the presence (red) of Dox. Each dot represents measurements in individual replicates as in (B). **(D)** Representative smFISH image using intronic *Nanog* probes before and after 72h of Dox induction. **(E)** Normalised levels of eRNA production from the -5kb enhancer presented as in (C). In all panels, n=4; 2 with each independent SunTag clone.

Next, we introduced to C1 and C2 a vector expressing a gRNA targeting the minimal Nanog promoter and validated binding of dCas9/VP64 with good specificity and inducibility (Fig. S1E). This was accompanied by increased histone H3 acetylation around the promoter (Fig. 1A, B), as expected given the ability of VP64 to recruit histone acetyltransferases (**Utley et al., 1998**), in the context of presumably unaltered nucleosomal organisation as evaluated by total H3 analysis (Fig. 1B). We also assessed *Nanog* expression over the course of 6 days of induction, and showed that both Nanog pre- and mRNA were induced from day 1 onwards

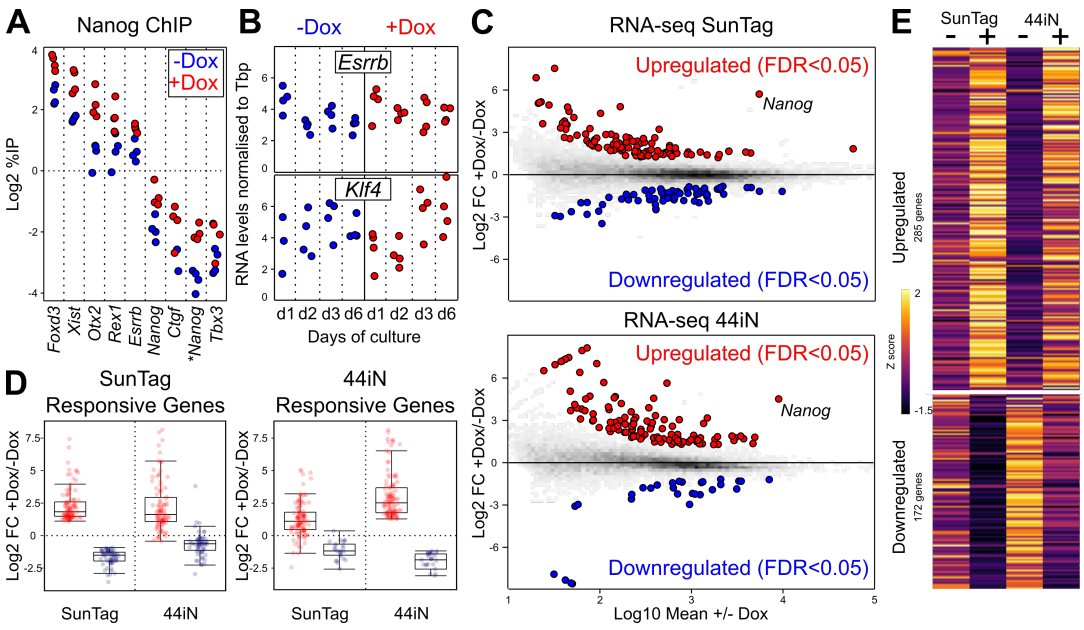


Fig. 2. Identification of Nanog-responsive genes. (A) ChIP analysis of Nanog binding across a set of targets, as indicated on the X-axis. Each dot represents measurements in individual replicates (n=4; 2 for each independent SunTag clone) **(B)** Normalised levels of *Esrrb* (top) and *Klf4* (bottom) mRNA after the indicated number of days in the absence/presence of Dox. Each dot represents measurements in individual replicates as in (A). **(C)** MA Plots displaying log2 fold changes as indicated on the Y-axis as a function of average expression. RNA-seq was performed in both SunTag (72h Dox induction) and 44iN cells (24h Dox withdrawal). Red and blue represent differentially expressed genes (FDR<0.05). **(D)** Boxplot of the log2 fold change for the genes identified in (C) as upregulated (red) or downregulated (blue) by Nanog, in either SunTag (left) or 44iN cells (right), measured in both inducible systems as indicated on the X-axis. **(E)** Heat map representing gene expression z scores of all transcripts identified by combining SunTag (FDR < 0.05), 44iN (FDR <0.05) and SunTag/44iN likelihood ratio test (FDR < 0.05) datasets.

(Fig. 1C), leading to an increase of Nanog protein levels (Fig. S1C). We also found that the increase of *Nanog* expression was due both to stronger and more frequent transcriptional bursts (Fig. 1D and Fig. S1F, G). Finally, we analysed the effects of Dox administration at the proximal -5kb enhancer of *Nanog*: upon induction, we found both sense and anti-sense enhancer transcription to be increased (Fig. 1E). Whether this is due to the proximity of these two regulatory elements or to a functional influence of the promoter on the enhancer, remains to be determined. In conclusion, we have generated Dox-inducible SunTag ES cells to activate endogenous promoters and dissect the subsequent consequences.

Definition of Nanog responsive genes. Upon Dox induction of *Nanog* in our SunTag cells we observed a 2-fold increase of Nanog binding to a panel of regulatory elements displaying a wide range of enrichment levels (Fig. 2A). This suggests that Dox induction may lead to functional consequences. However, the two main targets of Nanog that have been previously identified, *Esrrb* and *Klf4* (**Festuccia et al., 2012**), did not show any variation in expression levels over the course of 6 days of endogenous *Nanog* induction (Fig. 2B). Prompted by this unexpected observation, we performed RNA-seq to comprehensively study the global response to Dox treatment. We found a small number (163) of transcripts that were either up- or down-regulated (Fig. 2C top); neither *Klf4* nor *Esrrb* were among the induced genes (Table S1 and Fig. S2A, B). Nevertheless, the vast majority of genes that have been previously identified as responding to Nanog levels (**Nishiyama et al., 2009; Festuccia et al., 2012**), do exhibit the appropriate expression changes in our SunTag cells (Fig. S2C). To further validate our list of Nanog-responsive genes, we performed a complementary analysis using previously established *Nanog*-null cells (44iN) expressing a Dox-inducible *Nanog* transgene (**Festuccia et al., 2012**). The cells were grown in the continuous presence of Dox, which was then removed for 24 hours leading to a nearly complete loss of *Nanog* expression (Fig. 2C bottom). The number of responsive genes (141) observed with this strategy was also small (Fig. 2C and Table S1); they intersected with excellent statistical significance with the genes identified in the SunTag cells (p<1e-53). Moreover, we found the expression of genes significantly regulated in only one system, to nevertheless display highly coherent expression changes in the other system (Fig. 2D). Hence, to improve statistical power and expand on Nanog targets we combined the SunTag and 44iN datasets to test for those genes with coherent Nanog response across both systems (Fig. 2E and Fig. S2D). Combining with those genes already identified, this resulted in 457 genes (Fig. 2E and Fig. S2D), which generally display extremes of expression differences between Dox-treated SunTag (high *Nanog*) and untreated 44iN cells (low/absent *Nanog*); they globally behave in a concordant way when long-term *Nanog*-null cells are compared to wild-type cells (Fig. S2D), or when their expression is analysed in published datasets (Fig. S2E). Genes activated by Nanog are enriched in regulators of stem cell maintenance (FDR<2.99e-16), while repressed genes are enriched in differentiation processes such as nervous system development (FDR<6.89e-10). Hence, we have defined the compendium of Nanog-responsive genes in undifferentiated ES cells with unprecedented completion. Strikingly, neither *Klf4* nor *Esrrb* belong to our list of Nanog-responsive genes.

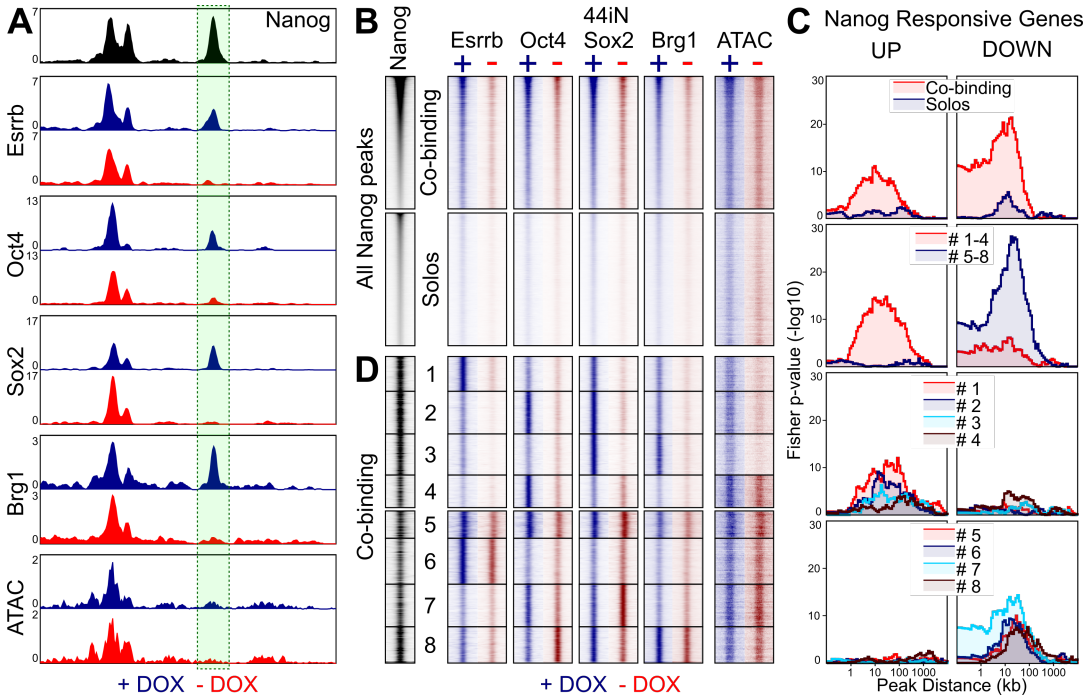


Fig. 3. Nanog rewires the pluripotency network. (A) Representative enrichment profiles (reads per million) of the indicated TFs (ChIP-seq) and chromatin accessibility (ATAC-seq) across an 8kb-long genomic region (mm9: chr10:33,952,000-33,960,000). In black, average signal from 6 publicly available Nanog datasets. In blue and red, average signal of 44iN cells cultured in the presence or absence of Dox, respectively. The peak highlighted by a green box shows reduced signal in the absence of Dox; note an upstream peak displays increased Sox2 binding. **(B)** Heat map of average enrichment levels in each indicated condition (+/-Dox) across 0.5kb centred on the Nanog peak summit at all Nanog sites (27,782), ranked from high to low Nanog. The regions were split in two groups depending on the presence (co-binding) or absence (Solos) of other TFs. **(C)** Plots displaying the -log10 Fisher p-value (Y-axis) of the enrichment of genes up- (left) or downregulated (right), at a given distance (X-axis) from specific groups of Nanog binding sites identified in (B, D), as indicated in the colour-coded insets. **(D)** Heat map corresponding to 8 clusters identified on the basis of TF co-binding and the effects of Dox withdrawal, presented as in (B) without ranking for Nanog binding levels.

Nanog rewires the pluripotency network to control its targets. Having established a list of Nanog-responsive genes, we aimed at exploring the mechanisms by which Nanog influences their expression, focusing on a potential role of Nanog in modulating binding of other regulators such as pluripotency TFs (Oct4, Sox2 and Esrrb) and the chromatin remodeller Brg1 that is functionally associated with self-renewal (**Ho et al., 2009; Kidder et al., 2009**). To do this, we first established a list of 27,782 regulatory elements bound by Nanog using six datasets derived from four independent published studies (Table S2) and used 44iN cells to address how Nanog impacts TF binding and chromatin accessibility at these sites. We noticed that at some Nanog binding regions, Esrrb, Oct4, Sox2, and Brg1, display a strong reduction of binding and decreased chromatin accessibility, after 24 hours of Dox withdrawal (Fig. 3A). This observation can be generalised to a large proportion of regions and is particularly prominent in the case of Esrrb (Fig. 3B). We then divided Nanog binding regions in two major groups Fig. 3B and Fig. S3, based on the presence of other TFs (regions of co-binding) or not (Nanog-solo regions). In Nanog-solo regions, which display lower levels of Nanog binding, the chromatin is less accessible irrespective of Nanog (Fig. 3B and Fig. S3), indicating that additional factors may be recruited. Strikingly, when we computed the number of Nanog-responsive genes as a function of the distance to Nanog binding regions, we observed that both activated and downregulated genes are particularly enriched in the vicinity of co-binding regions and not of Nanog-solos (Fig. 3C). Moreover, while activated genes tend to be located distally (within 10 to 100kb), downregulated genes also show a significant enrichment over closer distances (<10kb). To further explore the relationships between Nanog and other TFs, we used k-means clustering to identify 8 subgroups of Nanog binding sites (Fig. 3D and Fig. S3). In the first 4 clusters, the depletion of Nanog leads to an acute loss of TF binding; collectively these regions are strongly associated with the activation of Nanog targets (Fig. 3C). In contrast, clusters 5-8 are significantly associated with genes repressed by Nanog and the effects of its depletion are more nuanced (Fig. 3C, D). More specifically, clusters 1 to 3 display a nearly total loss of TF binding in the absence of Nanog, along with a marked decrease in chromatin accessibility (Fig. 3D and Fig. S3). These 3 clusters, in particular clusters 1 and 2, are associated with genes activated by Nanog (Fig. 3C). At cluster 4, however, chromatin accessibility shows minimal variations and, while very strong Oct4 binding is nearly completely lost upon Nanog depletion, Sox2 is not affected. Since Brg1 is particularly low across this cluster, Sox2 may recruit other chromatin remodellers to render the chromatin accessible at these regions, in a Nanog-independent manner. Accordingly, the correlation with Nanog-responsive genes of cluster 4 is weaker (Fig. 3C). Overall, at more than 6000 regions (clusters 1 to 3), Nanog plays a chief role in establishing functional and accessible regulatory regions capable of recruiting different combinations of TFs to activate its targets. Con-

versely, at clusters 5 to 8, the effects of the loss of Nanog are rather small both at the level of TF binding and of chromatin accessibility (Fig. 3D and Fig. S3). This suggests that Nanog-mediated repression uses radically different mechanisms, which are not based on the increased recruitment of Esrrb, Oct4 and Sox2. Rather, clusters 7 and 8 display increased Oct4 and Sox2 binding in the absence of Nanog, respectively, suggesting that Nanog downregulates the genes functionally linked to these two clusters by blocking Oct4 or Sox2 recruitment (Fig. 3D and Fig. S3). At other enhancers associated with genes repressed by Nanog, showing no alteration of Oct4 and Sox2 occupancy (clusters 5 and 6), Nanog may block the otherwise activatory function of other TFs. In conclusion, Nanog wires the pluripotency network by fostering TF recruitment and chromatin accessibility at distal regulatory elements to act as an activator, and uses different mechanisms, including the impairment of Oct4/Sox2 recruitment, both at promoter-proximal and distal regulatory elements of the genes it represses.

Nanog confers LIF-independent self-renewal in the absence of *Esrrb* induction. The strong influence of Nanog on the efficiency of self-renewal has been proposed to be largely mediated by Esrrb (Festuccia et al., 2012). Therefore, we were not expecting our SunTag cells endogenously activating *Nanog* to exhibit increased self-renewal capacity, given that *Esrrb* and other genes involved in self-renewal, such as *Klf4* (Niwa et al., 2009), are not strongly induced (Fig. 2B). To test this, we initially plated our SunTag lines at clonal density together with the parental controls lacking the *Nanog* gRNA, and cultured them in the presence or absence of Dox for 6 days (Fig. 4A). In the presence of LIF, we could not observe any major change in the efficiency of self-renewal. In contrast, in the absence of LIF, when virtually all the colonies display complete or partial signs of differentiation in all controls, cells with enhanced endogenous *Nanog* expression generated a substantial proportion of undifferentiated colonies (Fig. 4A, B). To further validate that these cells are bona-fide ES cells, we harvested them at the end of the clonal assay and performed two complementary assays. First, we re-plated them in 2i medium lacking serum (**Ying et al., 2008**), where only truly undifferentiated cells proliferate: both clones gave rise to typical spherical and undifferentiated colonies (Fig. 4C). Second, we re-plated them at clonal density in the absence of LIF and the presence/absence of Dox: only in the presence of Dox did we recover undifferentiated colonies; in the absence, all the cells differentiated (see below). This demonstrates that the exposure to Dox does not alter the differentiation capacity of our cell lines upon its withdrawal. We conclude, therefore, that Dox-induction of *Nanog* confers to our SunTag lines the ability to self-renew in the absence of LIF, a definitive proof of the efficiency of our CRISPR-ON strategy to study Nanog function. Strikingly, LIF-independent self-renewal was attained in the absence of any apparent induction of Klf4 and Esrrb mRNAs (Fig. S4A) or proteins (Fig. 4D). Therefore, to explore both the magnitude of the differentiation blockade at the molecular level, and to identify potential Klf4/Esrrb-independent mechanisms underlying Nanog-mediated self-renewal, we performed transcriptomic analyses. In control cells that were not stimulated by Dox, a large number of genes responded to LIF withdrawal (>5000) and exhibited important quantitative differences (Fig. S4B, C, D). In the presence of Dox, the magnitude of the expression changes of these LIF-responsive genes was globally diminished (Fig. S4B), even though the vast majority of pluripotency genes remained strongly downregulated (Fig. S4C, D). In fact, not all genes that respond to LIF withdrawal were rescued by *Nanog* induction to the same extent, with only around 20% being efficiently rescued (Fig. 4E, Fig. S4E and Table S1). This argues against the idea that the presence of substantial numbers of undifferentiated cells may explain all the expression changes measured upon Dox induction in the absence of LIF.

Increased regulatory potential of Nanog in the absence of LIF, due to a loss of redundancy. In the absence of LIF, the effects of endogenous Nanog induction are largely maximised: if in the presence of LIF we identified 285 up and 172 downregulated genes, in its absence these numbers raised to 856 and 589, respectively (Table S1). It appears, therefore, that LIF signalling attenuates the relative impact of Nanog on the ES cell transcriptome. To explore this further, we established the associations between the clusters of Nanog binding regions we have identified (Fig. 3), with four groups of genes: genes down- or upregulated upon LIF withdrawal and, among these two categories, those that Nanog can or cannot partially rescue by activating or repressing them, respectively (Fig. 4E). We observe that only one group, constituted of genes repressed by Nanog in the absence of LIF, is not enriched in any Nanog binding region that we have studied (Fig. 4F and Fig. S4F). In contrast, the group of genes downregulated upon LIF withdrawal, and rescued by Nanog, is similarly enriched for Nanog binding regions where Nanog leads the recruitment of other TFs (clusters 1 to 4) than for those where it does not (clusters 5 to 8; Fig. 4F and Fig. S4F). The activatory potential of Nanog through clusters 1 to 4 was already established from the previous analysis. However, the vast majority of genes upregulated by Nanog in the absence of LIF were not activated in LIF containing medium (>700 genes, Table S1). This suggests that in the presence of LIF, these genes are redundantly controlled by other LIF-dependent TFs, either through regions belonging to clusters 5 to 8 or through other regulatory elements where Nanog does not bind. Since clusters 5 to 8 were previously associated with genes repressed by Nanog in the presence of LIF, their enrichment in the vicinity of genes activated by Nanog in the absence of LIF implies that they are constituted by at least two functional categories: enhancers that are blocked by Nanog in the presence of LIF, leading to the downregulation of Nanog targets, and enhancers where Nanog also acts as an activator but redundantly to other LIF-dependent factors, most likely Esrrb. In agreement, this group is strongly enriched in genes of the pluripotency network (Fig. 4G), which are known to be controlled by several pluripotency TFs and from multiple distinct enhancers. The level of upregulation of these Nanog-activated genes is, however, relatively minor

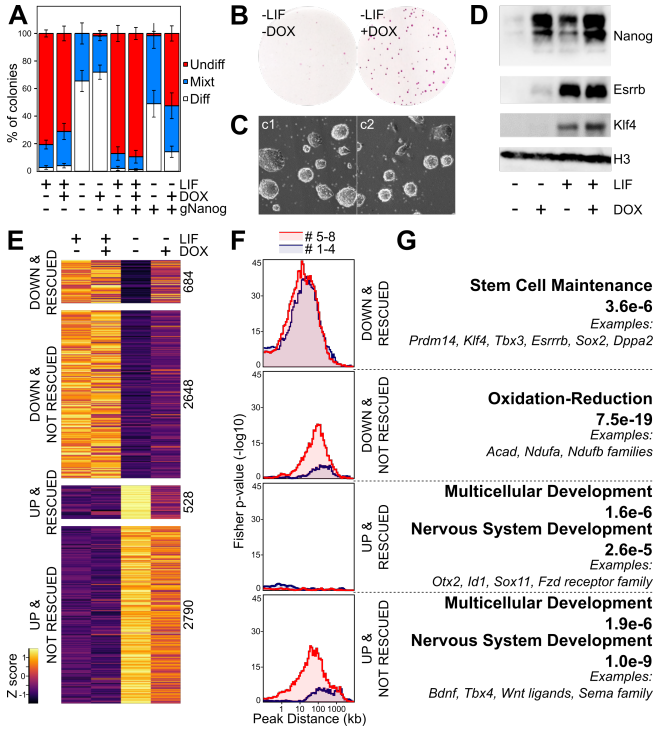


Fig. 4. Endogenous induction of *Nanog* mediates LIF-independent self-renewal by non-canonical mechanisms. (A) Histogram representing the percentage of undifferentiated (red), mixt (blue) and differentiated (white) colonies (Y-axis) counted after 6 days of clonal growth in the indicated conditions (X-axis). Error bars represent std. dev. (n=4). **(B)** Representative image of Alkaline Phosphatase stained colonies after 6 days of clonal growth in the presence/absence of Dox. **(C)** Representative image of the two SunTag clones after culturing them in 2i for 3 days following a 6-day clonal assay in the absence of LIF and the presence of Dox. **(D)** Representative Western-Blot of Nanog, Esrrb and Klf4 after 3 days of culture in the indicated conditions. **(E)** Heat map representing gene expression z scores across 4 groups of transcripts, as indicated in the left. UP/DOWN refers to their expression changes after 3 days of LIF deprivation (FDR<0.05); rescued versus not rescued indicates whether Nanog significantly alleviates their misregulation (FDR<0.05). **(F)** Cumulative plots displaying the -log10 Fisher p-value (Y-axis) of the enrichment of genes belonging to the groups identified in (E) at a given distance (X-axis) from Nanog binding clusters 1 to 4 (red) and 5 to 8 (blue), as established in Fig. 3. **(G)** Representative Gene Ontology terms enriched in each group of genes. The FDR is indicated together with selected examples.

and the rescue of pluripotent TFs, including *Esrrb* and *Klf4*, marginal (Fig. 4D and Fig. S4A, C). Finally, LIF-responsive genes that are not rescued by Nanog are associated with clusters 5 to 8 (Fig. 4F). This indicates that a large number of regulatory regions of the pluripotency network for which Nanog has a modest functional impact are present within these clusters. These regions activate or repress genes prior to differentiation in a Nanog-independent manner; upon LIF-withdrawal, their activity is likely invalidated with the ensuing consequences on gene expression even when *Nanog* is induced. These Nanog-independent, LIF-responsive genes, are closely associated with cluster 6 (Fig. S4F), which is dominated by Esrrb (Fig. 3D and Fig. S3), a prominent LIF target. This is particularly true for genes downregulated upon LIF withdrawal; satisfactorily, given the known role of Esrrb as a general regulator of metabolism and energy production (**Sone et al., 2017**), these genes are enriched for related terms such as oxidation-reduction (Fig. 4G). Notably, the inability of Nanog to rescue these genes further supports the notion that our SunTag cells have acquired LIF-independent self-renewal in the absence of functional Esrrb. In conclusion, these analyses underscore the complexity of the pluripotency network: whilst Nanog activates genes both in the presence and absence of LIF through regulatory regions belonging to clusters 1 to 4, at other enhancers the activation of Nanog can only be unmasked in the absence of LIF, when other pluripotency TFs are downregulated and their functional redundancy with Nanog, abolished. Unexpectedly, however, the genes repressed by Nanog in the absence of LIF, which exhibit a robust rescue of higher magnitude than that observed for the genes that Nanog activates (Fig. S4E), appear largely disconnected from the Nanog binding regions that we have studied here (Fig. 4F and Fig. S4F). These genes are associated, among other categories, with signalling and molecular pathways linked to differentiation (Fig. 4G). Hence, the probably indirect repression mediated by Nanog over these genes may underlie LIF-independent self-renewal in Dox-treated SunTag cells, despite the lack of Esrrb and Klf4.

Nanog sustains H3K27me3 to maintain silent differentiation genes upon LIF withdrawal. Gene set enrichment analysis (Fig. S5A) indicated that Nanog-rescued genes that are normally upregulated upon LIF withdrawal are enriched for targets of Polycomb Group proteins and for one of the marks they deposit to trigger facultative heterochromatin, H3K27me3 (FDR<6e-43). Hence, we profiled H3K27me3 in SunTag cells grown in the presence/absence of LIF and Dox. Overall, the patterns of H3K27me3 were found similar among all conditions, with notable exceptions (Fig. 5A). We identified three broad classes of H3K27me3 domains: those with high levels across all conditions and those that show either a loss or a gain of H3K27me3 upon LIF withdrawal (Fig. 5B and Table S3). Strikingly, the regions losing H3K27me3 in the absence of LIF maintained significant levels when endogenous Nanog expression was induced with Dox (Fig. 5B). In the presence of LIF, however, the induction of Nanog had minor consequences on H3K27me3, if any. This indicates that Nanog and LIF use parallel pathways to maintain H3K27me3 at a subset of H3K27me3 domains, and suggests that Nanog may confer LIF-independent self-renewal by sustaining H3K27me3 at these regions. Notably, the genes upregulated upon LIF withdrawal display differential enrichment among these three classes of H3K27me3 domains, depending on their Nanog-responsiveness: while nearly 70% of Nanog-rescued genes are marked by H3K27me3, which tends to decrease upon LIF-withdrawal except when Nanog is induced, only 30% of genes that are not rescued by Nanog show a similar pattern (Fig. S5B). This confirms that Nanog-rescued genes are particularly enriched in LIF-dependent H3K27me3. More specifically, H3K27me3 concentrates around the promoters of these genes, and displays a reduction in levels upon LIF withdrawal, exclusively in the absence of *Nanog* induction (Fig. 5C). Hence, Nanog stimulates the maintenance of H3K27me3 at a large subset of the promoters it represses in the absence of LIF. Nevertheless, a third of the genes repressed by Nanog in the absence of LIF is not embedded within H3K27me3; conversely, a third of the genes that are upregulated upon LIF withdrawal regardless of

Nanog expression are enriched in H3K27me3 (Fig. S5B) and maintain higher levels around their promoters in the presence of Nanog (Fig. 5C). Thus, we determined whether quantitative differences regarding the effect of Nanog over these groups could be measured. Among the genes repressed by Nanog, the higher magnitude of rescue is observed for those genes that are embedded in H3K27me3 (Fig. S5C). Similarly, even though their gene expression changes upon Dox induction were not statistically significant, within the group of genes not rescued by Nanog, those enriched in H3K27me3 show a clear tendency to be downregulated (Fig. S5C). A clear, global pattern can be inferred from these analyses: the ability of Nanog to rescue genes that are upregulated upon LIF withdrawal is directly correlated with H3K27me3 levels. Accordingly, ordering the heatmap of genes upregulated upon LIF withdrawal (regardless of the ability of Nanog to rescue them) by their enrichment levels for H3K27me3 in the presence of LIF and the absence of Dox, naturally orders the genes from efficient to poor rescue (Fig. 5D). Hence, H3K27me3 levels before LIF withdrawal are highly predictive of the efficiency of Nanog to block gene upregulation during differentiation. Overall, these analyses indicate that in the absence of LIF, Nanog mediates its repressive function by other means than those described in its presence (Fig. 3): by maintaining high levels of H3K27me3.

Nanog represses *Otx2* to confer increased self-renewal efficiency regardless of LIF signalling. Among the genes downregulated by Nanog in the absence of LIF are several developmental TFs such as *Sox11*, *Id1* and *Pou3f1*, among others (Table S1). Therefore, it may be possible that LIF-independent self-renewal is attained by the simultaneous inhibition of several developmental pathways. However, within the list of Nanog-repressed genes characterised by Nanog-dependent H3K27me3, we could also identify *Otx2* (Fig. 5A and Table S3), a key regulator of the earliest stages of ES cell differentiation (**Acampora et al., 2013; Buecker et al., 2014**). Several lines of evidence point to *Otx2* downregulation being an important mediator of Nanog function. First, *Otx2* has been already identified as an important negative target of Nanog (**Festuccia et al., 2012; Acampora et al., 2017**); accordingly, we observe its downregulation at the mRNA and protein levels upon *Nanog* induction (Fig. S5D, E). Moreover, further expression analyses indicate *Otx2* expression is closely controlled by Nanog levels (Fig. S5F). Second, the genes that are upregulated in the absence of LIF and that are rescued by Nanog, are enriched in genes activated by Otx2, while non-rescued genes are not (Fig. S5A). Third, the ectopic expression of Otx2 drives ES cells into differentiation (**Buecker et al., 2014; Acampora et al., 2017**), even in the presence of LIF, as we show here using our SunTag system targeted to the *Otx2* promoter (Fig. S5E, G). Therefore, it may be possible that its Nanog-mediated downregulation contributes to LIF-independent self-renewal in the context of our endogenous *Nanog* activation, despite the lack of strong upregulation of *Klf4* and *Esrrb*. To test this, we exploited the flexibility of the SunTag system to simultaneously activate *Nanog* and *Otx2* and perform clonal assays. Upon the

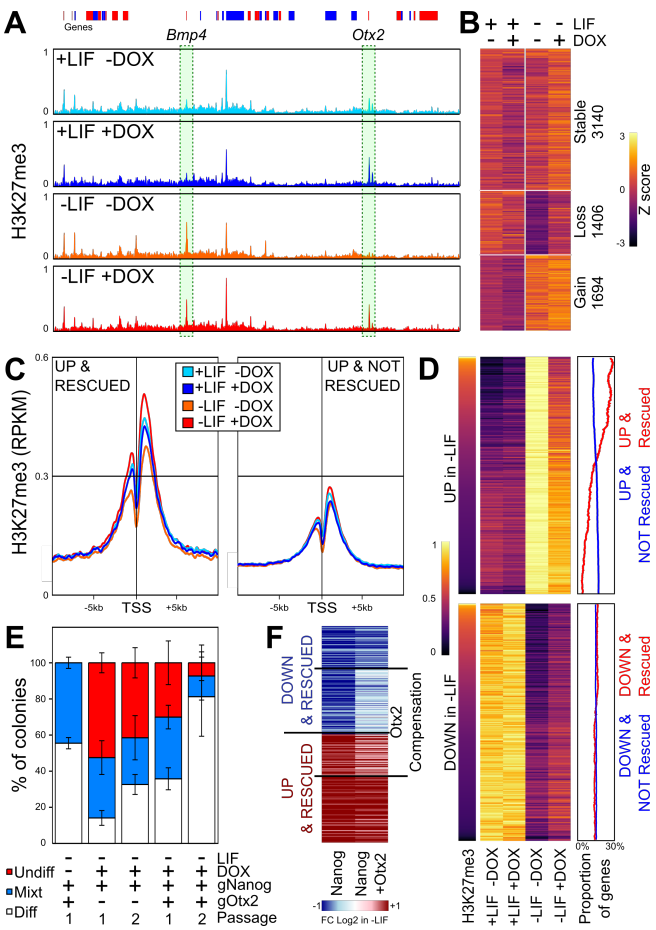


Fig. 5. A *Nanog*/H3K27me3/*Otx2* axis controls LIF-independent self-renewal. (A) H3K27me3 average levels (reads per million) across 5.15Mb (mm9: chr14:45346123-50415313) encompassing the *Bmp4* and *Otx2* genes that display the two stereotypical behaviors observed in SunTag cells cultured in the presence/absence of LIF/Dox for 3 days. **(B)** Heat map of H3K27me3 z scores of 6,240 H3K27me3 domains identified in SunTag cells cultured as indicated. **(C)** Average H3K27me3 (reads per million) across promoter regions of genes upregulated in the absence of LIF and either rescued (left) or not (right) by Nanog. **(D)** Correlative analysis of ranked H3K27me3 (left) with gene expression changes (arbitrary units; middle) for transcripts upregulated (top) or downregulated (bottom) in SunTag cells cultured in the presence/absence of LIF/Dox for 3 days. The percentage of genes belonging to rescued (red line) versus non-rescued genes (blue line) in a 500 gene sliding window across the regions of the heat map is shown on the right. **(E)** Histogram representing the percentage of undifferentiated (red), mixt (blue) and differentiated (white) colonies (Y-axis) counted after 6 (Passage 1) or 12 (Passage 2) days of clonal growth in the indicated conditions (X-axis). Error bars represent std. dev. (n=4). **(F)** Heat map of the average log2 fold change in gene expression between SunTag cells activating either *Nanog* or *Nanog* and *Otx2*, grown in the absence of LIF and the presence/absence of Dox for 3 days. Genes for which the simultaneous activation of *Otx2* compensates the changes observed when only *Nanog* is induced (FDR<0.05), are highlighted.

additional induction of *Otx2*, the proportion of undifferentiated colonies decreased in the absence of LIF, compared to cells activating *Nanog* only, in particular after two successive rounds of clonal growth (Fig. 5E). These results clearly place *Otx2* as a key factor that needs to be repressed by Nanog in order to obtain efficient LIF-independent self-renewal. Next, we performed transcriptomic analyses upon *Nanog*/*Otx2* induction in the absence of LIF to identify the set of genes that were effectively compensated by the action of Otx2. We observed that around 40% of the genes repressed by Nanog, and 70% of the genes activated by Nanog, displayed similar

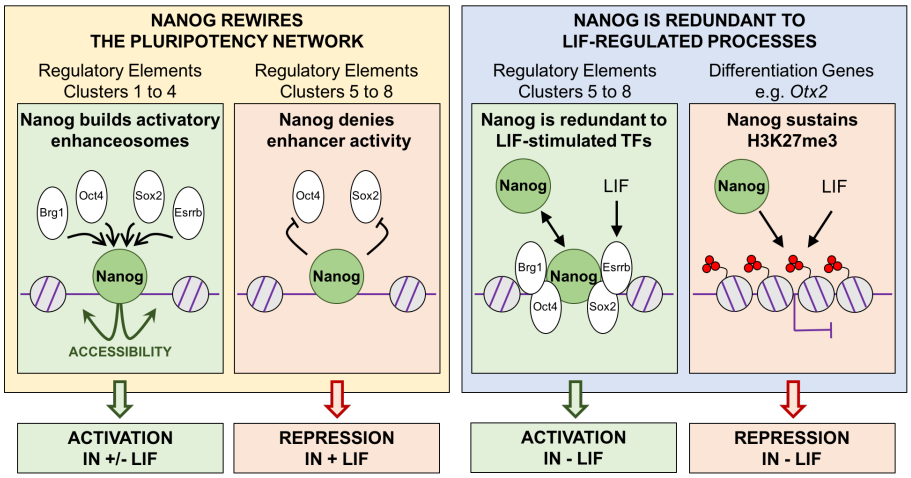


Fig. 6. Nanog is a versatile TF impacting the pluripotency network and epigenome. The function of Nanog at stereotypical clusters of regulatory elements targeted by the pluripotency network, as well as at differentiation genes, is shown. Briefly, Nanog displays four major behaviours (left to right): 1/ recruitment of other factors (Oct4, Sox2 and Esrrb, together with Brg1) to promote chromatin accessibility and activate gene transcription; 2/ inhibiting enhancer activity, leading to gene repression either by blocking Oct4/Sox2 recruitment (shown) or by other mechanisms (not shown for simplicity; see text for details); 3/ complementing enhancer activity redundantly with other factors which are controlled by LIF (such as Esrrb) – in this case, its activatory role can only be appreciated in the absence of LIF; 4/ Nanog and LIF act in parallel to sustain H3K27me3 at differentiation genes such as Otx2. This latter role of Nanog is particularly important in the context of Nanog-mediated, LIF-independent self-renewal.

levels to control cells grown in the absence of LIF and Dox, when both *Otx2* and *Nanog* were induced (Fig. 5F). Hence, at the molecular level, *Otx2* induction partially compensates the gene expression changes induced by Nanog overexpression in the absence of LIF, underscoring the antagonistic effect of Nanog and Otx2 over a large set of common genes (**Acampora et al., 2017**), including developmental TFs such as *Sox3*, *Sox11*, *Id1* and *Pou3f1*, among others (Table S1), which tend to be expressed in somatic cells and more particularly in neuroectodermal derivatives. In combination with the previous section, these results indicate strongly that Nanog controls H3K27me3 at key nodes in the differentiation network, such as *Otx2*, to indirectly repress a large set of genes involved with differentiation.

Discussion

Nanog rewires the pluripotency gene regulatory network. Gene regulatory networks constituted of master TFs are characterised by the capacity of individual factors to act over the same sets of regulatory elements, which together define and specify the molecular and transcriptional identity of each cell type (**Shlyueva et al., 2014; Spitz and Furlong, 2012**). However, we still have a relatively poor understanding of how single TFs impact globally on the recruitment of other members of a given network to impact its activity. Recently, the role of Oct4 has been suggested to rely on its ability to recruit Brg1 to render the chromatin accessible for other TFs to bind (**King and Klose, 2017**), matching a subset of the mechanisms we propose here for Nanog. However, it is unclear how much the initiation of differentiation that follows Oct4 depletion (**Niwa et al., 2000**) influenced the interpretations regarding how Oct4 directly impinges upon the pluripotency network. In our case, we have focused on Nanog, a factor that can be depleted from ES cells while preserving pluripotency (**Chambers et al., 2007**). Hence, it is likely that the rewiring of the network that we observe shortly after depleting Nanog, is due to primary and direct effects. Strikingly, our analyses suggest that the simplified view positing that pluripotency TFs bind cooperatively at regulatory elements to collectively control transcription, may need to be partially revisited: at least from the perspective of Nanog, the combinations of binding, their dependencies on Nanog, and their association to responsive genes, are more complex than we had previously anticipated (Fig. 6).

Although we observe, as expected, that Nanog-bound regions where other pluripotency TFs are also recruited are more strongly associated with Nanog-responsive genes compared to regions where Nanog binds alone, we also find that only a small subset displays similarly high binding levels of all three pluripotency TFs that we tested (Esrrb, Oct4 and Sox2). Indeed, the binding of one or two factors, in addition to Nanog, tends to dominate the others. This is valid even for Oct4/Sox2, which are believed to bind together at Oct4/Sox2 chimeric motifs (**Chew et al., 2005**). Therefore, different stoechiometries and/or residence times of individual factors seem to apply at distinct sets of regions. This produces a level of complexity that surpasses a simple model in which all factors bind together at key enhancers. Moreover, the effect of Nanog over these factors is also highly variable, with two clear groups of regions being easily identifiable: those in which Nanog plays a leading role; those where it does not. Strikingly, these two groups of regions display sharp distinctions regarding their association to Nanog-responsive genes. Regions where Nanog preserves chromatin accessibility and drives the recruitment of other pluripotency TFs and Brg1, are strongly biased to genes activated by Nanog. Conversely, the regions where Nanog does not promote TF binding or accessibility are associated either with genes repressed by Nanog in the presence of LIF, or with genes where Nanog acts as a redundant activator with LIF-stimulated TFs (Fig. 6).

To activate its targets, Nanog seems to use an expected mechanism essentially based on establishing a permissive chromatin architecture associated with the recruitment of other TFs and the formation of functional regulatory complexes. In contrast, while some precedents already pointed in the past into the direction of Oct4/Sox2-independent Nanog repression (**Navarro et al., 2008, 2012**), it is remarkable that Nanog-mediated repression is so strongly associated with regulatory elements where Nanog does not facilitate Esrrb,

Oct4 and Sox2 binding and the chromatin remains equally accessible irrespectively of Nanog. This differential association between gene activation/repression and the role of Nanog as a nucleation factor of functional complexes, suggests that Nanog displays cooperative binding primarily to promote gene expression. At repressive sites, Nanog may either turn the whole complex into repressive or inhibit its otherwise transactivation potential. Additionally, at a large number of regions, the loss of Nanog leads to increased binding of either Oct4 or Sox2 (Fig. 6). It is therefore questionable that Nanog/Oct4 and Nanog/Sox2 bind at the same time over the same DNA molecules, at least over these regions. Thus, caution must be taken when extrapolating molecular functions from generic binding profiles: even though Nanog/Oct4 and Nanog/Sox2 appear to bind together, the binding of Nanog is in fact detrimental to that of the other two factors. Remarkably, these regions are closely associated with genes repressed by Nanog, indicating that to repress its targets Nanog interferes with the binding or the activity of other pluripotency TFs. Whether the alternate behaviours of Nanog to activate or repress transcription (Fig. 6) represents a general rule or a specific property of Nanog should be thoroughly investigated.

Nanog-dependent H3K27me3 and *Otx2* repression, an alternative route for self-renewal. The ability of Nanog to block differentiation and promote LIF-independent self-renewal, something that not every pluripotent TF, including Oct4 and Sox2, is capable of doing, represents a defining property of Nanog (**Chambers et al., 2003**). However, this is not a unique characteristic of Nanog: a plethora of additional TFs, exemplified by Klf4 and Esrrb, have been progressively identified and demonstrated to provide LIF-independent self-renewal (**Niwa et al., 2009; Festuccia et al., 2012**). Hence, over the years, the importance of Nanog has been somehow equilibrated with that of other TFs, most notably Esrrb, which can replace Nanog in several contexts (**Festuccia et al., 2012; Zhang et al., 2018**). More strikingly, Esrrb has been proposed to be an obligatory mediator of the promotion of LIF-independent self-renewal by Nanog (**Festuccia et al., 2012**). Therefore, our observation of LIF-independent self-renewal in the absence of strong Esrrb upregulation, and of many other pluripotency TFs, is particularly enthralling. This is not the first time, however, that the role of a TF within the pluripotency network needs to be nuanced. Nanog itself was initially thought to be strictly required for germ cell development (**Chambers et al., 2007**) and for the induction of pluripotency via somatic cell reprogramming (**Silva et al., 2009**), conclusions that were subsequently invalidated (**Carter et al., 2014; Schwarz et al., 2014; Zhang et al., 2018**). Moreover, major experimental differences between our and previous studies may underlie the different conclusions regarding the mandatory requirement for Esrrb. Indeed, while others ectopically expressed Nanog constitutively and at high levels in *Esrrb* knock-out cells (**Festuccia et al., 2012**), we have used an inducible CRISPR-ON system to activate endogenous *Nanog* concomitantly with LIF withdrawal and the ensuing progressive downregulation of *Esrrb*. Therefore, the dynamic aspects of the two experimental setups are drastically different: it is possible that *Esrrb* is downregulated after Nanog has already impacted on other genes, which may independently block differentiation even when *Esrrb* is subsequently silenced. Besides this, however, our findings and their confrontation to previous conclusions highlight that different factors and mechanisms can potentially lead to similar phenotypic outcomes (**Konstantinides et al., 2018**).

Identifying the genes that mediate LIF-independent self-renewal in the absence of Esrrb may be particularly challenging because several prominent developmental regulators, from TFs to signalling molecules, are enriched among the genes that normally respond to LIF withdrawal but that endogenous *Nanog* induction is able to block. However, the genes that are upregulated when LIF is removed, and that Nanog is able to keep in check, display a blatant property: they tend to be targets of Polycomb PRC2 complexes and are embedded within H3K27me3 domains (**Azuara et al., 2006; Bernstein et al., 2006; Boyer et al., 2006**). At these genes, either LIF or Nanog are required to maintain H3K27me3 (Fig. 6). Therefore, this study identifies at least two modes of gene regulatory redundancy between Nanog and the LIF pathway: one directly based on LIF-stimulated TFs, such as Esrrb and Klf4, and another one based on the activity of Polycomb Group proteins (Fig. 6). We anticipate that identifying the exact molecular mechanisms used by Nanog to modulate H3K27me3 will be of great interest, and propose here that they are likely to be indirect. Overall, our observations argue for the existence of an alternative pathway to promote LIF-independent self-renewal through a previously unanticipated role of Nanog in the maintenance of H3K27me3 at differentiation-associated genes, thereby inhibiting the capacity of the cells to readily differentiate. This type of compensatory, chromatin-based mechanism, enables individual TFs to have broad impact by targeting key chromatin regulators with a more generic and systemic function. This remains so even when a regulatory network is largely dismantled, as is the case of the pluripotency network in the absence of LIF. This novel mechanism that we have unveiled may have the potential to dramatically increase the robustness and temporal integration of complex gene regulatory systems.

Whether the promotion of LIF-independent self-renewal associated with the inappropriate maintenance of H3K27me3 at differentiation genes results from the sum of many partial effects or, conversely, is based on specific and potent effects mediated by one or a few regulators, needs now specific attention. Given that the genes repressed by Nanog in the absence of LIF are strongly enriched for targets of Otx2, a driver of differentiation (**Acampora et al., 2013; Buecker et al., 2014**) that belongs to the category of bivalent genes in ES cells, we explored the possibility that *Otx2* downregulation may be the sole explanation for LIF-independent self-renewal in the absence of induced expression of other pluripotency TFs than Nanog itself. Using our SunTag cells to simultaneously activate both *Nanog* and *Otx2*, we could indeed observe that the efficiency of self-renewal was lower compared to the

exclusive induction of *Nanog*. However, the effects became robust after two continuous phases of clonal growth, indicating that, for this level of upregulation of *Otx2*, the functional compensation takes time to be fully established. Nevertheless, our transcriptomic studies after three days of induction already show that a large fraction of the genes normally misregulated upon *Nanog* induction, show expression levels similar to control cells. In accord with the developmental role of Otx2 (**Acampora et al., 1995**), this seems to be particularly true for neuroectodermal genes. It should now be investigated whether a relationship similar to Nanog-Otx2 exists between Nanog and other lineage specific determinants targeted by H3K27me3. Overall, this work underlines the ability of *Nanog* to convey its function through remarkably distinct molecular mechanisms in different environmental conditions (Fig. 6). Extrapolating our work to other TFs of the pluripotency network and more generally to other gene regulatory systems, would be of great interest.

Acknowledgements. We thank **Dr. Pentau Liu** for kindly providing PiggyBack plasmids and **Dr. Domingos Henrique** for fruitful discussions regarding the involvement of Polycomb activity in Nanog function. This work was supported by recurrent funding from the **Institut Pasteur**, the **CNRS**, and **Revive** (Investissement d'Avenir; ANR-10-LABX-73). P.N. acknowledges financial support from the **Fondation Schlumberger** (FRM FSER 2017), the **Fondation ARC** pour la recherche sur le cancer (ARC NAVARRO CA 14/12/2016), the **Agence Nationale de la Recherche** (ANR 16 CE12 0004 01 MITMAT), and the **Ligue contre le Cancer** (LNCC EL2018 NAVARRO). V.H is supported by the Fondation ARC and Revive. N.O. is supported by Revive.

Author contributions. V.H. performed the experimental work with help from **I.G.** (ChIP and ATAC-seq), **F.M.** and **P.C.** (smFISH), **C.P.** (RNA-seq), **D.M.** (gRNA design) and **A.D.** (cell culture). **N.O.** analysed sequencing data. **V.H.**, **N.O.** and **P.N.** analysed and interpreted the results. **P.N.** designed the project and wrote the manuscript with inputs from **V.H.** and **N.O.**

Declaration of interests. The authors declare no competing interests.

References

Acampora, D., Acampora, D., Mazan, S., Lallemand, Y., Avantaggiato, V., Maury, M., Simeone, A., and Brûlet, P. (1995). Forebrain and midbrain regions are deleted in Otx2-/- mutants due to a defective anterior neuroectoderm specification during gastrulation. **Development** 121, 3279–3290.

Acampora, D., Di Giovannantonio, L.G., and Simeone, A. (2013). Otx2 is an intrinsic determinant of the embryonic stem cell state and is required for transition to a stable epiblast stem cell condition. **Development** 140, 43–55.

Acampora, D., Di Giovannantonio, L.G., Garofalo, A., Nigro, V., Omodei, D., Lombardi, A., Zhang, J., Chambers, I., and Simeone, A. (2017). Functional Antagonism between OTX2 and NANOG Specifies a Spectrum of Heterogeneous Identities in Embryonic Stem Cells. **Stem Cell Reports** 9, 1642–1659.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jørgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., et al. (2006). Chromatin signatures of pluripotent cell lines. **Nat. Cell Biol.** 8, 532–538.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. **Cell** 125, 315–326.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. **Cell** 122, 947–956.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. **Nature** 441, 349–353.

Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T., and Wysocka, J. (2014). Reorganization of Enhancer Patterns in Transition from Naive to Primed Pluripotency. **Cell Stem Cell 14**, 838–853.

Carter, A.C., Davis-Dusenbery, B.N., Koszka, K., Ichida, J.K., and Eggan, K. (2014). Nanog-independent reprogramming to iPSCs with canonical factors. **Stem Cell Reports** 2, 119–126.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional Expression Cloning of Nanog, a Pluripotency Sustaining Factor in Embryonic Stem Cells. **Cell** 113, 643–655.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. **Nature** 450, 1230–1234.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008). Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. **Cell** 133, 1106–1117.

Chew, J.-L., Loh, Y.-H., Zhang, W., Chen, X., Tam, W.-L., Yeap, L.-S., Li, P., Ang, Y.-S., Lim, B., Robson, P., et al. (2005). Reciprocal Transcriptional Regulation of Pou5f1 and Sox2 via the Oct4/Sox2 Complex in Embryonic Stem Cells. **Mol. Cell. Biol.** 25, 6031–6046.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.-H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., et al. (2002). A genomic regulatory network for development. **Science** 295, 1669–1678.

Festuccia, N., Osorno, R., Halbritter, F., Karwacki-Neisius, V., Navarro, P., Colby, D., Wong, F., Yates, A., Tomlinson, S.R., and Chambers, I. (2012). Esrrb Is a Direct Nanog Target Gene that Can Substitute for Nanog Function in Pluripotent Cells. **Cell Stem Cell** 11, 477–490.

