Nuclear organization and regulation of gene expression in mouse Embryonic Stem Cells by long non-coding RNAs

Alexandra Tachtsidi

Nuclear organization and regulation of gene expression in mouse Embryonic Stem Cells by long non-coding RNAs.

Alexandra TACHTSIDI

Thesis presented for the degree of Doctor of Philosophy
Directed by Pablo Navarro Gil

Epigenetics of stem cells Laboratory - Pasteur Institute- Paris
Ecole Doctorale Complexité du Vivant (ED 515)
Sorbonne Université

December 11, 2018

Thesis Defense Committee:

President : Pr MORILLON Antonin, Research director

Examiner : Pr FEIL Robert, Research director

Examiner : Pr CIAUDO Constance, Research director

Member : Dr FRANCASTEL Claire, Research director

Member : Dr CHAUMEIL Julie, Principal Investigator

Supervisor : Dr NAVARRO GIL Pablo, Principal Investigator

Acknowledgements

I am greatly thankful to my thesis director, Pablo Navarro Gil, for giving me the chance to work in his laboratory. Thank you for always pushing me further and redefining my limits, shaping my perseverance and endurance. Thank you for the opportunity to use many different experimental techniques that were particularly enriching for me growing as a scientist while the experience of supervising a master student has been instructive for my managerial skills. I am also grateful to you for the possibility to participate in the fascinating experience of a Cold Spring Harbor Course and to attend very interesting international conferences during these years.

I would like to thank all the past and present members of our laboratory, the ECS team. It was a real pleasure to work with all of you guys these last four years, and to develop personal relationships outside the laboratory too. Thank you for all the help, advice and thoughts that we have exchanged and the moments of joy and laughter that we have shared. Agnes, thank you for keeping the lab running functional, for always being a kind and helpful colleague but also a friend. Thank you Philippe for always answering my questions and sharing so generously your wide knowledge and scientific expertise, and your office the last few months. Thank you Nicola for being an inspiring colleague and a friend from the very beginning of my thesis. I wish to express my gratitude to Thaleia for the liberating feeling of speaking daily in our mother-tongue, for bringing an extra sense of organization to the lab, for her support, suggestions and corrections. Thank you Nick for all our interesting discussions, the explanations and the analyses that you have done on my sequencing results. Inma, thank you for your positivity, support and kindness. I thank Nancy for our collaboration during her master internship. And Laurence, thank you for taking care of our administrative work.

I thank past and present members of the Department of Stem Cells and Developmental Biology of Pasteur Institute for creating a supportive and fun working environment but also Frank, Françoise and Masha for our discussions.

I would also like to express my gratitude to my Thesis Committee Members, Deborah Bourc'his, Antonin Morillon and Marco Vignuzzi, for their support and instructions during the years of the thesis.

Many friends have been surrounding me these last years to whom I am deeply grateful for their friendship and support. Especially, thank you Elma for suggesting I should come to Paris to pursue my Master and PhD studies, for being helpful and caring through the good and tough moments of our long student life.

A big, special thank you goes to you Victor, for your unlimited support and encouragement, for being always by my side loving and caring. Thank you for believing in me when I didn't believe in myself; for making me see science and the world through a different perspective.

Finally, all my gratitude goes to my parents, Eirini and Michail. В заключение, хотела бы выразить всю мою благодарность моим родителям. Папа, Мама спасибо вам огромное за вашу бесконечную любовь, поддержку и веру в меня. За все что вы сделали для меня, и продолжаете делать до сих пор, за все чему научили меня и предоставили. Все чего я достигла, является и вашей заслугой.

Abstract

The nucleus is a highly structured organelle and its complex architectural organization enables and facilitates different biological processes to take place at distinct subnuclear domains. The implication of long non-coding RNAs (lncRNAs) in nuclear organization by establishing and maintaining nuclear compartmentalization is now widely accepted. Numerous examples have been shown to either participate in the structuration of subnuclear domains or in the establishment of long range interactions in the three-dimensional nuclear space. However, a robust approach for the identification of "nuclear organizers" molecules such as Xist, Neat1, and Firre that shape the nucleus is currently lacking. To that end, we established an experimental approach that would allow us to identify such "structural" lncRNAs on a genome-scale level. Based on the biochemical property of known nuclear organizing lncRNAs to resist the so called nuclear matrix preparation, where most of the DNA and soluble molecules are removed, we performed nuclear matrix fractionation on mouse Embryonic Stem Cells (mESCs), purified the RNA fraction and explored its constituents by RNA-sequencing. We identified in such a way, a subset of transcripts (non-extracted RNAs, nextRNAs) potentially involved in the functional compartmentalization of the nucleus. The group of nextRNAs identified by RNA-seq was validated by RT-qPCR and contained few transcripts that are already known and described to be "nuclear organizers" (e.g. Xist, Firre, Neat1). Notably, we detected previously non-annotated transcripts thanks to our original RNA-seq datasets and focused our work on two of them: NextC1 (Next Candidate 1) and NextC2.

We extensively described and characterized the identified NextC1 and NextC2 on a functional and phenotypical level. The expression profile of the transcripts was studied in pluripotent and differentiating culturing conditions, in mutant cell lines for pluripotency transcription factors (TFs) as well as in different embryo-derived cell types. The subcellular localization of both lncRNAs was assessed by RNA-FISH. Loss- and gain-of-function assays were performed by targeting the promoter regions of NextC1 and 2 with the canonical CRISPR/Cas9 system for genome editing and CRISPR-derived systems for transcription inhibition or activation. Many of these functional assays were subsequently RNA-sequenced and an integrative data analysis is currently under investigation.

To my parents, to whom I am deeply grateful

For everything they have done for me and keep doing to date.

Table of Contents

Index of abbreviations

BFP: Blue Fluorescent Protein

ChIP: Chromatin Immunoprecipitation

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

CTCF: CCCTC-binding factor

DAPI: 4',6-Diamidino-2-Phenylindole

dCas9: catalytically dead Cas9 protein

DE: Differentially Expressed

EB: Embryoid Body

EpiLC: Epiblast-Like Cells

FACS: Fluorescence Activated Cell Sorter

FC: Fold Change

FCS: Fetal Calf Serum

FISH: Fluorescence In Situ Hybridization

GFP: Green Fluorescent Protein

KO: Knock Out

LAD: Lamina-Associated Domain

LIF: Leukaemia Inhibiting Factor

lincRNA: Long intergenic non-coding RNA

lncRNA: Long non-coding RNAs

mESC: Mouse Embryonic Stem Cells

miRNA: microRNA

ORF: Open Reading Frame

PRC2: Polycomb Repressive Complex 2

RA: Retinoic Acid

RFP: Red Fluorescent Protein

RNA-seq: RNA Sequencing

RT-qPCR: Real Time quantitatice Polymerase Chain Reaction

SE: Super Enhancer

smFISH: Single Molecule Fluorescence In Situ Hybridization

TAD: Topologically Associating Domain

TF: Transcription Factor

Tpm: Transcripts per million

TSS: Transcriptional Start Site

UCSC: University of California Santa Cruz (Database)

VP64: Four copies of VP16 (viral protein of 16 amino acids) transcriptional activator protein

WT: Wild Type

Xist: X Inactive Specific Transcript

Introduction

I. Nuclear organization

The characteristic feature of eukaryotic cells is the presence of a nucleus; an organelle which has a complex and dynamic organization. The nuclear architecture has triggered the research from the spectrum of cellular and developmental biology for almost a century now. Yet, the nuclear compartmentalization and its functionality have not been fully characterized. The high-order organization of the mammalian nucleus allows different biological processes to take place in distinct subnuclear compartments and it serves for a precise regulation of gene expression during different developmental stages through chromatin modifications and architectural rearrangements.

The analysis of the nuclear organization was inaugurated by E. Heitz in 1928, with the observation that the transcriptionally active euchromatin is decondensed and no longer visible during interphase, whereas heterochromatin -that is transcriptionally less active- is still visible following mitosis. Later on, the establishment of electron microscopy (EM) supported his conception and also showed that nuclear organization markedly varies in different cell types or developmental stages, however, it is similar between cells of a given cell type (Pueschel et al., 2016). The first big step forward in the study of the nuclear organization was achieved with fluorescence-based microscopy that allowed the precise localization of proteins and genes in relation to nuclear landmarks. Many nuclear bodies were discovered in that way and their consisting proteins and genes were identified, such as the nucleolus, the speckles, the paraspeckles and even smaller structures like the transcription factories (Bond and Fox, 2009; Eskiw et al., 2008; Spector and Lamond, 2011). At a larger scale, chromosome territories and gene-positioning in association with transcriptional activity were described (Cremer et al., 2006; Croft et al., 1999). With the development of the chromosome conformation capture technology (Dekker et al., 2002) we have gained more insight into the higher-order organization of the chromosomes and it has been shown that chromatin has different levels of organization, ranging from the typical 10nm chromatin fiber to topologically associating domains (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012). Today, we start to have an understanding of the three-dimensionally (3D) organized genome that is extensively compartmentalized while allowing long-range interactions to occur for gene regulation (Dekker and Mirny, 2016).

A. Chromosome organization and transcription

Interphase chromosomes occupy discrete territories of the nuclear space, the so-called Chromosome Territories (CT) that have a specific positioning in the nucleus depending on their gene composition; the gene-rich chromosomes have a more central position compared to the gene-poor chromosomes that are located closer to the nuclear periphery (Bolzer et al., 2005; Croft et al., 1999). In the same manner, gene positioning is tightly correlated with the transcriptional activity. Heterochromatin tends to be located at the nuclear periphery in vicinity to the nuclear lamina, whereas the distribution of euchromatin localizes in the center of the nucleus or close to nuclear pores. The transition through different developmental stages during cell differentiation causes gene reposition, depending on their transcriptional activity (Kosak et al., 2007; Meister et al., 2010; Takizawa et al., 2008; Williams et al., 2006). To that direction, genes can move to the nuclear periphery once being silenced whereas upon activation they can move to the interior of the nucleus or loop out of their CTs to the interchromosomal space (Chambeyron et al., 2005; Kosak et al., 2007; Meister et al., 2010). The first Hi-C study performed in human cells molecularly confirmed the existence of chromosome territories. The defined spatial positioning of chromosomes was shown by obtaining far more frequent interactions between distant sequences located on the same chromosome, compared to any other loci in the rest of the genome (Lieberman-Aiden et al., 2009).

B. A and B compartments

The aforementioned study additionally identified the existence of two classes of genomic compartments, the first one being gene rich, transcriptionally active, and hypersensitive to DNase I digestion, while the second was relatively gene poor, transcriptionally silent, and DNase I insensitive (Lieberman-Aiden et al., 2009). This was highly resembling the EM-observed euchromatin and heterochromatin regions in interphase cells. These two major compartments are termed compartments A and B. The A compartment, similar to euchromatin, contains more open and active chromatin whereas B is more closed, compact, harboring repressive chromatin marks, similar to heterochromatin (Pueschel et al., 2016; Rao et al., 2014). A compartment is associated with histone marks such as H3K4me3, H3K36me3, and hyperacetylation while B compartment, on the other hand, is bound by Polycomb Group (PcG) proteins and heterochromatin proteins (HP1) thus associated with repressive marks (Nagano et al., 2013; Sexton et al., 2012). Inter-chromosomal contacts

between domains from the same compartments (A/A, B/B) are more frequent than those between different compartments (A/B), and A compartments make more contacts than B ones (Gibcus and Dekker, 2013; Lieberman-Aiden et al., 2009).

C. Topologically associating domains (TADs)

The A/B compartments are comprised of sub-megabase-scale domains which constitute the primary units of interphase chromosome folding and are termed topologically associating domains (TADs) (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012). The TADs are large self-associating domains of chromosomes ranging several hundreds of kilobases. TADs that show similar chromatin states, *i.e*. active or repressed, tend to associate with each other *in cis* and *in trans*, with TADs on the same or other chromosomes respectively, to form two genomic compartments (Dekker and Heard, 2015). They are constant throughout development and are largely conserved across different mammalian cell types. Genes within the same TAD share more similar regulation than genes in different TADs during embryonic stem cell differentiation (Nora et al., 2012) and reorganization of the genome architecture occurs at the sub-megabase scale (within TAD) during differentiation (Phillips-Cremins et al., 2013). TADs are demarcated by constitutive occupancy of CTCF (CCCTC-binding factor, an insulator protein that blocks communication between adjacent regulatory elements in an orientation-dependent manner), Cohesin complexes, transcription start sites of housekeeping genes, transfer RNAs, and short interspersed element (SINE) retrotransposons (Dixon et al., 2012; Nora et al., 2012). However, most of the CTCF-binding sites (around 85%) are actually found within the TADs, delimiting intra-TADs of an intermediate size of 100 kb–1 Mb. The spacing and orientation of CTCF-binding sites is responsible for the formation of individual loops or larger TADs was revealed by genome-wide CTCF ChIA-PET analysis (Chromatin Interaction Analysis by Paired-End Tag Sequencing) which also showed that gene regulatory interactions between promoters and their distal regulatory elements occur mostly within TADs (Tang et al., 2015).

CTCF and Cohesin co-occupied sites define and anchor the long-range interactions, *i.e*. the TADs and inter-TAD communications while Mediator and Cohesin co-binding establishes short-range, cell-type specific interactions within a TAD such as for enhancer-promoter interactions (Phillips-Cremins et al., 2013). Depletion of CTCF results in loop elimination between CTCF sites, and in disrupted insulation between neighboring TADs, surprisingly

though the genomic compartments are unaffected and no aberrant gene activation is observed (Kubo et al., 2017; Nora et al., 2017). On the same line, Cohesin depletion disrupts looping between CTCF sites and reduces intra-TAD interactions yet leaves compartmentalization intact (Rao et al., 2017; Sofueva et al., 2013). Therefore, compartmentalization of mammalian chromosomes emerges independently of proper insulation of TADs. Chromosome folding beyond the TAD scale is disrupted only when of both CTCF and Cohesin are depleted leading to general chromatin compaction (Tark-Dame et al., 2014). The inversion of an individual CTCF motif from a convergent looping pair abrogates the loop, underscoring the importance of the orientation for interactivity between remote pairs of CTCF sites (Guo et al., 2015; de Wit et al., 2015). In addition, inversion of clustered CTCF sites (at the protocadherin and β-globin loci) has been shown to disrupt local chromatin folding and allowed the inverted CTCF cluster to contact previously insulated regions downstream of the CTCF site (Guo et al., 2015). Collectively, TADs appear to represent functional domains with boundaries that do not allow enhancers to reach genes located in adjacent TADs (**Fig.1.1**).

Compartments and TADs are not a constitutive feature of chromosomes, as TADs are depleted along mitotic chromosomes (Naumova et al., 2013), and compartmentalization is lost at the inactive X chromosome in mammalians (Minajigi et al., 2015; Nora et al., 2012). This implies that there are mechanisms continuously instructing the chromosome organization for the proper 3D genome folding.

D. Nuclear lamina and Lamina-associated domains (LADs)

In mammalian cells, a network of intermediate filament proteins (of lamins and lamin-binding associated proteins) exists between the nuclear membrane and the chromatin (Gruenbaum et al., 2005). This structure, known as the nuclear lamina (NL), is implicated in a broad range of biological functions such as nuclear architecture, chromatin organization, and gene expression (Goldman et al., 2002). The development of genome-wide mapping techniques has made it possible to assess the molecular interactions between the chromatin and the NL. DamID is a genome-wide assay where NL proteins are fused to a DNA adenine methyltransferase (Dam) protein from the bacteria *Escherichia coli,* which will methylate any piece of DNA that is in molecular contact with the NL *in vivo* (van Steensel and Henikoff, 2000). Through the use of DamID, chromatin-NL interaction maps have been generated for mouse, and human cells (Guelen et al., 2008; Peric-Hupkes et al., 2010) and have revealed that

very large (median size of 500Kb) chromosomal domains engage in interactions with the NL. These domains are termed lamina-associated domains (LADs) and over a thousand of them exist in mouse and human cells.

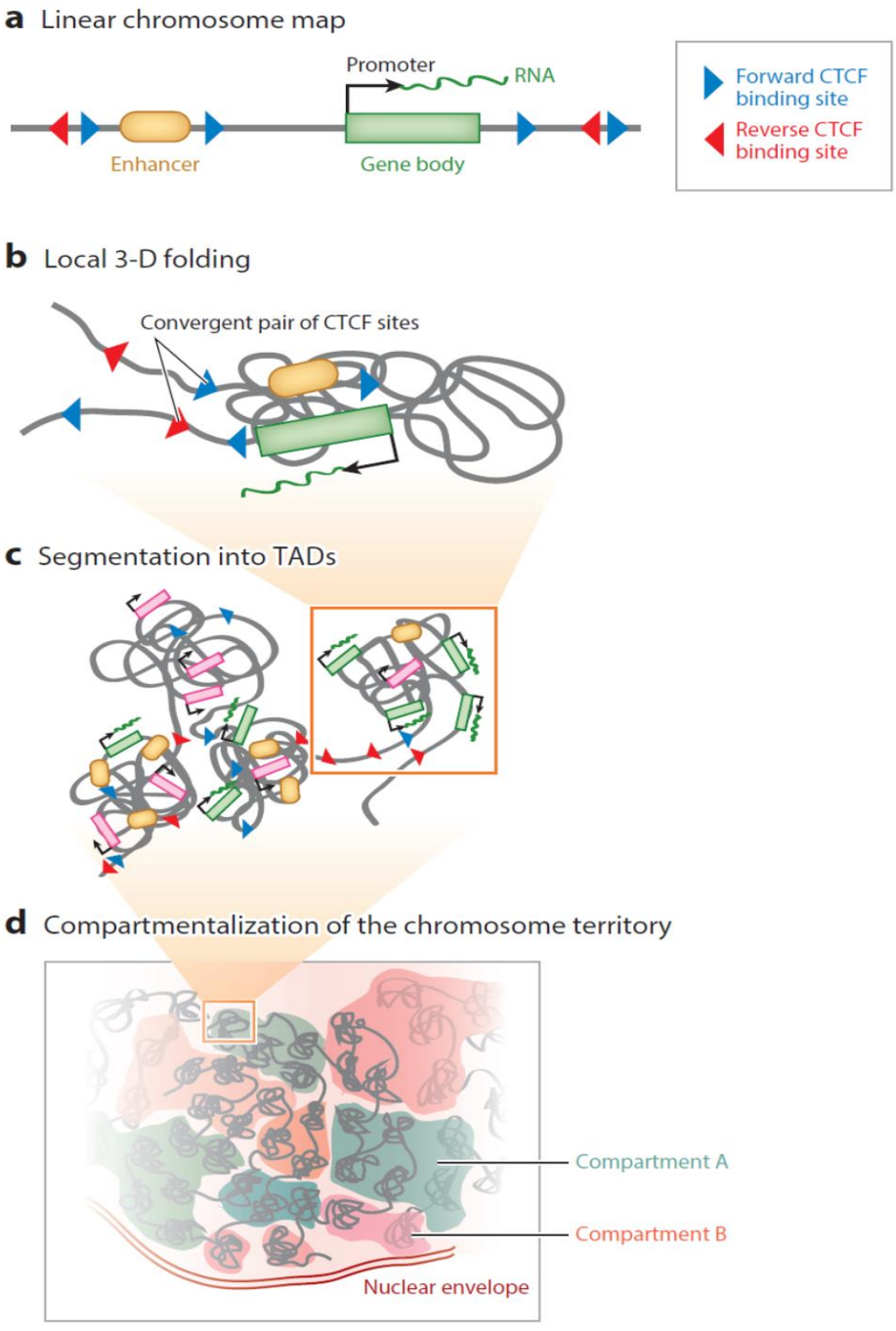


Figure 1.1. Illustration of the different levels of genome folding**.** a) Linear chromosome map. Representation of a genomic locus with an enhancer, a gene and its promoter, and CTCF binding sites in forward (blue) or reverse (red) orientation. b) Local 3-D folding. Convergent pairs of CTCF sites are brought into close spatial proximity forming a loop and enabling chromosomal contacts between the enhancer and the target promoter in the intervening domain. c) Segmentation into into topologically associating domains (TADs). This level of folding packages enhancers and promoters (resulting in transcriptional activation, in green) from the same domain together while insulating them from the

regulatory elements of neighboring domains. TADs also contain inactive genes (pink), which are not responsive to surrounding enhancers. d) Compartmentalization of the chromosome territory. The association of TADs from the same or different chromosomes defines two main compartments; A (blue) and B (red), which roughly correspond to the transcriptionally active and inactive fractions of the genome *(with permission from Merkenschlager and Nora, 2016)*.

LADs are typically gene-poor and transcriptionally inert, enriched in repressive (H3K9me3, H3K27me3) and devoid of active histone marks (H3K4me3, H3K36me3) (Kind et al., 2015). Surprisingly, recent studies have demonstrated that LADs associate also with euchromatin regions and are important for the epithelial-to-mesenchymal transition in mouse (Pascual-Reguant et al., 2018). Finally, lamins have been shown to have an important role for the global three-dimensional genome organization, as their loss has been shown to cause decompaction and detachment of some LADs from the NL in mouse embryonic stem cells (mESCs) which in turn disrupts the 3D chromatin interactions of LADs and inter-TADs (Zheng et al., 2018).

E. Nuclear bodies

The mammalian nucleus is further compartmentalized into membraneless subnuclear organelles which are specialized domains ("nuclear bodies") supporting distinct biological processes. They are defined mostly by the specific proteins and RNAs that they contain, at high local concentrations (Dundr, 2012; Dundr and Misteli, 2010; Mao et al., 2011). Some of the reported nuclear bodies with specialized functions are the nucleolus, the nuclear speckles, the paraspeckles, the Cajal and Promyelocytic (PML) bodies. The nucleolus is the largest nuclear structure where ribosomal RNAs are transcribed and ribosomes assembled (Boisvert et al., 2007). The nuclear speckles harbor the pre-mRNA splicing machinery and tend to localize near large clusters of active genes (Spector and Lamond, 2011). The paraspeckles are the domains were RNAs are sequestered for nuclear retention (Clemson et al., 2009). In the Cajal bodies the biogenesis and maturation of small nuclear RNA (snRNAs) takes place as well as the processing of histone mRNAs (Caudron-Herger and Rippe, 2012). In PML bodies divers regulatory proteins aggregate and these structures are required for heterochromatin integrity although their exact way of function is not yet deciphered (Pueschel et al., 2016). Many of these nuclear bodies are enriched in non-coding RNAs and some particularly in long non-coding RNAs (lncRNAs); the implication of lncRNAs in the formation of nuclear domains will be described more thoroughly in the **section II.B.**

Fluorescence recovery after photobleaching (FRAP) experiments have shown rapid and dynamic exchange of major protein components of nuclear bodies with the nucleoplasm,

suggesting an ordered assembly of these nuclear sub-organelles (Phair and Misteli, 2001). The formation and structural maintenance of nuclear bodies relies on the protein-protein and protein-RNA interactions (Dundr and Misteli, 2010; Mao et al., 2011). Very recent studies on liquid-liquid phase separation have proposed that different nuclear bodies behave like liquid-phase droplets that can condense through concentration-dependent phase separation. Nucleoli, paraspeckles and Cajal bodies form liquid-like condensates which are able to compartmentalize and concentrate proteins of similar biochemical properties and RNAs (Berry et al., 2015; Fox et al., 2018; Sawyer et al., 2018).

F. Nuclear matrix

Evidence for a non-chromatin scaffold within the nucleus first came to light half a century ago when electron microscopy (EM) and two-dimensional gel analysis revealed that the nucleus contains a large amount of non-chromatin insoluble protein and heterogeneous RNA resistant to extensive biochemical extraction (Berezney and Coffey, 1974; Capco et al., 1982; Herman et al., 1978). This fibrogranular structure revealed by DNAseI digestion and ammonium sulfate extraction, was termed nuclear matrix, and was proposed to form an architectural scaffold to support internal organization of the nucleus (He et al., 1990). Nuclear RNA was showed to be a key component of the nuclear matrix, since transcription inhibition or RNase treatment was causing the matrix fibers to collapse leaving what appeared to be largely hollow nuclei (as observed by EM) (Herman et al., 1978; Nickerson et al., 1989). After DNase digestion, further extraction with high salts of matrix proteins revealed a core filament network of RNA nature that was depleted upon RNase digestion (He et al., 1990). Different protocols have been used for the extraction of the components of the nuclear matrix, essentially differing in the salts and salt concentrations of the washes that follow the permeabilization with nonionic detergents and DNaseI digestion (Engelke et al., 2014).

The nuclear matrix was present in all the cells and tissues examined and was shown to organize the chromatin by attaching to the bases of DNA loops (Matrix/Scaffold attachment region, S/MAR) (Nickerson, 2001; Razin et al., 1981). Chromosome territories have been shown to be anchored to the nuclear matrix through the S/MARs and disruption of the matrix with RNase treatment results in the disruption of higher-order chromosome territory architecture (Ma et al., 1999). The S/MAR elements can be found inside genes and even inside exons. Some of the nuclear matrix proteins that preferentially bind to S/MAR elements are

lamins (Fiorini et al., 2006), SATB1 (special AT-rich binding protein 1) (de Belle et al., 1998), and SAFA/hnRNPU (Romig et al., 1992).

II. Long non-coding RNAs

A. Long non-coding RNAs identification and classification

Two-thirds of the mammalian genome has been shown to be pervasively transcribed, however only less than 2% is finally translated into proteins (Bertone et al., 2004; Carninci et al., 2005; Dinger et al., 2009; Djebali et al., 2012). During the last years, the advent of next generation sequencing techniques (NGS) has enabled the identification of thousands of non-coding genes that are subdivided into two categories: long and small non-coding RNAs. Long non-coding RNAs (lncRNAs) comprise a large family of transcripts larger than 200 nucleotides in size that often share many properties with the protein-coding mRNAs as being RNA polymerase II transcribed, spliced, capped and poly-adenylated (Quinn and Chang, 2016; Rinn and Chang, 2012). They mostly do not encode proteins, however some annotated lncRNAs have been reported to give rise to small peptides (Anderson et al., 2015; Cohen, 2014; Nelson et al., 2016). LncRNAs definition being based only on the length of the transcript and the lack of coding potential results in a broad heterogeneous family of molecules with highly diverse functional properties (Ulitsky and Bartel, 2013a).

The discovery of two long RNA molecules that had typical mRNA properties yet did not encode a protein, the first identified lncRNAs H19 and Xist, can be traced back to the early 1990's (Brannan et al., 1990; Brockdorff et al., 1992; Brown et al., 1991). Both of these lncRNAs were shown to be functional, H19 was involved in parental imprinting (Bartolomei et al., 1991; Gabory et al., 2010) while Xist was found to be orchestrating the inactivation of X chromosome, for dosage compensation in female and male mammals, by coating the inactive X chromosome from which it is transcribed (Brockdorff et al., 1992; Brown et al., 1991, 1992; Penny et al., 1996). However, it was more than a decade later that the identification of a huge number of lncRNAs was accomplished. High-throughput RNA sequencing (RNA-seq), chromatin immunoprecipitation sequencing (ChIP-seq) analysis and *ab initio* transcriptome reconstruction performed in multiple cell lines in the mouse and human, resulted in the identification of thousands of lncRNAs (Guttman et al., 2009, 2010a; Mikkelsen et al., 2007; Mortazavi et al., 2008; Sultan et al., 2008; Trapnell et al., 2009; Yassour et al., 2009). To date, only a few of this plethora of molecules have been characterized, while the vast majority of them remain largely unstudied.

LncRNAs are commonly expressed at lower levels than mRNAs, are slightly shorter and exhibit more tissue- or cell-type specific patterns of expression (Cabili et al., 2011; Dinger et al., 2009; Guttman et al., 2009; Mercer et al., 2008; Ulitsky and Bartel, 2013b). The sequence conservation is on average much lower for lncRNAs than for their coding counterparts (Kutter et al., 2012; Necsulea et al., 2014; Ulitsky and Bartel, 2013a). Evolutionary conservation of a lncRNA suggests often functional relevance (Chen et al., 2016; Ulitsky, 2016; Ulitsky et al., 2011). Nevertheless, although exonic sequences of lncRNA are not highly conserved, many studies have shown that numerous lncRNAs are localized in syntenic regions and exhibit a conserved location in respect to adjacent orthologous coding genes (Carninci et al., 2005; Dinger et al., 2008; Hezroni et al., 2015; Ulitsky et al., 2011). This finding suggests that lncRNAs might have a function independent of their sequence and this has shown to be the case for some lncRNAs with synteny conservation which regulate their neighboring protein-coding genes (Amaral et al., 2009; Bell et al., 2016; Wang et al., 2011). Another characteristic that can be independent of the primary sequence conservation is the formation of secondary structures, since some mutations can alter the primary sequence of an RNA but still preserve base pairing (Pegueroles and Gabaldón, 2016; Washietl et al., 2005). LncRNAs often fold into complex and thermodynamically stable secondary and tertiary structures which can be of crucial importance for their function (Mercer and Mattick, 2013; Zhang et al., 2010).

A functional prediction and classification method for lncRNAs is still lacking, however a categorization based on their relative position to neighboring coding-genes is used to group this large family of transcripts. Therefore, in respect to their location to a nearby coding-gene (**Fig.1.2**), lncRNAs can be named antisense when they are transcribed from the opposite direction to their coding counterpart or divergent lncRNAs when they originate from bivalent promoters that also control protein-coding genes (Katayama et al., 2005). Intronic lncRNAs are transcribed from introns within protein-coding genes and are transcribed to the same direction transcripts with intronic and/or exonic overlaps (Rinn and Chang, 2012). Finally, many lncRNAs are transcribed from loci devoid of protein-coding genes, at distance at least 5kb, and are named long intergenic noncoding RNAs, lincRNAs (or large intervening) (Guttman et al., 2009).

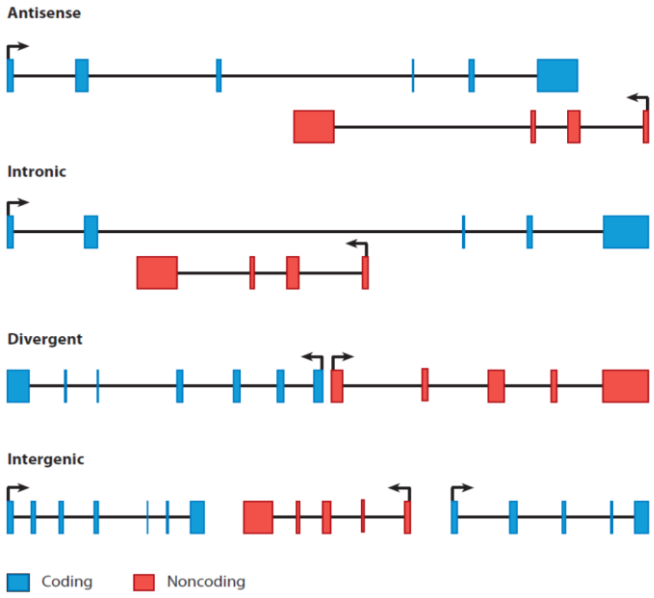


Figure 1.2. Classification of lncRNAs based on their position relative to a neighboring protein-coding gene. Antisense lncRNAs are transcribed in the opposite direction of protein-coding genes at its 3' extremity. Intronic lncRNAs initiate their transcription within introns of protein-coding genes. Divergent lncRNAs are transcribed from bivalent promoters at a reverse direction in respect to the protein-coding gene. Intergenic lncRNAs have distinct transcriptional units from protein-coding genes, they are transcribed from gene deserts, *i.e.* genomic locations between protein-coding genes of more than 5kb distance *(with permission from Rinn and Chang, 2012)*.

B. Functional mechanisms of lncRNAs

The number of studies describing the involvement of individual lncRNAs in diverse biological functions is ever growing, however, the functional relevance and the mechanisms of actions for the vast majority of them is largely unknown, and a functional classification is yet to be established. Computational analyses aiming at providing tools to improve our ability to predict the functionality of a given lncRNA just start to evolve (Kirk et al., 2018). One way of summarizing lncRNAs mode of action is to divide them based on their localization and function: (i) nuclear lncRNAs regulating gene expression in *cis*, (ii) nuclear lncRNAs acting in *trans* on distant genes, and (iii) regulatory lncRNAs acting in the cytoplasm. Few examples that demonstrate different functional activities exerted by lncRNAs follow.

Cis-acting lncRNAs

The Xist RNA is the paradigm of RNA-mediated regulation of transcription. It is one of the first and definitely the most studied lncRNA to date. At the onset of X-chromosome inactivation, Xist is transcribed from the future inactive X, spreads (in *cis*) across the entire chromosome and forms a subnuclear compartment devoid of active transcription marks and

enriched in repressive (Brockdorff et al., 1992; Brown et al., 1991; Chaumeil et al., 2006; Clemson et al., 1996). Xist RNA is indispensable for silencing, compaction and repositioning of the X-chromosome to the nuclear periphery (Chaumeil et al., 2006; Plath et al., 2002). The mechanism of action of Xist for the regulation of transcription is based on its interactions with chromatin regulatory complexes and its localization to the chromatin, through its interaction with nuclear matrix protein hnRNP U (SAFA) (Hasegawa et al., 2010; McHugh et al., 2015). Few recent studies have identified the proteins that interact with Xist RNA and revealed its multilayered repression activity. More specifically, it has been shown that Xist interacts with repressor proteins SHARP and SMRT that activate the histone deacetylase HDAC3 leading to the eviction of RNA polymerase II from the X chromosome (McHugh et al., 2015) but also that Xist is interacting with cohesins in order to repulse them from the inactive and establish a chromosomal architecture that disfavors transcription (Minajigi et al., 2015). In addition, Polycomb repressive complex 2 (PRC2) is recruited by Xist RNA (through debated direct or indirect interactions) for the deposition of the H3K27me3 repressive mark and transcriptional silencing (McHugh et al., 2015; Minajigi et al., 2015).

Another process where lncRNAs can regulate transcription in *cis* is genomic imprinting, when a certain gene is mono-allelically expressed in a parent-of-origin specific manner (Ferguson-Smith, 2011). It has been shown that a lncRNA is often transcribed on the opposite allele of the one that is producing the mRNA suggesting a repressive role for the lncRNA. Few notable examples are Air, Kcnq1ot1 and Nespas lncRNAs at the Igfr2, Kcnq and Gnas imprinted loci respectively (Nagano et al., 2008; Pandey et al., 2008; Williamson et al., 2011). The repression can be mediated through different mechanisms by the lncRNA; the act of transcription itself of the Air lncRNA has a repressive effect on the overlapping Igfr2 gene (Latos et al., 2012) while the recruitment of chromatin modifying proteins (G9a) by Air to the locus leads to the silencing of the Slc22a3 neighboring gene within the imprinted cluster (Nagano et al., 2008).

On the other hand, lncRNAs have been shown to activate transcription in *cis* by recruiting histone modifying complexes, as in the case of HOTTIP lncRNA that anchors the H3K4me3 methyltransferase MLL complex through its direct binding to WDR5 protein, in order to promote the transcriptional activation of the HOXA gene cluster (Wang et al., 2011).

Many enhancers are transcribed bidirectionally into molecules (eRNAs) that are unspliced, unstable and correlated with the expression of their neighboring genes (Kim et al., 2010a; Santa et al., 2010). Some enhancers are found in very close proximity to the promoters of lncRNAs and result in the unidirectional transcription of the enhancer-associated lncRNAs (termed ncRNA-activating, ncRNA-a), which are spliced, poly-adenylated, and stable transcripts (Lam et al., 2014; Ørom et al., 2010). These lncRNAs can activate the expression of neighboring genes (Li et al., 2013; Marques et al., 2013; Ørom et al., 2010) by facilitating enhancer-promoter interactions through their interaction with Mediator co-activator complex (Lai et al., 2013). Such an enhancer function has been shown to be the case for several lncRNAs (Fulco et al., 2016; Paralkar et al., 2016; Yin et al., 2015) and has been suggested to be the way of action for many lncRNA genes for transcription regulation (Bonasio and Shiekhattar, 2014). Very recently, it has been proposed that the splicing process of this class of lncRNA is crucial for their enhancer activity (Gil and Ulitsky, 2018; Tan et al., 2018).

Trans-acting lncRNAs

When the lncRNA transcripts are localized and have a functional role across the genome, they can act as scaffolds for the assembly of ribonucleoprotein complexes, as decoys for proteins to prevent their binding to their targets or as guides for proteins to mediate their localization at specific genomic loci (Rinn and Chang, 2012). The TERC lncRNA belongs to the first category as it is required for structural integrity of the telomerase complex by serving as a scaffold for the protein components of the complex (Zappulla and Cech, 2006). Gas5 lncRNA is induced upon growth factor starvation and interacts directly with the DNA-binding domain of the glucocorticoid receptor, thus acting as a decoy glucocorticoid response element and inhibiting glucocorticoid-regulated transcription in growth-arrested cells (Kino et al., 2010). The lncRNA HOTAIR (HOX transcript antisense intergenic RNA) acts as a guide for the recruitment of PRC2 complex to the HOXD gene cluster. HOTAIR is expressed from the HOXC locus and it interacts with the PRC2 complex, enabling its localization and induction of H3K27me3-mediated transcriptional repression of the HOXD locus which is located on a different chromosome (Rinn et al., 2007). LncRNAs can also accumulate in specific nuclear bodies and exert their functions there; such RNAs will be detailed below in section “lncRNAs and nuclear organization”.

lncRNAs acting in the cytoplasm

This class of lncRNAs need to be translocated to the cytoplasm in order to exert their biological function. In the cytoplasm the principles of their action are similar to those described for nuclear lncRNAs acting *in trans*, *i.e*. as scaffold and decoy for proteins. A lncRNA that acts as a decoy is the cytoplasm is NORAD which sequestrates Pumilio proteins and therefore prevents them from binding to their target mRNAs, effectively modulating their abundance (Lee et al., 2016; Tichon et al., 2016). linc-RoR acts as a microRNA sponge, modulating the concentration of miR-145. It shares miRNA-response elements with the pluripotency TFs Nanog, Oct4 and Sox2 and prevents these TFs from miRNA-mediated suppression in self-renewing human ES cells (Wang et al., 2013b)(Wang et al., 2013).

Collectively, the cases of several lncRNAs that have been studied point out to the fact that these genes can act on any level of gene regulation, summarized in **Fig.1.3**.

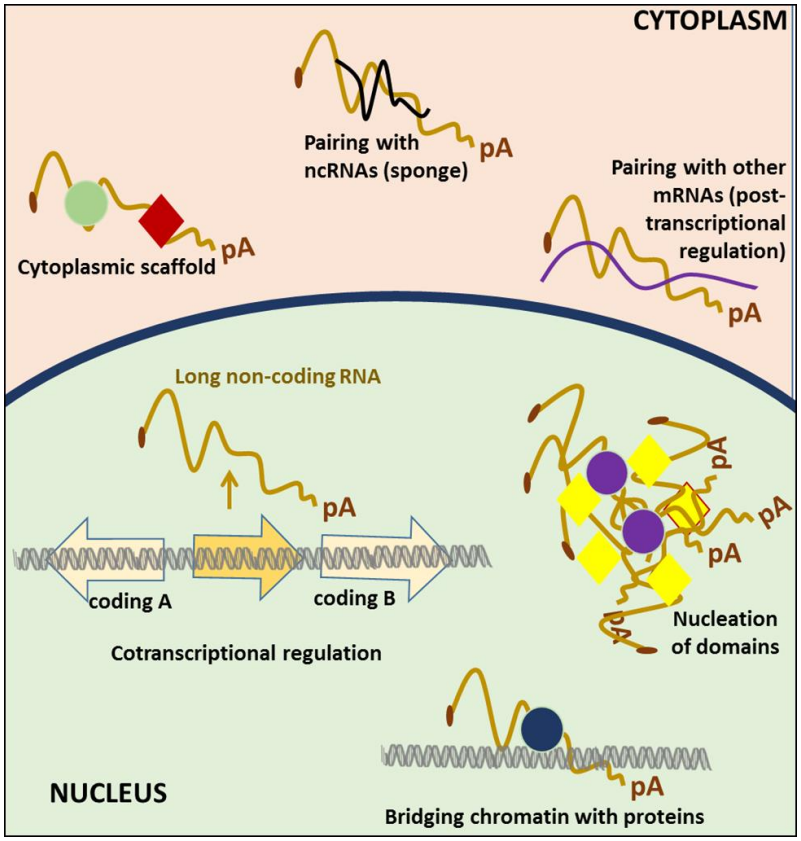


Figure 1.3. Diverse mechanisms of lncRNAs function. LncRNAs can act (i) on a co-transcriptional level *in cis* where either the transcript itself or the act of transcription is important, (ii) by recruitment of proteins or molecular complexes to specific genomic loci (*in cis* or *in trans*), (iii) by forming functional nuclear domains, (iv) serving as scaffold for ribonucleoprotein complexes, or (v) as protein decoy. Depicted as green, purple and blue circles as well as red and yellow rombs are different proteins.

LncRNAs and nuclear organization

LncRNAs have been shown to be involved in the dynamic nuclear organization as key player molecules, either by their participation in the formation of nuclear subcompartments or by the establishment of three-dimensional interactions of genomic loci for the regulation of gene expression (**Fig.1.4**). The nuclear, highly abundant lncRNAs Neat1 and Neat2/Malat1 are the most prominent cases of lncRNA that localize to and participate in the structural integrity of nuclear bodies (Hutchinson et al., 2007). Neat1 lncRNA scaffolds the formation of paraspeckles, nuclear bodies involved in the retention of mRNAs that undergo Adenosine-to-Inosine editing, possibly by relying on Neat1's continued transcription. Neat1 interacts with paraspeckles proteins and is required for the formation and stability of this nuclear compartment (Chen and Carmichael, 2009; Clemson et al., 2009; Mao et al., 2011; Sasaki et al., 2009; Sunwoo et al., 2009). Malat1 lncRNA localizes to nuclear speckles, a repository of transcription and splicing factors, and facilitates the proper localization of some protein components of the speckle (Bernard et al., 2010; Hutchinson et al., 2007). Malat1 is not necessary for the formation of speckles, in contrast to Neat1 RNA which is a structural component of the paraspeckles (Clemson et al., 2009). Gomafu (also known as Miat or Rncr2) is another example of lncRNA that participates in the formation of nuclear bodies (Sone et al., 2007a). It binds two splicing factors and modulates their function by sequestering these proteins into separate nuclear bodies (Ishizuka et al., 2014).

Other lncRNAs are necessary for the establishment of nuclear subcompartements locally silenced and compacted. These include Xist for the formation of the Barr body (the inactive X-chromosome), Kcnq1ot1 and Air for their imprinted loci, as previously detailed.

In addition, lncRNAs can participate in the three-dimensional structure of the genome through the establishment of long-range interactions in the nuclear space. The lncRNA Firre (Functional Intergenic Repeating RNA Element) has been demonstrated to form a domain localizing *in cis* over its own locus and *in trans* bringing five trans-chromosomal loci to proximity and orchestrating their gene expression (Hacisuleyman et al., 2014). Moreover, Firre is involved in anchoring the inactive X-chromosome to the perinucleolar region through the CTCF/cohesin complex (Yang et al., 2015). The lncRNA Charme also forms a domain in which it stabilizes long-range chromosomal interactions (*in cis*) on a region on its own chromosome and regulates the expression of the genes of that region (Ballarino et al., 2018). Many cases exist of chromosome looping or pairing where lncRNAs have been implicated to (Ma et al.,

2015; Wang et al., 2011; Yang et al., 2013). The association of lncRNA function with higher order chromatin structure by bringing distal sites into proximity have been verified by genome-wide chromosome conformation capture (Hi-C) data (Engreitz et al., 2013; Lieberman-Aiden et al., 2009).

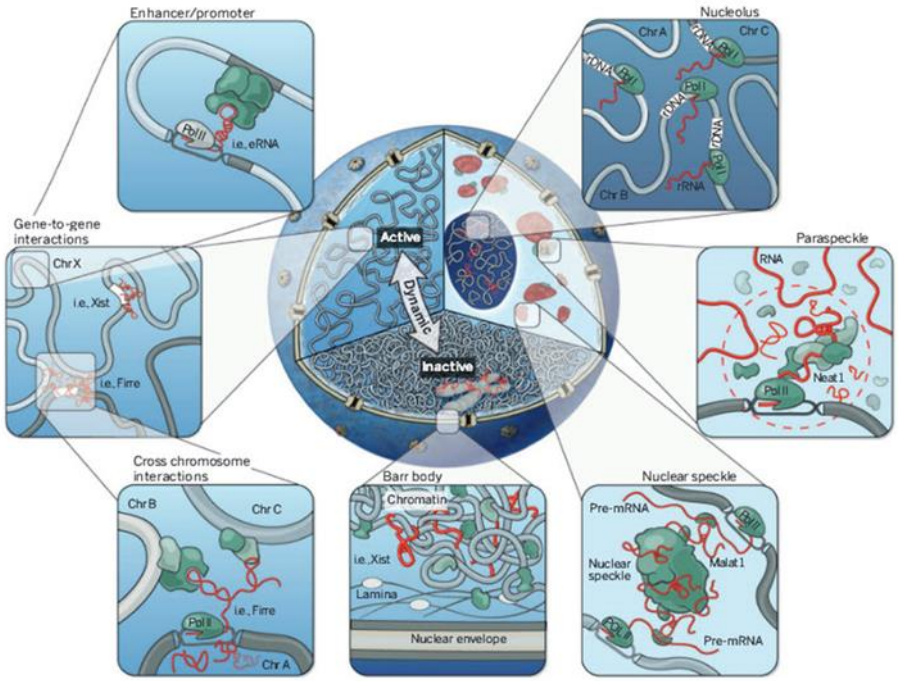


Figure 1.4. Cross-section of nucleus showing different organizational levels: active (A) and inactive (B) compartments, nuclear bodies and their associated ncRNAs, short- and long-range genomic interactions. Clockwise: nucleolus with the rRNAs, paraspeckle with Neat1 RNAs, nuclear speckles with Malat1 RNAs, the inactive X chromosome (Barr body) repositioned to B compartment and localized to the nuclear periphery, Firre RNA established long-range inter-chromosomal interactions, short-range interactions of enhancer-promoter elements *(with permission from Rinn and Guttman, 2014).*

LncRNAs and nuclear matrix

Many of the aforementioned lncRNAs, which actively participate in the nuclear organization, seem to share a common characteristic: their association with the nuclear matrix. Xist RNA resists the biochemical fractionation procedure of the removal of cellular DNA, protein and mRNAs, resulting in the insoluble nuclear matrix (Clemson et al., 1996). Additionally, Xist directly interacts with nuclear matrix hnRNPU/SAF-A protein (Hasegawa et al., 2010; McHugh et al., 2015) and requires the SATB1 nuclear matrix protein for the X-inactivation initiation (Agrelo et al., 2009). Therefore, Xist RNA seems to be strongly associated to the nuclear matrix. Furthermore, Firre RNA has been shown to interact with the hnRNPU/SAF-A protein and this interaction is necessary for the formation of the Firre domain

(Hacisuleyman et al., 2014), and Gomafu RNA has been shown to be fractionating with the insoluble nuclear matrix upon nuclear matrix preparation (Sone et al., 2007a). It is thus possible, that the association of a lncRNA with the nuclear matrix could be an indication that they might participate in higher-level chromatin organization.

III. Mouse Embryonic Stem Cells

The Embryonic Stem (ES) cells are isolated from the inner cell mass (ICM) of pre-implantation embryos at the blastocyst stage, and can be kept in culture under defined media. ES cells are characterized by two defining properties: self-renewal and pluripotency. These abilities to undergo unlimited cell divisions while maintaining their identity and to give rise to cells of all three germ layers (endoderm, mesoderm and ectoderm) as well as the germ lineage *in vitro* (Evans and Kaufman, 1981; Thomson et al., 1998) make them an interesting model system to study pluripotency and cellular differentiation. Mouse ES cells are able to give rise to teratocarcinomas when injected in adult compartments, and to produced chimeric animals when injected back in pre-implantation embryos contributing also to the germ line of the newborn (Bradley et al., 1984).

A. Extrinsic pathways regulating pluripotency

Mouse ES cells were originally cultured on feeder layers derived from mouse embryonic fibroblasts (MEF). Later it was found that Leukaemia Inhibitory Factor (LIF), a member of the Interleukin-6 cytokines produced by MEFs, was the key factor to maintain pluripotency of mouse ES cells by inhibiting their differentiation (Smith et al., 1988; Williams et al., 1988). ES cells require extrinsic growth factors to maintain their pluripotency in culture. These extrinsic modulators act on different signaling pathways to regulate intrinsic transcription factor networks to sustain ES cells in the undifferentiated state.

Binding of LIF to its receptor induces the activation of the JAK (Janus tyrosine kinase) pathway which subsequently can activate different signaling pathways: STAT3 (signal transducer and activator of transcription), Ras/ERK1/2 (extracellular-signal-related kinases 1/2) and PI3K (phosphoinositide 3-kinase). STAT3 is the effector of the self-renewal response (Han et al., 2013). In mouse ES cells, LIF can substitute MEF feeder layers in maintaining pluripotency in the presence of animal serum, by activating STAT3. However, in serum-free cultures, LIF is insufficient to block neural differentiation and maintain pluripotency. To that direction, it was found that BMP is able to replace serum short-term to maintain pluripotency of mouse ES cells in the presence of LIF, by activating inhibitors of differentiation (Id) genes, which block neural differentiation by promoting endo- and mesoderm differentiation (Ying et al., 2003).

FGF/MAPK (fibroblast growth factor/mitogen-activated protein kinases) signaling pathway triggers differentiation in mouse ES cells and its activation is antagonistic to self-renewal. The use of inhibitors of FGF receptor and ERK results in promotion of mouse ES cell pluripotency (Burdon et al., 1999; Kunath et al., 2007; Stavridis et al., 2007). Independence from the Erk pathway is a defining feature of mouse ES cells because it is a basic signaling module that is essential in many cell types.

In the canonical Wnt pathway, β-Catenin binds to Tcf3 in the nucleus and leads to the abolition of its repressing effects on stemness genes, by dissociating Tcf3 from its DNA-binding sites (Shy et al., 2013; Wu et al., 2012). Tcf3 has been shown to colocalize with many pluripotency factors on key regulatory regions to directly repress pluripotency factors as Nanog and is defined as an important factor to instruct early differentiation in mouse ES cells (Cole et al., 2008; Guo et al., 2011; Leeb et al., 2014; Martello et al., 2012; Pereira et al., 2006). Intracellular β-catenin is targeted for proteasome destruction due to phosphorylation by GSK3 (glycogen synthase kinase-3). Inhibition/deletion of GSK3 in presence of LIF allows efficient ES cell self-renewal, indicating GSK3 as its antagonist (Sato et al., 2004).

The chemical inhibition of Erk (by PD0325901) sustains robust ES cell self-renewal in the presence of LIF and chemical inhibition of GSK3β (by CHIR99021) mainly prevents β-catenin degradation. These two inhibitors constitute the 2i medium, in which spontaneous differentiation of mouse ES cells is abolished, LIF and serum stimulation become facultative, and the pluripotent state is considered as naïve, or ground state of pluripotnecy (Ying et al., 2008).

B. Transcription factor-mediated pluripotency regulation

The intrinsic regulators of pluripotency are forming a complex network of pluripotency factors. Pluripotency factors orchestrate the maintenance of the ES cell state and have been identified by relatively specific expression in ES cells and early embryos, and through genetic screens. The central pluripotency factor is Oct4 (a POU-domain transcription factor, POU5F1). Oct4 is expressed in oocytes, during the first cleavages of the embryo, then restricted to the inner cell mass of the blastocyst and later, in the post-implantation embryo, exclusively detected in the germ cell lineage. Its repression results in loss of pluripotency and dedifferentiation to trophectoderm lineage (Niwa et al., 2000). Interestingly, Oct4 overexpression does not reinforce pluripotency. Even a mild overexpression of Oct4 causes

differentiation into primitive endoderm and mesoderm (Niwa et al., 2000). Consequently, its level needs to be tightly regulated to sustain ES cell self-renewal since its up- or downregulation induce divergent developmental programs. Another essential TF for ES cell self-renewal is Sox2, and as for Oct4, its levels should be constrained for efficient self-renewal. Sox2 depletion results in trophoblast differentiation while overexpression might lead in neuroectoderm, mesoderm, and trophectoderm but not endoderm differentiation (Kopp et al., 2008; Masui et al., 2007; Zhao et al., 2004) Sox2 physically interacts with Oct4 protein and binds DNA together with Oct4 at Oct/Sox elements (Chen et al., 2008; Pardo et al., 2010). It acts synergistically with Oct4 to activate Oct-Sox enhancers, which regulate the expression of pluripotent stem cell-specific genes, including Oct4 and Sox2 themselves, and Nanog, the third core pluripotency TF. Although Nanog is not required for maintenance of pluripotency of ES cells (Chambers et al., 2007) it is necessary for the *in vivo* pluripotency to develop the ICM (Silva et al., 2009). These core TFs are highly interconnected and interdependent with one another and with others such as Esrrb, Klf4, Tfcp2l1, and Tbx3 forming the complex pluripotency transcription network (Chen et al., 2008; Marson et al., 2008).

C. Pluripotency states

Mouse ES cells are derived from the early blastocyst between E3.5 and E4.5 originating from the ICM and the pre-implantation epiblast, respectively. Mouse pluripotent cell lines could as well be derived from post-implantation epiblasts, called Epiblast stem cells (EpiSCs). EpiSCs show dependency to different signaling pathways, exhibit a distinct expression profile of pluripotent markers and a different epigenetic profile 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). 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 (Huang et al., 2012; Kojima et al., 2014). Therefore, ES cells and EpiSC constitute two pluripotency states, the former are naïve and the latter are primed pluripotent cells (Smith, 2017).

Reprograming of differentiated cells has been achieved through the artificial expression of pluripotency-associated TFs (reprogramming factors) (Takahashi and Yamanaka, 2006). The induced pluripotent stem cells (iPSCs) 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.

D. ES cells and nuclear organization

Self-renewal requires that the ES cell genome maintains a cellular memory that defines its pluripotent capacity. At the same time, the ES cell genome must exhibit high plasticity to be able to enter any distinct differentiation pathway. However, the architectural integrity of the nucleus is important for faithful genome function (Francastel et al., 2000; Lamond and Earnshaw, 1998). The gene repositioning through different developmental stages and during differentiation occurs during mitosis, when the nuclear-envelope breaks down and the condensation of chromosomes disrupts the organization of the nuclear architecture, and thereby allows repositioning of chromosomal regions (Walter et al., 2003). Many nuclear compartments with specific biological functions are subjected to massive rearrangement upon differentiation. The nuclear lamina, the nucleolus, heterochromatin structure and nuclear speckles are a few of them to undergo significant morphological changes comparing ESC to neuronal progenitor cells (NPC) (Meshorer and Misteli, 2006). Such changes include heterochromatin that is confined to fewer and larger foci in ESC, the nuclear speckles form smaller and more dispersed foci in ESC than in NPCs, while the ill-defined lamina of the ESCs becomes distinct and round in the NPCs.

Chromatin structure can influence gene function by affecting the accessibility of regulatory proteins to their target sites and by modulating the affinity of transcriptional regulators with their targets. In ES cells chromatin is globally decondensed, and as cells differentiate, regions of condensed heterochromatin are formed (Francastel et al., 2000; Melcer and Meshorer, 2010). ES cell chromatin is overall more active, and differentiation is accompanied by a transition to transcriptionally less-permissive chromatin by increase of H3K9me3 and decrease in H3, H4 acetylation (Meshorer et al., 2006).

Pluripotency transcription factors have been shown to have a role in the establishment of interactions that take place within a TAD (intra-TAD level). High levels of Oct4, Nanog, Sox2, Klf4, Esrrb are co-binding with Mediator coactivator (Med1) and RNA pol II at certain genomic hotspots, called super enhancers (SEs), which seem to control cell identity genes (Hnisz et al., 2013; Whyte et al., 2013). On the other hand, Sox2 enhancers form 3D-clusters that are segregated from heterochromatin but overlap with a subset of Pol II enriched regions. Such an enhancer clustering may increase the speed at which Sox2 finds its target sequences within individual clusters (Liu et al., 2014).

Pluripotent cells have been shown to contain almost entirely euchromatin, with highly mobile open chromatin and relative lack of nuclear structure (Dang-Nguyen and Torres-Padilla, 2015; Meshorer et al., 2006). In contrast, differentiated cells contain a heterochromatin footprint unique to their specific cell type that is visible by various imaging techniques; heterochromatin is compacted and relocated to the nuclear periphery (Hathaway et al., 2012). During differentiation, the organization and localization of chromocenters (heterochromatic clusters of centromeres from different chromosomes) changes and commonly they are observed on the nuclear periphery and the perinucleolar zone in differentiated cells (Mayer and Grummt, 2005; Wijchers et al., 2015). Pluripotent cells have smaller blocks of heterochromatin as imaged by microscopy and based on DAPI distribution, chromocentres are poorly compacted in ES and full iPS cells compared to MEFs (Ahmed et al., 2010; Fussner et al., 2011).

E. ES cells and lncRNAs

Many large-scale screens have been performed in ES cells in order to identify lncRNAs that are expressed in pluripotent cells and could be important for ES cell biology. Mostly loss-of-function techniques were applied and then the effect on pluripotency, differentiation or reprogramming properties of the mouse ES cells was monitored (Bergmann et al., 2015a; Bogu et al., 2016; Guttman, 2009; Guttman et al., 2010b, 2011a; Lv et al., 2015). Despite the high number of lncRNAs found to be specifically expressed in mouse ES cells, very few of them have been individually studied and functionally characterized so far. The few examples that have been studied demonstrated that lncRNAs can act through different mechanisms to exhibit a functional role relevant to ES cell biology. TUNA lncRNA (Tcl1 upstream neuron-associated lncRNA) has been shown important for the ES cells self-renewal and neural differentiation but also for reprogramming efficiency when overexpressed (Lin et al., 2014). Panct1 lncRNA was shown to associate with Tobf1 protein and affecting the recruitment of Oct4 at common gene targets (Chakraborty et al., 2017). The Linc-RoR (regulator of reprogramming) lncRNA, as mentioned before, acts as a sponge for miRNAs thus eliminating their negative effect on the core pluripotency TFs (Wang et al., 2013b). Last, the lincU lncRNA has been demonstrated to repress the ERK1/2 signaling pathway by stabilizing Dusp9 ERK-specific phosphatase (Jiapaer et al., 2018).

Thesis objectives

A number of studies have revealed the implication of lncRNAs in nuclear organization by forming functional domains in the nucleus. Xist, Neat1, and Firre lncRNAs are a few of the known functional lncRNAs that actively participate in the compartmentalization of the nucleus. To date, there is no robust way of identifying such molecules. The objective of my thesis is to be able to identify and characterize lncRNAs with a functional relevance for 3D genome organization.

Nuclear organization is tightly related to gene expression regulation and severely affected during differentiation of ES cells. Upon differentiation, mouse ES cells show massive genome architecture reconstruction. Combining the fields of lncRNA and ES cell biology under the prism of nuclear organization, the aim of my studies is to identify lncRNA genes that would play a role in the establishment of nuclear domains orchestrating in that way the necessary changes that need to occur for differentiation or maintenance of pluripotency.

Materials and methods

Cell culture

Culture media

FCS/LIF medium

Dulbecco's Modified Eagle Medium DMEM + GlutaMAX-I (Gibco, cat. 31966-021)

10% fetal calf serum FCS (Gibco, cat. 10270-098)

1X MEM non-essential amino acids (Gibco, cat. 1140- 035)

0.1 mM 2-mercaptoethanol (Gibco, cat. 31350-010)

10 ng/ml recombinant LIF (MILTENYI BIOTEC, 130-099-895)

2i/LIF medium

50% DMEM/F-12(1:1v/v, Gibco, cat. 31331-028)

50% Neurobasal (Gibco, cat. 21103-049)

0.5X N2 supplement (Gibco, cat. 17502-048)

0.5X B27 supplement (Gibco, cat. 17504-044).

1X L-Glutamine (Gibco, cat. 25030-024)

10µg/mL Insulin (Sigma I1882-100MG)

37.5µg/mL BSA (Sigma A3311-10G)

0.1 mM 2-mercaptoethanol (Gibco, cat. 31350-010)

1µM PD0325901 (Axon, cat. 1408)

3µM CHIR99021 (Axon, cat. 1386)

10 ng/ml recombinant LIF (MILTENYI BIOTEC, 130-099-895)

ES cell passaging

ES cells were cultured on plastic coated with 0.1% gelatin (SIGMA, cat. G1890-100G) in FCS/LIF media (or 2i/LIF when mentioned) and incubated at 37°C in 7% CO2. For the "matrix prep" (see below) cells were grown on gelatinized slides (Superfrost Plus, Thermo Fisher, cat. 4951PLUS4) placed in 15cm plates. Cells were passaged every 2-3 days, when they reached

70-80% confluence. Medium was changed every one or two days. Culture plates/flasks were treated with 0.1% gelatin in PBS 1X for 10 min before use. ES cells were washed with pre-warmed PBS and incubated with 1X trypsin-EDTA 0.05% (Thermo 25300062) at 37°C for 3 min. ES cells were quickly resuspended by pipetting up and down and a volume of DMEM/FCS medium equivalent to 5 times the volume of trypsin was added to block the reaction. Cells were transferred to a falcon tube and centrifuged for 5 min at 1000rpm. The pellet was resuspended in FCS/LIF (or 2i/LIF) and cells were split 1:5 or 1:10 at each passage.

Colony formation assay

After collecting the ES cells by trypsinization (as described for cell passaging), the cell pellets were resuspended in DMEM/FCS and counted. 600 cells were plated in a gelatinized well of a six well plate. Cells were cultured for 7 days in DMEM/FCS media with or without LIF and/or Doxycycline. Cells were washed with DMEM/FCS once and medium was replaced every day. Following, cells were washed in PBS and incubated for 45 sec in fixative solution (25ml of citrate solution, 8ml of formaldehyde solution and 65 ml of acetone). Fixed plates were washed in distilled water and stained for alkaline phosphatase activity using a leukocyte alkaline phosphatase kit (AP staining) (Sigma, cat. 86R-1KT). After a last water wash and an air drying step, the number of undifferentiated, mixed and differentiated colonies was assessed on a stereo-microscope (NIKON-SMZ1500).

Proliferation rate assay

0.3 million cells were counted and plated in appropriate medium in a well of a six-well plate. After 3 days, cells were trypsinized and counted to evaluate the total number of cells obtained. This procedure was repeated 4 times for each assessed cell line. Cell lines that were planned to be compared were always cultured in parallel to ensure comparable culture condition.

ES cell retinoic acid (RA) differentiation

Cells from FCS/LIF culture were counted and 10^5 cells were replated in 25cm2 flasks in DMEM/FCS media without LIF. 24h later, media was changed and retinoic acid (RA, final concentration 10^{-6}M) was added to DMEM/FCS. DMEM/FCS+RA medium was changed every day for 3 days. RNA samples were collected over the course of the assay with Day 0 being the plating point when cells come from +LIF medium and days 1, 2, 3 corresponding to RA treatment.

ES cell differentiation upon LIF withdrawal

Cells from FCS/LIF culture were counted and 0.3 x 10^6 cells/well were replated in six-well plates in two media conditions: DMEM/FCS with or without LIF. Culture lasted 48h with every day media change.

Embryoid bodies differentiation

Cells were washed once with PBS 1X and treated with pre-warmed trypsin. Partial dissociation of ES cells colonies was evaluated under the microscope and inactivation with a large volume of DMEM/FCS was done 1 or 2 min after trypsinization to allow small clumps of cells to be maintained. Cells were carefully recovered with a 10 mL pipette and transferred in a 50 mL falcon tube to avoid further dissociation. After few minutes, when the clumps progressively reached the bottom of the tube, as much as possible supernatant was gently removed without perturbing the accumulated clumps of cells. 10 mL of DMEM/FCS medium was gently added to the tube and the cells clumps were precociously resuspended and transferred into bacterial Petri dishes thus precluding cell adhesion (Day 0). Dishes were incubated at 37°C in 7% CO2. Medium was changed every day by carefully collecting the clumps of cells with a 10 mL pipette and replacing them into bacterial Petri Dishes in DMEM/FCS medium for 6 additional days and by splitting them, if necessary, into several dishes. At Day 6, the biggest embryoid bodies were collected by allowing the clumps to decant for a short time followed by the quick aspiration of the supernatant. They were subsequently replated on gelatinized surfaces at low density to allow for bodies adhesion and cell differentiation. One day later, adhesion of the embryoid bodies was checked under the microscope. Medium was changed every day and differentiating samples were collected at Day 6 and Day 10 for RNA extraction and gene expression analysis.

EpiLC differentiation

Cells cultured in FCS/LIF medium were adapted to 2i/LIF medium for 3 passages (9 days in total) before starting the differentiation protocol. The EpiLCs differentiation was induced by plating 0.23 million ES cells on a well of a 6-well plate coated with human plasma fibronectin (16.7 mg/ml) in N2B27 medium containing activin A (20 ng/ml), bFGF (12 ng/ml), and KSR (1%). Medium was changed every day until day 3 of differentiation. Cells were harvested along the assay (FCS/LIF, 2i/LIF 1st passage, 2i/LIF 3 passages, Day 1, 2 and 3 EpiLCs) for RNA extraction and gene expression analysis or in situ hybridization experiments.

Oct4 depletion assay

Zhbtc4 cells (Hitoshi Niwa, Miyazaki, and Smith, 2000) were used for Oct4 loss of function assay. This transgenic cell line was generated from the WT E14Tg2a line. Both endogenous loci of Oct4 gene have been invalidated and replaced by antibiotics resistance. In addition, two exogenous transgenes have been randomly integrated in the genome: one constitutively expressing the Doxycycline-repressed transcriptional activator (tTA) and another one harboring a tTA responsive promoter driving Oct4 cDNA expression. Therefore, upon Dox addition in the medium, the constitutive expression of Oct4 is quickly abolished to be already undetectable by 12 hours after treatment. 3 million cells were plated in FCS/LIF medium in a T75 flask and treated or not with Dox for 12, 24 of 48 hours and collected at the end of the treatment for RNA extraction and gene expression analysis.

Nanog depletion assay and Nanog KO cell line

44iN cells (Festuccia et al., 2012) were used for Nanog loss of function assay. This transgenic cell line was generated from the WT E14Tg2a line. Both endogenous loci of Nanog gene have been invalidated and replaced by antibiotics resistance. In addition, two exogenous transgenes have been randomly integrated in the genome: one constitutively expressing the Doxycycline-activated transcriptional activator (rtTA) and another one harboring an rtTA responsive promoter driving Nanog cDNA expression. Therefore, upon Dox withdrawal from the medium, the constitutive expression of Nanog is quickly abolished to be already undetectable by 12 hours after Dox removal. 3 million cells were plated in FCS/LIF medium in a T75 flask and treated or not with Dox for 12, 24 of 48 hours and collected at the end of the treatment for RNA extraction and gene expression analysis. To culture 44iN cells in the absence of Nanog long-term (Nanog KO cells), the cells were maintained under G418 selection, as previously described (Festuccia et al. 2012).

Nanog overexpression cell line

EF4 cells (Chambers et al., 2003) were used as Nanog overexpressing cells. This transgenic cell line was generated from the WT E14Tg2a line. An exogenous transgene have been randomly integrated in the genome harboring a Nanog cDNA cassette downstream of a CAG promoter leading to the stable and strong overexpression of Nanog. 1 million EF4 cells were plated in a T25 in FCS/LIF medium and lysed after 3 days of culture for RNA extraction and gene expression analysis.

Cytospin cells on glass slides for immunostaining (IF) or RNA/DNA FISH

After collecting the ES cells by trypsinization (as described for cell passaging), the cell pellets were resuspended in DMEM/PBS1X 1:1 at a concentration of 1million cells/mL. Cells were then cytospun (on Cytospin 2, Shandon Southern Products) at 400rpm with low acceleration mode for 4 min, onto glass superfrost slides and were subsequently subjected to fixation with 3% PFA (paraformaldehyde) for 10 min at room temperature followed by cold PBS washes. Next, cells were permeabilized on ice with 0.5% Triton X-100 in PBS for 5 min and washed twice in cold PBS. Slides could then be stored at 4°C in PBS for further use for IF or in ethanol 70% for RNA/DNA FISH experiment.

Extraction protocol: "matrix prep"

Cells that were used for the nuclear matrix extraction protocol were grown on glass slides until reaching ~80% confluency. For that aim, 10 million cells were plated in a B15 dish where 5 sterile glass superfrost slides were previously placed and gelatinized. When 80% confluency was reached, the slides were carefully removed from the dish with appropriate tweezers and further washed with PBS 1X at RT in coplin jars. Cells were first permeabilized with cold 0.5% Triton X-100 (SIGMA, cat.T8787) in PBS1X (for 5 min on ice). The slides were carefully washed twice with cold PBS on ice (for 2 and 1 min) and followed by DNase I (Qiagen, cat. 79254) treatment for 30 min at 37°C performed in a humid chamber. Slides were covered by a clean parafilm layer to ensure equal repartition of the DNase mix and minimize evaporation. After chromatin digestion, cells were subjected to another 0.5% Triton/PBS1X (for 3 min on ice) treatment to get rid of digested debris and washed twice again in 1XPBS (for 3 and 1 min on ice). Then cells were used either for RNA extraction by direct addition of Trizol reagent on the slides (within a 50mL falcon tube) or for immunostaining (IF), RNA or DNA fluorescence in situ hybridization (RNA/DNA FISH) by fixation with 3% cold PFA (for 10 min on ice) and subsequent PBS wash.

Transcription inhibition assay

1 million cells were plated in a single well of a 6-well plate and treated the next day with either Flavopiridol (400 nM) or Actinomycin D (5 µg/mL) for the indicated period of time following a 6-hour kinetics. All samples were harvested at the end of the assay for RNA extraction and gene expression analysis.

RNA isolation and Reverse Transcription (RT)

Total RNA from pelleted cells was isolated in Trizol according to manufacturer protocol (Thermo Fisher Scientific, cat. 15596-026). Resuspended RNAs were then treated with DNase I (Qiagen, cat. 79254) at 37°C for 30 min to ensure absence of genomic DNA contamination and re-purified with phenol/chloroform extraction, precipitated with ethanol and sodium acetate and resuspended in RNase-free water. The quality of the RNAs was systematically controlled by migration on 1% agarose gel, as well as the concentration, A260/A280 and A260/A280 absorbance determined by Nanodrop.

Reverse transcription reactions were performed on 200ng or 500ng of total RNA in a final volume of 20µL using a First Strand cDNA kit (Roche, cat. 04379012001) and 60µM random hexamers incubating at 25°C for 10 min, 50°C for 30 min or 2.5µM of anchored-oligo(dT)18 primers incubating at 55°C for 30 min. The amounts 200ng or 500ng of total RNA transcribed were then diluted in 500µL and 1ml of water, respectively.

Quantitative real-time PCR

Real-time PCR reactions were performed in duplicate for RNA expression analysis in 384-wells plates with a 480 LightCycler (Roche) using LightCycler 480 SYBR Green I Master (Roche, cat. 04877352001). Reaction volume was 10uL; cDNA was added 1:1 to the SYBR and primer mix (1µM). Primers were designed using the Primer3 online software (**Table 1**). Standard curves of all primers were performed to check for efficient amplification (above 90%), and all melting curves were generated to verify production of single DNA species with each primer pair. Values for gene expression were normalized to the levels of Tbp mRNA.

Target gene	**Sequence**
Tbp-F	GGGGAGCTGTGATGTGAAGT
Tbp-R	CCAGGAAATAATTCTGGCTCA
Nanog-F	AGGATGAAGTGCAAGCGGTG
Nanog-R	TGCTGAGCCCTTCTGAATCAG
Xist-f	GGTTCTCTCTCCAGAAGCTAGGAAAG
Xist-r	TGGTAGATGGCATTGTGTATTATATGG
Tsix3'-f	TGACCAGTACCTCGCAAGTTC
Tsix3'-r	CTAAGAGCACCTGGCTCCAC
NEAT1-f	GAAGATGCAGCAGTCGAACG
NEAT1-r	CAGGAGGCCATCGTTGAAGT
Rmrp-5'-f	CTTCTTGGCGGGCTAACAGT

Rmrp-5'-r	ACATGTTCCTTATCCTTTCGCC
dxz4-f	CACCGGAACTCATATGGAGA
dxz4-r	CCAGTCATCTGTCCAAATCA
Gm12690-intron-f	CCGTTTGCCTTCCCAGAGAT
Gm12690-intron-r	GCCAGAGGAGCCAGTGAATT
Vaultrc5-f	AGCTCAGCGGTTACTTCGAC
Vaultrc5-r	TCGAACCAAACACTCACGGG
Gm13067-exon-f	TGGAGGAGGATCGAGCAGG
Gm13067-exon-r	ATCTGTGCAGTTACCGAGGG
Kis2-f	AGCCACTCGGAAGGTCTCTA
Kis2-r	TAGGCCATCTGTGCGGTATG
Gm27000-f	GAATCCGTGCTCCTTGGCTA
Gm27000-r	CATCGGCTCACACCAGTCTT
Gm26917-f	TTGCATACCCTTCCCGTCTG
Gm26917-r	AGAAGCAGGCACCTAGGAGA
Gm26924-f	GCTTTTCTACGTTGGCTGGG
Gm26924-r	CTAGGTACCCGGGACAGGAG
Nespas-transexonic-f	CTACCTGGGTTGGCAGACAG
Nespas-transexonic-r	ATCCCTTGGGCTCATGATGG
Rpph1-f	GTGCCTCACCTCAGCCATT
Rpph1-r	AGGTGAGTTCCCAGAGAGCA
Gm26788-f	TGATGCCAAAGAAGCCACTGA
Gm26788-r	TTGTGAATGCTGCTGGACCT
Gigyf1-f	GGTGGATGAAGAGAGGCCTG
Gigyf1-r	GCTCTTTCTCCACTGCCTCA
Meg3-f	AACACGTTGCAACCCTCCT
Meg3-r	GGTGTCTGTGTCCGTGTGTC
Terc-f	CGCTGCAGGTCTGGACTTT
Terc-r	AGCTGTGGGTTCTGGTCTTT
Srrm2-f	CCAAGACGCAGAAGAAGTCC
Srrm2-r	TCTTGGTACGGGGAGAAGAA
Etl14a-f	AAAGCAGAAGTACCCCCACC
Etl14a-r	GGGACAGCTTGGGACTTTTT
Firre-5'-f	ACTGAAGATGAAGCCGGCAA
Firre-5'-r	GCCCATCGCCAATTTAACCC
Rere-f	CCGTCAGTCTGAAGTCCCAG
Rere-r	CCCGATTCTTGATGACTGGA
Ttn-exon-f	GCCTGCCATGTTTTCTGCAA
Ttn-exon-r	CTCTGGCAGACTGTGAGCAA
Nphs1as-f	GGCTCAGCTGTCAGTCCTAC
Nphs1as-r	CTGTCTCATCCGTGGGCAG
Rtel1-exon-f	GCCTGCTGTGAGTGACTACC
Rtel1-exon-r	TTGTATGCCCTCAGAGCTGC

Gm11611-f	ACTCTGGCTTCCACACTGTG
Gm11611-r	AGTGGGAAGGCTTGGAATGG
Gm15726-f	CGCTTCCTACGACACCATCA
Gm15726-r	GGAAGGGAATCAGCCGAGAG
Gm15247-f	GATCGTGCCCAACATCAAGC
Gm15247-r	AGGCCCCAATTGAGAACCTC
2900056M20Rik_1-f	GGACAGCCTCTTCAGTAGCC
2900056M20Rik_1-r	TGGAACACAGACTCCTCCGA
Gm11946-f	GTGGAGACGAGACAGGGTTG
Gm11946-r	TTCACGCACTCCCATCAGAC
Gm26542 int-f	TCTGCCTCCTCTCCCTACAC
Gm26542 int-r	GGAGGGAGGGCTTAGAGACA
L1orf2-f	GGAGGGACATTTCATTCTCATCA
L1orf2-r	GCTGCTCTTGTATTTGGAGCATAGA
MERVL_int-519-rv	CTAGAACCACTCCTGGTACCAAC
MERVL_LTR-365-fw	CTTCCATTCACAGCTGCGACTG
MMERVK10C_int_481-rv	TCGCTCRTGCCTGAAGATGTTTC
MMERVK10C_LTR_344-fw	TTCGCCTCTGCAATCAAGCTCTC
majSATrt-f	TGGAATATGGCGAGAAAACTG
majSATrt-r	AGGTCCTTCAGTGGGCATTT
nextC1_A-f	CCAGGGTTCCAACAGCTGAA
nextC1_A-r	GGCCCCTTTCCCATGCTAAT
nextC1_B-f	AACTGCAGAGCTCGTGACTC
nextC1_B-r	CACCCATCTCTCCCACTTCG
nextC1_ChIP2-f	GTGTGTACTGCAGGCCCTTA
nextC1_ChIP2-r	GGCCCCTCACTTCCTGAATAC
nextC1_ChIP1-f	CTGCACTTCACAGCTTGTCTT
nextC1_ChIP1-r	ATTCCCATCCGAGCTCAGTG
nextC1_D-f	CCATGAAGGCGCAGTACTGA
nextC1_D-r	AGACTCTCAGGGCTAGGCAA
nextC1_I-f	AGTTCTGTGTGCGGAGAGTG
nextC1_I-r	CAGGGAAGGATGACAGGCAG
nextC1_K-f	AGTATGTGCACACTCCAGCC
nextC1_K-r	CAGGCCAGTATGTTGCAGGA
nextC1_L-f	ATCACCACAGCCCATGACTG
nextC1_L-r	GGGCCAAGAGACGTGAATGA
nextC1_N-f	GTCAGGTAGGGCTAGGGACA
nextC1_N-r	TAAGAGCTGGGTGGGAAGGT
nextC1_O-f	ATTAGGCACCCCAAGCTGAC
nextC1_O-r	GTAAGGAACAGTGTGGCCGA
nextC1_P-f	GTGGAGGTGAGAGGAGGCTA
nextC1_P-r	ACAGAAACCCTTTGCCCCAA
nextC1_Q-f	ACAAATGGGGCCTTGTTCCA

nextC1_Q-r	CCACACCCTGCCACTGTATT
nextC1_R-f	AAGCCAGTCAGACGCATTGA
nextC1_R-r	TCTCTCACGACTGACCGACT
nextC1_S-f	CGAGTGCCTGAGTTGGAGTT
nextC1_S-r	AGCTCCCCTATCCTGTCGTT
nextC1_T-f	CTGTCCTTGCTCTCCAGGTG
nextC1_T-r	AGGGAAGAGGTGAGCTTGGA
nextC1_W-f	GCCATGTTGCTTTCCTTGGG
nextC1_W-r	AGGCAAGTGAGGGGACAATG
18S-f	GGATCCATTGGAGGGCAAGT
18S-r	CCCAAGATCCAACTACGAGCTT
28S-f	GAAGGCAAGATGGGTCACCA
28S-r	GAACTTCCGTGGGTGACTCC
Nanog premRNA F	GGTGATACGTTGGCCTTCTAGT
Nanog premRNA R	TTCTCAAATACACACAAGAGCCTTA
Malat1 F	CACACTGGCATGCTGGTCTA
Malat1 R	TACGGATGTGGTGGTGAAGC
Klf4-f	CAAGTCCCCTCTCTCCATTATCAAGAG
Klf4-r	CCACTACGTGGGATTTAAAAGTGCCTC
Klf4_int4-f	CCCGAAGACTAGTGGGGAAC
Klf4_int4-r	CTCTACAGCCTTCCGAGGTG
Tbra-F	GTGACTGCCTACCAGAATGA
Tbra-R	ATTGTCCGCATAGGTTGGAG
Esrrb-f	CGATTCATGAAATGCCTCAA
Esrrb-r	CCTCCTCGAACTCGGTCA
Fgf5-F	GTTTCCAGTGGAGCCCTTC
Fgf5-R	GAGACACAGCAAATATTTCCAAAA
Actc1 F	AGGGCGACGTAACACAGTTT
Actc1 R	ATCATGCGCCTGGATCTAGC
Dab2 F	TCTCAGCCTGCATCTTCTGA
Dab2 R	GAGCGAGGACAGAGGTCAAC
Fam53a F	AGCTGCCACTTGAGACCTTC
Fam53a R	GTGGCCATTTGTTCCCTTTGG
Maea F	TGAGTAGTTGCCCAGCTGTG
Maea R	CTACGTTTGCAGAGCTTGGC
Uvssa F	ACAGAGAAAGTGCAGACCGG
Uvssa R	AGGAGCATAGCCCTCTGTCA
NextC2 long spe F	GGACTGAGGGACACTGTCGT
NextC2 long spe R	TCAGAAGCCACACAGACTGG
NextC2 short spe F	TCTCTTCCCACTCACCGTCT
NextC2 short spe R	CAGGGCTCTGCATTTAGCTC
NextC2 common F	AAGTATGGAGCTAAATGCAGAGC
NextC2 common R	AATGTAAGGTTATAGTTTGGGGACA

nextC2_A-f	CTGTAGTGTGCCGTCCTGAA
nextC2_A-r	AGAATGGCTTCCATCCTCCT
nextC2_B-f	GGGGACAGCCTGTAAAACAA
nextC2_B-r	TCTCACACCCTGTCCTTTCC
nextC2_ex1-f	CCATCCAACACCATCTTTCC
nextC2_ex1-r	ACGACAGTGTCCCTCAGTCC
nextC2_int1-f	AGCATTGGATAAGGCCTGTG
nextC2_int1-r	TGTTCACTGGTTTCCAACGA
nextC2_E-f	TGCACACAAGAGCATGACAA
nextC2_E-r	CTTGGTGATCTCCCCTTGAA
nextC2_F-f	CTGGAAGGGTCTGGATTGAA
nextC2_F-r	CACTTGGTCTCCTGGCTCTC
nextC2_G-f	GCATGTGACCCTTTTGGACT
nextC2_G-r	GCTCTTCCTGAGGACCTGTG
nextC2_I-f	TGGTGGTGACTAGCAAGACG
nextC2_I-r	ATGATGTCGGAAAGCCACTC
nextC2_ex3-f	TGTCCCCAAACTATAACCTTACATTA
nextC2_ex3-r	AATGTAAGGTTATAGTTTGGGGACA
nextC2_J-f	GGTCAGGAGCTGAAGGACTG
nextC2_J-r	GCAAAATATGGCCTCTTGGA
nextC2_K-f	CAGCTTGGCTTGGAGGTTAG
nextC2_K-r	ACCACGTTGAGACACCTTCC
nextC2_L-f	CCGGTCATCAGACCAGTTTT
nextC2_L-r	AACCCAGGTGTTCCTGTCTG
nextC2_M-f	GTCGGTGTGGTTCCTGCTAT
nextC2_M-r	CTGCCACTGATAGACCAGCA
Rex1-f	CAGCTCCTGCACACAGAAGA
Rex1-r	ACTGATCCGCAAACACCTG
Sox2-f	CACAGATGCAACCGATGCA
Sox2-r	GGTGCCCTGCTGCGAGTA
Wnt3_F	CAAGCACAACAATGAAGCAGGC
Wnt3_R	TCGGGACTCACGGTGTTTCTC
Gata6_F	TGCAAGATTGCATCATGACAGA
Gata6_R	TGACCTCAGATCAGCCACGTTA
Sox17-f	CACAACGCAGAGCTAAGCAA
Sox17-r	CGCTTCTCTGCCAAGGTC
Oct4-f	CCCCAATGCCGTGAAGTTG
Oct4-r	TCAGCAGCTTGGCAAACTGTT
Hspb8-F	CCCTAAGGTCTGGCATGGTA
Hspb8-R	TTGGAGACAATCCCACCTTC
Otx2_F	AATCAACTTGCCAGAATCCAGGG
Otx2_R	GCTGTTGGCGGCACTTAGC
Pax6-F	AACAACCTGCCTATGCAACC

Pax6-R	CATAACTCCGCCCATTCACT
lsm6 F	GGATGAACGCATCTCCGTA
lsm6 R	GCAAACCCCTAGCGACTTCT
tmem184c F	ACGGGTATGGTTAGCAGCAG
tmem184c R	TGCCAAATTCTTGGTTTGGT

Table 1. List of used primer pairs

RNA-sequencing

RNA-sequencing was performed for three matrix preps of E14Tg2a (control-total and "next"). Before being sent for sequencing the RNA material was validated by RT-PCR for efficient depletion in the next samples. Libraries for RNA-Seq were prepared after ribosomal (rRNA) depletion at the genomic platform of the Institut Pasteur. Strand-specific, paired-end sequencing (100bp) of fragments of 300bp size, with an average depth of 30million reads per sample, was performed on HiSeq2500 instrument (Illumina). Reads were aligned to the mm9 mouse genome using Bowtie. Read counts were quantified and normalized per gene size and per million reads in SeqMonk. These quantifications were subjected to a DESEQ2 statistical filter with a FDR threshold of 5%.

All other RNA-sequencing were performed by paired-end poly(A) specific RNA-seq of 100bp on a HiSeq2500 instrument (Illumina) by Novogene company. Pre-validation of the RNA samples by RT-qPCR was routinely done as well as checking for RNA integrity.

RNA-sequencing data analysis

The computational analysis of the initial nuclear matrix RNA-seq datasets was conducted by Pablo Navarro. All the RNA-seq data for the NextC1 KO and SunTag clones as well as the NextC2 KO, SunTag and Krab clones were analyzed be Nick Owens, the computational biologist in our laboratory.

Karyotyping

Two millions of cells were plated in a T25 flask. The next day cells were arrested in metaphase with colcemid (4 h; 100 ng/ml−1; Gibco, 15212-012), trypsinized and resuspended in 10 ml of hypotonic solution (NaCitrate 0.017 M, KCl 0.03 M) for 10 min at room temperature. The cells were resuspended in 2.5 ml of ice-cold hypotonic solution. Five millilitres of cold fixative acetic acid–methanol 1:3 solution were added dropwise at 4 °C. Cells were collected and

resuspended in 2.5 ml of the supernatant and the above procedure was repeated three times. Fixed cells were spread by dropping on pre-heated glass slides and mounted (Vectashield; VectorLab, H1200) and imaged as for wide-field fluorescence.

Immuno-fluorescence

Slides covered with grown cells, processed or not according to the matrix prep protocol, were blocked in PBS1X/0,05%Tween20 (PBST)/1% donkey serum for 15 min at RT. Several antibodies were used; goat anti-Oct4 (Santa Cruz, sc-8628, dilution 1:500), rat anti-Nanog (eBioScience, cat. 14-5761-80), mouse anti-Esrrb (R&D, PP-H6705-00, dilution 1:500), goat anti-Sox2 (Santa Cruz, sc-17320, dilution 1:500), rabbit anti-Sall4 (abcam, ab29112, dilution 1:500), rabbit anti-Klf4 (Santa Cruz, sc-20691, dilution 1:500), anti-Pol II (Santa Cruz, sc-899, dilution 1:500) in blocking solution and the samples were incubated 1h at RT. After washing 3 times with PBST, samples were incubated with the appropriate secondary antibody (Alexa 488 or 594; Jackson ImmunoResearch) in 1:500 dilution for 45min at RT in a dark and humid chamber. 3 washes of PBST followed and the slides were then mounted with Vectashield containing DAPI (VectorLab, cat. H1200) and imaged using either 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 or 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.

RNA/DNA FISH

Hybridization and detection of nick translated probes was performed according to previously established protocols (Chaumeil et al., 2006). All DNA probes (1 ug/reaction) were nick-translated using the Vysis Nick Translation Kit (Abbott, cat. 32-801300). The Tsix/Xist probe is the “p510 Xist probe” that covers most of the Xist gene (provided by a colleague, Philippe Clerc). For NextC1 detection a probe was generated of a fosmid clone (from Children’s Hospital Oakland Research Institute, bacpac.chori.org) covering almost the entire NextC1 genomic locus and for Firre a fosmid was also used from the same source.

Slides prior to the hybridization were dehydrated through a series of ethanols - 5min in each of 70%, 90%, 100% EtOH and air dried. The probes containing 0.3 ng of individual probe, 3ul of

mouse Cot1 DNA (Invitrogen, cat 18440016) and salmon sperm DNA (Invitrogen, cat.15632011) were hybridized in 50% Formamide/2X Hybridization buffer at 37°C overnight.

After overnight hybridization of the probes, the slides were washed 3 times in 50%Formamide/2XSSC buffer at 40 °C for 5 min and 3 times in 2XSSC buffer at 40 °C for 5 min. Mounting with Vectashield containing DAPI followed and the imaging was done immediately after.

For DNA FISH or sequential RNA/DNA FISH a DNA denaturation step is required. The slides after the serial ethanol dehydration were treated with RNaseA 10U/ml (Invitrogen, cat. EN0531) in 2XSSC at 37°C for 1h to remove primary transcripts at the gene locus. Then, DNA denaturation was performed in 50% formamide/2XSSC at 80°C for 30min. slides were dehydrated in cold ethanol and left for overnight hybridization, same as for RNA FISH.

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). NextC2 probes were designed using Stellaris Probe Designer version 4.2 on Biosearch Technologies website with the maximum masking level (5) and were synthetized by the same company. Image stacks (0.3 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.

Images deconvolution

Deconvolution was performed using the iterative restoration function of Volocity 5.4.0 imposing a 99% confidence and 23 max iteration limit, using the appropriate calculated PSFs and considering a 0.11µm effective pixel size.

Cell fractionation

ES cells were trypsinized and counted and 10 million cells were lysed in 200µL of 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. The nuclei were isolated from the cytoplasmic fraction by centrifugation at 1300 g for 5 min at 4°C. The supernatant was collected and then re-centrifuged 5 min at 20 000 g at 4°C, 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 pellet was washed an additional time in 200µL of lysis buffer and RNA was extracted by TRIzol® addition following manufacturer's protocol (Adapted from X. Q. D. Wang and Dostie 2017).

sgRNA design and cloning

gRNAs were designed using the online CRISPR Design Tool (http://crispr.mit.edu/) (Hsu et al., 2013). Two oligonucleotides corresponding to the 20 bp of the sgRNA sequence preceded by the following overhangs were synthesized (**Table 2**):

5' – CACCNNNNNNNNNNNNNNNNNN – 3'

3' – NNNNNNNNNNNNNNNNNNNCAAA – 5'

gRNA	**Oligo sequence**
Nextc1 del 1 F	caccGGCCCTTAGTAGCCGTGGGG
Nextc1 del 1 R	aaacCCCCACGGCTACTAAGGGCC
Nextc1 del 2 F	caccGGTGTGGGGCCCGAAGGTTC
Nextc1 del 2 R	aaacGAACCTTCGGGCCCCACACC
gNextC1-KRABA-F	caccGCTGCAGGCCCTTAGTAGCCG

gNextC1-KRABA-R	aaacCGGCTACTAAGGGCCTGCAGC
gNextC1-KRABB-F	caccGCATGGGAGCAGCGGCGATCT
gNextC1-KRABB-R	aaacAGATCGCCGCTGCTCCCATGC
gNextC1-KRABC-F	caccGAACGCCACTGAGCTCGGAT
gNextC1-KRABC-R	aaacATCCGAGCTCAGTGGCGTTC
gNextC1-SunTag1-F	caccGGCCCTTAGTAGCCGTGGGG
gNextC1-SunTag1-R	aaacCCCCACGGCTACTAAGGGCC
gNextC1-SunTag2-F	caccGCAGGAATGCCTAGTAATCA
gNextC1-SunTag2-R	aaacTGATTACTAGGCATTCCTGC
gNextC1-SunTag3-F	caccGGTGTGGGGCCCGAAGGTTC
gNextC1-SunTag3-R	aaacGAACCTTCGGGCCCCACACC
NextC2 big del 1 F	caccGGGTAGTCTAGCATGGGCGT
NextC2 big del 1 R	aaacACGCCCATGCTAGACTACCC
NextC2 big del 2 F	caccGTCATTGCTTCTTCACGACGC
NextC2 big del 2 R	aaacGCGTCGTGAAGAAGCAATGAC
NextC2 small del 1 F	caccGAAGATCTTTGCCCGTCACC
NextC2 small del 1 R	aaacGGTGACGGGCAAAGATCAAG
gNextC2-long KRAB-F	caccGTTTATCCTAGACAGGGATTA
gNextC2-long KRAB-F	aaacTAATCCCTGTCTAGGATAAAC
gNextC2-long SunTag1-F	caccGGGTAGTCTAGCATGGGCGT
gNextC2-long SunTag1-R	aaacACGCCCATGCTAGACTACCC
gNextC2-long SunTag2-F	caccGTTTATCCTAGACAGGGATTA
gNextC2-long SunTag2-R	aaacTAATCCCTGTCTAGGATAAAC
gNextC2-short KRAB1-F	caccGCAGAAAGGCTGATCGCGGTT
gNextC2-short KRAB1-R	aaacAACCGCGATCAGCCTTTCTGC
gNextC2-short KRAB2-F	caccGTGAGGCAAGCCTGCCGTGT
gNextC2-short KRAB2-R	aaacACACGGCAGGCTTGCCTCAC
gNextC2-short KRAB3-F	caccGCAGGGCTGGCCTACACGGC
gNextC2-short KRAB3-R	aaacGCCGTGTAGGCCAGCCCTGC
gNextC2-short SunTag1-F	caccGAAGATCTTTGCCCGTCACC
gNextC2-short SunTag1-R	aaacGGTGACGGGCAAAGATCAAG
gNextC2-short SunTag2-F	caccGCAGAAAGGCTGATCGCGGT
gNextC2-short SunTag2-R	aaacACCGCGATCAGCCTTTCTGC
gNextC2-short SunTag3-F	caccGCAGAAAGGCTGATCGCGGTT
gNextC2-short SunTag3-R	aaacAACCGCGATCAGCCTTTCTGC
gNextC2-short SunTag4-F	caccGCCAGCGCCTCCCACGGTGG
gNextC2-short SunTag4-R	aaacCCACCGTGGGAGGCGCTGGC

Table 2. List of used sgRNA

1ug of gRNA-expression 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.

CRISPR/Cas9-mediated deletion of NextC1 promoter

Cloning of both gRNAs was performed in the pGL3-U6-sgRNA-PGK-puromycin vector obtained from Addgene (cat. 51133). E14Tg2a ES cells were collected by trypsinisation, counted and 106 cells/well were plated in a six-well plate in 2ml FCS/LIF and incubated for 1h. 1ug for each of the two gRNA plasmid and 2ug of the dCas9 plasmid (pCas9_GFP, addgene, cat. 44719) were added to 250ul of DMEM, mixed by flicking the tube and incubated at RT for 20 min. 4ul of Lipofectamine 2000 reagent (Thermo Fisher, cat. 11668027) (1uL per 1ug of transfected plasmid) was diluted in 250ul of DMEM and incubated for 5 min at RT. The DNA and Lipo2000 solutions were mixed and added dropwise to the cells in the plates. After 24 hours the medium was replaced by fresh medium containing puromycin (1ug/ml) for selection of the cells having received the gRNA plasmids. Cells were cultured for one additional day before sorting GFP-positive cells. ES cells transfected with the two gRNAs and Cas9-GFP plasmids were trypsinized and resuspended in FCS/LIF, filtered through a 40µm cell strainer and kept on ice. The highest GFP-positive ES cells were sorted as single cell per well in 4 gelatinized 96-well plates (containing FCS/LIF medium) using a FacsARIA III cell sorter (Becton-Dickinson), while keeping samples on ice. These single cells were let for growth until forming colonies for 14 days with routinely media change. The surviving clones of the 4 plates were passaged by splitting in half onto 2 new 96-well plates. After 4 additional days one plate was used for DNA extraction/screening and the other for freezing down. The freezing of the plate was executed by trypsinizing the cells and adding directly to the plate FCS and DMSO, to final concentration 10%DMSO/FCS. The plate was then sealed with tissues and stored at -80°C.

For the PCR screen of the deletion amplicon, DNA was extracted directly in the plate, following the protocol described in Wettstein et al., 2016. Primers internal and surrounding the

deletion were used to assess the deleted or WT status of each clone. The expected size of the depletion was of 450bp. PCR reaction was performed in 96-well plates with a Taq DNA Polymerase (Thermo Fisher, cat. #EP0402). 200ng of the extracted DNA was mixed for a reaction following manufacturer's instructions. The PCR conditions were: (i) 95 °C for 3min, (ii) 95 °C for 30 s, (iii) 60 °C for 30 s, (iv) 72 °C for required extension, (v) 72 °C for 5 min and (vi) 4 °C final, with 30 cycles repeating steps (ii)–(iv). The products were run on 2% agarose gel.

The deletion bands were gel-purified using a PCR clean-up Gel extraction kit (MACHEREY-NAGEL, cat. 740609) following manufacturer's protocol. The PCR products were then cloned and bacteria were transformed using the Zero Blunt TOPO PCR Cloning Kit (Thermo Fisher, cat. K2800-J10); the steps outlined in the manual were followed. The transformation plates were incubated overnight at 37 °C, single colonies were picked and the vector they contained was sequenced (10 bacterial clones per "KO" clone) through conventional Sanger sequencing.

CRISPR/Cas9-mediated deletion of NextC2 isoforms

One of the 2 gRNAs was cloned in the pGL3-U6-sgRNA-PGK-puromycin vector obtained from Addgene (cat. 51133) and the other one in the pU6-(BbsI)_CBh-Cas9-T2A-mCherry (Addgene #64324). 1 million E14Tg2a WT cells were plated in a 6-well plate at Day 0. Meanwhile, 1 µg of the Cas9-mCherry-gRNA vector and 3 µg of the gRNA-puro plasmid were pre-mixed 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) for 30 min at room temperature to allow for complexes formation. Finally, complexes were added to the culture medium to allow for cell delivery. Puromycin (1 µg/mL) selection was performed from Day 2 to Day 5 and homogenous apparition of mCherry expression was checked under an epifluorescence microscope. Transiently selected cells for Puromycin resistance were further plated at clonal density at Day 5, and single clones were picked 10 days later. After expansion and freezing of the isolated clones, genomic DNA was isolated with NucleoSpin Tissue DNA extraction Kit (Macherey-Nagel, 740952.50), and screened by qPCR and PCR for proper genomic deletion. PCR was performed with LongAmp Taq PCR kit (BioLabs, E5200S) following manufactory's instructions. PCR products were run on an agarose gel, purified thanks to NucleoSpin Gel and PCR Clean-Up (Macherey-Nagel, 740609.50). Only homozygous clones for both big and small deletions were kept for further experiments.

Inducible Knock-down assay - dCas9-KRAB

PiggyBac vectors containing an rtTA trans-activator (PB-CAG-rtTA) expression cassette, the Dox-inducible dCas9-KRAB-BFP cassette, and the PiggyBac integrase vector were kindly provided by Dr. Pentao Liu (Gao et al., 2013; Gao et al., 2014). The PiggyBac expressing the gRNA and a Puromycin resistance was synthetized in the laboratory (Heurtier et al., 2018). 1 million E14Tg2a WT ES cells were plated at Day 0 in a P6 well. The next day, 0.5 µg of the gRNA-puro and PBAse vectors and 2µg of the rtTA and dCas9-KRAB vectors were co-transfected using 5µL of Lipofectamine 2000 (Invitrogen, 11668-019). At Day 2, the cells were treated with Puromycin (1µg/mL) for two additional days before the cells were trypsinized and plated at clonal density. 10 days later single clones were picked and expanded. In parallel, each clone was plated apart in presence of Dox to check for dCas9-KRAB-BFP induction under the microscope. Only BFP inducing clones under Dox treatment were kept for further gene expression analysis.

Inducible induction assay - CRISPRa SunTag

The CRISPRa SunTag system vectors were obtained from Addgene Company, modified in the laboratory to be Dox-inducible and inserted in E14Tg2a WT ES cells through PiggyBac vectors integration (Heurtier et al., 2018). Two independent clones were generated. 1 million cells of both clones were plated in a 6-well plate at D0. Meanwhile, 1 µg of PiggyBac transposase coding plasmid and 1 µg of PiggyBac plasmid containing the sgRNA with a Puromycin cassette were pre-mixed in 250 µL of DMEM Medium 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) for 30 min at room temperature to allow complexes formation. Finally, complexes were added to the culture medium to allow cell delivery. At Day 2, the cells were treated with Puromycin (1µg/mL) for two additional days before the cells were trypsinized and plated at clonal density. 10 days later single clones were picked and expanded. In parallel, each clone was plated apart in presence of Dox to check for BFP and GFP (linked to dCas9 and VP64 parts) induction under the microscope. Only BFP/GFP inducing clones under Dox treatment were kept for further gene expression analysis.

Results

I. Identification of structural long non-coding RNAs

A. Establishment of the experimental approach

Unlike proteins that can be classified based on their amino acid composition and presence of specific domains like RNA-binding, zing finger or chromo domains, and whose function might thus be predicted, a functional classification is not yet established in the field of long non-coding RNAs than often lack linear sequence homology. Although thousands of lncRNAs have been identified to be transcribed from the genome of many mammalian species, the diverse mechanisms of their action are far from being fully understood. Few examples of functional lncRNAs have been shown to participate in gene silencing, stem cell pluripotency or topological organization of subnuclear regions for instance (Bergmann et al., 2015b; Guttman et al., 2011b; Hacisuleyman et al., 2014; Rinn et al., 2007), but a general way to link given characteristics to function has not yet been attained.

The implication of lncRNAs in nuclear organization by establishing and maintaining nuclear compartmentalization is well established nowadays (Rinn and Guttman, 2014). However, a robust approach for the identification of molecules such as Xist, Neat1, Firre that shape the nucleus is currently lacking. Interestingly, previous findings have shown that few of the known "nuclear organizers" lncRNAs share the biochemical property of being resistant to the so called nuclear matrix fractionation or are associated with a nuclear matrix protein component. The Xist RNA territory has been shown to remain intact after nuclear matrix preparation (Clemson et al., 1996) and Gomafu RNA is highly insoluble, remaining unperturbed after removal of most chromatin (Sone et al., 2007b). Firre and Xist RNAs have been demonstrated to be interacting with the major nuclear matrix component hnRNPU/SAF-A protein (Hacisuleyman et al., 2014; Hasegawa et al., 2010). Since lncRNAs that participate in the functional shaping of the nuclear space interact with or are part of the nuclear matrix, we reasoned that in order to identify novel such lncRNAs, we would have to focus on those that physically associate to the nuclear matrix. Therefore, we decided to build on that observation an experimental approach that would allow us to identify such "structural" lncRNAs on a genome-scale level. For this reason, we performed nuclear matrix fractionation on cells, expecting most of the DNA, soluble proteins and RNAs to be removed and the remaining RNAs - that will be identified by RNA-sequencing – to be meaningful candidates of functional RNAs that participate in the nuclear organization (see workflow in **Fig.1.1**).

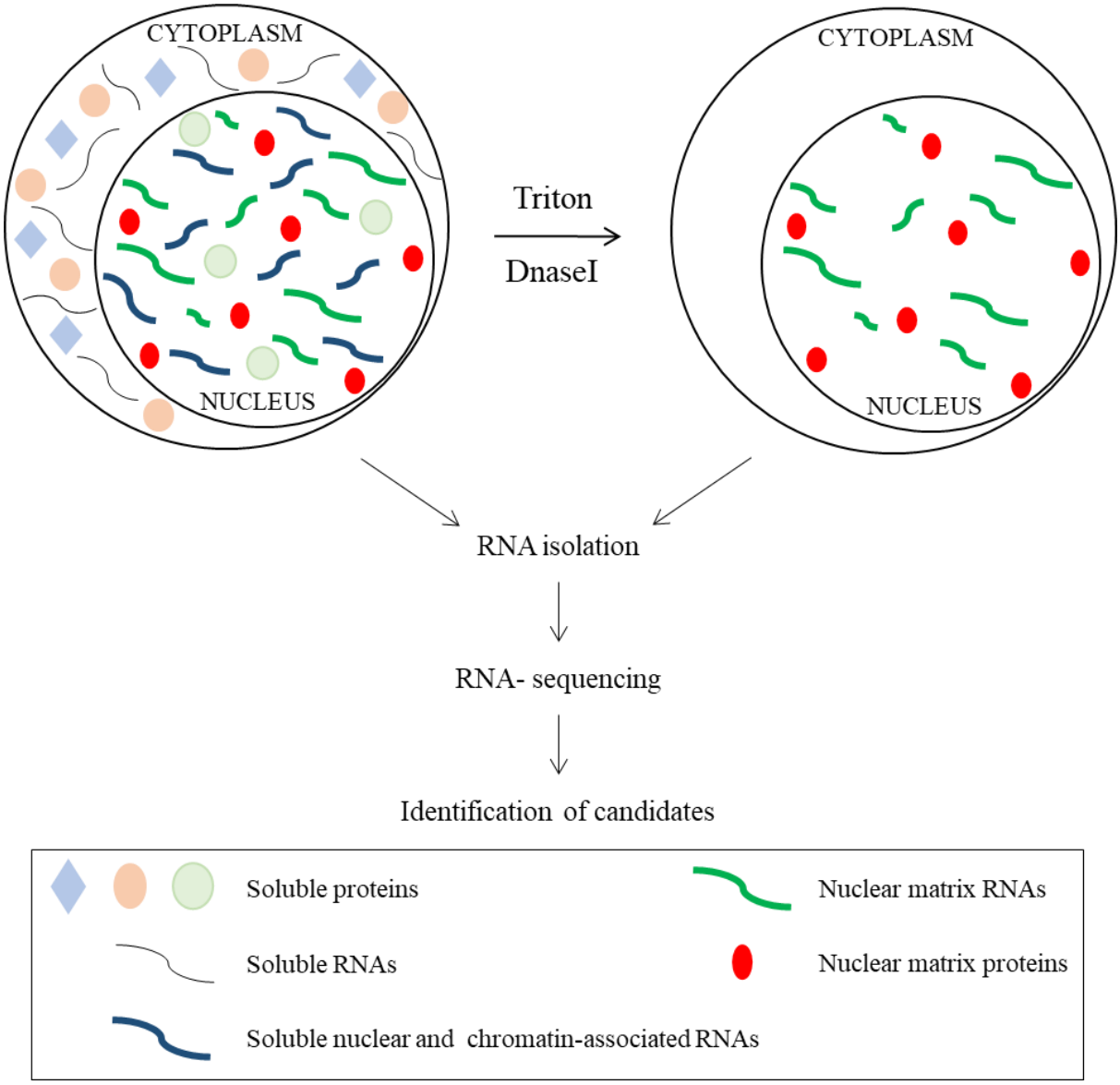


Figure 1.1. Experimental design of Matrix Preparation. A simplified nuclear matrix fractionation protocol is applied on cells using Triton X-100 and DNase I treatments. After removal of cytoplasmic and nuclear soluble proteins and RNAs, RNA isolation follows from control (non-treated) cells and matrix (treated) samples. RNA-sequencing then is conducted in order to identify a set of candidates that would be possibly participating in functionally structuring the nucleus.