Frum, T., and Ralston, A. (2015). Cell signaling and transcription factors regulating cell fate during formation of the mouse blastocyst. **Trends Genet.** 31, 402–410.

Ho, L., Jothi, R., Ronan, J.L., Cui, K., Zhao, K., and Crab-

tree, G.R. (2009). An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. **PNAS** 106, 5187–5191.

Huang, D., Wang, L., Duan, J., Huang, C., Tian, X. (Cindy), Zhang, M., and Tang, Y. (2018). LIF-activated Jak signaling determines Esrrb expression during late-stage reprogramming. **Biology Open** 7, bio029264.

Huang, S., Eichler, G., Bar-Yam, Y., and Ingber, D.E. (2005). Cell fates as high-dimensional attractor states of a complex gene regulatory network. **Phys. Rev. Lett.** 94, 128701.

Kidder, B.L., Palmer, S., and Knott, J.G. (2009). SWI/SNF-Brg1 Regulates Self-Renewal and Occupies Core Pluripotency-Related Genes in Embryonic Stem Cells. **STEM CELLS** 27, 317–328.

Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. (2008). An extended transcriptional network for pluripotency of embryonic stem cells. **Cell** 132, 1049–1061.

King, H.W., and Klose, R.J. (2017). The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. **ELife** 6.

Konstantinides, N., Kapuralin, K., Fadil, C., Barboza, L., Satija, R., and Desplan, C. (2018). Phenotypic Convergence: Distinct Transcription Factors Regulate Common Terminal Features. **Cell**. 174, 1-14.

Leeb, M., Pasini, D., Novatchkova, M., Jaritz, M., Helin, K., and Wutz, A. (2010). Polycomb complexes act redundantly to repress genomic repeats and genes. **Genes Dev** 24, 265–276.

Loh, Y.-H., Wu, Q., Chew, J.-L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., et al. (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. **Nature Genetics** 38, 431–440.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. **Nature** 448, 553–560.

Navarro, P., Chambers, I., Karwacki-Neisius, V., Chureau, C., Morey, C., Rougeulle, C., and Avner, P. (2008). Molecular Coupling of Xist Regulation and Pluripotency. **Science** 321, 1693–1695.

Navarro, P., Festuccia, N., Colby, D., Gagliardi, A., Mullin, N.P., Zhang, W., Karwacki-Neisius, V., Osorno, R., Kelly, D., Robertson, M., et al. (2012). OCT4/SOX2-independent Nanog autorepression modulates heterogeneous Nanog gene expression in mouse ES cells. **The EMBO Journal** 31, 4547–4562.

Nishiyama, A., Xin, L., Sharov, A.A., Thomas, M., Mowrer, G., Meyers, E., Piao, Y., Mehta, S., Yee, S., Nakatake, Y., et al. (2009). Uncovering Early Response of Gene Regulatory Networks in ESCs by Systematic Induction of Transcription Factors. **Cell Stem Cell** 5, 420–433.

Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. **Nature Genetics** 24, 372–376.

Niwa, H., Ogawa, K., Shimosato, D., and Adachi, K. (2009). A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. **Nature** 460, 118–122.

Parfitt, D.-E., and Shen, M.M. (2014). From blastocyst to gastrula: gene regulatory networks of embryonic stem cells and early mouse embryogenesis. **Philos. Trans. R. Soc. Lond., B, Biol. Sci.** 369.

Pasini, D., Bracken, A.P., Hansen, J.B., Capillo, M., and Helin, K. (2007). The Polycomb Group Protein Suz12 Is Required for Embryonic Stem Cell Differentiation. **Mol. Cell. Biol.** 27, 3769–3779.

Schwarz, B.A., Bar-Nur, O., Silva, J.C.R., and Hochedlinger, K. (2014). Nanog Is Dispensable for the Generation of Induced Pluripotent Stem Cells. **Current Biology** 24, 347–350.

Shinozaki, K., Yamaguchi-Shinozaki, K., and Seki, M. (2003). Regulatory network of gene expression in the drought and cold stress responses. **Curr. Opin. Plant Biol.** 6, 410–417.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. **Nat. Rev. Genet.** 15, 272–286.

Silva, J., Nichols, J., Theunissen, T.W., Guo, G., van Oosten, A.L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog Is the Gateway to the Pluripotent Ground State. **Cell** 138, 722–737.

Sone, M., Morone, N., Nakamura, T., Tanaka, A., Okita, K., Woltjen, K., Nakagawa, M., Heuser, J.E., Yamada, Y., Yamanaka, S., et al. (2017). Hybrid Cellular Metabolism Coordinated by Zic3 and Esrrb Synergistically Enhances Induction of Naive Pluripotency. **Cell Metabolism** 25, 1103-1117.e6.

Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. **Nat. Rev. Genet.** 13, 613–626.

Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S., and Vale, R.D. (2014). A protein tagging system for signal amplification in gene expression and fluorescence imaging. **Cell** 159, 635–646.

Utley, R.T., Ikeda, K., Grant, P.A., Côté, J., Steger, D.J., Eberharter, A., John, S., and Workman, J.L. (1998). Transcriptional activators direct histone acetyltransferase complexes to nucleosomes. **Nature** 394, 498–502.

Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. **Nature** 453, 519–523.

Zhang, M., Leitch, H.G., Tang, W.W.C., Festuccia, N., Hall-Ponsele, E., Nichols, J., Surani, M.A., Smith, A., and Chambers, I. (2018). Esrrb Complementation Rescues Development of Nanog-Null Germ Cells. **Cell Reports** 22, 332–339.

SUPPLEMENTARY INFORMATION

Five supplementary figures accompany this manuscript:

Fig.S1: Characterisation of Dox-inducible SunTag ES cells.

Fig.S2: Validation of Nanog-responsive genes identified in this study.

Fig.S3: Heterogeneity of TF binding and chromatin accessibility throughout Nanog binding regions.

Fig.S4: Global and gene-specific responses of the transcriptome to Nanog induction in the absence of LIF.

Fig.S5: H3K27me3 and Otx2 are involved in LIF-independent self-renewal as established by Nanog.

They can be found at the end of this document.

Five Supplementary Tables are available online:

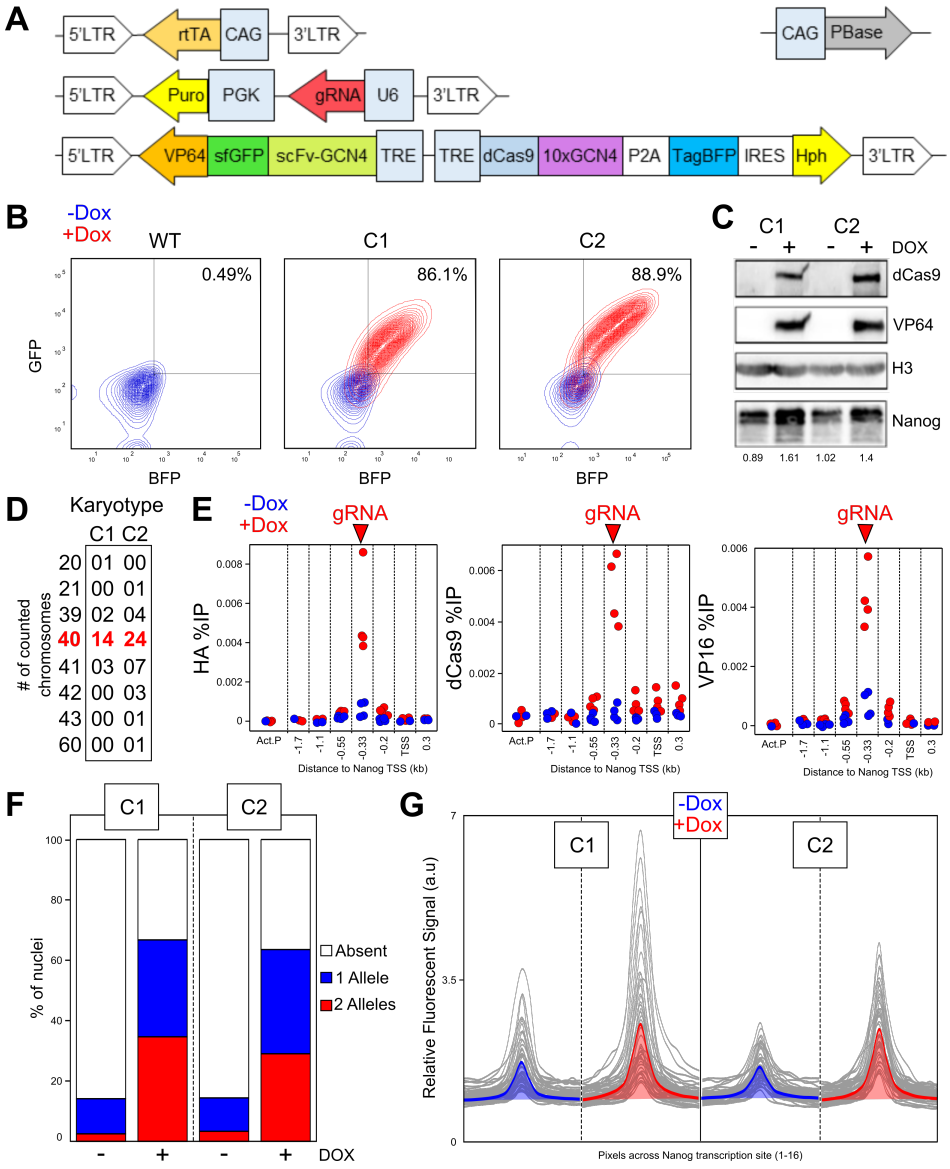
Table S1: The Nanog-centred transcriptome. In this table, all RNA-seq results are provided and annotated.

Table S2: Compendium of Nanog binding sites and their properties. In this table, all TF ChIP-seq and ATAC-seq results are provided and annotated.

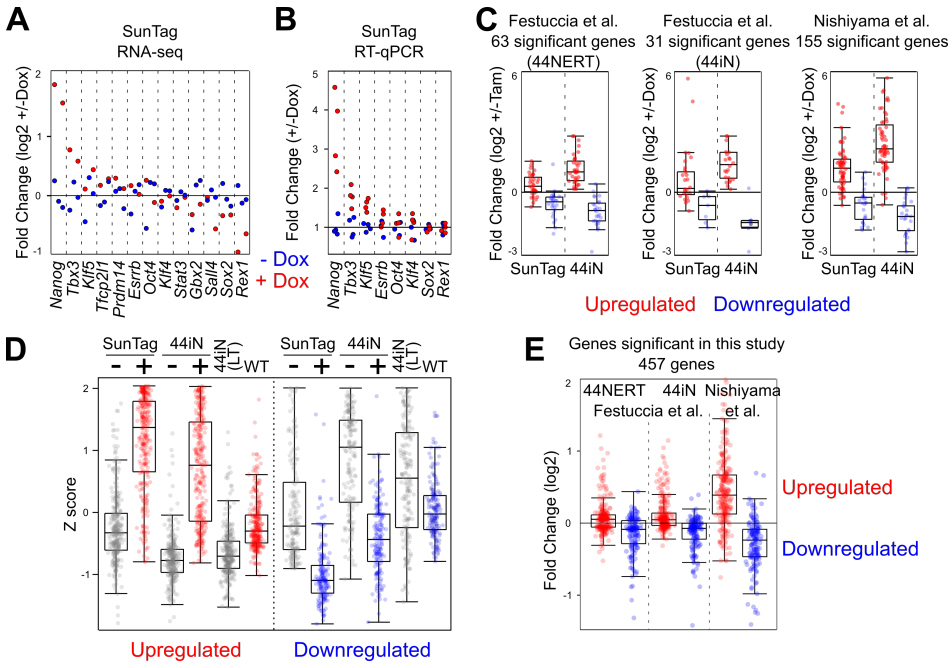
Table S3: Nanog and LIF-dependent H3K27me3. In this table, all H3K27me3 domains are provided and annotated.

Table S4: Primer sequences and antibodies used in this study.

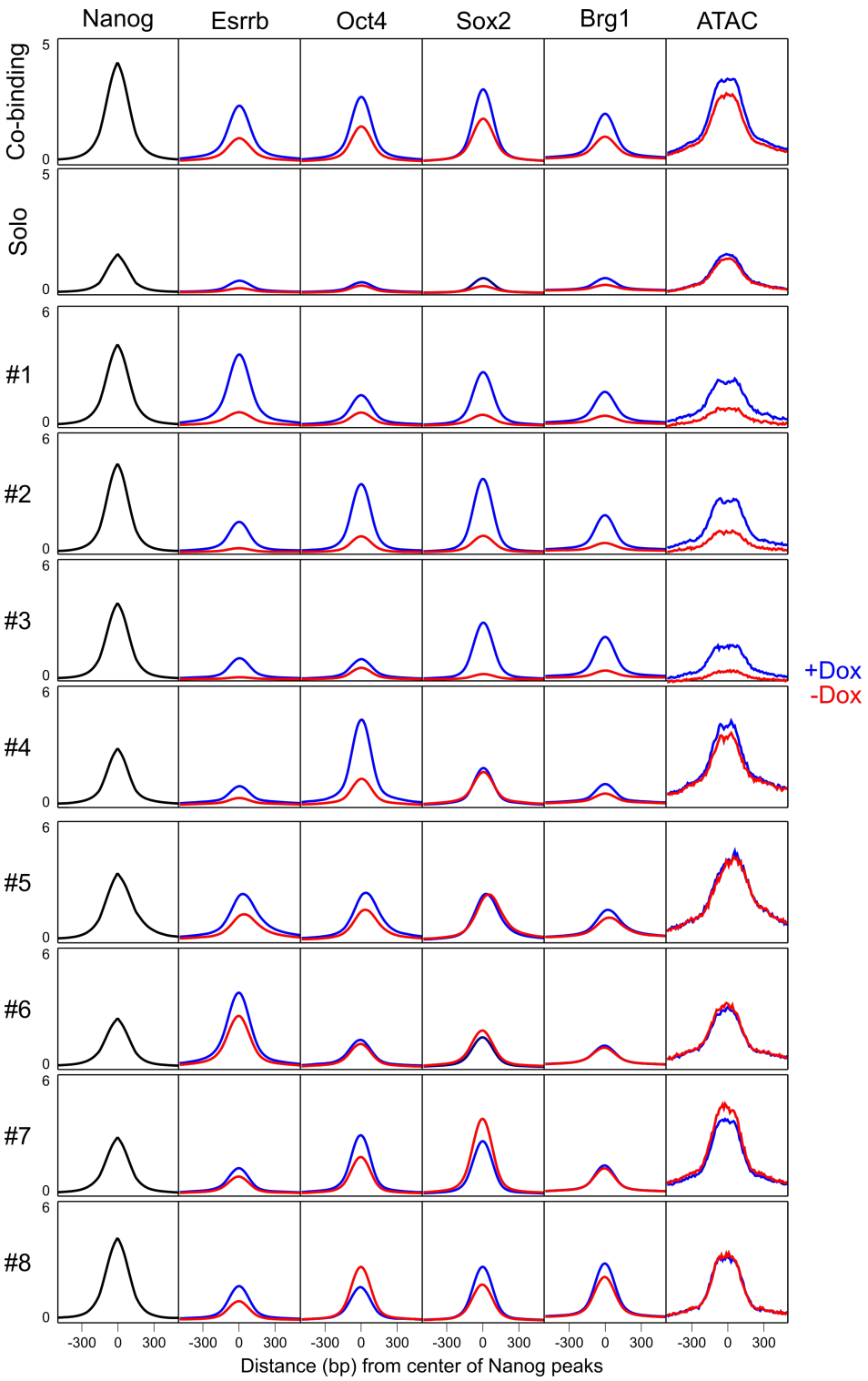
Table S5: Data Overview. In this table, all samples collected for RNA-seq, ChIP-seq and ATAC-seq are described along with Nanog public data sets.



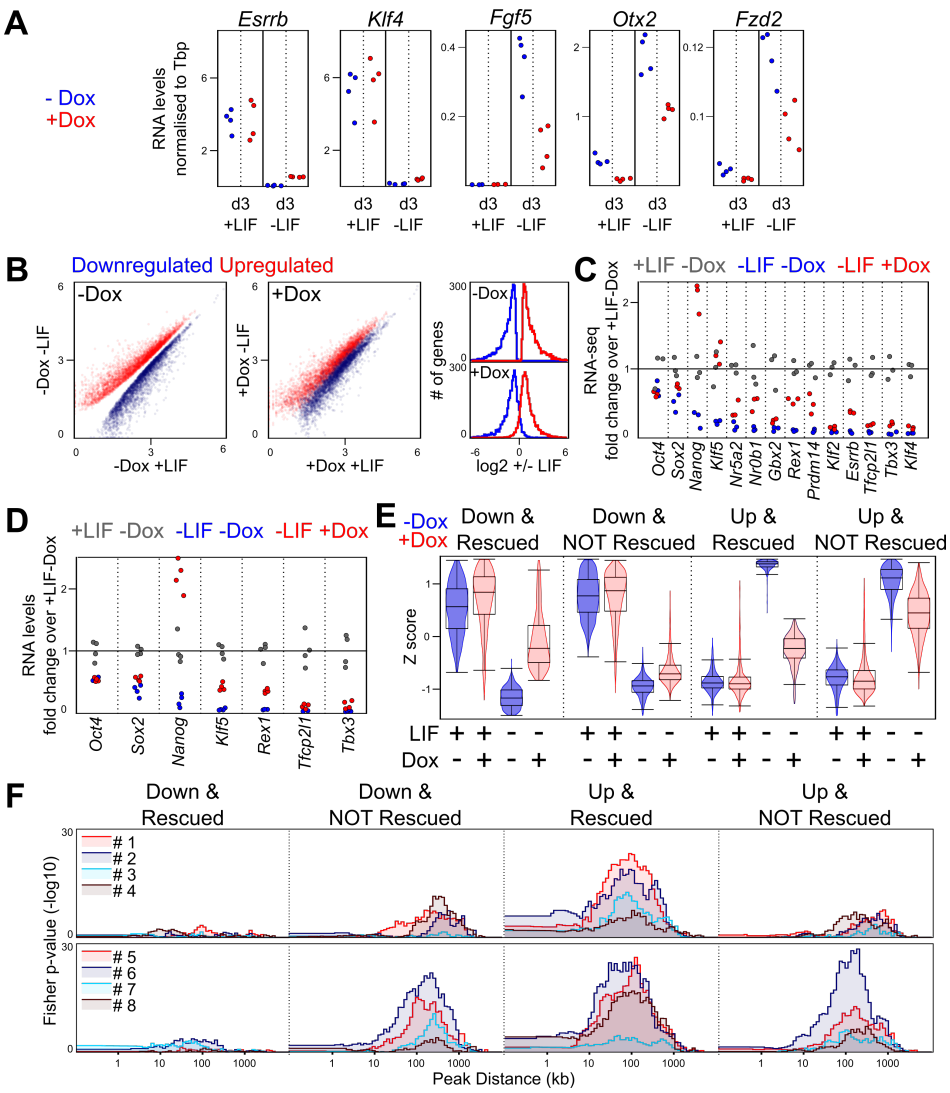
Supplementary Information, Fig. S 1. Characterisation of Dox-inducible SunTag ES cells. **(A)** Schematic representation of the Piggybac vectors used to generate ES cells expressing inducible SunTag transactivators. LTR: long terminal repeat; rtTA: reverse tetracycline-controlled transactivator; PBase: Piggybac transposase (non-integrative vector); CAG: constitutive RNAPII promoter; Puro: Puromycin resistance cassette; PgK: constitutive RNAPII promoter; U6: RNAPIII promoter for gRNA transcription; VP64: tetrameric fusion of Herpes simplex virus transactivation domain; sfGFP: super-folder green fluorescent protein; scFv-GCN4: Single-chain variable fragment antibody directed against the GCN4 yeast epitope; TRE: tetracycline responsive element; dCas9: enzymatically dead Cas9; 10xGCN4: 10 copies in tandem of the yeast GCN4 epitope; P2A: self-cleaving peptide; TagBFP: monomeric blue fluorescent protein; IRES: internal ribosome entry site; Hph: Hygromycin resistance cassette. **(B)** GFP versus BFP FACS profiles of wild-type cells (WT) and of two SunTag clones (C1 and C2) in the absence (blue) and the presence of Dox (red). The percentage indicates the proportion of double-positive cells. **(C)** Western-Blot analysis of the indicated proteins (right) in the indicated conditions (top). The numbers underneath indicate relative Nanog levels. **(D)** Karyotypes of SunTag ES cells. **(E)** ChIP analysis of the indicated proteins (the HA epitope is present in the two moieties of the SunTag system) at different positions along the *Nanog* locus (X-axis) with (red) and without Dox (blue). The amplicon overlapping the gRNA targeted sequence is indicated (top). Each data point represents an independent replicate (n=4; in both C1 and C2). **(F)** Proportion of SunTag cells displaying one or two *Nanog* active transcription sites as established by smFISH in the presence/absence of Dox for the two SunTag clones (n>500 nuclei in each clone/condition). **(G)** Relative quantification of the fluorescent signal measured along a line crossing Nanog transcription sites (n=40 sites in each clone/condition).



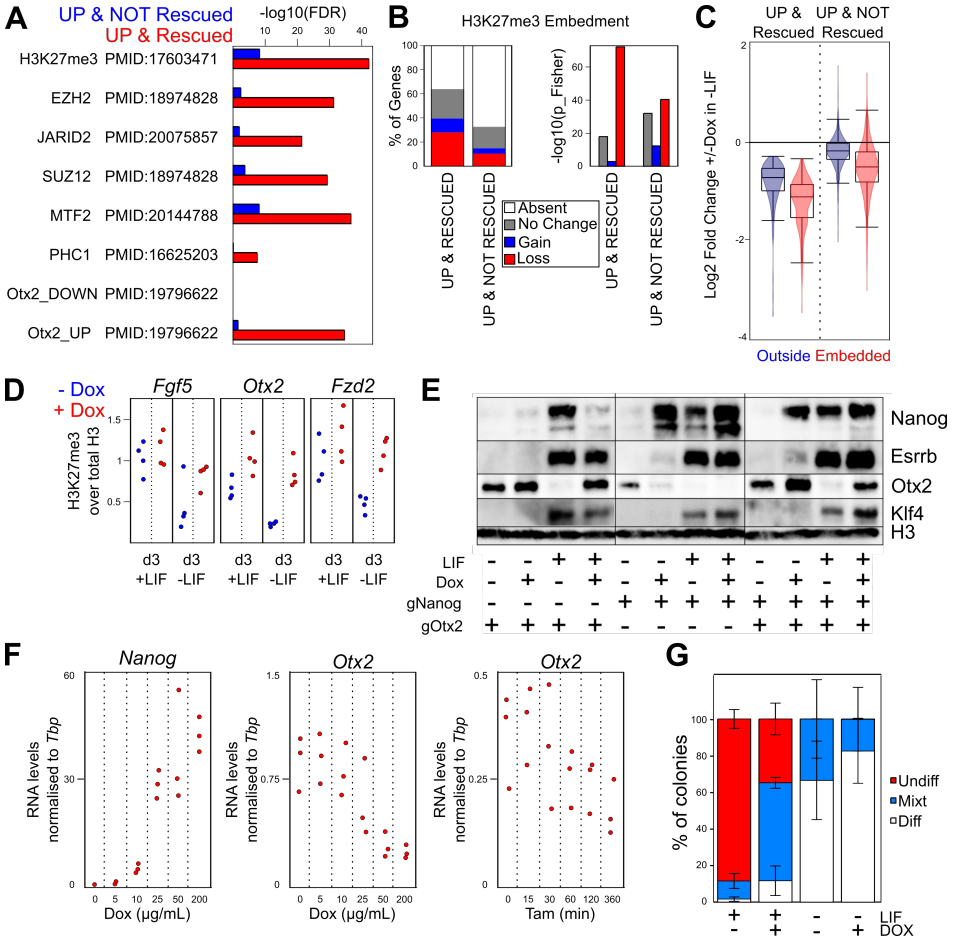
Supplementary Information, Fig. S 2. Validation of Nanog-responsive genes identified in this study. (A) Fold change (log2) of a set of pluripotency genes upon Dox treatment of SunTag cells. For each transcript, each value in the absence (blue) and in the presence (red) of Dox was normalised to the average of independent replicates. **(B)** RT-qPCR validation of a subset of the transcripts analysed in (A). **(C)** Confrontation of published results, as indicated, with our Nanog-responsive genes identified in SunTag and 44iN cells. **(D)** Box plot representation (z score, as in Fig. 2E) of expression levels of the genes identified in this study when considering both SunTag and 44iN together. 44iN(LT) indicate long-term culture in the absence of Dox (i.e. *Nanog* knock-out). **(E)** Quantification in published datasets, as indicated, of our extended list of Nanog-responsive genes.



Supplementary Information, Fig. S 3. Heterogeneity of TF binding and chromatin accessibility throughout Nanog binding regions. The average binding profile of each factor (A.U. correcting for TF occupancy as described in Methods), as labelled on the top, is shown across Nanog peaks (summit at 0bp) for each Nanog subgroup identified in Fig. 3.



Supplementary Information, Fig. S 4. Global and gene-specific responses of the transcriptome to *Nanog* induction in the absence of LIF. (A) RT-qPCR of the indicated mRNAs across the conditions shown on the X-axis, in the absence (blue) and the presence (red) of Dox. Each data point represents individual measurements in both C1 and C2 clones (2 for each). **(B)** Left: scatter plot of normalised mRNA levels (DESeq2 normalised) of differentially expressed genes in the presence/absence of LIF, identified in the absence of Dox (FDR<0.05). Middle: Identical representation of the same set of transcripts but measured in the presence of Dox. Right: histogram showing the distribution of LIF-responsive genes across a range of fold changes (X-axis) in the absence (top) and the presence (bottom) of Dox. **(C)** Relative mRNA levels of a set of pluripotency factors normalised to the average expression of all replicates measured in the presence of LIF and the absence of Dox (n=3). **(D)** Validation by RT-qPCR of gene expression genes for a subset of pluripotency factors, presented like in (B). **(E)** Z score violin plots of the four groups of genes identified by RNA-seq in SunTag cells cultured in the absence/presence of LIF/Dox across the conditions indicated on the X-axis. **(F)** Association of the same four groups with the individual Nanog-binding clusters identified in Fig. 3, presented as in Fig. 3.



Supplementary Information, Fig. S 5. H3K27me3 and Otx2 are involved in LIF-independent self-renewal as established by Nanog. (A) Enrichment (-log10FDR) of UP and NOT rescued and UP and rescued genes for Polycomb Group targets and Otx2 responsive genes. The specific publication used by the Enrichr software (http://amp.pharm.mssm.edu/Enrichr/) to compute the enrichment is shown. **(B)** Proportion (left) and statistical significance (right) of H3K27me3-embedded genes activated without LIF and displaying differential rescue by Nanog (X-axis). **(C)** Violin plot of the log2 fold change of expression for the two previous categories of LIF-responsive genes split in function of their embedment (red) or not (blue) within H3K27me3 domains. **(D)** Normalised H3K27me3 ChIP-qPCR at three genes showing Nanog-dependent H3K27me3 enrichment in the absence of LIF. **(E)** Western Blot of the indicated factors (right) across multiple conditions (bottom). Note other WB presented in the principal figures correspond exactly to the blots shown here to facilitate direct comparisons between all analysed conditions. **(F)** Normalised mRNA levels of *Nanog* and *Otx2* over a Dox dose-response assay in 44iN and during a time-course of Tamoxifen treatment in Nanog-ERT2 fusion cells (44NERT), measured by RT-qPCR (n=3). 44NERT cells are *Nanog* knock-out cells expressing Nanog-ERT2 transgene (**Navarro et al., 2012**). (G) Clonal assay (n=4) in the indicated conditions (X-axis) using SunTag cells that activate *Otx2* exclusively.

IX. LASER: LncRNAs Associated with SElf-Renewal of mouse ES cells

A. Introduction

LncRNAs have been functionally associated with all layers of gene expression regulation (F. Kopp and Mendell 2018). Many of those have been shown to be specifically expressed in pluripotent cells when compared to their differentiating offspring (Dinger et al. 2008; Mitchell Guttman et al. 2009, 2011; N. Lin et al. 2014; Bergmann et al. 2015; Fort et al. 2015; Bogu et al. 2016). Therefore, a functional role in regulating mouse ES cells fate was hypothesized for this class of transcripts as done in the past for protein coding genes found exclusively in pluripotent cells (Okamoto et al. 1990). But, despite intensive efforts, only few of them were described as shielding ES cell self-renewal with mostly limited impact. Of note, all of this large scale studies were based on loss-of-function strategies screening for impaired self-renewal capacities. The idea that lncRNAs might mostly display fine-tuning regulatory functions of gene expression with a possible high redundancy is progressively emerging (L. Li and Chang 2014; Ransohoff, Wei, and Khavari 2018). As evidenced in the previous section, some "ancillary" pluripotency regulators show redundant functions in mouse ES cells and have been shown to be dispensable for pluripotency maintenance in different contexts (Chambers et al. 2007; J. Jiang et al. 2008; Hitoshi Niwa et al. 2009; Martello et al. 2012; Hackett and Surani 2014; Yeo et al. 2014; Yamane et al. 2018). Yet, as perfectly illustrated by Nanog, they can harbour strong regulatory functions. We thus speculate that some lncRNAs may display a comparable behaviour that couldn't be captured by loss of function experiments scrutinizing for strong differentiation phenotypes.

Interestingly, the network of ancillary pluripotency factors (Nanog, Klf2, Esrrb, Klf4, Prdm14, Sall4, Tfcp2l1, Tbx3) has been shown to be strongly interconnected, each factor having in general a positive impact on the others (Dunn et al. 2014). Moreover, their expression is stabilized when mouse ES cells reach the ground state of pluripotency (Ying et al. 2008; Wray et al. 2011; Ghimire et al. 2018). Interestingly, it was reported that the most responsive transcriptional targets of Nanog were enriched in these factors (Festuccia et al. 2012), a notion that might need to be revisited in regards to our recent work (see previous section). Therefore, in order to identify "ancillary pluripotency lncRNAs" we aimed at identifying such molecules whose expression was reinforced in ground state condition and were transcriptionally

controlled by Nanog. A way to functionally characterize those non-coding transcripts then consists in overexpressing them in differentiation condition and assess whether this gives rise to sustained self-renewal: a property shared by most of the aforementioned pluripotency factors (Chambers et al. 2003; Hitoshi Niwa et al. 2009; Hall et al. 2009b; Festuccia et al. 2012). We thus took advantage of CRISPRa as an appropriate means to enhance their endogenous expression since, as aforementioned, transcription site is an important parameter for lncRNAs mediated function.

B. Preliminary selection of lncRNAs candidates

Hence, in order to select Nanog-regulated lncRNAs we took advantage of the 44iN cell line where the endogenous loci of Nanog have been invalidated but expression can be artificially restored by transgene induction under Dox treatment (Festuccia et al. 2012). Those cells were cultured for a long term in the absence of Nanog (44iN) before expression was rescued for twenty-four hours (44iN + Dox). WT cells routinely cultured in serum-containing medium (WT FCS) were placed in 2i/LIF condition for three days **(Fig. 2.15 A)**. RNAs from four biological replicates of each condition were prepared, RT-qPCR experiments were done for validation of few Nanog target or 2i responsive genes (not shown). The four replicates of each condition were pre-mixed prior to library preparation. Total RNAs after ribosomal depletion were further sequenced by high-throughput RNA-seq capturing even non- or poorly poly-adenylated transcripts (Y. Zhang, Yang, and Chen 2014). Since lncRNAs have been shown to be, on average, lower expressed than their coding counterparts (M. Guttman 2009; Mitchell Guttman et al. 2010; Ulitsky et al. 2011; Derrien et al. 2012; Sigova et al. 2013), RNA-seq was performed with a relatively important depth (more than 150 million reads/sample).

After alignment, data were further analysed with Seqmonk program (Babraham Institute - for detailed procedure see Material and Methods section). Since unique sequencing samples were generated for each condition, precluding any statistical approach, we first evaluated the reliability of our RNA-seq results through their comparison with published datasets of similar samples. For this aim, we looked at the expression of the genes reported as tightly regulated by Nanog (Festuccia et al. 2012) or differentially expressed between FCS/LIF and 2i/LIF condition (Marks et al. 2012) in our samples. It appeared that the response of those genes was highly consistent with their previously reported behaviour **(Fig. 2.16 A, B and C)**. More precisely, the percentage of genes showing the expected tendency (fold change

treated/control > or < 1) in our dataset was high for each kind of response (89% (1737/1947) serum higher, 79% (1169/1479) 2i higher, 71% (346/489) Nanog induced, 91% (390/430) Nanog repressed). This result thus allowed us to gain further confidence in our datasets.

In order not to restrict our analysis to already annotated lncRNAs we combined three different blind analyses to identify our gene candidates. First, using Seqmonk analysis software, we divided the genome in small overlapping adjacent bins and quantified RNA-seq signal on both strands separately **(Fig. 2.15 B)**. Genomic bits with sufficient signal level showing a Nanog-dependent expression and being upregulated in 2i condition were selected **(Fig. 2.15 A)**. Since only one replicate was sequenced for each condition, our selection was only based on fold change thresholding. All selected segments overlapping an annotated CDS (coding DNA sequence) on the same strand were trimmed out. This analysis resulted in the selection of about 900 genomic pieces actually representing 343 independent loci. Second, an integrated Seqmonk tool called 'contig probe generator' identifying continuous stretches of reads was used to blindly bait active transcriptional blocks. CDS overlapping probes were discarded and 63 clusters with relevant expression profile were finally obtained. Finally, our sequencing data were sent to a collaborating group (Ulitsky I. laboratory, Noa Gil, Weizmann Institute of Science, Israel) who identified high confidence lncRNAs thanks to their previously published pipeline integrating conservation and coding potential features (Hezroni et al. 2015). From these high confidence transcripts, 31 regions fitting into our researched expression profile were selected.

Finally, the pool of all these transcripts was manually curated in order to remove alignment artefacts (mainly pseudogenes and repeated region misalignments), further selected on clear exonic structure, expression level and presence of active chromatin marks and led to a final list of 24 transcripts **(see Table 2)**. Given their expected regulation by Nanog and increased expression in the ground state of pluripotency, these lncRNAs will thereafter be referred to as **LASER** for **L**ncRNA **A**ssociated with **SE**lf-**R**enewal.

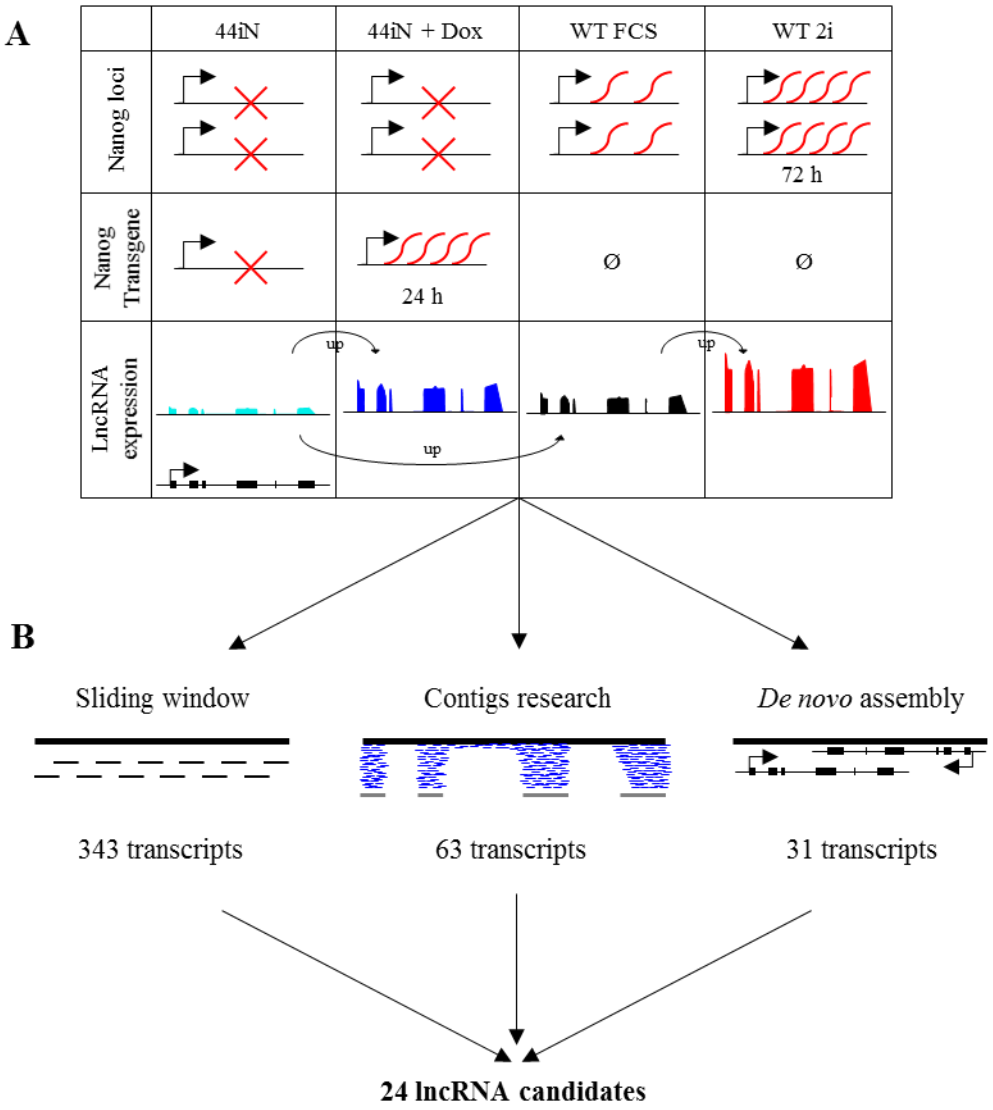


Fig. 2.15. A. Schematic representation of the four RNA samples sent for RNA-seq. 44iN Nanog KO cells were continuously kept in the absence of induction before Dox treatment was applied for 24 hours (44iN + Dox) shortly rescuing Nanog expression. WT E14 cells were cultured in FCS/LIF (WT FCS) and placed in 2i/LIF (WT 2i) medium for 3 days. Last raw corresponds to the expression profile that should be followed by lncRNA candidates. **B.** Analysis pipeline used to select our 24 LASER.

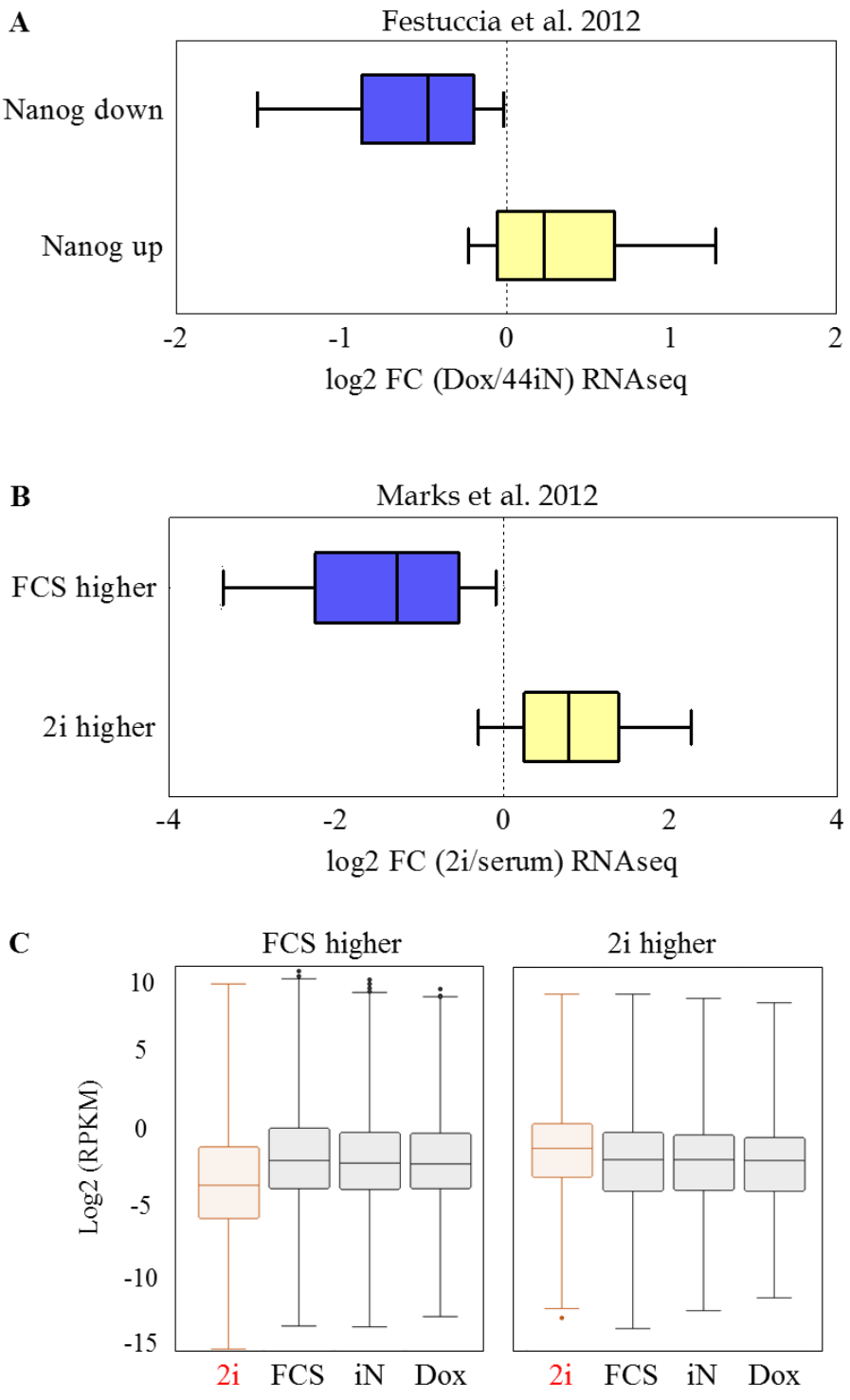


Fig. 2.16. External validation of the RNA-seq data **A.** Variation of expression between the 44iN+Dox and 44iN conditions for the 489 genes up regulated and 430 genes downregulated upon Dox treatment in 44iN cells from Festuccia et al. 2012. **B.** Variation of expression between 2i and FCS conditions for the 1947 'serum higher' genes and 1479 '2i higher' genes from Marks et al. 2012. **C.** Expression of the serum and 2i higher genes from Marks et al. 2012 in our four sequenced samples showing consistent tendency of the two sets of genes in all our serum samples compared to 2i condition.

C. Characterization of the 24 LASER

After manually defining the 5' and 3' extremity of each of the LASER from our RNA-seq data, RNA-seq quantification was done on the 24 loci. While their selection was sometimes only based on subsections of the full transcript, most of them showed the expected pattern of expression **(Fig. 2.17 A)**. Their expression profile was further validated by RT-qPCR for 23 of them (no acceptable primers could be designed for LASER5) **(Fig. 2.17 B)** on two separated replicates and showed an excellent consistency with the RNA-seq results therefore validating the reliability of our approach.

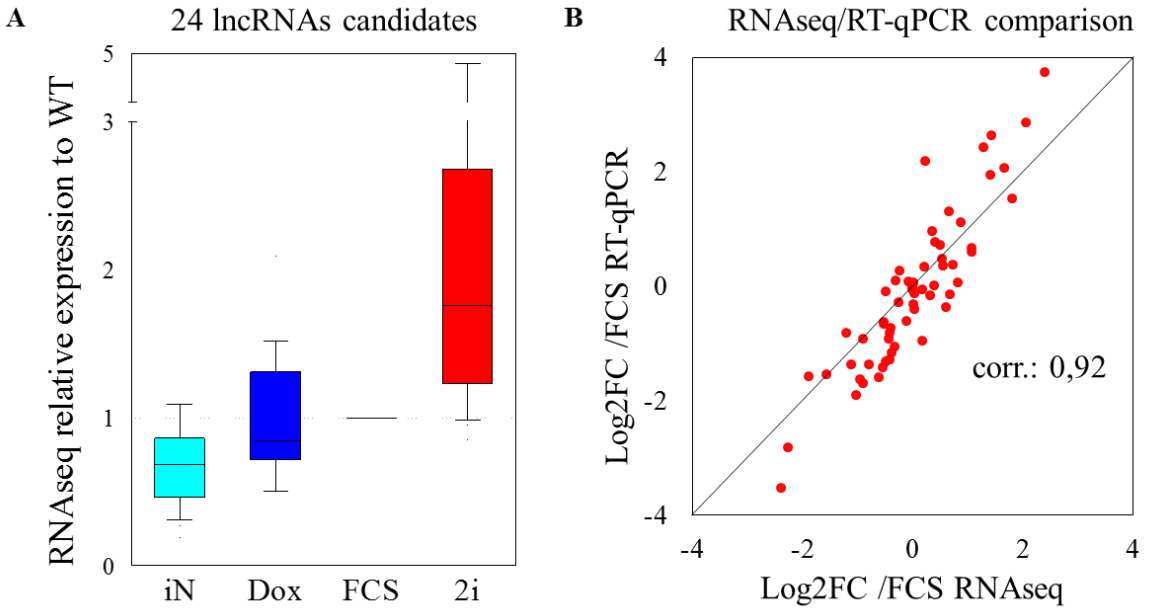


Fig. 2.17. Validation of the LASER expression **A.** RNA-seq expression profile of all LASER normalized to their expression in FCS/LIF condition showing low expression in 44iN cells, upregulation upon Dox treatment and upregulation in 2i condition. **B**. Log2 fold changes relative to FCS/LIF condition were calculated separately for the RNA-seq and the RT-qPCR data (3 per LASER) and plotted against each other. (RNA-seq n=1, RT-qPCR n=2). corr.= Pearson correlation coefficient.

From our 24 LASER, 13 were already annotated in ENSEMBL (Zerbino et al. 2018) as non-coding RNAs, some of the others were partially covered by fragmentary annotations found in large non-coding transcripts dedicated databases (Y. Zhao et al. 2016; Zheng et al. 2016). In order to check the absence of coding potential of the LASER, but since their exact sequence could not always be clearly inferred from our sequencing data, the PhyloCSF (M. F. Lin, Jungreis, and Kellis 2011) track of the UCSC genome browser was scrutinized all along

our LASER loci. None of them showed positive PhyloCSF score over their whole gene body in any of the possible frames, thus suggesting the absence of coding capacities of these transcripts.

D. LASER gRNAs design and test

After validating the expression profile and the absence of coding potential of our LASER we aimed at functionally testing their capacity to promote sustained self-renewal. We therefore needed to design gRNAs targeting each of our 24 promoters for subsequent overexpression with our SunTag cell line. A new algorithm for gRNAs design was set up in collaboration with the Bioinformatics platform of the Institut Pasteur (Damien Mornico) for rapid and multiplexed gRNAs design (see methods). After running this algorithm on the upstream regions (from -50bp to -1kb) preceding the TSS of our 24 lncRNA candidates, the two (or one if two acceptable gRNAs couldn't be designed) gRNAs showing the lowest off-target probability were selected. All selected gRNAs were further cloned separately into the PiggyBac gRNA expressing vector.

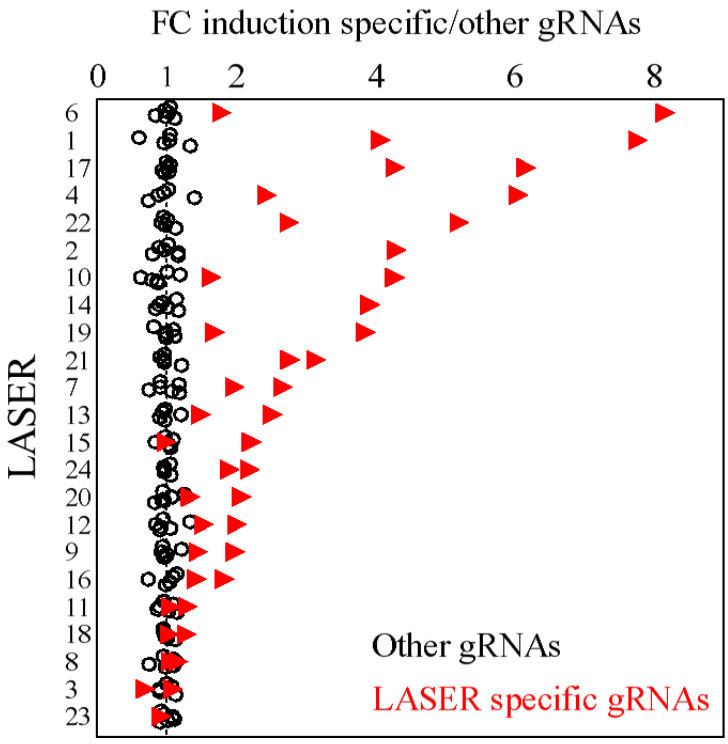


Fig. 2.18. All gRNAs were transfected separately in our SunTag cell line. Cells were treated with Dox for 2 days and expression of each LASER was measured by RT-qPCR in the samples having received its targeting gRNA as well as in samples with gRNAs targeting another LASER. Those latter values were used to normalize each LASER expression separately. (n=1)

We further tested the efficiency of all successfully cloned gRNAs separately (43 in total) **(Fig. 2.18)**. This showed that near half of them gave rise to a higher than twofold upregulation while around 25% of them completely failed to induce any transcriptional response. This led to the successful activation of more than half of our LASER although at least five of them didn't get activated at all in this single assay. LASER 5 gRNAs efficiency couldn't be assessed since no efficient qPCR primers were obtained. No specific bias could obviously be found for the non-working gRNAs in term of sequence or distance to TSS. Therefore, more than half of our LASER could be moderately overexpressed in our SunTag cell line allowing for functional test in differentiation condition.