A variety of protocols exist for a nuclear matrix fractionation resulting in a slightly different degrees of preservation of its underlying structure. The general approach has three steps that use (i) Triton-X-100 detergent to remove membranes and soluble proteins, 2) DNAse I digestion to fragment the DNA, and 3) hypertonic salt washes to remove the digested chromatin (>90% of DNA and >86% histones are removed) (Capco et al., 1982; He et al., 1990; Nickerson, 2001). This process leaves behind ribosomal precursor, heterogeneous nuclear

RNA and insoluble proteins (Berezney and Jeon, 1995; Capco et al., 1982; Herman et al., 1978). In our case, we use a comparable extraction procedure (extended in Material and Methods section) – thereafter matrix prep – where the last hypertonic buffer wash has been replaced by a second detergent washing step. Mouse Embryonic Stem (ES) cells grown on glass slides were subjected to Triton X-100 treatment for removal of diffusible molecules, DNase I digestion for removal of DNA and an additional Triton treatment to wash away the released chromatin-associated proteins and RNAs, nuclear debris and chromatin clots. The protein and RNA material which was left after our matrix preps is termed "matrix fraction" (**Fig. 1.2**)

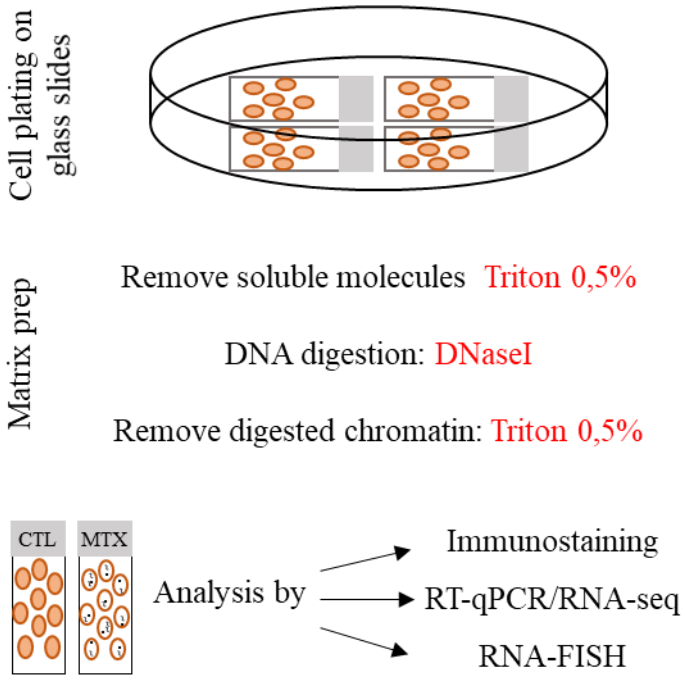


Figure 1.2. Matrix prep and following analysis. Glass slide-grown mES cells are subjected to a series of Triton and DNaseI treatments. Control and Matrix samples are then analyzed in parallel for protein and RNA content.

At first, in order to validate our ability to successfully perform nuclear matrix fractionation, we investigated the level of chromatin digestion and the removal of soluble proteins by bright-field microscopy and immunostaining. As shown in **Fig. 1.3**, we could see by phase-contrast microscopy that the extracted nuclei are brighter when compared to control nuclei, most likely due to the digestion and removal of chromatin along with the wash-out of diffusible molecules. The darker structures which are visible in control and much more

prominent in the extracted nuclei clearly corresponds to nucleoli which have been shown not to be extracted upon matrix preparation protocol (Berezney and Coffey, 1974). In addition, to evaluate the efficiency of the removal of diffusible molecules, we used two pluripotency transcription factors (TF), Nanog and Oct4, as specific examples of soluble proteins and compared their abundance in our control and matrix prep samples. Both proteins were effectively eliminated after our biochemical treatment. Few other diffusible proteins were verified by immunofluorescence to be efficiently extracted, like Sox2, Esrrb, Klf4, Sall4 and RNA polymerase II (data not shown).

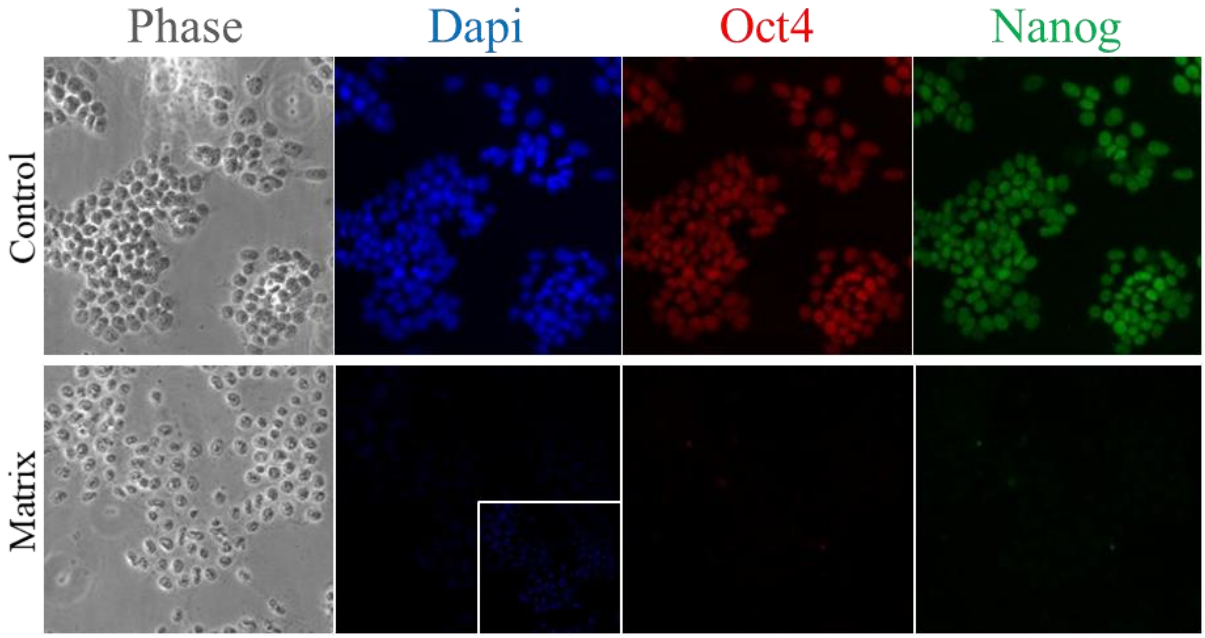


Figure 1.3. Extraction validation by phase-contrast and fluorescence microscopy. In the upper panel, left to right, control cells are imaged by phase contrast, DNA is stained by DAPI (blue), and stainings are shown for Oct4 (red) and Nanog (green). The lower panel shows the same imaging in extracted (matrix) cells, where nuclei are devoid of chromatin, Oct4 and Nanog proteins. Exposure times are equal for control and matrix images. The inset image of DAPI channel of the matrix sample corresponds to an overexposed image in order to visualize the remaining very faint DAPI staining.

Conversely, we investigated the level of nuclear matrix preservation within our matrix prep samples by looking into the detectability of proteins previously shown to be enriched within this nuclear structure. As the nuclear lamina was the first factor to be identified as a constitutive component of the nuclear matrix (Berezney and Coffey, 1974; Herman et al., 1978), we thus stained for a key component of the nuclear lamina, LaminB1, in our control and matrix samples (**Fig.1.4A**). In order to estimate histone depletion efficiency, we additionally stained for the histone 3 lysine 4 trimethylation mark (H3K4me3) (**Fig.1.4A**). We readily visualized that LaminB1 was perfectly retained in the matrix prep samples in contrast to the

histone modification mark that was strongly depleted. However, we could still detect some weak signal for H3K4me3 that focalized in puncta, reminiscent of its localization in the control sample, suggesting that few histone molecules might be still retained after our matrix prep. This could be explained by the slightly modified protocol we use where the common high-salt extraction step is replaced by a Triton wash, likely resulting in a milder histone extractability. SAFA has been demonstrated to be a major component of the nuclear matrix (Romig et al., 1992) and speckles have been shown to be preserved in matrix prep protocols (Mintz et al., 1999). For that reason, we then performed a staining for SAFA protein and a splicing speckle protein, SC35. We observed that SAFA was detectable when DNA is efficiently removed in the matrix to a slightly lower extent compared to control sample (**Fig.1.4B**). SC35 seemed to be moderately washed away, yet still detectable in all cells of the extracted sample.

We therefore validated the efficiency of our established nuclear matrix prep in regards to the removal of DNA and soluble proteins as well as satisfactory nuclear matrix preservation. Further, we proceeded to RNA isolation from control (whole cell RNA recovered) and matrix cells to subsequently perform RNA-sequencing.

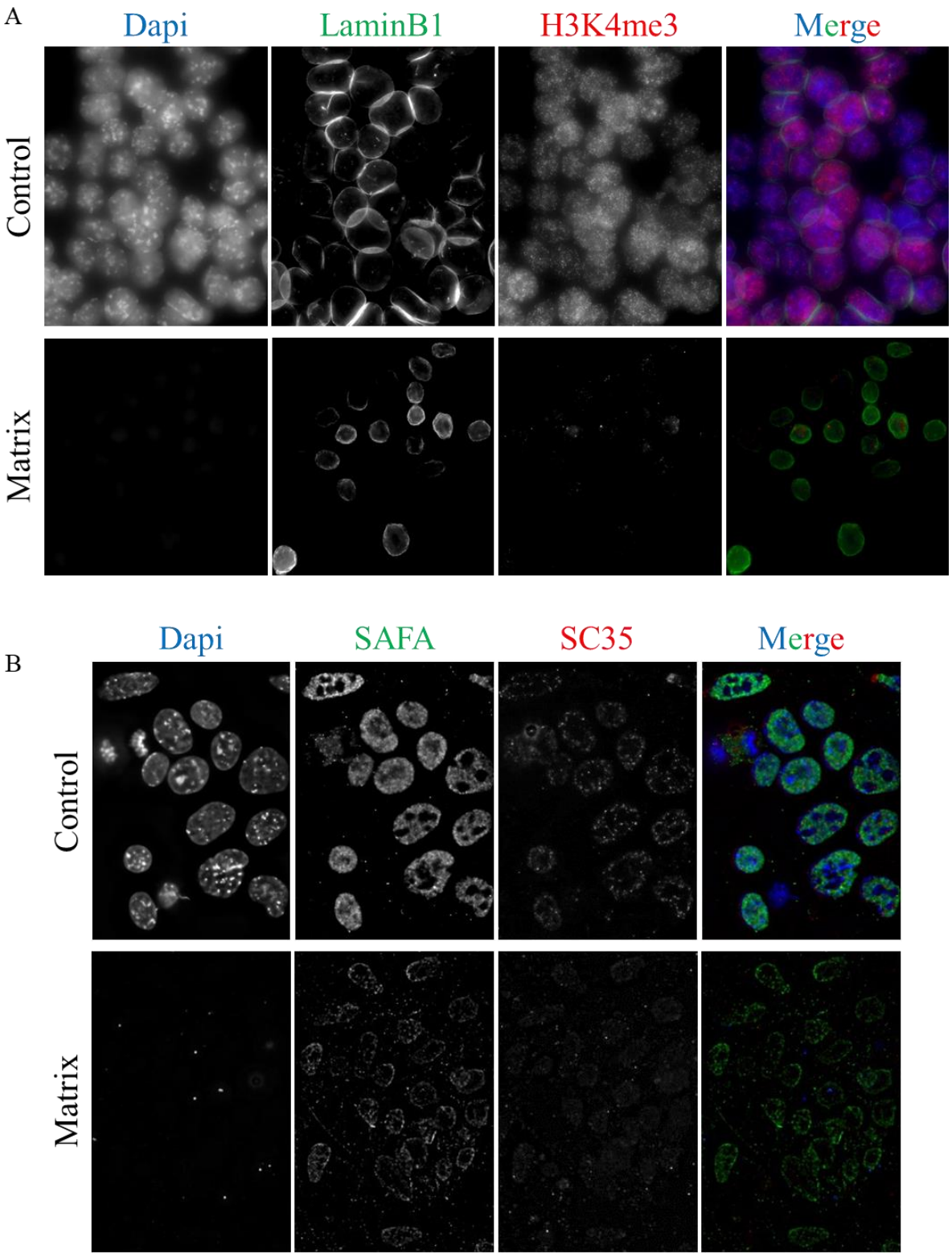


Figure 1.4. Nuclear matrix preservation assessed by immunostaining. A-B) Upper panel corresponds to intact nuclei and lower panel corresponds to extracted nuclei. Exposure times are equal for control and matrix images. A) Good retention of LaminB1 is obtained while H3K4me3 loss is quite massive – apart from few foci in some cells- when DNA is efficiently removed. B) SAFA and SC35 are detected in control and matrix nuclei, SC35 being somewhat extracted. Digested-chromatin clots can precipitate

on the glass slides and result in bright foci in the DAPI channel, however the drastic removal of DNA can be appreciated.

B. Matrix-associated transcript identification

In order to identify RNA molecules that could be implicated in the nuclear organization of mouse ES cells we performed RNA-seq on control and matrix samples. A wild-type male ES cell line (Tg2a) was used and three independent matrix preps were performed. Total RNAs after ribosomal depletion were sequenced by strand specific, paired-end RNA-seq capturing even non- or poorly poly-adenylated transcripts. Reads were aligned to the mouse genome using TopHat alignment tool and were quantified over 26127 annotated transcripts using Seqmonk software (Babraham Institute). DESeq2 analysis was conducted to detect differentially expressed (DE) genes between the control and matrix samples with a FDR cut-off of 5%. 5713 DE genes were identified, falling into two subgroups of genes: (i) 2976 genes significantly reduced in the matrix fraction samples compared to total RNAs and (ii) 2737 genes specifically enriched in the matrix RNA fraction. The log2-scaled RPKM values of the control samples were plotted against those of the matrix samples in the scatterplot depicted in **Fig.1.5**, showing a global decrease in the representation of annotated transcripts. Genes significantly enriched in the total RNA fraction (depicted in orange) were termed "depleted RNAs" because they were massively extracted from our matrix sample whereas genes significantly enriched in the matrix RNA fraction (depicted in red) were termed "Next RNAs", for Non-extracted RNAs. Amongst the Next RNAs we observed the lncRNAs Xist, Neat1, Malat1 and Firre, previously shown to nucleate functional domains in the nucleus. Thus, by retrieving molecules that are known to be able to participate in the formation of subnuclear domains, we confirmed the reliability of our approach.

Subsequently, we tried to characterize the Next RNA population by looking at distinct general features such as expression level, protein or non-coding annotation and gene structure. On average Next transcripts are expressed at relatively low levels compared to the highly expressed depleted transcripts in control samples (**Fig.1.6A**). Since lncRNAs have been shown to be typically expressed at low levels (Guttman et al., 2009, 2010a; Khalil et al., 2009), we thought that the Next RNA fraction could be belong to this non-coding RNA family. Nonetheless, when looking into the genes that constitute the two categories, Next and depleted RNAs, we found that around 90% of RNAs from both groups were reported to be coding for

proteins. Thus the annotated Next RNAs appear not to be particularly enriched in (long) non-coding transcripts. Furthermore, we investigated whether a bias in the size of the depleted or Next transcripts could be found. It was revealed that Next RNAs tend to be bigger in size than the depleted ones, since transcripts longer than 50kb are better retained (highest percentage in Next RNAs as compared to depleted) (**Fig.1.6B**). A possible explanation for that would be that the Next RNAs are enriched in long multi-exonic transcripts that undergo extensive splicing. Such transcripts are often associated with splicing speckles that resist the nuclear matrix fractionation.

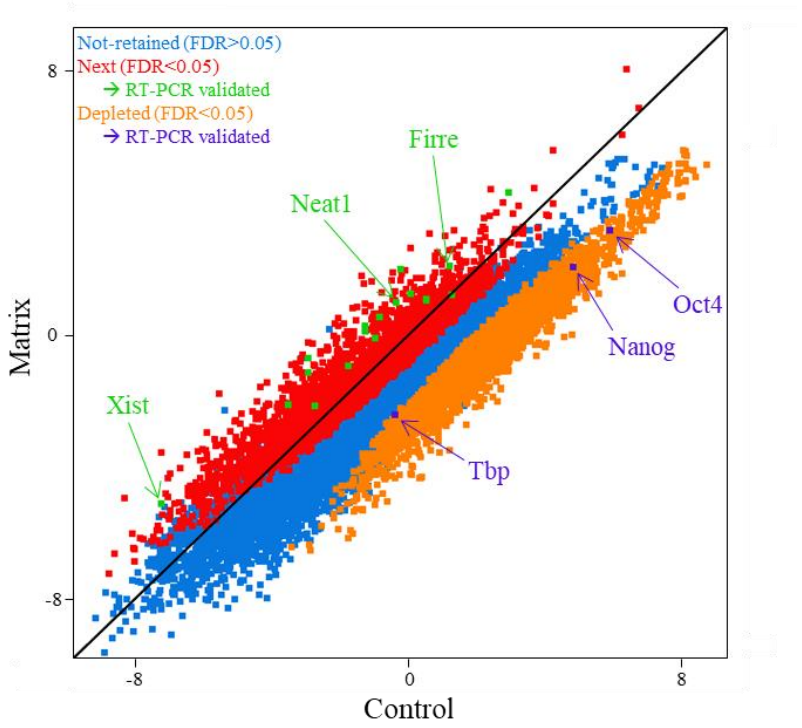


Figure 1.5. Scatterplot from RNA-seq analysis of 26127 genes in control and matrix samples. X axis corresponds to the averaged control samples and Y axis corresponds to averaged matrix RPKM values (in log2 scale). Enriched transcripts (FDR<0.05) in the control sample shown in orange are termed depleted RNAs; enriched transcripts (FDR<0.05) in the matrix sample shown in red are termed Next RNAs (Non-extracted); non-retained transcripts in blue are not significantly enriched (FDR>0.05) in either of the two samples. Few transcripts that were afterwards validated by RT-PCR of the depleted RNAs are shown in purple and of the Next RNAs are shown in green.

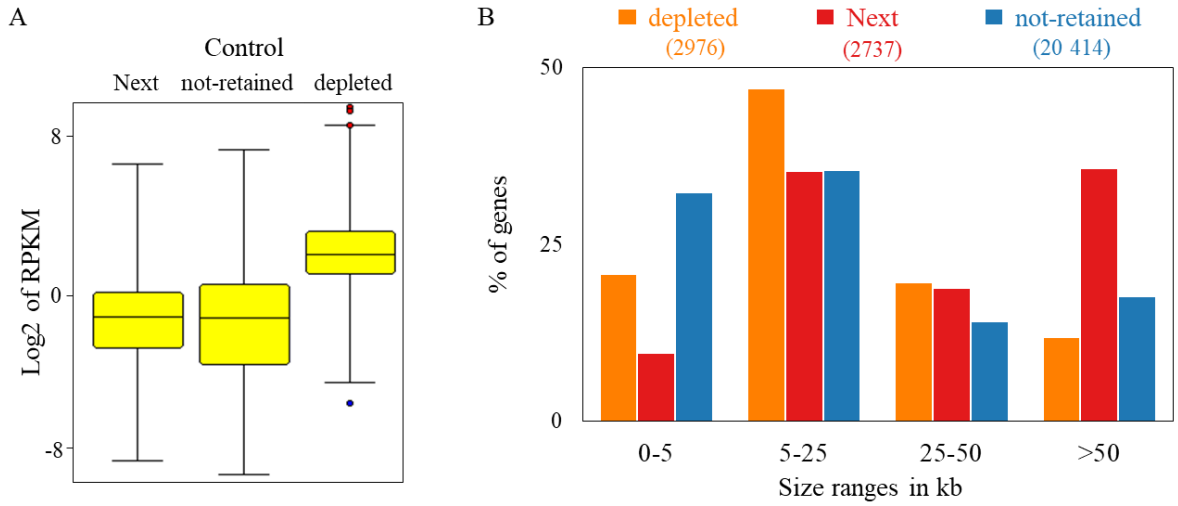


Figure 1.6. Analysis of expression level and size of the three subgroups of genes: depleted, Next and non-retained. A) Box plot of expression level of next, non-retained and depleted genes in control sample (one replicate shown). The depleted RNAs have on average a higher expression level compared to the next RNAs in a whole cell RNA population. B) Size distribution of the genes of each subgroup. X axis shows four size ranges of the transcripts in kb; Y axis shows the percentage of genes of each group that fall into each size range. Long transcripts (>50kb) seem to be preferentially retained.

Since repetitive elements have been proposed to play a role in shaping the nucleus *(Casanova et al., 2013; Chow et al., 2010; Hall et al., 2014a; Probst et al., 2010)*, we explored the possibility of them being enriched within our Next RNAs. We analyzed the enrichment in Line1 (L1), ERVK, ERVL and Alu elements as annotated in Repbase database and found that there is a greater overlap of repeats from all these repeat families with Next genes rather than with depleted genes. This finding is congruent with what has been shown by Hall et al., 2014 regarding Line1-rich Cot-1 RNAs being able to resist the matrix preparation and moreover having a potential role in genome packaging. These taken together with the finding that Alu-rich sequences are driving nuclear localization of mRNAs and more commonly of lncRNAs *(Lubelsky and Ulitsky, 2018)* suggests that repetitive elements might have an important role for nuclear organization.

We continued by looking further into the 2737 Next genes. In order to validate the RNA-seq datasets that we generated and create a short candidate list for further studies, we verified the retention and the depletion of some transcripts by RT-qPCR (**Fig1.5**). We manually sub-selected a few transcripts of depleted (three) and Next (thirty one) RNAs to be validated by RT-qPCR in additional matrix preps independent of the sequenced samples. We chose these

transcripts based on their level of depletion/retention in our RNA-seq data, their (decent) level of expression, and mostly non-coding annotation (see examples in **Fig.1.7**).

To perform this RT-qPCR validation, we generated matrix preps using Tg2a male ES cells and Lf2 female ES cells to rule out that any of the identified Next RNAs would be cell line or sex-specific. Four biological replicates of matrix preps were prepared two for each cell line- and the expression levels of thirty four genes (**Table 1**) was assessed. Among the Next RNAs that we selected to investigate by RT-qPCR few have an annotation of protein-

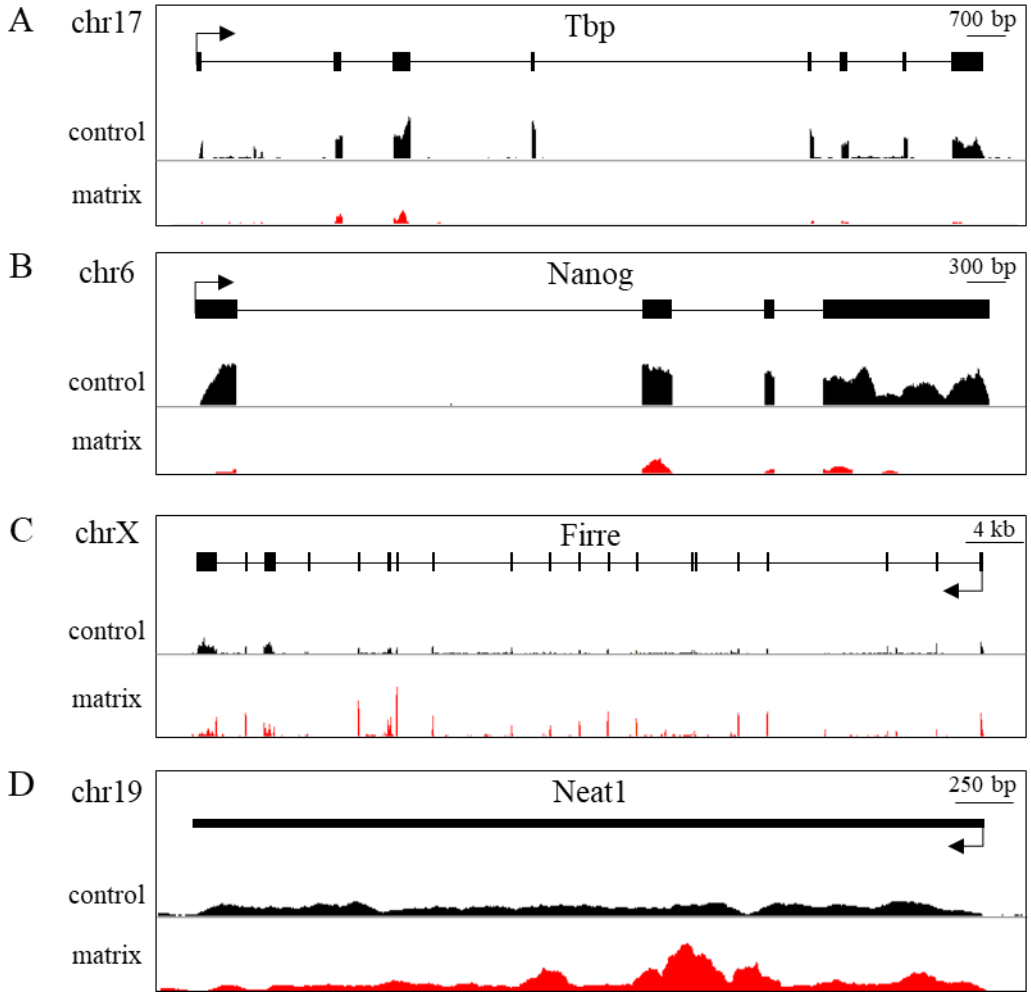


Figure 1.7. Screenshots of IGV browser showing RNA-seq coverage in control (black) and matrix (red) samples. Two examples are shown for deleted and two for Next transcripts. Reads coverage has been group-auto scaled to visualize the level of depletion and retention in each case. Tbp and Nanog show around 80% of loss in the matrix compared to the control samples. Firre and Neat are detected approximately 2,5 fold higher in the matrix than in control samples.

coding genes. A closer look to the read coverage of these particular transcripts indicates that only part of the full mRNAs qualifies them to be grouped within the Next population. In some cases (like Rere) only few exons seem to be retained and could potentially be forming circular RNAs (circRNAs) (Hansen et al., 2013; Memczak et al., 2013). The class of circRNAs usually resulting from intronic sequences circularization, has been reported to have functional roles in mammals but has not been yet largely studied. Another hypothesis could be that this kind of transcripts might also arise from specific isoforms of protein-coding genes that exert yet unknown non-coding RNA function, as reported in the case of the long isoform of the Zdbf2 gene (Liz) that acts as a *cis*-regulatory element for the transcription of the canonical Zdbf2 isoform (Duffié et al., 2014; Greenberg et al., 2017).

Gene	Coding or Non-coding	Depleted or Next in RNA-seq	Depleted or Next by RN-qPCR
Tbp	protein coding	Depleted	Depleted
Nanog	protein coding	Depleted	Depleted
Oct4	protein coding	Depleted	Depleted
Xist	non-coding	Next	Next
Tsix	non-coding	Next	Next
Gm12690	non-coding	Next	Next
Vaultrc5	non-coding	Next	Next
Gm13067	non-coding	Next	Next
Kis2	non-coding	Next	Next
Gm27000	non-coding	Next	Next
Gm26917	non-coding	Next	Next
Gm26924	non-coding	Next	Next
Dxz4	non-coding	Next	Next
Neat1	non-coding	Next	Next
Rmrp	non-coding	Next	Next
Rpph1	non-coding	Next	Depleted
Nespas	non-coding	Next	Next
Gm26788	non-coding	Next	Next
Gigyf1	protein coding	Next	Next
Meg3	non-coding	Next	Next
Terc	non-coding	Next	Next
Srrm2	protein coding	Next	Next
Etl4	protein coding	Next	Next
Firre	non-coding	Next	Next
Rere	protein coding	Next	Next
Titin	protein coding	Next	Next
Nphs1as	non-coding	Next	Next
Rtel1	protein coding	Next	Next
Gm11611	non-coding	Next	Next
Gm15726	non-coding	Next	Next
Gm15247	non-coding	Next	Next
2900056M20Rik	non-coding	Next	Next
Gm11946	non-coding	Next	Next
Gm26542	non-coding	Next	Next

Table 1. Genes verified by RT-PCR. Nine protein coding and twenty five non-coding genes.

The genes whose expression was assessed by RT-qPCR showed high consistency with the RNA-seq results. All three depleted RNAs and all but one Next RNAs (except for the non-coding Rpph1 gene) were indeed validated as depleted or retained in the matrix samples, in accordance with the RNA-seq results. The depletion or retention of the tested genes was represented as a ratio of expression in matrix to control samples (**Fig.1.8).**

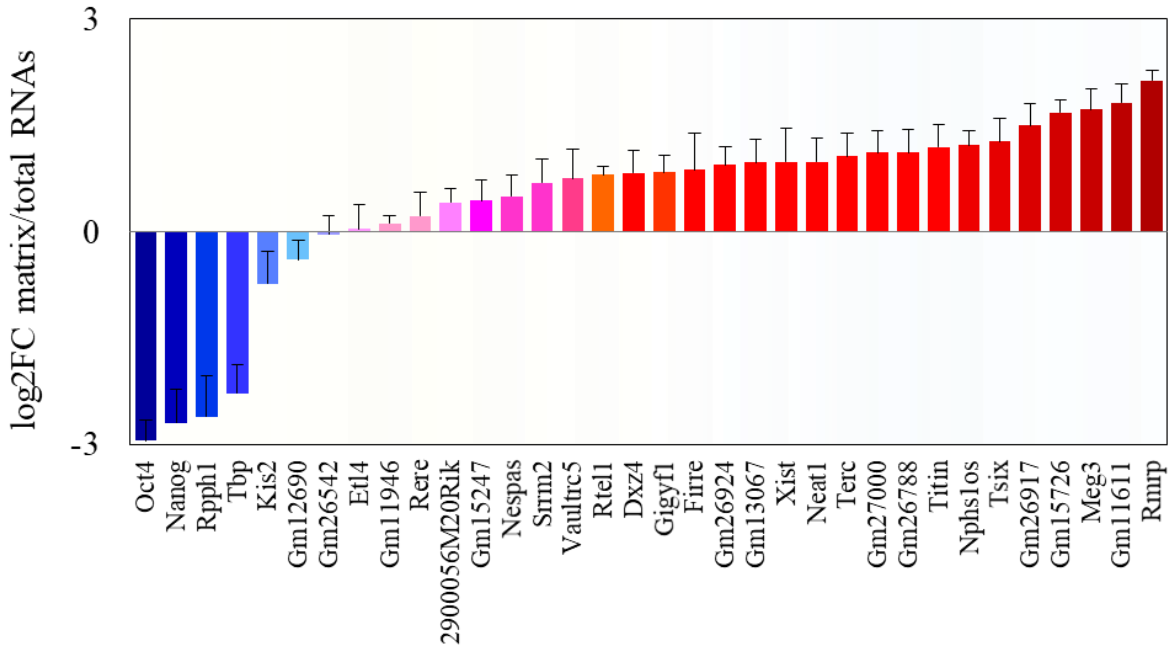


Figure 1.8. RT-qPCR validation of some depleted and Next genes. Expression level is shown in log2 ratio of matrix to control expression level; genes are order from highest depletion to highest retention values. Data are shown as means ± s.e.m. from four matrix prep replicates (n=4).

We further wanted to confirm by RT-qPCR the enrichment of repetitive elements in the matrix fraction that was observed from the analysis of our RNA-seq datasets. For this purpose we measured the amount of Line1, ERVK (including different subfamilies of Line1 and ERVK), ERVL retrotransposons but also major satellite repeats. All of them were found to be enriched in the matrix fraction (**Fig.1.9**). Remarkably, the major satellites were the repetitive elements to be the highest enriched in our matrix preps. Interestingly, they have been shown to have a pivotal role in the formation and reorganization of the heterochromatin in the early mouse embryo development (Casanova et al., 2013; Probst et al., 2010), suggesting comparable functions is ES cells.

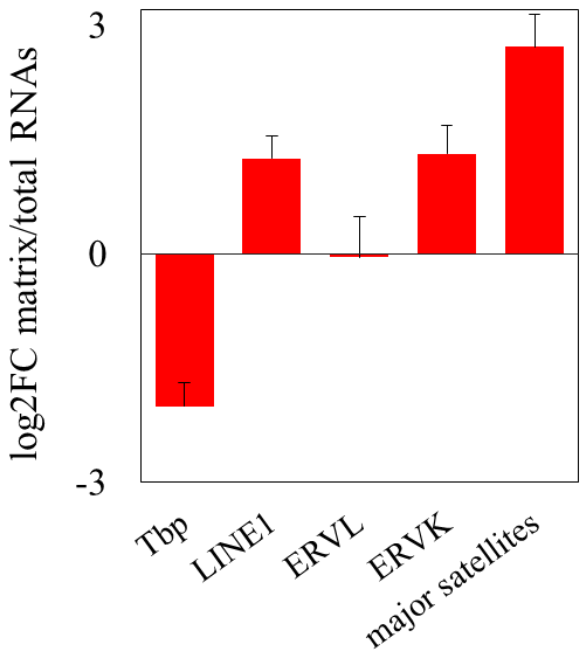


Figure 1.9. Repetitive elements are enriched in the matrix prep by RTqPCR. Expression level is shown in log2 of matrix to control ratio. Tbp is shown to be depleted whereas Line1, ERVL, ERVk and major satellite repeats are highly retained in the matrix samples (n=4, ±s.e.m.).

As mentioned previously (**Fig.1.4B**), remaining nuclear debris or fragments of digested chromatin can be sometimes precipitating on the glass slides we use in our experimental procedure. In order to preclude the possibility that the detection of the Next RNAs in our RNA-seq and RT-qPCR is due to such contaminants that did not get properly washed away after the DNA digestion, we added an additional layer of validation. We therefore performed RNA-FISH to evaluate the detectability of a several Next RNA transcripts in control and extracted cells. Moreover, in order to assess the efficient digestion of the genomic locus of the transcripts in question we performed DNA-FISH, to visually assess its elimination. Indeed, we were able to detect by RNA-FISH three tested Next RNAs, Xist, Firre and major satellite RNA, in both control and matrix samples while the DNA-FISH for the respective transcripts was no longer visible in the matrix cells. As (nicely) illustrated in **Fig.1.10**, bright foci of major satellite RNAs are detected in matrix cells where DNA is very efficiently digested and removed, as seen by DAPI staining and more importantly there is no signal for major satellite DNA (as by the DNA-FISH for major satellite). In control cells, major satellite DNA foci are perfectly colocalizing with the chromocenters (dense heterochromatic foci as seen by DAPI staining) that represent the coalescence of the major satellites from different chromosomes, and their RNA is localized

at the transcription sites. Likewise, RNA/DNA-FISH showed the retention of Firre and Xist RNAs in the matrix samples (data not shown) upon complete removal of their genomic loci and chromatin. This findings suggest that (undoubtedly) Next RNAs even when chromatin-associated, do rely on non-chromatin nuclear substructures to be able to resist matrix extraction.

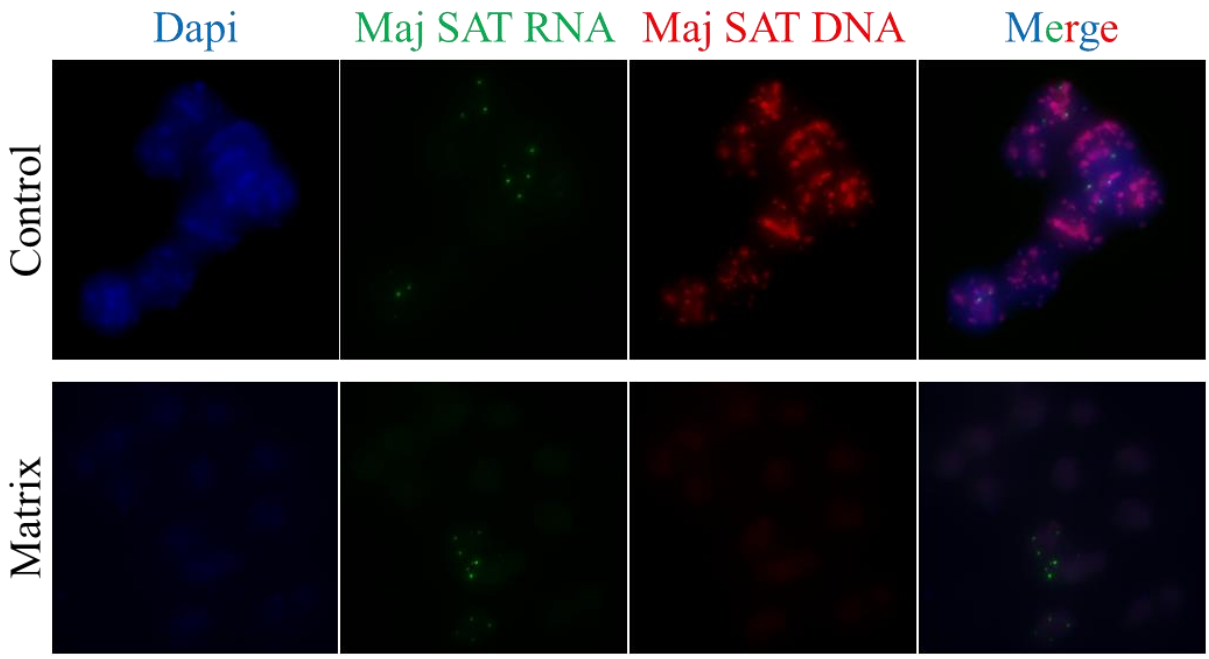


Figure 1.10. Sequential RNA/DNA-FISH in undifferentiated male mouse ES cells (Tg2a). Major satellites (MajSAT) RNA-FISH (green) followed by a MajSAT DNA FISH (red); DNA is counterstained with DAPI (blue). Upper panel control sample, lower panel matrix prep. In control sample MajSAT DNA colocalizes with the chromocenters (DAPI dense heterochromatic foci) and MajSAT RNA is detected accumulating at the transcription sites (on different chromocenters). MajSAT RNA is detected even upon complete DNA digestion and removal, as shown in matrix samples.

Having analyzed and validated our RNA-seq datasets by two independent methods (RT-qPCR and RNA-FISH), showed us the validity of our experimental approach in identifying RNA molecules that would be novel candidates for participating in the functional organization of the mouse ES cell nucleus.

C. Selecting candidates for functional characterization

The list of the Next RNAs that we obtained from our RNA-seq datasets is quite extended (2737 genes). As a consequence, we decided to focus on the short list of (mostly) non-coding transcripts that were validated by RT-qPCR and more specifically those that were ES cell specific. We speculated that genes that are specifically transcribed in ES cells but not in differentiated cells would be more probable to have a functional relevance in respect to ES cell biology. To this end, we decided to proceed with a commonly used differentiation assay of ES cells and monitor the dynamics of the expression of our Next RNAs every twenty four hours over the course of a three-day kinetics. For this purpose, we conducted retinoic acid differentiation assay which promotes stem cell neural lineage specification (Bain et al., 1995; Fraichard et al., 1995; Strübing et al., 1995). Driven by the idea that basal gene expression can be biased according to the background of the mouse of origin or the sex of distinct ES cell lines (Choi et al., 2017; Schulz et al., 2014; Sharova et al., 2007), we chose to check the dynamics of the expression of our Next RNAs in different mouse ES cell lines. Therefore, we used common wild-type ES cell lines; Tg2a, R1 (both male) and Lf2 (female). Total RNAs were collected daily and analyzed at the end of the three-day kinetics by RT-qPCR. First, in order to confirm a successful differentiation, we analyzed the expression of pluripotency markers (like Nanog, Oct4, Sox2, Klf4) that were expected to be downregulated and differentiation markers that were expected to be upregulated in this kind of differentiation assay (like Hoxb1, Id1, Sox17) (Zhang et al., 2015). As expected, we observed a rapid decrease in expression level of multiple pluripotency transcription factors (**Fig.1.11A**) but also an increase in expression level of differentiation markers (data not shown). We then assessed the expression levels of all our confirmed Next RNAs (Table 1, right column). Based on their transcriptional response we could assign them in three groups: (i) Next RNAs that are downregulated upon differentiation (**Fig.1.11B**), (ii) Next RNAs that are upregulated upon differentiation (**Fig.1.11C**) and (iii) Next RNAs the expression of which does not change during the three days of retinoic acid differentiation. Only four Next RNAs fell into the first category: Kis2, Gm12690, Nphs1as and Titin. Of these, only Gm12690 appears to have higher expression in our female cell line that also shows higher expression of pluripotency markers compared to ES male cell lines (Schulz et al., 2014). Interestingly, three of these Next RNAs, Kis2, Gm12690 and Nphs1as, have been annotated as ES cell specific transcripts in an independent study (Hussein et al., 2014), validating independently our conclusion. Of note, eleven Next RNAs underwent an upregulatation upon differentiation and fifteen Next RNAs did not show a particular response

upon retinoic acid treatment. Therefore, we decided to focus on the four ES cell specific Next transcripts to study their potential to regulate the genome of pluripotent ES cells.

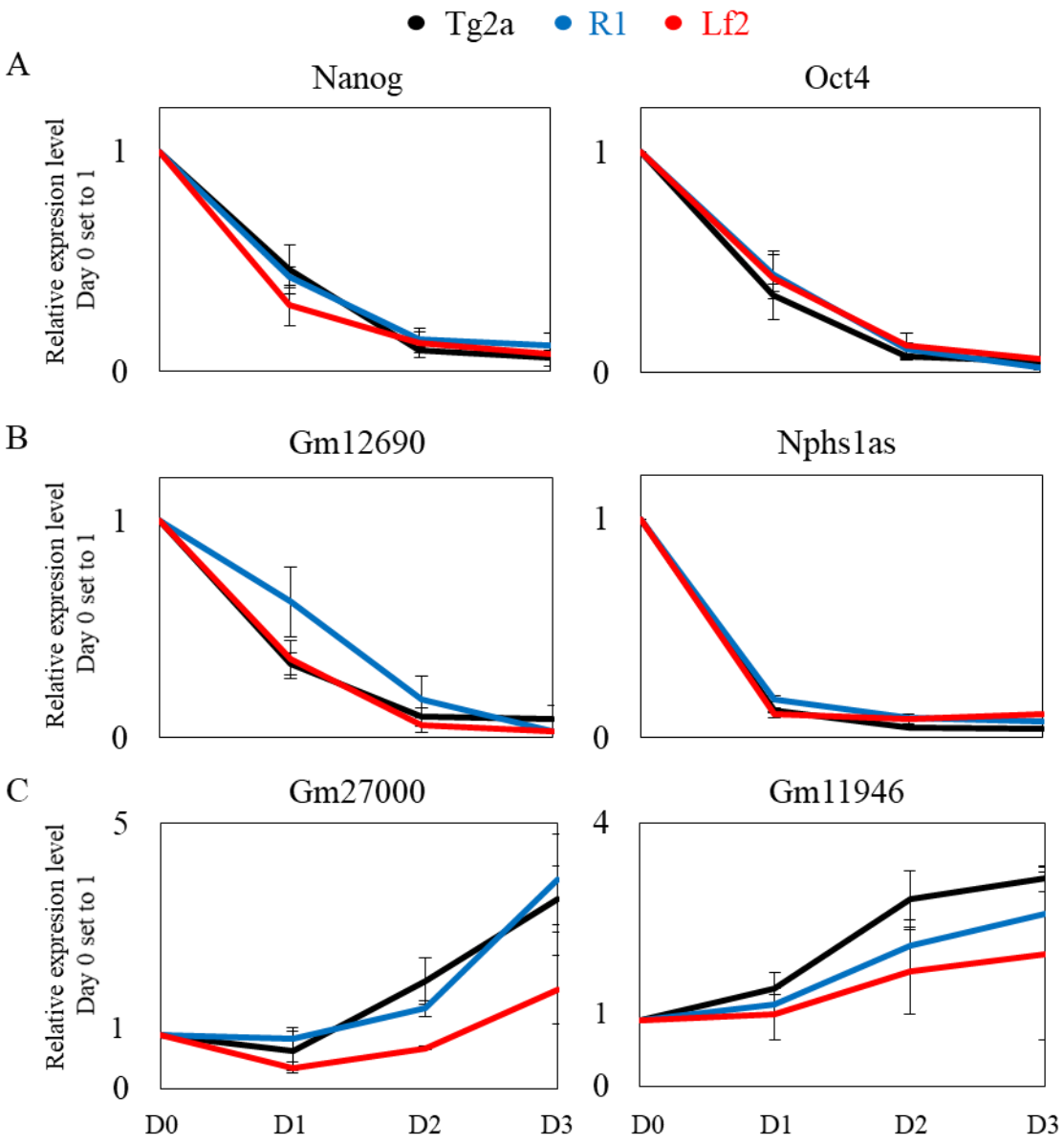


Figure 1.11. Kinetics of a three-day differentiation assay with retinoic acid in Tg2a (E14), R1 and Lf2 ES cell lines as measured by RT-qPCR. Values were normalized to Tbp and are expressed as the fold change to Day 0 (D0). Data represent mean ± SEM from three biological replicates for each cell line. A) Pluripotency markers are downregulated upon differentiation, B) Next RNAs that are downregulated upon differentiation, C) Next RNAs that are upregulated upon differentiation.

To that end, we first investigated the subcellular distribution and localization of Kis2, Gm12690, Nphs1as and Titin. RNA-FISH was conducted for all four of these genes using fosmid clone generated probes. We first observed that these four Next RNAs are strictly nuclear. Unfortunately, none of them does show a focalization in domains resembling those of

Xist cloud or Neat1 paraspeckles. While Kis2 (**Fig.1.12A**), Gm12690 and Titin RNAs are mostly detected at their transcription sites, Nphs1as (**Fig.1.12B**) diffusible transcripts were additionally detected in the nuclei of few cells. In addition, Kis2 and Nphs1as RNAs (**Fig.1.12**) are detected in the majority of the cell population (around 80%) whereas Gm12690 and Titin expression is detected in a smaller fraction of the cell population (less than 40%). We also performed RNA-FISH on matrix samples and were able to visually confirm the retention of the four Next RNAs in extracted nuclei (**Fig.1.12**). However, none of the four Next RNAs could form subnuclear structures such as the Xist cloud, the Malat-enriched speckles or the strong transcriptional foci of Firre. Therefore we decided not to proceed to further functional characterization of these transcripts but rather continue with the unannotated transcripts that were uncovered by our original sequencing material.

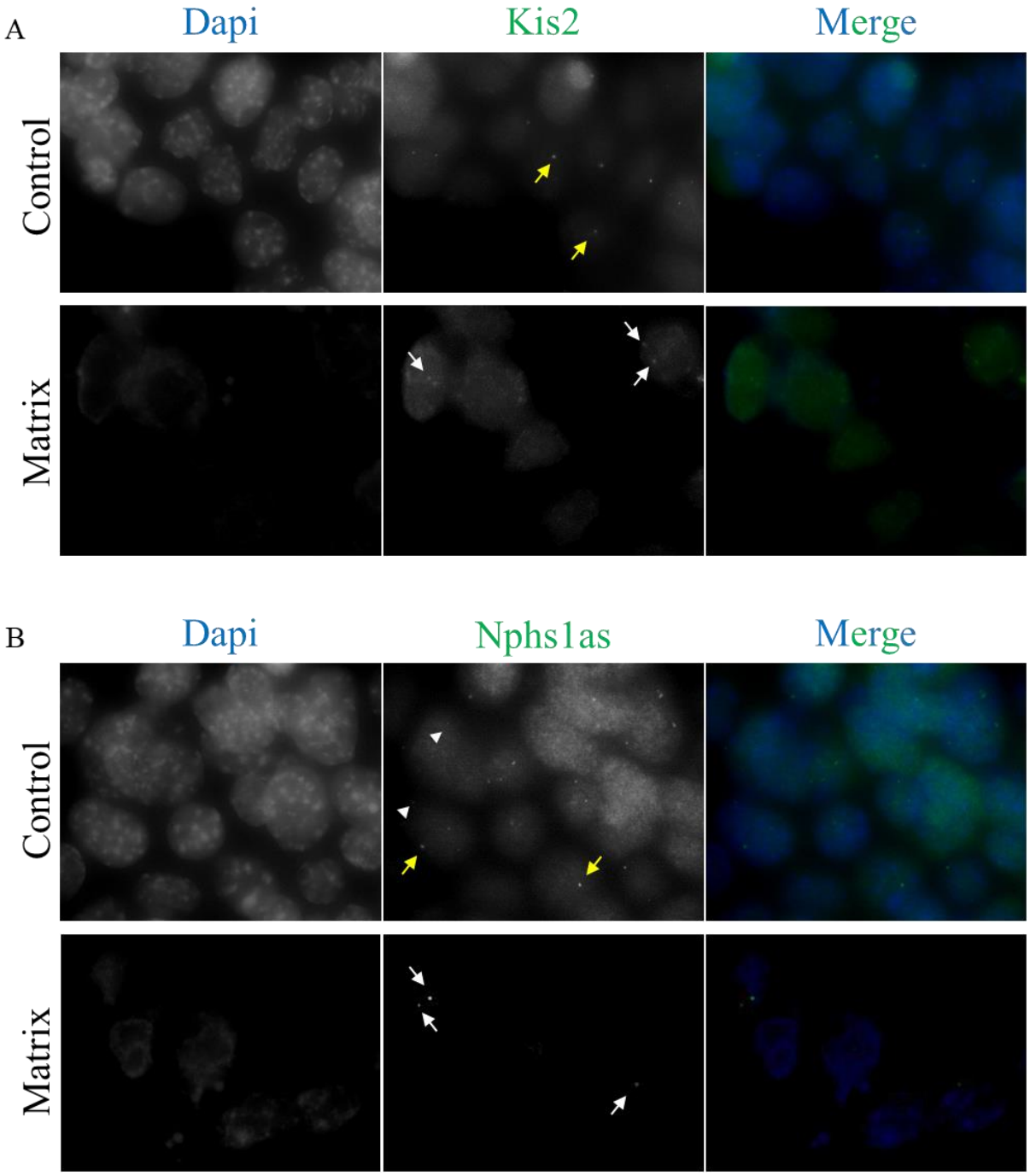


Figure 1.12. RNA-FISH for Kis2 and Nphs1as RNAs in control (upper panel) and matrix (lower panel) samples. Yellow arrows point to active transcription sites; white arrowheads point to diffusible RNA molecules; white arrows point to retained transcripts upon matrix extraction. DNA is counterstained with DAPI. A) Kis2 is detected at its transcription site in the majority of the visualized control cells and is also detected in nuclei devoid of chromatin. B) Nphs1as is detected at active transcription sites but also as diffusible molecules in the nuclei of control cells however it is detected only at foci that would correspond to transcription sites in the extracted nuclei.

D. Discovery of novel long non-coding RNAs

The first part of the study of our RNA-seq datasets was to identify which of the known or predicted transcripts of the mouse genome were able to resist the matrix prep and therefore represent potential candidates of RNAs with a role in nuclear compartmentalization. Additionally, we further investigated the possibility of discovering new, previously non-annotated transcripts by exploring deeper our datasets. With our experimental setup, a substantial part of the transcriptome is drastically washed out, thus increasing sequencing depth of the transcripts that resist the extraction and are commonly poorly represented in usual total RNA samples. This hypothesis seems to be validated when taking into account the low expression level of the identified Next RNAs in the control samples, where they are normally barely detectable (**Fig.1.6A**). We therefore performed a blind investigation of our datasets in search of novel RNAs that might have been overlooked so far in previous studies.

To this end, we undertook two independent analyses and later intersected the obtained results. First, using the Seqmonk analysis software, the whole genome was binned in regions of a 20kb size with a 4kb step size and the aligned reads were separately counted for each region (**Fig.1.13A**). We isolated the regions that were significantly enriched (displayed >2fold change and FDR<0.05) in the Next fraction compared to whole cell RNAs and were manually trimmed for reasonable read coverage density. These criteria led us to 80 regions, of which the majority represented annotated transcripts that were already discussed in the section I.B., such as Xist, Neat1 and Firre RNAs. When filtering out the known or predicted transcripts we narrowed down our list to 12 loci that did not correspond to any annotated gene. For our second blind approach, a Seqmonk feature called 'contig probe generator' was used in order to identify distinct active transcriptional blocks (**Fig.1.13.B**). Probes smaller than 5kb were discarded and a minimal coverage of 10 mapped reads in each matrix samples was used as a filtering. Overlapping probes were fused together. This resulted in a total number of 745 selected contigs. The contigs overlapping with annotated genes were trimmed out, shortlisting the probes to 39 that were located in gene deserts. These 39 regions finally corresponded to 10 genomic loci. From the candidates coming up from these two analyses there is an overlap of five loci, from which we sub-selected two, named Non-extracted Candidates 1 and 2 (thereafter NextC1 and NextC2), to perform more extensive analysis and characterization. The particular interest for each of the two candidates will be discussed in their corresponding section.

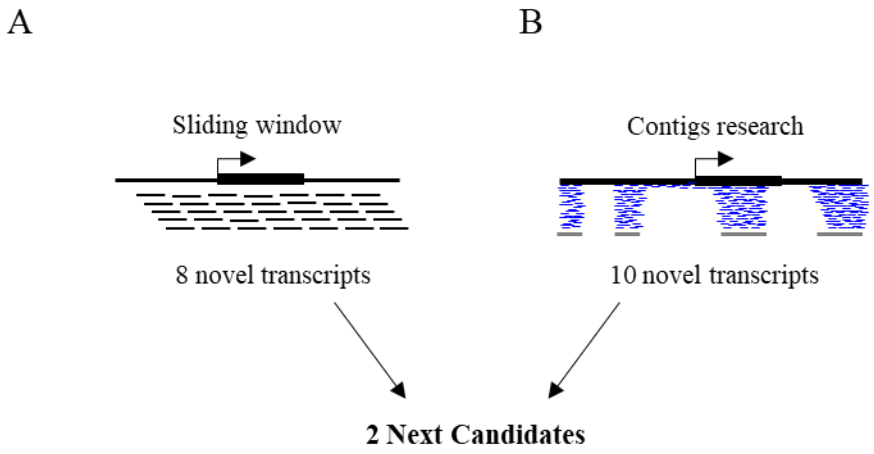


Figure 1.13 Pipelines for identification of novel transcripts using Seqmonk analysis software. A) Binning the genome in 20kb regions with a 4kb sliding window, selecting for those significantly enriched in matrix opposed to control samples and removing those overlapping annotated genes, led to the identification of 8 novel transcripts. B) Contigs of collapsed 5kb probes, highly represented in matrix samples, were generated and filtered for those overlapping annotated genes, revealed 10 novel transcripts. Combination of the two approaches resulted in the selection of two Non-extracted Candidates: NextC1 and NextC2.

E. Discussion

Although during the last years the field of long non-coding RNA biology has made a huge progress and thousands of such molecules have been identified, a systematic approach to predict lncRNA biological functions is still lacking. Computational analyses aiming at providing tools to improve our ability to predict the functionality of a given lncRNA just started to evolve (Kirk et al., 2018). In addition, the notion that lncRNAs actively participate in the nuclear organization has been well-established (Bergmann and Spector, 2014; Cheng et al., 2016; Engreitz et al., 2016; Ip and Nakagawa, 2012; Joung et al., 2017; Maass et al., 2018; Rinn and Guttman, 2014). We sought to identify new lncRNA molecules that would be functionally relevant for the shaping of the nuclear space. In order to tackle this issue, we got inspired by the concept of the nuclear matrix that claims the existence of a nuclear substructure consisting of proteins and RNAs independently of chromatin (Berezney and Coffey, 1974; Pienta and Coffey, 1985). The RNA component of the matrix has been shown to be crucial for the preservation of the nuclear matrix (He et al., 1990; Nickerson et al., 1989). We therefore wanted to purify the RNA fraction of the nuclear matrix believing that it would be of key importance for the structural organization of the nucleus.

With this approach, we confirmed our ability to identify RNA molecules that would be potential nuclear-organizers since known transcripts with this property emerged from our dataset. However, we faced technical issues with our matrix preparations as our protocol appeared to be quite harsh for cells grown on glass surface. We observed that many cells detached, resulting in having few cells available for immunostainings or RNA-FISH, and respectively for RNA isolation. One possible option to overcome this issue we had was to establish the same protocol in suspension for cells previously grown on plastic plates. However, the latter modification was not successful since the DNA digestion step resulted in a viscous precipitate that could not be further processed. Then, we decided to try an alternative approach that has been used for the isolation of transcription factories and the insoluble nuclear fraction (Melnik et al., 2011). According to the "transcription factories preparation", nuclei are isolated with a physiological, isotonic buffer, and chromatin is digested by DNaseI. The chromatin-free nuclei are then treated with Caspases in order to solubilize the transcription factories leaving as a precipitate the insoluble matrix. In order to establish this experimental technique in our laboratory, we started a collaboration with the group of Dr. A. Papantonis (Chromatin Systems Biology Lab, Centre for Molecular Medicine, University of Cologne, DE), who has extensively used and is currently working with this transcription factories isolation protocol. However, for unknown reasons, we could not successfully reproduce this experimental setup in our laboratory on mouse ES cells.

During our research work, another study came out aiming at identifying architectural ncRNAs. The experimental approach in this case was based on the observation that RNAs that are participating in the formation of nuclear bodies are entrapped in the protein phase during regular RNA extraction methods, due to strong RNA-protein interactions (Chujo et al., 2017). The use of differential extraction methods by sample needle shearing or heating, and subsequent RNA-seq of the recovered RNAs resulted in the identification of ncRNAs which exert a subnuclear granule-like distribution. Interestingly, in this study they could retrieve known ncRNAs forming nuclear bodies, such as Neat1, Line1 and Gomafu RNAs, in addition to the identification of novel ncRNAs with a nuclear body-like subnuclear distribution. Therefore, this approach was conceptually comparable to ours, since an alternative RNA purification would lead to the determination of a subset of RNA molecules that could have an architectural role in the nucleus. It would be of interest to compare their datasets with ours in order to identify any common transcripts that could come up from such an analysis.

As previously mentioned, our Next RNA genes were found to be expressed at relatively low levels in total RNA samples. This strongly corroborates the idea that we indeed identified transcripts specifically enriched in the matrix fraction, and not the most abundant RNAs that would be detectable non-specifically, after the extensive depletion of the transcriptome upon our matrix preparation protocol. Interestingly, a high percentage of the Next RNAs had a size bigger than 50kb that might be explained by the presence of long premature RNAs retained at nuclear speckles. Indeed, long transcripts have a higher probability to be undergoing splicing at a given time compared to shorter RNAs at equivalent transcriptional rates. Moreover, it is worth investigating the average size of introns of our Next RNAs since large introns tend to be subjected to recursive splicing (Georgomanolis et al., 2016; Hayashi et al., 2018; Pai et al., 2018) which might additionally increase the time they are being spliced after finishing to be transcribed.

During the last years our knowledge and understanding of the complex 3D genome organization and the nuclear architecture is growing rapidly. Recently, it has been proposed that the formation and maintenance of different nuclear bodies relies on liquid-liquid phase separation principles based on differences in concentration. Nucleoli, paraspeckles, Cajal bodies but also super enhancers and transcription factor hubs can form liquid-like condensates which are able to compartmentalize and concentrate proteins of similar biochemical properties (Berry et al., 2015; Cho et al., 2018; Fox et al., 2018; Hnisz et al., 2017; Larson et al., 2017; Mangan et al.; Sabari et al., 2018; Sawyer et al., 2018). Intriguingly, RNAs have been also shown to be able to promote phase separation depending on their local concentration and/or secondary structure (Langdon et al., 2018; Maharana et al., 2018). More interestingly, Xist RNA has been proposed to induce the heterochromatinization of the inactive X-chromosome via phase separation mechanism of accumulation of repressive proteins (Cerase et al., 2018). We could therefore speculate, that Next RNAs that are found at high local concentration in distinct regions of the nucleus could potentially drive the nucleation of domains via phase separation properties. In this direction, Line1 and major satellite RNAs are quite abundant in the nuclei of ES cells, with focal enrichment at euchromatic or pericentric heterochromatin respectively (Hall et al., 2014b; Percharde et al., 2018; Tosolini et al., 2018; Velazquez Camacho et al., 2017). Despite the lack of experimental evidence that these repeats can phase separate, we could hypothesize that the focal accumulation of these transcripts could be a source of phase separation that might help for the structural organization of the domains they are associated with.

A different experimental approach we could apply in order to identify the nuclear matrix associated RNAs would be to determine the lncRNA-interactome of SAFA. SAFA has been reported not only to be interacting with Xist and Firre lncRNAs but also to be a crucial mediator for their function (Hacisuleyman et al., 2014; Hasegawa et al., 2010). We thus reason that it would be of great interest to establish the list of the lncRNAs that are bound by the SAFA protein. Furthermore, SAFA has been shown to physically interact with Oct4 and Sox2 in ES cells (Vizlin-Hodzic et al., 2011) suggesting that SAFA cooperation with RNA might have a direct impact on the pluripotency network activity. In addition, it has been proposed that the DNA-binding activity of Oct4 might be partially dependent on its interaction in complex with the Panct1 lncRNA showing that non-coding transcripts can directly modulate the regulatory functions of pluripotency TFs (Chakraborty et al., 2017). Therefore, an intriguing experiment would be to perform a sequential immunoprecipitation assay (IP) of Oct4 and then SAFA proteins, and afterwards sequence the RNAs bound by the two proteins in complex. In this way we would be able to identify lncRNAs that might be associated with the nuclear matrix through their interaction with SAFA but also fulfill important roles for the regulation of the pluripotent transcriptomic signature.

In the workflow that we followed after the identification of our Next RNAs, we prioritized the study of those that were downregulated upon differentiation. The nuclear architecture changes drastically upon differentiation of ES cells (Meshorer and Misteli, 2006; Meshorer et al., 2006) and we decided as a first approach to focus on transcripts that might be responsible for the specific nuclear organization of the pluripotent stage. However, the Next RNAs that were found to be upregulated upon differentiation could also be very interesting candidates for the establishment of the changing nuclear organization and might be worth studying further. We could first assess their subcellular localization in both undifferentiated and differentiating cells and monitor their potential to form domains.

In order to simplify the computational analysis of our matrix RNA-seq samples where repeats are abundant, and to perform robust statistical analysis of the non-repetitive Next RNA fraction we recently decided to proceed with additional RNA sequencing after poly(A) selected libraries preparation. The new datasets are currently under statistical analysis.

II. NextC1 (Non-extracted Candidate 1)

A. Validation of NextC1 RNA and matrix retention

NextC1 is the first Next RNA candidate that was selected for an extensive characterization. From our RNA-seq data NextC1 appears as a very long stretch of reads spreading over a nearly 60kb long genomic locus located on mouse chromosome 5. It is transcribed from the negative strand within a gene desert with its closest neighboring gene being located 44kb away from it. No evidence of splicing within the transcript could be identified despite our paired-end long read sequencing (**Fig.2.1A**). We first aimed at validating Nexct1 transcription unit structure, unveiled by our RNA-seq results, by RT-qPCR. For that purpose, we designed a number of primer pairs covering the entire region of dense reads coverage with an extent to the 5' and 3' extremities and performed RT-qPCR in the three ES cell lines previously used (**Fig.2.1B**). We confirmed the transcription start site (TSS) of the transcript as well as two putative termination sites corresponding to the drop of signal of the RT-qPCR. The two termination sites were additionally confirmed by RT-qPCR performed after oligo(dT) priming in the RT reaction, *i.e.* with a bias towards the 3' polyA+ transcripts end (data not shown). In addition, we observed a higher level of expression in our female cells (Lf2) compared to the two male lines (R1 and Tg2a), reminiscent of the naïve pluripotency markers' expression profile (Schulz et al., 2014). Of note, for all the further RT-qPCRs performed on NextC1 RNA, three primer pairs are used across the locus; one closer to the beginning of the transcript, one in the middle and one towards the end of the gene body.

Next, we investigated the local chromatin environment of the NextC1 locus by looking for the presence of histone modifications and RNA polymerase enrichment at the locus. To that end, we used existing chromatin immunoprecipitation sequencing (ChIP-seq) data of the ENCODE/LICR track of UCSC genome browser generated from mouse ES cells (**Fig.2.1C**). An unambiguous enrichment of RNA polymerase II at the TSS as well as in the gene body shows that NextC1 is transcribed by RNA pol II. This was confirmed by a strong enrichment in histone post-translational modifications typically found at RNA pol II transcription units such as H3K36me3 that marks transcriptional elongation in gene bodies (Baubec et al., 2015) and a very clear H3K4me3 peak around the TSS (Bernstein et al., 2002; Santos-Rosa et al., 2002). Interestingly, we noticed that the NexctC1 promoter was also surrounded by strong enrichments in H3K4me1 and H3K27ac histone modifications that mark active enhancers

(Creyghton et al., 2010; Heintzman et al., 2007). This suggests that NextC1 could be classified as a unidirectional enhancer RNA (eRNA) (Kim et al., 2010b; Koch and Andrau, 2011; Santa et al., 2010). However, its very large size, relatively high level of expression as well as the high ratio of H3K4me3/me1 at its promoter argue in favor of its assignment to the lncRNA family (Lam et al., 2014).

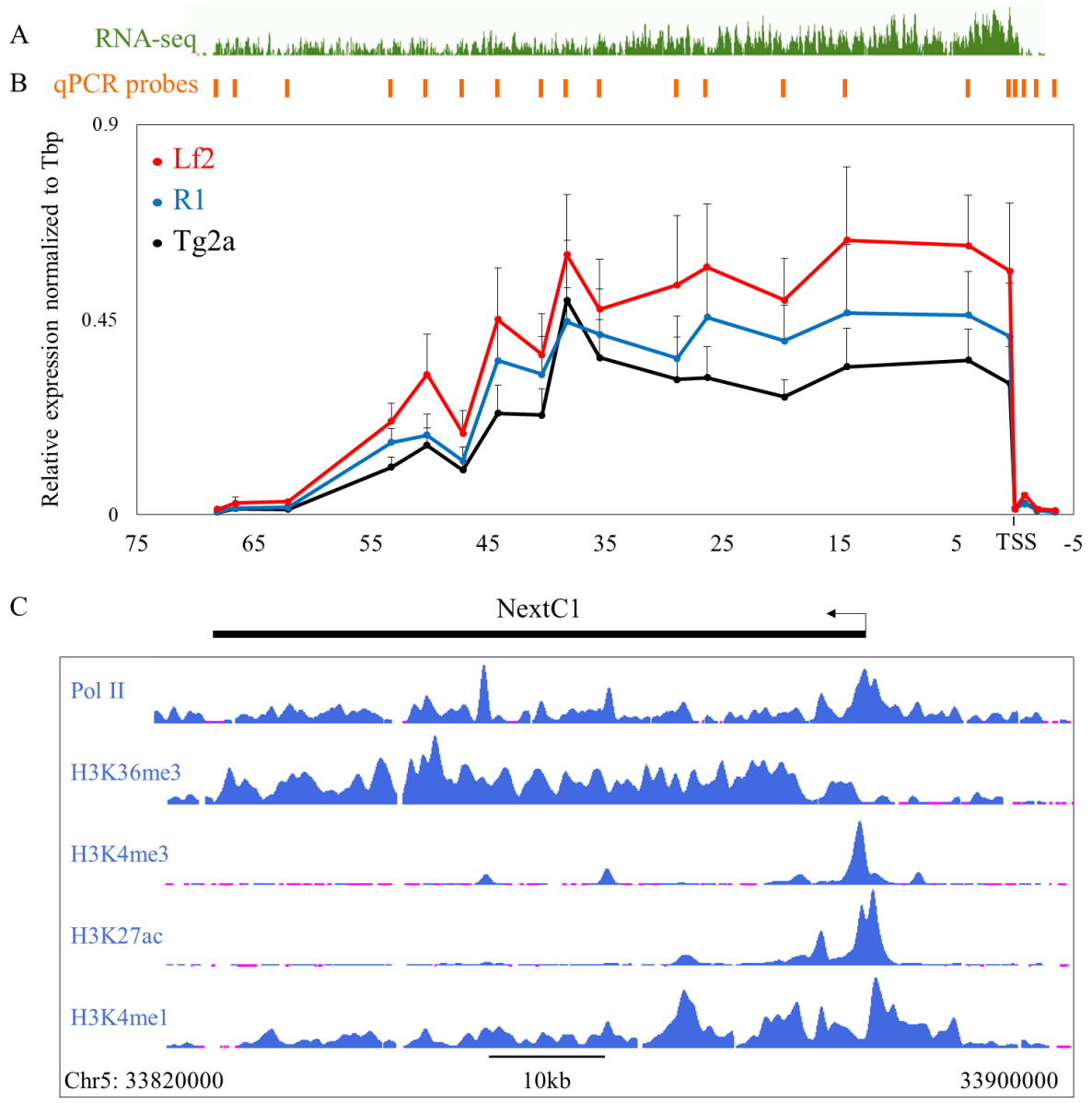


Figure 1.1 NextC1 locus characterization. A) Screenshot of IGV browser showing RNA-seq coverage from a control Tg2a sample. NextC1 is transcribed from the negative strand. B) RT–qPCR analysis of NextC1 expression in Tg2a, R1 and Lf2 ES cell lines. Values are normalized to Tbp mRNA and are shown as means ± s.e.m. from five independent culture replicates. X axis corresponds to the position of the primers along the transcript in kb; 0 corresponds to the transcription start site (TSS) and the orientation is inverted to be correlated with the transcription from the (-) strand. C) Schematic representation of NextC1 locus with a screenshot of UCSC genome browser (ENCODE/LICR track,

mm9 assembly) showing transcription by RNA polymerase II, the active transcriptional unit defined by H3K36me3, the promoter region defined by H3K4me3 and the enhancer region marked by H3K27ac and H3K4me1.

After validating NextC1 transcript expression in regular samples of total RNAs, we sought to validate the retention of the transcript in matrix samples as detected in RNA-seq (**Fig.2.2**). Therefore, we performed RT-qPCR in two different ES cell lines (Tg2a and Lf2) in control and matrix samples. We used Tbp as a reference gene to demonstrate the efficient depletion in each matrix prep and alongside measured NextC1 expression in the matrix and control fractions. We ascertained NextC1 retention in every matrix sample upon all the conditions tested (**Fig.2.3**) definitely stating NextC1 RNA as a reproducible matrix fraction associated transcript.

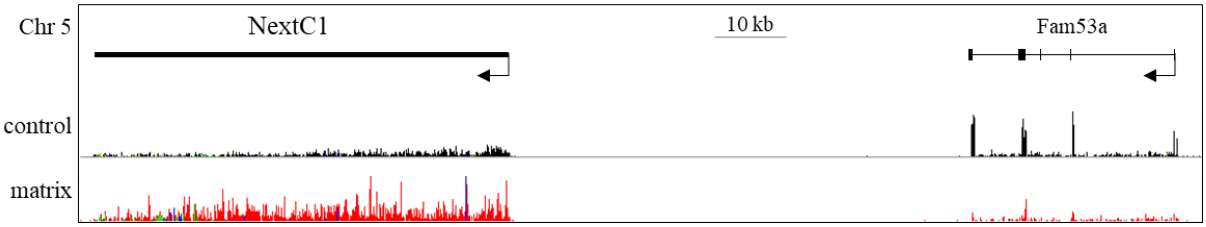


Figure 2.2. Screenshot of IGV browser of NextC1 locus with its neighboring gene Fam53a showing RNA-seq coverage in control (black) and matrix (red) samples. Reads coverage have been group-auto scaled to visualize the level of depletion and retention. NextC1 is enriched in the matrix sample whereas Fam53a (a depleted RNA) is massively lost in the matrix sample.

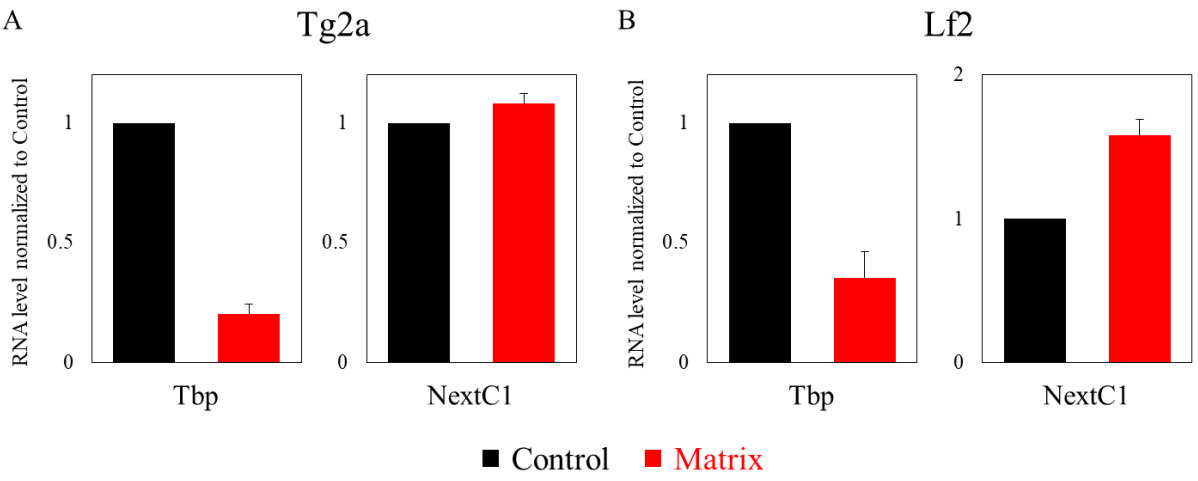


Figure 2.3. NextC1 is retained in nuclear matrix preps while Tbp is extracted in ES cells. A) Tg2a (n=6, ±SEM). B) Lf2 (n=3, ±SEM).

B. NextC1 coding potential and conservation

To date, NextC1 does not have an annotation neither in the ENSEMBL database (Zerbino et al., 2018) nor in the largest databases dedicated to mouse non-coding transcripts such as NONCODE (Zhao et al., 2016) or deepBase (Zheng et al., 2016). Therefore, to confirm that our gene is a *bona fide* lncRNA we questioned its coding potential using the PhyloCSF track of the UCSC genome browser (Lin et al., 2011). The full locus of NextC1 shows a negative coding potential in all of the three possible frames all along this large region. We also used the web tool Coding Potential Assessing Tool (CPAT) (Wang et al., 2013a) to interrogate the coding potential of NextC1. The coding probability given by the algorithm was of 0.089 (when the positive coding probability cutoff is >0.44) leading to the non-coding labelling of NextC1.

Last, we investigated the conservation of the gene in mouse and human genomes. We found that there is very little sequence conservation (2-3kb out of a 58kb sequence) between the two organisms apart from small, highly conserved regions, most likely corresponding to transcription factor binding sites in the syntenic human region (according to transcription factors binding site annotation from the ENCODE ChIP-seq database, UCSC human genome browser). In parallel, we tried to investigate the existence of local repeats internal of NextC1 that might have a functional role for a lncRNA, as it has been reported for the Firre lncRNA (Hacisuleyman et al., 2016). Thus, we used the BLAST tool to map 10 kb long bins of Nextc1 full sequence to the mouse genome but could not identify any specific repeated sequences within NextC1 whereas a comparable method easily retrieved short internal repeats within Firre.

C. NextC1 RNA stability

NextC1 is transcribed by RNA pol II (as shown before, **Fig.2.1C**) with no evidence of splicing events and exonic structure in our RNA-seq data (**Fig.2.2**). In addition, the transcription termination of the transcript does not seem to be dominated by a strong polyA signal. These distinct features, taken together with the decoration of NextC1 promoter with enhancer marks prompted us to determine the stability of such a long transcript as an interesting indicator of potential functionality. We thus performed transcription inhibition assays with the help of two commonly used drugs, Flavopiridol, an RNA polymerase II elongation inhibitor, and Actinomycin D, a DNA intercalating agent (Bensaude, 2011). We monitored over 6h the

stability of few mature and premature transcripts alongside with the NextC1 RNA levels. We found that the pre-mRNA levels of Nanog are already dramatically affected as soon as 30 minutes after drug treatment whereas its mature mRNA level shows a maximum of 50% decrease after 6h of flavopiridol treatment and no response upon Actinomycin D treatment (**Fig.2.4**). Of note, Nanog mRNA has been shown to have a half-life of ~5h in Actinomycin D (Abranches et al., 2013) but in our case, likely due to milder inhibition, its apparent stability upon this treatment was increased. Finally, NextC1 appeared to be relatively unstable, with a half-life of less than 2h, when compared to Xist or Firre lncRNAs that have been shown to be stable after more than 5 or 6h of transcription inhibition respectively (Clemson et al., 1996; Hacisuleyman et al., 2014) but still showed a higher stability than rapidly degraded intronic sequences. The relative instability of NextC1 could be explained by the fact that it is an unspliced lncRNA with a strictly nuclear localization and both of these features characterize lncRNAs more likely to be unstable (Clark et al., 2012).

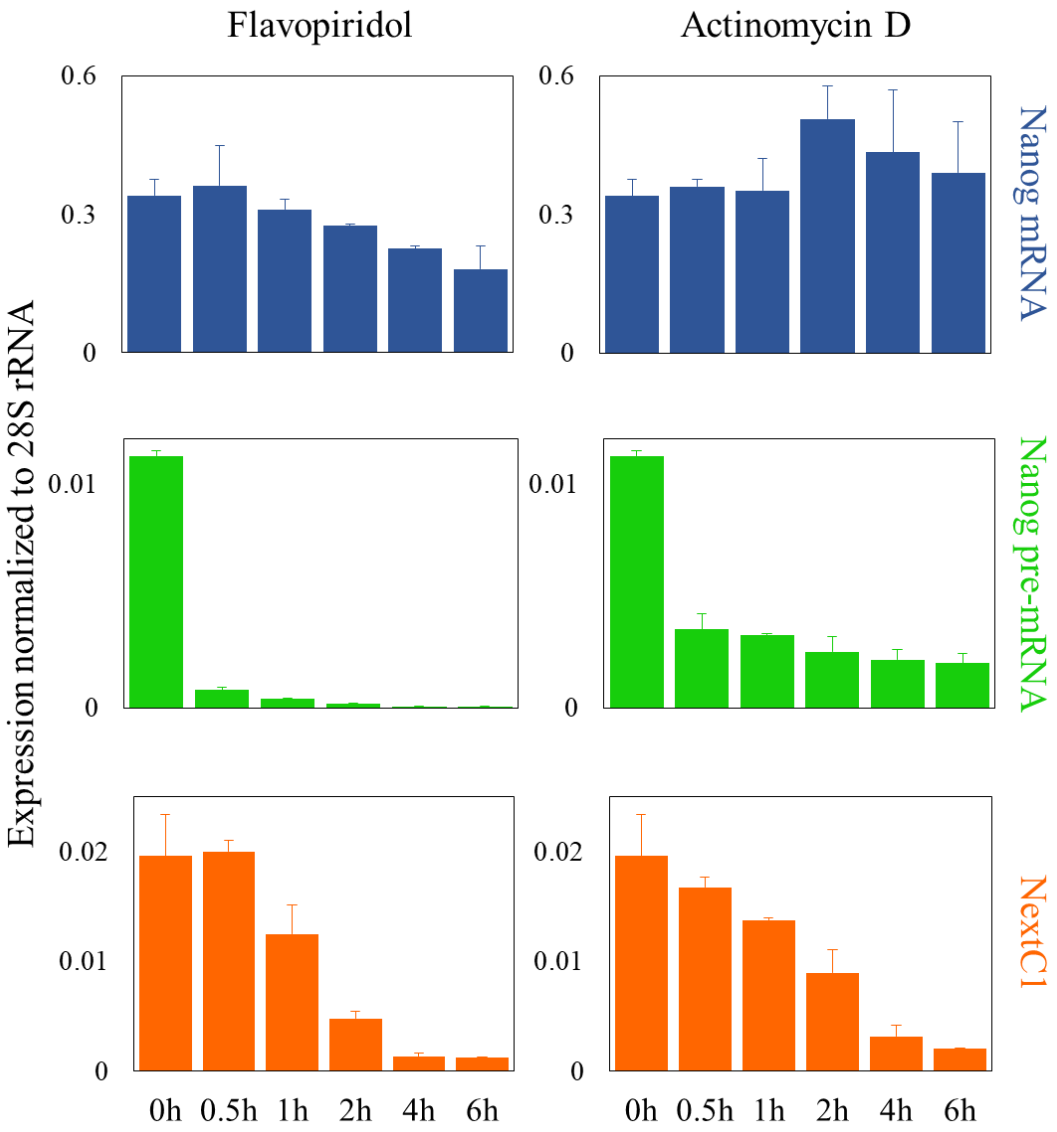


Figure 2.4. Transcription inhibition by Flavopiridol (left panels) and Actinomycin D (right panels) calculated over 6 hours. Representative RT-qPCR graphs for Nanog mRNA (blue), Nanog pre-mRNA (green) and NextC1 (orange). Data are normalized to 28S rRNA and are shown as means and ±sem of 2 independent experiments.

D. NextC1 expression regulation by the pluripotency network

Intrigued by the expression profile of NextC1, which was reminiscent of the pluripotency markers pattern in our three cell lines (**Fig.2.1B**) and the strong enrichment for markers of distal regulatory elements within the NextC1 promoter (**Fig.2.1C**), we wondered whether its genomic locus is targeted by pluripotency transcription factors (TFs). For that aim, we took advantage of the publicly available ChIP-seq data in ES cells for multiple pluripotency TFs (Chen et al., 2008; Handoko et al., 2011; Ma et al., 2011) and visualized them on the UCSC

browser through the online compendium CODEX (http://codex.stemcells.cam.ac.uk/). Strikingly, we ascertained a plethora of TFs involved in pluripotency to be bound to NextC1's promoter (**Fig.2.5**). The core pluripotency TFs Nanog, Oct4 and Sox2 as well as Esrrb, Klf4, Tcfcp2l1, Prdm14, Tbx3, exert a strong binding site right at the promoter of NextC1 gene. Due to this high TFs occupancy, we further continued our investigation of the NextC1 locus by looking into a published study on the strong *cis* regulatory elements called super enhancers identified in mouse ES cells (Whyte et al., 2013). Super enhancers are described as dense clusters of enhancers with high levels of Oct4, Nanog, Sox2, Klf4, Esrrb and Mediator coactivator (Med1) co-binding. These elements often trigger local transcription activation and distally regulate the expression of genes involved in the control of cell identity (Hnisz et al., 2013; Whyte et al., 2013). Interestingly, NextC1 promoter region has been identified as a 7kb long super enhancer in mouse ES cells (**Fig.2.5**) suggesting potential key regulatory functions of this locus in the maintenance of ES cell identity.

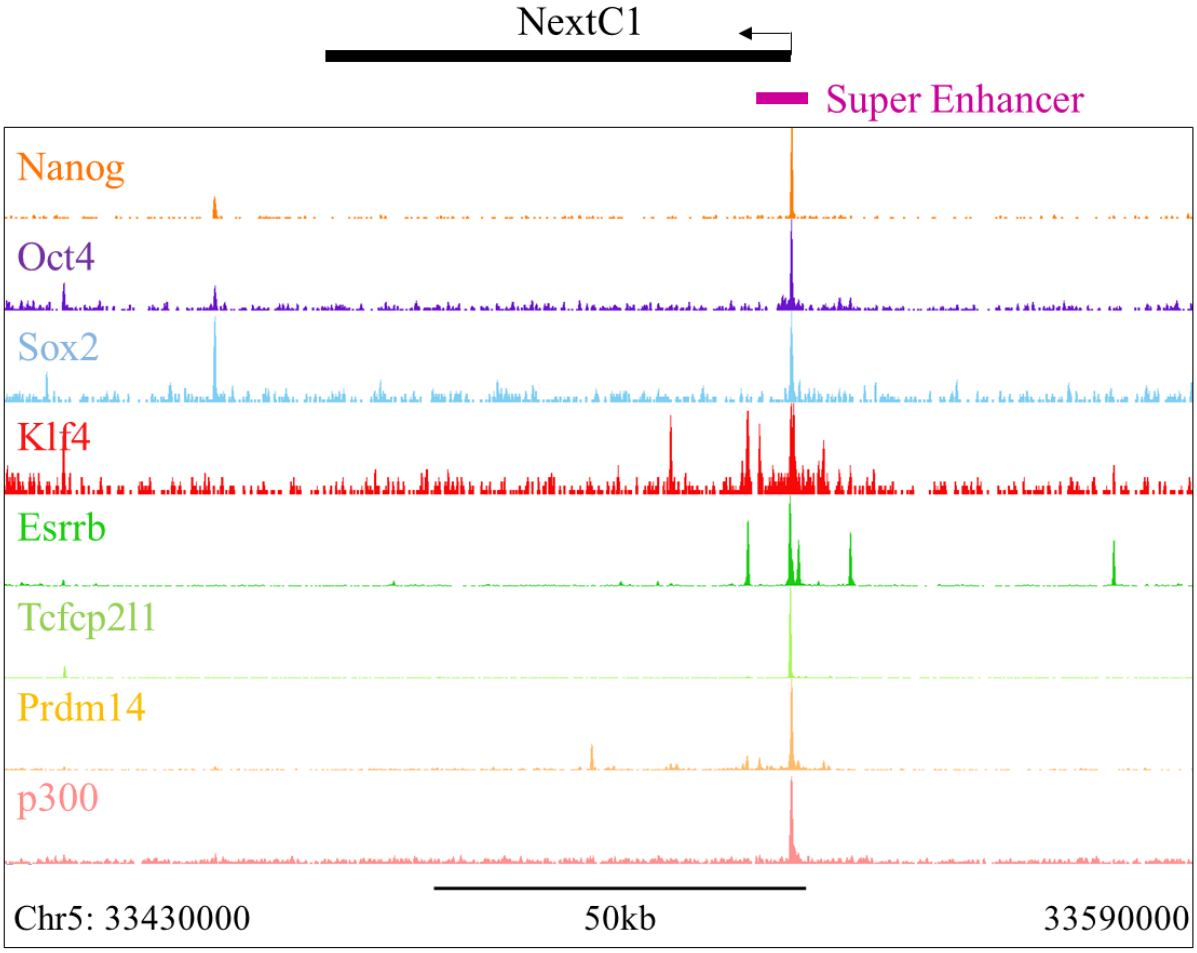


Figure 2.5. Transcription factors binding profiles at NextC1 locus as defined by ChIP-seq (UCSC browser, mm10 assembly). A least one strong binding site is found at the promoter region of the gene

for Nanog, Oct4, Sox2, Klf4, Esrrb, Tcfc2l1 and Prdm14 pluripotency TFs as well as the p300 enhancer binding protein.

Consequently, the dual nature of NextC1 promoter prompted us to study its regulation by the pluripotency factors as well as its expression in pluripotent and differentiated cells. First, we compared its expression levels in ES cells cultured in usual serum-containing medium supplemented with LIF and in the so called "ground state of pluripotency" where cells were cultured in serum-free medium in the presence of LIF and the two kinases inhibitors (2i) for 72h. We observed that, like Nanog and other pluripotency genes which display higher expression levels in 2i culturing condition, NextC1 expression increases during the serum to 2i transition (**Fig.2.6A**). Comparably, upon 48h LIF withdrawal, when ES cells undergo early differentiation and factors such as Nanog, Klf4 and Esrrb exhibit decreasing levels, NextC1 shows a comparable response (**Fig.2.6B**). Moreover, we examined NextC1 response upon inducible Oct4 depletion which leads to trophectodermal differentiation of ES cells. To do so, we used the Zhbtc4 cell line in which doxycycline (DOX) treatment results in rapid and complete extinction of Oct4 expression (Niwa et al., 2000). We remarked that upon acute loss of Oct4 (12h) the expression of NextC1 rapidly dropped until it got gradually abolished after two days of Oct4 absence while ES cells readily differentiate (**Fig.2.6C**). In addition, we performed another differentiation assay with retinoic acid treatment and LIF withdrawal over the course of three days driving mouse ES cells towards neuronal lineages (Bain et al., 1995; Fraichard et al., 1995; Strübing et al., 1995). Interestingly, we saw that NextC1 is rapidly downregulated in this context in our three WT ES cell lines (**Fig.2.6D**). Overall, these results collectively show that NextC1 follows the expression dynamics of other known stemness markers in culturing conditions that promote (2i) or impair (LIF withdrawal, Oct4 depletion, retinoic acid differentiation) self-renewal and pluripotency strongly suggesting the specific expression of NextC1 in naïve ES cells.

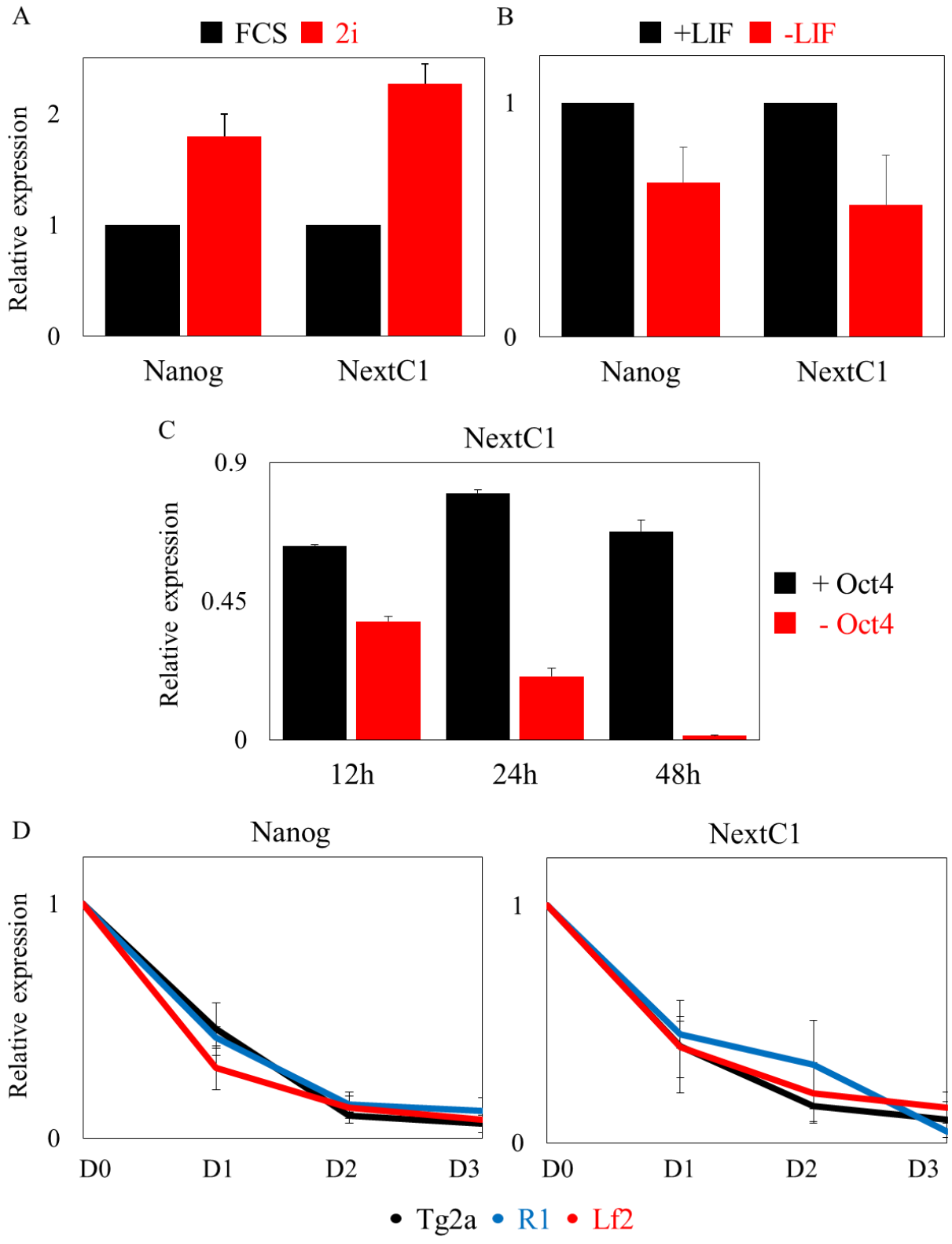


Figure 2.6. NextC1 expression and pluripotency. A) Nanog and NextC1 show increased expression levels in ES cells grown for 72h in 2i/LIF compared to FCS/LIF media. Values were normalized to Tbp mRNA levels and are expressed as the fold change to FCS (n=3). B) Nanog and NextC1 are downregulated upon 48h of LIF withdrawal. Values were normalized to Tbp mRNA levels and are expressed as the fold change to +LIF (n=4). C) NextC1 is progressively lost upon Oct4 depletion.

Values were normalized to Tbp mRNA levels (n=3). D) Nanog and NextC1 are downregulated upon retinoic acid differentiation (n=3, for each cell line). Data represent mean ± SEM from n biological replicates.

To more precisely define the pattern of NextC1 expression we assessed its RNA levels in well-characterized embryo-derived cell types representing the main lineages of the early embryo development. To begin with, we measured NextC1 expression in trophoblast stem cells (TS), extraembryonic endoderm stem cells (XEN), and mouse embryonic fibroblasts (MEF). RT-qPCR analysis further confirmed the exclusive expression of NextC1 in ES cells (**Fig.2.7A**). Additionally, we induced Epiblast-like cells (EpiLCs) differentiation from ES cells cultured in 2i medium. EpiLCs resemble the post-implantation epiblast of the embryo and represent the transition from naïve (2i) to primed pluripotency (Hayashi et al., 2011). Remarkably, while mouse EpiLCs are still considered as pluripotent cells, NextC1 expression drastically decreases in EpiLCs compared to 2i cultured ES cells (**Fig.2.7B**) or even to ES cultured in FCS/LIF (**Fig.2.7A**). All these results strongly support the assumption that NextC1 expression is restricted to pluripotent cells *in vitro*, and more precisely to the naïve state of pluripotency.

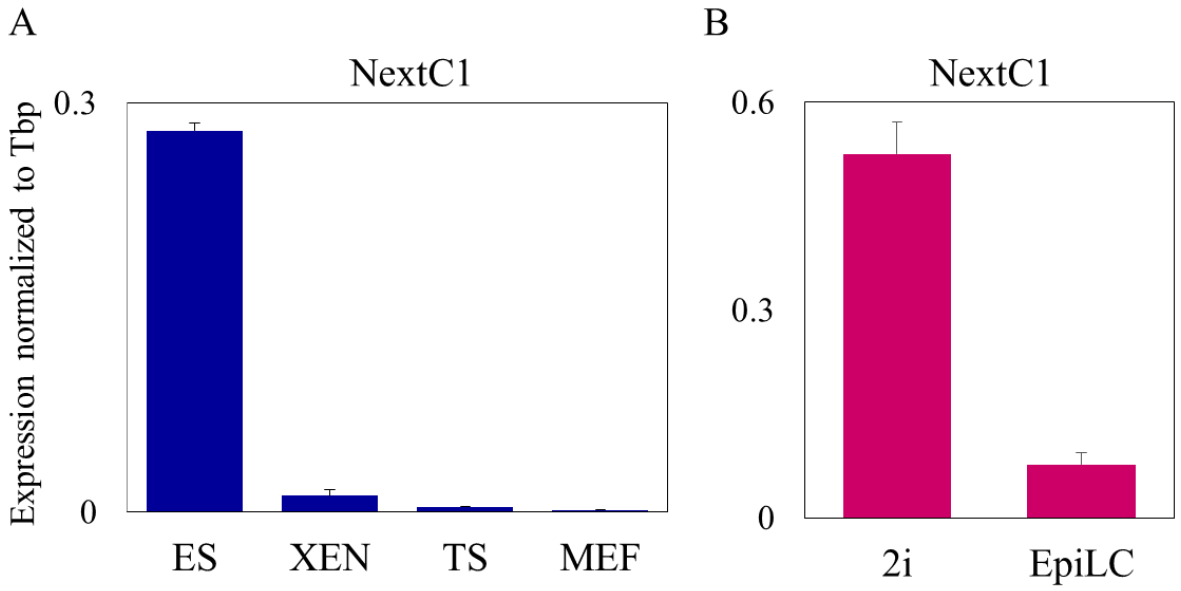


Figure 2.7. NextC1 in different embryonic cell types. A) NextC1 is expressed in Embryonic Stem cells (ES) (n=3) cultured in serum-containg medium but not in extraembryonic endoderm stem cells (XEN) (n=2), trophoblast stem cells (TS) (n=5) or mouse embryonic fibroblast cells (MEF) (n=3). B) NextC1 is strongly decreased in EpiLCs compared to 2i grown ES cells from which EpiLCs are induced after three days of treatment with ActivinA and FGF factors (n=5). Data are shown as means ± s.e.m. from n cell culture replicates.

Furthermore, in order to get insight about NextC1 expression *in vivo* during the early stages of mouse embryo development, we looked for available RNA-seq datasets conducted in different developmental stages between the zygotic and blastocyst stages. In a first RNA-seq study (Boroviak et al., 2015), we found NextC1 to be expressed mostly at ICM and pre-implantation Epiblast stages compared to morula and later stages of embryo development (not shown). We additionally analyzed the data of a single cell RNA-seq (Deng et al., 2014) that revealed the highest expression of NextC1 to be in mid– and late 2 cell stages, when the zygotic genome gets activated for the first time, and around the mid blastocyst stage which corresponds to the naïve mouse ES cells state (**Fig.2.8**). Moreover, NextC1 expression profile correlated at a much higher degree with pluripotency markers (corr. coeff.: Sox2:0.6; Nanog:0.32; Oct4:0.24) than with differentiation markers (corr. coeff.: Gata6:-0.05; Cdx2:-0.14) at the single cell level. Altogether, these data clearly indicate that NextC1 expression is associated *in vivo* with the naïve state of pluripotency.

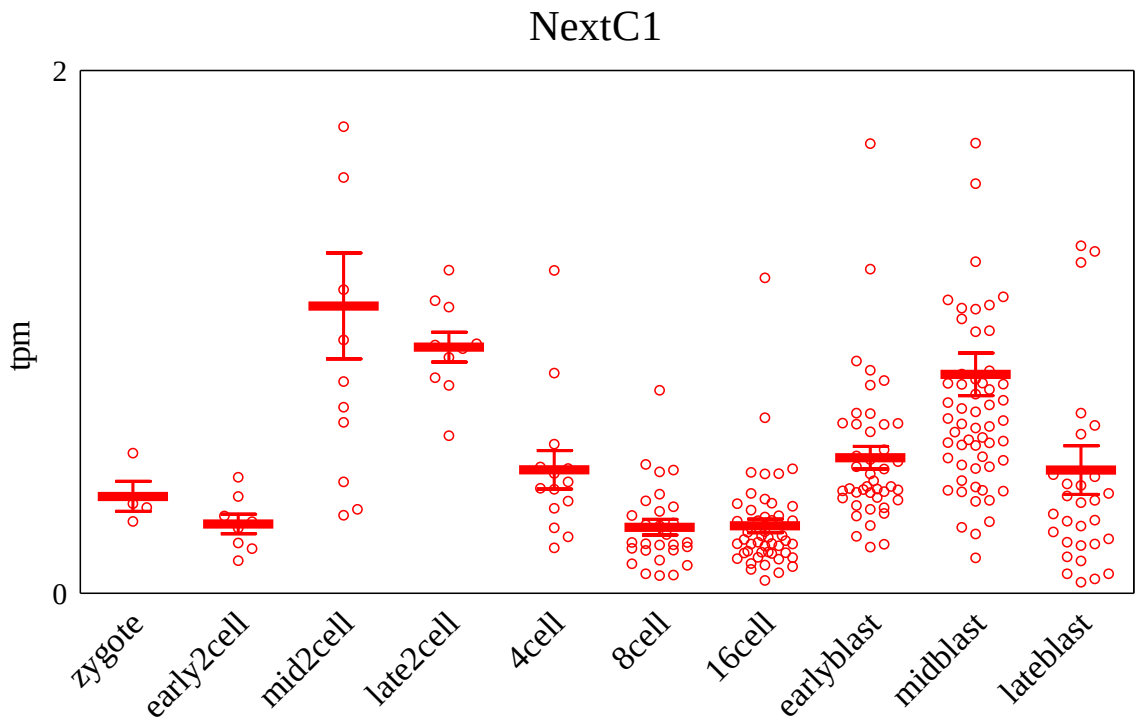


Figure 2.8. NextC1 expression revealed by single cell RNA-seq during mouse early embryo development. Transcripts per million (tpm) counts are shown for each cell at each developmental stage.

Finally, in order to assess the expression of NextC1 in later contexts *in vivo* we visualized available RNA-seq data performed in a multitude of different cell types and tissues at embryonic and adult stages in the mouse. The ENCODE/LICR RNA-seq track of the UCSC genome browser allowed us to look through diverse cell types and tissues in addition to ES

cells (RNA-seq data from Shen et al., 2012). Interestingly, NextC1 expression is strictly restricted to mouse ES cells with no detectable signal arising from any other dataset (**Fig.2.9**).

Collectively, all these results markedly point out to the conclusion that NextC1 expression is tightly related to ES cell identity along with known pluripotency factors raising the question of its potential role in the maintenance of the naïve ES cell state.

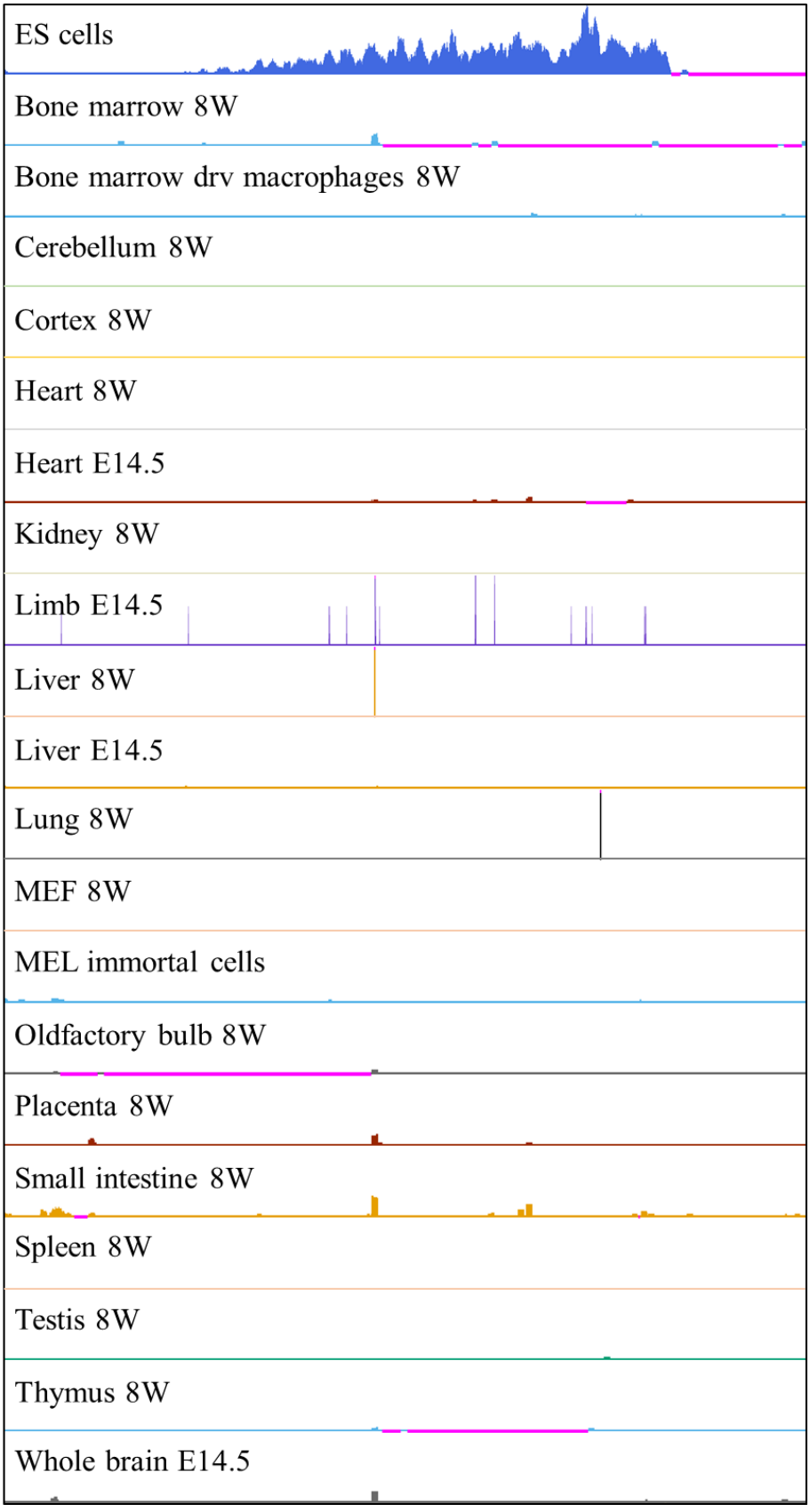


Figure 2.9. RNA-seq reads coverage of different cell types and tissues (not exhaustive) visualized on UCSC browser (ENCODE/LICR track) on NextC1 locus. 8W=adult week 8; E=embryonic day. NextC1 is detected exclusively in mouse Embryonic Stem cells.

E. NextC1 subcellular localization

We initially selected NextC1 as a potential lncRNA candidate to participate in the nuclear organization of mouse ES cells. However, an obvious prerequisite for this to happen is that the transcript shows a nuclear localization. To determine the cellular distribution of NextC1 transcript we first performed RT-qPCR in cytoplasmic and nuclear fractions after cell fractionation assays. We attested that NextC1 has a high enrichment in the nuclear compartment compared to the cytoplasmic fraction similar to that of Malat and Neat1 transcripts that serve as the epitome of nuclear localized RNAs. In parallel, effective fractionation was demonstrated by the respective nuclear and cytoplasmic enrichment of pre-mRNAs (intronic primers were used) and mature mRNAs of Nanog and Klf4 transcripts (**Fig.2.10**).

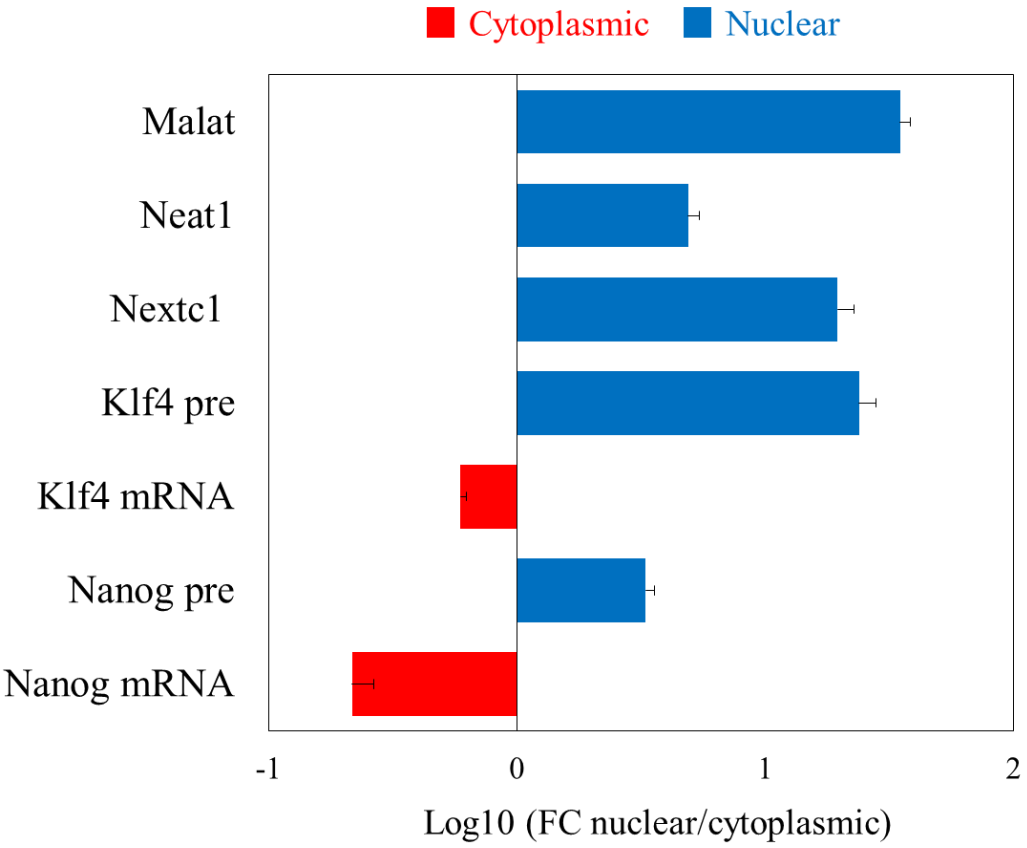


Figure 2.10. Cell fractionation analysis by RT-qPCR. Primary transcripts and nuclear lncRNAs are enriched in the nucleus whereas spliced transcripts of TFs are enriched in the cytoplasm. Data are represented as ratio of nuclear to cytoplasmic expression levels in log10 scale (n=2, ±SEM).