E. LASER overexpression upon LIF withdrawal

From our previous experiment, we picked for further experiments the gRNAs leading to the highest response for the 18 LASER showing transcriptional induction under SunTag activation. We additionally selected the four gRNAs showing an induction fold change greater than two (LASER 1, 17, 4, 22). All of them were transfected for a second time (independently of previous section) in our SunTag clones along with a gRNA targeting Nanog promoter (see section II) and a non-targeting gRNA (LacZ). Cells were plated at low density in the absence of LIF, where mouse ES cells massively differentiate (Austin G. Smith et al. 1988), and treated with Dox. After a week, cells were fixed and stained for alkaline phosphatase activity, the gold standard to assess the undifferentiated state of mouse ES cells. While, as expected, Nanog gRNA transfected cells gave rise to a homogenous layer of stained colonies, our LacZ negative control only showed the presence of few coloured colonies **(Fig. 2.19)**. This confirmed the ability of our experimental setup to discriminate between factors conferring LIF-independence and those that do not. Unfortunately, all cells transfected with LASER's gRNAs essentially resembled our negative control, in both SunTag clones (see two examples in **Fig. 2.19**). We therefore concluded that none of our tested LASER (18/24) gave rise to LIF-independent self-renewal when overexpressed.

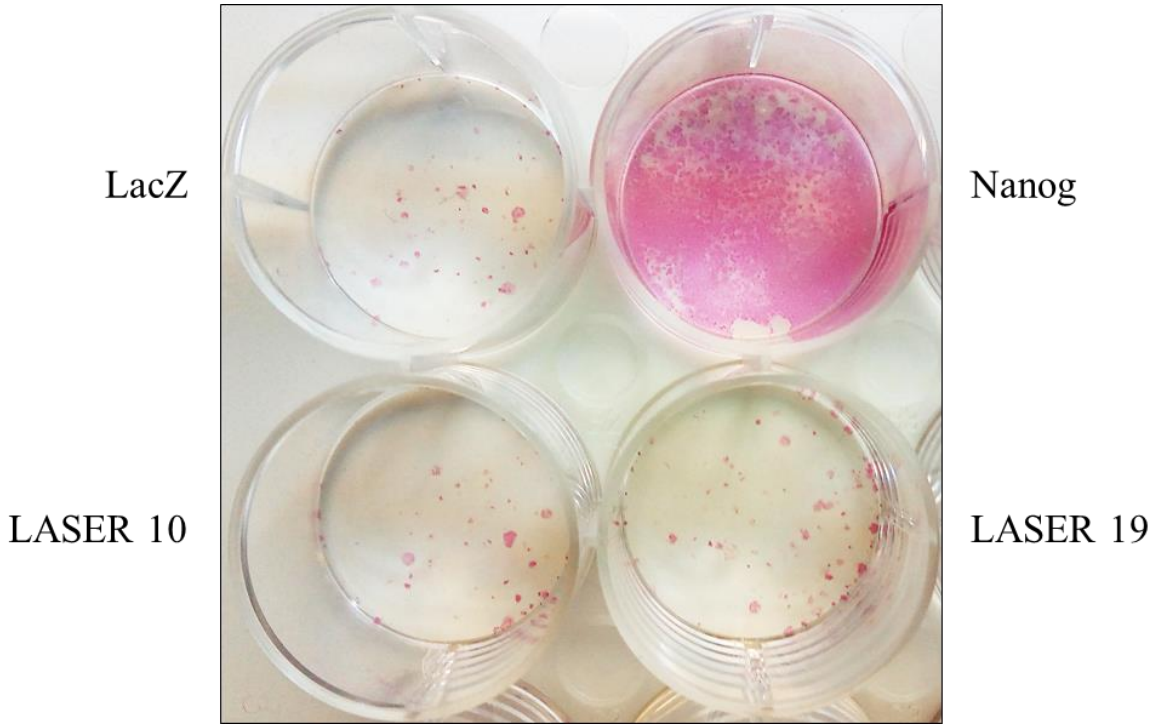


Fig. 2.19. Representative picture of Alkaline Phosphatase staining including LacZ, Nanog, and two LASER gRNAs after 6 days in –LIF +Dox condition. Undifferentiated colonies are stained in pink while differentiated colonies are transparent (not visible here).

F. Transcriptomic response upon three LASER overexpression

Following the negative conclusion we faced in the previous functional test, we decided to focus our attention on three of the LASER. One of them, LASER1, is located 50kb away from Esrrb gene and shows great upregulation in 2i medium as well as low expression in Nanog KO cell line but is not responsive to Nanog rescue **(Fig. 2.20)**. The second one, LASER 10, shows both a nice response to Nanog rescue and a strong response in 2i medium. It is located close to Tfap2c gene **(Fig. 2.21 A)** which has been largely associated both with trophectoderm and PGCs differentiation process (Magnúsdóttir et al. 2013; Schemmer et al. 2013; Z. Cao et al. 2015). The last one, LASER 17, initiates in a mouse ES cells super enhancer and is adjacent to Cd9 locus **(Fig. 2.21 A)**. Cd9 has been shown to be highly expressed in mouse ES cells despite its dispensability for mouse ES cells self-renewal (Akutsu et al. 2009). LASER 17 shows about 3 fold upregulation both upon Nanog rescue and FCS/2i switch and is also lowly expressed in Nanog KO cells **(Fig. 2.20)**. All three appeared to be efficiently spliced and largely detected in polyA selected datasets **(Fig. 2.21 A)**. Therefore, their precise spliced sequences could be determined and submitted to coding potential evaluation tool CPAT (L. Wang et al. 2013). They all got labelled as non-coding with low coding probabilities (from 4 to 25%, a

range where the discrimination specificity of CPAT is still greater that 90%). Of note, none of them had an annotated human orthologue in GENCODE database. Therefore, the relevance of their genomic location, their *bona fide* characteristics of mature RNA-Pol II transcribed RNAs as well as their very low coding probability prompted us to assess the transcriptional effects of their endogenous overexpression.

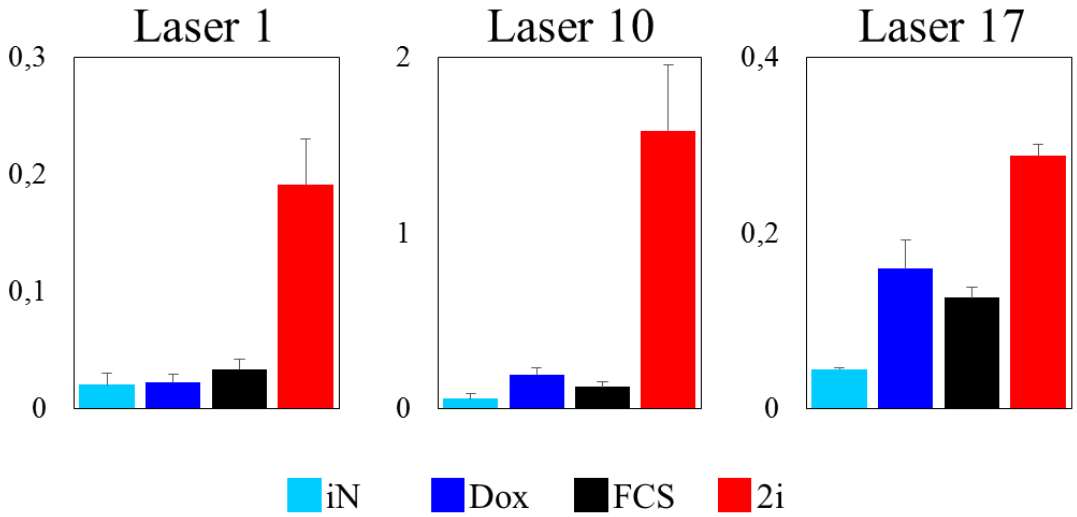


Fig. 2.20. RT-qPCR validation of the expression profile of LASER1, 10 and 17 in the samples listed before showing strong response upon FCS to 2i medium switch for the 3 lncRNAs but rather different levels of response to Nanog expression. Values are normalized to Tbp level. (n=2, SEM).

The two gRNAs targeting each of these three LASER were among the most efficient ones (**Fig. 2.18**). Thus, we transfected again the six corresponding gRNAs separately in our two SunTag clones and manually picked single subclones originating from each of these transfections. In total, 21 subclones originating from the two parental SunTag clones and harbouring one of the six gRNAs were kept for further experiments (each gRNA being represented by 2 to 4 subclones). This ensured that a clonal or SunTag off-target effect would be easily identified and separated from any significant result. We treated all these clones with Dox for 72 hours before harvesting them for RNA purification. SunTag induction of the three LASER was validated by RT-qPCR and compared to three additional subclones that received the LacZ non-targeting gRNA **(Fig. 2.21 B)**. This confirmed the strong upregulation of our three lncRNAs with a consistent difference in efficiency within the couples of gRNAs targeting the same LASER. Therefore, in order to assess the transcriptomic consequences of the overexpression of these three self-renewal associated transcripts, we sent these 48 samples for RNA-seq (21 LASER subclones, 3 LacZ clones, -/+ Dox).

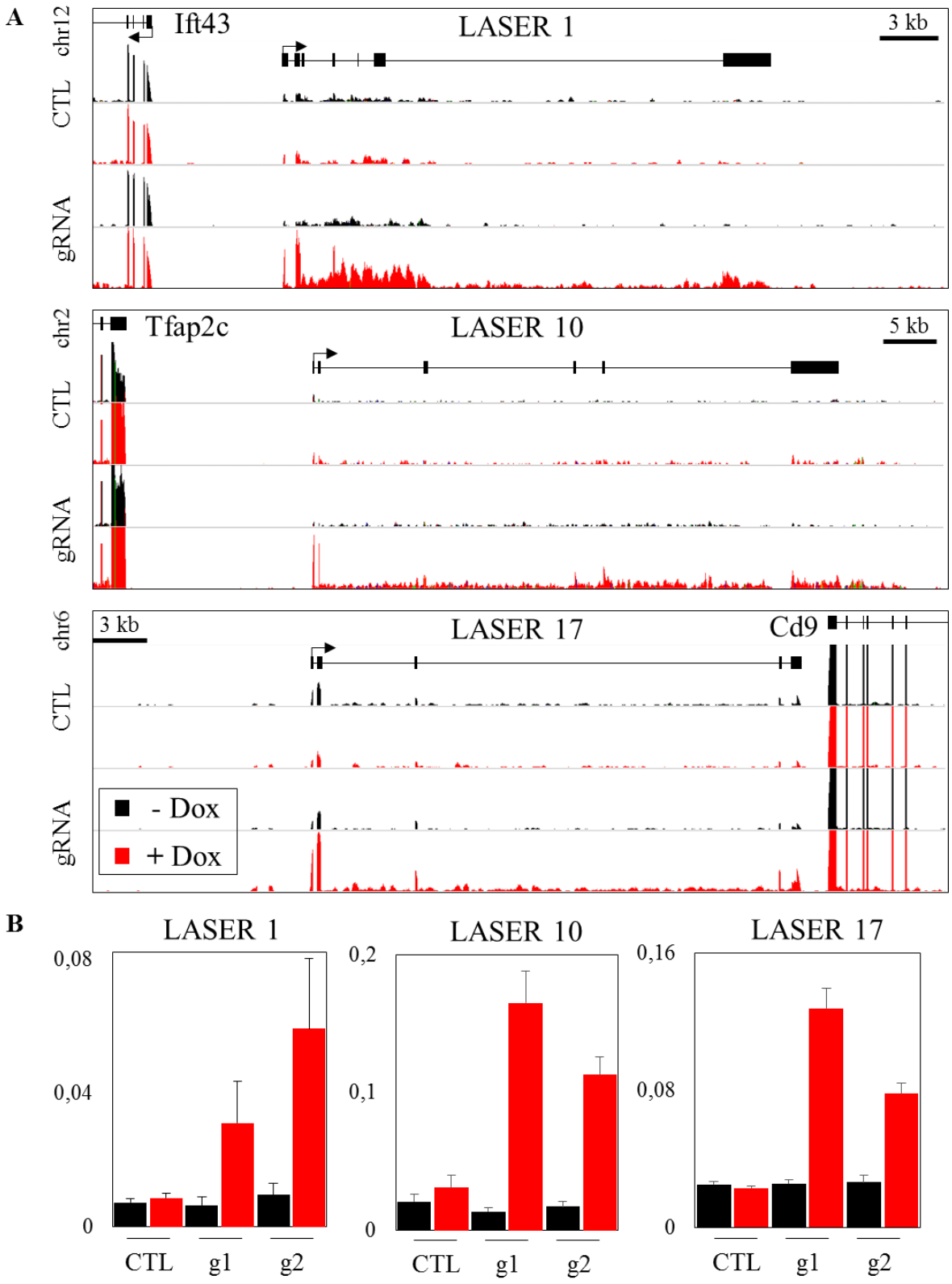


Fig. 2.21. **A.** Schematic representation of the three LASER loci with RNA-seq coverage profile upon SunTag induction. CTL = LacZ non targeting gRNA. **B.** RT-qPCR measure of the induction with SunTag for 72h in FCS/LIF, values are normalized to Tbp level (n=4-8, Mean ± SEM).

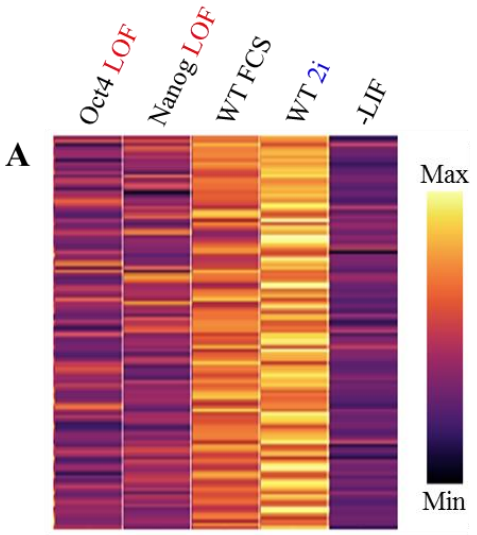
The analysed RNA-seq signal (aligned by N. Owens) validated the upregulation of our three LASER from their endogenous loci **(Fig. 2.21 A)**. Interestingly, transcription initiation and termination as well as splicing didn't seem to be affected by the artificial activation produced by the recruitment of the SunTag complex. Expression of the surrounding genes didn't seem to be affected neither (not shown). We then performed a statistical analysis with DESEQ program to identify the differentially expressed genes upon activation of the three LASER. Few differentially expressed genes (DEGs) with relatively low expression level could be identified between none treated and Dox treated samples for each gRNA separately (FDR<0.05). However, when intersecting the lists of DEGs obtained with the two gRNAs targeting the same LASER, it appeared that each overlap was empty. If performing the same DESEQ analysis by pooling together the samples obtained with the two gRNAs, even less DEGs were significantly (FDR<0.05) identified showing usually inconsistent expression profile along all our samples. It is most likely that the few DEGs identified with a single gRNA were due either to spurious off-target binding of the SunTag complex or were randomly pulled out and further eliminated when pooling together the two gRNAs data by increasing the power of our statistical analysis. However, one of the gRNAs (LASER 1, gRNA2) produced a number of DEGs that largely exceeded what one would expect from random fluctuation of gene expression. This observation will be discussed in section IV.

Consequently, it seems that the overexpression of these three non-coding RNAs didn't result in the modification of expression of another transcript at the genome-wide level. However, this doesn't exclude the possibility that these lncRNAs might have a functional role at any level of gene regulation.

G. Additional lncRNAs candidates selection

Despite its unsuccessful outcome, our low-scale screening of LASER overexpression upon LIF withdrawal demonstrated the feasibility of our approach. This prompted us to select a higher number of candidates in order to be able to perform a similar screening at a higher magnitude. However, a wide list of lncRNAs candidates cannot be curated manually unless countless human and time resources. Therefore, a greater number of conditions and replicates was needed to greatly increase our confidence in the selected lncRNAs. We thus augmented our previous RNA datasets with loss of function experiments of Nanog, Esrrb and Oct4

transcription factors for 24 hours with the help of cell lines using Tet-ON or OFF systems **(Fig. 2.22 B)** (Hitoshi Niwa, Miyazaki, and Smith 2000; Festuccia et al. 2012, 2016).



B

Cell line/Gene	WT		Nanog					Esrrb		Oct4	
Condition/ Expression	FCS	2i	CTL	LOF	KO	CTL/GOF		CTL	LOF	CTL	LOF
						+LIF	-LIF				

Fig.2.22 A. New candidates selected on the basis of their expression profiles in our new RNA-seq data, this cluster was selected for the following pattern: down-regulation upon loss of Oct4 or Nanog and LIF-withdrawal. **B.** New biological samples analysed by RNA-seq with polyA selected libraries and about 30 million reads per sample. 2 to 3 replicates of each condition have been sequenced. LOF = loss of function; GOF = gain of function.

Remarkably, it appeared that all our LASER were fairly detectable in RNA-seq datasets performed on poly-adenylated RNAs (see examples: **Fig. 2.21**. **A**). We thus decided to restrict our analysis to this kind of transcripts and performed polyA selected RNA-seq providing increased coverage for spliced and mature RNA-pol II transcribed RNA at the expense of intronic and none poly-adenylated RNAs signal. Nanog SunTag dataset (see section II.) in which Nanog overexpression was performed in the presence and absence of LIF was included as well in our analysis.

A new high-confidence database of 3490 lncRNAs including all LASER, all non-coding transcripts identified by our collaborators (Igor Ulitsky laboratory) and all annotated lncRNAs in ENSEMBL mouse database supported by minimal coverage in a good fraction of our samples was generated by Nick Owens, a computational biologist in the laboratory. Thanks

to a statistical tool he set up for that aim, we already selected several clusters of lncRNAs showing relevant patterns of expression among all our conditions **(Fig. 2.22. A)**. These transcripts will be further screened for their capacity to provide LIF-independence.

H. LASER 23 (Gm14820) characterization

LASER 23 fairly fulfilled the criteria previously mentioned for lncRNAs candidates' selection **(Fig. 2.23)**. But it further captured our attention because of its antisense relationship with Suv39h1 gene in a 'tail-to-tail' orientation **(Fig. 2.24 A)**. Suv39h1 is a histone methyl transferase associated with the formation and the function of constitutive heterochromatin on genomic repeated elements such as satellites, tandem repeats or retroviruses but is also involved in the repression of lineage-inappropriate genes in the trophectoderm or S-phase associated genes in terminally differentiated cells (Firestein et al. 2000; Lehnertz et al. 2003; Ait- Si- Ali et al. 2004; Alder et al. 2010; Fritsch et al. 2010; Bulut-Karslioglu et al. 2014, 2012; Velazquez Camacho et al. 2017). It has been shown to recognize and propagate H3K9 trimethylation through its chromo and SET domains respectively and interact with other H3K9 methyltransferases as well as Hp1 proteins and RNAs to form and expand heterochromatin domains. The loss of Suv39h1 and 2 in mouse ES cells leads to the loss of H3K9me3 at pericentric regions but is associated with a gain in polycomb-mediated repressive histone mark H3K27me3 at the same places with no major functional consequences, a plasticity which seems also present at the zygotic stage (Peters et al. 2003; Puschendorf et al. 2008; Marks et al. 2012). In contrast, Suv39h1/2 depletion in somatic tissues leads to genomic instability inducing higher tumor risk and defects in male germ cells meiosis (Peters et al. 2001).

Interestingly, constitutive heterochromatin structure differs between male and female genomic material at the time of fertilization and persists during the first cleavages until the 8 cell stage (Arney et al. 2002; Santos et al. 2005; van der Heijden et al. 2006; Govin et al. 2007; Puschendorf et al. 2008). Indeed, in mice, at these stages, male chromosomes are fully deprived from Suv39h-mediated H3K9me3 repressive mark but rather enriched in PRC2-mediated H3K27me3 compared to female genome (H. Liu, Kim, and Aoki 2004; Puschendorf et al. 2008). This asymmetry is progressively lost with the formation of compact repressive hubs, the chromocenters, where centromeric and pericentromeric regions cluster together, and gradual trimethylation of H3K9 residues on the paternal genome (Mayer et al. 2000; Kuznetsova et al. 2016; Puschendorf et al. 2008). However, this transition is tightly regulated and is critical for the establishment of proper nuclear structure later on (Burton and Torres-

Padilla 2010; Nakamura et al. 2012; De La Fuente, Baumann, and Viveiros 2015; Ma et al. 2015). Thus, appropriate expression of heterochromatin organizers as Suv39h1 seems to be critical in these very early steps of development.

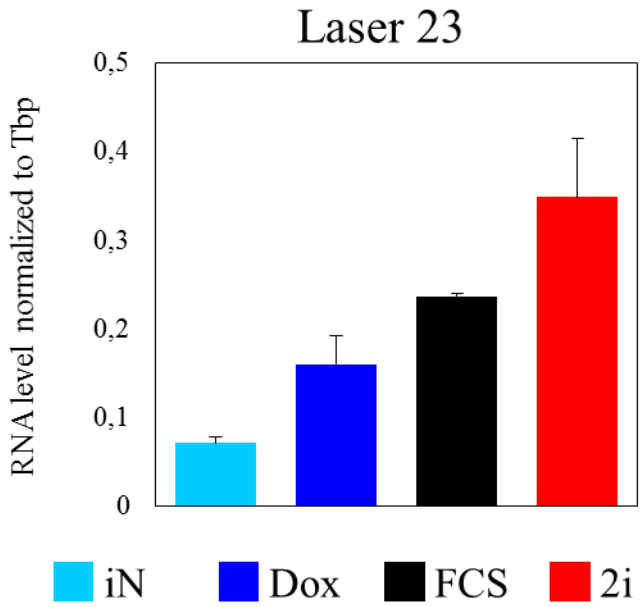


Fig. 2.23. RT-qPCR measure of RNA expression of LASER 23 in the four aforementioned samples used for LASER selection. LASER 23 shows both response to Nanog expression as well as pluripotency ground state attainment. Values are normalized to Tbp level (n=2, Mean ± SEM).

Given the known ability of antisense lncRNAs to regulate the expression of their opposite gene (Magistri et al. 2012), we examined the expression of Laser23 and Suv39h1 in undifferentiated and differentiating ES cells. It appeared that both expressions were negatively correlated **(Fig. 2.24 B)**. While LASER 23 was strongly repressed two days after cells being cultured in the absence of LIF or 24 hours after Oct4 depletion, Suv39h1 precisely showed the reverse pattern. We thus wondered whether these opposite expression profiles could also be found *in vivo* during early development, when heterochromatin organisation faces dramatic rearrangement. We therefore looked at single cell RNA-seq data (Q. Deng et al. 2014) performed from the zygote to the blastocyst stage of mouse embryo development (with help of Nick Owens). It appears that LASER 23 is already expressed at the zygotic stage before peaking at the early 2 cell stage and progressively decreases until blastocyst stage where only very few cells keep residual expression. Suv39h1 is oppositely getting highly induced at the mid 2 cell stage to be intermediately expressed throughout the following steps of development

(Fig. 2.24 C). Therefore, LASER 23 and Suv39h1 seem to have mutually exclusive expression profiles which prompted us to dig into the relationship between these two transcripts.

We then explored LASER 23 conservation between mouse and human as this has been shown to be a good prediction factor for functionality (Ulitsky 2016). It appeared that a similarly structured transcript was present in the syntenic region of the human genome (AF196970) annotated as an lncRNA. It correspondingly initiates downstream of human SUV39H1 in the opposite direction and also shows a large overlap with SUV39H1 exon 3. Sequence conservation between LASER 23 and AF196970 is only restricted to their overlapping region with suv39h1/SUV39H1 exon 3, therefore not exhibiting a real sequence similarity for the non-coding transcripts. This might indicate that the act of transcription rather than the sequence of the antisense RNA is important for the regulation of this locus. We further evaluated the coding potential of both human and mouse lncRNAs with the CPAT tool (L. Wang et al. 2013) and obtained very low coding probabilities of 13 and 1% for human and mouse transcripts respectively indicating that in both species the antisense transcript to suv39h1/SUV39H1 most likely doesn't have coding functions.

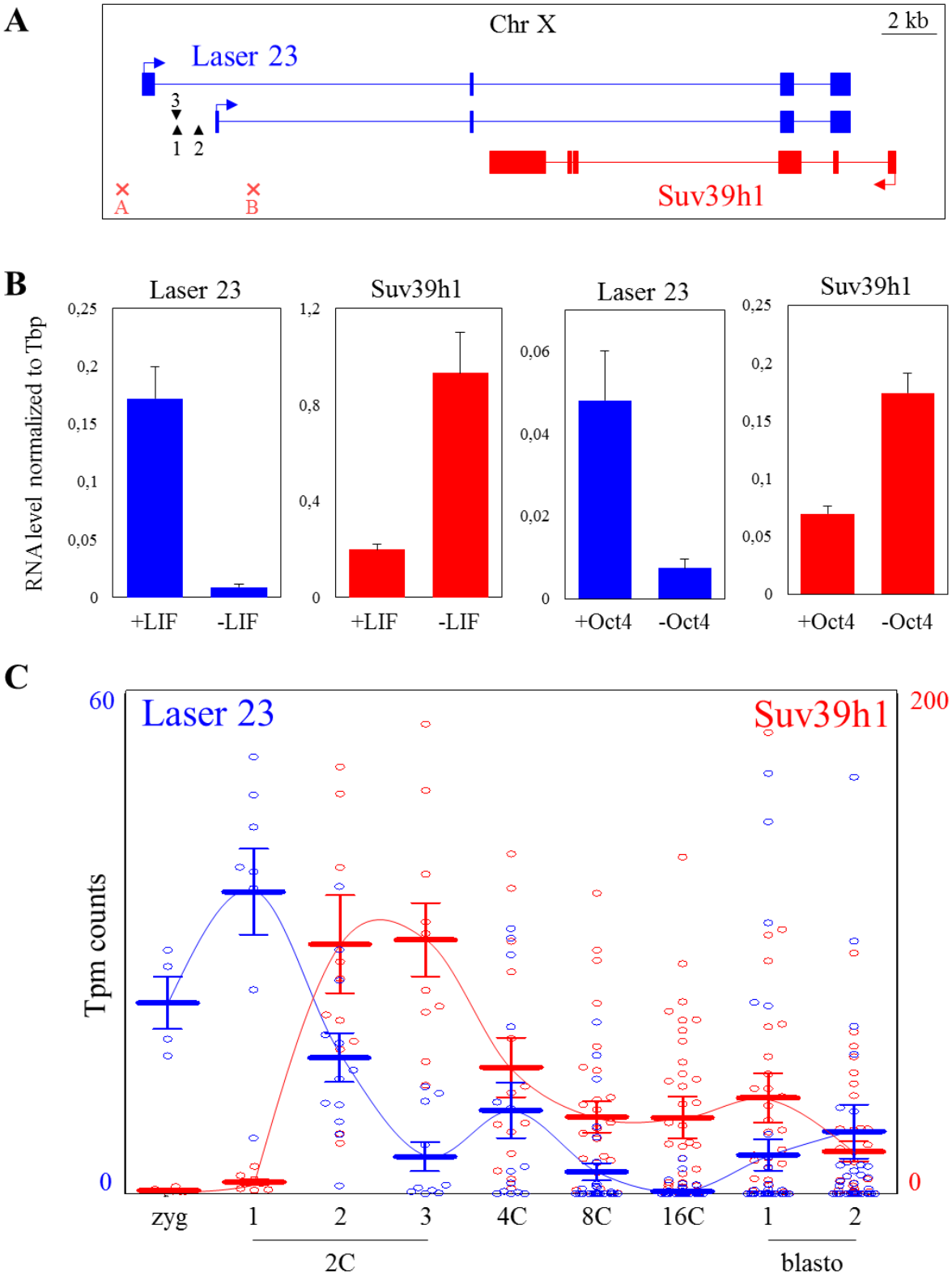


Fig. 2.24. **A.** Schematic representation of LASER 23/Suv39h1 locus. LASER 23 shows two alternative promoters both active in mouse ES cells. gRNAs used for SunTag overexpression are represented as black arrowheads. gRNAs used for CRISPR/Cas9 deletion are represented by red crosses. **B.** RT-qPCR measure of LASER 23 and Suv39h1 RNAs expression in E14 WT cells cultured for 2 days in presence or absence of LIF (n=6) and in Zhbtc4 cells treated (Oct4-

) or not (Oct4+) with Dox for 12 hours (n=2) (Hitoshi Niwa, Miyazaki, and Smith 2000) Values are normalized to Tbp level (Mean ± SEM). **C.** Single cell RNA-seq expression during mouse embryo early development, from Deng et al. 2014. (zyg = zygote, 2C, 4C, 8C, 16C = 2, 4, 8 and 16 cells stage, blasto = blastocyst, 1 = early, 2 = middle, 3 = late). Tpm counts are shown for LASER 23 (left) and Suv39h1 (right).

To assess whether the opposite expression profile observed in the early embryo also exists in mouse ES cells at the single cell level, we performed single molecule RNA-FISH on LASER 23 and Suv39h1 transcripts (non-overlapping exonic probes) in WT male ES cells **(Fig. 2.25)**. LASER 23 molecules could be detected at its transcription site but also as few diffusible molecules in the nucleus while Suv39h1 RNAs were detected both in the cytoplasm and the nucleus. We quantified the relative distribution of both RNAs with a collaborator (F. Mueller, Institut Pasteur) using FISHQuant program (Mueller et al. 2013). This clearly showed a negative correlation between the distribution of Suv39h1 and its antisense since cells that are high for one transcript are low for the other **(Fig. 2.25)**. However, few hypothetically co-transcriptional events were detected, maybe suggesting a model where transcription can initiate from both promoters, but one direction rapidly overtakes.

We consequently decided to actively interfere with the transcriptional activity of the locus in order to assess the functional relevance of LASER 23 expression. Since the first gRNA designed towards one of its two alternative promoters didn't work in our previous study **(Fig. 2.18)**, we designed three additional ones targeting its second promoter **(Fig. 2.24. A)**. It is worth mentioning that both promoters seem to be equally active in mouse ES cells according to RNA-seq data (not shown). We therefore transfected these three gRNAs in our SunTag clones and picked individual subclones harbouring each of them. Cells were then treated with Dox for 2 days followed by RNA extraction. This experiment was performed both in presence and absence of LIF as Suv39h1 was shown to be upregulated upon LIF withdrawal **(Fig. 2.24. B)**. The induction of LASSER 23 was moderate in +LIF condition but reached very low level in – LIF where it naturally gets strongly down-regulated probably due to the high content in ERVK elements of its promoter (as seen for AK031828 in section I) **(Fig. 2.26)**. However, we didn't assess yet the consequences of the recruitment of the SunTag complex within the first intron of the longest isoform of LASER 23 on its expression. Indeed, we cannot exclude a subsequent transcriptional interference which would lessen the impact of our assay. Nevertheless, this low overexpression of LASER 23 had no significant impact on Suv39h1 expression **(Fig. 2.26)**.

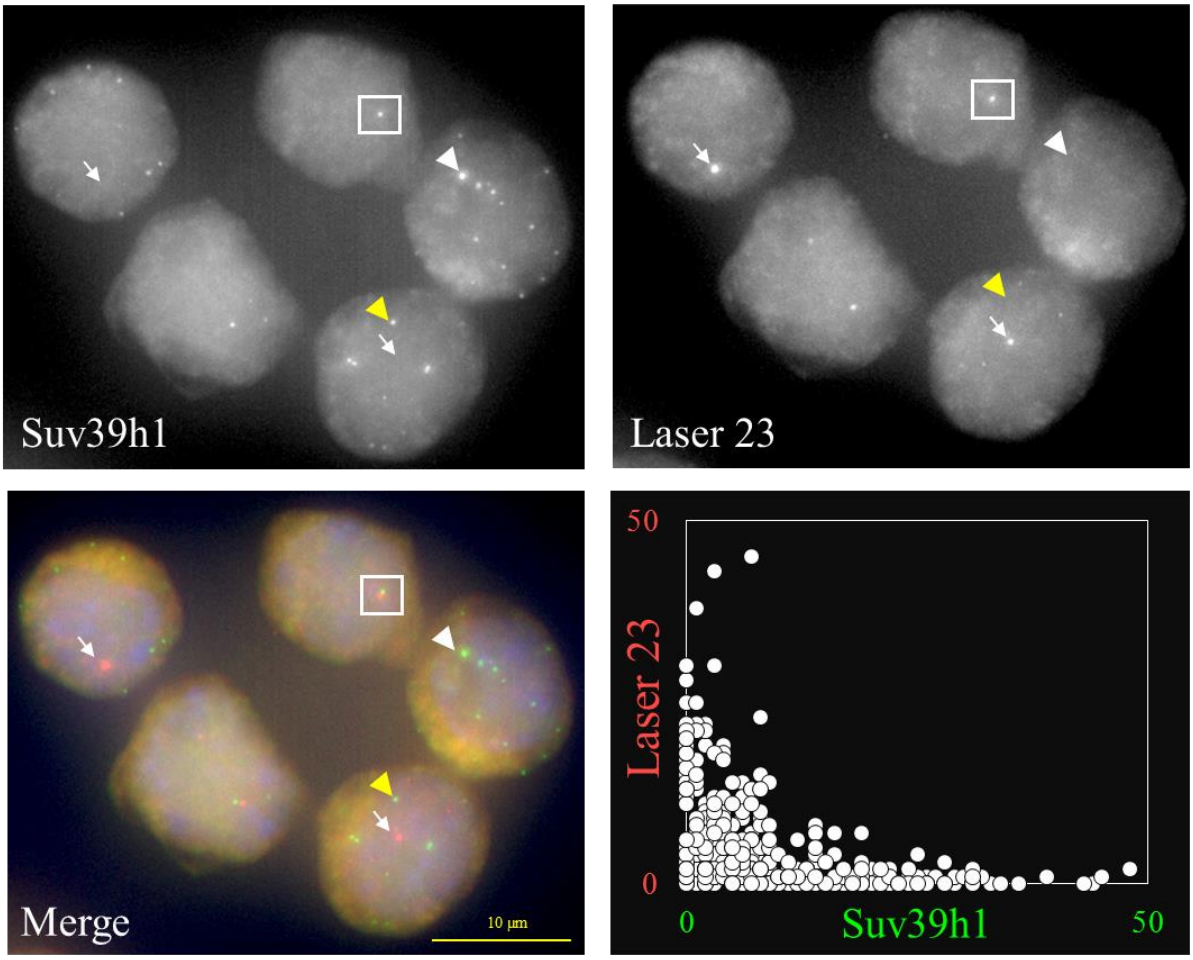


Fig. 2.25. Single molecule-RNA-FISH of Suv39h1 and LASER 23 in FCS/LIF WT male ES cells. Representative pictures of Suv39h1 (green) and LASER 23 (red) transcripts in single cells. Transcriptional sites for Suv39h1 (white arrowhead), LASER 23 (white arrow) and potential co-transcriptional sites (white square) could be identified as well as diffusing Suv39h1 single RNA molecule (yellow arrow head). Bottom right: detected single RNA molecules of Suv39h1 and Laser 23 in single cells quantified by FISHQuant pipeline.

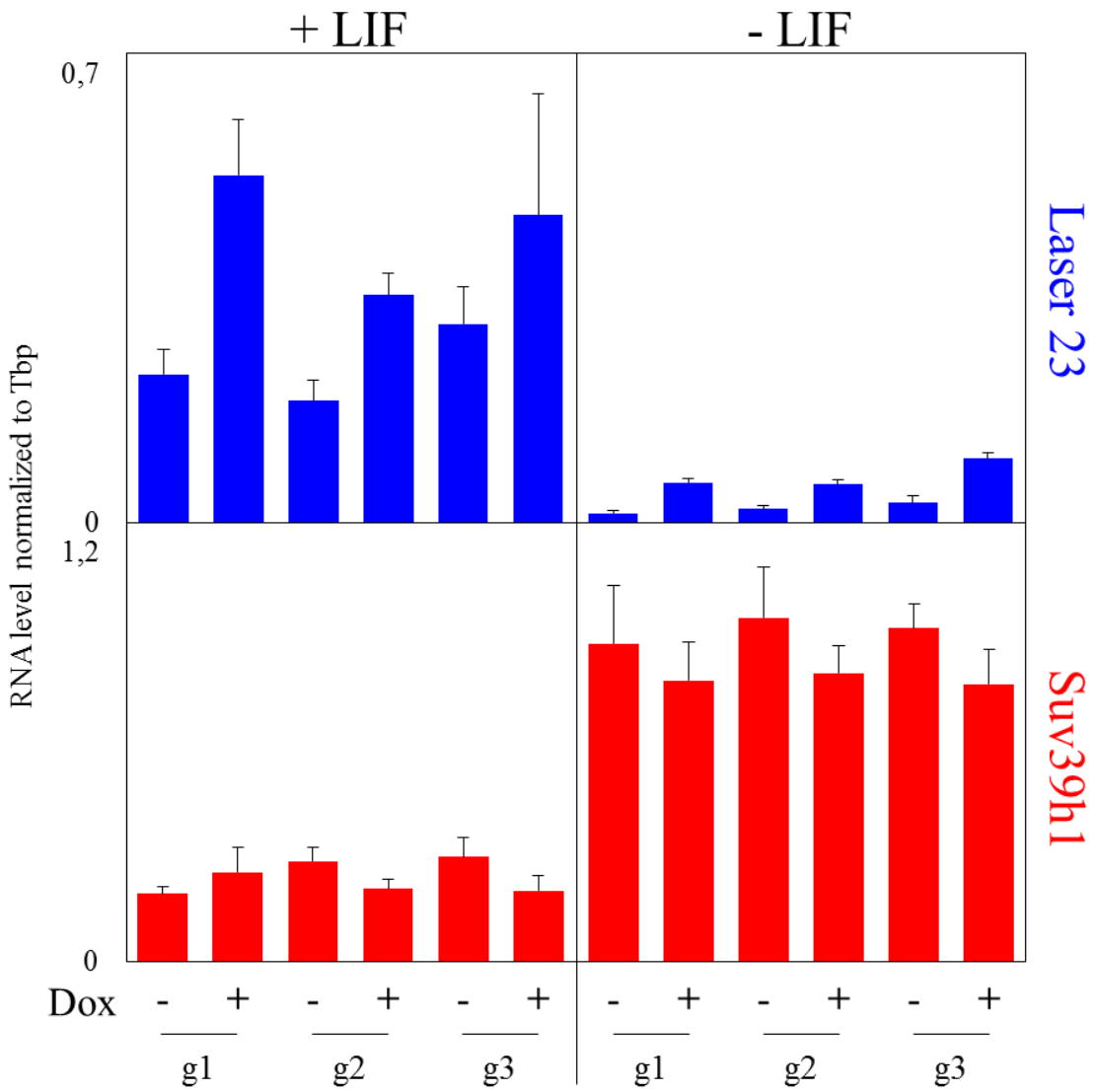


Fig. 2.26. RT-qPCR measure of LASER 23 and Suv39h1 RNA expression during SunTag activation of LASER 23 in the presence and absence of LIF with the three gRNAs previously depicted (Fig. 2.22). Two clones per gRNA originating each from one of the 2 SunTag clones. (n=4, Mean ± SEM).

We further assessed the effect of the depletion of LASER 23 on Suv39h1 expression. In that aim, we used CRISPR/Cas9 system to delete the two alternative promoters of LASER 23. Two gRNAs were subsequently designed at each side of these two promoters potentially generating a 6 kb-long deletion **(Fig. 2.24 A)**. This experiment was performed in male E14 WT ES cells and several KO clones were obtained with no noticeable phenotype. The absence of residual expression of LASER 23 in the KO cells was confirmed by smFISH while Suv39h1 signal looked more homogenous among the cell population with a slightly increased number of detected dots **(Fig. 2.27 A)**. However, this still needs to be properly quantified. We further

evaluated Suv39h1 and LASER 23 expression by RT-qPCR **(Fig. 2.27 C)**. While LASER 23 RNA could not be detected at all in our KO clones, it appeared that Suv39h1 expression was increased in the same cells. This was particularly true in 2i medium where we could confirm this result at the protein level **(Fig. 2.27 B)**.

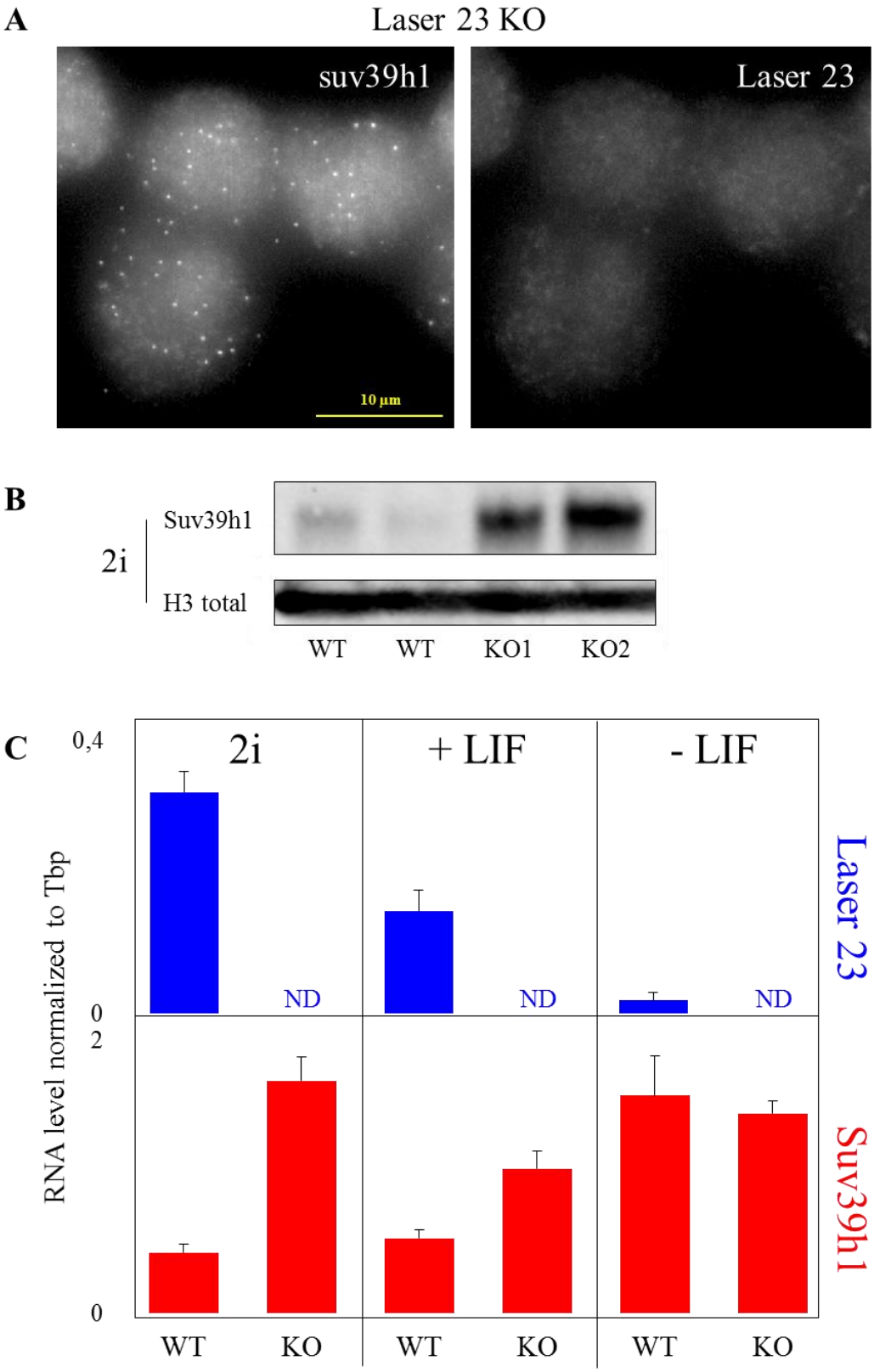


Fig. 2.27. **A.** Single molecule RNA FISH of Suv39h1 (left) and LASER 23 (right) RNAs in one of our KO clones. Only background signal can be observed in LASER 23 channel. **B.** Western Blot for Suv39h1 and total-H3 loading control performed with lysates of WT cells (left) and two KO clones (right) after 2 passages in 2i medium. **C.** RT-qPCR measure of LASER 23 and Suv39h1 RNAs expression level in WT cells and our KO clones in 2i medium (2 passages), FCS/LIF and after 2 days of LIF withdrawal, ND=non-detectable (n=4, Mean ± SEM).

Taken together, all these results suggest a repressive effect of LASER 23 transcription, associated with the undifferentiated state of ES cells and the first cleavage state of the embryo, on Suv39h1 expression. However, many experiments are still to be done in order to assess the precise mechanism of this inhibition.

I. Discussion

Our preliminary approach allowed us to select a small number of lncRNAs whose expression is enhanced in the ground state of pluripotency and associated with Nanog activity. We now greatly extended our capacity of selection by generating a large amount of sequencing samples representing a challenging framework for the pluripotency network. We additionally built a comprehensive list of high confidence lncRNAs expressed in mouse pluripotent cells. We further identified hundreds of transcripts responsive to drastic variations of expression of the pluripotency factors (Nanog, Esrrb, Oct4) as well as signalling pathways inhibition (LIF, MEK/ERK, GSK3β). These reliable lists of lncRNAs can now be functionally investigated using CRISPRa at a higher scale. This will more likely require the design and cloning of more than two gRNAs per promoter given the high rate of inefficient ones (close to 50% in our hands). We additionally generated a gRNA expressing vector inserted within a PiggyBac transposon which allows for efficient transfection and integration of large size libraries of vectors with single integrant per cell (G. Guo et al. 2009; G. Guo and Smith 2010). Hence, a library of thousands of gRNAs can be properly delivered to our SunTag clones followed by serial passages in the absence of LIF, a condition that would be less stringent than a plating at clonal density where the LIF can be properly washed out, but allows for the emergence of clones with enhanced proliferative capacities in the absence of the cytokine. Downstream identification of the overrepresented gRNAs in the self-renewing cell population by DNA sequencing would lead to the characterization of lncRNAs conferring reduced requirement to LIF signalling. Such kind of phenotypic advantage-based CRISPRa screening already gave

quite promising results (Gilbert et al. 2014; Konermann et al. 2015; Joung et al. 2017; S. J. Liu et al. 2017).

Indeed, the screening that we performed on the 18 LASER where cells were placed at low density in the absence of LIF is relatively stringent if we consider that a powerful transcription factor such as Klf4 (one of the Yamanaka factors) has been reported to confer variable self-renewal sustaining capacities according to different studies (J. Jiang et al. 2008, 4; Hitoshi Niwa et al. 2009; Hall et al. 2009b; P. Zhang et al. 2010; Martello, Bertone, and Smith 2013; Jeon et al. 2016; Yamane et al. 2018). Together with the previously mentioned idea that lncRNAs might mainly confer fine-tuning regulatory functions on gene expression, it seems reasonable to reduce the selective strength of our strategy. We could as well screen for a synergistic effect of the upregulation of our lncRNAs with one of the ancillary pluripotency factors in shielding mouse ES cells against differentiation. This could be achieved by creating a basal vector harbouring a gRNA targeting one of these pluripotency factors that would then serve to build a large library of vectors by adding gRNAs targeting the lncRNAs candidates.

It was shown in section II than Nanog confers LIF-independent self-renewal through, among other mechanisms, maintaining H3K27me3 level at many repressed loci by an indirect mechanism. Moreover, this effect is clearly specific to a subset of genomic locations but doesn't seem to be related to Nanog bound regions. So, how are all these specific loci recognized by the PRC2 complex for proper repressive mark deposition? PRC2 complex recruitment in mouse ES cells has been extensively linked with its RNA interacting property (J. Zhao et al. 2010; Kaneko et al. 2013; Bonasio et al. 2014; Kaneko, Bonasio, et al. 2014b, 2014b). We can therefore hypothesize that one or several lncRNAs that get activated by Nanog may drive the PRC2 complex to specific positions for repressive state maintenance (S. C. Wu, Kallin, and Zhang 2010; Kaneko, Bonasio, et al. 2014b; Ng, Johnson, and Stanton 2012). Such molecule should be potentially identified by the analysis of the group of our lncRNAs that get strongly induced by Nanog in the absence of LIF. Additionally, we can take advantage of the different ChIP-seq clusters identified in section II to refine our selection to identify direct and truly dependent targets of Nanog function.

In contrast, LASER 1, 10 and 17 differently respond to Nanog induction but were all upregulated when ES cells were placed in 2i medium. Their low level of expression in serum/LIF culture led us to speculate that they might be expressed only in a small portion of the cells and that their strong increase in 2i medium may correlate with the higher and more

homogenous expression of some pluripotency factors appearing in this condition. We therefore hypothesised that increasing their expression in every cell in serum/LIF would recapitulate their potential effect in the ground state of pluripotency. Surprisingly, none of our three LASER overexpressions did produce any modification of gene expression at the transcriptome-wide level. Therefore, wouldn't it be more relevant to assess the effect of their depletion in 2i medium? Not really, supposing the same redundancy exists within lncRNAs than among the ancillary pluripotency factors, especially when considering the milder phenotypes usually resulting from lncRNAs depletion compared to mRNA (Nakagawa 2016). Additionally, while small families of protein were identified based on sequence homologies and shared functional domains and led to simultaneous inactivation studies unmasking some redundancies (J. Jiang et al. 2008), a functional classification of lncRNAs is still inexistent. Therefore, it seems more promising to assess the effect of their ectopic expression in a situation where they are not or very lowly expressed assuming that their potential redundancy with other lncRNAs is there inexistent. Considering the effects of Nanog overexpression in presence and absence of LIF (section II), this consideration seems even more relevant.

The remarkable technological improvements recently achieved in the RNA biology field bring a hope for and will surely greatly influence lncRNAs classification and characterization. In particular in regards to the identification of the partners (proteins, other RNAs) (Aw et al. 2016; Lu et al. 2016; McHugh and Guttman 2018), the structure (Denny et al. 2018) and the cellular localization (Abudayyeh et al. 2017) of these transcripts which represent essential phases for their unbiased description. Indeed, some studies focus on the interaction between a lncRNA of interest and a given protein, which was sometimes never shown to have RNA interacting properties, for quite obscure reasons (Jiapaer et al. 2018). One could therefore anticipate a big qualitative step towards the understanding of this still mysterious class of transcripts in the next decade.