To further examine the subcellular localization of NextC1 we performed RNA-FISH on WT mouse ES cells with a fosmid clone generated probe. The generated probe (~38kb) covered a large part of the NextC1 locus (~58kb) (**Fig.2.11A**). RNA-FISH analysis confirmed an exclusively nuclear and mostly focal localization of NextC1 in the majority of the cell population in undifferentiated male and female ES cells (**Fig.2.11**). Moreover, a sequential RNA/DNA FISH carried out after nuclear matrix preparation authenticated NextC1 RNA retention *on site* while its genomic locus was digested by DNaseI treatment (**Fig.2.12**). In addition, we further validated the loss of NextC1 expression upon targeted differentiation with retinoic acid treatment (**Fig.2.11C**). Indeed, in differentiating female cells, we could see that within the cells that have fully differentiated and have formed the Xist cloud, NextC1 is no longer detectable whereas in cells that probably have not undergone differentiation yet, NextC1 is detected while no Xist domain is visible (**Fig.2.11C**). In undifferentiated ES cells, NextC1 expression was detected in >80% of the cells with a strong focalization around its transcription site, extending beyond its genomic locus, as attested by RNA/DNA-FISH (**Fig.2.12**). Of the NextC1 positive cells, around 80% express NextC1 in a biallelic manner which leads to the detection of two (or four in cells that are in G2 phase of the cell cycle) bright foci corresponding to the two active transcription sites. Around 20% of NextC1 positive cells express it from a single allele and in the rest <5% of cells, we detect NextC1 as diffusible molecules scattered in the nucleus with no visible active transcription site (**Fig.2.11D**). The nuclear localization of NextC1, which is mostly focalizing around is transcription sites, is strongly reminiscent to those of Firre and Charme lncRNAs. Interestingly, these two lncRNAs contribute to the orchestration of tridimensional domains by establishing long-range chromosomal interactions that result in specific gene activation (Ballarino et al., 2018; Hacisuleyman et al., 2014).

Collectively, all the data analyzed so far, suggest that NextC1 is an abundant, naïve ES cell specific, lncRNA exerting strong transcription foci in the nucleus. To determine its functional relevance, we set up a series of loss- and gain-of-function assays to gain insight into any phenotypical or molecular consequences of such NextC1 manipulation.

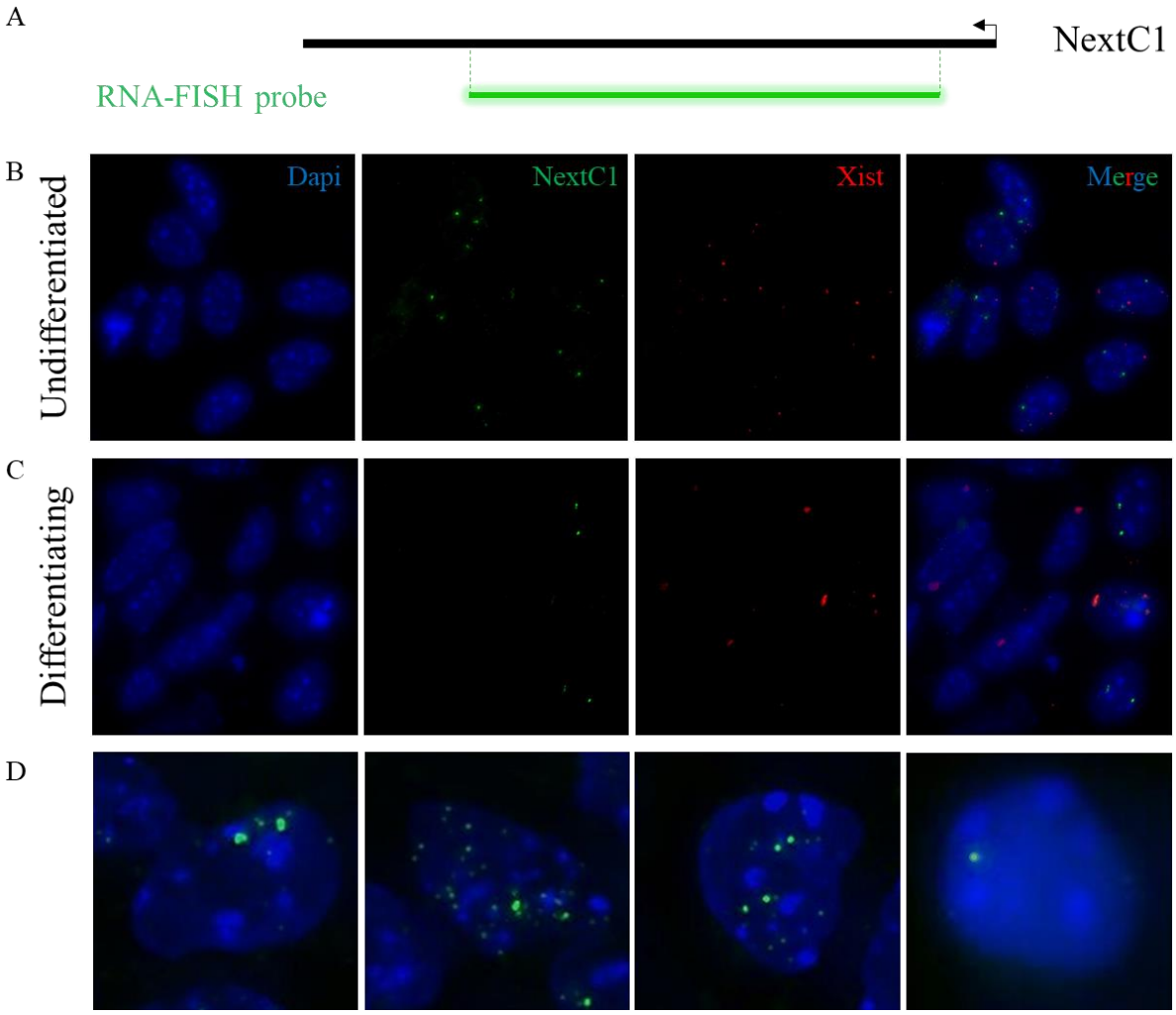


Figure 2.11. NextC1 subcellular localization revealed by RNA-FISH. A) RNA-FISH probe was generated from a fosmid covering the biggest part of NextC1 genomic sequence. B) The majority of undifferentiated Lf2 ES cells express NextC1 (green) in a biallelic manner with an accumulation at the transcription sites; Xist (red) is seen as pinpoints at its transcription sites. C) In differentiating Lf2 cells when Xist forms a cloud, NextC1 is not detectable and in cells that probably are not yet differentiated, NextC1 is detected and Xist is visible as pinpoints. D) Different patterns of NextC1 nuclear distribution; from left to right: both alleles active with local accumulation of the transcript, both transcription foci with scattered molecules in the nucleoplasm, four transcription foci of probably G2 phase cells, single active allele with local accumulation of NextC1. DNA is stained with DAPI.

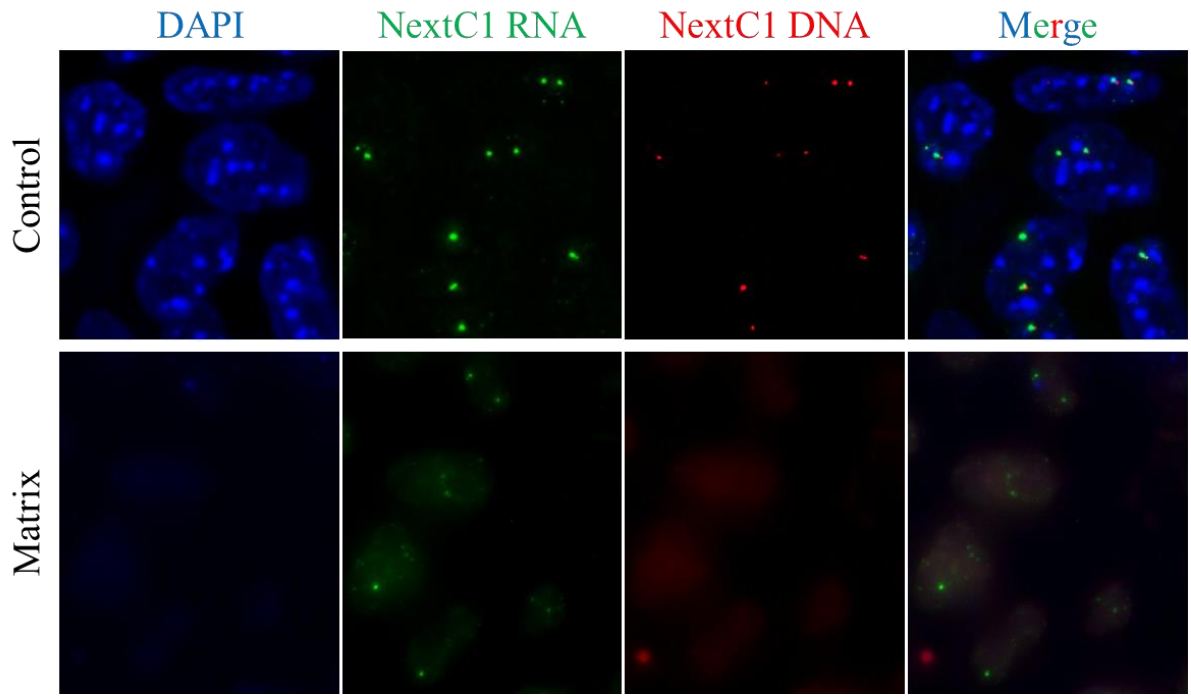


Figure 2.12. NextC1 is accumulating at its transcription sites and is retained in extracted nuclei as seen by RNA/DNA-FISH. In control sample (upper panel) NextC1 RNA (green) is colocalizing with NextC1 genomic locus (red). In matrix prep (lower panel) NextC1 RNA is detected even when DNA has been efficiently digested and removed. DNA is counterstained with DAPI (blue).

F. Functional assays

i. Loss of function

Promoter deletion with CRISPR/Cas9 system

We first decided to address the functionality of NextC1 by a loss-of-function assay using the CRISPR/Cas9 toolkit. In order not to interfere massively with the genomic context of NextC1 locus and the regulatory DNA elements it might contain by deleting the full NextC1 transcription unit (58kb) we decided to perform a rather short deletion within its promoter region that would eliminate its downstream expression. We therefore generated a male NextC1 promoter-knockout ES cell line by deleting a 454bp long genomic region, where a multitude of pluripotency TFs have been shown to bind in ES cells (**Fig.2.13A, Fig.2.5**). Two single guide RNAs (sgRNA) surrounding this common TFs binding site were designed in that region using the online CRISPR Design Tool (http://crispr.mit.edu/) (Hsu et al., 2013) (**Fig.2.13A**). Plasmids with the two sgRNAs and the Cas9-GFP transgene were lipofectamine-transfected in

Tg2a ES cells and first validated by PCR on genomic DNA to efficiently produce deleted allele on a batch of cells. After this validation step, we repeated the transfection and selected the properly transfected cells by FACS sorting high GFP-positive cells (GFP expression is linked to Cas9 transgene) and puromycin selection (puromycin resistance was present in the sgRNA plasmids). We acquired single cells from FACS sorting, cultured them and obtained sixty five clones which were further screened by genomic PCR. Eight homozygous mutant clones were identified by PCR to carry the expected size deletion of about 454bp. The PCR products were inserted in bacterial vectors and analyzed by Sanger sequencing (10 bacterial clones sequenced per KO clone). This revealed a slightly larger deletion in some of our KO clones but confirmed their mutated status (**Fig.2.13B)**. Surprisingly, we found a unique sequencing profile for each clone suggesting that both alleles bear the exact same mutation. This might be explained by Homology Directed Repair occurring between both alleles after Cas9 targeting such that the first deleted allele serves as a repairing matrix for the second one producing identical mutant alleles.

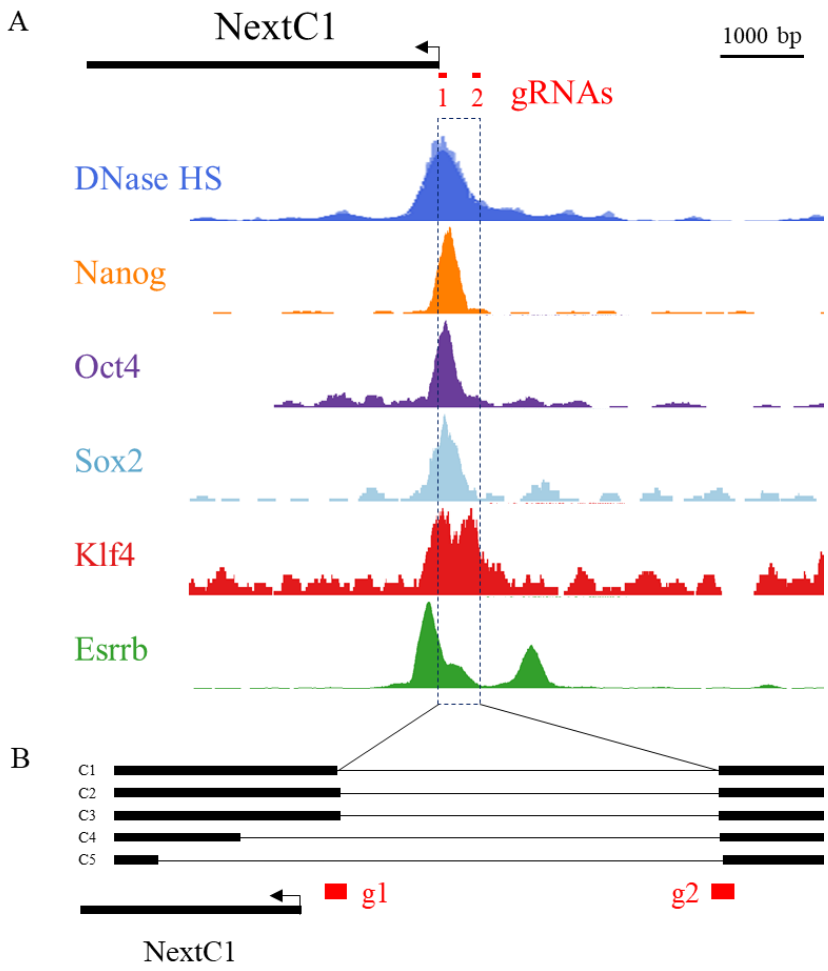


Figure 2.13. NextC1 promoter deletion with CRISPR/Cas9 system. A) Localization of the designed sgRNAs (g1, g2) in respect to the promoter of NextC1 and to the binding sites of few pluripotency TFs. Highlighted in black box is the targeted region with g1 and g2. B) PCR amplicons of the five generated homozygous mutant clones were sequenced; deleted sequence depicted as a thin line. g1, g2 and the promoter of NextC1 gene are also shown.

The karyotype of the eight mutant clones was checked and showed regular modal number of forty chromosomes for five clones that we thus kept for further analyses. The expression level of NextC1 was measured by RT-qPCR in the generated homozygous mutants (named thereafter C1 to C5) cultured in parallel with wild-type (WT) cells in FCS and 2i conditions. The results showed that NextC1 transcription significantly decreased to almost non-detectable levels in all the KO clones (**Fig.2.14A)** despite a slight increase of signal in 2i medium, most likely due to the reinforcement of the remaining pluripotency TF binding around NextC1 promoter leading to negligible activation. RNA-FISH was performed on these clones, to further validate NextC1 depletion showing very rare (<10%) and weak transcription ongoing in the KOs clones (**Fig.2.14B)**. The mutant clones did not show any morphological difference when compared to WT ES cells in either the FCS nor 2i conditions.

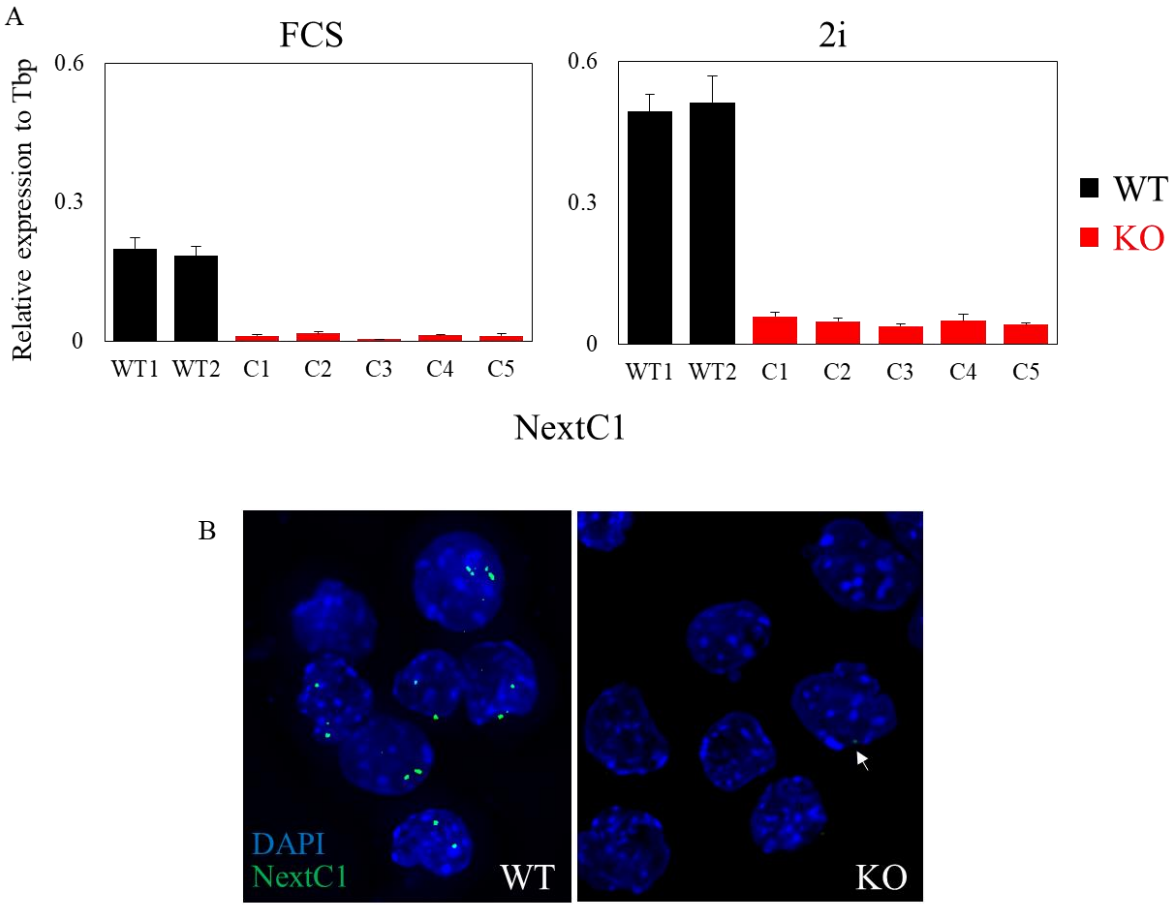


Figure 2.14. NextC1 expression in mutant clones. A) Representative RT-qPCR graphs for NextC1 expression in WT and mutant clones in FCS (left) and 2i (right) cultures. Data are normalized to Tbp mRNA and are shown as means and ±sem of nine independent experiments. B) RNA-FISH of NextC1 (green) in a WT and a KO clone verified the suppressed expression in the later. White arrow shows a weak signal of NextC1 expression that is found in <10% of KO cells. DNA is stained with DAPI.

Given, the restricted expression of NextC1 to the naïve pluripotent state, we first decided to assess the potential impact of NextC1 depletion on ES cell self-renewal. To this end, we performed clonal growth assay followed by alkaline phosphatase staining (AP staining) on the WT and KO clones in self-renewing (+LIF) or differentiation (-LIF) conditions. After six days of growth at clonal density, cells were fixed and stained for AP activity (specific of the undifferentiated colonies) and fully stained, mixed and unstained colonies were manually counted. We did not notice any significant difference in the number or size of the AP positive colonies consistent in all the KO clones compared to WT cells in the +LIF condition. In –LIF condition, no purely undifferentiated colonies were detected in neither of the WT or KO cells while the ratio between mixed and fully differentiated colonies was similar for the two types of cells (**Fig.2.15A**). This result suggests that NextC1 loss has no major effect on the self-renewal ability of ES cells. Nevertheless, we sought to evaluate the expression levels of few

pluripotency and differentiation markers in the KO clones in FCS and 2i conditions. The expression of pluripotency markers (as a representative example Nanog is shown, **Fig.2.15B**) were found to be comparable in WT and KO cells in both conditions. Differentiation markers show low and inconsistent expression between WT and KO cells in both conditions as exemplified by Brachyury gene, in **Fig.2.15C**. Therefore, we could not conclude that NextC1 depletion had any strong impact on the self-renewal and pluripotent status of mouse ES cells in FCS or 2i conditions.

Second, we hypothesized that a potential effect of NextC1 absence in the undifferentiated state might be revealed in the differentiation potential of our KO cells. Thus, we decided to study their capacity to give rise to the three germ layers; ectoderm, mesoderm and endoderm. To that end, we performed an Embryoid Body (EB) differentiation assay and assessed by RT-qPCR the expression of differentiation markers of the three lineages (three markers of each), and the downregulation of pluripotency markers. We observed the downregulation of NextC1 in the WT cells over the days of the differentiation and confirmed its absence of expression in the mutant cells. Pluripotency markers got similarly silenced during this assay in the WT and KO clones. Moreover, the differentiation potential in the three germ layers did not seem to be impaired in our KO clones as each marker of the different lineages got respectively induced (**Fig.2.16**). However, we could observe that the NextC1 KO clones showed a mild delay of induction of the endoderm markers (Dab2 and Gata6) at day 6 during the differentiation assay that got compensated at day 10.

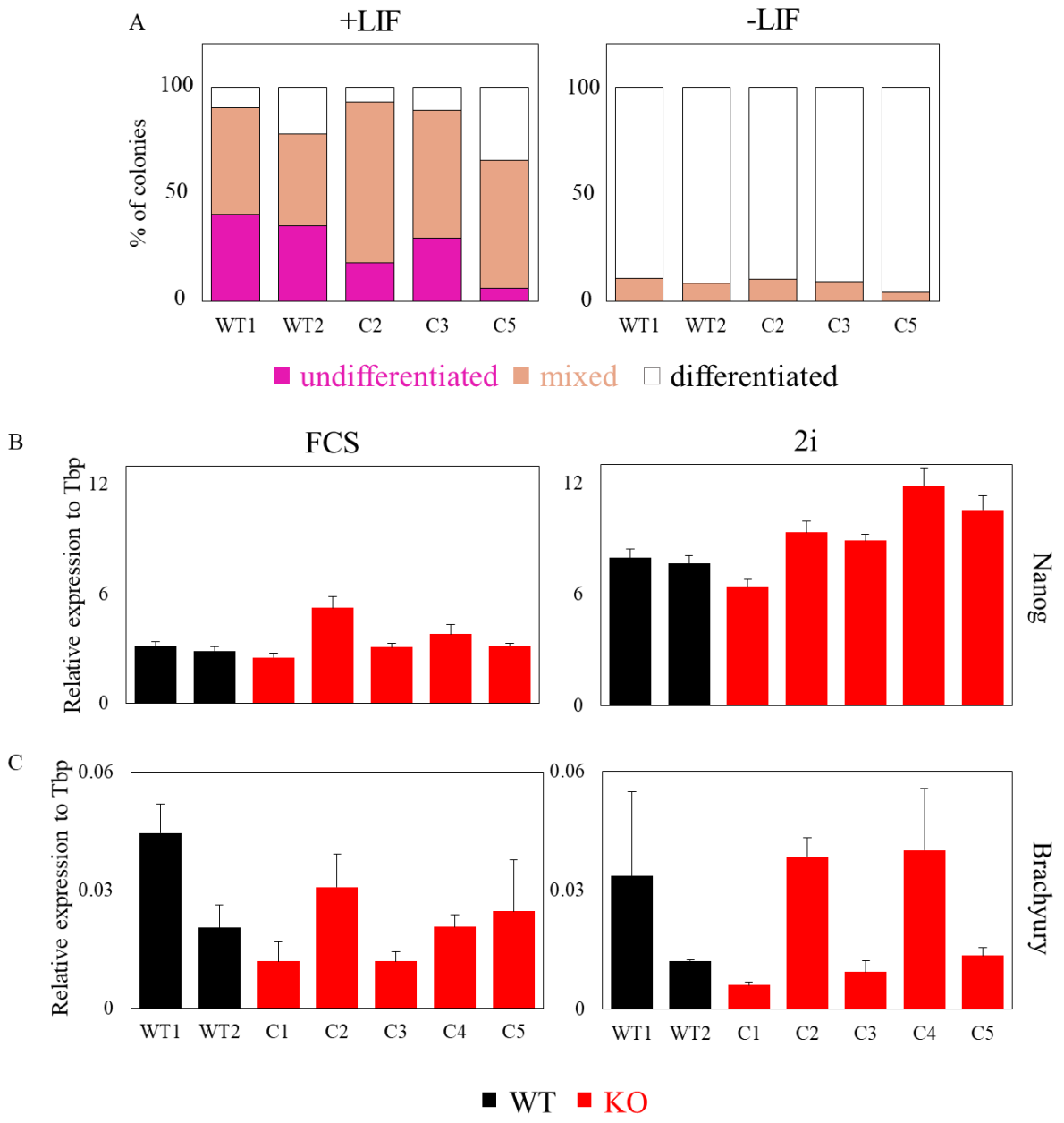


Figure 2.15. A) Alkaline phosphatase staining in NextC1 mutant clones. Diagram representing the percentage of undifferentiated (pink), mixed (orange) and differentiated colonies in each clone, in cultures with or without LIF. B-C) RT-qPCR analysis of Nanog (B) and Brachyury (C) expression levels in WT and KO clones in FCS (left) and 2i (right) conditions. Error bars represent the mean ± SEM of nine independent experiments.

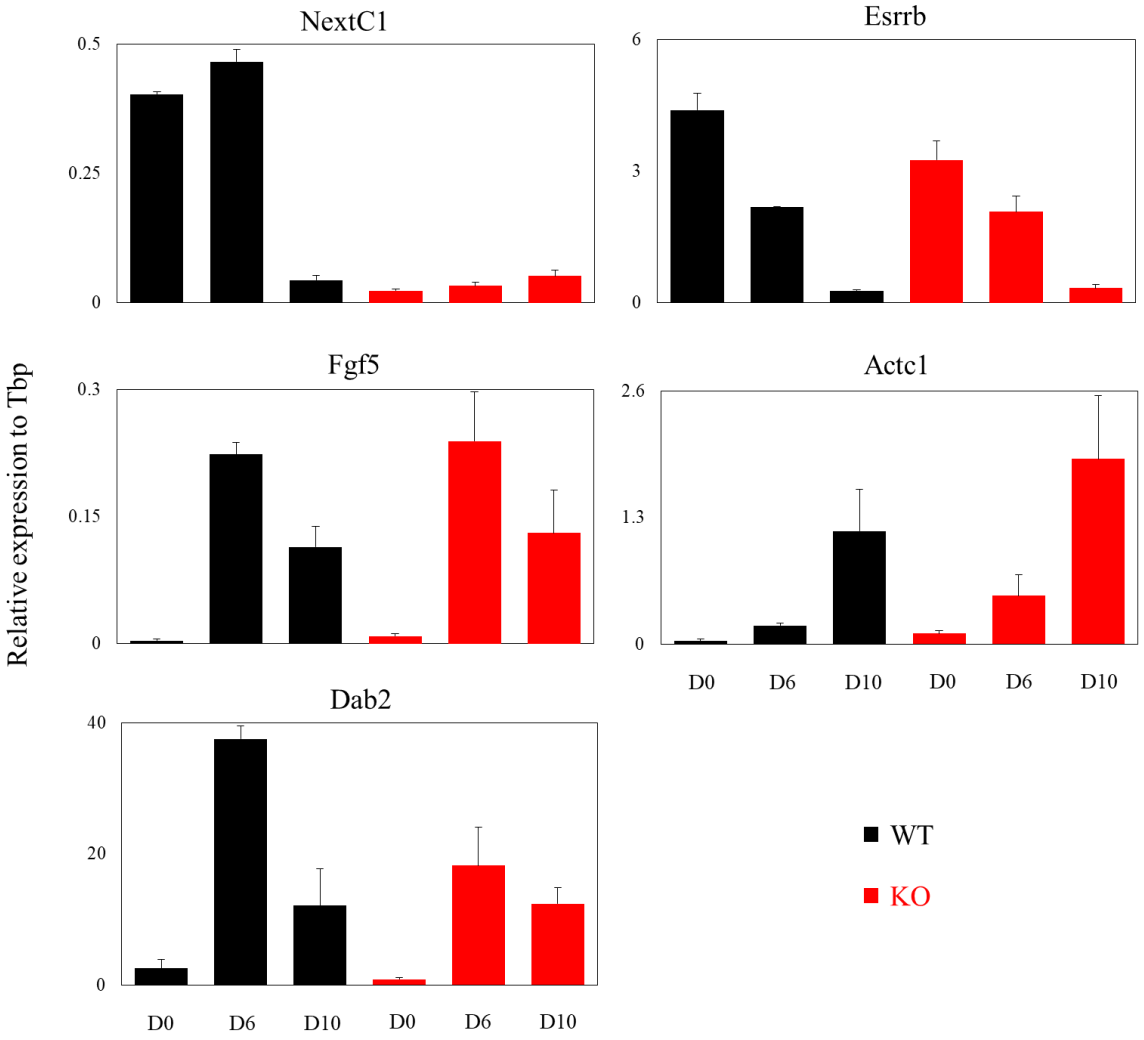


Figure 2.16. Differentiation capacity of NextC1 mutants. RT-qPCR analysis of NextC1, pluripotency (Esrrb) and differentiation markers (ectoderm: Fgf5; mesoderm: Actc1; endoderm: Dab2) in WT and NextC1 KO ES cells during Embryoid Bodies differentiation. The time points represent undifferentiated ES colonies (D0), Embryoid Bodies plating (D6) and multilineage progenitors (D10). Data are shown normalized to Tbp. Error bars represent the mean ± SEM of two WT clones and five KO clones of a single experiment.

First, we examined the possibility that the proliferation rate would be affected in our NextC1 mutant clones. For that, we assessed the growth rate of the KO clones in FCS and 2i conditions by plating four times iteratively the same number of cells and counting the obtained number of grown cells at day three. In FCS condition, we noticed that the number of cells we recovered at the time of passage was similar in WT and KO cells. However, in 2i condition we observed a slight retardation of the KO clones (apart from one) compared to WT cells (**Fig.2.17**). Given the fact that NextC1 expression is increased in 2i condition, it is be possible that an effect on cell proliferation might be visible only in this particular medium.

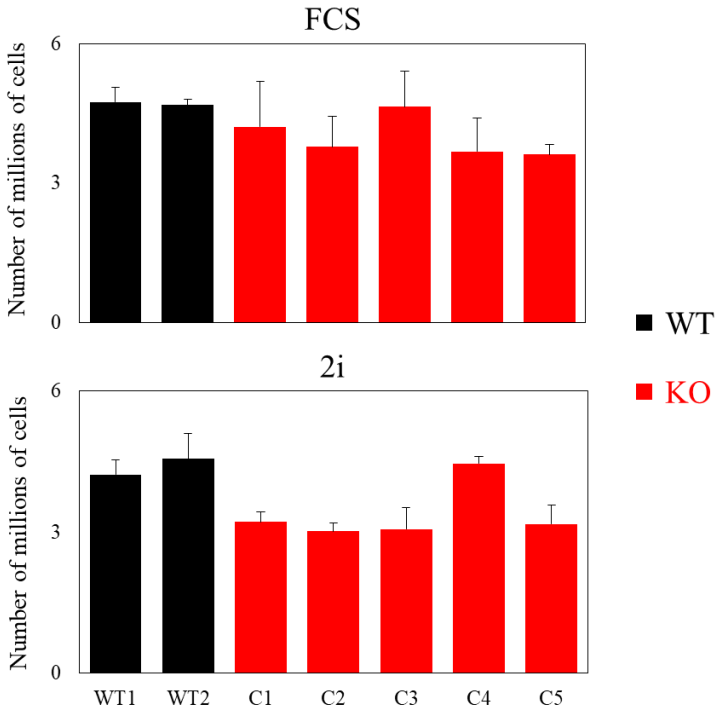


Figure 2.17. Proliferation rate of NextC1 mutants in FCS and 2i culturing conditions. Growth rate was assessed by counting the cells at each passage and replating the same number of cells. Data are represented by number of millions of cells as means and ±SEM of four countings in FCS and 2i conditions.

On one hand, due to the histone modification marks present at NextC1 promoter (of enhancer function) and on the other hand, with the knowledge that lncRNAs act often in *cis* in regulating expression of neighboring genes (Ørom et al., 2010; Paralkar et al., 2016; Yin et al., 2015) we sought to explore the effect of NextC1 depletion on the expression of the genes located in its proximity. More specifically, we looked by RT-qPCR analysis the expression levels of genes that are located up to 300kb away from the NextC1 locus within the same topologically associated domain (TAD). In order to define the TAD where NextC1 is lying we used publicly available Hi-C data obtained from the 3D genome browser of the Yen Lab (http://promoter.bx.psu.edu/hi-c/view.php) (data for mouse ES cells from (Dixon et al., 2012)). Surprisingly, none of the seven tested neighboring genes showed any kind of modification upon NextC1 loss of expression (see two of them in **Fig.2.18**). This suggests that, despite the local accumulation of NextC1 RNA at its transcription site, neither the NextC1 RNA nor its promoter sequence are participating in the in *cis* regulation of expression at least of the genes that could interact and/or be in close proximity with the NextC1 locus. This indicates that NextC1 might be involved in long range distance interactions as demonstrated for the aforementioned Firre lncRNA.

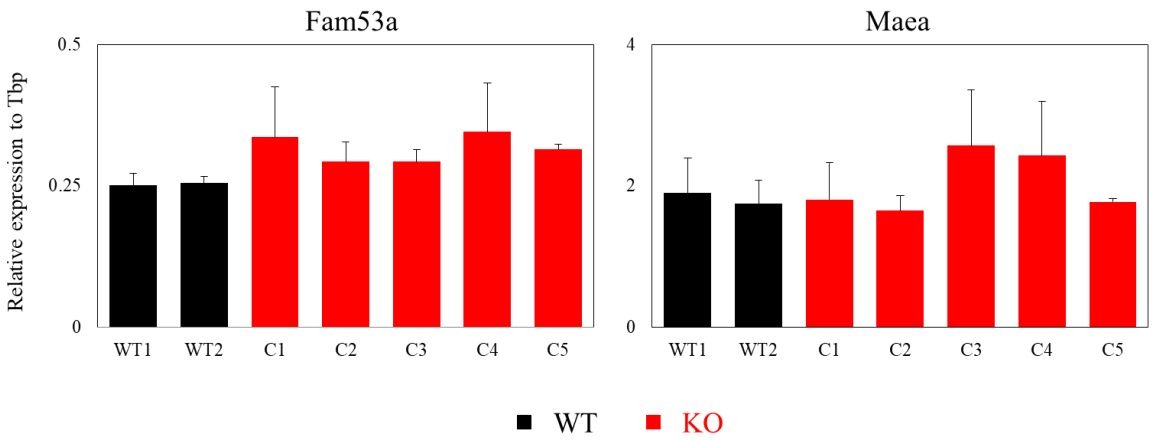


Figure 2.18. Effect of NextC1 loss of expression on neighboring genes located in the same TAD. RT-qPCR analysis on two genes flanking NextC1 locus. Data are normalized to Tbp mRNA and shown as means and ±SEM of three biological replicates.

With the experiments and the analysis performed so far on NextC1 mutant clones we could only target few genes at a time in each condition to assess the molecular consequence of NextC1 depletion. In order to assess the transcriptional consequences of the loss-of-function assay on a genome-wide level, we proceeded by conducting RNA-seq on two WT and five KO clones cultured in FCS, 2i and EB differentiation (at day 6). The RNA-seq analysis validated the depletion of NextC1 expression in all the KO clones **(Fig.2.19)**. We next performed a statistical analysis with DESeq program to identify the differentially expressed genes (DEGs) upon loss of NextC1expression. We identified 752, 842, 596 and 2195 genes being differentially expressed in FCS, 2i (short and long term) and EB conditions respectively (FDR<0.05). The analysis of these genes is currently ongoing in order to decipher any functional potential of NextC1 gene.

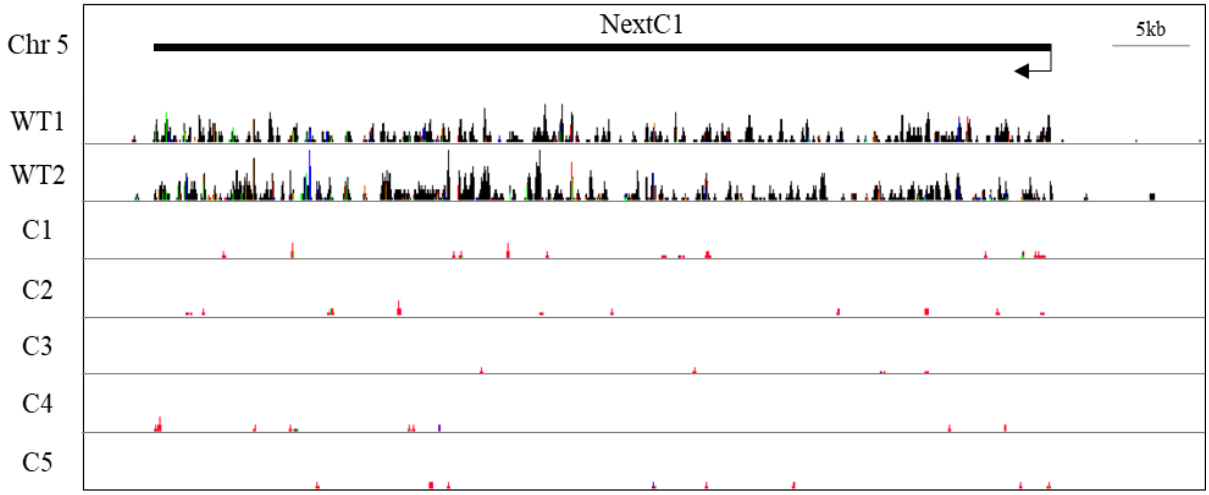


Figure 2.19. Screenshot of IGV browser of NextC1 locus. RNA-seq coverage in WT (black) and mutant clones (red) samples. Reads coverage have been group-auto scaled. No expression of NextC1 is detected in any of the five KO clones.

Inducible transcript knockdown by CRISPRi

The previously described technique of the promoter deletion of NextC1 and the RNA-seq on the generated mutant clones should unmask genes that get misregulated upon this loss-of-function approach. The misregulated identified genes might either result from the deletion of the locus and/or the absence of a functional NextC1 transcript. To be able to distinguish between these two scenarios, we generated a cell line where an inducible knockdown of NextC1 would be possible without intervening with the genomic locus. To that end, we used the CRISPR interference (CRISPRi) system in which a catalytically inactive Cas9 (dCas9) protein is fussed to a repressive chromatin modifier KRAB domain (Gilbert et al., 2013, 2014). Single guide RNAs (sgRNA) were designed in a window of -50bp to +300bp in respect to the TSS of NextC1 using the online CRISPR Design Tool (http://crispr.mit.edu/) (**Fig.2.20A**) as this has been shown to be the optimal region to target for CRISPRi inhibition (Gilbert et al., 2014). We initially tested the efficiency of the three designed sgRNAs by co-transfecting them with the dCas9-KRAB-BFP transgene that is expressed under the control of a Doxycycline (Dox) inducible promoter and linked to a blue fluorescent protein (BFP). We observed that on the bulk of the transfected cells, gA and gC efficiently repressed NextC1 transcription, with a decrease in expression of more than 50% upon DOX treatment (**Fig.2.20B**). Afterwards, forty eight clones were manually picked for each transfection with guide RNAs gA, gC and a pool of gA-gB-gC. Five clones of each transfected bulk of cells were selected based on strong BFP signal under DOX induction and RT-qPCR analysis was done in order to sub select three clones that show an efficient downregulation of NextC1 expression up to >90% to almost undetectable

levels (**Fig.2.20C**). The generated clones will be used for the validation of the DEGs identified by the RNA-seq of the mutant clones of NextC1 promoter deletion (detailed in the previous section) and the determination on the functionality of the transcript or promoter locus of NextC1 gene.

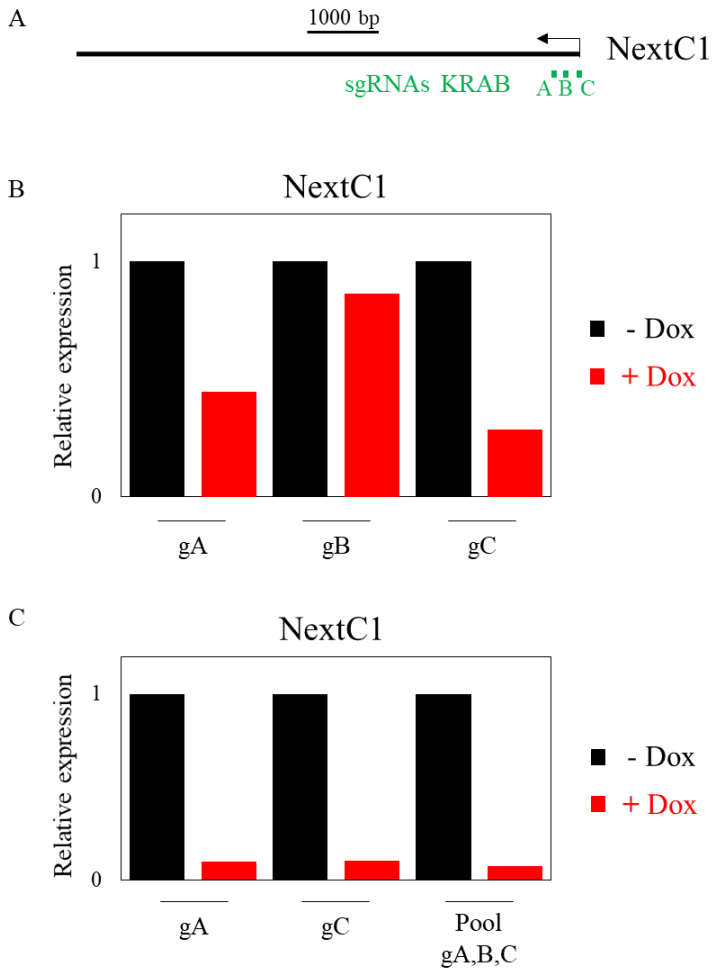


Figure 2.20. CRISPRi for NextC1 knockdown. A) Localization of three designed sgRNAs (gA, gB, gC) in respect to the promoter of NextC1 (+45bp to +340bp). B-C) RTq-PCR analysis on NextC1 expression upon DOX treatment (+DOX, 72h). Data are normalized to Tbp mRNA and shown as fold change to untreated samples (-DOX). B) Bulk of transfected ES cells with gA, gB, gC. C) Picked clones that shown the best KD effect using gA, gC or pool of gA-gB-gC.

ii. Gain of function

In order to untangle the functional relevance of NextC1 we also followed a gain-of-function approach in parallel to the loss-of-function assay. When studying a lncRNA function, it is important to be able to recapitulate the local activity of the RNA molecule especially for transcripts acting in *cis* or when the mere act of transcription is important for its function

(Engreitz et al., 2016; Li et al., 2013; Ørom et al., 2010). In that aim, we used the SunTag CRISPR activation system (Gilbert et al., 2014; Tanenbaum et al., 2014) in which the enzymatically deficient Cas9 (dCas9) protein recruits through an epitope-antibody couple many copies of a potent trans-activator domain (VP64) that enables the overexpression of a given gene from its endogenous locus. The SunTag system components have been adapted by a colleague in the lab who had generated two stable clones (c1 and c2 in **Fig.2.21C**) expressing the CRISPR SunTag system upon DOX induction in ES cells (Heurtier et al., 2018).

For the endogenous overexpression of NextC1 we used the two sgRNAs previously mentioned for the deletion assay (section **II.F.i**) as well as an additional one between those two, all of them being in the optimal window of proximity to the TSS of the targeted gene (-50bp to -400bp upstream to the TSS) for an efficient activation (Gilbert et al., 2014) (**Fig.2.21A**). At first, we tested the efficiency of the three sgRNAs in one of the SunTag clones. RT-qPCR analysis was performed on transfected cells with each sgRNA separately and DOX treatment of 24h, 48h or 72h. We observed that the sgRNA that was located closer to the TSS (g1) could not trigger the overexpression of NetxC1 (**Fig.2.21B**) and that the upregulation was gradual over the course of the DOX induction peaking at 72h (data not shown). Therefore, we used both working sgRNAs (g2 and g3) in the two SunTag (C1, C2) clones to exclude clone or sgRNA off-target artefactual effects and generated batches of C1g2, C1g3, C2g3 cells. For each batch we picked 12 clones, monitored their GFP and BFP signals (GFP and BFP are linked to the different components of the SunTag system, VP64 and dCas9 respectively) upon DOX treatment and proceeded to RT-qPCR analysis of the five clones with the highest GFP/BGP signal per batch (**Fig.2.21C**). We finally selected the single subclone per batch of cells that reached the highest induction of NextC1 (highlighted in **Fig.2.21C**) which thereafter will be simply called C1g2, C1g3, and C2g3.

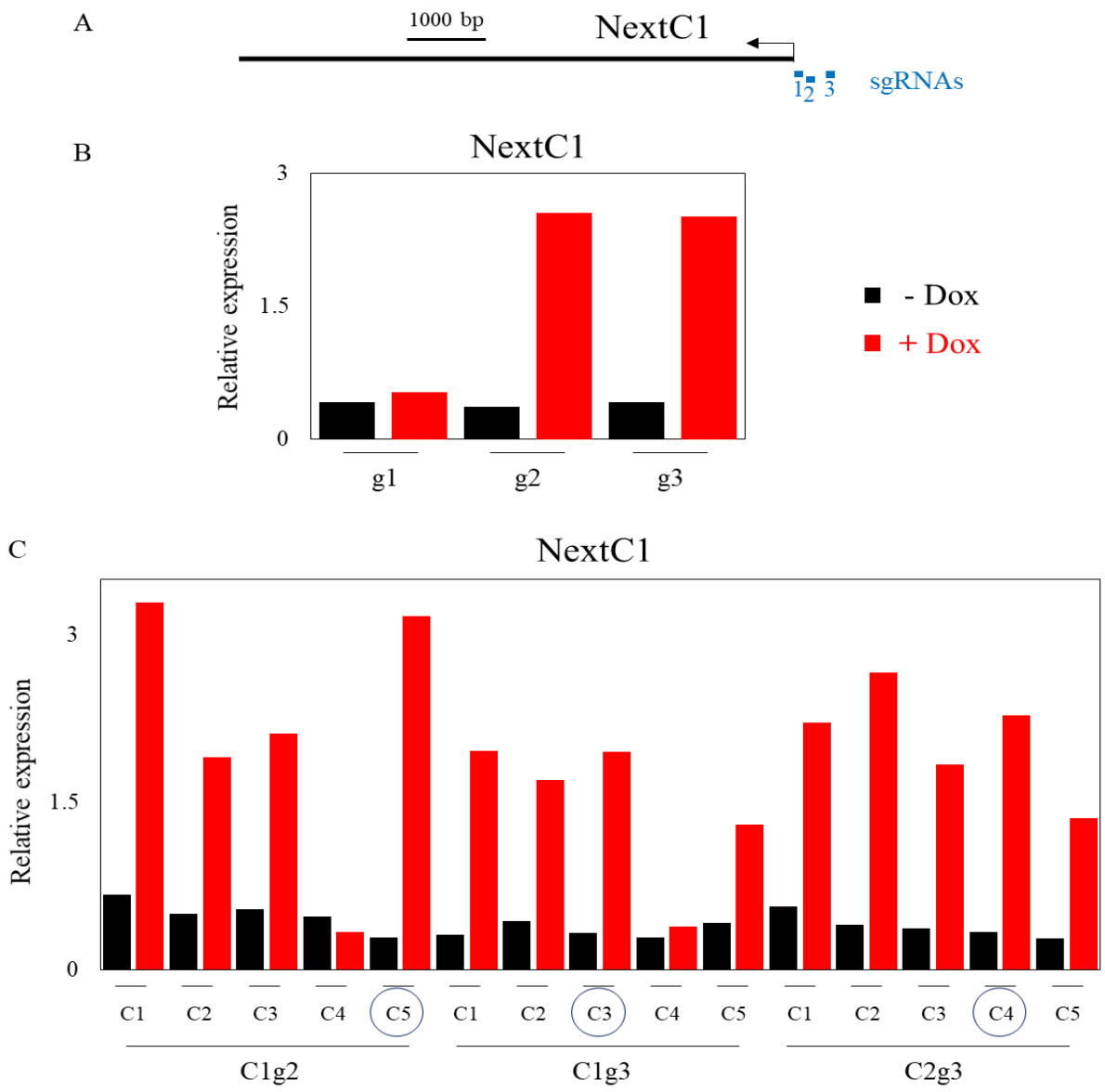


Figure 2.21. NextC1 endogenous overexpression with CRISPRa SunTag system. A) Localization of three sgRNAs (g1, g2, g3) in respect to the promoter of NextC1 (-150bp to -580bp). B) RTq-PCR analysis of NextC1 expression with the three sgRNAs upon DOX treatment (+DOX, 72h). Data are normalized to Tbp mRNA and shown as fold change to untreated samples (-DOX). C) RT-qPCR analysis of NextC1 expression of picked clones. Final clone selection for further studies is highlighted: C5 for C1g2, C3 for C1g3, and C4 for C2g3.

Before proceeding with functional assays on the generated NextC1 overexpressing clones, we wanted first to determine the time point at which the induction of the activation would be the highest since it would be more practical to unveil the effect of NextC1 overexpression at the moment that it is the most highly expressed. Therefore, we performed kinetics of DOX induction over five days and at three passages (nine days) and established that NextC1 expression reaches its zenith at 72h of DOX treatment (**Fig.2.22**). Onwards, the cells were always induced with DOX for 72h.

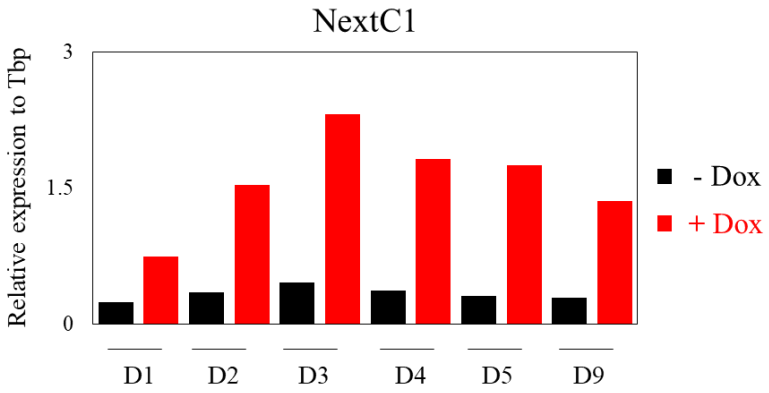


Figure 2.22. Kinetics of DOX induction to determine the best time point of NextC1 upregulation. Data are normalized to Tbp mRNA.

In addition, we checked that NextC1 levels upon DOX induction are comparable in the three clones that we used. We observed that we achieved three fold higher expression in DOX treated cells for the three used clones while in the two parental SunTag clones (that have not received any sgRNA) NextC1 remains at basal expression level (**Fig.2.23A**). We have previously observed the strong accumulation of NextC1 transcripts around its transcription sites in ES cells by RNA-FISH, so given the upregulation we attained with the SunTag system we wondered what would be the consequence on the subnuclear localization of the transcript. We estimated by RNA-FISH that upon overexpression, NextC1 RNA molecules remain mostly at the transcription sites. We observed that these foci tend to be brighter and bigger in the DOX treated cells compared to the non-treated (**Fig.2.23B**). Interestingly, we did not find substantial changes in the percentage of cells actively expressing NextC1 between the two conditions. This suggests that the overexpression affects the cells in which NextC1 is already active meaning that the SunTag activation increases the size of the transcriptional bursts rather than their frequency.