Nonetheless, LASER23 relationship with Suv39h1 seems to be on the way of being characterized. We indeed accumulated many clues suggesting a repressive effect of LASER 23 upon its opposite coding gene. However, this still needs proper confirmation as well as a mechanistic explanation. Indeed, we plan to validate the upregulation of Suv39h1 in the absence of LASER 23 through other experimental approaches. First, a strong polyA signal will be knocked-in the first intron of LASER 23 transcript to separate a possible regulatory function of the underlying genomic sequence including LASER 23 transcription initiation sites from the RNA *per se* or its transcription. Histone modifications along its locus as well as splicing

efficiency will be evaluated for Suv39h1 in the context of truncated (polyA signal), absent (KO) and WT LASER 23 transcript. The same experiments will be repeated in the context of a strong induction of LASER 23 expression upon knock-in of a potent constitutive promoter upstream of its TSS. Finally, to decipher whether any regulatory effect is mediated through the specific tethering of LASER 23 RNA to Suv39h1 locus, the overexpression from an exogenous cDNA vector will be performed in WT and KO cells. Indeed, the few diffusible molecules that could be observed by sm-RNA-FISH and the accumulation at its transcription site rather suggest a locally mediated effect of LASER 23.

Furthermore, the downstream effects of this potential repression in mouse ES cells will need to be addressed. Given the demonstrated functions of Suv39h1, a particular attention should be placed on the expression of repeated sequences, as retrotransposons and satellites, as well as on the chromatin modifications at those loci. This will be performed upon, acute or prolonged, induction or repression of LASER 23 in order to assess the potential dynamics of any effect as short and long term compensatory mechanisms exist in the silencing of repeated elements in mouse pluripotent cells (Peters et al. 2003; Berrens et al. 2017). Moreover, Suv39h1 has also been shown to play an important role in the silencing of bivalent promoters during trophoblast stem like (TSL) cells differentiation (Alder et al. 2010). We previously showed that LASER 23 is strongly repressed upon this directed differentiation in zhbtc4 cells **(Fig. 2.24 B)** (Hitoshi Niwa, Miyazaki, and Smith 2000). Therefore, we can speculate that the ectopic expression of LASER 23 in that context would impair the proper establishment of a trophoblastic identity through the repression of Suv39h1, an experimental approach that is planned to be done.

Finally, a major aspect of LASER 23 characterization might be coming from the early embryo development. Indeed, the expression of LASER 23 was shown to be even higher prior or towards fertilization than in the following stages of zygotic development and cleavages **(Fig. 2.28)** (Karlic et al. 2017), therefore at the moment where Suv39h1 has been shown to be totally absent (Puschendorf et al. 2008). One hypothesis would be that the strong expression of LASER 23 in the mouse oocyte allows for robust and long-lasting repression of Suv39h1 until the 2-cell-stage assuring the absence of its expression and downstream function towards fertilization where paternal deprivation of H3K9me3 mark is precociously maintained. Consequently, assessing the effect of LASER 23 deprivation in oocytes on zygotic development and early development chromatin structure would represent an essential experiment to question the function of this lncRNA *in vivo*. However, such investigation would

require LASER 23 KO female mice to be viable or the generation of a new model of tissue-(oocyte) specific depletion of LASER 23 through Cre-mediated, tissue-specific, recombination (Lan, Xu, and Cooney 2004).

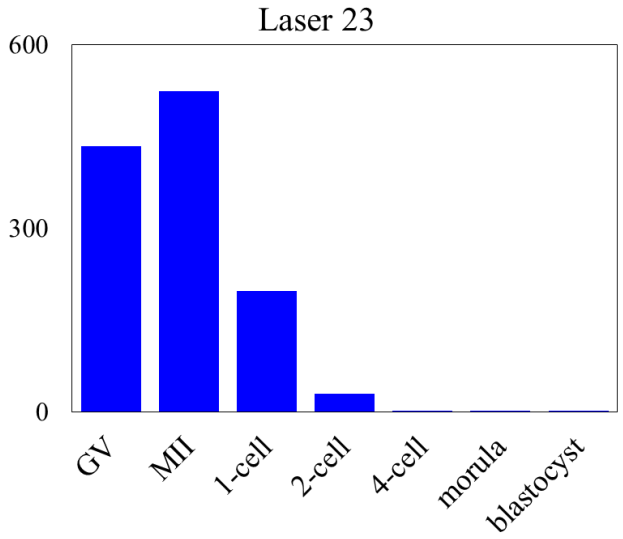


Fig. 2.28. RNA-seq expression profile of LASER 23 in oocytes and early developing mouse embryo from Karlic et al. 2017 GV=germinal vesicle, MII= meiosis II oocytes (Y-scale represents FPKM).

X. A serendipity-driven approach

A. Introduction

The 2 cell-like state has been described as a rare population of cells arising from mouse ES cells culture showing a transcriptomic profile resembling the first cleavage stage of the embryo including expression of the retroviral MERVL elements and the Zscan4 protein family members (Macfarlan et al. 2012). This state has also been associated with numerous additional transient features as rapid telomere elongation, genome-wide chromatin decompaction and DNA demethylation, dispersed chromocenters, histones hyper acetylation, global translation block and Oct4 protein depletion as well as extended potency (Zalzman et al. 2010; Macfarlan et al. 2012; Dan et al. 2013; Ishiuchi et al. 2015; Eckersley-Maslin et al. 2016; Choi et al. 2017). It was additionally shown that every mouse ES cells is likely to undergo such event every 9 passages (18-27 days) (Zalzman et al. 2010). Different ways have been discovered to artificially enhance the entry in this specific state of mouse ES cells: Dux and Zscan4 induction, CAF-1 complex or LINE1 RNAs knock down or mir34a depletion (Zalzman et al. 2010; Ishiuchi et al. 2015; Choi et al. 2017; Hendrickson et al. 2017; Percharde et al. 2018). However, the natural mechanism, if unique (Rodriguez-Terrones et al. 2018), leading to this original state has not been clarified.

B. An unexpected 2 cell-like state induction

As exposed in the previous section, an RNA-seq experiment was performed under induction of LASER 1, 10 and 17. No differentially expressed gene (DEG) could be identified showing a consistent response between the two tested gRNAs for each lncRNA. Yet, while one of the gRNA (gRNA 1) targeting the promoter of LASER 1 only led to the identification of few DEGs, it appeared that activation of the SunTag system combined with the expression of the second one (gRNA 2) produces a very large number of responsive genes. Indeed, 900 DEGs could be identified by a DESEQ analysis ($-1<log2FC<1$, $padj<0.05$) (performed by Olivier Piau). These 900 DEGs included 742 upregulated and 158 downregulated genes. We rapidly understood that this list of genes strongly resembled the typical subset of transcripts induced upon the 2 cell-like state as the most famous markers of it (Zscan4 family, Zfp352, Eif1a-like family) were among the strongest induced genes. In addition, activation of many ERVL repeat elements could be easily visualised on the sequencing data **(Fig. 2.29)**.

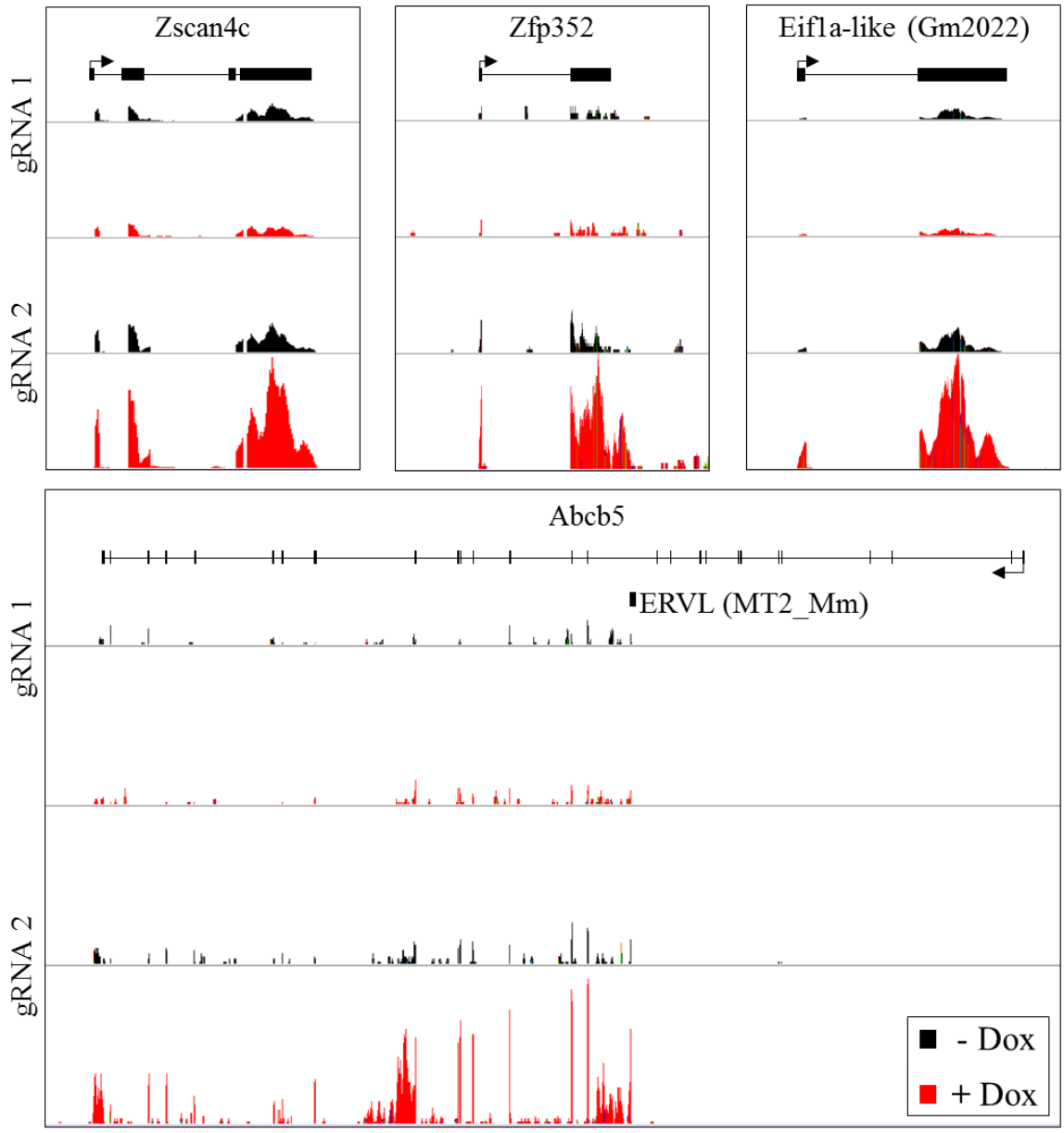


Fig. 2.29. Screenshots of IGV browser showing RNA-seq coverage obtained before (-Dox) and after SunTag induction (+Dox) with the two different gRNAs targeting LASER 1 promoter. Strong induction of Zscan4d, Zfp352, Gm2022 as well as a chimeric transcript of Abcb5 gene initiating in an ERVL element can be seen.

When comparing these results with published datasets dealing with the induction of the 2 cell-like state (Ishiuchi et al. 2015; Eckersley-Maslin et al. 2016; Hendrickson et al. 2017), it appeared that half (434) of our DEGs have been shown to be misregulated in at least one of this three studies **(Fig. 2.30)**. In addition, about 50% of the DEGs identified in two of these studies (Eckersley-Maslin et al. 2016; Hendrickson et al. 2017) were included in our dataset. Among the 38 genes that were shared between all these studies and us, Zscan4 family genes

and Zfp352 could be found. The range of misregulated genes in our case was not particularly high if comparing to CAF-1 complex depletion study (Ishiuchi et al. 2015) in which the use of two different shRNAs led to 2517 or 1676 upregulated (1498 common) and 96 or 31 downregulated genes respectively (15 common). We further validated some of our DEGs by RT-qPCR **(Fig. 2.31 A).**

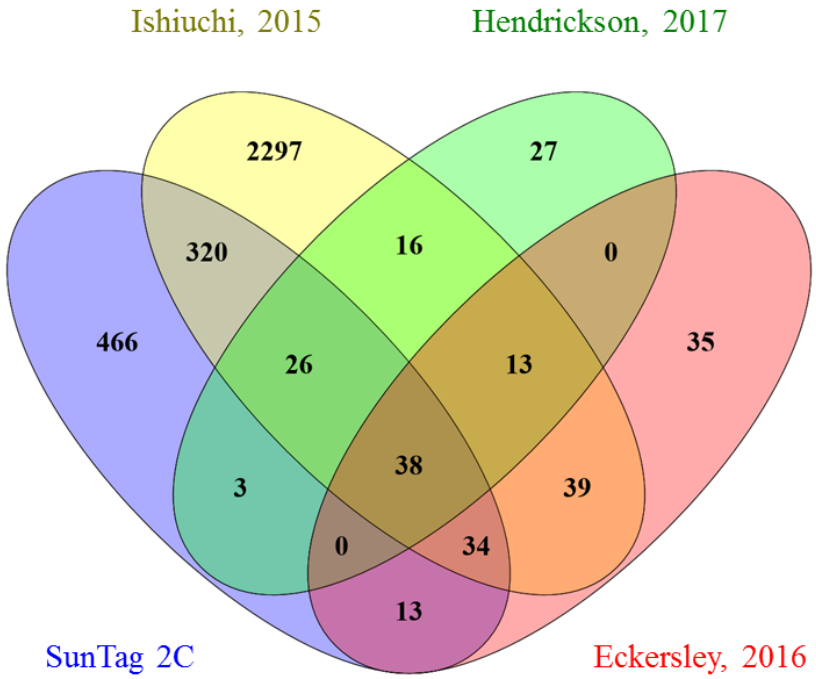


Fig. 2.30. Differentially expressed genes overlap between our genes list (SunTag 2C) and other genome-wide published datasets reporting induction of the 2-cell like state (Ishiuchi et al. 2015; Hendrickson et al. 2017; Eckersley-Maslin et al. 2016).

We further performed gene set analysis with the lists of our upregulated and downregulated genes separately using the ESCAPE online database containing many functional studies performed in pluripotent cells. It revealed that our upregulated genes were highly enriched in genes shown to be upregulated by Gata3 (padj 3.10^{-59}) and Zscan4c (padj 6.10^{-43}) transcription factors overexpression in mouse ES cells (Nishiyama et al. 2009) and that our downregulated genes were strongly enriched in Gata3 (padj 3.10^{-54}) downregulated genes in the same study. In fact, when digging into these genes lists, it appears that half of the misregulated genes obtained by Nishiyama et al. 2009 under Zscan4c overexpression are shared with Gata3. Interestingly, Gata3 binding site motif has been shown to be highly enriched in open chromatin regions in single cell at the 2 cell stage of mouse embryo development (F. Guo et al. 2017). In addition, Gata2 and 3 have been shown to have an identical binding motif

in human ESCs (Krendl et al. 2017) and share redundant functions and targets during mouse pre-implantation embryo development (Home et al. 2017). Moreover, Gata2 de-repression has been shown to strongly induce 2 cell-like state features in mouse ES cells (Choi et al. 2017). Consequently, it is very likely that a 2 cell-like state gene expression signature was induced upon Gata3 increase by Nishiyama et al. 2009 explaining the similarity of this dataset with ours. Of note, while Gata3 is not included in our list of DEGs and doesn't show any significant change in our datasets, Gata2 is upregulated upon induction of SunTag expression with LASER 1 gRNA 2 **(Fig. 2.31. B)**.

As a result, the set of DEGs obtained in our dataset is comparable with the list of the genes affected in the two cell-like state of ES cells.

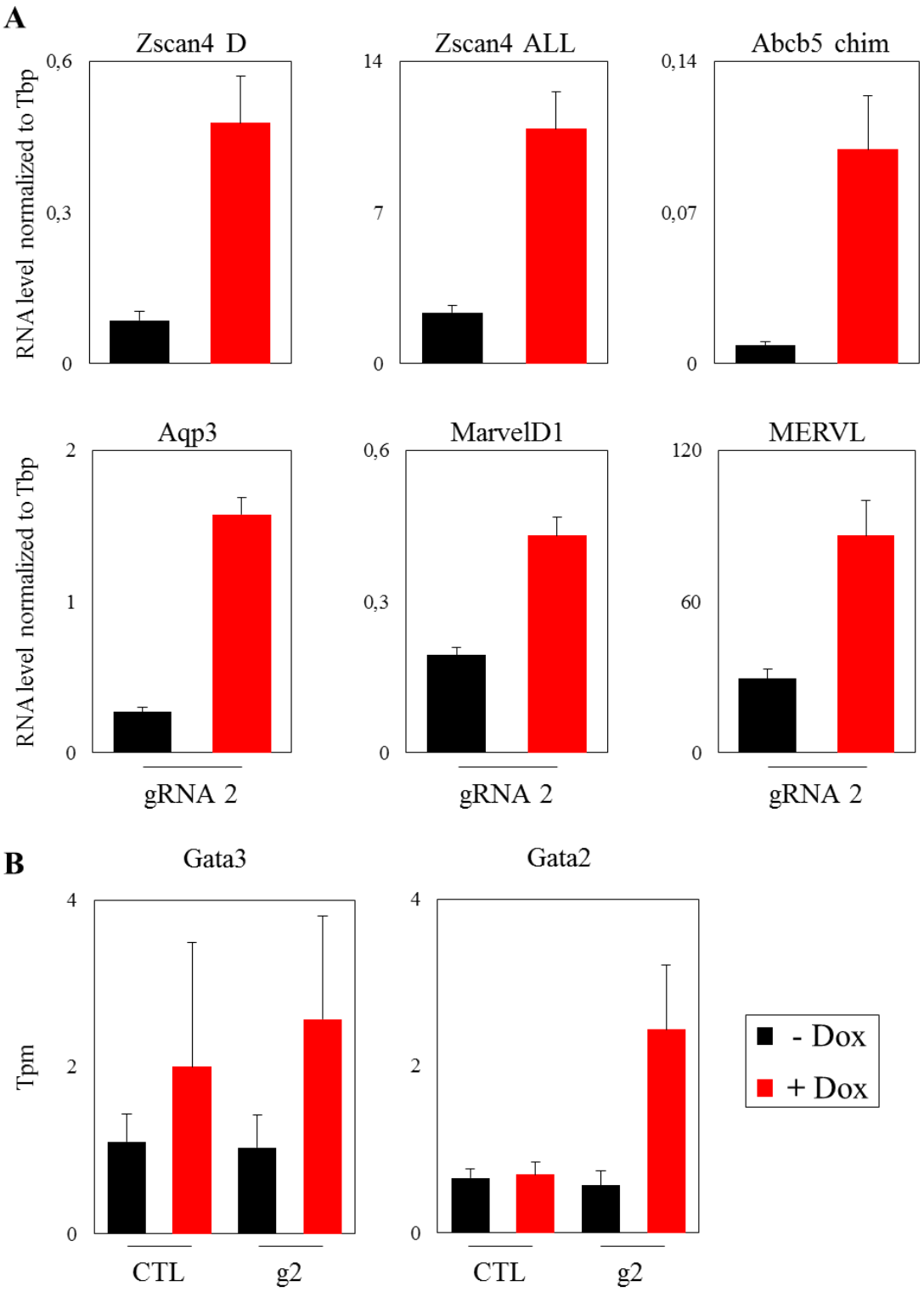


Fig. 2.31. **A.** RT-qPCR validation of differentially expressed genes expression (n=4, Mean ± SEM). **B.** RNA-seq quantification in tpm of Gata2 and 3 expression. CTL corresponds to LacZ non targeting gRNA (n=2, Mean ± SEM).

C. First hypothesis: a LASER 1-mediated effect

Although the hypothesis of an off-target effect induced by the gRNA 2 was the most compelling one, we first wanted to validate or eliminate the possibility that such an effect could be mediated by LASER 1 induction. Then, one possibility explaining the discrepancy between the two gRNAs effect could result from the expression of a different form of LASER 1 (alternative splicing, transcription start site…) under recruitment of the SunTag complex through gRNA 1 or 2. Indeed, a slight modification in the RNA sequence could potentially provide novel functions through many different ways: apparition of a new open reading frame, modified localization, new partners/targets.

We assessed the structure of the induced transcript by RT-qPCR **(Fig. 2.32 A and B)**. The closest forward primer to the TSS of LASER 1 showed transcriptional induction only with gRNA 2. This revealed that the recruitment of SunTag mediated by the two gRNAs led to two different sites of transcription initiation. Indeed, gRNA 1 was located near by the annotated TSS of LASER 1 and it is likely that dCas9 binding prevented its accurate transcription initiation to happen. This might also explain the lower upregulation obtained with gRNA 1 compared to gRNA 2. However, downstream splicing of the transcript didn't seem to be affected. Thus, we wondered whether this different sequence content could impact cellular localization of Laser 1. We performed cellular fractionation experiment under Dox treatment with both gRNA but neither of the primary transcript or the spliced RNA localization seemed to be affected by this different transcription start sites **(Fig. 2.33 A)**.

We finally assessed the consequence of the induction of LASER 1 with two novel gRNAs located around gRNA 2 targeting site **(Fig. 2.32 A).** These two new gRNAs (3 and 4) were cloned into expression vectors and integrated in parallel with the two previous ones (1 and 2) in our two SunTag clones. gRNA 4 appeared to be a little bit more efficient than gRNA 2 while the other new gRNA didn't produce any activation (gRNA 3). Since gRNA 4 was almost exactly antisense to gRNA 2 it was very unlikely that they would lead to any difference in LASER 1 transcription initiation site. Finally, while induction of the SunTag system in gRNA 2 containing cells led to a slight increase of Zscan4d expression, none of the other gRNAs showed the same effect **(Fig. 2.33 B)**.

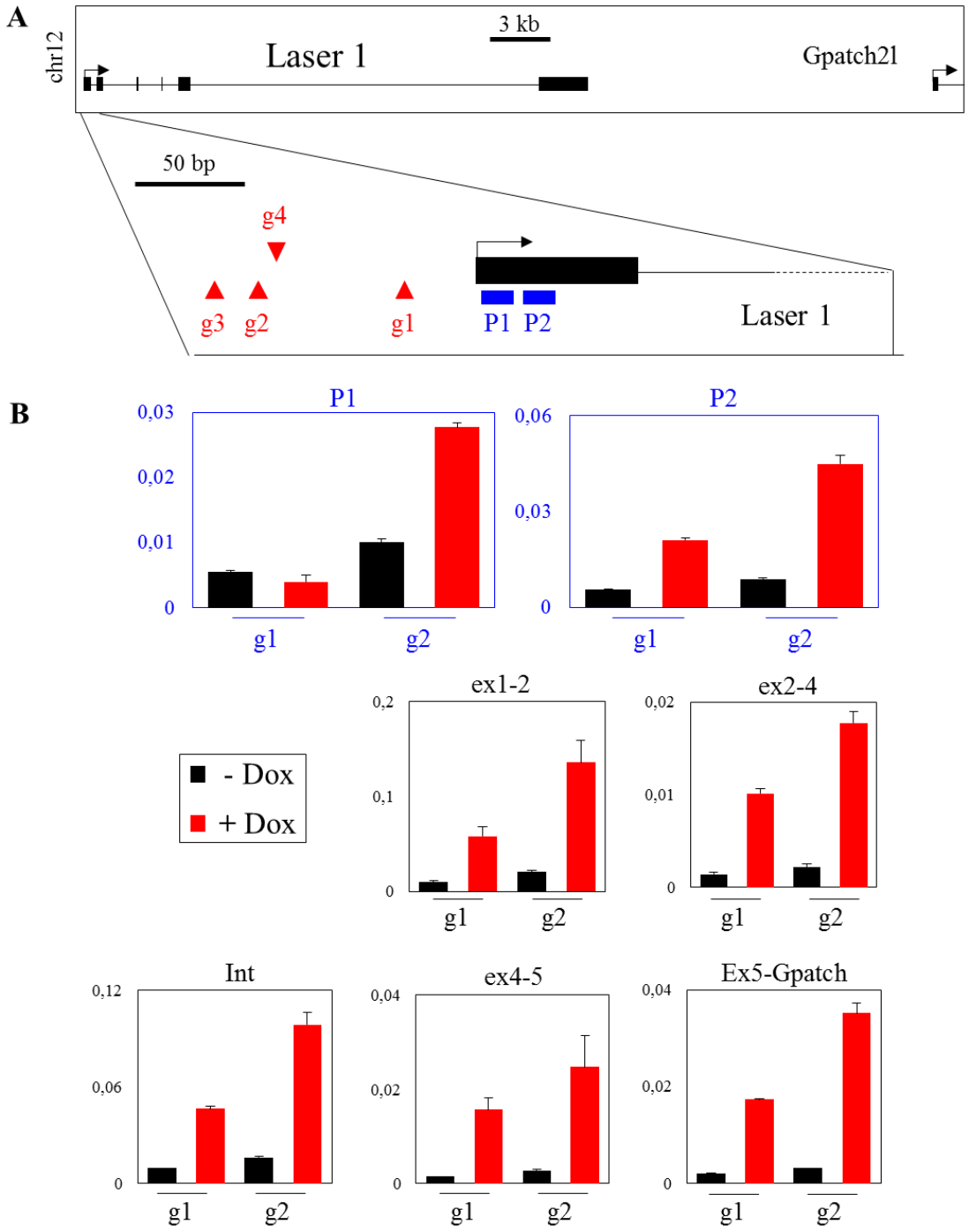


Fig. 2.32. **A.** Schematic representation of LASER 1 locus. The different gRNAs targeting sites are shown with red arrowheads. Forward primers for TSS mapping are represented in blue on the first LASER 1 exon. **B.** RT-qPCR measure of LASER 1 expression upon gRNA 1 and 2 induction. P1 and P2 correspond to the primer depicted in **A.** with a reverse primer lying within LASER 1 exon 2. ex1-2 = exons 1-2 spanning primers pair (identical for ex2-4 and ex4-5). Int = intronic primers located in intron 2. Ex5-Gpacth corresponds to a rare splicing detected between the last exon of LASER 1 and the second exon of its adjacent gene Gaptch2l. Values are normalized to Tbp level (n=2, Mean ± SEM).

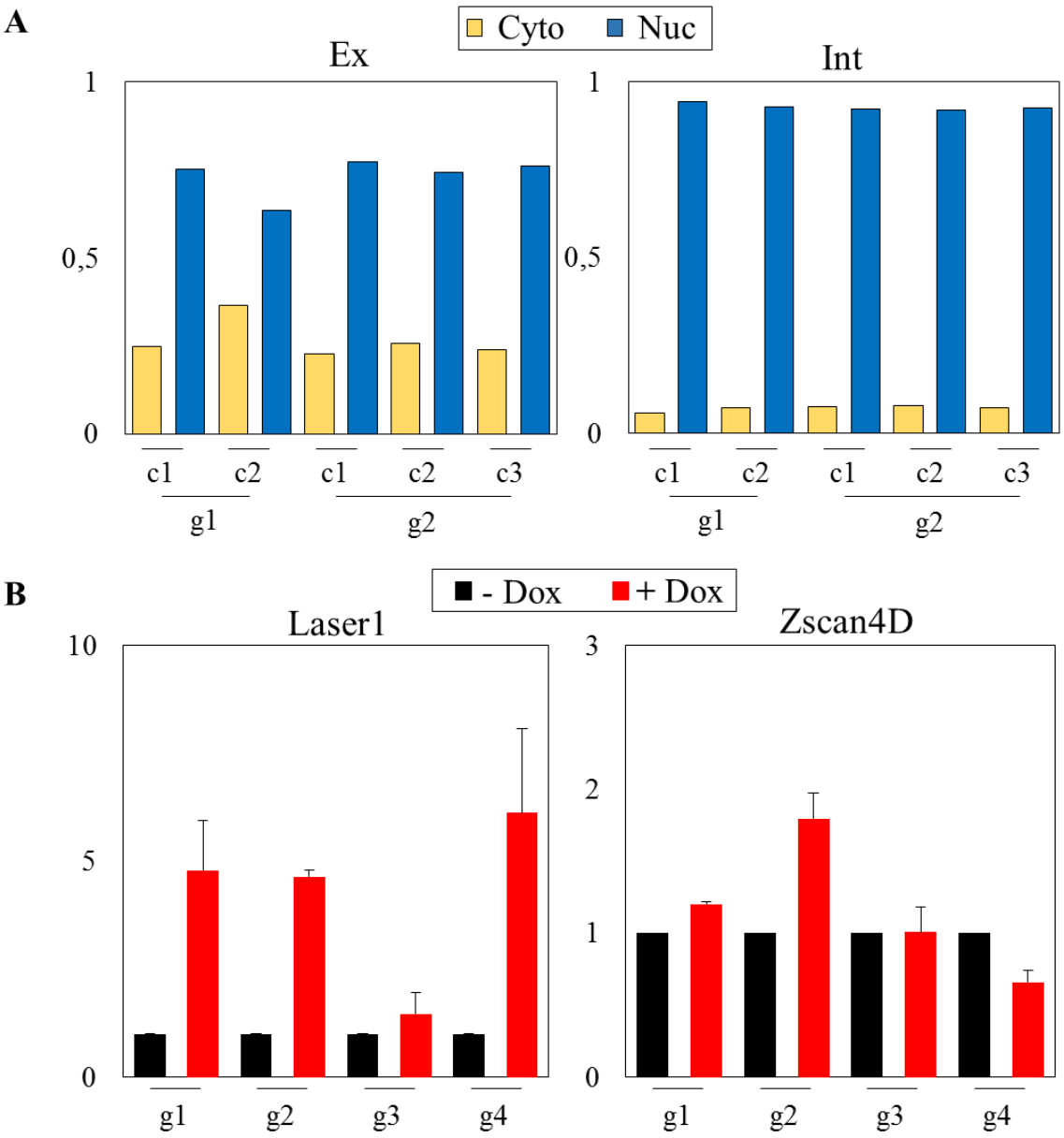


Fig. 2.33. **A.** RT-qPCR measure of RNA abundance in nuclear and cytoplasmic fractions of LASER 1 spliced RNA (Ex) or primary transcript (Int) showing unaffected nuclear distribution of LASER 1 under induction with both gRNAs. Two or three independent clones (c1, c2, c3) harbouring gRNA 1 or 2 were used for this experiment. **B.** RT-qPCR measure of expression of LASER 1 and Zscan4d RNAs with or without induction of LASER 1 with the four different gRNAs shown in Fig. 2.30 A. The two SunTag clones were transfected with the four gRNAs and, after selection for stable integration, the batch of cells was used for Dox treatment and RNA extraction (n=2, Mean ± SEM).

It is worth mentioning that the small effect on Zscan4 expression seen in this experiment can be associated with the reported extensive heterogeneity of cellular clones in inducing the 2 cell-like state (Hendrickson et al. 2017). Indeed, gene expression analysis was

this time measured on a batch of transfected cells and not on previously picked individual clones, thus increasing the intrinsic heterogeneity of the examined population compared to our previous RT-qPCR and RNA-seq data **(Fig. 2.31 A)**. Hence, all these results collectively argued against a LASER 1-mediated effect.

D. Second hypothesis: an off-target effect

After accumulating evidence that the two cell-like markers induction was not due to LASER 1 locus induction we decided to map the binding sites of the SunTag system genome-wide when expressed along with gRNA 2. We therefore repeated the induction experiment with our LASER 1 gRNAs 1 and 2 clones and prepared chromatin samples for ChIP-sequencing. We immuno-precipitated the HA Tag as it is fused to both parts of the SunTag complex. We included samples from the two parental SunTag clones without gRNAs as negative controls **(Fig. 2.34)**.

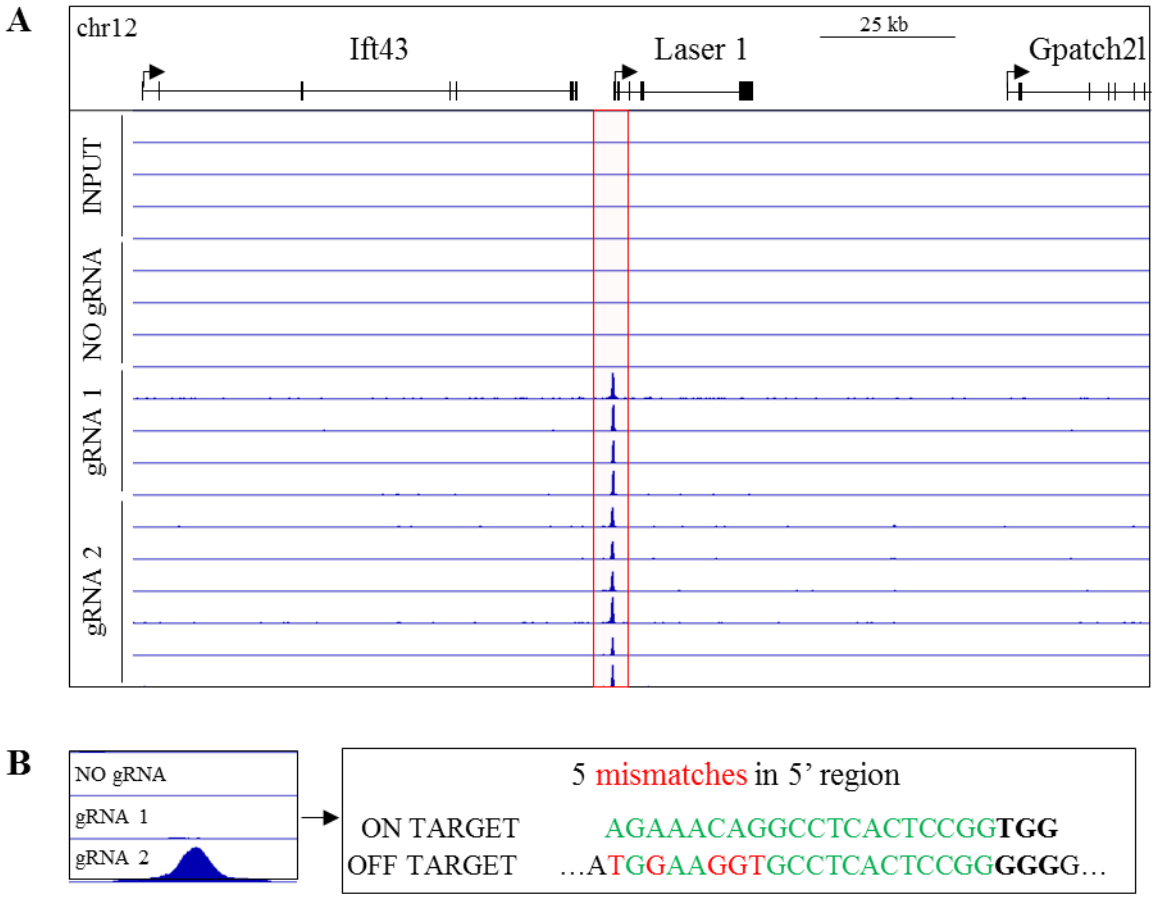


Fig. 2.34. **A.** Screenshot of IGV browser showing ChIP-sequencing signal around LASER 1 gene. This dataset includes: 4 replicates of Input samples, 4 replicates of the two SunTag clones without gRNA, 4 replicates of SunTag subclones with gRNA 1, 6 replicates of SunTag

subclones with gRNA 2. **B.** Screenshot of one of the strongest off-target sites showing canonical off-target sequence with 5 mismatches in the 5' region of the gRNA2 sequence.

Binding at the on-target LASER 1 locus was very sharp and strong compared to background **(Fig. 2.34 A)**. It was quite obvious at first glance that binding profile of SunTag with gRNA 2 was showing off-target binding at multiple positions even if signal at those places was weaker than on the on-target site. One of the strongest one we could manually identify was closer inspected **(Fig. 2.34 B)**. It revealed, at the summit of the peak, a conventional off-target site with 5 mismatches situated in the 5' region of the gRNA that has been shown to be the less critical part for mismatched allowing for Cas9 binding (Hsu et al. 2013). The ChIP-seq results were further analysed with MACS to identify high confidence binding sites throughout the genome (performed by Olivier Piau). Without any gRNA we identified 7 consistent peaks within the two SunTag clones corresponding to weak binding at very broadly open regions as super enhancers where dCas9 scanning likely happens more frequently with a probably higher residency time than randomly expected. With gRNA 1, only 5 off-target sites could be identified showing canonical off-target sequences with few mismatches in the 5' region of the gRNA. Finally, 928 off-target peaks were identified by MACS in at least 4 of the 6 replicates we had with gRNA 2.

Strikingly, no peaks were seen in the vicinity of identified inducers of the two cell-like state (Dux, Zscan4 cluster, Gata factors, Eif1a-like clusters…) thus not allowing for a simple explanation of the observed effect.

One very broad off-target binding site caught our attention. This signal was almost 2kb long with detected binding all along it **(Fig. 2.35 A)**. This SunTag binding domain was actually located in the second exon of a coding gene, the Loricrin gene (Lor). When looking for the sequence leading to such a broad binding it appeared that this region was mainly composed of a (CCG)n repeat and a high GC percentage but didn't contain any sequence resembling the full gRNA 2 sequence. However we noticed that the end of the gRNA sequence contained a CCGG motif and subsequently wondered whether such a short kind of motif could trigger dCas9 binding towards so many places. We found in the literature that such small motifs contained in the seed region of the gRNA (last 6 to 8 bp before the PAM) were reported to be able to trigger genome-wide binding of dCas9 into open regions. However, such binding is usually weaker (or faster) than compared to the on-target site and rarely gave rise to indels when using the active Cas9 protein (X. Wu et al. 2014b; Polstein et al. 2015). We thus used MEME-ChIP tool

(Machanick and Bailey 2011) to identify a possible recurrent motif in our 928 identified peaks (performed by Olivier Piau). A CTCCGGNGG motif corresponding to the six last nucleotides of the gRNA followed by the PAM sequence **(Fig. 2.34 B)** was identified and found in 86% of the peaks **(Fig. 2.35 B)** therefore explaining the off-target activity of our gRNA 2.

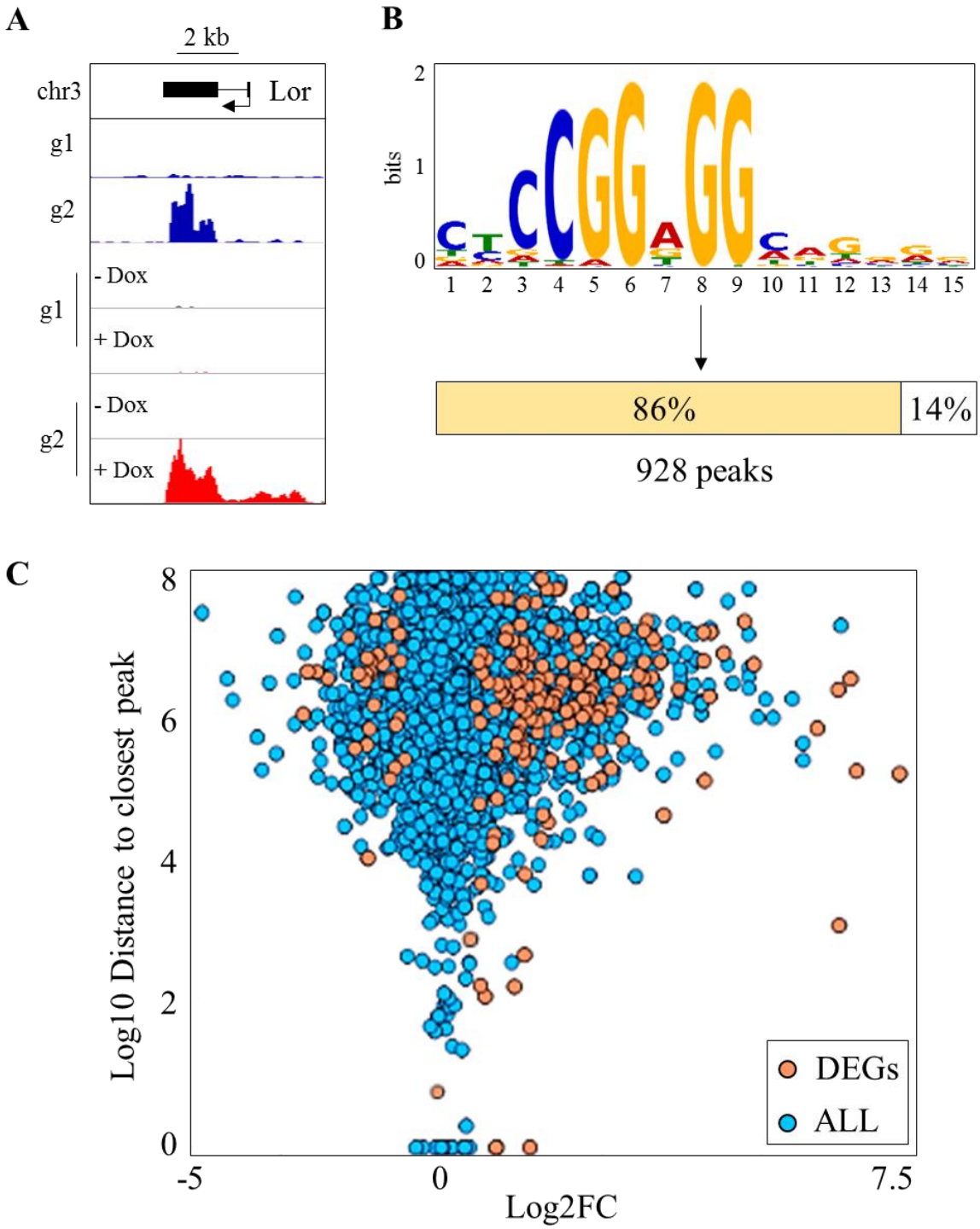


Fig. 2.35. **A.** Screenshot of IGV browser showing ChIP-seq binding profile (blue) of SunTag and transcriptional response (black and red) from RNA-seq experiment with gRNA 1 and 2

over Loricrin gene locus **B.** Motif identified from our 928 peaks regions **C.** Transcriptional response of DEGs and all other genes represented in function of the distance to the closest SunTag binding peak.

In the case of the Loricrin gene, an extensive transcriptional induction resulted from the broad recruitment of the SunTag complex at the locus (Log2FC around 7) **(Fig. 2.35 A).** We thus expected our 742 upregulated genes to show such a binding towards their promoter. We therefore investigated the relationship between the transcriptional response at the genome-wide level and the binding activity of the SunTag complex guided by LASER 1 gRNA 2 (performed by Olivier Piau) **(Fig. 2.35 C)**. It surprisingly showed that most of our DEGs are not in the proximity of the identified binding sites, and are consequently most likely, not directly induced by the SunTag system. This raises the possibility that one intermediary activator would be stimulated by one of these binding events and that this effector would, downstream, activate the 2 cell like state specific transcriptome. This analysis also showed that, despite off-target binding of the SunTag complex towards many promoters, a transcriptional response can only be observed for very few of the downstream genes, suggesting that the efficiency of SunTag activation at off-target sites is low.

E. Identification of candidate genes

Following the hypothesis that one of the DEGs might be responsible for the activation of the 2 cell-like state, we selected all of them showing a gRNA 2 SunTag binding peak at less than 3 kb away from their TSS. We subselected the ones presenting a log2 fold change of expression greater than 1 in at least one the three aforementioned datasets upon 2 cell-like state induction (Ishiuchi et al. 2015; Eckersley-Maslin et al. 2016; Hendrickson et al. 2017) **(Fig. 2.37. B)**. Only ten DEGs passed these two filters. From those, only two showed great potential based on the fact that they are not expressed in the absence of and show a high induction upon Dox treatment in our SunTag 2 cell-like state samples, and are known regulators of transcription: Dmrt2 and Ebf3 **(Fig. 2.36)**.

Dmrt2 is a transcription factor that has been shown to be involved in the skeletal and muscular development of the mouse embryo (Seo et al. 2006). Dmrt2 null embryos show somite patterning defects and die soon after birth due to abnormal rib and sternal development, leading to breathing inability. Its expression in early embryo, from the previously used single cell RNA-seq, shows that it is expressed in the zygote and rapidly decreases after the early 2-

cell stage. Due to its lethal consequence at birth, the consequence of the absence of Dmrt2 expression in oocyte has never been assessed. Therefore, its possible function in zygotic genome activation, where a similar transcriptome to the 2-cell like state is established *in vivo*, can't be excluded **(Fig. 2.37. A)**.

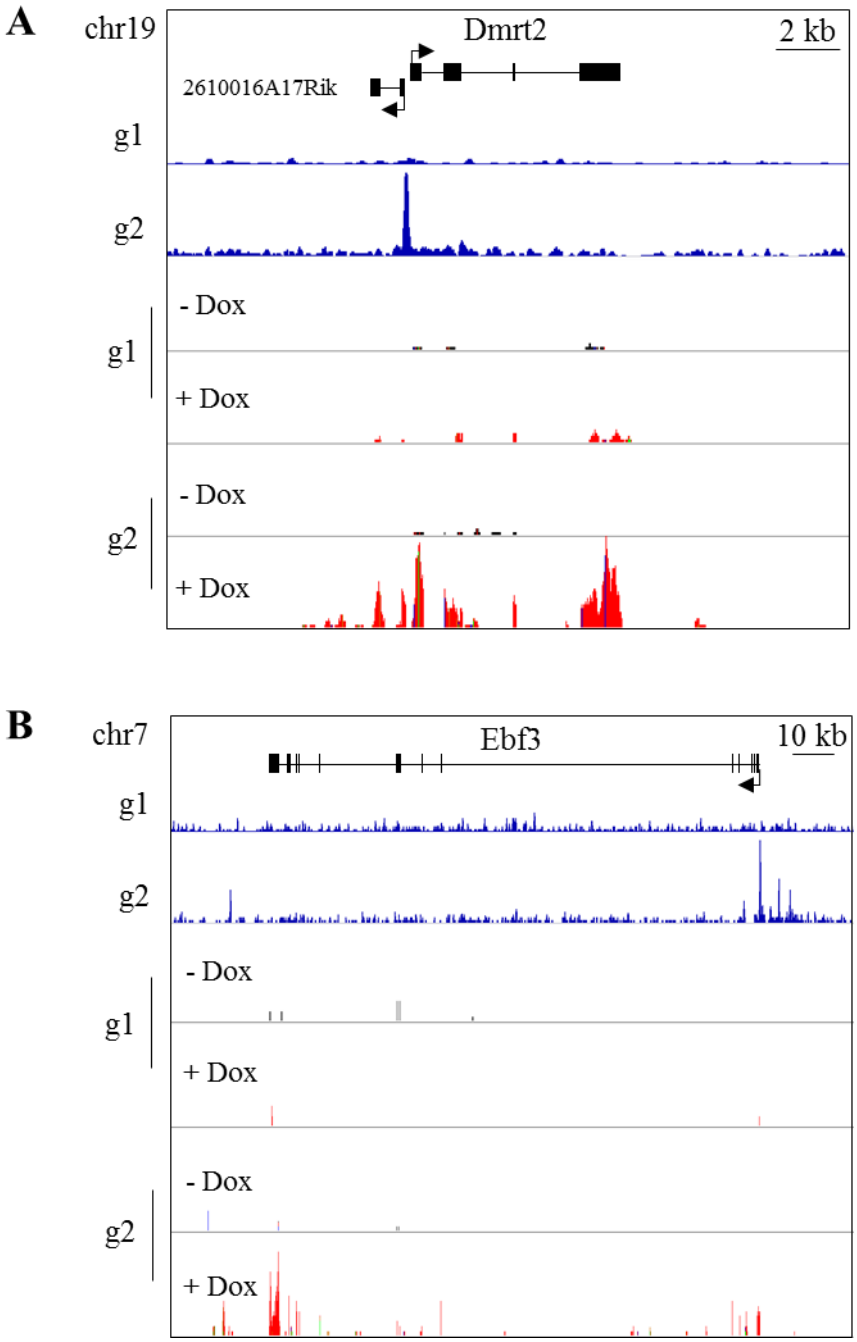
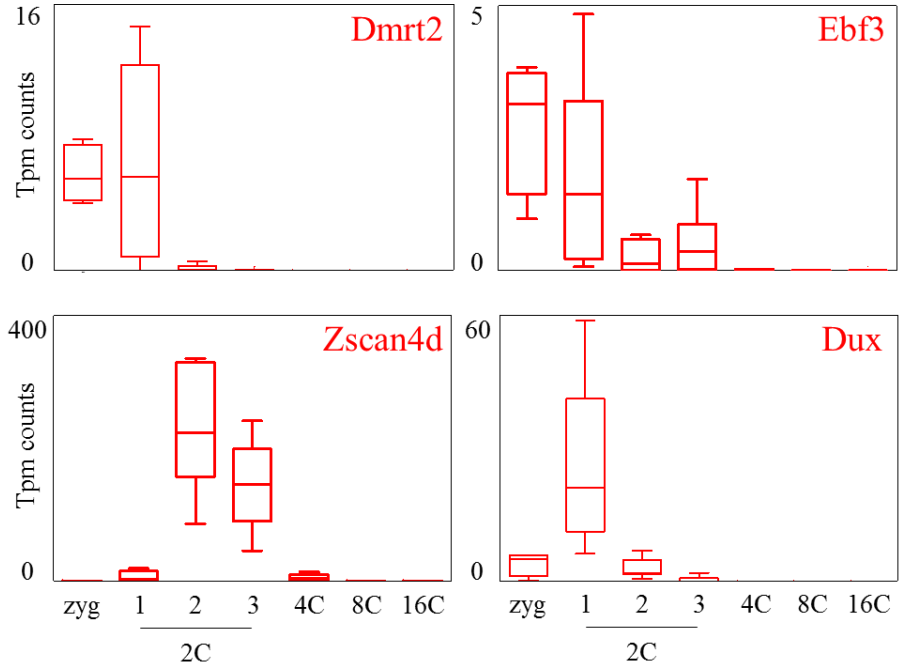


Fig. 2.36. Screenshots of IGV browser showing RNA-seq coverage obtained before (-Dox, black) and after induction (+Dox, red) with the two different gRNAs targeting LASER 1 promoter. **A.** Dmrt2 locus showing a strong SunTag binding (blue) with gRNA 2 on its promoter inducing both Dmrt2 and its divergent neighbour expression. **B.** Ebf3 locus showing

multiple SunTag binding (blue) upstream of its TSS and a strong associated transcriptional response.

A comparable expression profile was observed for Ebf3 during early development **(Fig. 2.37. A)**. Ebf3, as well, is a transcription factor and has been shown to play a large panel of regulatory functions in multiple tissues, from development of the cortical neurons to terminal differentiation of specific muscles and maintenance of the integrity of the hematopoietic stem cell niche (Blackburn et al. 2017; Sleven et al. 2017; Harms et al. 2017; Seike et al. 2018; Iwai et al. 2018). Its deletion has therefore been associated with neurodevelopmental disease in human (Lopes et al. 2017; Harms et al. 2017, 3) but Ebf3 null mice are viable and fertile with a, however, dramatically and slightly reduced mating efficiency in males and females respectively (S. S. Wang et al. 2004). Consequently, its maternal depletion doesn't seem to strongly affect early development.

A



B

		p60 KD	p150 KD	Dux OE	MERVL/Zscan+
Dmrt2	Log2FC	0,88	2,02	1,95	1,39
	adjP	0,23	0,00073	0,00011	NA
Ebf3	Log2FC	0,78	0,84	2,64	0,35
	adjP	0,29	0,28	0,00016	NA

Fig. 2.37. **A.** Single cell RNA-seq expression during mouse embryo early development, from Deng et al. 2014. (zyg = zygote, 2C, 4C, 8C, 16C = 2, 4, 8 and 16 cells stage, 1 = early, 2 = middle, 3 = late). Tpm counts values are represented for each cell. **B.** Log2 Fold Change of expression and corresponding adjusted p-value in 4 different datasets involving 2-cell like state

induction (Eckersley-Maslin and Spector 2014; Ishiuchi et al. 2015; Hendrickson et al. 2017) showing that both candidates are significantly induced upon Dux overexpression but also have a positive Log2FC among all datasets.

F. Discussion

The short motif-driven, binding of dCas9, consequently tethering the SunTag complex to many genomic places, induced the expression of *bona fide* markers of the 2-cell like state in our SunTag clones. The fact that many copies of this motif exist in the genome (about 28.000 occurrences) but that only a small subset of them are actually bound, is comparable to what can be seen with conventional transcription factors and mostly depends on chromatin accessibility (X. Wu et al. 2014a; Polstein et al. 2015). What was more surprising for us was that many genes present a peak close to their TSS, but only few of them show any response to SunTag activation. We should first eliminate the possibility that all these genes are not simply highly expressed genes in ES cells whose the spurious binding of the SunTag complex in the vicinity of their promoter doesn't manage to overexpress. But if this is not the case, it would be very informative for us and others to understand what rules the fact that a gene will be responsive or not to CRISPRa stimulation. Therefore, chromatin context, transcription factors binding, and chromatin modifiers enrichment should be correlated with the responsiveness of all these bound TSS.

A clear mechanism for the induction of our phenotype first needs to be found before further characterization of our cells showing a 2-cell like state expression profile. Indeed, many different aspects are characteristic of this transient behaviour. First, validation of MERVL and Zscan4 inductions in single cell needs to be determined by RNA FISH and immuno staining. This should allow us to assess the percentage of our cells entering the 2-cell like state at a given time. The chromatin state of these cells will need to be evaluated in term of histone acetylation, histone motility, chromatin accessibility, as well as DNA methylation status and translation activity that are the main features getting affected towards the appearance of this state (Hung et al. 2013; Ishiuchi et al. 2015; Rodriguez-Terrones et al. 2018). Moreover, extended potency of the cells will have to be assessed as ES cells transiting through this state have been shown to have totipotent abilities (Macfarlan et al. 2012; Choi et al. 2017).

As previously mentioned and shown in **Fig. 2.29**, the repeated cluster of Eif1a-like genes are strongly induced upon our 2-cell like state induction and has been reported in almost

all studies presenting a similar expression profile. These proteins have been shown to be responsible for the induction of a translational block (Hung et al. 2013; Eckersley-Maslin et al. 2016) appearing during the 2-cell like state. However, an interesting observation is that MERVL elements, as many retroviral elements, harbour a functional protease in absence of accumulated mutations and have been shown to produce virus-like particles in the 2-cell stage embryo (Bénit et al. 1997; Ribet et al. 2008). Remarkably, viral proteases have been shown to strongly interfere with host cell translation (Balvay et al. 2007; Walsh, Mathews, and Mohr 2013) to redirect the endogenous machinery towards the production of viral proteins. Therefore, MERVL protease expression should perhaps be investigated during the 2-cell like state and its potential effect on translation assessed. Since retroviral protease inhibitors are well-developed drugs (Kuhelj et al. 2001; Tyagi et al. 2017), the functional role of their expression could be easily tested. A translational block has never been reported so far at the 2-cell-stage embryo but seems to be important to test in regards to what has been observed in the 2-cell like state. Indeed, this might be an important feature impacting the maternal-zygotic transition characterized by the turnover of maternal factors progressively replaced by the zygotic ones (Tadros and Lipshitz 2009). Such a mechanism would also propose a new functional explanation for the brief expression of these hijacked retroviral elements as a mean for the host organism to transiently impact one of its main cellular machinery. Another interesting aspect of the regulation of the translational activity is its direct impact on the regulation of chromatin structure and accessibility as shown recently in mouse ES cells (Bulut-Karslioglu et al. 2018).

As shown above **(Fig. 2.37 A)** Dux and Zscan4 genes get activated during the early and mid 2-cell stage of the embryo respectively. It would thus be surprising that those factors might be *in vivo* responsible for the induction of the transcription wave to which they belong to. Therefore, we can speculate that upstream activators of the specific 2-cell stage transcriptional profile might exist. Such factors should be already expressed at the zygotic stage, most likely inherited from the maternal stock of mRNAs. Remarkably, that is the case of the two candidates we selected, Dmrt2 and Ebf3. Nevertheless, if Dux and Gata2/3 factors are able to induce this response *in vitro*, what is the common property they share to do so? Moreover, would this property also be shared with our candidate factors? Strikingly, if looking at the reported binding motif of Dmrt2, it appears as a composite of Gata3 (closely related to Gata2) and Dux motifs **(Fig. 2.38)** suggesting that it might be able to activate a comparable set of targets. In contrast, Ebf3 binding motif has not been robustly characterized so far. Finally, we now aim at validating

whether both candidates are able to recapitulate the transcriptional activation of the 2-cell like state signature by their overexpression in mouse ES cells. If such a result would be confirmed, given their expression pattern in the early embryo, it would suggest an upstream role of these factors in initiating the establishment of the 2 cell stage transcriptional landscape.

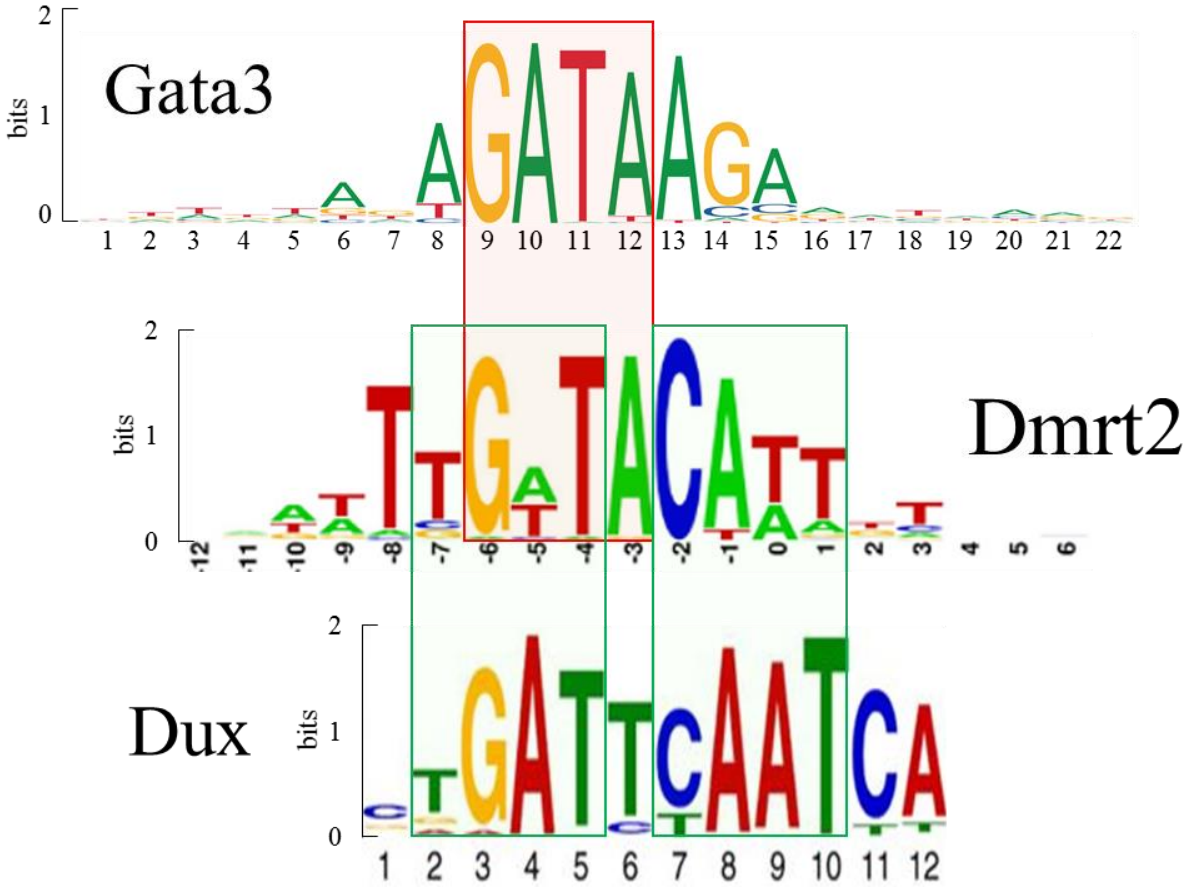


Fig. 2.38. Comparison of Dmrt2 binding motif with Gata3 and Dux factors showing a substantial overlap of similarity with both factors (Jaspar database, (Hendrickson et al. 2017; Murphy, Zarkower, and Bardwell 2007).

gRNAs		RT-qPCR		Antibodies	Dilution used for WB
Nanog-F	caccGCTTCCCACTAGAGATCGCCA	Tbp-F	GGGGAGCTGTGATGTGAAGT	H3 (Abcam ab1791)	1:5000
Nanog-R	aaacTGGCGATCTCTAGTGGGAAGC	Tbp-R	CCAGGAAATAATTCTGGCTCA	anti-rabbit IgG HRP (Thermo RB230254)	1:10000
Laser10-1-F	caccGGCGGTACCAACTGTTGAGA	Nanog-F	aggatgaagtgcaagcggtg	anti-mouse IgG HRP (Thermo QB213868)	1:10000
Laser10-1-R	aaacTCTCAACAGTTGGTACCGCC	Nanog-R	tgctgagcccttctgaatcag	Suv39h1 (Millipore #07-550)	1:1000
Laser10-2-F	caccGGGGTCTGACTGGCCCCTAG	Laser1-F	TCTCCAAGCAGGAGAAGGAA		
Laser10-2-R	aaacCTAGGGGCCAGTCAGACCCC	Laser1-R	CTTGGTCTTGGGATCAGGAG		
Laser11-1-F	caccGATGGAAAGAGGGCGTCGCTC	Laser2-F	CTGCTAGGTCGGGGATGTAG		
Laser11-1-R	aaacGAGCGACGCCCTCTTTCCATC	Laser2-R	TCCTCTTAACAAGCCCCAAA		
Laser11-2-F	caccGCGTCCCGCGCTGAGCGTAAA	Laser3-F	CTCGGGTCAACATCCTGTTT		
Laser11-2-R	aaacTTTACGCTCAGCGCGGGACGC	Laser3-R	GGACAGAGAACCACCTCCTG		
Laser12-1-F	caccGTTCTCGCCTAGCCCGTGCTG	Laser4-F	GGTCCCAAGACACTTCAGGA		
Laser12-1-R	aaacCAGCACGGGCTAGGCGAGAAC	Laser4-R	CAGGGTGAAGGGACAGGTT		
Laser12-2-F	caccGTAGGTAGCCTGCACCAAAAT	Laser5-F	CACGGGATGTGGAGTTCTTT		
Laser12-2-R	aaacATTTTGGTGCAGGCTACCTAC	Laser5-R	GAAACAACCCAGCACAGGAT		
Laser13-1-F	caccGAAAGCGGAAGTTGTAGTAC	Laser6-F	AGGCCTGCCAGTTTCAGAG		
Laser13-1-R	aaacGTACTACAACTTCCGCTTTC	Laser6-R	CCCCTTTTCTGGGACTTGTT		
Laser13-2-F	caccGTCGTTGTGTTAAACACTTAT	Laser7-F	TTCCCTGTCATGGGGATAAA		
Laser13-2-R	aaacATAAGTGTTTAACACAACGAC	Laser7-R	ACAGACAACCCAAGTCAGCA		
Laser14-1-F	caccGCCAGAAGCTGGTCGCTGGTT	Laser8-F	GGCAAGGCTTGAGATCTGTC		
Laser14-1-R	aaacAACCAGCGACCAGCTTCTGGC	Laser8-R	AGAGGTTCACAGGCAGCAGT		
Laser14-2-F	caccGGCGCATGCATTATTCAGCG	Laser9-F	AAAGACCCAGGGCAAGAACT		
Laser14-2-R	aaacCGCTGAATAATGCATGCGCC	Laser9-R	CACAGTGGCATCCTGTCTCA		
Laser15-1-F	caccGCTTGCTCTCTCGGCCTAGG	Laser10-F	ACTGGTCCCCACACAAGAGA		
Laser15-1-R	aaacCCTAGGCCGAGAGAGCAAGC	Laser10-R	TTCCACTCTTGGCTCTCCAT		
Laser15-2-F	caccGACAATAGCGCTGAAAGGCCG	Laser11-F	TTTGTGGACTTGCAGGATGA		
Laser15-2-R	aaacCGGCCTTTCAGCGCTATTGTC	Laser11-R	CTCAAGGTACCGAGCTCCAC		
Laser16-1-F	caccGGGATATCTTCCGCCGCTGG	Laser12-F	TGTCCCCTTCTGTGAACCTT		
Laser16-1-R	aaacCCAGCGGCGGAAGATATCCC	Laser12-R	AAGCATGAGTGGGAGGTGAG		
Laser16-2-F	caccGACGGAGCGTGTCTAAGCTCG	Laser13-F	AAGAAAAGCAGGAGGGAACAA		
Laser16-2-R	aaacCGAGCTTAGACACGCTCCGTC	Laser13-R	GAGCGAGCCTCAGACACTCTA		
Laser17-1-F	caccGAATGTAGCCAGGCCACAGTG	Laser14-F	CAGAAGGAACGTGGGACAAG		
Laser17-1-R	aaacCACTGTGGCCTGGCTACATTC	Laser14-R	AGACGGTGGACCCTTTTCTT		
Laser17-2-F	caccGCTAGATCGACTTCTCAGAAT	Laser15-F	CATGGGTGAGGAGTCAAGAAA		
Laser17-2-R	aaacATTCTGAGAAGTCGATCTAGC	Laser15-R	CCAAATTGGAATCGCAGAAA		
Laser18-1-F	caccGCGGCGAGCGCGCGGATAGCC	Laser16-F	TACCACGTTATCGGGGTCTC		
Laser18-1-R	aaacGGCTATCCGCGCGCTCGCCGC	Laser16-R	GCTAGGTAGCCAAGCACAGC		
Laser18-2-F	caccGAATTGGTCCCCTCCGCCTCG	Laser17-F	TGCCACTCTGTAGGCTGAAA		
Laser18-2-R	aaacCGAGGCGGAGGGGACCAATTC	Laser17-R	CAACCTCAATGAAAGGCAGAA		
Laser19-1-F	caccGTTACGAATCCTTCGGTCCAC	Laser18-F	TGCACAGATCGGACACAACT		
Laser19-1-R	aaacGTGGACCGAAGGATTCGTAAC	Laser18-R	TCAGAACTCCACAGGCATCTC		
Laser19-2-F	caccGCCAAAACTCAGTTAAGCGCG	Laser19-F	TGGCCAGAGGTACTAGGAAGG		
Laser19-2-R	aaacCGCGCTTAACTGAGTTTTGGC	Laser19-R	GCCAACAGGAGACAGAAGGA		
Laser1-1-F	caccGCCAGTGGCTTAGCCGTCTGT	Laser20-F	AGGAGCCATGGACAGTGAGT		
Laser1-1-R	aaacACAGACGGCTAAGCCACTGGC	Laser20-R	GCCTATGGCCTTGGTTTTTC		
Laser1-2-F	caccGAGAAACAGGCCTCACTCCGG	Laser21-F	GATGCTAGCGGCTTTTGTCT		
Laser1-2-R	aaacCCGGAGTGAGGCCTGTTTCTC	Laser21-R	TGTGGCTGCATATTCTCTGC		
Laser20-1-F	caccGACCACGGCTTGTTAGATTAT	Laser22-F	CTGTGAAGCCAGACAAGCAG		
Laser20-1-R	aaacATAATCTAACAAGCCGTGGTC	Laser22-R	GCTTCTGCCTCCTGCTCTT		
Laser20-2-F	caccGCCATGGCGTGCCAGTAACC	Laser23-F	CACCATGCTTCCTGCCATA		
Laser20-2-R	aaacGGTTACTGGCACGCCATGGC	Laser23-R	TTCCAGATCCCAAAGTCTCCT		
Laser21-1-F	caccGCGTCTAGCTGCGTGGCGTC	Laser24-F	TCCCCCAAAAATCATCAAAA		
Laser21-1-R	aaacGACGCCACGCAGCTAGACGC	Laser24-R	ATGCATTACATGGGCATTCA		
Laser21-2-F	caccGCTGCGTGGCGTCTGGATCGT	L23ex23-F	AGGAGACTTTGGGATCTGGAA		
Laser21-2-R	aaacACGATCCAGACGCCACGCAGC	L23ex23-R	CGCTACGATCTCTGCATCTTC		
Laser22-1-F	caccGCATCACAAAACTCGATCGAT	Aqp3-f	ccctctggacacttggacat		
Laser22-1-R	aaacATCGATCGAGTTTTGTGATGC	Aqp3-r	gttgacggcatagccagaat		
Laser22-2-F	caccGAACTCGATCGATCGGCTTAG	Marveld1-F	GTGAGTTCTGTCTGCCACCA		
Laser22-2-R	aaacCTAAGCCGATCGATCGAGTTC	Marveld1-R	CTCTACCACGGCTCTTCCAC		
Laser23-1-F	caccGCCACCACTATCCAGTGAAA	Abcb5-chim-F	gaaccattgatgggaattgg		
Laser23-1-R	aaacTTTCACTGGATAGTGGTGGC	Abcb5-chim-R	gcaatggcaatcctctgttt		
Laser24-1-F	caccGCACATGCGCAGGATCGTCG	MERVL_int-519-rv	CTAGAACCACTCCTGGTACCAAC		
Laser24-1-R	aaacCGACGATCCTGCGCATGTGC	MERVL_LTR-365-fw	CTTCCATTCACAGCTGCGACTG		
Laser24-2-F	caccGAAGTTTCCGCATGCGCAGCA	Esrrb-f	cgattcatgaaatgcctcaa		
Laser24-2-R	aaacTGCTGCGCATGCGGAAACTTC	Esrrb-r	cctcctcgaactcggtca		
Laser2-1-F	caccGCCGATGTTCTGCCTTAGTCC	L1B-F	tctccctttgttgctcaca		
Laser2-1-R	aaacGGACTAAGGCAGAACATCGGC	L1B-R	AGTGAGGCCTGTTTCTTCCTC		
Laser3-1-F	caccGTTGGGTCATGTGATCGGGTC	L1C-F	GCAATGGAGGTGTGGCTAGT		
Laser3-1-R	aaacGACCCGATCACATGACCCAAC	L1C-R	TTTGGTACCCACAATCCTCTG		
Laser3-2-F	caccGATCGGGTCTGGCCCATAGG	08rik_2-f	AGAGGAAAATTTGGGCCAGT		
Laser3-2-R	aaacCCTATGGGCCAGACCCGATC	08rik_2-r	GGAGCATGAGGTTGTTCCAT		
Laser4-1-F	caccGTCCCAAGGACAACTTGTCGC	Egf-2-F	TCGAGAGAAGCGAGAGAAGC		
Laser4-1-R	aaacGCGACAAGTTGTCCTTGGGAC	Egf-2-R	TGTTCCATCTGGGTCAATCC		
Laser4-2-F	caccGAATCGGTAGCCTCCCGCCCA	42TE2-f	AGAAACGGAGAGCCGAAGAT		
Laser4-2-R	aaacTGGGCGGGAGGCTACCGATTC	42TE2-r	TGGATTTAGTATGGAGGTGATGG		
Laser5-1-F	caccGAGGGAAGTGGAGTTAGACCG	Pax6-F	aacaacctgcctatgcaacc		
Laser5-1-R	aaacCGGTCTAACTCCACTTCCCTC	Pax6-R	cataactccgcccattcact		

Laser5-2-F	caccGTTATTCAGTGTCCGTCGGA	Pax6intron-F	GGAAGGGCTGAGGAGATAGG					
Laser5-2-R	aaacTCCGACGGACACTGAATAAC	Pax6intron-R	TGCTCTTGGGTAAACCTGCT					
Laser6-1-F	caccGTTAAACCCTCAGACAATCGA	Pax6eRNA_3'-F	CTGCTTTGCCTAGAGGGTTG					
Laser6-1-R	aaacTCGATTGTCTGAGGGTTTAAC	Pax6eRNA_3'-R	GTGCAGTGGGATTTGAACCT					
Laser6-2-F	caccGTTGGAGTGTCCTGACAATCT	Pax6eRNA_5'-F	TACAGCCGCACAATTTCTGA					
Laser6-2-R	aaacAGATTGTCAGGACACTCCAAC	Pax6eRNA_5'-R	ACCTCTTTCACGGTGTCAGC					
Laser7-1-F	caccGCAGTTGACAGCGGCAAGTC	ELP4-F	TCTGACAAGTCTGGAGGCAA					
Laser7-1-R	aaacGACTTGCCGCTGTCAACTGC	ELP4-R	TGTGATGAATCGGATGTCAAA					
Laser7-2-F	caccGACAGGTGGATAGGACGGACG	RCN1-F	ACAAGGCCATCACCATCACT					
Laser7-2-R	aaacCGTCCGTCCTATCCACCTGTC	RCN1-R	CTCCAAGACCTTCGATCAGC					
Laser8-1-F	caccGGGACTAAAGCCTATTATTC	IMMP1L-F	AAAGCCCAAGTGATCCAAAA					
Laser8-1-R	aaacGAATAATAGGCTTTAGTCCC	IMMP1L-R	ACATGACCTGTTGGCACGTA					
Laser8-2-F	caccGAACCAACTATAACTGGCGT	DNAJC24-F	GTACAGCATTCTGGGTGCAG					
Laser8-2-R	aaacACGCCAGTTATAGTTGGTTC	DNAJC24-R	CTGGCACATCTGCACTTTGT					
Laser9-1-F	caccGCGGAGACTTGGGCAGCCGGC							
Laser9-1-R	aaacGCCGGCTGCCCAAGTCTCCGC							
Laser9-2-F	caccGCTGCAGGTGGAAGAGTCCG							
Laser9-2-R	aaacCGGACTCTTCCACCTGCAGC							
Laser1-3-F	caccCTA**TTT**CCAACTCGGCTTGG							
Laser1-3-R	aaacCCAAGCCGAGTTGGAAATAG							
Laser1-4-F	caccAGAGCAAGACCACCGGAGTG							
Laser1-4-R	aaacCACTCCGGTGGTCTTGCTCT							
gNanog2-f	caccGTGGGGCGTGGGTGCCGCCT							
gNanog2-r	aaacAGGCGGCACCCACGCCCCAC							
gNanog4-f	caccGGGATTAACTGTGAATTCAC							
gNanog4-r	aaacGTGAATTCACAGTTAATCCC							
gEsrrb2-f	caccGTGAGTTTTTCCCCGTGGTC							
gEsrrb2-r	aaacGACCACGGGGAAAAACTCAC							
gEsrrb5-f	caccGGCAGGTTGGCCAAATATTT							
gEsrrb5-r	aaacAAATATTTGGCCAACCTGCC							
guc008rik1-f	caccGTTGTTCTGACAACCTAATCG							
guc008rik1-r	aaacCGATTAGGTTGTCAGAACAAC							
guc008rik2-f	caccGAAATAGGGTGACCCTCGATT							
guc008rik2-r	aaacAATCGAGGGTCACCCTATTTC							
guc008rik3-f	caccGGAGGAGTAGTAGTACAATG							
guc008rik3-r	aaacCATTGTACTACTACTCCTCC							
glinc1242-1-f	caccGTCAATCCCCTCGATTAGCT							
glinc1242-1-r	aaacAGCTAATCGAGGGGATTGAC							
glinc1242-5-f	caccGTCCTCTTCCACATGTGCGA							
glinc1242-5-r	aaacTCGCACATGTGGAAGAGGAC							
Ng1-F	caccGGTGGGGCGTGGGTGCCGCC							
Ng1-R	aaacGGCGGCACCCACGCCCCACC							
Ng2-F	caccGCTTCCCACTAGAGATCGCCA							
Ng2-R	aaacTGGCGATCTCTAGTGGGAAGC							
Ng3-F	caccGACTTCCCACTAGAGATCGCC							
Ng3-R	aaacGGCGATCTCTAGTGGGAAGTC							
Ng4-F	caccGCTGTAAGGTGACCCAGACT							
Ng4-R	aaacAGTCTGGGTCACCTTACAGC							
Ng5-F	caccGATCTGAAGGCCAACGGCTCA							
Ng5-R	aaacTGAGCCGTTGGCCTTCAGATC							
Pax6p-1-F	caccGCATCCAATCGGCTGGCGCG							
Pax6p-1-R	aaacCGCGCCAGCCGATTGGATGC							
Pax6p-2-F	caccGTCCCGCTCTGGTTCAGGCGC							
Pax6p-2-R	aaacGCGCCTGAACCAGAGCGGGAC							
Pax6e-5-F	caccGAAGTGTAACACTGGGGCTAT							
Pax6e-5-R	aaacATAGCCCCAGTGTTACACTTC							
Pax6e-6-F	caccGTAATCGCAGAGTCGATGAGC							
Pax6e-6-R	aaacGCTCATCGACTCTGCGATTAC							

Table 1. gRNAs, primers and additional antibodies list

LASER	Name	ID	chr	start (mm9)	end (m9)	strand	Closest gene 1	Closest gene 2	Ulitsky ID
Laser1	AK144886	Gm805	chr12	87510291	87534552	+	Ift43	Gpatch2l	MSTRG.6328
Laser2			chr10	60481089	60525845	-	Pcbd1	Sgpl1	MSTRG.2467-68-69-70
Laser3		ENSMUSG00000008600	chr11	93994795	94017090	+	Spag9		MSTRG.4766
Laser4		Gm53	chr11	96112711	96132320	+	Hoxb9	mir196a	MSTRG.4842-43
Laser5		AK145614	chr13	99870660	99877535	+	Tnpo1	Ptcd2	MSTRG.7684
Laser6			chr14	73220187	73247655	-	Fndc3a	Cysltr2	MSTRG.8714
Laser7			chr14	76915745	76930545	+	Tsc22d1	Serp2	MSTRG.8767
Laser8			chr15	74467046	74483289	-	Mroh4	Arc	MSTRG.9480
Laser9	linc p21	Trp53cor1	chr17	29194418	29215906	-	cdkn1a	Srsf3	MSTRG.11475
Laser10			chr2	172403283	172457326	+	Tcfap2c	Bmp7	MSTRG.16143
Laser11		Ppp4r1l-ps	chr2	173405938	173485040	-	Rab22a	Ankrd60	MSTRG.16163
Laser12			chr4	127007227	127020770	+	gjb3	Gjb4	MSTRG.19269
Laser13			chr4	137437061	137439857	-	Ece1		MSTRG.19532
Laser14		E130006D01Rik	chr5	112163304	112190748	-	Mn1	Miat	MSTRG.21071
Laser15	2410137F16Rik	srrm4as	chr5	116883580	116916823	+	Srrm4	Hsbp8	MSTRG.21213
Laser16	E230016M11Rik	AK054076	chr6	66986593	67030646	+	Gadd45a	A430010J10Rik	MSTRG.22545
Laser17			chr6	125382601	125408973	+	Cd9	Plekhg6	MSTRG.23240
Laser18			chr6	128388774	128453299	+	Fkbp4	Pzp	MSTRG.23287
Laser19		Lockd	chr6	134879103	134906816	+	Cdkn1b		MSTRG.23335
Laser20			chr7	6929957	6943798	+	Aurkc	Zim3	MSTRG.23626
Laser21		2900076A07Rik	chr7	88668436	88676384	+	Fsd2	Ap3b2	MSTRG.24775
Laser22			chr9	75316281	75334487	+	Leo1	Tmod3	MSTRG.28179
Laser23		AK010638	chrX	7626298	7651745	+	Suv39h1	Glod5	MSTRG.28939
Laser24		Dleu2	chr14	62243159	62267663	-	Kcnrg	Trim13	MSTRG.8598

Table 2. LASER list

References

Abranches, Elsa, Ana M. V. Guedes, Martin Moravec, Hedia Maamar, Petr Svoboda, Arjun Raj, and Domingos Henrique. 2014. 'Stochastic NANOG Fluctuations Allow Mouse Embryonic Stem Cells to Explore Pluripotency'. *Development* 141 (14): 2770–79. https://doi.org/10.1242/dev.108910.

Abudayyeh, Omar O., Jonathan S. Gootenberg, Patrick Essletzbichler, Shuo Han, Julia Joung, Joseph J. Belanto, Vanessa Verdine, et al. 2017. 'RNA Targeting with CRISPR-Cas13a'. *Nature* 550 (7675): 280–84. https://doi.org/10.1038/nature24049.

Acampora, D., L. G. Di Giovannantonio, and A. Simeone. 2013. 'Otx2 Is an Intrinsic Determinant of the Embryonic Stem Cell State and Is Required for Transition to a Stable Epiblast Stem Cell Condition'. *Development* 140 (1): 43–55. https://doi.org/10.1242/dev.085290.

Acampora, D., S. Mazan, Y. Lallemand, V. Avantaggiato, M. Maury, A. Simeone, and P. Brûlet. 1995. 'Forebrain and Midbrain Regions Are Deleted in Otx2-/- Mutants Due to a Defective Anterior Neuroectoderm Specification during Gastrulation'. *Development (Cambridge, England)* 121 (10): 3279–90.

Acampora, Dario, Luca Giovanni Di Giovannantonio, Arcomaria Garofalo, Vincenzo Nigro, Daniela Omodei, Alessia Lombardi, Jingchao Zhang, Ian Chambers, and Antonio Simeone. 2017. 'Functional Antagonism between OTX2 and NANOG Specifies a Spectrum of Heterogeneous Identities in Embryonic Stem Cells'. *Stem Cell Reports* 9 (5): 1642–59. https://doi.org/10.1016/j.stemcr.2017.09.019.

Acampora, Dario, Daniela Omodei, Giuseppe Petrosino, Arcomaria Garofalo, Marco Savarese, Vincenzo Nigro, Luca Giovanni Di Giovannantonio, Vincenzo Mercadante, and Antonio Simeone. 2016. 'Loss of the Otx2-Binding Site in the Nanog Promoter Affects the Integrity of Embryonic Stem Cell Subtypes and Specification of Inner Cell Mass-Derived Epiblast'. *Cell Reports* 15 (12): 2651–64. https://doi.org/10.1016/j.celrep.2016.05.041.

Ait- Si- Ali, Slimane, Valentina Guasconi, Lauriane Fritsch, Hakima Yahi, Redha Sekhri, Irina Naguibneva, Philippe Robin, Florence Cabon, Anna Polesskaya, and Annick Harel- Bellan. 2004. 'A Suv39h- dependent Mechanism for Silencing S- phase Genes in Differentiating but Not in Cycling Cells'. *The EMBO Journal* 23 (3): 605–15. https://doi.org/10.1038/sj.emboj.7600074.

Akutsu, Hidenori, Takumi Miura, Masakazu Machida, Jun-ichi Birumachi, Aki Hamada, Mitsutoshi Yamada, Stephen Sullivan, Kenji Miyado, and Akihiro Umezawa. 2009. 'Maintenance of Pluripotency and Self-Renewal Ability of Mouse Embryonic Stem

Cells in the Absence of Tetraspanin CD9'. *Differentiation*, Special Issue on Stem Cells, 78 (2): 137–42. https://doi.org/10.1016/j.diff.2009.08.005.

Alder, Olivia, Fabrice Lavial, Anne Helness, Emily Brookes, Sandra Pinho, Anil Chandrashekran, Philippe Arnaud, Ana Pombo, Laura O'Neill, and Véronique Azuara. 2010. 'Ring1B and Suv39h1 Delineate Distinct Chromatin States at Bivalent Genes during Early Mouse Lineage Commitment'. *Development (Cambridge, England)* 137 (15): 2483–92. https://doi.org/10.1242/dev.048363.

Ambrosetti, D. C., C. Basilico, and L. Dailey. 1997. 'Synergistic Activation of the Fibroblast Growth Factor 4 Enhancer by Sox2 and Oct-3 Depends on Protein-Protein Interactions Facilitated by a Specific Spatial Arrangement of Factor Binding Sites.' *Molecular and Cellular Biology* 17 (11): 6321–29. https://doi.org/10.1128/MCB.17.11.6321.

Ambrosetti, Davide-Carlo, Hans R. Schöler, Lisa Dailey, and Claudio Basilico. 2000. 'Modulation of the Activity of Multiple Transcriptional Activation Domains by the DNA Binding Domains Mediates the Synergistic Action of Sox2 and Oct-3 on the Fibroblast Growth Factor-4Enhancer'. *Journal of Biological Chemistry* 275 (30): 23387–97. https://doi.org/10.1074/jbc.M000932200.

Anders, Carolin, Ole Niewoehner, Alessia Duerst, and Martin Jinek. 2014. 'Structural Basis of PAM-Dependent Target DNA Recognition by the Cas9 Endonuclease'. *Nature* 513 (7519): 569–73. https://doi.org/10.1038/nature13579.

Apostolou, Effie, and Matthias Stadtfeld. 2018. 'Cellular Trajectories and Molecular Mechanisms of IPSC Reprogramming'. *Current Opinion in Genetics & Development*, Cell reprogramming, regeneration and repair, 52 (October): 77–85. https://doi.org/10.1016/j.gde.2018.06.002.

Arney, Katharine L., Siqin Bao, Andrew J. Bannister, Tony Kouzarides, and M. Azim Surani. 2002. 'Histone Methylation Defines Epigenetic Asymmetry in the Mouse Zygote'. *The International Journal of Developmental Biology* 46 (3): 317–20.

Avilion, Ariel A., Silvia K. Nicolis, Larysa H. Pevny, Lidia Perez, Nigel Vivian, and Robin Lovell-Badge. 2003. 'Multipotent Cell Lineages in Early Mouse Development Depend on SOX2 Function'. *Genes & Development* 17 (1): 126–40. https://doi.org/10.1101/gad.224503.

Aw, Jong Ghut Ashley, Yang Shen, Andreas Wilm, Miao Sun, Xin Ni Lim, Kum-Loong Boon, Sidika Tapsin, et al. 2016. 'In Vivo Mapping of Eukaryotic RNA Interactomes Reveals Principles of Higher-Order Organization and Regulation'. *Molecular Cell* 62 (4): 603–17. https://doi.org/10.1016/j.molcel.2016.04.028.

Azuara, Véronique, Pascale Perry, Stephan Sauer, Mikhail Spivakov, Helle F. Jørgensen, Rosalind M. John, Mina Gouti, et al. 2006. 'Chromatin Signatures of Pluripotent Cell Lines'. *Nature Cell Biology* 8 (5): 532–38. https://doi.org/10.1038/ncb1403.

Ballaré, Cecilia, Martin Lange, Audrone Lapinaite, Gloria Mas Martin, Lluis Morey, Gloria Pascual, Robert Liefke, et al. 2012. 'Phf19 Links Methylated Lys36 of Histone H3 to Regulation of Polycomb Activity'. *Nature Structural & Molecular Biology* 19 (12): 1257–65. https://doi.org/10.1038/nsmb.2434.

Balvay, Laurent, Marcelo Lopez Lastra, Bruno Sargueil, Jean-Luc Darlix, and Théophile Ohlmann. 2007. 'Translational Control of Retroviruses'. *Nature Reviews Microbiology* 5 (2): 128–40. https://doi.org/10.1038/nrmicro1599.

Barrangou, Rodolphe, Christophe Fremaux, Hélène Deveau, Melissa Richards, Patrick Boyaval, Sylvain Moineau, Dennis A. Romero, and Philippe Horvath. 2007. 'CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes'. *Science* 315 (5819): 1709–12. https://doi.org/10.1126/science.1138140.

Bénit, L., N. De Parseval, J. F. Casella, I. Callebaut, A. Cordonnier, and T. Heidmann. 1997. 'Cloning of a New Murine Endogenous Retrovirus, MuERV-L, with Strong Similarity to the Human HERV-L Element and with a Gag Coding Sequence Closely Related to the Fv1 Restriction Gene.' *Journal of Virology* 71 (7): 5652–57.

Bergmann, Jan H., Jingjing Li, Mélanie A. Eckersley-Maslin, Frank Rigo, Susan M. Freier, and David L. Spector. 2015. 'Regulation of the ESC Transcriptome by Nuclear Long Noncoding RNAs'. *Genome Research* 25 (9): 1336–46. https://doi.org/10.1101/gr.189027.114.

Bernstein, Bradley E., Tarjei S. Mikkelsen, Xiaohui Xie, Michael Kamal, Dana J. Huebert, James Cuff, Ben Fry, et al. 2006. 'A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells'. *Cell* 125 (2): 315–26. https://doi.org/10.1016/j.cell.2006.02.041.

Berrens, Rebecca V., Simon Andrews, Dominik Spensberger, Fátima Santos, Wendy Dean, Poppy Gould, Jafar Sharif, et al. 2017. 'An EndosiRNA-Based Repression Mechanism Counteracts Transposon Activation during Global DNA Demethylation in Embryonic Stem Cells'. *Cell Stem Cell* 21 (5): 694-703.e7. https://doi.org/10.1016/j.stem.2017.10.004.

Bessonnard, Sylvain, Laurane De Mot, Didier Gonze, Manon Barriol, Cynthia Dennis, Albert Goldbeter, Geneviève Dupont, and Claire Chazaud. 2014. 'Gata6, Nanog and Erk Signaling Control Cell Fate in the Inner Cell Mass through a Tristable Regulatory Network'. *Development (Cambridge, England)* 141 (19): 3637–48. https://doi.org/10.1242/dev.109678.

Bikard, David, Wenyan Jiang, Poulami Samai, Ann Hochschild, Feng Zhang, and Luciano A. Marraffini. 2013. 'Programmable Repression and Activation of Bacterial Gene Expression Using an Engineered CRISPR-Cas System'. *Nucleic Acids Research* 41 (15): 7429–37. https://doi.org/10.1093/nar/gkt520.

Black, Joshua B., Andrew F. Adler, Hong-Gang Wang, Anthony M. D'Ippolito, Hunter A. Hutchinson, Timothy E. Reddy, Geoffrey S. Pitt, Kam W. Leong, and Charles A. Gersbach. 2016. 'Targeted Epigenetic Remodeling of Endogenous Loci by CRISPR/Cas9-Based Transcriptional Activators Directly Converts Fibroblasts to Neuronal Cells'. *Cell Stem Cell* 19 (3): 406–14. https://doi.org/10.1016/j.stem.2016.07.001.

Blackburn, Patrick R., Sarah S. Barnett, Michael T. Zimmermann, Margot A. Cousin, Charu Kaiwar, Filippo Pinto e Vairo, Zhiyv Niu, et al. 2017. 'Novel de Novo Variant in EBF3 Is Likely to Impact DNA Binding in a Patient with a Neurodevelopmental Disorder and Expanded Phenotypes: Patient Report, in Silico Functional Assessment, and Review of Published Cases'. *Cold Spring Harbor Molecular Case Studies* 3 (3). https://doi.org/10.1101/mcs.a001743.

Blake, Judith A., and Melanie R. Ziman. 2014. 'Pax Genes: Regulators of Lineage Specification and Progenitor Cell Maintenance'. *Development* 141 (4): 737–51. https://doi.org/10.1242/dev.091785.

Blaschke, Kathryn, Kevin T. Ebata, Mohammad M. Karimi, Jorge A. Zepeda-Martínez, Preeti Goyal, Sahasransu Mahapatra, Angela Tam, et al. 2013. 'Vitamin C Induces Tet-Dependent DNA Demethylation and a Blastocyst-like State in ES Cells'. *Nature* 500 (7461): 222–26. https://doi.org/10.1038/nature12362.

Boeuf, Hélène, Charlotte Hauss, Fabienne De Graeve, Nathalie Baran, and Claude Kedinger. 1997. 'Leukemia Inhibitory Factor–Dependent Transcriptional Activation in Embryonic Stem Cells'. *The Journal of Cell Biology* 138 (6): 1207–17.

Bogu, Gireesh K., Pedro Vizán, Lawrence W. Stanton, Miguel Beato, Luciano Di Croce, and Marc A. Marti-Renom. 2016. 'Chromatin and RNA Maps Reveal Regulatory Long Noncoding RNAs in Mouse'. *Molecular and Cellular Biology* 36 (5): 809–19. https://doi.org/10.1128/MCB.00955-15.

Bolotin, Alexander, Benoit Quinquis, Alexei Sorokin, and S. Dusko Ehrlich. 2005. 'Clustered Regularly Interspaced Short Palindrome Repeats (CRISPRs) Have Spacers of Extrachromosomal Origin'. *Microbiology (Reading, England)* 151 (Pt 8): 2551–61. https://doi.org/10.1099/mic.0.28048-0.

Bonasio, Roberto, Emilio Lecona, Varun Narendra, Philipp Voigt, Fabio Parisi, Yuval Kluger, and Danny Reinberg. 2014. 'Interactions with RNA Direct the Polycomb Group Protein SCML2 to Chromatin Where It Represses Target Genes'. *ELife* 3 (July). https://doi.org/10.7554/eLife.02637.

Bonev, Boyan, Netta Mendelson Cohen, Quentin Szabo, Lauriane Fritsch, Giorgio L. Papadopoulos, Yaniv Lubling, Xiaole Xu, et al. 2017. 'Multiscale 3D Genome Rewiring during Mouse Neural Development'. *Cell* 171 (3): 557-572.e24. https://doi.org/10.1016/j.cell.2017.09.043.

Boroviak, Thorsten, Remco Loos, Paul Bertone, Austin Smith, and Jennifer Nichols. 2014. 'The Ability of Inner Cell Mass Cells to Self-Renew as Embryonic Stem Cells Is Acquired upon Epiblast Specification'. *Nature Cell Biology* 16 (6): 516–28. https://doi.org/10.1038/ncb2965.

Boroviak, Thorsten, and Jennifer Nichols. 2014. 'The Birth of Embryonic Pluripotency'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369 (1657). https://doi.org/10.1098/rstb.2013.0541.

Bourillot, Pierre-Yves, Irène Aksoy, Valerie Schreiber, Florence Wianny, Herbert Schulz, Oliver Hummel, Norbert Hubner, and Pierre Savatier. 2009. 'Novel STAT3 Target Genes Exert Distinct Roles in the Inhibition of Mesoderm and Endoderm Differentiation in Cooperation with Nanog'. *Stem Cells (Dayton, Ohio)* 27 (8): 1760–71. https://doi.org/10.1002/stem.110.

Boyer, Laurie A., Kathrin Plath, Julia Zeitlinger, Tobias Brambrink, Lea A. Medeiros, Tong Ihn Lee, Stuart S. Levine, et al. 2006. 'Polycomb Complexes Repress Developmental Regulators in Murine Embryonic Stem Cells'. *Nature* 441 (7091): 349–53. https://doi.org/10.1038/nature04733.

Boyle, Kristy, Jian-Guo Zhang, Sandra E. Nicholson, Evelyn Trounson, Jeffery J. Babon, Edward J. McManus, Nicos A. Nicola, and Lorraine Robb. 2009. 'Deletion of the SOCS Box of Suppressor of Cytokine Signaling 3 (SOCS3) in Embryonic Stem Cells Reveals SOCS Box-Dependent Regulation of JAK but Not STAT Phosphorylation'. *Cellular Signalling* 21 (3): 394–404. https://doi.org/10.1016/j.cellsig.2008.11.002.

Bradley, A., M. Evans, M. H. Kaufman, and E. Robertson. 1984. 'Formation of Germ-Line Chimaeras from Embryo-Derived Teratocarcinoma Cell Lines'. *Nature* 309 (5965): 255–56.

Brien, Gerard L., Guillermo Gambero, David J. O'Connell, Emilia Jerman, Siobhán A. Turner, Chris M. Egan, Eiseart J. Dunne, et al. 2012. 'Polycomb PHF19 Binds H3K36me3 and Recruits PRC2 and Demethylase NO66 to Embryonic Stem Cell Genes during Differentiation'. *Nature Structural & Molecular Biology* 19 (12): 1273–81. https://doi.org/10.1038/nsmb.2449.

Brockdorff, Neil. 2013. 'Noncoding RNA and Polycomb Recruitment'. *RNA* 19 (4): 429–42. https://doi.org/10.1261/rna.037598.112.

———. 2017. 'Polycomb Complexes in X Chromosome Inactivation'. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 372 (1733). https://doi.org/10.1098/rstb.2017.0021.

Brons, I. Gabrielle M., Lucy E. Smithers, Matthew W. B. Trotter, Peter Rugg-Gunn, Bowen Sun, Susana M. Chuva de Sousa Lopes, Sarah K. Howlett, et al. 2007. 'Derivation of

Pluripotent Epiblast Stem Cells from Mammalian Embryos'. *Nature* 448 (7150): 191–95. https://doi.org/10.1038/nature05950.

Brookes, Emily, Inês de Santiago, Daniel Hebenstreit, Kelly J. Morris, Tom Carroll, Sheila Q. Xie, Julie K. Stock, et al. 2012. 'Polycomb Associates Genome-Wide with a Specific RNA Polymerase II Variant, and Regulates Metabolic Genes in ESCs'. *Cell Stem Cell* 10 (2): 157–70. https://doi.org/10.1016/j.stem.2011.12.017.

Brouns, Stan J. J., Matthijs M. Jore, Magnus Lundgren, Edze R. Westra, Rik J. H. Slijkhuis, Ambrosius P. L. Snijders, Mark J. Dickman, Kira S. Makarova, Eugene V. Koonin, and John van der Oost. 2008. 'Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes'. *Science (New York, N.Y.)* 321 (5891): 960–64. https://doi.org/10.1126/science.1159689.

Buecker, Christa, Rajini Srinivasan, Zhixiang Wu, Eliezer Calo, Dario Acampora, Tiago Faial, Antonio Simeone, Minjia Tan, Tomasz Swigut, and Joanna Wysocka. 2014. 'Reorganization of Enhancer Patterns in Transition from Naive to Primed Pluripotency'. *Cell Stem Cell* 14 (6): 838–53. https://doi.org/10.1016/j.stem.2014.04.003.

Bulut-Karslioglu, Aydan, Steffen Biechele, Hu Jin, Trisha A. Macrae, Miroslav Hejna, Marina Gertsenstein, Jun S. Song, and Miguel Ramalho-Santos. 2016. 'Inhibition of MTOR Induces a Paused Pluripotent State'. *Nature* 540 (7631): 119–23. https://doi.org/10.1038/nature20578.

Bulut-Karslioglu, Aydan, Inti A. De La Rosa-Velázquez, Fidel Ramirez, Maxim Barenboim, Megumi Onishi-Seebacher, Julia Arand, Carmen Galán, et al. 2014. 'Suv39h-Dependent H3K9me3 Marks Intact Retrotransposons and Silences LINE Elements in Mouse Embryonic Stem Cells'. *Molecular Cell* 55 (2): 277–90. https://doi.org/10.1016/j.molcel.2014.05.029.

Bulut-Karslioglu, Aydan, Trisha A. Macrae, Juan A. Oses-Prieto, Sergio Covarrubias, Michelle Percharde, Gregory Ku, Aaron Diaz, Michael T. McManus, Alma L. Burlingame, and Miguel Ramalho-Santos. 2018. 'The Transcriptionally Permissive Chromatin State of Embryonic Stem Cells Is Acutely Tuned to Translational Output'. *Cell Stem Cell* 22 (3): 369-383.e8. https://doi.org/10.1016/j.stem.2018.02.004.

Bulut-Karslioglu, Aydan, Valentina Perrera, Manuela Scaranaro, Inti Alberto de la Rosa-Velazquez, Suzanne van de Nobelen, Nicholas Shukeir, Johannes Popow, et al. 2012. 'A Transcription Factor–Based Mechanism for Mouse Heterochromatin Formation'. *Nature Structural & Molecular Biology* 19 (10): 1023–30. https://doi.org/10.1038/nsmb.2382.

Burdon, T., C. Stracey, I. Chambers, J. Nichols, and A. Smith. 1999. 'Suppression of SHP-2 and ERK Signalling Promotes Self-Renewal of Mouse Embryonic Stem Cells'. *Developmental Biology* 210 (1): 30–43. https://doi.org/10.1006/dbio.1999.9265.

Burton, Adam, and Maria-Elena Torres-Padilla. 2010. 'Epigenetic Reprogramming and Development: A Unique Heterochromatin Organization in the Preimplantation Mouse Embryo'. *Briefings in Functional Genomics* 9 (5–6): 444–54. https://doi.org/10.1093/bfgp/elq027.

Cabili, Moran N., Cole Trapnell, Loyal Goff, Magdalena Koziol, Barbara Tazon-Vega, Aviv Regev, and John L. Rinn. 2011. 'Integrative Annotation of Human Large Intergenic Noncoding RNAs Reveals Global Properties and Specific Subclasses'. *Genes & Development* 25 (18): 1915–27. https://doi.org/10.1101/gad.17446611.

Cai, Ling, Scott B. Rothbart, Rui Lu, Bowen Xu, Wei-Yi Chen, Ashutosh Tripathy, Shira Rockowitz, et al. 2013. 'An H3K36 Methylation-Engaging Tudor Motif of Polycomb-like Proteins Mediates PRC2 Complex Targeting'. *Molecular Cell* 49 (3): 571–82. https://doi.org/10.1016/j.molcel.2012.11.026.