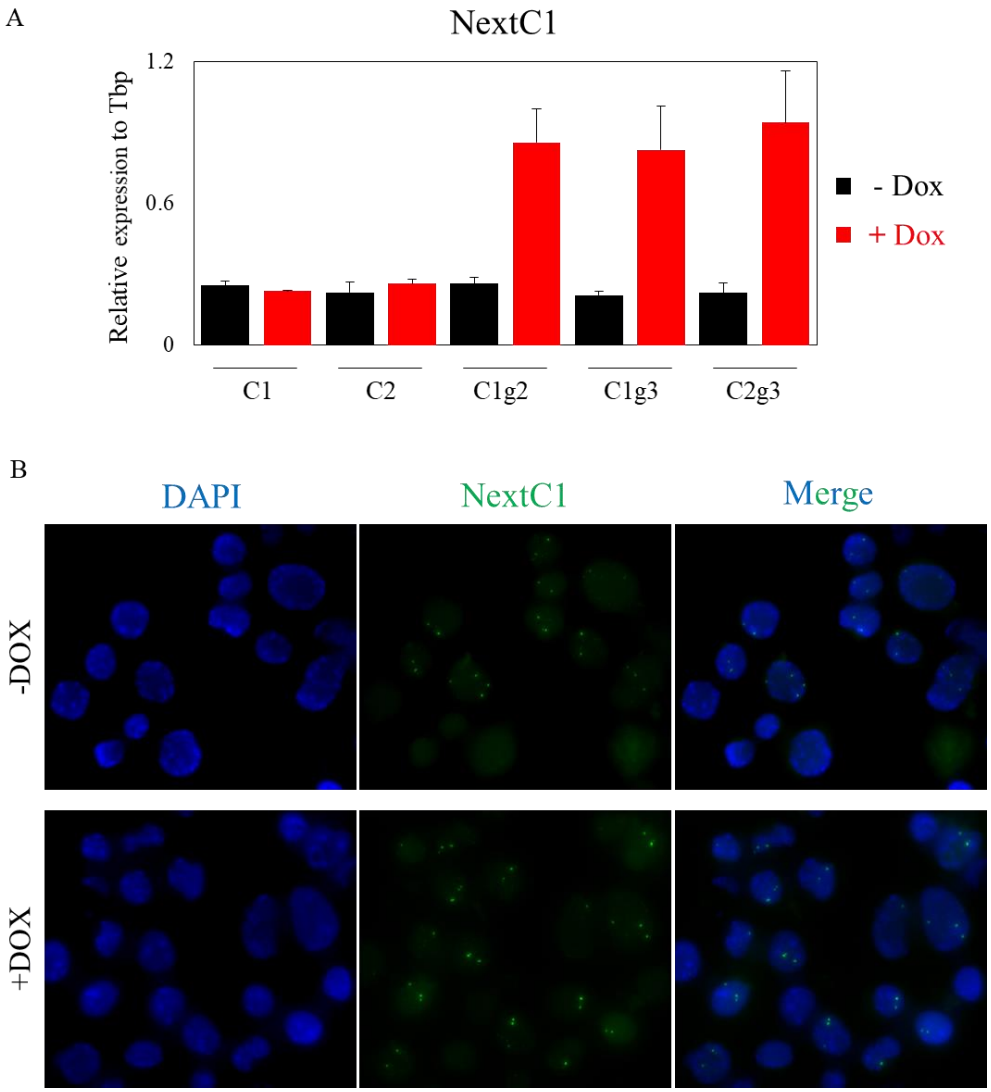


Figure 2.23. NextC1 activation assessment by RT-qPCR and RNA-FISH. A) RT-qPCR analysis on the two control (without sgRNA) SunTag clones and the three NextC1 targeted clones. Data normalized to Tbp and shown as means and ±SEM of three biological replicates. B) RNA-FISH in C1g3 clone in DOX untreated and treated cells confirms the overexpression of NextC1 by the visibly brighter and bigger transcriptional foci.

Then, we estimated the effect of NextC1 activation on ES cell self-renewal. For that, AP staining assay was conducted in +LIF and –LIF conditions. The staining showed no difference in the percentage of undifferentiated, mixed and differentiated colonies between the induced and non-induced cells in both conditions (**Fig.2.23A**). Thus, no impact of NextC1 gain-of-function was observed on the self-renewal ability of ES cells. This result was also supported by RT-qPCR analysis of pluripotency and differentiation factors (Nanog, Sox2, Oct4, Klf4, Esrrb, Id1, Hoxb1, Fgf5) that were found to be similarly expressed in -/+Dox treated cells (data not shown). Moreover, we assessed the growth rate of NextC1 overexpressing clones. Cell

counting over three passages showed that ES cells proliferate at the same rate both when NextC1 is upregulated and at basal level expression (**Fig.2.23B**).

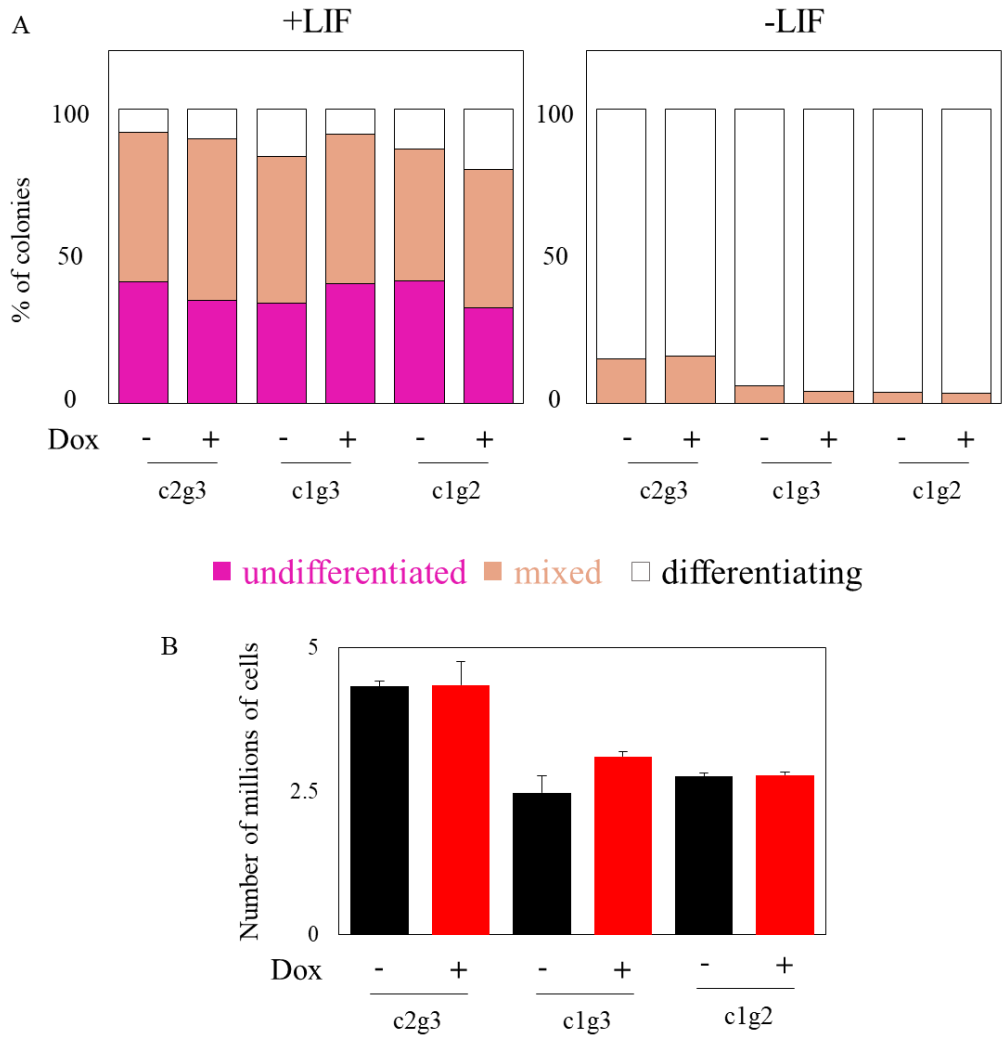


Figure 2.23. A) Alkaline phosphatase staining in NextC1 overexpressing clones. Diagram representing the percentage of undifferentiated (pink), mixed (orange) and differentiated colonies in each clone, in cultures with or without LIF. B) Proliferation rate of NextC1 overexpressing clones. Growth rate was assessed by counting the cells at each passage and replating the same number of cells. Data are represented by number of millions of cells as means and ±SEM of three countings.

Subsequently, we aimed at assessing the effect of the maintenance of NextC1 expression in a context where it normally gets repressed as shown before during differentiation. To that end, we performed a three-day retinoic acid differentiation (RA) assay where NextC1 expression was shown to be abolished (**Fig.2.6D**) and induced with DOX our three NextC1 SunTag cells from day 0. Strikingly, we observed by RT-qPCR that NextC1 was not upregulated in the differentiated samples while it was clearly upregulated in the samples that were cultured in parallel in +LIF condition (**Fig.2.24**). We then verified the expression of the

SunTag system components by looking at the GFP and BFP signal using fluorescence microscopy, corresponding to VP64 transactivator and dCas9 transgenes respectively. To our surprise, we observed that in the differentiating samples the cells under DOX induction were not expressing neither of the two. This implied that the transgenes are no longer expressed in differentiation conditions possibly due to their random integration in genomic regions that are primarily active in undifferentiated ES cells. Since we could not upregulate NextC1 in differentiation conditions with our inducible CRISPRa system we could not further assess any possible effects on the ES cell differentiation capacity.

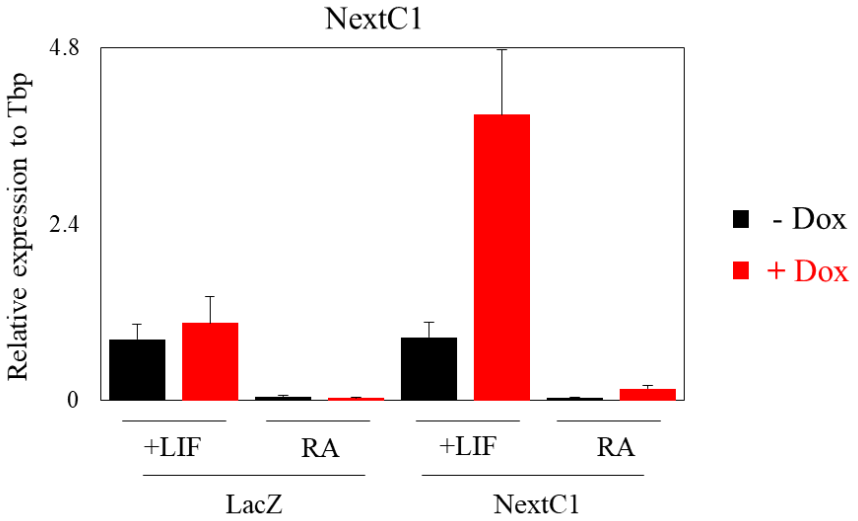


Figure 2.24. RT-qPCR measurement of the induction of NextC1 in +LIF and RA differentiation conditions when LacZ (non-targeting sgRNA) or NextC1 sgRNAs are used. Data are normalized to Tbp and are shown as means and ±SEM of three LacZ clones and three NextC1 clones, in two culture replicates (n=6).

In the light of a potential role of NextC1 in regulating the expression of neighboring genes in *cis*, we assessed the expression levels of few genes in the vicinity of NextC1 locus located within the same TAD. No effect was noticed at the activation of the seven tested genes while obtaining NextC1 three-fold overexpression (**Fig.2.25**). Nonetheless, this does not exclude the possibility that other genes that are located in the same TAD could be affected by NextC1 overexpression. To address the question whether NextC1 short-term (three days) and long-term (nine days) upregulation has an effect on the expression of neighboring genes but furthermore on a genome-wide level, we performed RNA-seq on SunTag clones with LacZ (non-targeting sgRNA) and NextC1 targeting sgRNAs. DESeq analysis led to the identification of 213 differentially expressed genes (DEGs) upon three days of NextC1 induction and 1204 DEGs upon nine-day induction. We are currently studying the list of the affected genes.

Combined with the generated RNA-seq datasets of NextC1 KO clones we aim at deciphering the functional relevance of NextC1.

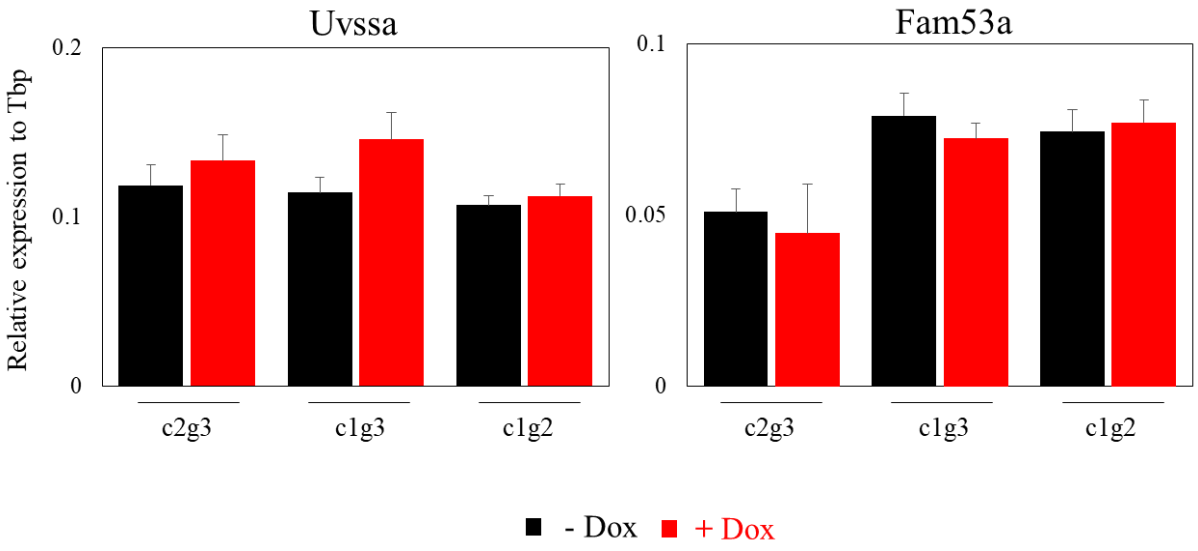


Figure 2.25. Effects of NextC1 activation on neighboring genes by RT-qPCR. Values are normalized to Tbp and represented as means and ±SEM of five culture replicates.

G. Discussion

The analysis of our nuclear matrix preparation samples allowed us to identify lncRNA candidates with a potential of structurally shaping the nuclear space. We chose to focus our functional studies on a non-annotated lncRNA, termed NextC1, which is characterized by a relatively good expression level restricted to naïve pluripotent cells.

A functional relevance of a lncRNA is often suggested when it is evolutionary conserved (Chen et al., 2016; Ulitsky, 2016; Ulitsky et al., 2011). Nevertheless, few cases have been reported of mouse specific lncRNAs that have no ortholog in human like Braveheart (Klattenhoff et al., 2013) and linc-Hoxa1 (or Haunt) RNAs (Maamar et al., 2013; Yin et al., 2015) while showing major functional roles. In addition, the functional orthology has been illustrated to be independent of sequence conservation in the case of the RSX gene that shows a similar function in dosage compensation in marsupials as Xist has in mouse and human (Grant et al., 2012). Therefore, the fact that NextC1 has no evident sequence conservation within mammals does not imply that it has no functionality. Another property which often indicates that a lncRNA molecule might be functional, is its transcriptional stability. However, the short-life of a transcript does not translate into non-functionality for a lncRNA (Clark et al., 2012;

Tani et al., 2012) but rather suggests the potential function to be rapidly mediated after transcription and involved in regulating gene expression (Clark et al., 2012; Dinger et al., 2009). We demonstrated that NextC1 is a rather unstable transcript, a property that could be explained by the absence of splicing and the strictly nuclear localization of the transcript, since it has been shown that lncRNAs tend to be more unstable when unspliced and nuclear (Clark et al., 2012). The absence of a strong poly-adenylation signal in NextC1 gene could also be connected with the instability of the transcript and might imply either that the end of the RNA molecule is not of importance for the activity of the transcript or that only the transcriptional activity of the locus matters and not the RNA molecule *per se*. Still, we found that NextC1 is more stable than intronic RNAs which are readily degraded after splicing (Kataoka et al., 2013; Masaki et al., 2015). Few ncRNAs which have been associated with nuclear organization, have been reported to get delocalized from their regular subnuclear distribution in transcription inhibition assays (Hacisuleyman et al., 2014; Hall et al., 2006, 2014b). Thus, it would be of interest, to monitor the subnuclear localization of NextC1 in such a context, to see whether its localization is also affected. In that case, we can hypothesize that such a displacement might be affecting the apparent stability of the RNA leading to the underestimation of the "physiological" half-life of the transcript.

One of the interesting characteristics of NextC1 gene is the presence of both promoter and enhancer marks at its TSS region. Distinction between enhancers and promoters has been a controversial subject since both can show dual functionality and ambiguous chromatin modification profiles (Andersson, 2015; Creyghton et al., 2010; Heintzman et al., 2007; Kim and Shiekhattar, 2015; Shen et al., 2012). In addition, it was shown very recently that enhancers and promoters could be rapidly switching functions during evolution (Carelli et al., 2018). NextC1 initiates from an ES cell Super Enhancer (SE) (Hnisz et al., 2013; Whyte et al., 2013) characterized by a high enrichment for p300-acetyl-transferase (**Fig.2.26**) and at the same time for H3K4me3 (active promoter) mark at its TSS, suggesting that the locus can exert both functions. It has been reported that RNAs which are produced from a SE element can in turn participate in the SE function (Hnisz et al., 2013; Kim et al., 2010a; Lam et al., 2014). In order to investigate whether that could be the case for NextC1 SE, it would be interesting to assess whether the production of the NextC1 RNA is necessary for the function of the SE. To address this question, we could abrogate the expression of NextC1 by inserting an early ectopic polyA signal shortly downstream its TSS and assess pluripotency TF binding as well as distal regulatory function of the SE. Although currently, we do not know the target genes of this SE,

it was shown that SEs usually regulate the expression of cell type specific genes (Hnisz et al., 2013; Whyte et al., 2013), so it would be reasonable, to hypothesize that the SE of NextC1 locus might be regulating pluripotency and/or developmental genes. A way to gain insight into the role of this SE DNA element could be to conduct a 4C (circularized chromosome conformation capture) assay and identify the ensemble of the genomic loci it interacts with (Zhao et al., 2006).

NextC1 transcript could also be functionally independent of the SE of its promoter region. LncRNAs that exert enhancer-like activity are usually stable and spliced, with the splicing itself playing a role in promoting their enhancer activity (Gil and Ulitsky, 2018; Tan et al., 2018). On the contrary, NextC1 is neither spliced nor stable, making this scenario quite unlikely. Regarding the functionality of the NextC1 RNA, it could be elucidated by applying one of the recently developed technologies for mapping of the genomic regions that a given lncRNA transcript interacts with. Chromatin isolation by RNA purification (CHIRP), capture hybridization analysis of RNA targets (CHART) and RNA antisense purification (RAP) are similar techniques aiming at identifying the genomic loci that can be bound by an RNA molecule of interest (Chu et al., 2011; Engreitz et al., 2013; Simon et al., 2011). Coupling of either method with mass spectrometry analysis, the protein interactome of NextC1 RNA could also be revealed providing us with more information to understand the surrounding molecular environment of this transcript. Moreover, it would be interesting to perform a 4C experiment using as bait regions the CTCF binding sites that are present across the locus of NextC1 and to investigate the genomic regions they could interact with. More specifically, binding sites to be checked first should be the three peaks that are specific to ES cells (highlighted in **Fig.2.26**), as shown from available ChIP-seq datasets in different cell lines but also supported by ChIP-seq data generated in our lab using ES cells, fibroblasts and myoblasts (unpublished). The binding sites of CTCF could be important for the formation of chromosomal looping between NextC1 and the target genes, and critical for the function of NextC1 (Engreitz et al., 2013; Ma et al., 2015). Therefore, it would be essential to include in such a 4C experiment our generated NextC1 mutant clones and examine whether potential interactions in WT ES cells are lost or new are established upon deletion of the promoter region and impairement of NextC1 expression.

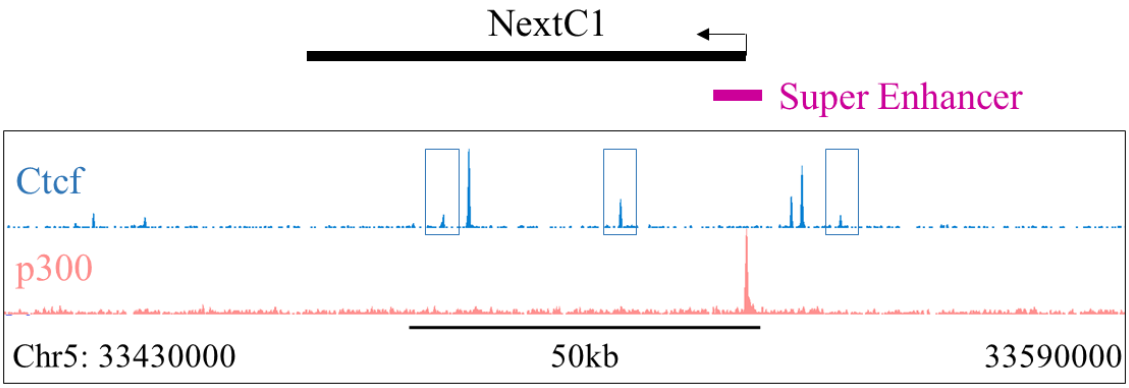


Figure 2.26. CTCF binding sites spread across the NextC1 locus. Highlighted are the three sites specific to ES cells. The identified super enhancer is annotated at the top of the figure as a violet bar and the p300 enhancer marker is depicted in pink.

The presence of an ES cell specific SE in the NextC1 locus was a good clue to orient our investigations towards the relationship between NextC1 and the pluripotency network. Indeed, by looking into the expression profile of NextC1 in different culturing conditions, cell types, and developmental stages we were able to show that NextC1 is tightly linked to naïve pluripotency. It is noteworthy that in 2i conditions its higher expression level detected by RT-qPCR was also validated by RNA-FISH experiments (data not shown) which revealed that NextC1 is expressed in almost all the cells (>95% of the population) in a biallelic manner (90%), resembling the homogeneity of pluripotency TFs attained in 2i media (Wray et al., 2010). To support our analyses, NextC1 was retrieved as an unannotated ES cell-stage specific long non-coding transcript in an independent study scrutinizing the reprograming process from mouse fibroblasts to induced pluripotent cells (iPS) (Hussein et al., 2014). Interestingly, in accordance with our findings that NextC1 is highly expressed in naïve but not in formative or primed pluripotent cells, both *in vivo* and *in vitro*, we found out that the SE at NextC1 locus in ES cells is lost in EpiSC by looking into the data of Novo et al., 2018. This observation suggests that NextC1 locus and/or transcript are important (and potentially functional) only in naïve pluripotency. An additional cell type that would be interesting to look into for NextC1 expression would be the primordial germ cells (PGC) where most of the TFs bound to the SE are re-expressed in the post-implantation embryo (Hayashi et al., 2007).

The subcellular localization of an lncRNA is very important and can be indicative of its possible molecular roles. In our case, the step of identification of localization was especially crucial since we were looking for RNA molecules that could be participating in nuclear organization and therefore possibly showing particular subnuclear distribution. The strictly

nuclear localization and mostly focal enrichment at its transcription sites suggest that NextC1 could potentially have a regulatory role in proximity to its locus. This could be either acting in *cis* on neighboring genes (like Charme in Ballarino et al., 2018 or Kcnq1ot1 in Pandey et al., 2008) or in *trans* by bridging together inter-chromosomal loci (like Firre in Hacisuleyman et al., 2014). Of note, the transcription foci of NextC1 were found to be comparable or even bigger and brighter (with variations from cell-to-cell) than those of Firre lncRNA, when both were subjected to RNA-FISH with fosmid generated probes (**Fig.2.27**). This finding suggests that NextC1 could form local domains around its genomic site. Furthermore, the mostly biallelic expression of the majority of the cell population suggested that NextC1 is probably expressed during all the phases of the cell cycle and that its expression is rapidly reactivated in early G1 cells after mitosis. Taking into consideration that the transcript is rather unstable, we could presume that NextC1 transcription is characterized by high rate and low burst, a feature that tends to characterize key cellular components as housekeeping genes (Lionnet and Singer, 2012; López-Maury et al., 2008; Suter et al., 2011). Whether such characteristics support an importance of the NextC1 transcript or simply result from the strong binding of multiple TFs at the same region (SE) that increases the probability of a transcriptional activity at a given time, remains yet to be seen, if a functional role will be attributed to the transcript.

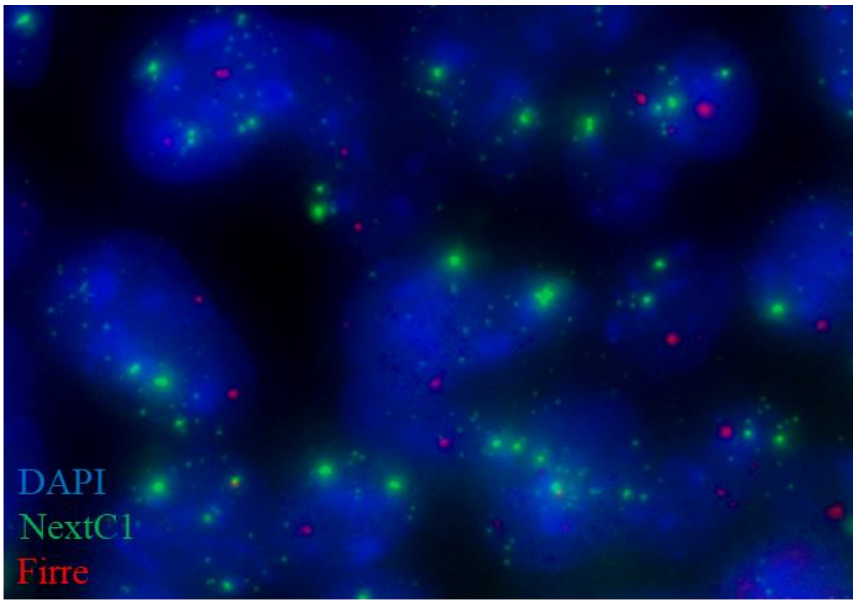


Figure 2.27. RNA-FISH for NextC1 (green) and Firre (red) in Lf2 ES cells. DNA is stained with DAPI.

The field of phase separation by liquid-droplet condensation has recently exploded with new studies constantly emerging which identify proteins that could potentially be involved in compartmentalization via their low complexity domains. Super enhancer (SE) formation and function has been proposed to be mediated by phase separation of RNA polymerase II, Mediator subunits and Brd4 proteins which contribute to concentrate the transcription machinery and control the expression of key cell identity genes (Cho et al., 2018; Hnisz et al., 2017; Sabari et al., 2018; Sawyer et al., 2018). RNAs have also been suggested to promote phase separation depending on their local concentration and/or their secondary structure (Langdon et al., 2018; Maharana et al., 2018). These recent discoveries along with the fact that NextC1 is locally accumulating at its transcription site (which equals high local concentration) that coincides with the SE region, made us wonder whether it could be involved in phase separation alone or by participating in the phase separation that is imposed by the proteins of the SE. Given the fact that different SEs can form hubs and interact in the 3D nuclear space, it would be interesting to perform a DNA-FISH of NextC1 and other SEs and visualize potential proximity or co-localization.

In order to shed light into the potential functionality of NextC1 (locus and transcript) we generated mutant clones where the promoter region, including the binding site of the pluripotency TFs, was deleted. The efficiency of obtaining homozygous mutants with the use of CRISPR/Cas9 system was remarkable and which really revolutionized and facilitated genome editing (Hsu et al., 2014). The KO clones although not expressing NextC1 in FCS growing conditions, showed very low transcription of NextC1 in 2i. One possible explanation for this could be that there is residual binding of TFs at sites that were not eliminated by the deletion triggering some transcription in 2i, when pluripotency TFs are more abundant. It would be thus interesting to perform a chromatin immunoprecipitation assay (ChIP) for few TFs like Esrrb, Klf4 and Prdm14 that show a wider binding region (or more than one distinct peak in ChIP-seq datasets, **Fig.2.5**) and to question their binding capacity at the NextC1 promoter in the KO clones. If proven that residual binding is still possible, it would suggest that the SE might be still active and functional. Therefore, we could distinguish the genes identified as differentially expressed upon NextC1 KO by our RNA-seq to be targets of NextC1 transcript and not of the SE element. In such a scenario, it would be intriguing to perform a larger deletion, removing the full sequence of the annotated SE and aiming at dissecting the possible roles between SE and NextC1 RNA.

So far the analysis we have performed on the mutant clones of NextC1 showed that the deletion of NextC1 and a SE in ES cells did not seemingly have an impact on the ESC self-renewal or potential for differentiation. Could it be possible though that due to redundancy and compensatory mechanisms their regulatory role was masked? In that case, our KO clones have probably adapted to NextC1 loss and they should be challenged in stress conditions in order to unmask the regulatory function of the lncRNA. Another hypothesis could be that the effects of NextC1 lncRNA regulatory role would be visible at later developmental stages so it would be impossible for us to capture the impact with our experimental conditions. On the same direction, this could be the reason why we could detect a mild effect in the Embryoid Body differentiation at an intermediary moment (day 6) in the KO clones compared to the WT, regarding some markers of the endodermal lineage. In that case, we might get important information, on genome-wide changes caused by NextC1 deletion, from the RNA-seq of the KO clones in EB differentiation. If promising results emerge from our RNA-seq datasets we could think of generating a NextC1 knockout mouse model to monitor the effects during the early mouse development based on the specificity of the transcript's expression in the naïve pluripotency. We should take into consideration though that even well-established functional lncRNAs, such as Malat1, Neat1, Hotair, *etc*., have been, strikingly, shown to be largely dispensable for viability and development (Amândio et al., 2016; Goudarzi et al., 2018; Han et al., 2018; Nakagawa et al., 2011; Zhang et al., 2012).

The NextC1 mutant clones that we generated do not all bear the exact same deletion as we discovered by PCR product sequencing (**Fig.2.13**). Interestingly, the clone that carried the biggest deletion (C5) was the one that showed the fewest undifferentiated colonies in the AP staining that we performed in +LIF conditions, and relatively fewer mixed colonies in–LIF conditions (**Fig.2.15**). It might be possible that the extended deletion has removed more efficiently the multiple Esrrb peaks of the region when compared to the other deletions. This will be addressed by an Esrrb ChIP (along with other pluripotency TFs) using the KO clones, as already mentioned above. We should also further investigate the results of the AP stainings, by performing additional replicates and carefully monitoring whether clones C4 and C5, which have the biggest deletions, reduce the self-renewal capacity of the mutant clones. It would also be necessary to take into consideration this difference in respect to the deletion size when analyzing the RNA-seq datasets of the NextC1 mutants.

The RNA-seq data will hopefully give us crucial information regarding the functional relevance of NextC1 while the differentially expressed (DE) genes after the DESeq statistical analysis already show a significant number of genes to be affected after NextC1 deletion of expression. We will be thoroughly looking at the impact of NextC1 deletion in tne transcriptional activity of (i) genes located within the NextC1 TAD to identify potential targets of in *cis* regulation, (ii) genes involved in cell cycle regulation, metabolism or apoptotic mechanisms to potentially explain the little growth retardation of the KO clones in 2i medium, and (iii) genes that have been listed as potential interactors with the NextC1 locus taken from the online repository of chromatin interaction data "4Dgenome" (https://4dgenome.research.chop.edu/). In addition, it would be interesting to check whether we could detect a specific enrichment for genes that are in close proximity to other SE (231 SE assigned to 210 genes in ES cells according to Whyte et al., 2013) among the genes that appear to be differentially expressed upon NextC1 depletion. Such a scenario could imply that the SEs are somehow interacting (by physical interactions and formation of SE hubs for instance) and perturbation of one could influence the function of another, a notion that is supported by several publications (Beagrie et al., 2017; Moorthy et al., 2017; Novo et al., 2018; Rao et al., 2017). The loci of the potential targets of NextC1 that will be identified by our RNA-seq analysis should be analyzed by DNA-FISH to assess their subnuclear localization and spatial relationship to the NextC1 locus. In that way, if NextC1 expression is necessary for the interaction between its own locus and its target loci, we could expect to find them colocalizing with NextC1 in WT cells and losing the colocalization in the KO cells.

The validation of the findings of our RNA-seq should be performed using the inducible knockdown clones that we generated with the CRISPR/dCas9-KRAB system. The inducible abrogation of NextC1 expression, with no modification of the locus, would help us to decipher whether the misregulated genes result from the abolished expression of the transcript or the deletion of the SE locus. However, the recruitment of the dCas9 protein along with the KRAB domain corresponds to the loading of a big cargo downstream the TSS of NextC1, rising the possibility that it could sterically compete with the binding of TFs around this region (more specifically the SE region). Nevertheless, the sgRNAs we have designed should be far enough from the TFs binding sites to avoid such an effect which could though be checked by ChIP experiment. Another option to validate the requirement of NextC1 RNA molecule to mediate the identified effects would be the insertion of a polyA signal shortly downstream of its TSS, as mentioned above.

In parallel, we plan to use the RNA-seq data from the NextC1 overexpressing SunTag clones, cross them with the list of DE genes of the KO clones and explore the genes that might be responding in a meaningful way in the two cases. It is noteworthy that the upregulation of NextC1 transcript that we acquired was around three- to four-fold, which could be considered not strong enough to produce any noticeable effect. Nonetheless, taking into account that NextC1 is expressed by the vast majority of the cells, in a biallelic manner, as well as the binding of many pluripotency TFs at its promoter, we can speculate that the upregulation of NextC1 is reaching a plateau. We did observe though that the transcription foci became bigger and brighter in the overexpressing cells, it is thus possible that we might capture the consequences of the reinforcement of its transcriptional activity, considering it is the RNA molecule that has a functional role. If we were to activate and corroborate the enhancer activity of NextC1 locus, the most appropriate system of CRISPRa to use would have been the CRISPR/dCas9-p300 core (Hilton et al., 2015) which has been shown to perform better than VP64 fusions in such cases.

Finally, it would have been quite compelling to maintain NextC1 expression in differentiation assays where it gets shut off, such as RA or EB differentiation. We were not able to examine the outcome of such a transcriptional activity though since the SunTag system did not work in differentiation conditions likely due to the extinction of the transgenes during the exit from pluripotency. Such a setup would enable us to see the effects of NextC1 RNA expression in an environment where the TFs normally bound to its promoter are no longer expressed thus eliminating the function of the SE. We could similarly try to express endogenously NextC1 by transfecting the SunTag system components directly in a cell line that does not express NextC1, like for instance MEFs. It would be interesting to observe its subnuclear localization in that context and assess the potential consequences of this ectopic expression on the genes that will be identified as being be responsive to NextC1 modulation in ES cells (from our RNA-seq data analysis).

To conclude, the fact that NextC1 has no annotation yet in gene databases results in it being overlooked in different large-scale screens and studies that aim at identifying functional lncRNAs and their global properties. Consequently, characteristic features allowing for the classifications of lncRNAs could not have been attributed to NextC1. With our work we opened the route towards a better understanding of the functionality of this gene in an attempt to enrich the knowledge we have on lncRNAs.

III. NextC2 (Non-extracted Candidate 2)

A. Validation of NextC2 RNA and matrix retention

From our RNA-seq data, NextC2 appears as a large transcriptional unit of 80kb located on mouse chromosome 8 (**Fig.3.1**). It is transcribed from the positive strand in a gene desert with its closest adjacent gene being located more than 100kb away. Using our paired-end, long read sequencing but also other RNA-seq datasets (poly(A) selected) (**Fig.3.2C**) that were generated in our lab we could computationally construct gene models of the most represented isoforms of NextC2. This analysis showed that the NextC2 locus actually produces two distinct isoforms, thereafter called long and short according to the length of their first intron, that initiate from two alternative promoters (**Fig.3.2A**). These two isoforms differ at their first exon but share their short second and very long third exon. Therefore, we first aimed at validating by RT-qPCR the NextC2 transcription unit structure, revealed by the RNA-seq results. To this end, we designed discriminating primer pairs for the two isoforms and a number of primer pairs targeting their common part but also the full locus of NextC2 (**Fig.3.2B**) and performed RT-qPCR in Tg2a ES cells grown in FCS. We confirmed the transcription start site (TSS), the putative termination site (**Fig.3.2D**) and the existence of the two isoforms. Of note, for all the further RT-qPCR performed on NextC2 RNA, three primer pairs will be shown corresponding to the long isoform (represented always in red color), the short isoform (in blue), and a common splicing junction (in green).

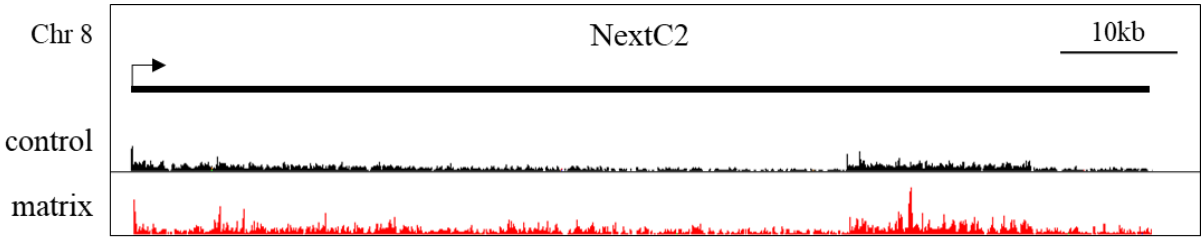


Figure 3.1. Screenshot of IGV browser showing the RNA-seq reads coverage of the NextC2 locus in control (black) and matrix (red) samples. Reads coverage have been group-auto scaled to visualize the level of NextC2 retention.

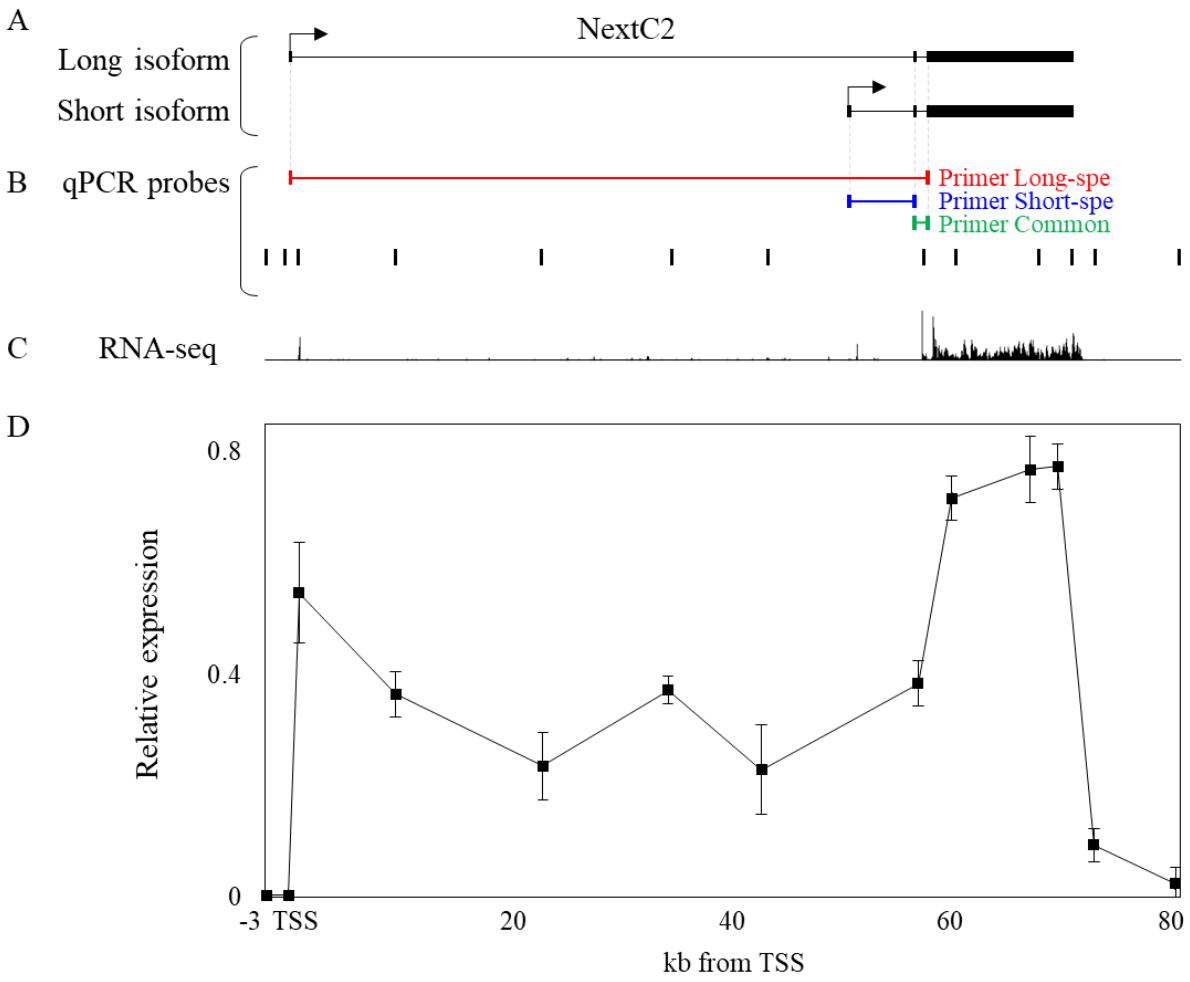


Figure 2.2. NextC2 locus characterization. A) NextC2 is transcribed from the positive strand and gives rise to two isoforms (long and short) initiating from two alternative promoters. B) RT–qPCR primer pairs designed to distinguish long from short isoform, a common part of the transcript and others covering the full genomic locus. C) Screenshot of IGV browser showing RNA-seq coverage of ES cells. D) RT-qPCR analysis of the primers representing the full locus of NextC2 in Tg2a ES cells. Values are normalized to Tbp mRNA and are shown as means ± s.e.m. from three biological replicates.

Once having validated the expression of NextC2 in regular samples of total RNAs, we proceeded with validating NextC2 retention in matrix samples as detected by RNA-seq (**Fig.3.1**). Therefore, we performed RT-qPCR in control and matrix samples of Tg2a ES cells. Tbp was used as a control to demonstrate the efficient depletion in the matrix preps and for NextC2 the distinguishing primers for the two isoforms were used to attest for its retention. Indeed, we observed a drastic depletion of Tbp transcript while both isoforms of NextC2 were enriched in the matrix compared to control fractions (**Fig.3.3**) demonstrating that NextC2 RNA is reproducibly a matrix-fraction associated transcript.

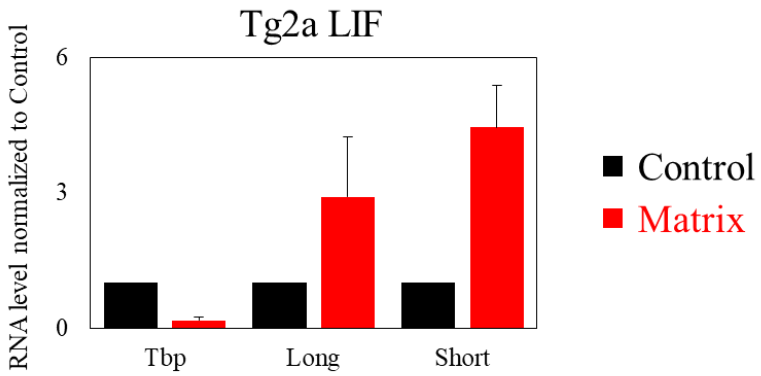


Figure 3.3. NextC2 retention in matrix prep. Tbp is efficiently depleted while Long and Short isoforms of NextC2 and enriched in matrix compared to control samples. Data are normalized to control samples for each transcript and shown as mean and ±SEM of four independent experiments.

B. NextC2 coding potential and conservation

NextC2 has no annotation in the latest ENSEMBL database (Zerbino et al., 2018). Nevertheless, unlike NextC1, NextC2 has few short annotations in deepBase (Zheng et al., 2016) and one annotation in NONCODE (Zhao et al., 2016). However, none of these annotations correspond to the transcripts that we recovered from our RNA-seq data nor by our RT-qPCR analysis and are most likely to be expressed in more differentiated tissues. Therefore, in order to confirm that NextC2 gene is a *bona fide* lncRNA we investigated its coding potential using the PhyloCSF track of the UCSC genome browser (Lin et al., 2011). The full locus of NextC2 shows a negative coding potential in any of the three possible frames all along its long sequence. Additionally, we used the web tool Coding Potential Assessing Tool (CPAT) (Wang et al., 2013a) to more precisely question the coding potential of the two isoforms. For both long and short isoforms the outcome was non-coding labelling, with a coding probability of -0.086 and 0.085 respectively and a positive coding probability cutoff at 0.44.

Next, we investigated the conservation of the gene in mouse and human genomes. We found that there is very poor sequence conservation between the two species. With the help of the mammalian conservation track of the UCSC genome browser (placental mammal conservation by PhastCons), we could identify few small, highly conserved regions which, as for NextC1, most likely correspond to transcription factors binding sites in the syntenic human region (according to transcription factors binding sites annotation from ENCODE ChIP-seq database, UCSC human genome browser). In addition, we searched for the existence of local repeats within the NextC2 gene. For this, we used the BLAST tool to map 10 kb long bins of

Nextc2 full sequence to the mouse genome but did not retrieve any specific local repeats of the NextC2 sequence. However, we noticed that the full NextC2 locus is relatively rich in transposable elements (TEs), a general feature that is already well-established for lncRNAs (Kapusta et al., 2013; Kelley and Rinn, 2012)

C. NextC2 expression regulation by the pluripotency network

Interestingly the long NextC2 isoform transcription starts site is located from within a retroelement of a mouse-specific subfamily (LTR9A) of the ERVK retroviruses family. This specific LTR subfamily has been shown to often be enriched in pluripotency TFs binding (Nanog, Oct4, Sox2, Esrrb, Klf4) and the signature of enhancers (H3K4me1, H3K27ac, p300) in mouse ES cells, and implicated in regulating ES-specific gene expression patterns (Sundaram et al., 2017). Therefore, we explored the publicly available ChIP-seq datasets for multiple pluripotency TFs in mouse ES cells (Chen et al., 2008; Handoko et al., 2011; Ma et al., 2011; Whyte et al., 2013) and visualized their binding profile on UCSC browser through the online compendium CODEX (http://codex.stemcells.cam.ac.uk/). Strikingly, we discovered multiple pluripotency TFs to be bound 2kb downstream of the proximal (short) promoter of NextC2 (**Fig.3.4**). The pluripotency TFs Nanog, Oct4, Sox2, Esrrb, Tcfcp2l1, and Prdm14, all exert a strong binding at this region that is also characterized by a p300 enhancer binding protein peak as well as strong H3K4me1 and H3K27ac enrichments (not shown). This profile strongly suggests an enhancer activity of this region. On the contrary, the distal promoter leading to the long isoform expression, while being mostly composed of LTR9A elements, with the aforementioned characteristics genome wide, did not show strong binding activity of pluripotency TFs; we only attested a small peak for Nanog, Esrrb, Tcfcp2l1 and Prdm14 upstream of the TSS.

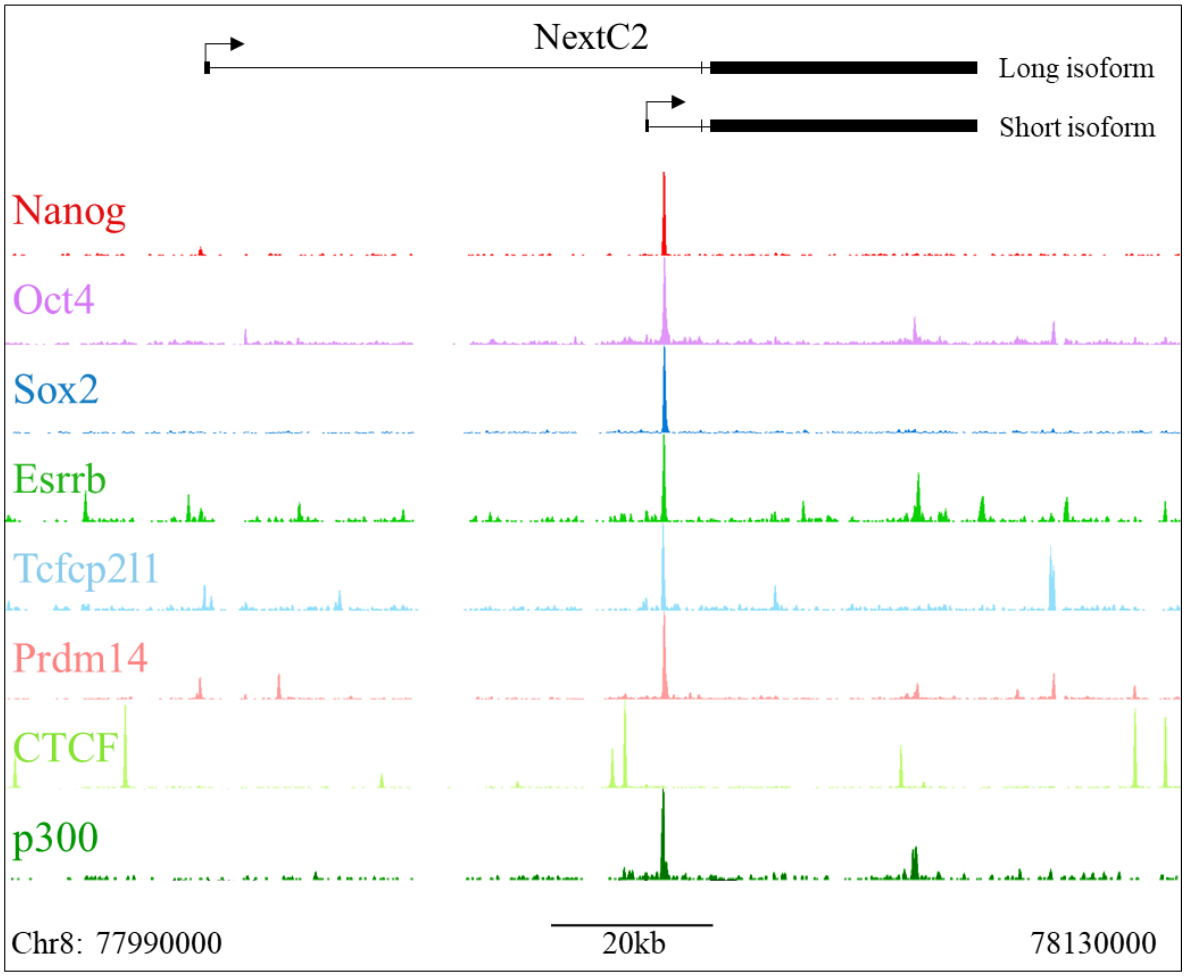


Figure 3.4. Pluripotency TFs and p300 coactivator binding profiles over the NextC2 locus as identified by ChIP-seq (UCSC browser, mm10 assembly). Short isoform's promoter has a strong binding site for Nanog, Oct4, Sox2, Esrrb, Tcfc2l1 and Prdm14 pluripotency TFs as well as the p300 enhancer binding protein. Few CTCF peaks are spread at the locus with two being at close proximity to the proximal promoter.

Given the strong binding of multiple pluripotency TFs downstream of the short isoform promoter (**Fig.3.4**), we decided to assess whether NextC2 expression was regulated by pluripotency factors. Thus, we first measured the expression of both isoforms in serum (FCS) and after long-term culture in 2i medium since pluripotency factor expression has been shown to get corroborated in this condition (Wray et al., 2010). Remarkably, this analysis revealed that the two isoforms follow a mirror-image expression profile between these two media (**Fig.3.5A**). More precisely, while the short NextC2 isoform was upregulated in 2i by two- to threefold, the expression of the long isoform was completely abolished in this medium resulting in an overall decrease of the exons shared by the two transcripts. We then sought to gain more insight into this opposite behavior by capturing the dynamics of response of the two promoters to 2i treatment. Thus, we performed a kinetics experiment of FCS to 2i transition over the

course of three days which confirmed the previous finding but additionally showed that both responses occur as soon as 24 hours after the transition (**Fig.3.5B**). We consequently carried out a shorter kinetics assay and specifically revealed that NextC2 transcripts' expression simultaneously react to 2i treatment after four hours, therefore being among the fastest responsive genes we could independently identify in the laboratory (**Fig.3.6**). This suggests a direct role of either serum stimulation or Fgf/Wnt pathways in the regulation of NextC2 locus.