Cao, Ru, Liangjun Wang, Hengbin Wang, Li Xia, Hediye Erdjument-Bromage, Paul Tempst, Richard S. Jones, and Yi Zhang. 2002. 'Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing'. *Science (New York, N.Y.)* 298 (5595): 1039–43. https://doi.org/10.1126/science.1076997.

Cao, Zubing, Timothy S. Carey, Avishek Ganguly, Catherine A. Wilson, Soumen Paul, and Jason G. Knott. 2015. 'Transcription Factor AP-2γ Induces Early Cdx2 Expression and Represses HIPPO Signaling to Specify the Trophectoderm Lineage'. *Development (Cambridge, England)* 142 (9): 1606–15. https://doi.org/10.1242/dev.120238.

Carleton, Julia B., Kristofer C. Berrett, and Jason Gertz. 2017. 'Multiplex Enhancer Interference Reveals Collaborative Control of Gene Regulation by Estrogen Receptor α-Bound Enhancers'. *Cell Systems* 5 (4): 333-344.e5. https://doi.org/10.1016/j.cels.2017.08.011.

Carlevaro-Fita, Joana, Anisa Rahim, Roderic Guigó, Leah A. Vardy, and Rory Johnson. 2016. 'Cytoplasmic Long Noncoding RNAs Are Frequently Bound to and Degraded at Ribosomes in Human Cells'. *RNA* 22 (6): 867–82. https://doi.org/10.1261/rna.053561.115.

Carpenter, Anne E., and David M. Sabatini. 2004. 'Systematic Genome-Wide Screens of Gene Function'. *Nature Reviews Genetics* 5 (1): 11–22. https://doi.org/10.1038/nrg1248.

Carter, Ava C., Brandi N. Davis-Dusenbery, Kathryn Koszka, Justin K. Ichida, and Kevin Eggan. 2014. 'Nanog-Independent Reprogramming to IPSCs with Canonical Factors'. *Stem Cell Reports* 2 (2): 119–26. https://doi.org/10.1016/j.stemcr.2013.12.010.

Carter, David, Lyubomira Chakalova, Cameron S. Osborne, Yan-feng Dai, and Peter Fraser. 2002. 'Long-Range Chromatin Regulatory Interactions *in Vivo*'. *Nature Genetics* 32 (4): 623–26. https://doi.org/10.1038/ng1051.

Cartwright, Peter, Cameron McLean, Allan Sheppard, Duane Rivett, Karen Jones, and Stephen Dalton. 2005. 'LIF/STAT3 Controls ES Cell Self-Renewal and Pluripotency by a Myc-Dependent Mechanism'. *Development (Cambridge, England)* 132 (5): 885–96. https://doi.org/10.1242/dev.01670.

Chakraborty, Debojyoti, Dennis Kappei, Mirko Theis, Anja Nitzsche, Li Ding, Maciej Paszkowski-Rogacz, Vineeth Surendranath, et al. 2012. 'Combined RNAi and Localization for Functionally Dissecting Long Noncoding RNAs'. *Nature Methods* 9 (4): 360–62. https://doi.org/10.1038/nmeth.1894.

Chakraborty, Debojyoti, Maciej Paszkowski-Rogacz, Nicolas Berger, Li Ding, Jovan Mircetic, Jun Fu, Vytautas Iesmantavicius, et al. 2017. 'LncRNA Panct1 Maintains Mouse Embryonic Stem Cell Identity by Regulating TOBF1 Recruitment to Oct-Sox Sequences in Early G1'. *Cell Reports* 21 (11): 3012–21. https://doi.org/10.1016/j.celrep.2017.11.045.

Chamberlain, Stormy J., Della Yee, and Terry Magnuson. 2008. 'Polycomb Repressive Complex 2 Is Dispensable for Maintenance of Embryonic Stem Cell Pluripotency'. *Stem Cells (Dayton, Ohio)* 26 (6): 1496–1505. https://doi.org/10.1634/stemcells.2008-0102.

Chambers, Ian, Douglas Colby, Morag Robertson, Jennifer Nichols, Sonia Lee, Susan Tweedie, and Austin Smith. 2003. 'Functional Expression Cloning of Nanog, a Pluripotency Sustaining Factor in Embryonic Stem Cells'. *Cell* 113 (5): 643–55. https://doi.org/10.1016/S0092-8674(03)00392-1.

Chambers, Ian, Jose Silva, Douglas Colby, Jennifer Nichols, Bianca Nijmeijer, Morag Robertson, Jan Vrana, Ken Jones, Lars Grotewold, and Austin Smith. 2007. 'Nanog Safeguards Pluripotency and Mediates Germline Development'. *Nature* 450 (7173): 1230–34. https://doi.org/10.1038/nature06403.

Chantzoura, Eleni, Stavroula Skylaki, Sergio Menendez, Shin-Il Kim, Anna Johnsson, Sten Linnarsson, Knut Woltjen, Ian Chambers, and Keisuke Kaji. 2015. 'Reprogramming Roadblocks Are System Dependent'. *Stem Cell Reports* 5 (3): 350–64. https://doi.org/10.1016/j.stemcr.2015.07.007.

Chavez, Alejandro, Jonathan Scheiman, Suhani Vora, Benjamin W. Pruitt, Marcelle Tuttle, Eswar Iyer, Shuailiang Lin, et al. 2015. 'Highly-Efficient Cas9-Mediated Transcriptional Programming'. *Nature Methods* 12 (4): 326–28. https://doi.org/10.1038/nmeth.3312.

Chavez, Alejandro, Marcelle Tuttle, Benjamin W Pruitt, Ben Ewen-Campen, Raj Chari, Dmitry Ter-Ovanesyan, Sabina J Haque, et al. 2016. 'Comparative Analysis of Cas9 Activators Across Multiple Species'. *Nature Methods* 13 (7): 563–67. https://doi.org/10.1038/nmeth.3871.

Chazaud, Claire, and Yojiro Yamanaka. 2016. 'Lineage Specification in the Mouse Preimplantation Embryo'. *Development* 143 (7): 1063–74. https://doi.org/10.1242/dev.128314.

Chazaud, Claire, Yojiro Yamanaka, Tony Pawson, and Janet Rossant. 2006. 'Early Lineage Segregation between Epiblast and Primitive Endoderm in Mouse Blastocysts through the Grb2-MAPK Pathway'. *Developmental Cell* 10 (5): 615–24. https://doi.org/10.1016/j.devcel.2006.02.020.

Chen, Xi, Han Xu, Ping Yuan, Fang Fang, Mikael Huss, Vinsensius B. Vega, Eleanor Wong, et al. 2008. 'Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells'. *Cell* 133 (6): 1106–17. https://doi.org/10.1016/j.cell.2008.04.043.

Chen, Xiaoyu, Marrit Rinsma, Josephine M. Janssen, Jin Liu, Ignazio Maggio, and Manuel A.F.V. Gonçalves. 2016. 'Probing the Impact of Chromatin Conformation on Genome Editing Tools'. *Nucleic Acids Research* 44 (13): 6482–92. https://doi.org/10.1093/nar/gkw524.

Cheng, Albert W, Haoyi Wang, Hui Yang, Linyu Shi, Yarden Katz, Thorold W Theunissen, Sudharshan Rangarajan, Chikdu S Shivalila, Daniel B Dadon, and Rudolf Jaenisch. 2013. 'Multiplexed Activation of Endogenous Genes by CRISPR-on, an RNA-Guided Transcriptional Activator System'. *Cell Research* 23 (10): 1163–71. https://doi.org/10.1038/cr.2013.122.

Chew, Guo-Liang, Andrea Pauli, John L. Rinn, Aviv Regev, Alexander F. Schier, and Eivind Valen. 2013. 'Ribosome Profiling Reveals Resemblance between Long Non-Coding RNAs and 5' Leaders of Coding RNAs'. *Development (Cambridge, England)* 140 (13): 2828–34. https://doi.org/10.1242/dev.098343.

Chew, Joon-Lin, Yuin-Han Loh, Wensheng Zhang, Xi Chen, Wai-Leong Tam, Leng-Siew Yeap, Pin Li, et al. 2005. 'Reciprocal Transcriptional Regulation of Pou5f1 and Sox2 via the Oct4/Sox2 Complex in Embryonic Stem Cells'. *Molecular and Cellular Biology* 25 (14): 6031–46. https://doi.org/10.1128/MCB.25.14.6031-6046.2005.

Choi, Yong Jin, Chao-Po Lin, Davide Risso, Sean Chen, Thomas Aquinas Kim, Meng How Tan, Jin B. Li, et al. 2017. 'Deficiency of MicroRNA MiR-34a Expands Cell Fate Potential in Pluripotent Stem Cells'. *Science*, January, aag1927. https://doi.org/10.1126/science.aag1927.

Choy, Bob, and Michael R. Green. 1993. 'Eukaryotic Activators Function during Multiple Steps of Preinitiation Complex Assembly'. *Nature* 366 (6455): 531–36. https://doi.org/10.1038/366531a0.

Chris I. Ace, Mike A. Dalrymple, Fiona H. Ramsay, Valerie G. Preston, and Chris M. Preston. 1988. 'Mutational Analysis of the Herpes Simplex Virus Type 1 Trans-Inducing Factor Vmw65'. J. Gen. Virol.

Clevers, Hans. 2006. 'Wnt/β-Catenin Signaling in Development and Disease'. *Cell* 127 (3): 469–80. https://doi.org/10.1016/j.cell.2006.10.018.

Cole, Megan F., Sarah E. Johnstone, Jamie J. Newman, Michael H. Kagey, and Richard A. Young. 2008. 'Tcf3 Is an Integral Component of the Core Regulatory Circuitry of Embryonic Stem Cells'. *Genes & Development* 22 (6): 746–55. https://doi.org/10.1101/gad.1642408.

Cong, Le, F. Ann Ran, David Cox, Shuailiang Lin, Robert Barretto, Naomi Habib, Patrick D. Hsu, et al. 2013. 'Multiplex Genome Engineering Using CRISPR/Cas Systems'. *Science (New York, N.Y.)* 339 (6121): 819–23. https://doi.org/10.1126/science.1231143.

Costa, Yael, Junjun Ding, Thorold W. Theunissen, Francesco Faiola, Timothy A. Hore, Pavel V. Shliaha, Miguel Fidalgo, et al. 2013. 'Nanog-Dependent Function of Tet1 and Tet2 in Establishment of Pluripotency'. *Nature* 495 (7441): 370–74. https://doi.org/10.1038/nature11925.

Cousens, D J, R Greaves, C R Goding, and P O'Hare. 1989. 'The C-Terminal 79 Amino Acids of the Herpes Simplex Virus Regulatory Protein, Vmw65, Efficiently Activate Transcription in Yeast and Mammalian Cells in Chimeric DNA-Binding Proteins'. *The EMBO Journal*, 6.

Dan, Jiameng, Minshu Li, Jiao Yang, Jiaojiao Li, Maja Okuka, Xiaoying Ye, and Lin Liu. 2013. 'Roles for *Tbx3* in Regulation of Two-Cell State and Telomere Elongation in Mouse ES Cells'. *Scientific Reports* 3 (December): 3492. https://doi.org/10.1038/srep03492.

Darr, Henia, Yoav Mayshar, and Nissim Benvenisty. 2006. 'Overexpression of NANOG in Human ES Cells Enables Feeder-Free Growth While Inducing Primitive Ectoderm Features'. *Development* 133 (6): 1193–1201. https://doi.org/10.1242/dev.02286.

Davies, Owen R., Chia-Yi Lin, Aliaksandra Radzisheuskaya, Xinzhi Zhou, Jessica Taube, Guillaume Blin, Anna Waterhouse, Andrew J. H. Smith, and Sally Lowell. 2013. 'Tcf15 Primes Pluripotent Cells for Differentiation'. *Cell Reports* 3 (2): 472–84. https://doi.org/10.1016/j.celrep.2013.01.017.

De La Fuente, Rabindranath, Claudia Baumann, and Maria M. Viveiros. 2015. 'ATRX Contributes to Epigenetic Asymmetry and Silencing of Major Satellite Transcripts in the Maternal Genome of the Mouse Embryo'. *Development (Cambridge, England)* 142 (10): 1806–17. https://doi.org/10.1242/dev.118927.

Deng, Qiaolin, Daniel Ramsköld, Björn Reinius, and Rickard Sandberg. 2014. 'Single-Cell RNA-Seq Reveals Dynamic, Random Monoallelic Gene Expression in Mammalian Cells'. *Science (New York, N.Y.)* 343 (6167): 193–96. https://doi.org/10.1126/science.1245316.

Deng, Wulan, Jeremy W. Rupon, Ivan Krivega, Laura Breda, Irene Motta, Kristen S. Jahn, Andreas Reik, et al. 2014. 'Reactivation of Developmentally Silenced Globin Genes by Forced Chromatin Looping'. *Cell* 158 (4): 849–60. https://doi.org/10.1016/j.cell.2014.05.050.

Denny, Sarah Knight, Namita Bisaria, Joseph David Yesselman, Rhiju Das, Daniel Herschlag, and William James Greenleaf. 2018. 'High-Throughput Investigation of Diverse Junction Elements in RNA Tertiary Folding'. *Cell* 0 (0). https://doi.org/10.1016/j.cell.2018.05.038.

Derrien, Thomas, Rory Johnson, Giovanni Bussotti, Andrea Tanzer, Sarah Djebali, Hagen Tilgner, Gregory Guernec, et al. 2012. 'The GENCODE v7 Catalog of Human Long Noncoding RNAs: Analysis of Their Gene Structure, Evolution, and Expression'. *Genome Research* 22 (9): 1775–89. https://doi.org/10.1101/gr.132159.111.

DeVeale, Brian, Irina Brokhman, Paria Mohseni, Tomas Babak, Charles Yoon, Anthony Lin, Kento Onishi, et al. 2013. 'Oct4 Is Required ~E7.5 for Proliferation in the Primitive Streak'. *PLOS Genetics* 9 (11): e1003957. https://doi.org/10.1371/journal.pgen.1003957.

Deveau, Hélène, Rodolphe Barrangou, Josiane E. Garneau, Jessica Labonté, Christophe Fremaux, Patrick Boyaval, Dennis A. Romero, Philippe Horvath, and Sylvain Moineau. 2008. 'Phage Response to CRISPR-Encoded Resistance in Streptococcus Thermophilus'. *Journal of Bacteriology* 190 (4): 1390–1400. https://doi.org/10.1128/JB.01412-07.

Dhaliwal, Navroop K., Kamelia Miri, Scott Davidson, Hala Tamim El Jarkass, and Jennifer A. Mitchell. 2018. 'KLF4 Nuclear Export Requires ERK Activation and Initiates Exit from Naive Pluripotency'. *Stem Cell Reports* 10 (4): 1308–23. https://doi.org/10.1016/j.stemcr.2018.02.007.

Di Croce, Luciano, and Kristian Helin. 2013. 'Transcriptional Regulation by Polycomb Group Proteins'. *Nature Structural & Molecular Biology* 20 (10): 1147–55. https://doi.org/10.1038/nsmb.2669.

Dietrich, Jens-Erik, and Takashi Hiiragi. 2007. 'Stochastic Patterning in the Mouse Pre-Implantation Embryo'. *Development (Cambridge, England)* 134 (23): 4219–31. https://doi.org/10.1242/dev.003798.

Ding, Sheng, Xiaohui Wu, Gang Li, Min Han, Yuan Zhuang, and Tian Xu. 2005. 'Efficient Transposition of the PiggyBac (PB) Transposon in Mammalian Cells and Mice'. *Cell* 122 (3): 473–83. https://doi.org/10.1016/j.cell.2005.07.013.

Dinger, Marcel E., Paulo P. Amaral, Tim R. Mercer, Ken C. Pang, Stephen J. Bruce, Brooke B. Gardiner, Marjan E. Askarian-Amiri, et al. 2008. 'Long Noncoding RNAs in Mouse Embryonic Stem Cell Pluripotency and Differentiation'. *Genome Research* 18 (9): 1433–45. https://doi.org/10.1101/gr.078378.108.

Dinger, Marcel E., Paulo P. Amaral, Timothy R. Mercer, and John S. Mattick. 2009. 'Pervasive Transcription of the Eukaryotic Genome: Functional Indices and Conceptual Implications'. *Briefings in Functional Genomics* 8 (6): 407–23. https://doi.org/10.1093/bfgp/elp038.

Diwan, S. B., and L. C. Stevens. 1976. 'Development of Teratomas from the Ectoderm of Mouse Egg Cylinders'. *Journal of the National Cancer Institute* 57 (4): 937–42.

Dixon, Jesse R., Siddarth Selvaraj, Feng Yue, Audrey Kim, Yan Li, Yin Shen, Ming Hu, Jun S. Liu, and Bing Ren. 2012. 'Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions'. *Nature* 485 (7398): 376–80. https://doi.org/10.1038/nature11082.

Do, D. V., J. Ueda, D. M. Messerschmidt, C. Lorthongpanich, Y. Zhou, B. Feng, G. Guo, et al. 2013. 'A Genetic and Developmental Pathway from STAT3 to the OCT4-NANOG Circuit Is Essential for Maintenance of ICM Lineages in Vivo'. *Genes & Development* 27 (12): 1378–90. https://doi.org/10.1101/gad.221176.113.

Doetschman, T., R. G. Gregg, N. Maeda, M. L. Hooper, D. W. Melton, S. Thompson, and O. Smithies. 1987. 'Targetted Correction of a Mutant HPRT Gene in Mouse Embryonic Stem Cells'. *Nature* 330 (6148): 576–78. https://doi.org/10.1038/330576a0.

Dunn, S.-J., G. Martello, B. Yordanov, S. Emmott, and A. G. Smith. 2014. 'Defining an Essential Transcription Factor Program for Naïve Pluripotency'. *Science* 344 (6188): 1156–60. https://doi.org/10.1126/science.1248882.

Duval, David, Béatrice Reinhardt, Claude Kedinger, and Hélène Boeuf. 2000. 'Role of Suppressors of Cytokine Signaling (Socs) in Leukemia Inhibitory Factor (LIF) - Dependent Embryonic Stem Cell Survival'. *The FASEB Journal* 14 (11): 1577–84. https://doi.org/10.1096/fj.99-0810com.

Ebata, Kevin T., Kathryn Mesh, Shichong Liu, Misha Bilenky, Alexander Fekete, Michael G. Acker, Martin Hirst, Benjamin A. Garcia, and Miguel Ramalho-Santos. 2017. 'Vitamin C Induces Specific Demethylation of H3K9me2 in Mouse Embryonic Stem Cells via Kdm3a/B'. *Epigenetics & Chromatin* 10: 36. https://doi.org/10.1186/s13072-017-0143-3.

Eckersley-Maslin, Mélanie A., and David L. Spector. 2014. 'Random Monoallelic Expression: Regulating Gene Expression One Allele at a Time'. *Trends in Genetics* 30 (6): 237–44. https://doi.org/10.1016/j.tig.2014.03.003.

Eckersley-Maslin, Mélanie A., Valentine Svensson, Christel Krueger, Thomas M. Stubbs, Pascal Giehr, Felix Krueger, Ricardo J. Miragaia, et al. 2016. 'MERVL/Zscan4 Network Activation Results in Transient Genome-Wide DNA Demethylation of MESCs'. *Cell Reports* 17 (1): 179–92. https://doi.org/10.1016/j.celrep.2016.08.087.

Endoh, Mitsuhiro, Takaho A. Endo, Tamie Endoh, Yu-ichi Fujimura, Osamu Ohara, Tetsuro Toyoda, Arie P. Otte, et al. 2008. 'Polycomb Group Proteins Ring1A/B Are Functionally Linked to the Core Transcriptional Regulatory Circuitry to Maintain ES Cell Identity'. *Development (Cambridge, England)* 135 (8): 1513–24. https://doi.org/10.1242/dev.014340.

Engreitz, Jesse M., Jenna E. Haines, Elizabeth M. Perez, Glen Munson, Jenny Chen, Michael Kane, Patrick E. McDonel, Mitchell Guttman, and Eric S. Lander. 2016. 'Local Regulation of Gene Expression by LncRNA Promoters, Transcription and Splicing'. *Nature* 539 (7629): 452–55. https://doi.org/10.1038/nature20149.

Epinat, Jean-Charles, Sylvain Arnould, Patrick Chames, Pascal Rochaix, Dominique Desfontaines, Clémence Puzin, Amélie Patin, Alexandre Zanghellini, Frédéric Pâques, and Emmanuel Lacroix. 2003. 'A Novel Engineered Meganuclease Induces Homologous Recombination in Yeast and Mammalian Cells'. *Nucleic Acids Research* 31 (11): 2952–62.

Evans, M. J., and M. H. Kaufman. 1981. 'Establishment in Culture of Pluripotential Cells from Mouse Embryos'. *Nature* 292 (5819): 154–56.

Faddah, Dina A., Haoyi Wang, Albert Wu Cheng, Yarden Katz, Yosef Buganim, and Rudolf Jaenisch. 2013. 'Single-Cell Analysis Reveals That Expression of Nanog Is Biallelic and Equally Variable as That of Other Pluripotency Factors in Mouse ESCs'. *Cell Stem Cell* 13 (1): 23–29. https://doi.org/10.1016/j.stem.2013.04.019.

Farzadfard, Fahim, Samuel D. Perli, and Timothy K. Lu. 2013. 'Tunable and Multifunctional Eukaryotic Transcription Factors Based on CRISPR/Cas'. *ACS Synthetic Biology* 2 (10): 604–13. https://doi.org/10.1021/sb400081r.

Ferrari, Karin J., Andrea Scelfo, SriGanesh Jammula, Alessandro Cuomo, Iros Barozzi, Alexandra Stützer, Wolfgang Fischle, Tiziana Bonaldi, and Diego Pasini. 2014. 'Polycomb-Dependent H3K27me1 and H3K27me2 Regulate Active Transcription and Enhancer Fidelity'. *Molecular Cell* 53 (1): 49–62. https://doi.org/10.1016/j.molcel.2013.10.030.

Festuccia, Nicola, Agnès Dubois, Sandrine Vandormael-Pournin, Elena Gallego Tejeda, Adrien Mouren, Sylvain Bessonnard, Florian Mueller, Caroline Proux, Michel Cohen-

Tannoudji, and Pablo Navarro. 2016. ‘Mitotic Binding of Esrrb Marks Key Regulatory Regions of the Pluripotency Network’. *Nature Cell Biology* 18 (11): 1139–48. https://doi.org/10.1038/ncb3418.

Festuccia, Nicola, Rodrigo Osorno, Florian Halbritter, Violetta Karwacki-Neisius, Pablo Navarro, Douglas Colby, Frederick Wong, Adam Yates, Simon R. Tomlinson, and Ian Chambers. 2012. ‘Esrrb Is a Direct Nanog Target Gene That Can Substitute for Nanog Function in Pluripotent Cells’. *Cell Stem Cell* 11 (4): 477–90. https://doi.org/10.1016/j.stem.2012.08.002.

Fidalgo, Miguel, Francesco Faiola, Carlos-Filipe Pereira, Junjun Ding, Arven Saunders, Julian Gingold, Christoph Schaniel, Ihor R. Lemischka, José C. R. Silva, and Jianlong Wang. 2012. ‘Zfp281 Mediates Nanog Autorepression through Recruitment of the NuRD Complex and Inhibits Somatic Cell Reprogramming’. *Proceedings of the National Academy of Sciences of the United States of America* 109 (40): 16202–7. https://doi.org/10.1073/pnas.1208533109.

Filipczyk, Adam, Konstantinos Gkatzis, Jun Fu, Philipp S. Hoppe, Heiko Lickert, Konstantinos Anastassiadis, and Timm Schroeder. 2013. ‘Biallelic Expression of Nanog Protein in Mouse Embryonic Stem Cells’. *Cell Stem Cell* 13 (1): 12–13. https://doi.org/10.1016/j.stem.2013.04.025.

Firestein, Ron, Xiangmin Cui, Phil Huie, and Michael L. Cleary. 2000. ‘Set Domain-Dependent Regulation of Transcriptional Silencing and Growth Control by SUV39H1, a Mammalian Ortholog OfDrosophila Su(Var)3-9’. *Molecular and Cellular Biology* 20 (13): 4900–4909. https://doi.org/10.1128/MCB.20.13.4900-4909.2000.

Forrai, Ariel, Kristy Boyle, Adam H. Hart, Lynne Hartley, Steven Rakar, Tracy A. Willson, Ken M. Simpson, et al. 2006. ‘Absence of Suppressor of Cytokine Signalling 3 Reduces Self-Renewal and Promotes Differentiation in Murine Embryonic Stem Cells’. *STEM CELLS* 24 (3): 604–14. https://doi.org/10.1634/stemcells.2005-0323.

Fort, Alexandre, Daisuke Yamada, Kosuke Hashimoto, Haruhiko Koseki, and Piero Carninci. 2015. ‘Nuclear Transcriptome Profiling of Induced Pluripotent Stem Cells and Embryonic Stem Cells Identify Non-Coding Loci Resistant to Reprogramming’. *Cell Cycle* 14 (8): 1148–55. https://doi.org/10.4161/15384101.2014.988031.

Frankenberg, Stephen, François Gerbe, Sylvain Bessonnard, Corinne Belville, Pierre Pouchin, Olivier Bardot, and Claire Chazaud. 2011. ‘Primitive Endoderm Differentiates via a Three-Step Mechanism Involving Nanog and RTK Signaling’. *Developmental Cell* 21 (6): 1005–13. https://doi.org/10.1016/j.devcel.2011.10.019.

Fraser, M. J., T. Clszczon, T. Elick, and C. Bauser. 1996. ‘Precise Excision of TTAA-Specific Lepidopteran Transposons PiggyBac (IFP2) and Tagalong (TFP3) from the Baculovirus Genome in Cell Lines from Two Species of Lepidoptera’. *Insect Molecular Biology* 5 (2): 141–51. https://doi.org/10.1111/j.1365-2583.1996.tb00048.x.

Fraser, Malcolm J., Lynne Cary, Kitima Boonvisudhi, and Hwei-Gene Heidi Wang. 1995. 'Assay for Movement of Lepidopteran Transposon IFP2 in Insect Cells Using a Baculovirus Genome as a Target DNA'. *Virology* 211 (2): 397–407. https://doi.org/10.1006/viro.1995.1422.

Friedrich, G., and P. Soriano. 1991. 'Promoter Traps in Embryonic Stem Cells: A Genetic Screen to Identify and Mutate Developmental Genes in Mice.' *Genes & Development* 5 (9): 1513–23. https://doi.org/10.1101/gad.5.9.1513.

Fritsch, Lauriane, Philippe Robin, Jacques R. R. Mathieu, Mouloud Souidi, Hélène Hinaux, Claire Rougeulle, Annick Harel-Bellan, Maya Ameyar-Zazoua, and Slimane Ait-Si-Ali. 2010. 'A Subset of the Histone H3 Lysine 9 Methyltransferases Suv39h1, G9a, GLP, and SETDB1 Participate in a Multimeric Complex'. *Molecular Cell* 37 (1): 46–56. https://doi.org/10.1016/j.molcel.2009.12.017.

Fujikura, Junji, Eiji Yamato, Shigenobu Yonemura, Kiminori Hosoda, Shinji Masui, Kazuwa Nakao, Jun-ichi Miyazaki, and Hitoshi Niwa. 2002. 'Differentiation of Embryonic Stem Cells Is Induced by GATA Factors'. *Genes & Development* 16 (7): 784–89. https://doi.org/10.1101/gad.968802.

Fulco, Charles P., Mathias Munschauer, Rockwell Anyoha, Glen Munson, Sharon R. Grossman, Elizabeth M. Perez, Michael Kane, Brian Cleary, Eric S. Lander, and Jesse M. Engreitz. 2016. 'Systematic Mapping of Functional Enhancer-Promoter Connections with CRISPR Interference'. *Science (New York, N.Y.)* 354 (6313): 769–73. https://doi.org/10.1126/science.aag2445.

Gagliardi, Alessia, Nicholas P. Mullin, Zi Ying Tan, Douglas Colby, Anastasia I. Kousa, Florian Halbritter, Jason T. Weiss, et al. 2013. 'A Direct Physical Interaction between Nanog and Sox2 Regulates Embryonic Stem Cell Self-renewal'. *The EMBO Journal* 32 (16): 2231–47. https://doi.org/10.1038/emboj.2013.161.

Gajović, Srećko, Luc St-Onge, Yoshifumi Yokota, and Peter Gruss. 1998. 'Retinoic Acid Mediates Pax6 Expression during in Vitro Differentiation of Embryonic Stem Cells'. *Differentiation* 62 (4): 187–92. https://doi.org/10.1046/j.1432-0436.1998.6240187.x.

Galonska, Christina, Jocelyn Charlton, Alexandra L. Mattei, Julie Donaghey, Kendell Clement, Hongcang Gu, Arman W. Mohammad, et al. 2018. 'Genome-Wide Tracking of DCas9-Methyltransferase Footprints'. *Nature Communications* 9 (1): 597. https://doi.org/10.1038/s41467-017-02708-5.

Galonska, Christina, Michael J. Ziller, Rahul Karnik, and Alexander Meissner. 2015. 'Ground State Conditions Induce Rapid Reorganization of Core Pluripotency Factor Binding before Global Epigenetic Reprogramming'. *Cell Stem Cell* 17 (4): 462–70. https://doi.org/10.1016/j.stem.2015.07.005.

Gao, Xuefei, Jason C.H. Tsang, Fortis Gaba, Donghai Wu, Liming Lu, and Pentao Liu. 2014. 'Comparison of TALE Designer Transcription Factors and the CRISPR/DCas9 in Regulation of Gene Expression by Targeting Enhancers'. *Nucleic Acids Research* 42 (20): e155. https://doi.org/10.1093/nar/gku836.

Garneau, Josiane E., Marie-Ève Dupuis, Manuela Villion, Dennis A. Romero, Rodolphe Barrangou, Patrick Boyaval, Christophe Fremaux, Philippe Horvath, Alfonso H. Magadán, and Sylvain Moineau. 2010. 'The CRISPR/Cas Bacterial Immune System Cleaves Bacteriophage and Plasmid DNA'. *Nature* 468 (7320): 67–71. https://doi.org/10.1038/nature09523.

Gasiunas, Giedrius, Rodolphe Barrangou, Philippe Horvath, and Virginijus Siksnys. 2012. 'Cas9–CrRNA Ribonucleoprotein Complex Mediates Specific DNA Cleavage for Adaptive Immunity in Bacteria'. *Proceedings of the National Academy of Sciences of the United States of America* 109 (39): E2579–86. https://doi.org/10.1073/pnas.1208507109.

Gearing, D. P., M. R. Comeau, D. J. Friend, S. D. Gimpel, C. J. Thut, J. McGourty, K. K. Brasher, J. A. King, S. Gillis, and B. Mosley. 1992. 'The IL-6 Signal Transducer, Gp130: An Oncostatin M Receptor and Affinity Converter for the LIF Receptor'. *Science (New York, N.Y.)* 255 (5050): 1434–37.

Gearing, D P, N M Gough, J A King, D J Hilton, N A Nicola, R J Simpson, E C Nice, A Kelso, and D Metcalf. 1987. 'Molecular Cloning and Expression of CDNA Encoding a Murine Myeloid Leukaemia Inhibitory Factor (LIF).' *The EMBO Journal* 6 (13): 3995–4002.

Gearing, D. P., C. J. Thut, T. VandeBos, S. D. Gimpel, P. B. Delaney, J. King, V. Price, D. Cosman, and M. P. Beckmann. 1991. 'Leukemia Inhibitory Factor Receptor Is Structurally Related to the IL-6 Signal Transducer, Gp130'. *The EMBO Journal* 10 (10): 2839–48.

Ghimire, Sabitri, Margot Van der Jeught, Jitesh Neupane, Matthias S. Roost, Jasper Anckaert, Mina Popovic, Filip Van Nieuwerburgh, et al. 2018. 'Comparative Analysis of Naive, Primed and Ground State Pluripotency in Mouse Embryonic Stem Cells Originating from the Same Genetic Background'. *Scientific Reports* 8 (1): 5884. https://doi.org/10.1038/s41598-018-24051-5.

Gil, Jesús, and Ana O'Loghlen. 2014. 'PRC1 Complex Diversity: Where Is It Taking Us?' *Trends in Cell Biology* 24 (11): 632–41. https://doi.org/10.1016/j.tcb.2014.06.005.

Gilbert, Luke A., Matthew H. Larson, Leonardo Morsut, Zairan Liu, Gloria A. Brar, Sandra E. Torres, Noam Stern-Ginossar, et al. 2013. 'CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes'. *Cell* 154 (2): 442–51. https://doi.org/10.1016/j.cell.2013.06.044.

Gilbert, Luke A., Max A. Horlbeck, Britt Adamson, Jacqueline E. Villalta, Yuwen Chen, Evan H. Whitehead, Carla Guimaraes, et al. 2014. 'Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation'. *Cell* 159 (3): 647–61. https://doi.org/10.1016/j.cell.2014.09.029.

Gossen, M., and H. Bujard. 1992. 'Tight Control of Gene Expression in Mammalian Cells by Tetracycline-Responsive Promoters.' *Proceedings of the National Academy of Sciences* 89 (12): 5547–51. https://doi.org/10.1073/pnas.89.12.5547.

Govin, Jérôme, Emmanuelle Escoffier, Sophie Rousseaux, Lauriane Kuhn, Myriam Ferro, Julien Thévenon, Raffaella Catena, et al. 2007. 'Pericentric Heterochromatin Reprogramming by New Histone Variants during Mouse Spermiogenesis'. *J Cell Biol* 176 (3): 283–94. https://doi.org/10.1083/jcb.200604141.

Grabole, Nils, Julia Tischler, Jamie A Hackett, Shinseog Kim, Fuchou Tang, Harry G Leitch, Erna Magnúsdóttir, and M Azim Surani. 2013. 'Prdm14 Promotes Germline Fate and Naive Pluripotency by Repressing FGF Signalling and DNA Methylation'. *EMBO Reports* 14 (7): 629–37. https://doi.org/10.1038/embor.2013.67.

Guilinger, John P., Vikram Pattanayak, Deepak Reyon, Shengdar Q. Tsai, Jeffry D. Sander, J. Keith Joung, and David R. Liu. 2014. 'Broad Specificity Profiling of TALENs Results in Engineered Nucleases With Improved DNA Cleavage Specificity'. *Nature Methods* 11 (4): 429–35. https://doi.org/10.1038/nmeth.2845.

Guo, Fan, Lin Li, Jingyun Li, Xinglong Wu, Boqiang Hu, Ping Zhu, Lu Wen, and Fuchou Tang. 2017. 'Single-Cell Multi-Omics Sequencing of Mouse Early Embryos and Embryonic Stem Cells'. *Cell Research* 27 (8): 967–88. https://doi.org/10.1038/cr.2017.82.

Guo, Ge, Yue Huang, Peter Humphreys, Xiaozhong Wang, and Austin Smith. 2011. 'A PiggyBac-Based Recessive Screening Method to Identify Pluripotency Regulators'. *PLOS ONE* 6 (4): e18189. https://doi.org/10.1371/journal.pone.0018189.

Guo, Ge, and Austin Smith. 2010. 'A Genome-Wide Screen in EpiSCs Identifies Nr5a Nuclear Receptors as Potent Inducers of Ground State Pluripotency'. *Development* 137 (19): 3185–92. https://doi.org/10.1242/dev.052753.

Guo, Ge, Jian Yang, Jennifer Nichols, John Simon Hall, Isobel Eyres, William Mansfield, and Austin Smith. 2009. 'Klf4 Reverts Developmentally Programmed Restriction of Ground State Pluripotency'. *Development* 136 (7): 1063–69. https://doi.org/10.1242/dev.030957.

Guttman, M. 2009. 'Chromatin Signature Reveals over a Thousand Highly Conserved Large Non-Coding RNAs in Mammals'. *Nature* 458: 223–27. https://doi.org/10.1038/nature07672.

Guttman, Mitchell, Ido Amit, Manuel Garber, Courtney French, Michael F. Lin, David Feldser, Maite Huarte, et al. 2009. 'Chromatin Signature Reveals over a Thousand Highly Conserved Large Non-Coding RNAs in Mammals'. *Nature* 458 (7235): 223–27. https://doi.org/10.1038/nature07672.

Guttman, Mitchell, Julie Donaghey, Bryce W. Carey, Manuel Garber, Jennifer K. Grenier, Glen Munson, Geneva Young, et al. 2011. 'LincRNAs Act in the Circuitry Controlling Pluripotency and Differentiation'. *Nature* 477 (7364): 295–300. https://doi.org/10.1038/nature10398.

Guttman, Mitchell, Manuel Garber, Joshua Z. Levin, Julie Donaghey, James Robinson, Xian Adiconis, Lin Fan, et al. 2010. 'Ab Initio Reconstruction of Cell Type-Specific Transcriptomes in Mouse Reveals the Conserved Multi-Exonic Structure of LincRNAs'. *Nature Biotechnology* 28 (5): 503–10. https://doi.org/10.1038/nbt.1633.

Hackett, Jamie A., and M. Azim Surani. 2014. 'Regulatory Principles of Pluripotency: From the Ground State Up'. *Cell Stem Cell* 15 (4): 416–30. https://doi.org/10.1016/j.stem.2014.09.015.

Haegele, Lorenz, Barbara Ingold, Heike Naumann, Ghazaleh Tabatabai, Birgit Ledermann, and Sebastian Brandner. 2003. 'Wnt Signalling Inhibits Neural Differentiation of Embryonic Stem Cells by Controlling Bone Morphogenetic Protein Expression'. *Molecular and Cellular Neuroscience* 24 (3): 696–708. https://doi.org/10.1016/S1044-7431(03)00232-X.

Haft, Daniel H., Jeremy Selengut, Emmanuel F. Mongodin, and Karen E. Nelson. 2005. 'A Guild of 45 CRISPR-Associated (Cas) Protein Families and Multiple CRISPR/Cas Subtypes Exist in Prokaryotic Genomes'. *PLOS Computational Biology* 1 (6): e60. https://doi.org/10.1371/journal.pcbi.0010060.

Hall, John, Ge Guo, Jason Wray, Isobel Eyres, Jennifer Nichols, Lars Grotewold, Sofia Morfopoulou, et al. 2009a. 'Oct4 and LIF/Stat3 Additively Induce Krüppel Factors to Sustain Embryonic Stem Cell Self-Renewal'. *Cell Stem Cell* 5 (6): 597–609. https://doi.org/10.1016/j.stem.2009.11.003.

———. 2009b. 'Oct4 and LIF/Stat3 Additively Induce Krüppel Factors to Sustain Embryonic Stem Cell Self-Renewal'. *Cell Stem Cell* 5 (6): 597–609. https://doi.org/10.1016/j.stem.2009.11.003.

Hamilton, William B., and Joshua M. Brickman. 2014. 'Erk Signaling Suppresses Embryonic Stem Cell Self-Renewal to Specify Endoderm'. *Cell Reports* 9 (6): 2056–70. https://doi.org/10.1016/j.celrep.2014.11.032.

Hanes, Jozef, Lutz Jermutus, Susanne Weber-Bornhauser, Hans Rudolf Bosshard, and Andreas Plückthun. 1998. 'Ribosome Display Efficiently Selects and Evolves High-Affinity

Antibodies in Vitro from Immune Libraries'. *Proceedings of the National Academy of Sciences* 95 (24): 14130–35. https://doi.org/10.1073/pnas.95.24.14130.

Harms, Frederike Leonie, Katta M. Girisha, Andrew A. Hardigan, Fanny Kortüm, Anju Shukla, Malik Alawi, Ashwin Dalal, et al. 2017. 'Mutations in EBF3 Disturb Transcriptional Profiles and Cause Intellectual Disability, Ataxia, and Facial Dysmorphism'. *American Journal of Human Genetics* 100 (1): 117–27. https://doi.org/10.1016/j.ajhg.2016.11.012.

Harrington, John J., Bruce Sherf, Stephen Rundlett, P. David Jackson, Rob Perry, Scott Cain, Christina Leventhal, et al. 2001. 'Creation of Genome-Wide Protein Expression Libraries Using Random Activation of Gene Expression'. *Nature Biotechnology* 19 (5): 440–45. https://doi.org/10.1038/88107.

Hart, Adam H., Lynne Hartley, Marilyn Ibrahim, and Lorraine Robb. 2004. 'Identification, Cloning and Expression Analysis of the Pluripotency Promoting Nanog Genes in Mouse and Human'. *Developmental Dynamics: An Official Publication of the American Association of Anatomists* 230 (1): 187–98. https://doi.org/10.1002/dvdy.20034.

Hastreiter, Simon, Stavroula Skylaki, Dirk Loeffler, Andreas Reimann, Oliver Hilsenbeck, Philipp S. Hoppe, Daniel L. Coutu, et al. 2018. 'Inductive and Selective Effects of GSK3 and MEK Inhibition on Nanog Heterogeneity in Embryonic Stem Cells'. *Stem Cell Reports*, May. https://doi.org/10.1016/j.stemcr.2018.04.019.

Heijden, G. W. van der, A. A. H. A. Derijck, L. Ramos, M. Giele, J. van der Vlag, and P. de Boer. 2006. 'Transmission of Modified Nucleosomes from the Mouse Male Germline to the Zygote and Subsequent Remodeling of Paternal Chromatin'. *Developmental Biology* 298 (2): 458–69. https://doi.org/10.1016/j.ydbio.2006.06.051.

Hendrickson, Peter G., Jessie A. Doráis, Edward J. Grow, Jennifer L. Whiddon, Jong-Won Lim, Candice L. Wike, Bradley D. Weaver, et al. 2017. 'Conserved Roles for Murine DUX and Human DUX4 in Activating Cleavage Stage Genes and MERVL/HERVL Retrotransposons'. *Nature Genetics* 49 (6): 925–34. https://doi.org/10.1038/ng.3844.

Hezroni, Hadas, David Koppstein, Matthew G. Schwartz, Alexandra Avrutin, David P. Bartel, and Igor Ulitsky. 2015. 'Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of Transcriptomes in 17 Species'. *Cell Reports* 11 (7): 1110–22. https://doi.org/10.1016/j.celrep.2015.04.023.

Hille, Frank, and Emmanuelle Charpentier. 2016. 'CRISPR-Cas: Biology, Mechanisms and Relevance'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 371 (1707). https://doi.org/10.1098/rstb.2015.0496.

Hilton, Isaac B., Anthony M. D'Ippolito, Christopher M. Vockley, Pratiksha I. Thakore, Gregory E. Crawford, Timothy E. Reddy, and Charles A. Gersbach. 2015. 'Epigenome

Editing by a CRISPR/Cas9-Based Acetyltransferase Activates Genes from Promoters and Enhancers'. *Nature Biotechnology* 33 (5): 510–17. https://doi.org/10.1038/nbt.3199.

Ho, Lena, Raja Jothi, Jehnna L. Ronan, Kairong Cui, Keji Zhao, and Gerald R. Crabtree. 2009. 'An Embryonic Stem Cell Chromatin Remodeling Complex, EsBAF, Is an Essential Component of the Core Pluripotency Transcriptional Network'. *Proceedings of the National Academy of Sciences* 106 (13): 5187–91. https://doi.org/10.1073/pnas.0812888106.

Ho, Lena, Erik L. Miller, Jehnna L. Ronan, Wen Qi Ho, Raja Jothi, and Gerald R. Crabtree. 2011. 'EsBAF Facilitates Pluripotency by Conditioning the Genome for LIF/STAT3 Signalling and by Regulating Polycomb Function'. *Nature Cell Biology* 13 (8): 903–13. https://doi.org/10.1038/ncb2285.

Hoffmeyer, Katrin, Dirk Junghans, Benoit Kanzler, and Rolf Kemler. 2017. 'Trimethylation and Acetylation of β-Catenin at Lysine 49 Represent Key Elements in ESC Pluripotency'. *Cell Reports* 18 (12): 2815–24. https://doi.org/10.1016/j.celrep.2017.02.076.

Holoch, Daniel, and Raphaël Margueron. 2017. 'Mechanisms Regulating PRC2 Recruitment and Enzymatic Activity'. *Trends in Biochemical Sciences* 42 (7): 531–42. https://doi.org/10.1016/j.tibs.2017.04.003.

Home, Pratik, Ram Parikshan Kumar, Avishek Ganguly, Biswarup Saha, Jessica Milano-Foster, Bhaswati Bhattacharya, Soma Ray, et al. 2017. 'Genetic Redundancy of GATA Factors in the Extraembryonic Trophoblast Lineage Ensures the Progression of Preimplantation and Postimplantation Mammalian Development'. *Development (Cambridge, England)* 144 (5): 876–88. https://doi.org/10.1242/dev.145318.

Horn, Carsten, Jens Hansen, Frank Schnütgen, Claudia Seisenberger, Thomas Floss, Markus Irgang, Silke De-Zolt, Wolfgang Wurst, Harald von Melchner, and Patricia Ruiz Noppinger. 2007. 'Splinkerette PCR for More Efficient Characterization of Gene Trap Events'. *Nature Genetics* 39 (8): 933–34. https://doi.org/10.1038/ng0807-933.

Hotta, Akitsu, and James Ellis. 2008. 'Retroviral Vector Silencing during IPS Cell Induction: An Epigenetic Beacon That Signals Distinct Pluripotent States'. *Journal of Cellular Biochemistry* 105 (4): 940–48. https://doi.org/10.1002/jcb.21912.

Hsu, Patrick D, David A Scott, Joshua A Weinstein, F Ann Ran, Silvana Konermann, Vineeta Agarwala, Yinqing Li, et al. 2013. 'DNA Targeting Specificity of RNA-Guided Cas9 Nucleases'. *Nature Biotechnology* 31 (9): 827–32. https://doi.org/10.1038/nbt.2647.

Huang, Delun, Ling Wang, Jingyue Duan, Chang Huang, Xiuchun (Cindy) Tian, Ming Zhang, and Young Tang. 2018. 'LIF-Activated Jak Signaling Determines Esrrb Expression

during Late-Stage Reprogramming'. *Biology Open* 7 (1): bio029264. https://doi.org/10.1242/bio.029264.

Huang, Yali, Rodrigo Osorno, Anestis Tsakiridis, and Valerie Wilson. 2012. 'In Vivo Differentiation Potential of Epiblast Stem Cells Revealed by Chimeric Embryo Formation'. *Cell Reports* 2 (6): 1571–78. https://doi.org/10.1016/j.celrep.2012.10.022.

Hung, Sandy S. C., Raymond C. B. Wong, Alexei A. Sharov, Yuhki Nakatake, Hong Yu, and Minoru S. H. Ko. 2013. 'Repression of Global Protein Synthesis by Eif1a-Like Genes That Are Expressed Specifically in the Two-Cell Embryos and the Transient Zscan4-Positive State of Embryonic Stem Cells'. *DNA Research* 20 (4): 391–402. https://doi.org/10.1093/dnares/dst018.

Hunkapiller, Julie, Yin Shen, Aaron Diaz, Gerard Cagney, David McCleary, Miguel Ramalho-Santos, Nevan Krogan, Bing Ren, Jun S. Song, and Jeremy F. Reiter. 2012. 'Polycomb-Like 3 Promotes Polycomb Repressive Complex 2 Binding to CpG Islands and Embryonic Stem Cell Self-Renewal'. *PLOS Genetics* 8 (3): e1002576. https://doi.org/10.1371/journal.pgen.1002576.

Huston, J S, D Levinson, M Mudgett-Hunter, M S Tai, J Novotný, M N Margolies, R J Ridge, R E Bruccoleri, E Haber, and R Crea. 1988. 'Protein Engineering of Antibody Binding Sites: Recovery of Specific Activity in an Anti-Digoxin Single-Chain Fv Analogue Produced in Escherichia Coli.' *Proceedings of the National Academy of Sciences of the United States of America* 85 (16): 5879–83.

Illingworth, Robert S., Jurriaan J. Hölzenspies, Fabian V. Roske, Wendy A. Bickmore, and Joshua M. Brickman. 2016. 'Polycomb Enables Primitive Endoderm Lineage Priming in Embryonic Stem Cells'. *ELife* 5 (October): e14926. https://doi.org/10.7554/eLife.14926.

Ishino, Y., H. Shinagawa, K. Makino, M. Amemura, and A. Nakata. 1987. 'Nucleotide Sequence of the Iap Gene, Responsible for Alkaline Phosphatase Isozyme Conversion in Escherichia Coli, and Identification of the Gene Product.' *Journal of Bacteriology* 169 (12): 5429–33. https://doi.org/10.1128/jb.169.12.5429-5433.1987.

Ishiuchi, Takashi, Rocio Enriquez-Gasca, Eiji Mizutani, Ana Bošković, Celine Ziegler-Birling, Diego Rodriguez-Terrones, Teruhiko Wakayama, Juan M. Vaquerizas, and Maria-Elena Torres-Padilla. 2015. 'Early Embryonic-like Cells Are Induced by Downregulating Replication-Dependent Chromatin Assembly'. *Nature Structural & Molecular Biology* 22 (9): 662–71. https://doi.org/10.1038/nsmb.3066.

Iwai, Ryota, Hidenori Tabata, Mayuko Inoue, Kei-Ichiro Nomura, Tadashi Okamoto, Masamitsu Ichihashi, Koh-Ichi Nagata, and Ken-Ichi Mizutani. 2018. 'A Prdm8 Target Gene Ebf3 Regulates Multipolar-to-Bipolar Transition in Migrating Neocortical Cells'. *Biochemical and Biophysical Research Communications* 495 (1): 388–94. https://doi.org/10.1016/j.bbrc.2017.11.021.

Jansen, Ruud, Jan D. A. van Embden, Wim Gaastra, and Leo M. Schouls. 2002. 'Identification of Genes That Are Associated with DNA Repeats in Prokaryotes'. *Molecular Microbiology* 43 (6): 1565–75. https://doi.org/10.1046/j.1365-2958.2002.02839.x.

Jauch, Ralf, Calista Keow Leng Ng, Kumar Singh Saikatendu, Raymond C. Stevens, and Prasanna R. Kolatkar. 2008. 'Crystal Structure and DNA Binding of the Homeodomain of the Stem Cell Transcription Factor Nanog'. *Journal of Molecular Biology* 376 (3): 758–70. https://doi.org/10.1016/j.jmb.2007.11.091.