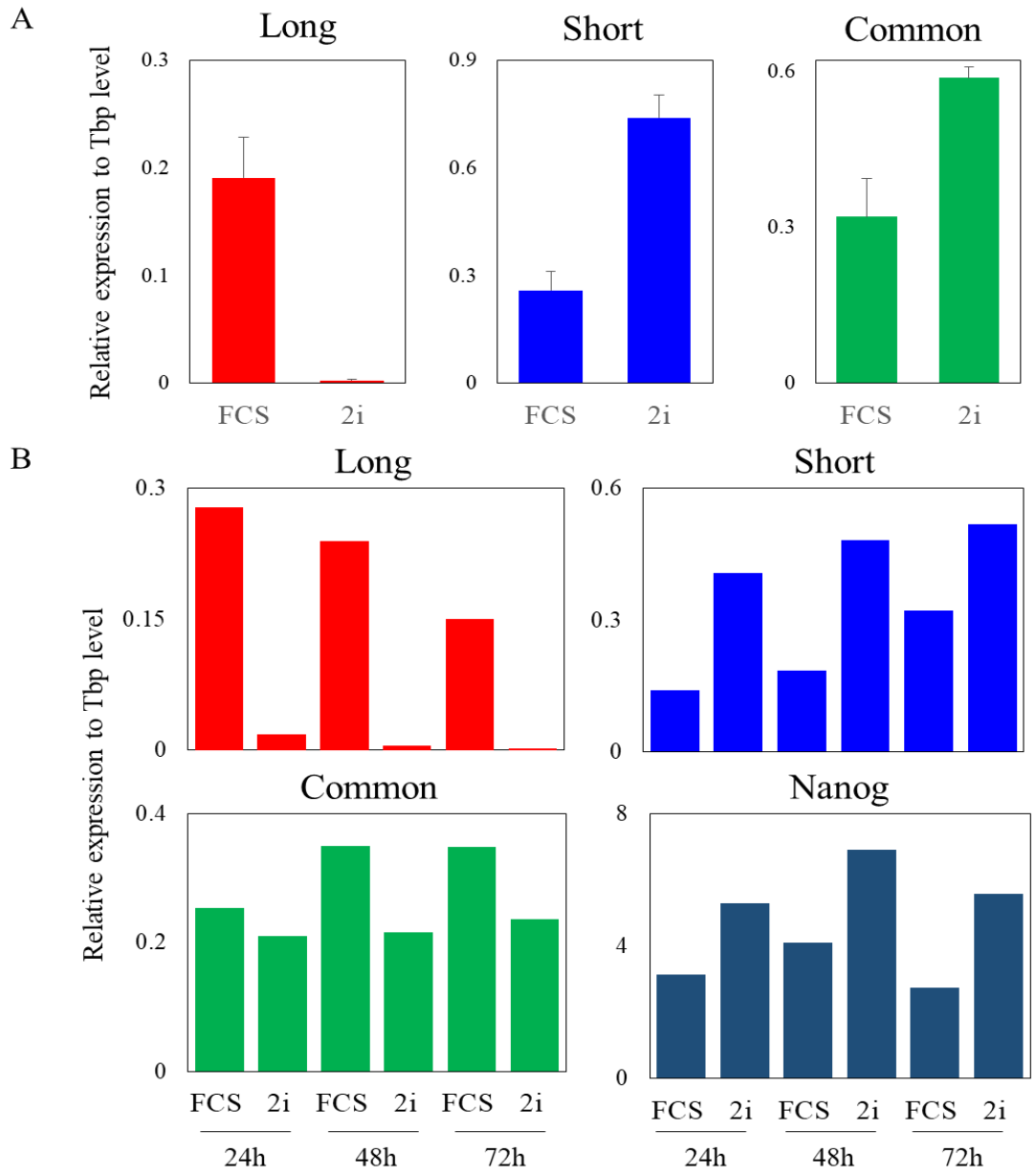


Figure 3.5. NextC2 expression under serum to 2i transition, over long-term 2i exposure (A) or three day kinetics (B). RT-qPCR analysis of NextC2 isoforms and Nanog expression levels. Data are normalized to Tbp mRNA levels and shown as mean and ±SEM of four independent experiments (A) or single experiment (B).

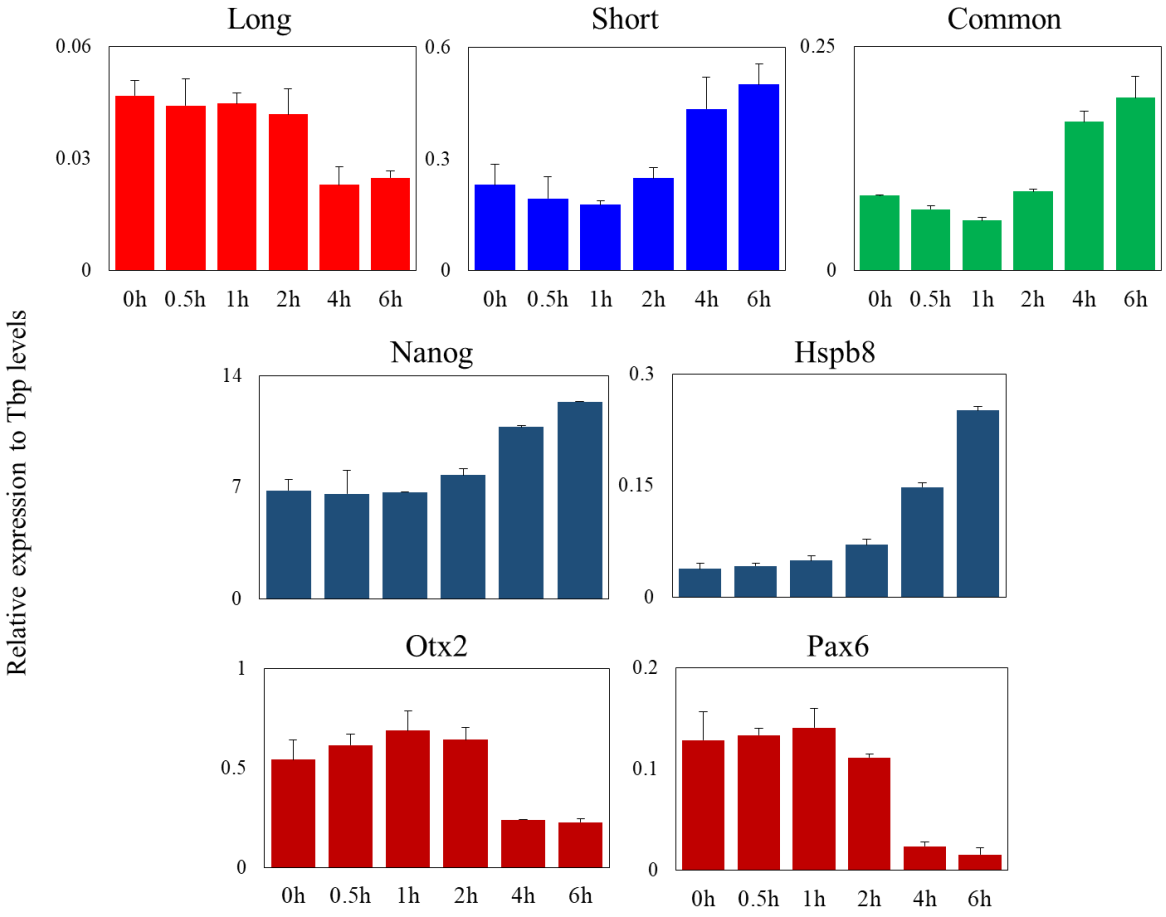


Figure 3.6. Short kinetics of FCS to 2i transition. RT-qPCR analysis of NextC2 isoforms and examples of fast responsive genes increasing (Nanog, Hspb8) and decreasing (Otx2, Pax6) in this transition. Data are normalized to Tbp mRNA levels and shown as mean and ±SEM of two independent experiments.

Furthermore, we sought to specifically address the regulation of the two isoforms by two core pluripotency factors binding at the NextC2 locus, Oct4 and Nanog. For this purpose, we took advantage of available mutant cell lines where the expression of both TFs can be artificially manipulated through DOX-responsive systems (Tet-ON and OFF) (Chambers et al., 2003; Festuccia et al., 2012; Niwa et al., 2000). Therefore, we performed loss of function assays for the two factors over a two day long timecourse. By RT-qPCR analysis, we observed that NextC2 expression was completely lost by 24h of Oct4 depletion, surprisingly showing a similar response of its two isoforms in that condition. However, while the short isoform already negatively responded at 12h, the expression of the long one was still unaffected or slightly increased at this early time point indicating a distinct response to the acute loss of Oct4 (**Fig.3.7A**). These results suggest that the regulatory function of Oct4 is not crucial at the locus since its rapid depletion does not lead to a fast and profound transcriptional response (12h) before the cells initiate the trophectodermal lineage commitment (24h) (Niwa et al., 2000). On the contrary, both transcripts responded very rapidly -in less than 12h- to Nanog depletion,

suggesting a rather key role for Nanog in the regulation of the NextC2 locus (**Fig.3.7B**). In this case, both isoforms once more showed a negatively correlated pattern of expression. This result was further confirmed in cell lines where Nanog is either depleted or overexpressed long term, leading to a strictly opposite pattern of expression of the two isoforms; the long and the short isoforms were no more detectable upon long term overexpression or depletion of Nanog respectively (**Fig.3.7C**).

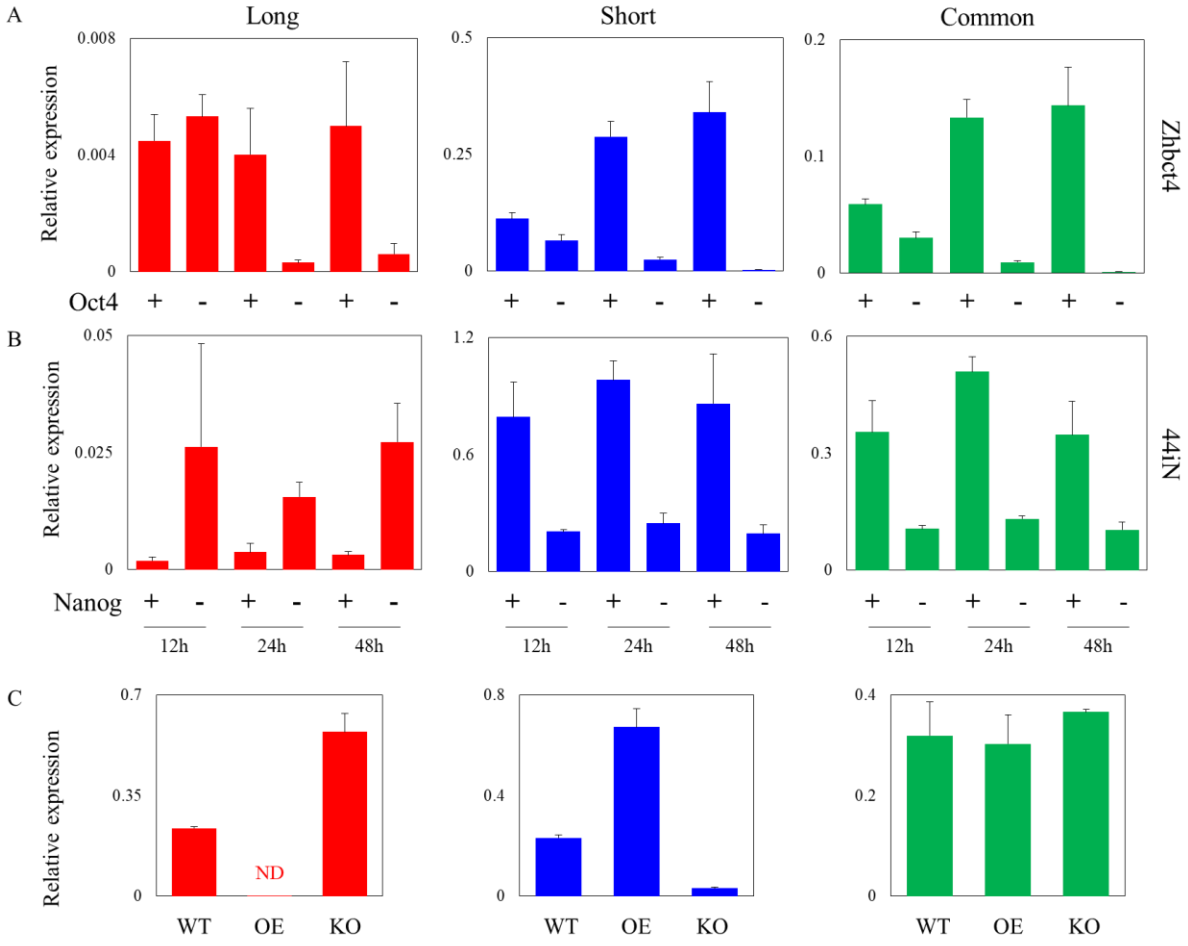


Figure 3.7. NextC2 expression under the regulation of pluripotency TFs Oct4 and Nanog. RT-qPCR analysis of NextC2 isoforms expression levels. A) Zhbtc4 cells treated or not with doxycycline (DOX) for 12, 24 and 48 hours *(Niwa et al., 2000)*. B) 44iN cells were maintained in the presence of DOX which was then removed for 12, 24 or 48 hours *(Festuccia et al., 2012)*. C) EF4 cells (OE) stably overexpress Nanog *(Chambers et al., 2003)* while 44iN cells (KO) were cultured in the absence of DOX, *i.e.* absence of Nanog *(Festuccia et al., 2012)*. Data are normalized to Tbp mRNA levels and shown as mean and ±SEM of three (A, B) or two (C) independent experiments.

Moreover, intrigued by the strong binding of Tcfcp2l1 at the NextC2 enhancer, a key factor downstream of the LIF pathway (Martello et al., 2013), we aimed at answering whether any of the two isoforms were sensitive to LIF withdrawal. We found that only a moderate

response of both transcripts was obtained after 48 hours of LIF withdrawal which might be associated with preliminary commitment rather than LIF responsiveness (**Fig.3.8A**). In addition, no response was observed for shorter time of LIF removal (**Fig.3.8B**). Therefore, we concluded that the NextC2 locus is not directly targeted by the LIF pathway.

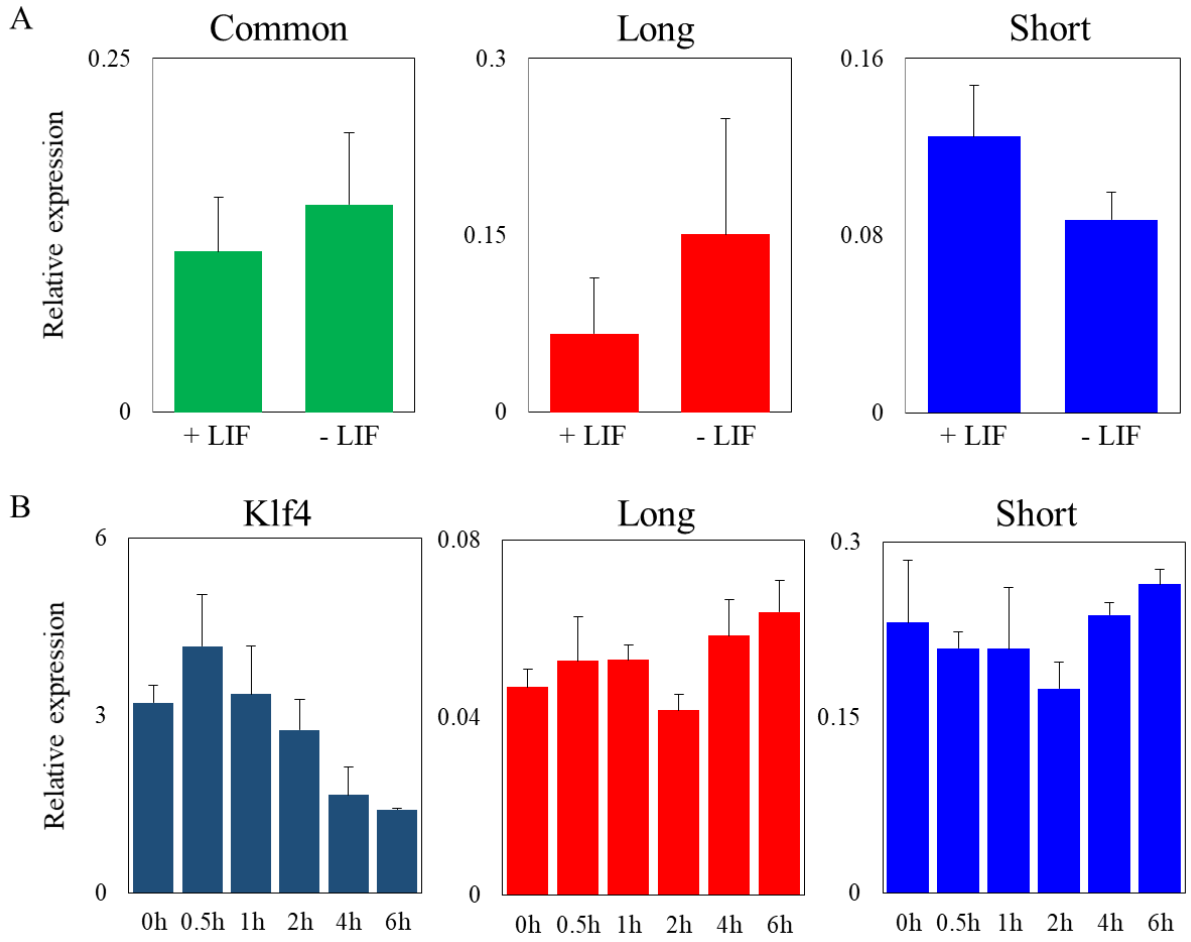


Figure 3.8. RT-qPCR analysis of LIF withdrawal and NextC2 response. A) 48h removal of LIF has moderate effect on NextC2 expression. B) Short kinetics of LIF removal indicate that NextC2 is not affected whereas Klf4 is already responsive at 4h. Data are normalized to Tbp mRNA levels and shown as mean and ±SEM of three (A) or two (B) independent experiments.

Finally, we addressed the transcriptional activity of both proximal and distal NextC2 promoters upon differentiation of mouse ES cells. For that aim, we first measured the expression of the two isoforms upon Embryoid Body (EB) differentiation. The short isoform followed Nanog expression over time while the long isoform was progressively increasing until the end of the differentiation protocol (**Fig.3.9A**). Comparable results were obtained upon retinoic acid (RA) induced differentiation. Accordingly, the short NextC2 transcript rapidly

decreasing until it got almost undetectable after three days of RA treatment similarly to Nanog mRNA. On the contrary, the long isoform was continuously expressed at day one before completely collapsing towards day three (**Fig.3.9B**). We additionally analyzed the expression of both NextC2 RNAs in our EpiLC differentiation assay. Interestingly, we showed that the long NextC2 isoform gets upregulated during the transition from mouse ES cells to EpiLCs, whereas the short RNA progressively decreases from day one to three (**Fig.3.9C**). Last, we investigated the expression of NextC2 locus in distinct cell lines representing extra-embryonic cell types (TS,XEN) as well as a later embryonic differentiation state (MEF) and found that both isoforms were absolutely undetectable in these three cell lines (**Fig.3.9D**).

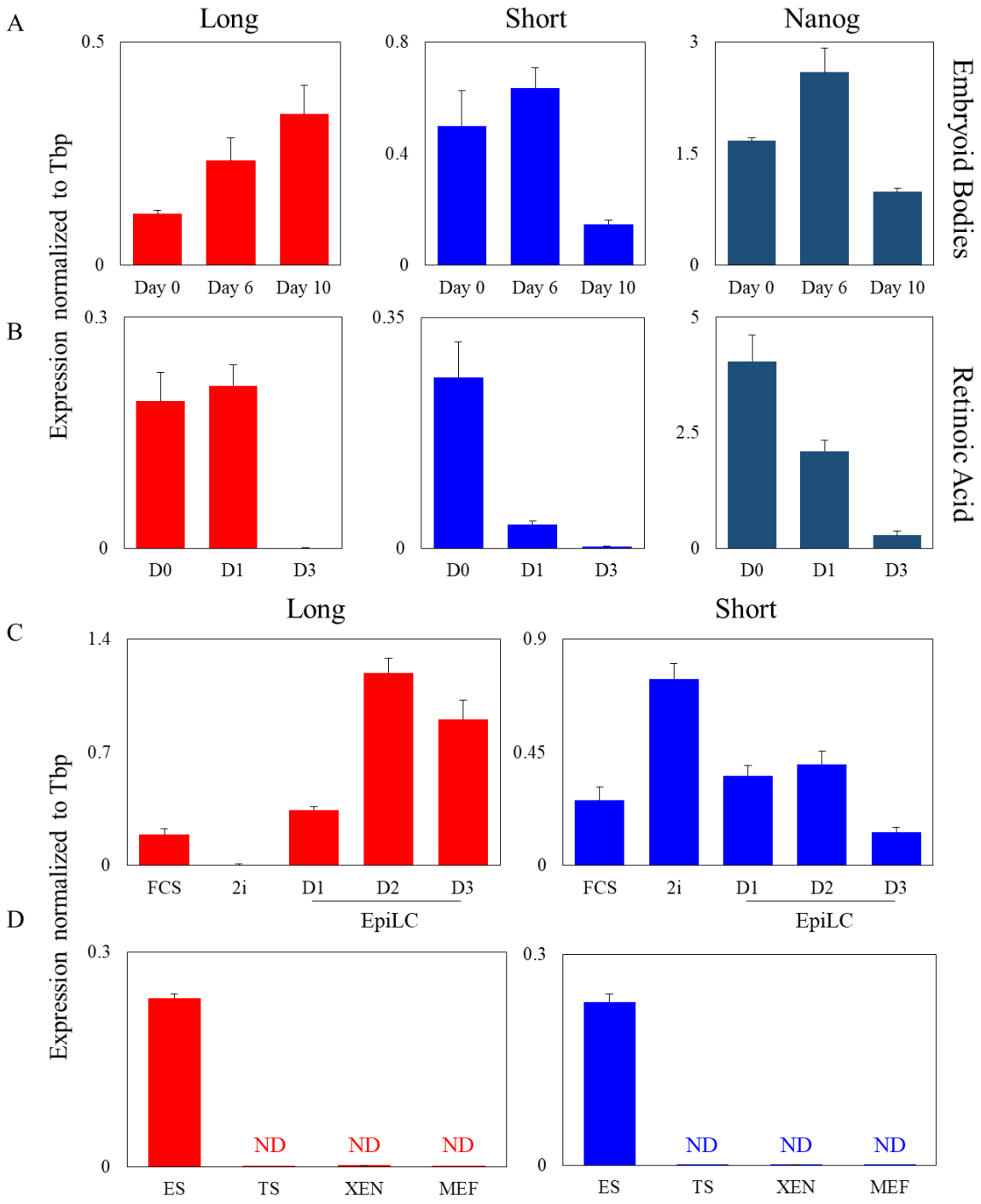


Figure 3.9. NextC2 expression in differentiation. A) Embryoid bodies differentiation (n=2), B) Retinoic acid differentiation (n=4). C) EpiLCs induction from ES cells (n=4). D) Embryo-derived cell types: ES cells (n=2), extraembryonic endoderm stem cells (XEN) (n=3), trophoblast stem cells (TS) (n=5) or mouse embryonic fibroblast cells (MEF) (n=1). Data are shown as means ± s.e.m. from n cell culture replicates.

Collectively, NextC2 transcription unit presents two distinct isoforms differing in their transcription initiation site and additionally showing differential pattern of expression. The short NextC2 isoform seems to be regulated by a downstream enhancer where multiple

pluripotency TFs bind and is therefore associated with the naïve pluripotent state. The long NextC2 isoform is rather associated with the formative pluripotent state and tends to be repressed by pluripotency factors like Nanog and in the naïve state of pluripotency. None of the transcripts is a downstream target of the LIF pathway but both are responsive to Fgf/Wnt or BMP alterations. Lastly, we could not detect any expression of these two transcripts neither upon neuronal commitment nor in extra-embryonic cell types and MEFs.

D. NextC2 subcellular localization

NextC2, as NextC1, was initially selected as a lncRNA candidate that could structurally organize the mouse ES cell nucleus. Therefore, it was essential to establish its subcellular distribution. To this end, we used the cell fractionation assays that were also used for NextC1 determination of localization and by RT-qPCR analysis measured NextC2 expression in the cytoplasmic and nuclear fractions. We ascertained that both isoforms of NextC2 are enriched in the nuclear fraction, however to a lesser extent than Malat or Neat1 RNAs. This could imply that some transcripts are also localized in the cytoplasm (**Fig.3.10**).

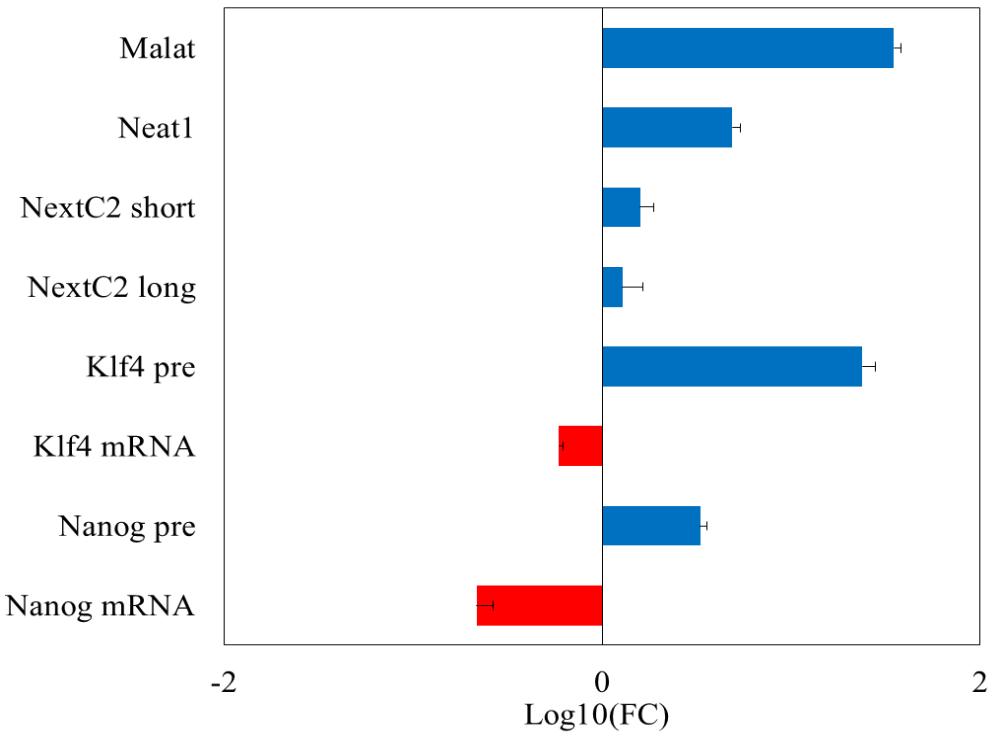


Figure 3.10. Cell fractionation assay shows that NextC2 long and short isoforms are notably localized in the nucleus. Primary transcripts and nuclear lncRNAs are markedly enriched in the nuclear fraction whereas mature transcripts of TFs are enriched in the cytoplasmic. Data are represented as ratio of nuclear to cytoplasmic expression levels in log10 scale (n=2, ±SEM).

In order to determine whether NextC2 RNAs are strictly nuclear or some diffuse to the cytoplasm and to investigate their potential ability to form any kind of nuclear domains, we assessed their subcellular localization using single molecule fluorescent in situ hybridization (smFISH). In that way, we were able to assess the expression of NextC2 isoforms in a single-cell and single-molecule resolution. Two oligonucleotide probes with distinct fluorophores were designed: one specific to the first intron of the long isoform and a second one targeting the exonic sequences of all the exons pulled together (including the two mutually exclusive first exons of the two isoforms) (**Fig.3.11A**). We performed the smFISH experiment on cytospun Tg2a cells grown in FCS, and 2i conditions as well as EpiLC differentiation day two and three (D2, D3) when the long isoform reaches its highest levels of expression (**Fig.3.11B**). We observed that NextC2 expression is detected in the majority of the cells grown in FCS (around 70%) of which 25% actively transcribe the long isoform. The spliced RNAs are mostly detected as single molecules scattered through the nucleoplasm. Whether these mature transcripts are derived from the short or the long isoform is unknown since cells that do not show active transcription from the long intron might still have transcribed it shortly before the cells were processed for cytospin. Few mature transcripts were detected in the cytoplasm explaining the relatively mild nuclear enrichment detected by RT-qPCRs whereas most of NextC2 molecules were clearly localized in the nucleus. In 2i cultured cells, as expected from our RT-qPCR results, the long isoform was not detected anymore while around 78% of the cells still express NextC2 short transcript (slightly more compared to FCS in agreement with previous RT-qPCRs). The distribution pattern of the mature short isoform is composed of dispersed RNA molecules in the nucleus with 23% of the cells having evident transcription pinpoints. In EpiLCs, we attested the upregulation of the long isoform, manifested by big, bright transcription foci represented by the intronic probe. At day two of the differentiation (D2) more than 80% of the cells are long isoform-positive and almost all of them (>95%) exhibit large transcription foci with some few diffused molecules; similarly at D3 around 70% of the cells actively transcribe the long isoform (**Fig.3.11B**). The dispersed molecules as well as the bright transcription foci detected with the exonic probe indicate that NextC2 isoforms might exert *in cis* and/or *in trans* functions.

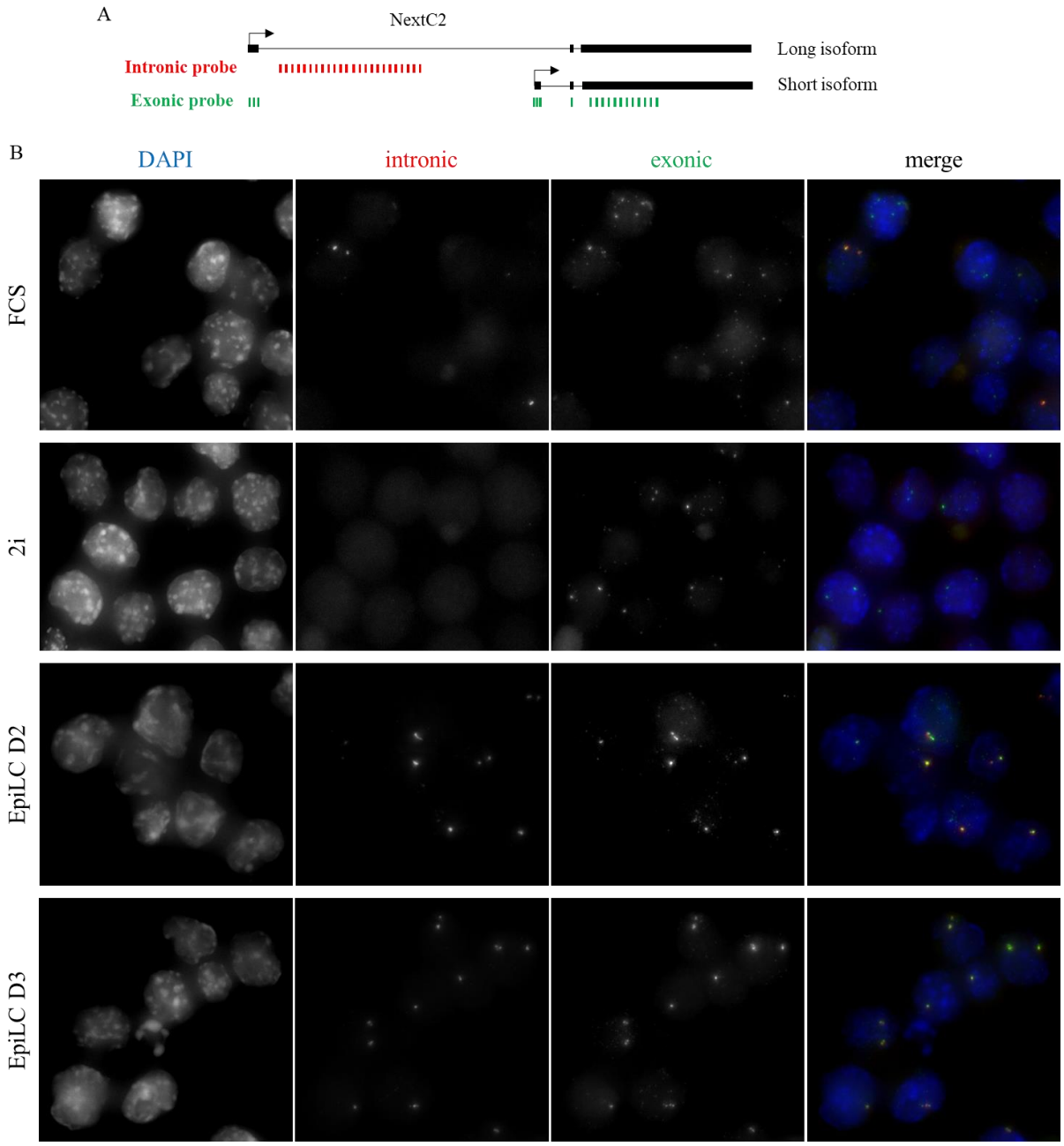


Figure 3.11. Subcellular localization of NextC2 assessed by smFISH. A) The designed probes target (i) the first intron of the long isoform and (ii) the exonic sequence without distinguishing the two isoforms. B) Panels showing smFISH for intronic sequence of long isoform (red) and exonic sequence of both isoforms (green) performed on ES cells grown in FCS, 2i conditions and EpiLC differentiation (D2, D3). smFISH demonstrated nuclear localization of NextC2. The long isoform is mainly accumulating at the transcription sites whereas the short isoform notably is scatter in the nucleus. DNA is stained by DAPI.

E. NextC2 RNA stability

If one of the two NextC2 isoforms was very unstable, it would be very unlikely that we could detect any diffusible molecules arising from this isoform by RNA-FISH. We thus

wondered whether the two isoforms have different stabilities, a property that would enable us to characterize which of the two isoforms might be producing the observed scattered molecules. NextC2 is an RNA pol II transcribed gene as shown by the enrichment of RNA polymerase II and presence of histone modification marks typical of Pol II-transcribed genes. Therefore, we analyzed by RT-qPCR the stability of NextC2 transcripts in the transcription inhibition assays described previously for NextC1 RNA stability (section **II.C.**), where Flavopiridol and Actinomycin D were used. Surprisingly, we noticed that NextC2 long isoform is very stable since it remained unaffected after 6h of treatment of either Flavopiridol or ActinomycinD (**Fig.3.12**). On the contrary, the short isoform appeared to be less stable in comparison, with a half-life of around 2h (with flavopiridol).

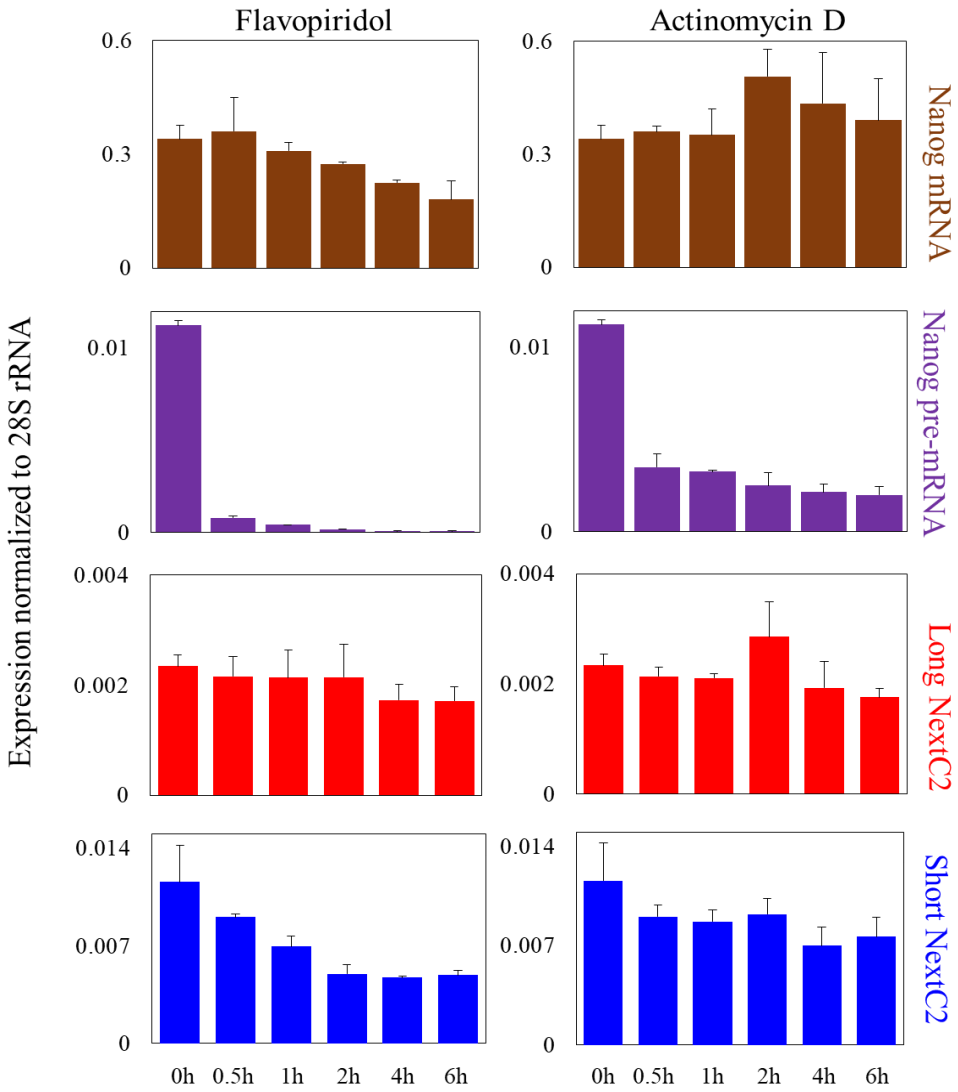


Figure 3.12. Transcription inhibition assay with Flavopiridol (left panels) and Actinomycin D (right panels). RT-qPCR analysis for Nanog mRNA (brown), Nanog pre-mRNA (purple) and NextC2 long

(red) and short (blue) transcripts. Data are normalized to 28S rRNA and are shown as means and ±sem of 2 independent experiments.

F. Functional assays

i. Loss of function

Constitutive locus deletion with CRISPR/Cas9 system

To address the functionality of NextC2 we first performed a loss-of-function assay using the CRISPR/Cas9 system. To that end, we targeted for deletion both promoters at once, so that we could eliminate completely NextC2 transcription. Thus, we generated male knockout ES cells by deleting a 55kb long genomic region covering the locus shortly upstream and downstream of the long and the short isoforms promoters respectively. We specifically avoided to affect the enhancer region located within the first intron of the short isoform and that is bound by multiple pluripotency TFs in ES cells, considering that it might be a regulatory region of other genes outside NextC2 locus. Thereafter, this deletion will be referred to as "big deletion" (**Fig.3.15**). As quantified by RT-qPCR, in 2i condition ES cells express exclusively the short isoform. Therefore, in order to limit confounding factors inherent to very large genomic deletions containing potential regulatory elements we generated additional male knockout ES cells targeting exclusively the short NextC2 isoform. More precisely, we deleted a 2kb long genomic region surrounding its first exon and we reasoned that we would be able to more accurately address the functional relevance of NextC2 short isoform in a context where it is the only one to be expressed (*i.e.* in 2i medium). Hereafter, this deletion will be referred to as "small deletion" (**Fig.3.15**), both deletions sharing the same 3' end. For these deletions, three single guide RNAs (sgRNA) were designed and used; two different ones for the 5' part of the deletions and the same sgRNA for the 3' extremity, with the use of the online CRISPR Design Tool (Hsu et al., 2013). Plasmids with the Cas9-Cherry and sgRNA transgenes were lipofectamine-transfected in Tg2a ES cells and first validated for efficient decrease of transcription of short or both isoforms on a batch of cells by RT-qPCR (data not shown). After this validation step, seventy single clones were manually picked for each of the two deletions and were further screened by PCRs on genomic DNA. Sixteen homozygous mutant clones for the small deletion (named small KO clones, SKO) were identified by PCR running on agarose gel and carried the expected size deletion of about 2kb. Accordingly, five homozygous mutant

clones were identified with a deletion of about 55kb for the big deletion (named big KO clones, BKO).

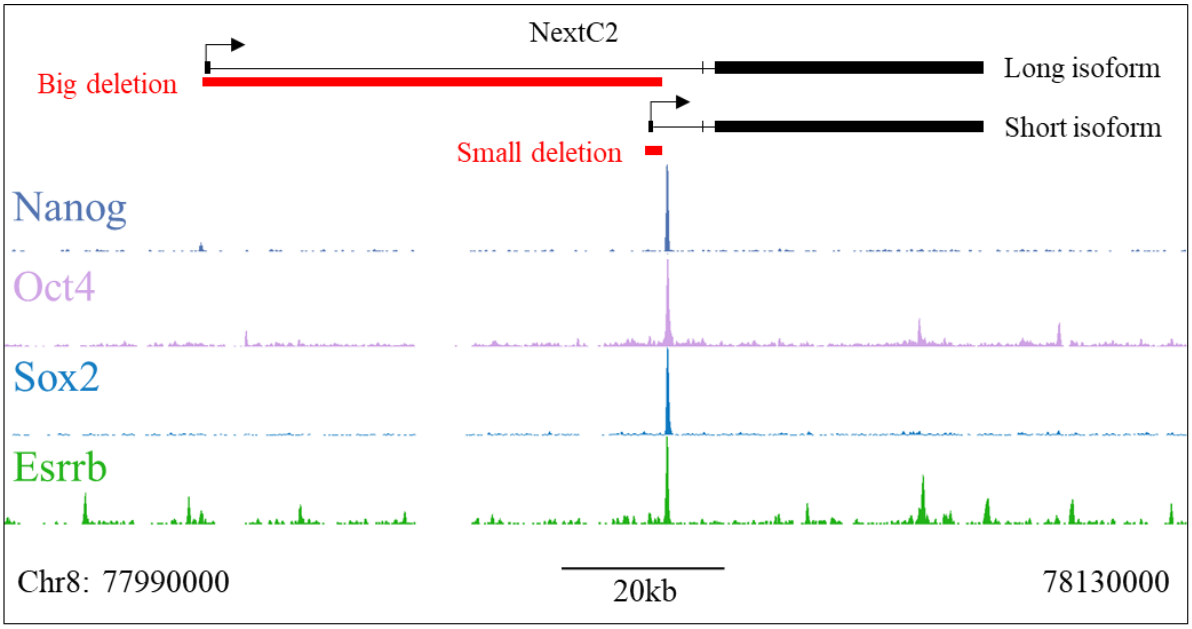


Figure 3.15. NextC2 deletion with CRISPR/Cas9 system targeting short isoform promoter (small deletion) or both isoforms promoters (big deletion). The enhancer region which is enriched in pluripotency TFs binding sites and located downstream the short isoform's first exon should remain intact by the deletions.

We analyzed the karyotype of all the mutant clones (21) and discarded those that did not show regular modal number of forty chromosomes. We finally kept four big KO clones and three small KO clones for further studies. Next, the expression level of the two isoforms was assessed by RT-qPCR assay in the generated mutants clones cultured in FCS conditions. We validated the complete depletion of NextC2 expression in all the big KO clones (BKO1 to BKO4) (**Fig.3.16**). In regard to the small KO clones, we observed that only the long isoform was still detected, slightly lower compared to the WT levels, while the short isoform was no longer expressed. This resulted in lower overall levels of the NextC2 shared exons as seen by the "common part" RT-qPCR in the SKO compared to the WT clones.

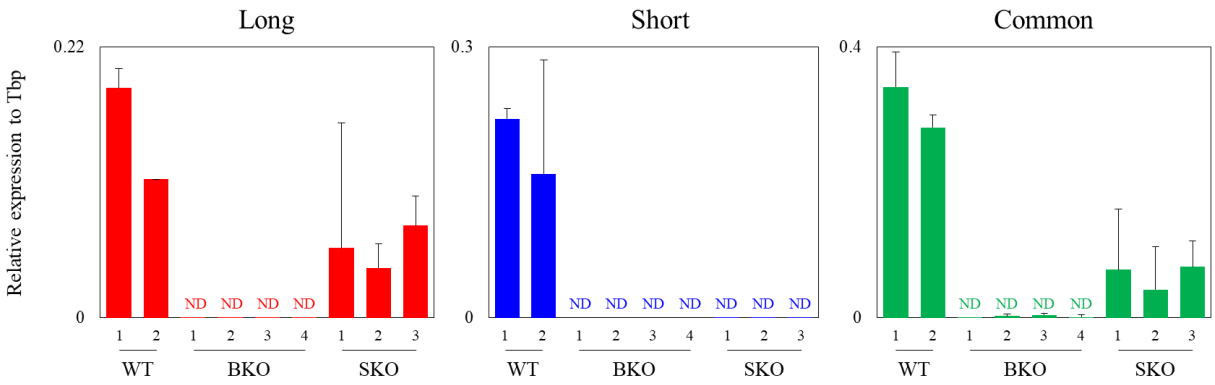


Figure 3.16. NextC2 expression measured by RT-qPCR in the mutant clones of big and small deletions. WT=wild-type Tg2a ES cells, BKO= big deletion KO clones, SKO= small deletion KO clones. Data are normalized to Tbp mRNA and are shown as means and ±sem of two biological replicates.

In addition, to validate the efficient depletion of the targeted transcript on a single cell level, we performed smFISH experiments on WT, BKO and SKO cells. No signal for either the intronic (corresponding long isoform) or the exonic (both isoforms) probes was detected in the big KO clones, validating the results obtained by RT-qPCR. In the small KO clones, the transcriptional activity of the long isoform was somewhat decreased compared to the WT cells (11% instead of 18% positive cells for the intronic probe) whereas the common exonic NextC2 probe was strongly affected since only 27% cells of the population still expressed NextC2 gene compared to the 71% of the cells in the WT population (**Fig.3.17**). This strongly correlated with previous RT-qPCR results and suggests that the small KO clones exclusively express the long isoform. Thus, the smFISH and RT-qPCR analyses of the KO clones validated the efficient CRISPR/Cas9-mediated NextC2 deletion.

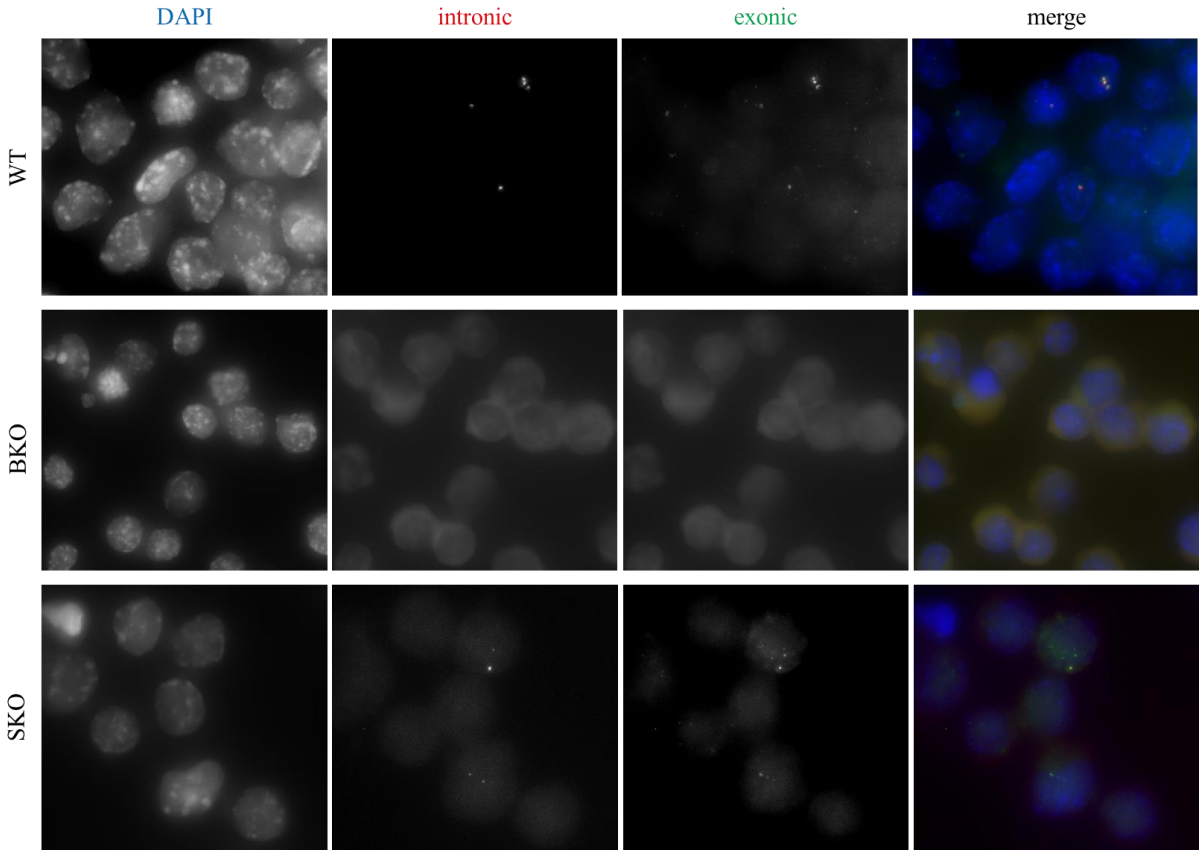


Figure 3.17. smFISH validation of NextC2 mutant clones. In BKO cells the signal is abolished for both probes intronic (red) of long isoform and exonic (green) of both isoforms. Compared to the WT cells, in SKO cells the levels of intronic signal are slightly decreased while the exonic signal is reduced by more than half. DNA is stained with DAPI.

Morphologically, the KO clones of both deletions looked similar to the WT ES cells with no obvious difference in cell death. Part of these observations was verified by a proliferation rate assay in which cells were plated and counted over serial passages. No significant difference was observed in the number of cells that were obtained at the end of each passage (**Fig.3.18**) suggesting that KO clones do not show growth defect phenotypes.

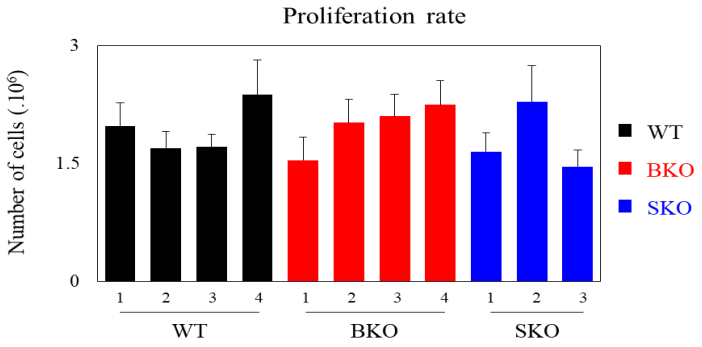


Figure 3.18. Proliferation rate assay of NextC2 mutants in FCS condition. Data are represented by number of millions of cells as means and ±SEM of five countings.

We have established that the short isoform is prevalently expressed in naïve pluripotent ES cells while the long isoform is rather linked to the formative state of pluripotency. Consequently, we reasoned that depending on the culture conditions of the cells, each one of the deletions would have a different impact. Thus, we cultured WT, BKO and SKO ES cells in FCS and 2i conditions and differentiated them in Epiblast-like cells (EpiLCs). We monitored the characteristic morphological changes that occur during the transition from one condition to the other. Globally, we did not notice any remarkable difference or retardation of the mutant cells to form the compact round-shaped colonies of 2i or the flattened EpiLC colonies. We further validated the expected presence or absence of the two isoforms in the WT, small and big KO cells by RT-qPCR analysis. In the SKO clones, the short isoform was no longer detected in any of the conditions while the expression of the long isoform remained mostly unaffected. In the BKO clones, both isoform expression was abolished. Surprisingly, we could still detect some signal in the common exons of the 2i condition in the two types of KO whereas the promoter and the first exon of the short isoform was absent. We thus surmised that a cryptic promoter might be appearing in the KO clones leading to ectopic initiation of transcription independently of the two regular promoters. Finally, we observed an overall decrease in the level of NextC2 transcription in FCS and 2i conditions in the SKO clones, where the short isoform is notably transcribed but not in EpiLCs where the long isoform is predominantly expressed (**Fig.3.19**).

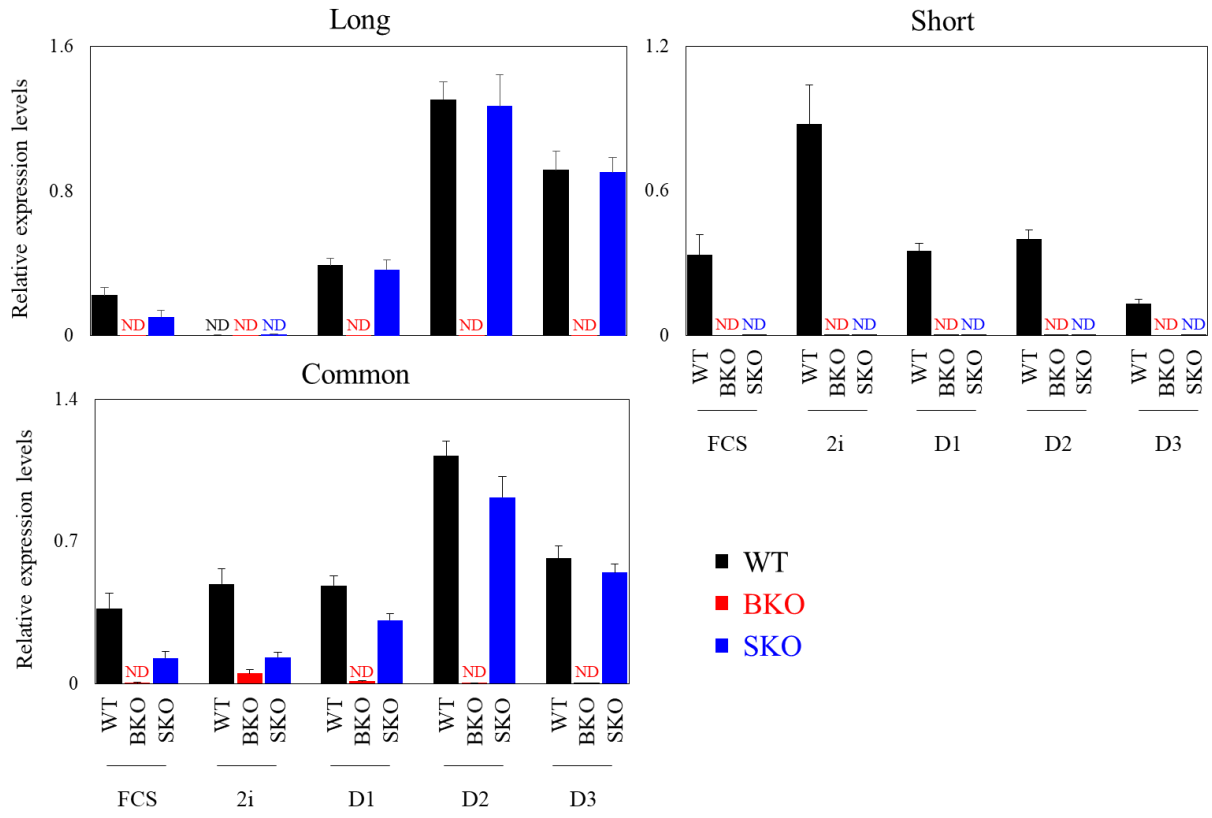


Figure 3.19. NextC2 expression in wild-type and mutant clones. Representative RT-qPCR graphs for NextC2 expression of the long and short isoforms, or their common part in WT (black), big KO (red) and small KO (blue) clones, in FCS, 2i conditions and EpiLC differentiation (D1, D2, D3). Data are normalized to Tbp mRNA and are shown as means and ±sem of two culture experiments and at least three independent clones per WT/BKO/SKO.