Jensen, Kristopher Torp, Lasse Fløe, Trine Skov Petersen, Jinrong Huang, Fengping Xu, Lars Bolund, Yonglun Luo, and Lin Lin. 2017. 'Chromatin Accessibility and Guide Sequence Secondary Structure Affect CRISPR-Cas9 Gene Editing Efficiency'. *FEBS Letters* 591 (13): 1892–1901. https://doi.org/10.1002/1873-3468.12707.

Jeon, Hyojung, Tsuyoshi Waku, Takuya Azami, Le Tran Phuc Khoa, Jun Yanagisawa, Satoru Takahashi, and Masatsugu Ema. 2016. 'Comprehensive Identification of Krüppel-Like Factor Family Members Contributing to the Self-Renewal of Mouse Embryonic Stem Cells and Cellular Reprogramming'. *PLoS ONE* 11 (3). https://doi.org/10.1371/journal.pone.0150715.

Ji, Zhe, Ruisheng Song, Aviv Regev, and Kevin Struhl. 2015. 'Many LncRNAs, 5'UTRs, and Pseudogenes Are Translated and Some Are Likely to Express Functional Proteins'. ELife. 19 December 2015. https://doi.org/10.7554/eLife.08890.

Jiang, Fuguo, and Jennifer A. Doudna. 2017. 'CRISPR–Cas9 Structures and Mechanisms'. *Annual Review of Biophysics* 46 (1): 505–29. https://doi.org/10.1146/annurev-biophys-062215-010822.

Jiang, Jianming, Yun-Shen Chan, Yuin-Han Loh, Jun Cai, Guo-Qing Tong, Ching-Aeng Lim, Paul Robson, Sheng Zhong, and Huck-Hui Ng. 2008. 'A Core Klf Circuitry Regulates Self-Renewal of Embryonic Stem Cells'. *Nature Cell Biology* 10 (3): 353–60. https://doi.org/10.1038/ncb1698.

Jiapaer, Zeyidan, Guoping Li, Dan Ye, Mingliang Bai, Jianguo Li, Xudong Guo, Yanhua Du, et al. 2018. 'LincU Preserves Naïve Pluripotency by Restricting ERK Activity in Embryonic Stem Cells'. *Stem Cell Reports*, July. https://doi.org/10.1016/j.stemcr.2018.06.010.

Jin, Jiali, Jian Liu, Cong Chen, Zhenping Liu, Cong Jiang, Hongshang Chu, Weijuan Pan, et al. 2016. 'The Deubiquitinase USP21 Maintains the Stemness of Mouse Embryonic Stem Cells via Stabilization of Nanog'. *Nature Communications* 7 (November): 13594. https://doi.org/10.1038/ncomms13594.

Jinek, Martin, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A. Doudna, and Emmanuelle Charpentier. 2012. 'A Programmable Dual-RNA–Guided DNA

Endonuclease in Adaptive Bacterial Immunity'. *Science* 337 (6096): 816–21. https://doi.org/10.1126/science.1225829.

Jinek, Martin, Fuguo Jiang, David W. Taylor, Samuel H. Sternberg, Emine Kaya, Enbo Ma, Carolin Anders, et al. 2014. 'Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation'. *Science* 343 (6176): 1247997. https://doi.org/10.1126/science.1247997.

Joo, Jin Young, Hyun Woo Choi, Min Jung Kim, Holm Zaehres, Natalia Tapia, Martin Stehling, Koo Sung Jung, Jeong Tae Do, and Hans R. Schöler. 2014. 'Establishment of a Primed Pluripotent Epiblast Stem Cell in FGF4-Based Conditions'. *Scientific Reports* 4 (December): 7477. https://doi.org/10.1038/srep07477.

Joung, Julia, Jesse M. Engreitz, Silvana Konermann, Omar O. Abudayyeh, Vanessa K. Verdine, Francois Aguet, Jonathan S. Gootenberg, et al. 2017. 'Genome-Scale Activation Screen Identifies a LncRNA Locus Regulating a Gene Neighbourhood'. *Nature* 548 (7667): 343–46. https://doi.org/10.1038/nature23451.

Kahan, B. W., and B. Ephrussi. 1970. 'Developmental Potentialities of Clonal in Vitro Cultures of Mouse Testicular Teratoma'. *Journal of the National Cancer Institute* 44 (5): 1015–36.

Kaneko, Syuzo, Roberto Bonasio, Ricardo Saldaña-Meyer, Takahaki Yoshida, Jinsook Son, Koichiro Nishino, Akihiro Umezawa, and Danny Reinberg. 2014a. 'Interactions between JARID2 and Noncoding RNAs Regulate PRC2 Recruitment to Chromatin'. *Molecular Cell* 53 (2): 290–300. https://doi.org/10.1016/j.molcel.2013.11.012.

———. 2014b. 'Interactions between JARID2 and Noncoding RNAs Regulate PRC2 Recruitment to Chromatin'. *Molecular Cell* 53 (2): 290–300. https://doi.org/10.1016/j.molcel.2013.11.012.

Kaneko, Syuzo, Jinsook Son, Roberto Bonasio, Steven S. Shen, and Danny Reinberg. 2014. 'Nascent RNA Interaction Keeps PRC2 Activity Poised and in Check'. *Genes & Development* 28 (18): 1983–88. https://doi.org/10.1101/gad.247940.114.

Kaneko, Syuzo, Jinsook Son, Steven S Shen, Danny Reinberg, and Roberto Bonasio. 2013. 'PRC2 Binds Active Promoters and Contacts Nascent RNAs in Embryonic Stem Cells'. *Nature Structural & Molecular Biology* 20 (11): 1258–64. https://doi.org/10.1038/nsmb.2700.

Kang, Minjung, Anna Piliszek, Jérôme Artus, and Anna-Katerina Hadjantonakis. 2013. 'FGF4 Is Required for Lineage Restriction and Salt-and-Pepper Distribution of Primitive Endoderm Factors but Not Their Initial Expression in the Mouse'. *Development (Cambridge, England)* 140 (2): 267–79. https://doi.org/10.1242/dev.084996.

Karlic, Rosa, Sravya Ganesh, Vedran Franke, Eliska Svobodova, Jana Urbanova, Yutaka Suzuki, Fugaku Aoki, Kristian Vlahovicek, and Petr Svoboda. 2017. 'Long Non-Coding RNA Exchange during the Oocyte-to-Embryo Transition in Mice'. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes* 24 (2): 129–41. https://doi.org/10.1093/dnares/dsw058.

Karwacki-Neisius, Violetta, Jonathan Göke, Rodrigo Osorno, Florian Halbritter, Jia Hui Ng, Andrea Y. Weiße, Frederick C. K. Wong, et al. 2013. 'Reduced Oct4 Expression Directs a Robust Pluripotent State with Distinct Signaling Activity and Increased Enhancer Occupancy by Oct4 and Nanog'. *Cell Stem Cell* 12 (5): 531–45. https://doi.org/10.1016/j.stem.2013.04.023.

Kearns, Nicola A, Hannah Pham, Barbara Tabak, Ryan M Genga, Noah J Silverstein, Manuel Garber, and René Maehr. 2015. 'Functional Annotation of Native Enhancers with a Cas9 -Histone Demethylase Fusion'. *Nature Methods* 12 (5): 401–3. https://doi.org/10.1038/nmeth.3325.

Keramari, Maria, Janet Razavi, Karen A. Ingman, Christoph Patsch, Frank Edenhofer, Christopher M. Ward, and Susan J. Kimber. 2010. 'Sox2 Is Essential for Formation of Trophectoderm in the Preimplantation Embryo'. *PloS One* 5 (11): e13952. https://doi.org/10.1371/journal.pone.0013952.

Kidder, Benjamin L., Stephen Palmer, and Jason G. Knott. 2009. 'SWI/SNF-Brg1 Regulates Self-Renewal and Occupies Core Pluripotency-Related Genes in Embryonic Stem Cells'. *STEM CELLS* 27 (2): 317–28. https://doi.org/10.1634/stemcells.2008-0710.

Kielman, Menno F., Maaret Rindapää, Claudia Gaspar, Nicole van Poppel, Cor Breukel, Sandra van Leeuwen, Makoto Mark Taketo, Scott Roberts, Ron Smits, and Riccardo Fodde. 2002. 'Apc Modulates Embryonic Stem-Cell Differentiation by Controlling the Dosage of Beta-Catenin Signaling'. *Nature Genetics* 32 (4): 594–605. https://doi.org/10.1038/ng1045.

Kim, Myoung Ok, Sung-Hyun Kim, Yong-Yeon Cho, Janos Nadas, Chul-Ho Jeong, Ke Yao, Dong Joon Kim, et al. 2012. 'ERK1 and ERK2 Regulate Embryonic Stem Cell Self-Renewal through Phosphorylation of Klf4'. *Nature Structural & Molecular Biology* 19 (3): 283–90. https://doi.org/10.1038/nsmb.2217.

Kim, Sung-Hyun, Myoung Ok Kim, Yong-Yeon Cho, Ke Yao, Dong Joon Kim, Chul-Ho Jeong, Dong Hoon Yu, et al. 2014. 'ERK1 Phosphorylates Nanog to Regulate Protein Stability and Stem Cell Self-Renewal'. *Stem Cell Research* 13 (1): 1–11. https://doi.org/10.1016/j.scr.2014.04.001.

Kim, Y. G., J. Cha, and S. Chandrasegaran. 1996. 'Hybrid Restriction Enzymes: Zinc Finger Fusions to Fok I Cleavage Domain'. *Proceedings of the National Academy of Sciences of the United States of America* 93 (3): 1156–60.

King, Hamish W, and Robert J Klose. 2017. 'The Pioneer Factor OCT4 Requires the Chromatin Remodeller BRG1 to Support Gene Regulatory Element Function in Mouse Embryonic Stem Cells'. *ELife* 6 (March). https://doi.org/10.7554/eLife.22631.

Kingston, Robert E., and John W. Tamkun. 2014. 'Transcriptional Regulation by Trithorax-Group Proteins'. *Cold Spring Harbor Perspectives in Biology* 6 (10). https://doi.org/10.1101/cshperspect.a019349.

Kleinsmith, L. J., and G. B. Pierce. 1964. 'MULTIPOTENTIALITY OF SINGLE EMBRYONAL CARCINOMA CELLS'. *Cancer Research* 24 (October): 1544–51.

Kojima, Yoji, Keren Kaufman-Francis, Joshua B. Studdert, Kirsten A. Steiner, Melinda D. Power, David A. F. Loebel, Vanessa Jones, et al. 2014. 'The Transcriptional and Functional Properties of Mouse Epiblast Stem Cells Resemble the Anterior Primitive Streak'. *Cell Stem Cell* 14 (1): 107–20. https://doi.org/10.1016/j.stem.2013.09.014.

Konermann, Silvana, Mark D. Brigham, Alexandro E. Trevino, Julia Joung, Omar O. Abudayyeh, Clea Barcena, Patrick D. Hsu, et al. 2015. 'Genome-Scale Transcriptional Activation by an Engineered CRISPR-Cas9 Complex'. *Nature* 517 (7536): 583–88. https://doi.org/10.1038/nature14136.

Koopman, P., and R. G. Cotton. 1984. 'A Factor Produced by Feeder Cells Which Inhibits Embryonal Carcinoma Cell Differentiation. Characterization and Partial Purification'. *Experimental Cell Research* 154 (1): 233–42.

Kopp, Florian, and Joshua T. Mendell. 2018. 'Functional Classification and Experimental Dissection of Long Noncoding RNAs'. *Cell* 172 (3): 393–407. https://doi.org/10.1016/j.cell.2018.01.011.

Kopp, Janel L., Briana D. Ormsbee, Michelle Desler, and Angie Rizzino. 2008. 'Small Increases in the Level of Sox2 Trigger the Differentiation of Mouse Embryonic Stem Cells'. *Stem Cells (Dayton, Ohio)* 26 (4): 903–11. https://doi.org/10.1634/stemcells.2007-0951.

Krebs, Danielle L., and Douglas J. Hilton. 2009. 'SOCS Proteins: Negative Regulators of Cytokine Signaling'. *STEM CELLS* 19 (5): 378–87. https://doi.org/10.1634/stemcells.19-5-378.

Krendl, Christian, Dmitry Shaposhnikov, Valentyna Rishko, Chaido Ori, Christoph Ziegenhain, Steffen Sass, Lukas Simon, et al. 2017. 'GATA2/3-TFAP2A/C Transcription Factor Network Couples Human Pluripotent Stem Cell Differentiation to Trophectoderm with Repression of Pluripotency'. *Proceedings of the National Academy of Sciences*, October, 201708341. https://doi.org/10.1073/pnas.1708341114.

Ku, Manching, Richard P. Koche, Esther Rheinbay, Eric M. Mendenhall, Mitsuhiro Endoh, Tarjei S. Mikkelsen, Aviva Presser, et al. 2008. 'Genomewide Analysis of PRC1 and

PRC2 Occupancy Identifies Two Classes of Bivalent Domains'. *PLoS Genetics* 4 (10): e1000242. https://doi.org/10.1371/journal.pgen.1000242.

Kuhelj, Robert, Christopher J. Rizzo, Chong-Hwan Chang, Prabhakar K. Jadhav, Eric M. Towler, and Bruce D. Korant. 2001. 'Inhibition of Human Endogenous Retrovirus-K10 Protease in Cell-Free and Cell-Based Assays'. *Journal of Biological Chemistry* 276 (20): 16674–82. https://doi.org/10.1074/jbc.M008763200.

Kunath, Tilo, Marc K. Saba-El-Leil, Marwa Almousailleakh, Jason Wray, Sylvain Meloche, and Austin Smith. 2007. 'FGF Stimulation of the Erk1/2 Signalling Cascade Triggers Transition of Pluripotent Embryonic Stem Cells from Self-Renewal to Lineage Commitment'. *Development (Cambridge, England)* 134 (16): 2895–2902. https://doi.org/10.1242/dev.02880.

Kundu, Tapas K., Vikas B. Palhan, Zhengxin Wang, Woojin An, Philip A. Cole, and Robert G. Roeder. 2000. 'Activator-Dependent Transcription from Chromatin In Vitro Involving Targeted Histone Acetylation by P300'. *Molecular Cell* 6 (3): 551–61. https://doi.org/10.1016/S1097-2765(00)00054-X.

Kuroda, Takao, Masako Tada, Hiroshi Kubota, Hironobu Kimura, Shin-ya Hatano, Hirofumi Suemori, Norio Nakatsuji, and Takashi Tada. 2005. 'Octamer and Sox Elements Are Required for Transcriptional Cis Regulation of Nanog Gene Expression'. *Molecular and Cellular Biology* 25 (6): 2475–85. https://doi.org/10.1128/MCB.25.6.2475-2485.2005.

Kuznetsova, Inna S., Dmitrii I. Ostromyshenskii, Alexei S. Komissarov, Andrei N. Prusov, Irina S. Waisertreiger, Anna V. Gorbunova, Vladimir A. Trifonov, Malcolm A. Ferguson-Smith, and Olga I. Podgornaya. 2016. 'LINE-Related Component of Mouse Heterochromatin and Complex Chromocenters' Composition'. *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology* 24 (3): 309–23. https://doi.org/10.1007/s10577-016-9525-9.

Lan, Zi-Jian, Xueping Xu, and Austin J. Cooney. 2004. 'Differential Oocyte-Specific Expression of Cre Recombinase Activity in GDF-9-ICre, Zp3cre, and Msx2Cre Transgenic Mice'. *Biology of Reproduction* 71 (5): 1469–74. https://doi.org/10.1095/biolreprod.104.031757.

Landeira, David, and Amanda G. Fisher. 2011. 'Inactive yet Indispensable: The Tale of Jarid2'. *Trends in Cell Biology* 21 (2): 74–80. https://doi.org/10.1016/j.tcb.2010.10.004.

Lee, Jeannie T., and Rudolf Jaenisch. 1997. 'Long-Range Cis Effects of Ectopic X-Inactivation Centres on a Mouse Autosome'. *Nature* 386 (6622): 275–79. https://doi.org/10.1038/386275a0.

Leeb, Martin, Sabine Dietmann, Maike Paramor, Hitoshi Niwa, and Austin Smith. 2014. 'Genetic Exploration of the Exit from Self-Renewal Using Haploid Embryonic Stem Cells'. *Cell Stem Cell* 14 (3): 385–93. https://doi.org/10.1016/j.stem.2013.12.008.

Leeb, Martin, and Anton Wutz. 2007. 'Ring1B Is Crucial for the Regulation of Developmental Control Genes and PRC1 Proteins but Not X Inactivation in Embryonic Cells'. *The Journal of Cell Biology* 178 (2): 219–29. https://doi.org/10.1083/jcb.200612127.

Lehnertz, Bernhard, Yoshihide Ueda, Alwin A. H. A. Derijck, Ulrich Braunschweig, Laura Perez-Burgos, Stefan Kubicek, Taiping Chen, En Li, Thomas Jenuwein, and Antoine H. F. M. Peters. 2003. 'Suv39h-Mediated Histone H3 Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin'. *Current Biology* 13 (14): 1192–1200. https://doi.org/10.1016/S0960-9822(03)00432-9.

Leitch, Harry G., Jennifer Nichols, Peter Humphreys, Carla Mulas, Graziano Martello, Caroline Lee, Ken Jones, M. Azim Surani, and Austin Smith. 2013. 'Rebuilding Pluripotency from Primordial Germ Cells'. *Stem Cell Reports* 1 (1): 66–78. https://doi.org/10.1016/j.stemcr.2013.03.004.

Li, Gang, Raphael Margueron, Manching Ku, Pierre Chambon, Bradley E. Bernstein, and Danny Reinberg. 2010. 'Jarid2 and PRC2, Partners in Regulating Gene Expression'. *Genes & Development*, February. https://doi.org/10.1101/gad.1886410.

Li, Lingjie, and Howard Y. Chang. 2014. 'Physiological Roles of Long Noncoding RNAs: Insights from Knockout Mice'. *Trends in Cell Biology* 24 (10): 594–602. https://doi.org/10.1016/j.tcb.2014.06.003.

Li, Meng Amy, Daniel J. Turner, Zemin Ning, Kosuke Yusa, Qi Liang, Sabine Eckert, Lena Rad, Tomas W. Fitzgerald, Nancy L. Craig, and Allan Bradley. 2011. 'Mobilization of Giant PiggyBac Transposons in the Mouse Genome'. *Nucleic Acids Research* 39 (22): e148–e148. https://doi.org/10.1093/nar/gkr764.

Li, Meng, Michael Sendtner, and Austin Smith. 1995. 'Essential Function of LIF Receptor in Motor Neurons'. *Nature* 378 (6558): 724–27. https://doi.org/10.1038/378724a0.

Li, Rongbo, Yuan Zhuang, Min Han, Tian Xu, and Xiaohui Wu. 2013. 'PiggyBac as a High-Capacity Transgenesis and Gene-Therapy Vector in Human Cells and Mice'. *Disease Models & Mechanisms*, January, dmm.010827. https://doi.org/10.1242/dmm.010827.

Li, Xianghong, Erin R. Burnight, Ashley L. Cooney, Nirav Malani, Troy Brady, Jeffry D. Sander, Janice Staber, et al. 2013. 'PiggyBac Transposase Tools for Genome Engineering'. *Proceedings of the National Academy of Sciences of the United States of America* 110 (25): E2279–87. https://doi.org/10.1073/pnas.1305987110.

Lim, Youngbin, So Young Bak, Keewon Sung, Euihwan Jeong, Seung Hwan Lee, Jin-Soo Kim, Sangsu Bae, and Seong Keun Kim. 2016. 'Structural Roles of Guide RNAs in the

Nuclease Activity of Cas9 Endonuclease'. *Nature Communications* 7 (November): 13350. https://doi.org/10.1038/ncomms13350.

Lin, Michael F., Irwin Jungreis, and Manolis Kellis. 2011. 'PhyloCSF: A Comparative Genomics Method to Distinguish Protein Coding and Non-Coding Regions'. *Bioinformatics* 27 (13): i275–82. https://doi.org/10.1093/bioinformatics/btr209.

Lin, Nianwei, Kung-Yen Chang, Zhonghan Li, Keith Gates, Zacharia A. Rana, Jason Dang, Danhua Zhang, et al. 2014. 'An Evolutionarily Conserved Long Noncoding RNA TUNA Controls Pluripotency and Neural Lineage Commitment'. *Molecular Cell* 53 (6): 1005–19. https://doi.org/10.1016/j.molcel.2014.01.021.

Liu, Guo-You, Guang-Nian Zhao, Xiao-Feng Chen, De-Long Hao, Xiang Zhao, Xiang Lv, and De-Pei Liu. 2016. 'The Long Noncoding RNA Gm15055 Represses Hoxa Gene Expression by Recruiting PRC2 to the Gene Cluster'. *Nucleic Acids Research* 44 (6): 2613–27. https://doi.org/10.1093/nar/gkv1315.

Liu, Honglin, Jin-Moon Kim, and Fugaku Aoki. 2004. 'Regulation of Histone H3 Lysine 9 Methylation in Oocytes and Early Pre-Implantation Embryos'. *Development* 131 (10): 2269–80. https://doi.org/10.1242/dev.01116.

Liu, Peng, Meng Chen, Yanxia Liu, Lei S. Qi, and Sheng Ding. 2018. 'CRISPR-Based Chromatin Remodeling of the Endogenous Oct4 or Sox2 Locus Enables Reprogramming to Pluripotency'. *Cell Stem Cell* 22 (2): 252-261.e4. https://doi.org/10.1016/j.stem.2017.12.001.

Liu, S. John, Max A. Horlbeck, Seung Woo Cho, Harjus S. Birk, Martina Malatesta, Daniel He, Frank J. Attenello, et al. 2017. 'CRISPRi-Based Genome-Scale Identification of Functional Long Noncoding RNA Loci in Human Cells'. *Science* 355 (6320): eaah7111. https://doi.org/10.1126/science.aah7111.

Liu, X Shawn, Hao Wu, Xiong Ji, Yonatan Stelzer, Xuebing Wu, Szymon Czauderna, Jian Shu, Daniel Dadon, Richard A. Young, and Rudolf Jaenisch. 2016. 'Editing DNA Methylation in the Mammalian Genome'. *Cell* 167 (1): 233-247.e17. https://doi.org/10.1016/j.cell.2016.08.056.

Lloret-Llinares, Marta, Christophe K. Mapendano, Lasse H. Martlev, Søren Lykke-Andersen, and Torben Heick Jensen. 2015. 'Relationships between PROMPT and Gene Expression'. *RNA Biology* 13 (1): 6–14. https://doi.org/10.1080/15476286.2015.1109769.

Loh, Yuin-Han, Qiang Wu, Joon-Lin Chew, Vinsensius B. Vega, Weiwei Zhang, Xi Chen, Guillaume Bourque, et al. 2006. 'The Oct4 and Nanog Transcription Network Regulates Pluripotency in Mouse Embryonic Stem Cells'. *Nature Genetics* 38 (4): 431–40. https://doi.org/10.1038/ng1760.

Lopes, Fátima, Gabriela Soares, Miguel Gonçalves-Rocha, Jorge Pinto-Basto, and Patrícia Maciel. 2017. 'Whole Gene Deletion of EBF3 Supporting Haploinsufficiency of This Gene as a Mechanism of Neurodevelopmental Disease'. *Frontiers in Genetics* 8 (October). https://doi.org/10.3389/fgene.2017.00143.

Lu, Zhipeng, Qiangfeng Cliff Zhang, Byron Lee, Ryan A. Flynn, Martin A. Smith, James T. Robinson, Chen Davidovich, et al. 2016. 'RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure'. *Cell* 165 (5): 1267–79. https://doi.org/10.1016/j.cell.2016.04.028.

Luo, Sai, J. Yuyang Lu, Lichao Liu, Yafei Yin, Chunyan Chen, Xue Han, Bohou Wu, et al. 2016. 'Divergent LncRNAs Regulate Gene Expression and Lineage Differentiation in Pluripotent Cells'. *Cell Stem Cell* 18 (5): 637–52. https://doi.org/10.1016/j.stem.2016.01.024.

Lv, Jie, Hui Liu, Shihuan Yu, Hongbo Liu, Wei Cui, Yang Gao, Tao Zheng, et al. 2015. 'Identification of 4438 Novel LincRNAs Involved in Mouse Pre-Implantation Embryonic Development'. *Molecular Genetics and Genomics: MGG* 290 (2): 685–97. https://doi.org/10.1007/s00438-014-0952-z.

Lyashenko, Natalia, Markus Winter, Domenico Migliorini, Travis Biechele, Randall T. Moon, and Christine Hartmann. 2011. 'Differential Requirement for the Dual Functions of β-Catenin in Embryonic Stem Cell Self-Renewal and Germ Layer Formation'. *Nature Cell Biology* 13 (7): 753–61. https://doi.org/10.1038/ncb2260.

Ma, Xue-Shan, Shi-Bin Chao, Xian-Ju Huang, Fei Lin, Ling Qin, Xu-Guang Wang, Tie-Gang Meng, et al. 2015. 'The Dynamics and Regulatory Mechanism of Pronuclear H3k9me2 Asymmetry in Mouse Zygotes'. *Scientific Reports* 5 (December): 17924. https://doi.org/10.1038/srep17924.

Maamar, Hédia, Moran N. Cabili, John Rinn, and Arjun Raj. 2013. 'Linc-HOXA1 Is a Noncoding RNA That Represses Hoxa1 Transcription in Cis'. *Genes & Development* 27 (11): 1260–71. https://doi.org/10.1101/gad.217018.113.

MacArthur, Ben D., Ana Sevilla, Michel Lenz, Franz-Josef Müller, Berhard M. Schuldt, Andreas A. Schuppert, Sonya J. Ridden, et al. 2012. 'Nanog-Dependent Feedback Loops Regulate Murine Embryonic Stem Cell Heterogeneity'. *Nature Cell Biology* 14 (11): 1139–47. https://doi.org/10.1038/ncb2603.

Macfarlan, Todd S., Wesley D. Gifford, Shawn Driscoll, Karen Lettieri, Helen M. Rowe, Dario Bonanomi, Amy Firth, Oded Singer, Didier Trono, and Samuel L. Pfaff. 2012. 'ES Cell Potency Fluctuates with Endogenous Retrovirus Activity'. *Nature* 487 (7405): 57–63. https://doi.org/10.1038/nature11244.

Machanick, Philip, and Timothy L. Bailey. 2011. 'MEME-ChIP: Motif Analysis of Large DNA Datasets'. *Bioinformatics* 27 (12): 1696–97. https://doi.org/10.1093/bioinformatics/btr189.

Maeder, Morgan L, Samantha J Linder, Vincent M Cascio, Yanfang Fu, Quan H Ho, and J Keith Joung. 2013. 'CRISPR RNA-Guided Activation of Endogenous Human Genes'. *Nature Methods* 10 (10): 977–79. https://doi.org/10.1038/nmeth.2598.

Magistri, Marco, Mohammad Ali Faghihi, Georges St Laurent, and Claes Wahlestedt. 2012. 'Regulation of Chromatin Structure by Long Noncoding RNAs: Focus on Natural Antisense Transcripts'. *Trends in Genetics* 28 (8): 389–96. https://doi.org/10.1016/j.tig.2012.03.013.

Magnúsdóttir, Erna, Sabine Dietmann, Kazuhiro Murakami, Ufuk Günesdogan, Fuchou Tang, Siqin Bao, Evangelia Diamanti, Kaiqin Lao, Bertie Gottgens, and M. Azim Surani. 2013. 'A Tripartite Transcription Factor Network Regulates Primordial Germ Cell Specification in Mice'. *Nature Cell Biology* 15 (8): 905–15. https://doi.org/10.1038/ncb2798.

Malaguti, Mattias, Paul A. Nistor, Guillaume Blin, Amy Pegg, Xinzhi Zhou, and Sally Lowell. 2013. 'Bone Morphogenic Protein Signalling Suppresses Differentiation of Pluripotent Cells by Maintaining Expression of E-Cadherin'. *ELife* 2 (December): e01197. https://doi.org/10.7554/eLife.01197.

Mali, Prashant, John Aach, P. Benjamin Stranges, Kevin M. Esvelt, Mark Moosburner, Sriram Kosuri, Luhan Yang, and George M. Church. 2013. 'CAS9 Transcriptional Activators for Target Specificity Screening and Paired Nickases for Cooperative Genome Engineering'. *Nature Biotechnology* 31 (9): 833–38. https://doi.org/10.1038/nbt.2675.

Mali, Prashant, Luhan Yang, Kevin M. Esvelt, John Aach, Marc Guell, James E. DiCarlo, Julie E. Norville, and George M. Church. 2013. 'RNA-Guided Human Genome Engineering via Cas9'. *Science (New York, N.Y.)* 339 (6121): 823–26. https://doi.org/10.1126/science.1232033.

Margueron, Raphaël, and Danny Reinberg. 2011. 'The Polycomb Complex PRC2 and Its Mark in Life'. *Nature* 469 (7330): 343–49. https://doi.org/10.1038/nature09784.

Marks, Hendrik, Tüzer Kalkan, Roberta Menafra, Sergey Denissov, Kenneth Jones, Helmut Hofemeister, Jennifer Nichols, et al. 2012. 'The Transcriptional and Epigenomic Foundations of Ground State Pluripotency'. *Cell* 149 (3): 590–604. https://doi.org/10.1016/j.cell.2012.03.026.

Marraffini, Luciano A., and Erik J. Sontheimer. 2008. 'CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA'. *Science (New York, N.Y.)* 322 (5909): 1843–45. https://doi.org/10.1126/science.1165771.

Martello, Graziano, Paul Bertone, and Austin Smith. 2013. ‘Identification of the Missing Pluripotency Mediator Downstream of Leukaemia Inhibitory Factor’. *The EMBO Journal* 32 (19): 2561–74. https://doi.org/10.1038/emboj.2013.177.

Martello, Graziano, Toshimi Sugimoto, Evangelia Diamanti, Anagha Joshi, Rebecca Hannah, Satoshi Ohtsuka, Berthold Göttgens, Hitoshi Niwa, and Austin Smith. 2012. ‘Esrrb Is a Pivotal Target of the Gsk3/Tcf3 Axis Regulating Embryonic Stem Cell Self-Renewal’. *Cell Stem Cell* 11 (4): 491–504. https://doi.org/10.1016/j.stem.2012.06.008.

Martin, G. R. 1980. ‘Teratocarcinomas and Mammalian Embryogenesis’. *Science (New York, N.Y.)* 209 (4458): 768–76.

———. 1981. ‘Isolation of a Pluripotent Cell Line from Early Mouse Embryos Cultured in Medium Conditioned by Teratocarcinoma Stem Cells’. *Proceedings of the National Academy of Sciences of the United States of America* 78 (12): 7634–38.

Martin, G R, and M J Evans. 1975. ‘Differentiation of Clonal Lines of Teratocarcinoma Cells: Formation of Embryoid Bodies in Vitro.’ *Proceedings of the National Academy of Sciences of the United States of America* 72 (4): 1441–45.

Martinez Arias, Alfonso, and Joshua M Brickman. 2011. ‘Gene Expression Heterogeneities in Embryonic Stem Cell Populations: Origin and Function’. *Current Opinion in Cell Biology*, Cell differentiation, Cell division, growth and death, 23 (6): 650–56. https://doi.org/10.1016/j.ceb.2011.09.007.

Masui, Shinji, Yuhki Nakatake, Yayoi Toyooka, Daisuke Shimosato, Rika Yagi, Kazue Takahashi, Hitoshi Okochi, et al. 2007. ‘Pluripotency Governed by Sox2 via Regulation of Oct3/4 Expression in Mouse Embryonic Stem Cells’. *Nature Cell Biology* 9 (6): 625–35. https://doi.org/10.1038/ncb1589.

Matsuda, T., T. Nakamura, K. Nakao, T. Arai, M. Katsuki, T. Heike, and T. Yokota. 1999. ‘STAT3 Activation Is Sufficient to Maintain an Undifferentiated State of Mouse Embryonic Stem Cells’. *The EMBO Journal* 18 (15): 4261–69. https://doi.org/10.1093/emboj/18.15.4261.

Matsui, Y., K. Zsebo, and B. L. Hogan. 1992. ‘Derivation of Pluripotential Embryonic Stem Cells from Murine Primordial Germ Cells in Culture’. *Cell* 70 (5): 841–47.

Mattick, John S., and John L. Rinn. 2015. ‘Discovery and Annotation of Long Noncoding RNAs’. *Nature Structural & Molecular Biology* 22 (1): 5–7. https://doi.org/10.1038/nsmb.2942.

Mayer, Wolfgang, Avril Smith, Reinald Fundele, and Thomas Haaf. 2000. ‘Spatial Separation of Parental Genomes in Preimplantation Mouse Embryos’. *The Journal of Cell Biology* 148 (4): 629–34. https://doi.org/10.1083/jcb.148.4.629.

McHugh, Colleen A., and Mitchell Guttman. 2018. 'RAP-MS: A Method to Identify Proteins That Interact Directly with a Specific RNA Molecule in Cells'. In *RNA Detection*, 473–88. Methods in Molecular Biology. Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-7213-5_31.

Messerschmidt, Daniel M., and Rolf Kemler. 2010. 'Nanog Is Required for Primitive Endoderm Formation through a Non-Cell Autonomous Mechanism'. *Developmental Biology* 344 (1): 129–37. https://doi.org/10.1016/j.ydbio.2010.04.020.

Mikkelsen, Tarjei S., Manching Ku, David B. Jaffe, Biju Issac, Erez Lieberman, Georgia Giannoukos, Pablo Alvarez, et al. 2007. 'Genome-Wide Maps of Chromatin State in Pluripotent and Lineage-Committed Cells'. *Nature* 448 (7153): 553–60. https://doi.org/10.1038/nature06008.

Mitsui, Kaoru, Yoshimi Tokuzawa, Hiroaki Itoh, Kohichi Segawa, Mirei Murakami, Kazutoshi Takahashi, Masayoshi Maruyama, Mitsuyo Maeda, and Shinya Yamanaka. 2003. 'The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES Cells'. *Cell* 113 (5): 631–42. https://doi.org/10.1016/S0092-8674(03)00393-3.

Mojica, Francisco J. M., Cesar Díez-Villaseñor, Elena Soria, and Guadalupe Juez. 2000. 'Biological Significance of a Family of Regularly Spaced Repeats in the Genomes of Archaea, Bacteria and Mitochondria'. *Molecular Microbiology* 36 (1): 244–46. https://doi.org/10.1046/j.1365-2958.2000.01838.x.

Mojica, Francisco J. M., Chc)sar Díez-Villaseñor, Jesús García-Martínez, and Elena Soria. 2005. 'Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements'. *Journal of Molecular Evolution* 60 (2): 174–82. https://doi.org/10.1007/s00239-004-0046-3.

Morey, Lluís, and Kristian Helin. 2010. 'Polycomb Group Protein-Mediated Repression of Transcription'. *Trends in Biochemical Sciences* 35 (6): 323–32. https://doi.org/10.1016/j.tibs.2010.02.009.

Mueller, Florian, Adrien Senecal, Katjana Tantale, Hervé Marie-Nelly, Nathalie Ly, Olivier Collin, Eugenia Basyuk, Edouard Bertrand, Xavier Darzacq, and Christophe Zimmer. 2013. 'FISH-Quant: Automatic Counting of Transcripts in 3D FISH Images'. *Nature Methods* 10 (4): 277–78. https://doi.org/10.1038/nmeth.2406.

Murphy, Mark W, David Zarkower, and Vivian J Bardwell. 2007. 'Vertebrate DM Domain Proteins Bind Similar DNA Sequences and Can Heterodimerize on DNA'. *BMC Molecular Biology* 8 (July): 58. https://doi.org/10.1186/1471-2199-8-58.

Nakagawa, Shinichi. 2016. 'Lessons from Reverse-Genetic Studies of LncRNAs'. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, SI: Clues to long noncoding RNA taxonomy, 1859 (1): 177–83. https://doi.org/10.1016/j.bbagrm.2015.06.011.

Nakamura, Toshinobu, Yu-Jung Liu, Hiroyuki Nakashima, Hiroki Umehara, Kimiko Inoue, Shogo Matoba, Makoto Tachibana, Atsuo Ogura, Yoichi Shinkai, and Toru Nakano. 2012. 'PGC7 Binds Histone H3K9me2 to Protect against Conversion of 5mC to 5hmC in Early Embryos'. *Nature* 486 (7403): 415–19. https://doi.org/10.1038/nature11093.

Nakashima, Kinichi, Makoto Yanagisawa, Hirokazu Arakawa, and Tetsuya Taga. 1999. 'Astrocyte Differentiation Mediated by LIF in Cooperation with BMP2'. *FEBS Letters* 457 (1): 43–46. https://doi.org/10.1016/S0014-5793(99)00997-7.

Navarro, Pablo, Nicola Festuccia, Douglas Colby, Alessia Gagliardi, Nicholas P. Mullin, Wensheng Zhang, Violetta Karwacki-Neisius, et al. 2012. 'OCT4/SOX2-independent Nanog Autorepression Modulates Heterogeneous Nanog Gene Expression in Mouse ES Cells'. *The EMBO Journal* 31 (24): 4547–62. https://doi.org/10.1038/emboj.2012.321.

Nelson, Benjamin R., Catherine A. Makarewich, Douglas M. Anderson, Benjamin R. Winders, Constantine D. Troupes, Fenfen Wu, Austin L. Reese, et al. 2016. 'A Peptide Encoded by a Transcript Annotated as Long Noncoding RNA Enhances SERCA Activity in Muscle'. *Science (New York, N.Y.)* 351 (6270): 271–75. https://doi.org/10.1126/science.aad4076.

Nett, Isabelle RE, Carla Mulas, Laurent Gatto, Kathryn S. Lilley, and Austin Smith. 2018. 'Negative Feedback via RSK Modulates Erk-dependent Progression from Naïve Pluripotency'. *EMBO Reports*, June, e45642. https://doi.org/10.15252/embr.201745642.

Ng, Shi-Yan, Rory Johnson, and Lawrence W. Stanton. 2012. 'Human Long Non-Coding RNAs Promote Pluripotency and Neuronal Differentiation by Association with Chromatin Modifiers and Transcription Factors'. *The EMBO Journal* 31 (3): 522–33. https://doi.org/10.1038/emboj.2011.459.

Niakan, Kathy K., Hongkai Ji, René Maehr, Steven A. Vokes, Kit T. Rodolfa, Richard I. Sherwood, Mariko Yamaki, et al. 2010. 'Sox17 Promotes Differentiation in Mouse Embryonic Stem Cells by Directly Regulating Extraembryonic Gene Expression and Indirectly Antagonizing Self-Renewal'. *Genes & Development* 24 (3): 312–26. https://doi.org/10.1101/gad.1833510.

Nichols, J., I. Chambers, T. Taga, and A. Smith. 2001. 'Physiological Rationale for Responsiveness of Mouse Embryonic Stem Cells to Gp130 Cytokines'. *Development (Cambridge, England)* 128 (12): 2333–39.

Nichols, Jennifer, Ian Chambers, and Austin Smith. 1994. 'Derivation of Gerlime Competent Embryonic Stem Cells with a Combination of Interleukin-6 and Soluble Interleukin-6 Receptor'. Experimental Cell Research.

Nichols, Jennifer, Branko Zevnik, Konstantinos Anastassiadis, Hitoshi Niwa, Daniela Klewe-Nebenius, Ian Chambers, Hans Schöler, and Austin Smith. 1998. 'Formation of Pluripotent Stem Cells in the Mammalian Embryo Depends on the POU Transcription Factor Oct4'. *Cell* 95 (3): 379–91. https://doi.org/10.1016/S0092-8674(00)81769-9.

Nicholson, Sandra E., David De Souza, Louis J. Fabri, Jason Corbin, Tracy A. Willson, Jian-Guo Zhang, Anabel Silva, et al. 2000. 'Suppressor of Cytokine Signaling-3 Preferentially Binds to the SHP-2-Binding Site on the Shared Cytokine Receptor Subunit Gp130'. *Proceedings of the National Academy of Sciences* 97 (12): 6493–98. https://doi.org/10.1073/pnas.100135197.

Nishimasu, Hiroshi, F. Ann Ran, Patrick D. Hsu, Silvana Konermann, Soraya I. Shehata, Naoshi Dohmae, Ryuichiro Ishitani, Feng Zhang, and Osamu Nureki. 2014. 'Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA'. *Cell* 156 (5): 935–49. https://doi.org/10.1016/j.cell.2014.02.001.

Nishioka, Noriyuki, Ken-ichi Inoue, Kenjiro Adachi, Hiroshi Kiyonari, Mitsunori Ota, Amy Ralston, Norikazu Yabuta, et al. 2009. 'The Hippo Signaling Pathway Components Lats and Yap Pattern Tead4 Activity to Distinguish Mouse Trophectoderm from Inner Cell Mass'. *Developmental Cell* 16 (3): 398–410. https://doi.org/10.1016/j.devcel.2009.02.003.

Nishiyama, Akira, Li Xin, Alexei A. Sharov, Marshall Thomas, Gregory Mowrer, Emily Meyers, Yulan Piao, et al. 2009. 'Uncovering Early Response of Gene Regulatory Networks in ESCs by Systematic Induction of Transcription Factors'. *Cell Stem Cell* 5 (4): 420–33. https://doi.org/10.1016/j.stem.2009.07.012.

Niwa, H., T. Burdon, I. Chambers, and A. Smith. 1998. 'Self-Renewal of Pluripotent Embryonic Stem Cells Is Mediated via Activation of STAT3'. *Genes & Development* 12 (13): 2048–60.

Niwa, Hitoshi, Jun-ichi Miyazaki, and Austin G. Smith. 2000. 'Quantitative Expression of Oct-3/4 Defines Differentiation, Dedifferentiation or Self-Renewal of ES Cells'. *Nature Genetics* 24 (4): 372–76. https://doi.org/10.1038/74199.

Niwa, Hitoshi, Kazuya Ogawa, Daisuke Shimosato, and Kenjiro Adachi. 2009. 'A Parallel Circuit of LIF Signalling Pathways Maintains Pluripotency of Mouse ES Cells'. *Nature* 460 (7251): 118–22. https://doi.org/10.1038/nature08113.

Niwa, Hitoshi, Yayoi Toyooka, Daisuke Shimosato, Dan Strumpf, Kadue Takahashi, Rika Yagi, and Janet Rossant. 2005. 'Interaction between Oct3/4 and Cdx2 Determines Trophectoderm Differentiation'. *Cell* 123 (5): 917–29. https://doi.org/10.1016/j.cell.2005.08.040.

Nora, Elphège P., Bryan R. Lajoie, Edda G. Schulz, Luca Giorgetti, Ikuhiro Okamoto, Nicolas Servant, Tristan Piolot, et al. 2012. 'Spatial Partitioning of the Regulatory Landscape

of the X-Inactivation Center'. *Nature* 485 (7398): 381–85. https://doi.org/10.1038/nature11049.

Novershtern, Noa, and Jacob H. Hanna. 2011. 'EsBAF Safeguards Stat3 Binding to Maintain Pluripotency'. *Nature Cell Biology* 13 (8): 886–88. https://doi.org/10.1038/ncb2311.

O'Carroll, Dónal, Sylvia Erhardt, Michaela Pagani, Sheila C. Barton, M. Azim Surani, and Thomas Jenuwein. 2001. 'The Polycomb-Group GeneEzh2 Is Required for Early Mouse Development'. *Molecular and Cellular Biology* 21 (13): 4330–36. https://doi.org/10.1128/MCB.21.13.4330-4336.2001.

Ohnishi, Yusuke, Wolfgang Huber, Akiko Tsumura, Minjung Kang, Panagiotis Xenopoulos, Kazuki Kurimoto, Andrzej K. Oleś, et al. 2014. 'Cell-to-Cell Expression Variability Followed by Signal Reinforcement Progressively Segregates Early Mouse Lineages'. *Nature Cell Biology* 16 (1): 27–37. https://doi.org/10.1038/ncb2881.

Ohtsuka, Satoshi, Satomi Nishikawa-Torikai, and Hitoshi Niwa. 2012. 'E-Cadherin Promotes Incorporation of Mouse Epiblast Stem Cells into Normal Development'. *PLoS ONE* 7 (9). https://doi.org/10.1371/journal.pone.0045220.

Okamoto, Koji, Hitoshi Okazawa, Akihiko Okuda, Masaharu Sakai, Masami Muramatsu, and Hiroshi Hamada. 1990. 'A Novel Octamer Binding Transcription Factor Is Differentially Expressed in Mouse Embryonic Cells'. *Cell* 60 (3): 461–72. https://doi.org/10.1016/0092-8674(90)90597-8.

Okashita, Naoki, Yoshiaki Suwa, Osamu Nishimura, Nao Sakashita, Mitsutaka Kadota, Go Nagamatsu, Masanori Kawaguchi, et al. 2016. 'PRDM14 Drives OCT3/4 Recruitment via Active Demethylation in the Transition from Primed to Naive Pluripotency'. *Stem Cell Reports* 7 (6): 1072–86. https://doi.org/10.1016/j.stemcr.2016.10.007.

Oksuz, Ozgur, Varun Narendra, Chul-Hwan Lee, Nicolas Descostes, Gary LeRoy, Ramya Raviram, Lili Blumenberg, et al. 2018. 'Capturing the Onset of PRC2-Mediated Repressive Domain Formation'. *Molecular Cell* 70 (6): 1149-1162.e5. https://doi.org/10.1016/j.molcel.2018.05.023.

Osorno, Rodrigo, Anestis Tsakiridis, Frederick Wong, Noemí Cambray, Constantinos Economou, Ronald Wilkie, Guillaume Blin, Paul J. Scotting, Ian Chambers, and Valerie Wilson. 2012. 'The Developmental Dismantling of Pluripotency Is Reversed by Ectopic Oct4 Expression'. *Development (Cambridge, England)* 139 (13): 2288–98. https://doi.org/10.1242/dev.078071.

Osumi, Noriko, Hiroshi Shinohara, Keiko Numayama-Tsuruta, and Motoko Maekawa. 2008. 'Concise Review: Pax6 Transcription Factor Contributes to Both Embryonic and Adult Neurogenesis as a Multifunctional Regulator'. *STEM CELLS* 26 (7): 1663–72. https://doi.org/10.1634/stemcells.2007-0884.

Paddison, Patrick J., Jose M. Silva, Douglas S. Conklin, Mike Schlabach, Mamie Li, Shola Aruleba, Vivekanand Balija, et al. 2004. 'A Resource for Large-Scale RNA-Interference-Based Screens in Mammals'. *Nature* 428 (6981): 427–31. https://doi.org/10.1038/nature02370.

Palmieri, Susan L., Werner Peter, Heike Hess, and Hans R. Schöler. 1994. 'Oct-4 Transcription Factor Is Differentially Expressed in the Mouse Embryo during Establishment of the First Two Extraembryonic Cell Lineages Involved in Implantation'. *Developmental Biology* 166 (1): 259–67. https://doi.org/10.1006/dbio.1994.1312.

Parisi, Silvia, Luca Cozzuto, Carolina Tarantino, Fabiana Passaro, Simona Ciriello, Luigi Aloia, Dario Antonini, Vincenzo De Simone, Lucio Pastore, and Tommaso Russo. 2010. 'Direct Targets of Klf5 Transcription Factor Contribute to the Maintenance of Mouse Embryonic Stem Cell Undifferentiated State'. *BMC Biology* 8 (September): 128. https://doi.org/10.1186/1741-7007-8-128.

Parisi, Silvia, Fabiana Passaro, Luigi Aloia, Ichiro Manabe, Ryozo Nagai, Lucio Pastore, and Tommaso Russo. 2008. 'Klf5 Is Involved in Self-Renewal of Mouse Embryonic Stem Cells'. *Journal of Cell Science* 121 (16): 2629–34. https://doi.org/10.1242/jcs.027599.

Pasini, Diego, Adrian P. Bracken, Jacob B. Hansen, Manuela Capillo, and Kristian Helin. 2007. 'The Polycomb Group Protein Suz12 Is Required for Embryonic Stem Cell Differentiation'. *Molecular and Cellular Biology* 27 (10): 3769–79. https://doi.org/10.1128/MCB.01432-06.

Pauli, Andrea, Eivind Valen, Michael F. Lin, Manuel Garber, Nadine L. Vastenhouw, Joshua Z. Levin, Lin Fan, et al. 2012. 'Systematic Identification of Long Noncoding RNAs Expressed during Zebrafish Embryogenesis'. *Genome Research* 22 (3): 577–91. https://doi.org/10.1101/gr.133009.111.

Pavletich, N. P., and C. O. Pabo. 1991. 'Zinc Finger-DNA Recognition: Crystal Structure of a Zif268-DNA Complex at 2.1 A'. *Science (New York, N.Y.)* 252 (5007): 809–17.

Peabody, D. S. 1993. 'The RNA Binding Site of Bacteriophage MS2 Coat Protein'. *The EMBO Journal* 12 (2): 595–600.

Pegueroles, Cinta, and Toni Gabaldón. 2016. 'Secondary Structure Impacts Patterns of Selection in Human LncRNAs'. *BMC Biology* 14 (July). https://doi.org/10.1186/s12915-016-0283-0.

Percharde, Michelle, Fabrice Lavial, Jia-Hui Ng, Vibhor Kumar, Rute A. Tomaz, Nadine Martin, Jia-Chi Yeo, et al. 2012. 'Ncoa3 Functions as an Essential Esrrb Coactivator to Sustain Embryonic Stem Cell Self-Renewal and Reprogramming'. *Genes & Development* 26 (20): 2286–98. https://doi.org/10.1101/gad.195545.112.

Percharde, Michelle, Chih-Jen Lin, Yafei Yin, Juan Guan, Gabriel A. Peixoto, Aydan Bulut-Karslioglu, Steffen Biechele, Bo Huang, Xiaohua Shen, and Miguel Ramalho-Santos. 2018. 'A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity'. *Cell*, June. https://doi.org/10.1016/j.cell.2018.05.043.

Pereira, Laura, Fei Yi, and Bradley J. Merrill. 2006. 'Repression of Nanog Gene Transcription by Tcf3 Limits Embryonic Stem Cell Self-Renewal'. *Molecular and Cellular Biology* 26 (20): 7479–91. https://doi.org/10.1128/MCB.00368-06.

Perez-Pinera, Pablo, D. Dewran Kocak, Christopher M. Vockley, Andrew F. Adler, Ami M. Kabadi, Lauren R. Polstein, Pratiksha I. Thakore, et al. 2013. 'RNA-Guided Gene Activation by CRISPR-Cas9-Based Transcription Factors'. *Nature Methods* 10 (10): 973–76. https://doi.org/10.1038/nmeth.2600.

Perez-Pinera, Pablo, David G. Ousterout, Jonathan M. Brunger, Alicia M. Farin, Katherine A. Glass, Farshid Guilak, Gregory E. Crawford, Alexander J. Hartemink, and Charles A. Gersbach. 2013. 'Synergistic and Tunable Human Gene Activation by Combinations of Synthetic Transcription Factors'. *Nature Methods* 10 (3): 239–42. https://doi.org/10.1038/nmeth.2361.

Peters, Antoine H. F. M., Stefan Kubicek, Karl Mechtler, Roderick J. O'Sullivan, Alwin A. H. A. Derijck, Laura Perez-Burgos, Alexander Kohlmaier, et al. 2003. 'Partitioning and Plasticity of Repressive Histone Methylation States in Mammalian Chromatin'. *Molecular Cell* 12 (6): 1577–89. https://doi.org/10.1016/S1097-2765(03)00477-5.

Peters, Antoine H. F. M., Dónal O'Carroll, Harry Scherthan, Karl Mechtler, Stephan Sauer, Christian Schöfer, Klara Weipoltshammer, et al. 2001. 'Loss of the Suv39h Histone Methyltransferases Impairs Mammalian Heterochromatin and Genome Stability'. *Cell* 107 (3): 323–37. https://doi.org/10.1016/S0092-8674(01)00542-6.

Piliszek, Anna, Joanna B. Grabarek, Stephen R. Frankenberg, and Berenika Plusa. 2016. 'Cell Fate in Animal and Human Blastocysts and the Determination of Viability'. *Molecular Human Reproduction* 22 (10): 681–90. https://doi.org/10.1093/molehr/gaw002.

Plusa, Berenika, Anna Piliszek, Stephen Frankenberg, Jérôme Artus, and Anna-Katerina Hadjantonakis. 2008. 'Distinct Sequential Cell Behaviours Direct Primitive Endoderm Formation in the Mouse Blastocyst'. *Development (Cambridge, England)* 135 (18): 3081–91. https://doi.org/10.1242/dev.021519.

Polstein, Lauren R., and Charles A. Gersbach. 2015. 'A Light-Inducible CRISPR/Cas9 System for Control of Endogenous Gene Activation'. *Nature Chemical Biology* 11 (3): 198–200. https://doi.org/10.1038/nchembio.1753.

Polstein, Lauren R., Pablo Perez-Pinera, D. Dewran Kocak, Christopher M. Vockley, Peggy Bledsoe, Lingyun Song, Alexias Safi, Gregory E. Crawford, Timothy E. Reddy, and Charles A. Gersbach. 2015. 'Genome-Wide Specificity of DNA Binding, Gene

Regulation, and Chromatin Remodeling by TALE- and CRISPR/Cas9-Based Transcriptional Activators'. *Genome Research* 25 (8): 1158–69. https://doi.org/10.1101/gr.179044.114.

Pourcel, C., G. Salvignol, and G. Vergnaud. 2005. 'CRISPR Elements in Yersinia Pestis Acquire New Repeats by Preferential Uptake of Bacteriophage DNA, and Provide Additional Tools for Evolutionary Studies'. *Microbiology* 151 (3): 653–63. https://doi.org/10.1099/mic.0.27437-0.

Puschendorf, Mareike, Rémi Terranova, Erwin Boutsma, Xiaohong Mao, Kyo-ichi Isono, Urszula Brykczynska, Carolin Kolb, et al. 2008. 'PRC1 and Suv39h Specify Parental Asymmetry at Constitutive Heterochromatin in Early Mouse Embryos'. *Nature Genetics* 40 (4): 411–20. https://doi.org/10.1038/ng.99.

Qi, Lei S., Matthew H. Larson, Luke A. Gilbert, Jennifer A. Doudna, Jonathan S. Weissman, Adam P. Arkin, and Wendell A. Lim. 2013. 'Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression'. *Cell* 152 (5): 1173–83. https://doi.org/10.1016/j.cell.2013.02.022.

Ramalho-Santos, Miguel, Soonsang Yoon, Yumi Matsuzaki, Richard C. Mulligan, and Douglas A. Melton. 2002. '"Stemness": Transcriptional Profiling of Embryonic and Adult Stem Cells'. *Science* 298 (5593): 597–600. https://doi.org/10.1126/science.1072530.

Ran, F. Ann, Patrick D. Hsu, Chie-Yu Lin, Jonathan S. Gootenberg, Silvana Konermann, Alexandro E. Trevino, David A. Scott, et al. 2013. 'Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity'. *Cell* 155 (2): 479–80. https://doi.org/10.1016/j.cell.2013.09.040.

Ransohoff, Julia D., Yuning Wei, and Paul A. Khavari. 2018. 'The Functions and Unique Features of Long Intergenic Non-Coding RNA'. *Nature Reviews. Molecular Cell Biology* 19 (3): 143–57. https://doi.org/10.1038/nrm.2017.104.

Rathjen, P D, J Nichols, S Toth, D R Edwards, J K Heath, and A G Smith. 1990. 'Developmentally Programmed Induction of Differentiation Inhibiting Activity and the Control of Stem Cell Populations.' *Genes & Development* 4 (12b): 2308–18. https://doi.org/10.1101/gad.4.12b.2308.

Resnick, J. L., L. S. Bixler, L. Cheng, and P. J. Donovan. 1992. 'Long-Term Proliferation of Mouse Primordial Germ Cells in Culture'. *Nature* 359 (6395): 550–51. https://doi.org/10.1038/359550a0.

Ribet, David, Sophie Louvet-Vallée, Francis Harper, Nathalie de Parseval, Marie Dewannieux, Odile Heidmann, Gérard Pierron, Bernard Maro, and Thierry Heidmann. 2008. 'Murine Endogenous Retrovirus MuERV-L Is the Progenitor of the "Orphan" Epsilon Viruslike

Particles of the Early Mouse Embryo'. *Journal of Virology* 82 (3): 1622–25. https://doi.org/10.1128/JVI.02097-07.

Riising, Eva Madi, Itys Comet, Benjamin Leblanc, Xudong Wu, Jens Vilstrup Johansen, and Kristian Helin. 2014. 'Gene Silencing Triggers Polycomb Repressive Complex 2 Recruitment to CpG Islands Genome Wide'. *Molecular Cell* 55 (3): 347–60. https://doi.org/10.1016/j.molcel.2014.06.005.

Rodda, David J., Joon-Lin Chew, Leng-Hiong Lim, Yuin-Han Loh, Bei Wang, Huck-Hui Ng, and Paul Robson. 2005. 'Transcriptional Regulation of Nanog by OCT4 and SOX2'. *The Journal of Biological Chemistry* 280 (26): 24731–37. https://doi.org/10.1074/jbc.M502573200.

Rodig, Scott J, Marco A Meraz, J. Michael White, Pat A Lampe, Joan K Riley, Cora D Arthur, Kathleen L King, et al. 1998. 'Disruption of the Jak1 Gene Demonstrates Obligatory and Nonredundant Roles of the Jaks in Cytokine-Induced Biologic Responses'. *Cell* 93 (3): 373–83. https://doi.org/10.1016/S0092-8674(00)81166-6.

Rodriguez-Terrones, Diego, Xavier Gaume, Takashi Ishiuchi, Amélie Weiss, Arnaud Kopp, Kai Kruse, Audrey Penning, Juan M. Vaquerizas, Laurent Brino, and Maria-Elena Torres-Padilla. 2018. 'A Molecular Roadmap for the Emergence of Early-Embryonic-like Cells in Culture'. *Nature Genetics* 50 (1): 106–19. https://doi.org/10.1038/s41588-017-0016-5.

Roost, Matthias S., Roderick C. Slieker, Monika Bialecka, Liesbeth van Iperen, Maria M. Gomes Fernandes, Nannan He, H. Eka D. Suchiman, et al. 2017. 'DNA Methylation and Transcriptional Trajectories during Human Development and Reprogramming of Isogenic Pluripotent Stem Cells'. *Nature Communications* 8 (October). https://doi.org/10.1038/s41467-017-01077-3.

Rose, Timothy M., Donna M. Weiford, Nancy L. Gunderson, and A.Gregory Bruce. 1994. 'Oncostatin M (OSM) Inhibits the Differentiation of Pluripotent Embryonic Stem Cells in Vitro'. *Cytokine* 6 (1): 48–54. https://doi.org/10.1016/1043-4666(94)90007-8.

Rossant, Janet, and Patrick P. L. Tam. 2009. 'Blastocyst Lineage Formation, Early Embryonic Asymmetries and Axis Patterning in the Mouse'. *Development (Cambridge, England)* 136 (5): 701–13. https://doi.org/10.1242/dev.017178.

Rostovskaya, Maria, Jun Fu, Mandy Obst, Isabell Baer, Stefanie Weidlich, Hailong Wang, Andrew J. H. Smith, Konstantinos Anastassiadis, and A. Francis Stewart. 2012. 'Transposon-Mediated BAC Transgenesis in Human ES Cells'. *Nucleic Acids Research* 40 (19): e150. https://doi.org/10.1093/nar/gks643.

Rouet, P., F. Smih, and M. Jasin. 1994. 'Expression of a Site-Specific Endonuclease Stimulates Homologous Recombination in Mammalian Cells'. *Proceedings of the National Academy of Sciences of the United States of America* 91 (13): 6064–68.

Rudin, N., and J. E. Haber. 1988. 'Efficient Repair of HO-Induced Chromosomal Breaks in Saccharomyces Cerevisiae by Recombination between Flanking Homologous Sequences'. *Molecular and Cellular Biology* 8 (9): 3918–28.

Sachs, Michael, Courtney Onodera, Kathryn Blaschke, Kevin T. Ebata, Jun S. Song, and Miguel Ramalho-Santos. 2013. 'Bivalent Chromatin Marks Developmental Regulatory Genes in the Mouse Embryonic Germline in Vivo'. *Cell Reports* 3 (6): 1777–84. https://doi.org/10.1016/j.celrep.2013.04.032.

Santos, Fátima, Antoine H. Peters, Arie P. Otte, Wolf Reik, and Wendy Dean. 2005. 'Dynamic Chromatin Modifications Characterise the First Cell Cycle in Mouse Embryos'. *Developmental Biology* 280 (1): 225–36. https://doi.org/10.1016/j.ydbio.2005.01.025.

Sapranauskas, Rimantas, Giedrius Gasiunas, Christophe Fremaux, Rodolphe Barrangou, Philippe Horvath, and Virginijus Siksnys. 2011. 'The Streptococcus Thermophilus CRISPR/Cas System Provides Immunity in Escherichia Coli'. *Nucleic Acids Research* 39 (21): 9275–82. https://doi.org/10.1093/nar/gkr606.

Schemmer, Jana, Marcos J. Araúzo-Bravo, Natalie Haas, Sabine Schäfer, Susanne N. Weber, Astrid Becker, Dawid Eckert, Andreas Zimmer, Daniel Nettersheim, and Hubert Schorle. 2013. 'Transcription Factor TFAP2C Regulates Major Programs Required for Murine Fetal Germ Cell Maintenance and Haploinsufficiency Predisposes to Teratomas in Male Mice'. *PLoS ONE* 8 (8). https://doi.org/10.1371/journal.pone.0071113.

Schmitz, Jochen, Manuela Weissenbach, Serge Haan, Peter C. Heinrich, and Fred Schaper. 2000. 'SOCS3 Exerts Its Inhibitory Function on Interleukin-6 Signal Transduction through the SHP2 Recruitment Site of Gp130'. *Journal of Biological Chemistry* 275 (17): 12848–56. https://doi.org/10.1074/jbc.275.17.12848.

Schöler, H. R., A. K. Hatzopoulos, R. Balling, N. Suzuki, and P. Gruss. 1989. 'A Family of Octamer-Specific Proteins Present during Mouse Embryogenesis: Evidence for Germline-Specific Expression of an Oct Factor'. *The EMBO Journal* 8 (9): 2543–50.

Schöler, Hans R., Siegfried Ruppert, Noriaki Suzuki, Kamal Chowdhury, and Peter Gruss. 1990. 'New Type of POU Domain in Germ Line-Specific Protein Oct-4'. *Nature* 344 (6265): 435–39. https://doi.org/10.1038/344435a0.

Schuettengruber, Bernd, Henri-Marc Bourbon, Luciano Di Croce, and Giacomo Cavalli. 2017. 'Genome Regulation by Polycomb and Trithorax: 70 Years and Counting'. *Cell* 171 (1): 34–57. https://doi.org/10.1016/j.cell.2017.08.002.

Schwarz, Benjamin A., Ori Bar-Nur, José C. R. Silva, and Konrad Hochedlinger. 2014. 'Nanog Is Dispensable for the Generation of Induced Pluripotent Stem Cells'. *Current Biology* 24 (3): 347–50. https://doi.org/10.1016/j.cub.2013.12.050.

Scognamiglio, Roberta, Nina Cabezas-Wallscheid, Marc Christian Thier, Sandro Altamura, Alejandro Reyes, Áine M. Prendergast, Daniel Baumgärtner, et al. 2016. 'Myc Depletion Induces a Pluripotent Dormant State Mimicking Diapause'. *Cell* 164 (4): 668–80. https://doi.org/10.1016/j.cell.2015.12.033.

Seike, Masanari, Yoshiki Omatsu, Hitomi Watanabe, Gen Kondoh, and Takashi Nagasawa. 2018. 'Stem Cell Niche-Specific Ebf3 Maintains the Bone Marrow Cavity'. *Genes & Development* 32 (5–6): 359–72. https://doi.org/10.1101/gad.311068.117.

Seo, Kwang Won, Yingdi Wang, Hiroki Kokubo, Jae R. Kettlewell, David A. Zarkower, and Randy L. Johnson. 2006. 'Targeted Disruption of the DM Domain Containing Transcription Factor Dmrt2 Reveals an Essential Role in Somite Patterning'. *Developmental Biology* 290 (1): 200–210. https://doi.org/10.1016/j.ydbio.2005.11.027.

Sexton, Tom, Eitan Yaffe, Ephraim Kenigsberg, Frédéric Bantignies, Benjamin Leblanc, Michael Hoichman, Hugues Parrinello, Amos Tanay, and Giacomo Cavalli. 2012. 'Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome'. *Cell* 148 (3): 458–72. https://doi.org/10.1016/j.cell.2012.01.010.

Shen, Xiaohua, Yingchun Liu, Yu-Jung Hsu, Yuko Fujiwara, Jonghwan Kim, Xiaohong Mao, Guo-Cheng Yuan, and Stuart H. Orkin. 2008. 'EZH1 Mediates Methylation on Histone H3 Lysine 27 and Complements EZH2 in Maintaining Stem Cell Identity and Executing Pluripotency'. *Molecular Cell* 32 (4): 491–502. https://doi.org/10.1016/j.molcel.2008.10.016.

Shi, Yanhong, Haruhisa Inoue, Joseph C. Wu, and Shinya Yamanaka. 2017. 'Induced Pluripotent Stem Cell Technology: A Decade of Progress'. *Nature Reviews Drug Discovery* 16 (2): 115–30. https://doi.org/10.1038/nrd.2016.245.

Shy, Brian R., Chun-I. Wu, Galina F. Khramtsova, Jenny Y. Zhang, Olufunmilayo I. Olopade, Kathleen H. Goss, and Bradley J. Merrill. 2013. 'Regulation of Tcf7l1 DNA Binding and Protein Stability as Principal Mechanisms of Wnt/β-Catenin Signaling'. *Cell Reports* 4 (1): 1–9. https://doi.org/10.1016/j.celrep.2013.06.001.

Sigova, Alla A., Alan C. Mullen, Benoit Molinie, Sumeet Gupta, David A. Orlando, Matthew G. Guenther, Albert E. Almada, et al. 2013. 'Divergent Transcription of Long Noncoding RNA/MRNA Gene Pairs in Embryonic Stem Cells'. *Proceedings of the National Academy of Sciences of the United States of America* 110 (8): 2876–81. https://doi.org/10.1073/pnas.1221904110.

Silva, José, Ian Chambers, Steven Pollard, and Austin Smith. 2006. 'Nanog Promotes Transfer of Pluripotency after Cell Fusion'. *Nature* 441 (7096): 997–1001. https://doi.org/10.1038/nature04914.

Silva, Jose M., Mamie Z. Li, Ken Chang, Wei Ge, Michael C. Golding, Richard J. Rickles, Despina Siolas, et al. 2005. 'Second-Generation ShRNA Libraries Covering the Mouse

and Human Genomes'. *Nature Genetics* 37 (11): 1281–88. https://doi.org/10.1038/ng1650.

Silva, Jose, Jennifer Nichols, Thorold W. Theunissen, Ge Guo, Anouk L. van Oosten, Ornella Barrandon, Jason Wray, Shinya Yamanaka, Ian Chambers, and Austin Smith. 2009. 'Nanog Is the Gateway to the Pluripotent Ground State'. *Cell* 138 (4): 722–37. https://doi.org/10.1016/j.cell.2009.07.039.

Singhal, Nishant, Daniel Esch, Martin Stehling, and Hans R. Schöler. 2014. 'BRG1 Is Required to Maintain Pluripotency of Murine Embryonic Stem Cells'. *BioResearch Open Access* 3 (1): 1–8. https://doi.org/10.1089/biores.2013.0047.

Sleven, Hannah, Seth J. Welsh, Jing Yu, Mair E. A. Churchill, Caroline F. Wright, Alex Henderson, Rita Horvath, et al. 2017. 'De Novo Mutations in EBF3 Cause a Neurodevelopmental Syndrome'. *The American Journal of Human Genetics* 100 (1): 138–50. https://doi.org/10.1016/j.ajhg.2016.11.020.

Smallwood, Sébastien A., and Gavin Kelsey. 2012. 'De Novo DNA Methylation: A Germ Cell Perspective'. *Trends in Genetics* 28 (1): 33–42. https://doi.org/10.1016/j.tig.2011.09.004.

Smith, A. G., and M. L. Hooper. 1987. 'Buffalo Rat Liver Cells Produce a Diffusible Activity Which Inhibits the Differentiation of Murine Embryonal Carcinoma and Embryonic Stem Cells'. *Developmental Biology* 121 (1): 1–9.

Smith, Austin G., John K. Heath, Deborah D. Donaldson, Gordon G. Wong, J. Moreau, Mark Stahl, and David Rogers. 1988. 'Inhibition of Pluripotential Embryonic Stem Cell Differentiation by Purified Polypeptides'. *Nature* 336 (6200): 688–90. https://doi.org/10.1038/336688a0.

Smith, Zachary D., and Alexander Meissner. 2013. 'DNA Methylation: Roles in Mammalian Development'. *Nature Reviews. Genetics* 14 (3): 204–20. https://doi.org/10.1038/nrg3354.

Solter, D., N. Skreb, and I. Damjanov. 1970. 'Extrauterine Growth of Mouse Egg-Cylinders Results in Malignant Teratoma'. *Nature* 227 (5257): 503–4.

Stadtfeld, Matthias, Nimet Maherali, and Konrad Hochedlinger. 2008. 'Defining Molecular Cornerstones during Fibroblast to IPS Cell Reprogramming in Mouse'. *Cell Stem Cell* 2 (3): 230–40. https://doi.org/10.1016/j.stem.2008.02.001.

Stavridis, Marios P., J. Simon Lunn, Barry J. Collins, and Kate G. Storey. 2007. 'A Discrete Period of FGF-Induced Erk1/2 Signalling Is Required for Vertebrate Neural Specification'. *Development (Cambridge, England)* 134 (16): 2889–94. https://doi.org/10.1242/dev.02858.

Sternberg, Samuel H., Sy Redding, Martin Jinek, Eric C. Greene, and Jennifer A. Doudna. 2014. 'DNA Interrogation by the CRISPR RNA-Guided Endonuclease Cas9'. *Nature* 507 (7490): 62–67. https://doi.org/10.1038/nature13011.

Stevens, L. C. 1968. 'The Development of Teratomas from Intratesticular Grafts of Tubal Mouse Eggs'. *Journal of Embryology and Experimental Morphology* 20 (3): 329–41.

Stevens, Leroy C., and C. C. Little. 1954. 'Spontaneous Testicular Teratomas in an Inbred Strain of Mice'. *Proceedings of the National Academy of Sciences of the United States of America* 40 (11): 1080–87.

Stewart, C. L., P. Kaspar, L. J. Brunet, H. Bhatt, I. Gadi, F. Köntgen, and S. J. Abbondanzo. 1992. 'Blastocyst Implantation Depends on Maternal Expression of Leukaemia Inhibitory Factor'. *Nature* 359 (6390): 76–79. https://doi.org/10.1038/359076a0.

Stock, Julie K., Sara Giadrossi, Miguel Casanova, Emily Brookes, Miguel Vidal, Haruhiko Koseki, Neil Brockdorff, Amanda G. Fisher, and Ana Pombo. 2007. 'Ring1-Mediated Ubiquitination of H2A Restrains Poised RNA Polymerase II at Bivalent Genes in Mouse ES Cells'. *Nature Cell Biology* 9 (12): 1428–35. https://doi.org/10.1038/ncb1663.

Stoop, Petra van der, Erwin A. Boutsma, Danielle Hulsman, Sonja Noback, Mike Heimerikx, Ron M. Kerkhoven, J. Willem Voncken, Lodewyk F. A. Wessels, and Maarten van Lohuizen. 2008. 'Ubiquitin E3 Ligase Ring1b/Rnf2 of Polycomb Repressive Complex 1 Contributes to Stable Maintenance of Mouse Embryonic Stem Cells'. *PloS One* 3 (5): e2235. https://doi.org/10.1371/journal.pone.0002235.

Stuart, Hannah T., Anouk L. van Oosten, Aliaksandra Radzisheuskaya, Graziano Martello, Anzy Miller, Sabine Dietmann, Jennifer Nichols, and José C. R. Silva. 2014. 'NANOG Amplifies STAT3 Activation and They Synergistically Induce the Naive Pluripotent Program'. *Current Biology* 24 (3): 340–46. https://doi.org/10.1016/j.cub.2013.12.040.

Tabansky, Inna, Alan Lenarcic, Ryan W. Draft, Karine Loulier, Derin B. Keskin, Jacqueline Rosains, José Rivera-Feliciano, et al. 2013. 'Developmental Bias in Cleavage-Stage Mouse Blastomeres'. *Current Biology: CB* 23 (1): 21–31. https://doi.org/10.1016/j.cub.2012.10.054.

Tadros, Wael, and Howard D. Lipshitz. 2009. 'The Maternal-to-Zygotic Transition: A Play in Two Acts'. *Development* 136 (18): 3033–42. https://doi.org/10.1242/dev.033183.

Tai, Chih-I., and Qi-Long Ying. 2013. 'Gbx2, a LIF/Stat3 Target, Promotes Reprogramming to and Retention of the Pluripotent Ground State'. *Journal of Cell Science* 126 (Pt 5): 1093–98. https://doi.org/10.1242/jcs.118273.

Takahashi, Kazutoshi, and Shinya Yamanaka. 2006. 'Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors'. *Cell* 126 (4): 663–76. https://doi.org/10.1016/j.cell.2006.07.024.

Takeda, Kiyoshi, Koichi Noguchi, Wei Shi, Takashi Tanaka, Makoto Matsumoto, Nobuaki Yoshida, Tadamitsu Kishimoto, and Shizuo Akira. 1997. 'Targeted Disruption of the Mouse Stat3 Gene Leads to Early Embryonic Lethality'. *Proceedings of the National Academy of Sciences* 94 (8): 3801–4. https://doi.org/10.1073/pnas.94.8.3801.

Tan, Frederick E., and Michael B. Elowitz. 2014. 'Brf1 Posttranscriptionally Regulates Pluripotency and Differentiation Responses Downstream of Erk MAP Kinase'. *Proceedings of the National Academy of Sciences of the United States of America* 111 (17): E1740-1748. https://doi.org/10.1073/pnas.1320873111.

Tanenbaum, Marvin E., Luke A. Gilbert, Lei S. Qi, Jonathan S. Weissman, and Ronald D. Vale. 2014. 'A Protein Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging'. *Cell* 159 (3): 635–46. https://doi.org/10.1016/j.cell.2014.09.039.

Tang, Thean-Hock, Jean-Pierre Bachellerie, Timofey Rozhdestvensky, Marie-Line Bortolin, Harald Huber, Mario Drungowski, Thorsten Elge, Jürgen Brosius, and Alexander Hüttenhofer. 2002. 'Identification of 86 Candidates for Small Non-Messenger RNAs from the Archaeon Archaeoglobus Fulgidus'. *Proceedings of the National Academy of Sciences of the United States of America* 99 (11): 7536–41. https://doi.org/10.1073/pnas.112047299.

Tang, Y. Amy, Derek Huntley, Giovanni Montana, Andrea Cerase, Tatyana B. Nesterova, and Neil Brockdorff. 2010. 'Efficiency of Xist-Mediated Silencing on Autosomes Is Linked to Chromosomal Domain Organisation'. *Epigenetics & Chromatin* 3 (May): 10. https://doi.org/10.1186/1756-8935-3-10.

Tee, Wee-Wei, Steven S. Shen, Ozgur Oksuz, Varun Narendra, and Danny Reinberg. 2014. 'Erk1/2 Activity Promotes Chromatin Features and RNAPII Phosphorylation at Developmental Promoters in Mouse ESCs'. *Cell* 156 (4): 678–90. https://doi.org/10.1016/j.cell.2014.01.009.

Tesar, Paul J., Josh G. Chenoweth, Frances A. Brook, Timothy J. Davies, Edward P. Evans, David L. Mack, Richard L. Gardner, and Ronald D. G. McKay. 2007. 'New Cell Lines from Mouse Epiblast Share Defining Features with Human Embryonic Stem Cells'. *Nature* 448 (7150): 196–99. https://doi.org/10.1038/nature05972.

The FANTOM Consortium, Alexandre Fort, Kosuke Hashimoto, Daisuke Yamada, Md Salimullah, Chaman A Keya, Alka Saxena, et al. 2014. 'Deep Transcriptome Profiling of Mammalian Stem Cells Supports a Regulatory Role for Retrotransposons in Pluripotency Maintenance'. *Nature Genetics* 46 (6): 558–66. https://doi.org/10.1038/ng.2965.

Theodorou, Elias, George Dalembert, Christopher Heffelfinger, Eric White, Sherman Weissman, Lynn Corcoran, and Michael Snyder. 2009. 'A High Throughput Embryonic Stem Cell Screen Identifies Oct-2 as a Bifunctional Regulator of Neuronal Differentiation'. *Genes & Development* 23 (5): 575–88. https://doi.org/10.1101/gad.1772509.

Thomas, K. R., and M. R. Capecchi. 1987. 'Site-Directed Mutagenesis by Gene Targeting in Mouse Embryo-Derived Stem Cells'. *Cell* 51 (3): 503–12.

Thomson, Matt, Siyuan John Liu, Ling-Nan Zou, Zack Smith, Alexander Meissner, and Sharad Ramanathan. 2011. 'Pluripotency Factors in Embryonic Stem Cells Regulate Differentiation into Germ Layers'. *Cell* 145 (6): 875–89. https://doi.org/10.1016/j.cell.2011.05.017.

Thormann, Verena, Maika C. Rothkegel, Robert Schöpflin, Laura V. Glaser, Petar Djuric, Na Li, Ho-Ryun Chung, Kevin Schwahn, Martin Vingron, and Sebastiaan H. Meijsing. 2018. 'Genomic Dissection of Enhancers Uncovers Principles of Combinatorial Regulation and Cell Type-Specific Wiring of Enhancer–Promoter Contacts'. *Nucleic Acids Research* 46 (6): 2868–82. https://doi.org/10.1093/nar/gky051.

Tyagi, Richa, Wenxue Li, Danelvis Parades, Mario A. Bianchet, and Avindra Nath. 2017. 'Inhibition of Human Endogenous Retrovirus-K by Antiretroviral Drugs'. *Retrovirology* 14 (March). https://doi.org/10.1186/s12977-017-0347-4.

Tycko, Josh, Vic E. Myer, and Patrick D. Hsu. 2016. 'Methods for Optimizing CRISPR-Cas9 Genome Editing Specificity'. *Molecular Cell* 63 (3): 355–70. https://doi.org/10.1016/j.molcel.2016.07.004.

Ulitsky, Igor. 2016. 'Evolution to the Rescue: Using Comparative Genomics to Understand Long Non-Coding RNAs'. *Nature Reviews Genetics* 17 (10): 601–14. https://doi.org/10.1038/nrg.2016.85.

Ulitsky, Igor, and David P. Bartel. 2013. 'LincRNAs: Genomics, Evolution, and Mechanisms'. *Cell* 154 (1): 26–46. https://doi.org/10.1016/j.cell.2013.06.020.

Ulitsky, Igor, Alena Shkumatava, Calvin H. Jan, Hazel Sive, and David P. Bartel. 2011. 'Conserved Function of LincRNAs in Vertebrate Embryonic Development despite Rapid Sequence Evolution'. *Cell* 147 (7): 1537–50. https://doi.org/10.1016/j.cell.2011.11.055.

Utley, Rhea T., Keiko Ikeda, Patrick A. Grant, Jacques Côté, David J. Steger, Anton Eberharter, Sam John, and Jerry L. Workman. 1998. 'Transcriptional Activators Direct Histone Acetyltransferase Complexes to Nucleosomes'. *Nature* 394 (6692): 498–502. https://doi.org/10.1038/28886.

Velazquez Camacho, Oscar, Carmen Galan, Kalina Swist-Rosowska, Reagan Ching, Michael Gamalinda, Fethullah Karabiber, Inti De La Rosa-Velazquez, et al. 2017. 'Major Satellite Repeat RNA Stabilize Heterochromatin Retention of Suv39h Enzymes by RNA-Nucleosome Association and RNA:DNA Hybrid Formation'. *ELife* 6. https://doi.org/10.7554/eLife.25293.

Vignali, Marissa, David J. Steger, Kristen E. Neely, and Jerry L. Workman. 2000. 'Distribution of Acetylated Histones Resulting from Gal4-VP16 Recruitment of SAGA and NuA4 Complexes'. *The EMBO Journal* 19 (11): 2629–40. https://doi.org/10.1093/emboj/19.11.2629.

Voigt, Philipp, Gary LeRoy, William J. Drury, Barry M. Zee, Jinsook Son, David B. Beck, Nicolas L. Young, Benjamin A. Garcia, and Danny Reinberg. 2012. 'Asymmetrically Modified Nucleosomes'. *Cell* 151 (1): 181–93. https://doi.org/10.1016/j.cell.2012.09.002.

Walsh, Derek, Michael B. Mathews, and Ian Mohr. 2013. 'Tinkering with Translation: Protein Synthesis in Virus-Infected Cells'. *Cold Spring Harbor Perspectives in Biology* 5 (1): a012351. https://doi.org/10.1101/cshperspect.a012351.

Wang, Jianlong, Sridhar Rao, Jianlin Chu, Xiaohua Shen, Dana N. Levasseur, Thorold W. Theunissen, and Stuart H. Orkin. 2006. 'A Protein Interaction Network for Pluripotency of Embryonic Stem Cells'. *Nature* 444 (7117): 364–68. https://doi.org/10.1038/nature05284.

Wang, Kevin C., and Howard Y. Chang. 2011. 'Molecular Mechanisms of Long Noncoding RNAs'. *Molecular Cell* 43 (6): 904–14. https://doi.org/10.1016/j.molcel.2011.08.018.

Wang, Kevin C., Yul W. Yang, Bo Liu, Amartya Sanyal, Ryan Corces-Zimmerman, Yong Chen, Bryan R. Lajoie, et al. 2011. 'Long Noncoding RNA Programs Active Chromatin Domain to Coordinate Homeotic Gene Activation'. *Nature* 472 (7341): 120–24. https://doi.org/10.1038/nature09819.

Wang, Liguo, Hyun Jung Park, Surendra Dasari, Shengqin Wang, Jean-Pierre Kocher, and Wei Li. 2013. 'CPAT: Coding-Potential Assessment Tool Using an Alignment-Free Logistic Regression Model'. *Nucleic Acids Research* 41 (6): e74. https://doi.org/10.1093/nar/gkt006.

Wang, Song S., Joseph W. Lewcock, Paul Feinstein, Peter Mombaerts, and Randall R. Reed. 2004. 'Genetic Disruptions of O/E2 and O/E3 Genes Reveal Involvement in Olfactory Receptor Neuron Projection'. *Development* 131 (6): 1377–88. https://doi.org/10.1242/dev.01009.

Wang, Sue-Hong, Ming-Shiun Tsai, Ming-Fu Chiang, and Hung Li. 2003. 'A Novel NK-Type Homeobox Gene, ENK (Early Embryo Specific NK), Preferentially Expressed in

Embryonic Stem Cells'. *Gene Expression Patterns* 3 (1): 99–103. https://doi.org/10.1016/S1567-133X(03)00005-X.

Wang, Wei, Chengyi Lin, Dong Lu, Zeming Ning, Tony Cox, David Melvin, Xiaozhong Wang, Allan Bradley, and Pentao Liu. 2008. 'Chromosomal Transposition of PiggyBac in Mouse Embryonic Stem Cells'. *Proceedings of the National Academy of Sciences* 105 (27): 9290–95. https://doi.org/10.1073/pnas.0801017105.

Wang, Xue Q. D., and Josée Dostie. 2017. 'Reciprocal Regulation of Chromatin State and Architecture by HOTAIRM1 Contributes to Temporal Collinear HOXA Gene Activation'. *Nucleic Acids Research* 45 (3): 1091–1104. https://doi.org/10.1093/nar/gkw966.

Wang, Yue, Zhenyu Xu, Junfeng Jiang, Chen Xu, Jiuhong Kang, Lei Xiao, Minjuan Wu, Jun Xiong, Xiaocan Guo, and Houqi Liu. 2013. 'Endogenous MiRNA Sponge LincRNA-RoR Regulates Oct4, Nanog, and Sox2 in Human Embryonic Stem Cell Self-Renewal'. *Developmental Cell* 25 (1): 69–80. https://doi.org/10.1016/j.devcel.2013.03.002.

Weaver, Jamie R., Martha Susiarjo, and Marisa S. Bartolomei. 2009. 'Imprinting and Epigenetic Changes in the Early Embryo'. *Mammalian Genome: Official Journal of the International Mammalian Genome Society* 20 (9–10): 532–43. https://doi.org/10.1007/s00335-009-9225-2.

Wei, Shu, Qingjian Zou, Sisi Lai, Quanjun Zhang, Li Li, Quanmei Yan, Xiaoqing Zhou, Huilin Zhong, and Liangxue Lai. 2016. 'Conversion of Embryonic Stem Cells into Extraembryonic Lineages by CRISPR-Mediated Activators'. *Scientific Reports* 6 (January): 19648. https://doi.org/10.1038/srep19648.

White, M. D., S. Bissiere, Y. D. Alvarez, and N. Plachta. 2016. 'Chapter Seven - Mouse Embryo Compaction'. In *Current Topics in Developmental Biology*, edited by Melvin L. DePamphilis, 120:235–58. Mammalian Preimplantation Development. Academic Press. https://doi.org/10.1016/bs.ctdb.2016.04.005.

Whyte, Warren A., David A. Orlando, Denes Hnisz, Brian J. Abraham, Charles Y. Lin, Michael H. Kagey, Peter B. Rahl, Tong Ihn Lee, and Richard A. Young. 2013. 'Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes'. *Cell* 153 (2): 307–19. https://doi.org/10.1016/j.cell.2013.03.035.

Wilder, P. J., D. Kelly, K. Brigman, C. L. Peterson, T. Nowling, Q. S. Gao, R. D. McComb, M. R. Capecchi, and A. Rizzino. 1997. 'Inactivation of the FGF-4 Gene in Embryonic Stem Cells Alters the Growth and/or the Survival of Their Early Differentiated Progeny'. *Developmental Biology* 192 (2): 614–29. https://doi.org/10.1006/dbio.1997.8777.

Wiles, M. V., and B. M. Johansson. 1999. ‘Embryonic Stem Cell Development in a Chemically Defined Medium’. *Experimental Cell Research* 247 (1): 241–48. https://doi.org/10.1006/excr.1998.4353.

Williams, R. L., 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. 1988. ‘Myeloid Leukaemia Inhibitory Factor Maintains the Developmental Potential of Embryonic Stem Cells’. *Nature* 336 (6200): 684–87. https://doi.org/10.1038/336684a0.

Wilusz, Jeremy E., Susan M. Freier, and David L. Spector. 2008. ‘3’ End Processing of a Long Nuclear-Retained Non-Coding RNA Yields a TRNA-like Cytoplasmic RNA’. *Cell* 135 (5): 919–32. https://doi.org/10.1016/j.cell.2008.10.012.

Wray, Jason, Tüzer Kalkan, Sandra Gomez-Lopez, Dominik Eckardt, Andrew Cook, Rolf Kemler, and Austin Smith. 2011. ‘Inhibition of Glycogen Synthase Kinase-3 Alleviates Tcf3 Repression of the Pluripotency Network and Increases Embryonic Stem Cell Resistance to Differentiation’. *Nature Cell Biology* 13 (7): 838–45. https://doi.org/10.1038/ncb2267.

Wu, Chun-I., Jackson A. Hoffman, Brian R. Shy, Erin M. Ford, Elaine Fuchs, Hoang Nguyen, and Bradley J. Merrill. 2012. ‘Function of Wnt/β-Catenin in Counteracting Tcf3 Repression through the Tcf3-β-Catenin Interaction’. *Development (Cambridge, England)* 139 (12): 2118–29. https://doi.org/10.1242/dev.076067.

Wu, Da Yong, and Zhen Yao. 2005. ‘Isolation and Characterization of the Murine *Nanog* Gene Promoter’. *Cell Research* 15 (5): 317–24. https://doi.org/10.1038/sj.cr.7290300.

Wu, Susan C., Eric M. Kallin, and Yi Zhang. 2010. ‘Role of H3K27 Methylation in the Regulation of LncRNA Expression’. *Cell Research* 20 (10): 1109–16. https://doi.org/10.1038/cr.2010.114.

Wu, Xuebing, David A. Scott, Andrea J. Kriz, Anthony C. Chiu, Patrick D. Hsu, Daniel B. Dadon, Albert W. Cheng, et al. 2014a. ‘Genome-Wide Binding of the CRISPR Endonuclease Cas9 in Mammalian Cells’. *Nature Biotechnology* 32 (7): 670–76. https://doi.org/10.1038/nbt.2889.

———. 2014b. ‘Genome-Wide Binding of the CRISPR Endonuclease Cas9 in Mammalian Cells’. *Nature Biotechnology* 32 (7): 670–76. https://doi.org/10.1038/nbt.2889.

Xenopoulos, Panagiotis, Minjung Kang, Alberto Puliafito, Stefano Di Talia, and Anna-Katerina Hadjantonakis. 2015. ‘Heterogeneities in Nanog Expression Drive Stable Commitment to Pluripotency in the Mouse Blastocyst’. *Cell Reports*, March. https://doi.org/10.1016/j.celrep.2015.02.010.

Xu, Chunlong, Xiaolan Qi, Xuguang Du, Huiying Zou, Fei Gao, Tao Feng, Hengxing Lu, et al. 2017. ‘PiggyBac Mediates Efficient in Vivo CRISPR Library Screening for

Tumorigenesis in Mice'. *Proceedings of the National Academy of Sciences* 114 (4): 722–27. https://doi.org/10.1073/pnas.1615735114.

Yamamizu, Kohei, Yulan Piao, Alexei A. Sharov, Veronika Zsiros, Hong Yu, Kazu Nakazawa, David Schlessinger, and Minoru S. H. Ko. 2013. 'Identification of Transcription Factors for Lineage-Specific ESC Differentiation'. *Stem Cell Reports* 1 (6): 545–59. https://doi.org/10.1016/j.stemcr.2013.10.006.

Yamamizu, Kohei, Alexei A. Sharov, Yulan Piao, Misa Amano, Hong Yu, Akira Nishiyama, Dawood B. Dudekula, David Schlessinger, and Minoru S. H. Ko. 2016. 'Generation and Gene Expression Profiling of 48 Transcription-Factor-Inducible Mouse Embryonic Stem Cell Lines'. *Scientific Reports* 6 (May): 25667. https://doi.org/10.1038/srep25667.

Yamane, Mariko, Satoshi Ohtsuka, Kumi Matsuura, Akira Nakamura, and Hitoshi Niwa. 2018. 'Overlapping Functions of Krüppel-like Factor Family Members: Targeting Multiple Transcription Factors to Maintain the Naïve Pluripotency of Mouse Embryonic Stem Cells'. *Development* 145 (10): dev162404. https://doi.org/10.1242/dev.162404.

Yang, Shen-Hsi, Tüzer Kalkan, Claire Morissroe, Hendrik Marks, Hendrik Stunnenberg, Austin Smith, and Andrew D. Sharrocks. 2014. 'Otx2 and Oct4 Drive Early Enhancer Activation during Embryonic Stem Cell Transition from Naive Pluripotency'. *Cell Reports* 7 (6): 1968–81. https://doi.org/10.1016/j.celrep.2014.05.037.

Yanmei Chen, Zhongwei Du, and Zhen Yao. 2006. 'Roles of the Nanog Protein in Murine F9 Embryonal Carcinoma Cells and Their Endoderm-Differentiated Counterparts | Cell Research'. 2006. https://www.nature.com/articles/7310067.

Yao, Shuyuan, Tanya Sukonnik, Tara Kean, Rikki R. Bharadwaj, Peter Pasceri, and James Ellis. 2004. 'Retrovirus Silencing, Variegation, Extinction, and Memory Are Controlled by a Dynamic Interplay of Multiple Epigenetic Modifications'. *Molecular Therapy* 10 (1): 27–36. https://doi.org/10.1016/j.ymthe.2004.04.007.

Ye, Shoudong, Ping Li, Chang Tong, and Qi-Long Ying. 2013a. 'Embryonic Stem Cell Self-Renewal Pathways Converge on the Transcription Factor Tfcp2l1'. *The EMBO Journal* 32 (19): 2548–60. https://doi.org/10.1038/emboj.2013.175.

———. 2013b. 'Embryonic Stem Cell Self-renewal Pathways Converge on the Transcription Factor Tfcp2l1'. *The EMBO Journal* 32 (19): 2548–60. https://doi.org/10.1038/emboj.2013.175.

Yeo, Jia-Chi, Jianming Jiang, Zi-Ying Tan, Guo-Rong Yim, Jia-Hui Ng, Jonathan Göke, Petra Kraus, et al. 2014. 'Klf2 Is an Essential Factor That Sustains Ground State Pluripotency'. *Cell Stem Cell* 14 (6): 864–72. https://doi.org/10.1016/j.stem.2014.04.015.

Ying, Qi-Long, Jennifer Nichols, Ian Chambers, and Austin Smith. 2003. 'BMP Induction of Id Proteins Suppresses Differentiation and Sustains Embryonic Stem Cell Self-Renewal in Collaboration with STAT3'. *Cell* 115 (3): 281–92. https://doi.org/10.1016/S0092-8674(03)00847-X.

Ying, Qi-Long, Marios Stavridis, Dean Griffiths, Meng Li, and Austin Smith. 2003. 'Conversion of Embryonic Stem Cells into Neuroectodermal Precursors in Adherent Monoculture'. *Nature Biotechnology* 21 (2): 183–86. https://doi.org/10.1038/nbt780.

Ying, Qi-Long, Jason Wray, Jennifer Nichols, Laura Batlle-Morera, Bradley Doble, James Woodgett, Philip Cohen, and Austin Smith. 2008. 'The Ground State of Embryonic Stem Cell Self-Renewal'. *Nature* 453 (7194): 519–23. https://doi.org/10.1038/nature06968.

Yoshida, K, T Taga, M Saito, S Suematsu, A Kumanogoh, T Tanaka, H Fujiwara, et al. 1996. 'Targeted Disruption of Gp130, a Common Signal Transducer for the Interleukin 6 Family of Cytokines, Leads to Myocardial and Hematological Disorders.' *Proceedings of the National Academy of Sciences of the United States of America* 93 (1): 407–11.

Yu, Junying, Maxim A. Vodyanik, Kim Smuga-Otto, Jessica Antosiewicz-Bourget, Jennifer L. Frane, Shulan Tian, Jeff Nie, et al. 2007. 'Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells'. *Science (New York, N.Y.)* 318 (5858): 1917–20. https://doi.org/10.1126/science.1151526.

Zalzman, Michal, Geppino Falco, Lioudmila V. Sharova, Akira Nishiyama, Marshall Thomas, Sung-Lim Lee, Carole A. Stagg, et al. 2010. 'Zscan4 Regulates Telomere Elongation and Genomic Stability in ES Cells'. *Nature* 464 (7290): 858–63. https://doi.org/10.1038/nature08882.

Zeng, Chao, Tsukasa Fukunaga, and Michiaki Hamada. 2018. 'Identification and Analysis of Ribosome-Associated LncRNAs Using Ribosome Profiling Data'. *BMC Genomics* 19 (May). https://doi.org/10.1186/s12864-018-4765-z.

Zerbino, Daniel R, Premanand Achuthan, Wasiu Akanni, M Ridwan Amode, Daniel Barrell, Jyothish Bhai, Konstantinos Billis, et al. 2018. 'Ensembl 2018'. *Nucleic Acids Research* 46 (Database issue): D754–61. https://doi.org/10.1093/nar/gkx1098.

Zhang, Kejing, Lingyu Li, Chengyang Huang, Chengyong Shen, Fangzhi Tan, Caihong Xia, Pingyu Liu, Janet Rossant, and Naihe Jing. 2010. 'Distinct Functions of BMP4 during Different Stages of Mouse ES Cell Neural Commitment'. *Development*, January, dev.049494. https://doi.org/10.1242/dev.049494.

Zhang, Man, Harry G. Leitch, Walfred W. C. Tang, Nicola Festuccia, Elisa Hall-Ponsele, Jennifer Nichols, M. Azim Surani, Austin Smith, and Ian Chambers. 2018. 'Esrrb Complementation Rescues Development of Nanog-Null Germ Cells'. *Cell Reports* 22 (2): 332–39. https://doi.org/10.1016/j.celrep.2017.12.060.

Zhang, Peilin, Rose Andrianakos, Yang Yang, Chunming Liu, and Wange Lu. 2010. 'Kruppel-like Factor 4 (Klf4) Prevents Embryonic Stem (ES) Cell Differentiation by Regulating Nanog Gene Expression'. *The Journal of Biological Chemistry* 285 (12): 9180–89. https://doi.org/10.1074/jbc.M109.077958.

Zhang, Xiaofei, Juan Zhang, Tao Wang, Miguel A. Esteban, and Duanqing Pei. 2008. 'Esrrb Activates Oct4 Transcription and Sustains Self-Renewal and Pluripotency in Embryonic Stem Cells'. *Journal of Biological Chemistry* 283 (51): 35825–33. https://doi.org/10.1074/jbc.M803481200.

Zhang, Xiao-Hui, Louis Y. Tee, Xiao-Gang Wang, Qun-Shan Huang, and Shi-Hua Yang. 2015. 'Off-Target Effects in CRISPR/Cas9-Mediated Genome Engineering'. *Molecular Therapy - Nucleic Acids* 4 (January). https://doi.org/10.1038/mtna.2015.37.

Zhang, Yang, Li Yang, and Ling-Ling Chen. 2014. 'Life without A Tail: New Formats of Long Noncoding RNAs'. *The International Journal of Biochemistry & Cell Biology* 54 (September): 338–49. https://doi.org/10.1016/j.biocel.2013.10.009.

Zhao, Jing, Toshiro K. Ohsumi, Johnny T. Kung, Yuya Ogawa, Daniel J. Grau, Kavitha Sarma, Ji Joon Song, Robert E. Kingston, Mark Borowsky, and Jeannie T. Lee. 2010. 'Genome-Wide Identification of Polycomb-Associated RNAs by RIP-Seq'. *Molecular Cell* 40 (6): 939–53. https://doi.org/10.1016/j.molcel.2010.12.011.

Zhao, Suling, Jennifer Nichols, Austin G. Smith, and Meng Li. 2004. 'SoxB Transcription Factors Specify Neuroectodermal Lineage Choice in ES Cells'. *Molecular and Cellular Neuroscience* 27 (3): 332–42. https://doi.org/10.1016/j.mcn.2004.08.002.

Zhao, Yi, Hui Li, Shuangsang Fang, Yue Kang, Wei wu, Yajing Hao, Ziyang Li, et al. 2016. 'NONCODE 2016: An Informative and Valuable Data Source of Long Non-Coding RNAs'. *Nucleic Acids Research* 44 (Database issue): D203–8. https://doi.org/10.1093/nar/gkv1252.

Zheng, Ling-Ling, Jun-Hao Li, Jie Wu, Wen-Ju Sun, Shun Liu, Ze-Lin Wang, Hui Zhou, Jian-Hua Yang, and Liang-Hu Qu. 2016. 'DeepBase v2.0: Identification, Expression, Evolution and Function of Small RNAs, LncRNAs and Circular RNAs from Deep-Sequencing Data'. *Nucleic Acids Research* 44 (Database issue): D196–202. https://doi.org/10.1093/nar/gkv1273.

Zhu, Qingqing, Lu Song, Guangdun Peng, Na Sun, Jun Chen, Ting Zhang, Nengyin Sheng, et al. 2014. 'The Transcription Factor Pou3f1 Promotes Neural Fate Commitment via Activation of Neural Lineage Genes and Inhibition of External Signaling Pathways'. *ELife* 3 (June). https://doi.org/10.7554/eLife.02224.

Zong, Xinying, Shinichi Nakagawa, Susan M. Freier, Jingyi Fei, Taekjip Ha, Supriya G. Prasanth, and Kannanganattu V. Prasanth. 2016. 'Natural Antisense RNA Promotes 3'

End Processing and Maturation of MALAT1 LncRNA'. *Nucleic Acids Research* 44 (6): 2898–2908. https://doi.org/10.1093/nar/gkw047.