The presence of an enhancer region bound by multiple pluripotency TFs downstream of the short isoform promoter (**Fig.3.4**) on one side, and the accumulation of the mature transcript around its transcription site in EpiLCs on the other side, prompted us to evaluate the possibility of a *cis*-regulatory enhancer-like function of NextC2 transcripts. Therefore, we aimed at investigating whether NextC2 depletion had an impact on the transcription of surrounding genes located within the same TAD in our different biological conditions. We looked into publicly available Hi-C data, obtained from the 3D genome browser of the Yen Lab (http://promoter.bx.psu.edu/hi-c/view.php) (data for mouse ES used cells from Dixon et al., 2012) in order to determine the NextC2 TAD and the genes located within. Then, we measured the expression levels of six out of the twelve TAD genes of the NextC2 locus by RT-qPCR. In contrast to our hypothesis, none of the tested adjacent genes showed any transcriptional response to the NextC2 deletions, neither in BKO nor in SKO mutant cells (representative two genes shown in **Fig.3.20**). Noteworthy, the expression of the two genes flanking the NextC2 locus have not been assessed yet due to difficulties in the design of PCR

primers. These results suggest that, NextC2 does not exert a regulatory function *in cis* on the expression of the genes that are found in close proximity to its locus. Nevertheless, this does not exclude the possibility that the NextC2 locus could be involved in regulatory long range interactions *in cis* or together with interchromosomal partners *in trans*.

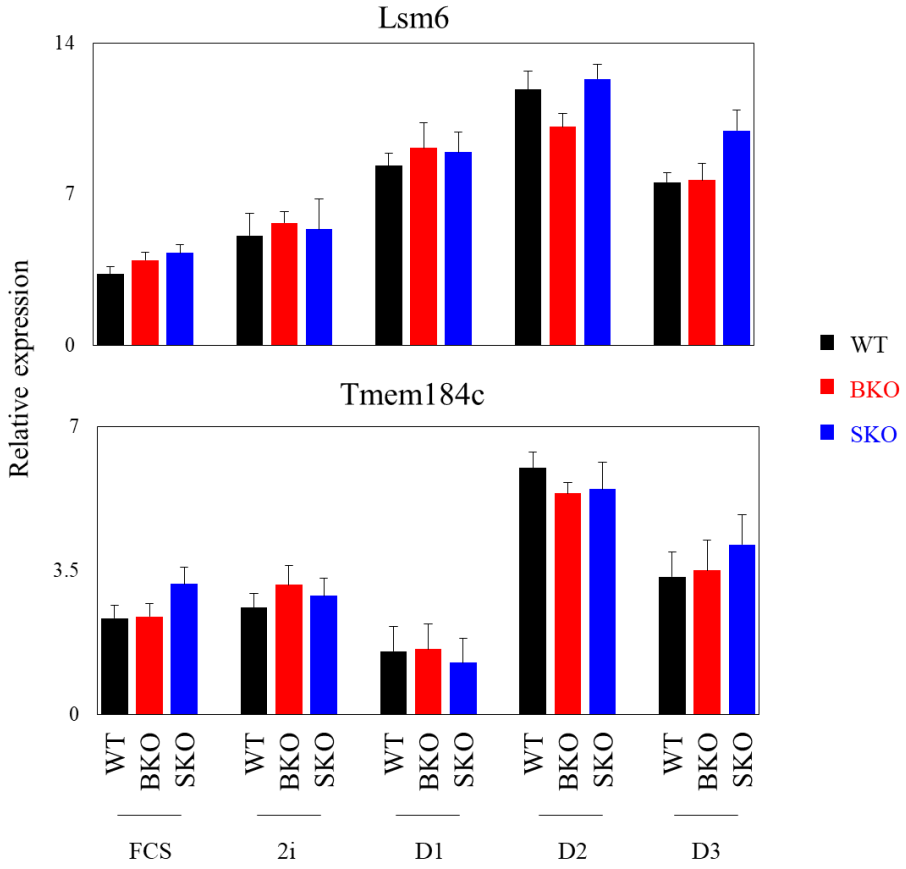


Figure 3.20. Genes within the NextC2 topologically associated domain (TAD). RT-qPCR analysis of expression levels of two neighboring genes of NextC2 in WT (black), big KO (red) and small KO (blue) clones, in FCS, 2i conditions and EpiLC differentiation (D1, D2, D3). Data are normalized to Tbp level and are represented as means and ±sem of two culture experiments and at least three independent clones per WT/BKO/SKO.

Subsequently, we decided to proceed with the assessment of pluripotency and differentiation genes expression across all our conditions and clones by RT-qPCR. In FCS and 2i conditions, we obtained similar levels among the WT, BKO and SKO clones for the pluripotency markers that we evaluated (Nanog, Oct4, Sox2, Esrrb, Klf4, Prdm14, and Rex1). Strikingly, while most of these genes had comparable levels during EpiLC differentiation in WT and KO clones, we noticed that Sox2, Rex1 and Prdm14 exhibited higher expression at day two and three (D2, D3) of the assay in all clones carrying the small deletion (SKO) compared to WT or BKO (**Fig.3.21A**). We next looked at epiblast marker genes that should be

activated during the course of the EpiLC differentiation. Interestingly, we observed once again a discrepancy for the SKO mutants, at EpiLCs D2 and D3, in inducing the expression of Wnt3 and Fgf5 genes -showing reduced levels of these genes- compared to the WT and BKO cells (**Fig.3.21B**). In addition, we checked few primitive endoderm marker genes (Gata4, Gata6, Sox17), and observed that the mutants bearing the small deletion stood out by expressing higher levels of Gata6 and Sox17 compared to WT and BKO cells, notably at D2 and D3 of EpiLC differentiation (**Fig.3.21C**). Collectively, these data suggest that the mutants, where only the short isoform was targeted for deletion (SKO), show a mild molecular delay in establishing the transcriptomic signature of epiblast-like cells corresponding to lower levels of Epiblast markers expression as well as higher expression of primitive endoderm and naïve pluripotency markers (Rex1, Prdm14) during EpiLC conversion. In order to address the transcriptional consequences of the deletions of the NextC2 isoforms on a genome-wide level and to subsequently infer the potential functionality of NextC2 gene, we decided to analyze our full dataset with RNA-seq. We conducted the sequencing on WT, BKO and SKO cells (two independent clones of each) in FCS, short- and long-term 2i culture, EpiLCs transition D1-D2-D3 and RA differentiation D1 and D3.

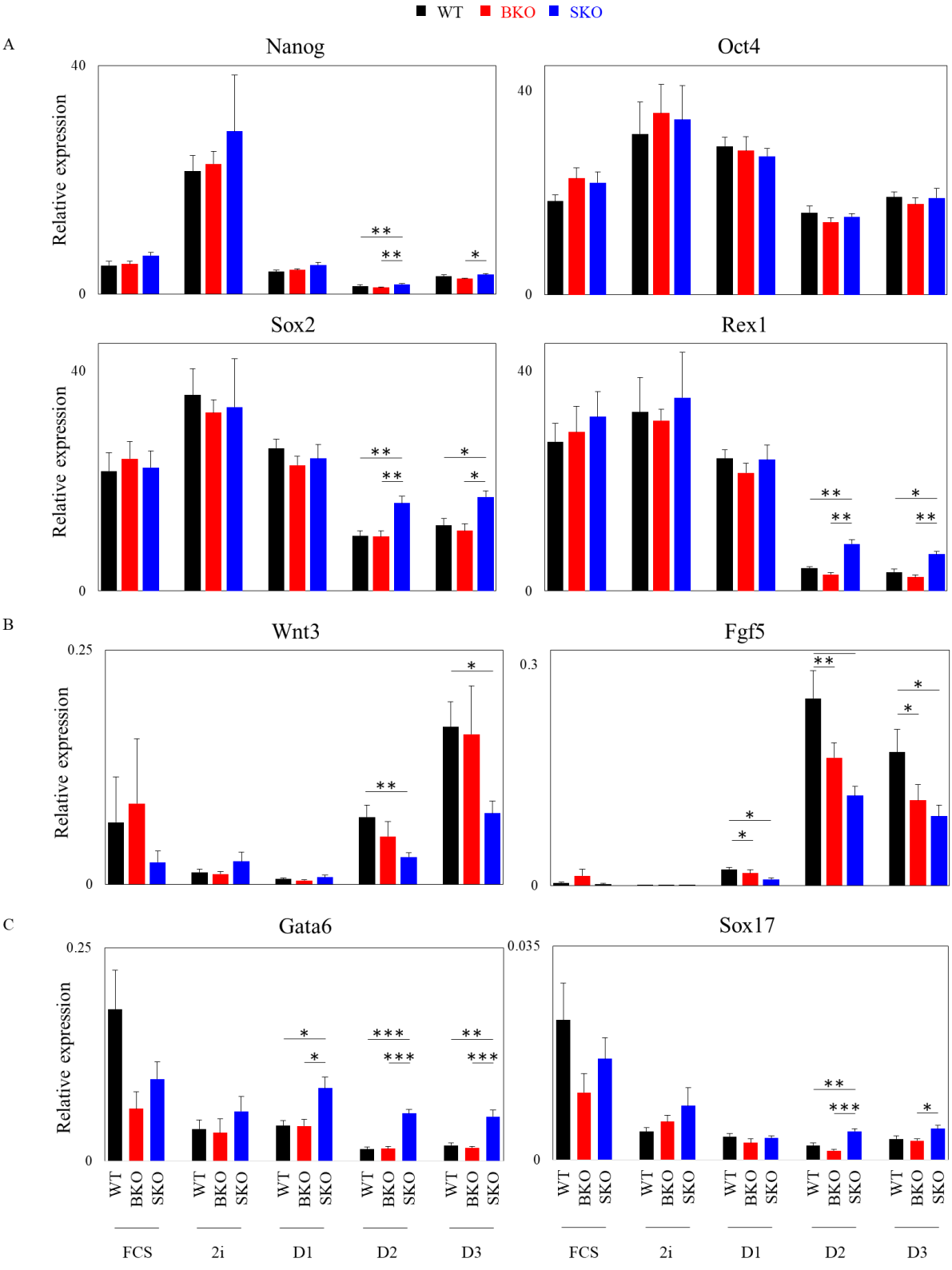


Figure 3.21. Expression levels of pluripotency, epiblast and primitive endoderm markers analyzed by RT-qPCR. WT (black), big KO (red) and small KO (blue) clones were cultured in FCS, 2i conditions and EpiLC differentiation (D1, D2, D3). Data are represented relative to Tbp expression level, as means and ±sem of two culture experiments and at least three independent clones per WT/BKO/SKO. Statistical analyses were performed on log transformed values by two-tailed unpaired t-test: *p<0.05, **p<0.01, ***p<0.001.

Inducible transcript KD by CRISPRi

The initial approach for a loss-of-function assay to investigate the molecular and phenotypical consequences of NextC2 deletion, was performed using the CRISPR/Cas9 system to delete parts of the NextC2 locus. However, like in the case of NextC1, we also aimed at generating a cell line where the genomic locus would remain intact and we could induce NextC2 knockdown, for each of the two isoforms separately, in an inducible manner (DOX). To this end, we used the CRISPR interference (CRISPRi) system, as previously for NextC1. Single guide RNAs (sgRNA) were designed around the promoters of the long and short isoforms, with the help of the CRISPR Design Tool (http://crispr.mit.edu/) (**Fig.3.22A**). The promoter region of the long isoform is harbored in an ERVK repetitive element (LTR9A) not allowing the design of sgRNAs in the close proximity to its TSS. First, we tested the efficiency of the designed sgRNAs by cotransfecting them with the dCas9-KRAB-BFP transgene (expressed under the control of DOX inducible promoter and linked to a blue fluorescent protein, BFP). On the bulk of transfected cells, we attested a decrease in expression of the long isoform, with gA, and a decrease in the expression of the short isoform, with g1 and g2 under DOX induction (**Fig.3.22B**). Furthermore, we manually picked twenty four clones for each transfection with the guide RNAs gA, g1 and g2. RT-qPCR analysis was performed on the clones that exhibited the highest BFP signal under DOX induction. The clones which showed the most efficient downregulation of NextC2 transcripts (to almost undetectable levels) were kept for further studies (**Fig.3.22C**). Upon DOX induction, the clones showed normal ES cell morphology and no differences were observed in the expression of few pluripotency and differentiation marker genes that were checked by RT-qPCR (data not shown). These generated clones were also included in the RNA-seq experiment that was conducted for the NextC2 deletion mutant clones (detailed in the previous section). By sequencing this set of loss-of-function clones we aimed at determining the genes the expression of which is regulated by NextC2, whether due to the transcript itself or the genomic locus of the NextC2 gene.

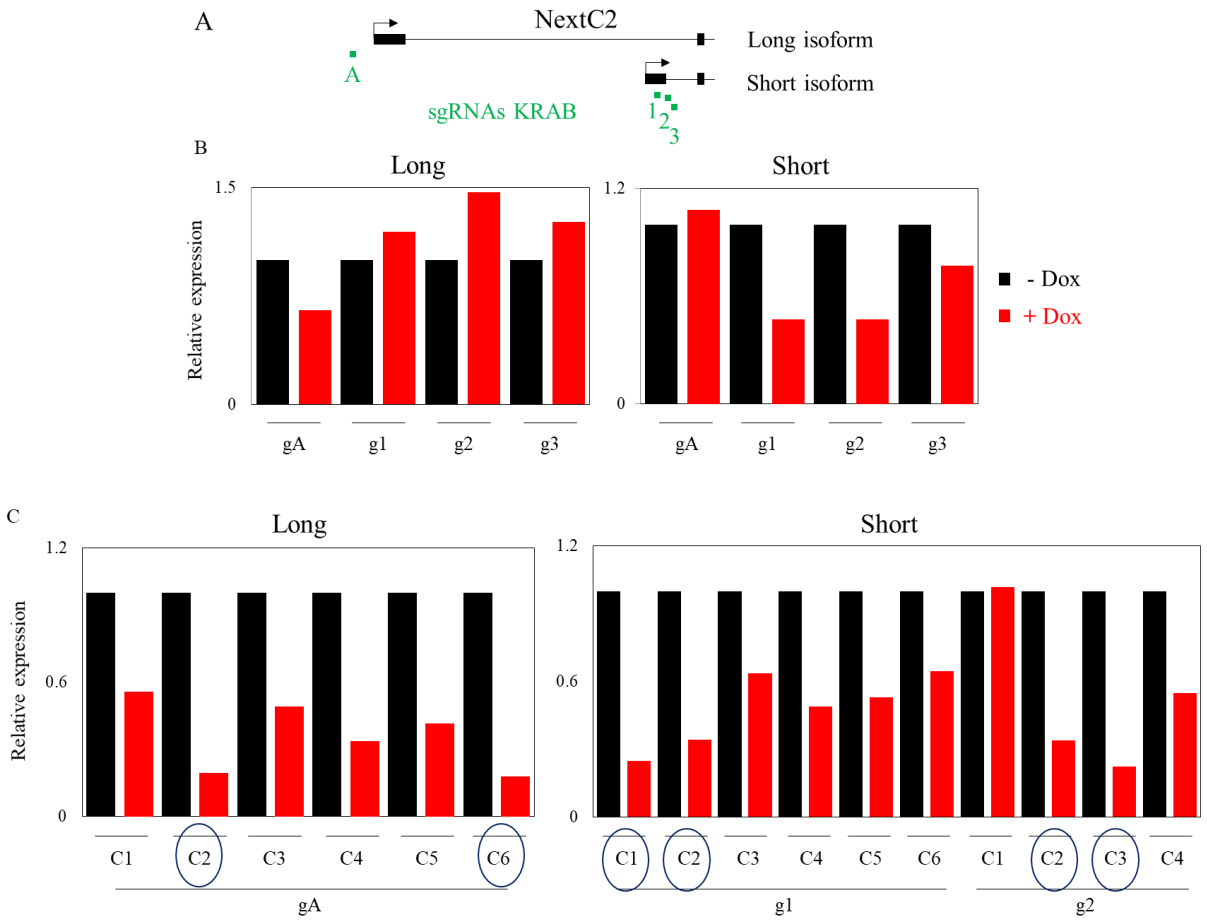


Figure 3.22. CRISPRi for NextC2 knockdown. A) Localization of the designed sgRNAs in respect to the two promoters of NextC2; gA for the long and g1, g2, g3 for the short isoform promoters. B-C) RTq-PCR analysis on expression of NextC2 isoforms upon DOX treatment (+DOX, 72h). Data are normalized to Tbp mRNA and shown as fold change to untreated samples (-DOX). B) Bulk of transfected ES cells with gA, g1, g2, g3. C) Picked clones that shown the best KD effect using gA, g1 and g2. Highlighted the clones that were finally chosen for further use.

ii. Gain of function

We used the SunTag CRISPR activation system, as shown before for NextC1, to perform a gain-of-function approach on NextC2 isoforms. We designed two and four sgRNAs targeting the long and the short isoform promoters respectively (**Fig.3.23A**). Initially, the efficiency of the sgRNAs was evaluated in one of our SunTag clones (C1 in **Fig.3.23B**). RT-qPCR analysis was performed on transfected cells with each sgRNA separately and DOX treatment for 72 hours. We decided to further use the two sgRNAs for the long (gA and gB) and the short isoforms (g1 and g3) and further sub select NextC2 overexpressing clones in the two SunTag (C1, C2) clones in order to exclude clone or sgRNA artefactual effects. We further

picked twelve clones from each transfection and assessed their GFP and BFP signals (linked to VP64 and dCas9 respectively) upon DOX induction. Then, we performed RT-qPCR analysis on the five clones with the highest GFP and BFP signal (**Fig.3.23C, D**). We finally selected the one subclone per batch that reached the highest induction of the targeted isoform (highlighted in **Fig. 3.23C, D**) for future studies. Upon DOX induction, the morphology of the growing ES cells was observed and the expression levels of few pluripotency and differentiation genes were monitored by RT-qPCR. Both assessments appeared to show normal behavior (data not shown). The generated NextC2 overexpressing clones were submitted for RNA-seq along with the loss-of-function samples previously mentioned. Collectively, with this massive RNA-seq dataset we hope to shed some light on the molecular consequences of the manipulation of expression of NextC2 and therefore unveil the functional relevance of this lncRNA. The integrated analysis of these complex dataset is currently under investigation.

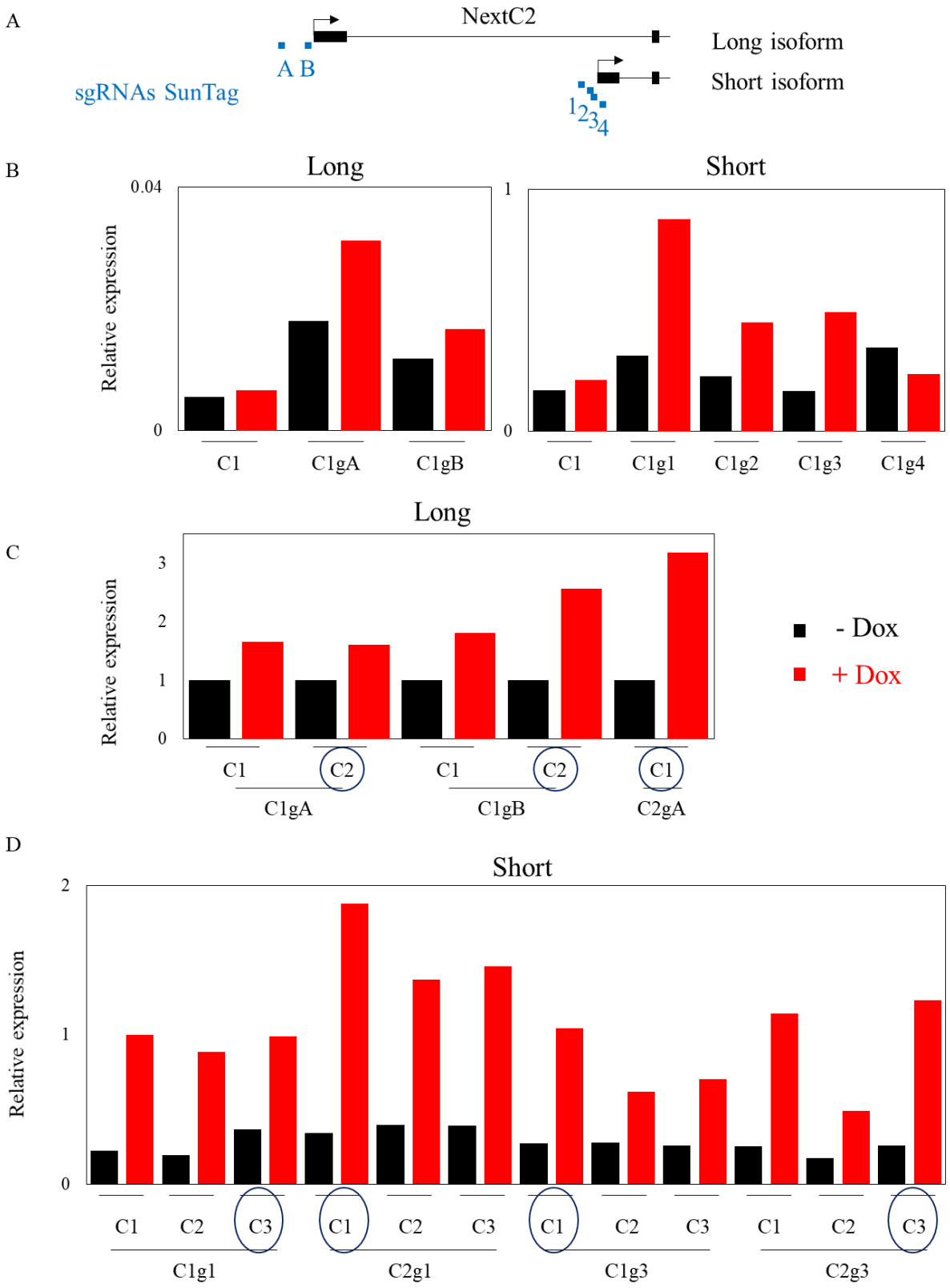


Figure 3.23. NextC2 endogenous overexpression with CRISPRa SunTag system. A) Localization of the sgRNAs (gA, gB, g1, g2, g3, g4) in respect to the two alternative promoters of NextC2. B-D) RTq-PCR analysis of NextC2 expression upon DOX treatment (+DOX, 72h). Data are normalized to Tbp mRNA. B) Batch of transfected cells with each sgRNA and control SunTag clone C1 without any

sgRNA. C) Clones with sgRNAs targeting the long isoform. D) Clones with sgRNAs targeting the short isoform. Final clone selection for further studies is highlighted in C and D.

G. Discussion

The second lncRNA candidate which we aimed at characterizing was NextC2. This gene initiates from two alternative promoters giving rise to two isoforms that have mutually exclusive first exons but share the rest of same exons. The two isoforms seem to have a mirror-image pattern of expression in different biological contexts where either one or the other is prevalently transcribed. The short NextC2 isoform was retrieved as an unannotated ES cell-stage specific long non-coding transcript in a study analyzing the reprograming process from mouse fibroblast to induced pluripotent cells (iPS) (Hussein et al., 2014).

We found that both isoforms have very rapid responses to Nanog fluctuations and while the short isoform follows the expression pattern of Nanog, the long isoform manifests exactly the opposite behavior. To reinforce this finding of ours, we could further look into the expression profile of the two isoforms in populations of Nanog-high and Nanog-low expressing cells, using a Nanog fluorescent reporter cell line and FACS sorting (Chambers et al., 2007). In that way, we could get an insight into how the two isoforms are expressed in more physiological cellular conditions where the Nanog gene is endogenously and rapidly switching from active to inactive state (and vice versa).

Furthermore, the expression of the long isoform being negatively regulated by Nanog, quickly downregulated in FCS to 2i transition, and upregulated upon EpiLC differentiation, resembled the expression profile of a TF, key driver of Epiblast differentiation, Otx2 (Acampora et al., 2013). Interestingly, we found in a published ChIP seq dataset performed in ES cells and EpiSC (Epiblast Stem Cells) (Acampora et al., 2016) that Otx2 binding, while already present in ES cells at low levels, was strongly strengthened in EpiSCs on the promoter of the long isoform (data not shown). We could therefore hypothesize that the long isoform might be tightly regulated by Otx2 while the short isoform is directly targeted by Nanog, among other TFs. The fact that Otx2 is a fast and strong target of Nanog, as well as the antagonistic nature between Otx2 and Nanog (Acampora et al., 2017; Festuccia et al., 2012; Heurtier et al., 2018) could possibly partially explain the switch of the alternative promoters that are active at different pluripotent states.

Apart from TF controlled regulation of expression, the switch of promoters in NextC2 expression could be potentially instructed by an epigenetic switch. In order to explore this idea, we looked for a possible enrichment in specific histone modification marks at the locus. We did not find at either of the two isoforms a H3K27me3 mark, neither in ES cells nor in EpiSCs (data from Factor et al., 2014; Marks et al., 2012). Furthermore, we looked at the DNA methylation levels (5mC mark) of the NextC2 gene and found that in ES cells the locus is not methylated while upon mesendodermal differentiation of ES cells the proximal promoter acquires some DNA methylation marks (data from Yu et al., 2013). Therefore, DNA methylation which is low in naïve cells and grows high in the primed pluripotent cells (Lee et al., 2014) could be regulating the transcriptional switch between the short and the long isoform. It would be interesting to investigate available datasets for DNA methylation profiles specifically during the ES cell to EpiLC transition, in order to verify that DNA methylation is established at the short isoform promoter upon the exit of the cells from naïve to primed pluripotency. Finally, chromatin architecture could also be involved in the promoter switching of NextC2 and more specifically the CTCF-mediated promoter-enhancer interactions (Greenberg et al., 2018; Murrell et al., 2004). There are multiple CTCF binding sites across the locus of NextC2, some of which ES specific (**Fig.3.4**), that could be potentially participating in partitioning the locus and establishing different interactions for each promoter with their respective enhancers, as it has been recently demonstrated to be the case of the Liz/Zbdf2 locus (Greenberg et al., 2018). First, it would be of interest to look into the orientations of the CTCF motifs at the locus and search whether there are any convergent positioning, as cohesin-mediated chromatin loops require convergent orientation for loop establishment (Rao et al., 2014). We could, additionally, look into the role of CTCF in establishing interactions at the NextC2 locus by performing 4C experiments using as viewpoints the different CTCF binding sites. Moreover, the 4C assays should be done in our big and small NextC2 KO clones, to assess the potential new interactions that could emerge for the remaining CTCF binding sites upon the deletion.

By smFISH assay we were able to verify that NextC2 is a nuclear lncRNA which is expressed in a big fraction of the cell population. We used a distiguishing probe that could specifically recognize the primary long isoform and a second probe that could recognize all the exons of NextC2. Given the exonic structure of the isoforms, we could not distinguish the two amongst them, in conditions that both isoforms are transcribed (FCS and EpiLC). Nevertheless, we could conclude that the mature transcript of both isoforms can be found dispersed in the

nucleus, since we detected scattered molecules with the exonic probe in 2i grown cells (hence corresponding only to the short isoform) and in FCS grown small deletion KO clones (hence corresponding only to the long isoform).

In order to gain insight into any potential functionality of NextC2, we proceeded to loss-of-function experiments by deleting either both promoters or only the proximal one. Surprisingly, we observed that removing the proximal promoter led to decreased production of the long isoform (SKO in FCS condition, **Fig.3.19**). This finding could suggest a potential regulatory role of the deleted region over the distal promoter. One possible hypothesis might be that the enhancer downstream the proximal promoter also participates in the activation of its distal counterpart. Moreover, this would imply that the enhancer function has been affected by the deletion. In order to address this issue, we should perform a ChIP experiment for few of the TFs known to bind to this region and control for their ability to bind in the mutant clones.

Another intriguing result we obtained from our NextC2 mutant clones was the transcription of the common exons of NextC2 isoforms in 2i conditions (**Fig.3.19**). We did not anticipate such an expression to occur since the only active promoter in 2i (short isoform) is deleted in both types of KO clones, BKO and SKO. Of note, it has been reported that in mammals a multitude of intragenic enhancers can serve as alternative promoters and produce new isoforms of a given gene, that is usually cell-type and developmental stage-specific, adding up to the complexity of the transcriptome (Kowalczyk et al., 2012). In NextC2, an enhancer element is located shortly downstream of the proximal promoter. We thus wondered whether this enhancer could be initiating the low transcription we observed by RT-qPCR. Therefore, we looked into our RNA-seq data of the KO clones, and indeed observed an alternative isoform being produced in 2i conditions (in both short- and long-term 2i culture) in the mutant clones (**Fig.3.24**). This new transcript indeed initiates from the intragenic enhancer region, producing a first exon and then following the exonic structure of the regular isoforms. However, the level of expression of this isoform can be considered as minor compared to WT samples. This observation can be linked to our previous comment on the promoter and enhancer marks of the NextC1 promoter, and whether an exclusive distinction between promoter and enhancer on the functional level is possible. In the case of NextC2, an intragenic enhancer element can easily substitute for the loss of its associated promoter in the expression of the downstream target gene. Therefore, the function of an enhancer or promoter DNA element is possibly dependent

on the context of the TFs bound around and the existence of a transcriptional unit in the vicinity for the transcription to be initiated.

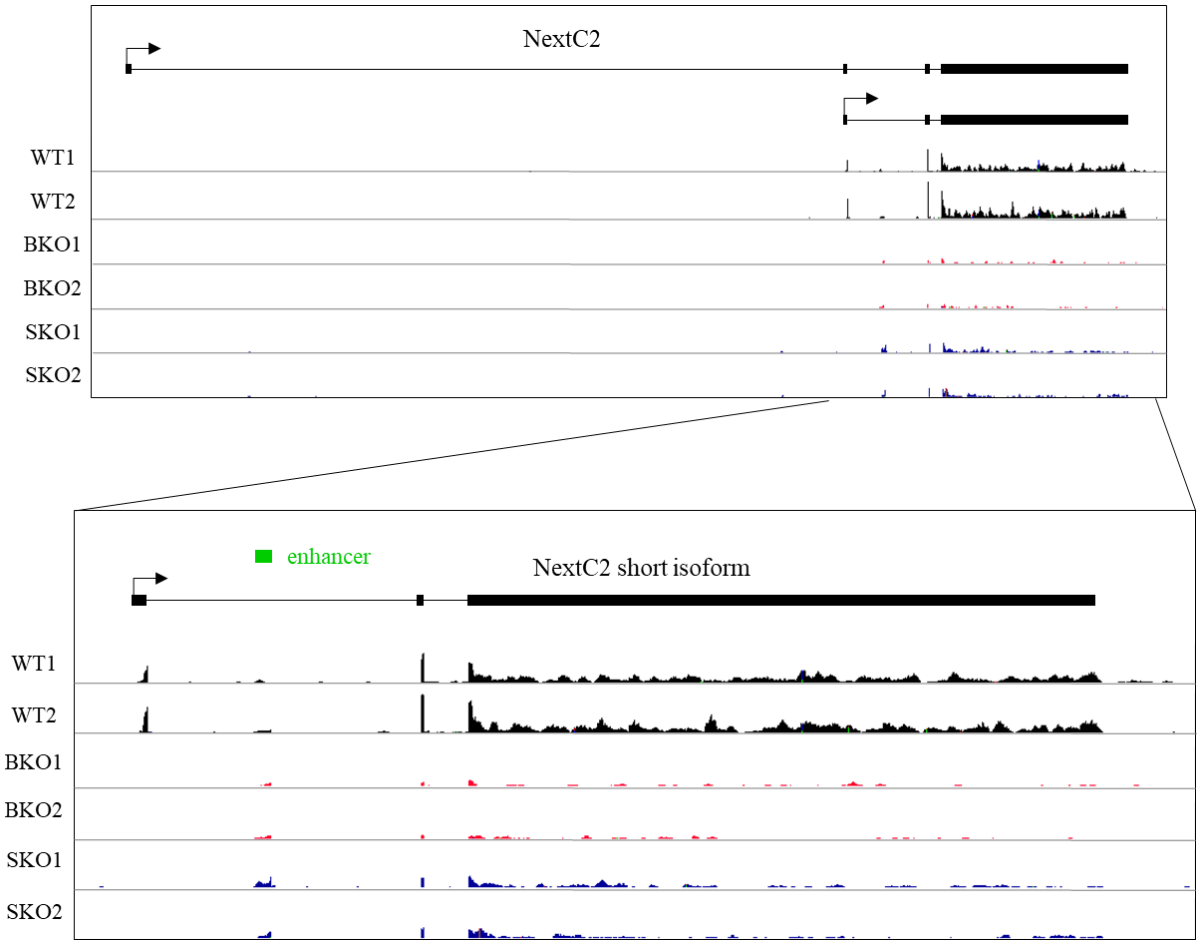


Figure 3.24. Screenshot of IGV browser of NextC2 locus. RNA-seq coverage in WT (black), big deletion mutant (red) and small deletion mutant (blue) clone cultured long-term in 2i. Reads coverage have been group-auto scaled. Upper panel: full locus of NextC2 to demonstrate that only short isoform is expressed in WT cells; lower panel zoom in on the short isoform. In mutant clones alternative transcription initiates from an intragenic enhancer, depicted in green.

Our data indicated a mild molecular delay of the clones bearing the small deletion in establishing the transcriptomic signature of epiblast-like cells. These mutants exhibit lower expression levels of Epiblast markers as well as higher levels of primitive endoderm and naïve pluripotency markers during EpiLC conversion. Strikingly, these molecular consequences are only observed in the SKO clones but not in the BKO, despite that fact that the deleted genomic locus in the SKO is also deleted in the BKO. One possible hypothesis we can make is that the small deletion (around 2kb) could have affected the regulation of one or several target genes but that this effect could be smoothened and somehow quenched by the big deletion, where a 55kb locus has been removed. We could imagine that upon the big deletion the chromatin

organization is rearranged in such a way that potentially new interactions are established that could substitute for the loss of the interactions provoked by the small deletion. The different interactions established could be on an intra-TAD level or on long-range inter-chromosomal scale, and could be elucidated by the 4C experiments proposed earlier using the CTCF binding sites of the locus as anchor points. Another possible explanation could be that the small deletion is producing a truncated NextC2 long isoform, which can no longer exert its function, resembling the dominant negative effect of protein mutations. In parallel, this hypothesis would suggest that the abolishment of both isoforms does not give an evident molecular phenotype on the few genes whose expression was assessed by RT-qPCR in the contexts that we examined, *i.e.* in FCS, 2i and EpiLC. In order to understand which one of these two or other scenarios is valid and whether NextC2 has any functional relevance, we will thoroughly analyze our RNA-seq data combining the KO clones and the inducible KD clones (CRISP/dCas9-KRAB). The KD clones will help us to distinguish whether it is the locus that is important for the functionality of NextC2 (first hypothesis), if the differentially expressed (DE) genes will be different in the two approaches of loss-of-function assays, or the transcript itself (second hypothesis) if the DE genes will be common in the two assays. The obtained results from the RNA-seq of the overexpressing NextC2 clones will build on an additional layer of information regarding the target genes of NextC2.

For the time being, we do not have a clear view whether the two isoforms of NextC2 are separate transcripts with possibly independent functions or simply two different isoforms with similar functions regulated to be expressed in different biological states. So far we have findings that could support either case; difference in RNA sequence and transcript stability would argue for the first case while the overall levels of the common exons that remain at comparable levels across the FCS, 2i and EpiLC samples -despite the different expression profiles of the two transcripts separately- would argue for the second. The latter would suggest that initiation from two alternative promoters might ensure the maintenance of the expression of NextC2 along naïve and early commitment steps of development. We aim at answering those questions with the analysis of our RNA-seq data and deciphering the role of NextC2 gene in embryonic stem cell biology.

Transcription factor activity and nucleosome organisation in mitosis

Nicola Festuccia[1,4], Nick Owens[1,4], Thaleia Papadopoulou[1], Inma Gonzalez[1], Alexandra Tachtsidi[1,3], Sandrine Vandoermel-Pournin[2], Elena Gallego[1], Nancy Gutierrez[1], Agnès Dubois[1], Michel Cohen-Tannoudji[2], 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]Mouse Functional Genetics, Department of Developmental and Stem Cell Biology, Institut Pasteur, CNRS UMR 3738, 25 rue du Docteur Roux, 75015 Paris, France.
[3]Sorbonne Université, Collège Doctoral, F-75005 Paris, France.
[4]Equal contribution

Mitotic bookmarking transcription factors (BFs) maintain the capacity to bind to their targets during mitosis, despite major rearrangements of the chromatin. While they were thought to propagate gene regulatory information through mitosis by statically occupying their DNA targets, it has recently become clear that BFs are highly dynamic in mitotic cells. This represents both a technical and a conceptual challenge to study and understand the function of BFs: first, formaldehyde has been suggested to be unable to efficiently capture these transient interactions, leading to profound contradictions in the literature; second, if BFs are not permanently bound to their targets during mitosis, it becomes unclear how they convey regulatory information to daughter cells. Here, comparing formaldehyde to alternative fixatives we clarify the nature of the chromosomal association of previously proposed BFs in embryonic stem cells: while Esrrb can be considered as a canonical BF that binds at selected regulatory regions in mitosis, Sox2 and Oct4 establish DNA sequence independent interactions with the mitotic chromosomes, either throughout the chromosomal arms (Sox2) or at pericentromeric regions (Oct4). Moreover, we show that ordered nucleosomal arrays are retained during mitosis at Esrrb bookmarked sites, whereas regions losing transcription factor binding display a profound loss of order. By maintaining nucleosome positioning during mitosis, Esrrb might ensure the rapid post-mitotic re-establishment of functional regulatory complexes at selected enhancers and promoters. Our results provide a mechanistic framework that reconciles dynamic mitotic binding with the transmission of gene regulatory information across cell division.

mitotic bookmarking | mitosis | nucleosome | transcription factor | fixation
Correspondence:
pnavarro@pasteur.fr

Introduction

During mitosis, the chromatin is drastically condensed and reconfigured to enable the equitable partition of the genetic material between the two daughter cells **(Ma et al. 2015)**. This leads to a strong decrease in transcriptional activity and to the general reduction of transcription factor (TF) binding throughout the genome. Loss of TF binding is further accentuated by the stereotypical phosphorylation of many regulators during mitosis, leading to an intrinsic reduction of their ability to bind DNA. This is particularly well illustrated by the systematic phosphorylation of C2H2 zinc finger TFs such as YY1 **(Rizkallah et al. 2011; Rizkallah and Hurt 2009)**, but has also been observed for other TFs such as Oct4 and Sox2 **(Shin et al. 2016; Qi et al. 2016)**. Moreover, the breakdown of the nuclear envelope, and the consequent increase of the volume that TFs can freely explore, leads to a decrease of TF concentration. This process naturally inhibits the ability of TFs to scan DNA for their binding motifs. Therefore, many processes occur simultaneously to temporarily halt gene regulation and transcription during mitosis. The mechanisms by which daughter cells accurately re-establish an environment permissive for efficient transcriptional activation early in interphase remain unknown **(de Castro et al. 2016)**. One potential mechanism is known as mitotic bookmarking: some TFs have the ability to interact with their DNA binding sites during cell division. These TFs, known as mitotic bookmarking factors (BFs), are believed to directly convey gene regulatory information from mother to daughter cells, as illustrated by Gata1 **(Kadauke et al. 2012)**, FoxA1 **(Caravaca et al. 2013)** and Esrrb **(Festuccia et al. 2016)**. Nonetheless, the molecular mechanisms underpinning this function remain to be elucidated **(Festuccia et al. 2017)**. BFs are highly dynamic during mitosis and often exhibit reduced residence times on the chromatin. Therefore, the function of BFs is not simply mediated by their stable retention at enhancers and promoters. Instead, their transient binding activity may preserve specific chromatin features at bookmarked sites. These features would represent the inherited properties driving and accelerating the reassembly of functional regulatory complexes early in the following interphase. Remarkably, even though the chromatin is highly condensed during mitosis, gene regulatory elements remain globally accessible **(Hsiung et al. 2015)**. This is particularly true at active promoters, perhaps reflecting their low but nevertheless significant mitotic activity, as recently reported **(Palozola et al. 2017)**. Enhancers, in contrast, show more variable degrees of chromatin accessibility. Yet, mitotic chromatin accessibility does not seem to correlate with mitotic binding, at least in the case of bookmarking by Gata1 in erythroblasts **(Kadauke et al. 2012)**. Moreover, the maintenance of chromatin accessibility does not preclude the possibility that nucleosome positioning in mitotic cells is highly modified, as previously suggested **(Kelly et al. 2010)**. Hence, further studies are required to clarify whether regulatory elements do indeed maintain a local chromatin architecture compatible with TF binding in mitotic cells, and how mitotic bookmarking correlates with and ultimately drives nucleosome organ-

isation. An essential condition to understand mitotic bookmarking processes is to accurately identify BFs and their mitotic binding sites. However, this has remained a difficult task because, as reported nearly 15 years ago **(Pallier et al. 2003)**, the most commonly used cross-linker, formaldehyde, leads to the artificial depletion of TFs from mitotic chromosomes **(Pallier et al. 2003; Teves et al. 2016)**. To circumvent this problem mitotic bookmarking activity has been explored using live imaging of tagged TFs. Even so, whether the global chromatin association of certain TFs detected by microscopy reflects the sum of site-specific interactions remains to be demonstrated. Diverse modes of binding, others than those involving base-specific interactions, may be responsible for the macroscopic decoration of the chromosomes by TFs, as we proposed earlier **(Festuccia et al. 2017)** and was clearly demonstrated for FoxA1 **(Caravaca et al. 2013)**. These interactions with the chromatin, or with other constituents of mitotic chromosomes, might be extremely transient and not easily captured by formaldehyde. In support of this distinction between global and site-specific interactions, several TFs have been efficiently captured at their mitotic binding sites using formaldehyde **(Festuccia et al. 2017)**, despite its seeming incapacity to cross-link TFs on mitotic chromosomes. Yet, it remains to be proven whether formaldehyde generally fails in capturing transient DNA-specific interactions, leading to the loss of enrichment of BFs on the chromosomes, or whether the interactions sustaining the global retention of BFs are distinct from those involved in TF binding to DNA. This does not only represent an important technical question; rather, it directly interrogates the nature and, hence, the function, of the interactions established between TFs and mitotic chromosomes: while global, dynamic and DNA sequence independent interactions may increase the concentration of TFs in the vicinity of DNA, possibly facilitating the re-establishment of binding in the following interphase, authentic mitotic bookmarking of promoters and enhancers may confer specificity to these processes and provide robustness to the post-mitotic resuscitation of gene regulatory networks.

In this manuscript, we combine (para-)Formaldehyde (PFA) with alternative fixatives that are able to capture hyper-dynamic protein-protein interactions, such as Disuccinimidyl Glutarate (DSG; **Tian et al. 2012**), to study the association of TFs with mitotic chromatin. We report that DSG, in contrast to PFA, preserves the global interactions of two TFs, Esrrb and Sox2, previously proposed to display strong mitotic bookmarking activity in embryonic stem (ES) cells **(Deluz et al. 2016; Festuccia et al. 2016; Teves et al. 2016; Liu et al. 2017)**. Nonetheless, irrespective of the fixative used, DNA sequence-specific interactions can be detected for Esrrb at thousands of target sites, but not for Sox2. Moreover, imaging after DSG fixation unmasks a particular behaviour of another TF previously suggested to act as a BF, Oct4: both in interphase and in mitosis, Oct4 is focally enriched on pericentric heterochromatin. In mitosis, Oct4 is also largely excluded from the chromatid arms and, similarly to what we observe for Sox2, loses its ability to engage in site-specific interactions. These observations strongly argue against the idea that the global decoration of the mitotic chromosomes can be taken as an indication of a sequence-specific bookmarking activity. Integrating our genome-wide localisation studies with additional assays interrogating chromatin accessibility, nucleosome positioning and stability (Table S1), we conclude that mitotic bookmarking, particularly by Esrrb, is strongly associated with the preservation of an interphase-like nucleosomal organisation. At bookmarked sites, both in interphase and mitosis, Esrrb motifs lie at the centre of nucleosome depleted regions surrounded by phased arrays of nucleosomes. At loci losing TF binding, the mitotic nucleosomes are vastly reorganised and display increased fragility as compared to interphase. We conclude that TFs display a range of behaviours in mitotic cells that can be captured with distinct fixatives. This variability reflects the ability of some TFs, like Esrrb, but not of all factors enriched on mitotic chromosomes, to bind at specific regulatory elements efficiently and to maintain a nucleosomal organisation compatible with the rapid re-establishment of regulatory complexes in interphase.

Results

The global localisation of TFs to mitotic chromosomes is preserved upon DSG fixation. Several TFs have been shown to seemingly coat the mitotic chromosomes when fusion proteins with fluorescent proteins or tags are used in live imaging approaches **(Festuccia et al. 2017)**. This is the case of Esrrb **(Festuccia et al. 2016)**, which we previously showed to decorate mitotic ES cell chromatin using GFP (Fig. 1A), Tdtomato and Snap-tag fusions. However, upon Formaldehyde fixation, several TFs capable of coating the mitotic chromosomes, seem to be globally delocalised and crosslinked outside of the chromosomes **(Pallier et al. 2003; Teves et al. 2016)**. We first aimed to test whether this is also the case for Esrrb. As expected, we observed a clear depletion on the mitotic chromosomes (Fig. 1B), which were identified by DAPI (Fig. S1) and Ki-67 staining (Fig. 1B), a protein enriched on their periphery **(Booth and Earnshaw 2017)**. We therefore aimed at identifying alternative cross-linking agents that would preserve the chromosomal enrichment of Esrrb. Among the different reagents and protocols that we tested, we found two that clearly allow to visualise Esrrb coating of the mitotic chromosomes. First, the homobifunctional crosslinker DSG (Fig. 1B), which has been used to capture hyper-dynamic protein-protein interactions due to its capacity to establish amide bonds via two NHS-ester groups **(Tian et al. 2012)**. Both followed by PFA post-fixation (Fig. 1 and Fig. S1) or not (not shown), the staining patterns were found identical. Second, Glyoxal (Fig. S1), a small bifunctional aldehyde that has been recently rediscovered for its use in fluorescent microscopy **(Richter et al. 2018)**. As a control, we stained ES cells for Nanog (Fig. 1A), a TF that is excluded from the mitotic chromatin **(Festuccia et al. 2016)**. Upon DSG or Glyoxal fixation, we did not observe retention of Nanog on mitotic chromosomes (Fig. 1B and Fig. S1), indicating that these two cross-linkers do not induce aspecific aggregation on the chromosomes. We also tested whether

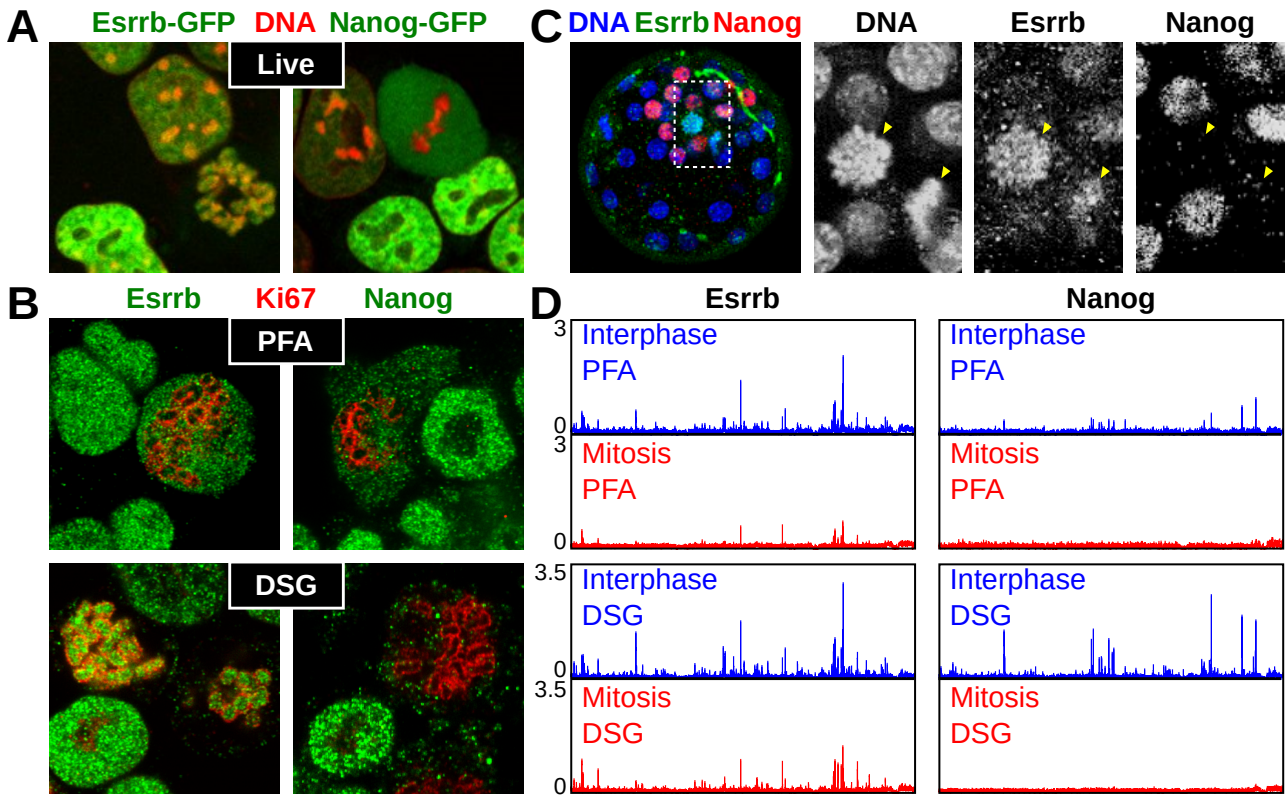


Fig. 1. Capturing global Esrrb binding on mitotic chromosomes. (A) Localisation of Esrrb-GFP (left) or Nanog-GFP (right) fusion proteins in live cells cultured with Hoechst 33342 (red). **(B)** Esrrb (left) and Nanog (right) immunofluorescence (green) after fixation with either PFA (top) or DSG+PFA (bottom; annotated as DSG). The chromosome periphery of mitotic chromosomes is identified by Ki67 (red). **(C)** Immunostaining for Nanog (red) or Esrrb (green) performed on a mouse blastocyst. Counterstain with Hoechst 33342 is shown in blue. Close-up on two mitotic cells (arrowheads) is shown in the right panels (dashed area delimits the selected region). **(D)** Representative binding profiles of Esrrb and Nanog in reads per million (RPM) in interphase (blue) or mitosis (red), obtained after fixation with either PFA (top) or DSG+PFA (bottom; annotated as DSG). The region shown corresponds to chr17:25954686-27500000 (1.5 Mb)

DSG would allow us to visualise the global chromosomal retention of Esrrb in vivo. We have shown before that upon microinjection of Esrrb-Tdtomato mRNA into mouse embryos, the produced fluorescent fusion proteins decorate the mitotic chromatin **(Festuccia et al. 2016)**. Accordingly, when we fixed mouse blastocysts with DSG we could observe mitotic figures with a clear coating of the chromosomes by Esrrb but not by Nanog (Fig. 1C). We conclude, therefore, that global coating of the chromosomes can be captured using alternative fixatives to PFA, in particular bifunctional molecules that are expected to increase the speed and efficiency of cross-linking of protein-protein interactions within complexes.

DSG fixation does not alter the profile of Esrrb binding in mitotic cells. Our finding that DSG and Glyoxal maintain the global association of Esrrb with the mitotic chromosomes opens the possibility to test whether this binding results from the sum of site-specific interactions or from other mechanisms. Indeed, if the global staining reflected site-specific interactions exclusively, one should expect to identify a much larger number of Esrrb binding sites by Chromatin Immuno-Precipitation (ChIP) after fixation with DSG or Glyoxal than with PFA. Yet, despite our efforts, we could not perform ChIP with these two reagents (data not shown). Since DSG or Glyoxal alone are sufficient in their own to globally retain Esrrb on mitotic chromosomes, this indicates that the underlying interactions are likely to be DNA-independent. After a double crosslinking with DSG followed by PFA (DSG+PFA), which is frequently used in biochemical approaches **(Tian et al. 2012)**, ChIP was instead particularly efficient (Figs. 1, 2, 4). Therefore, we performed ChIP-seq in asynchronous (thereafter interphase) and mitotic preparations of ES cells (>95% purity); after splitting the populations in two, we proceeded in parallel with either PFA or DSG+PFA crosslinking. We observed very similar profiles of Esrrb binding both in interphase and in mitosis, irrespective of whether the cells had been crosslinked with PFA or with DSG+PFA (Fig. 1D and Fig. 4). Therefore, whereas Esrrb is globally cross-linked outside or within the mitotic chromosomes by PFA and DSG respectively, the mitotic ChIP signal does not vary dramatically. We note however that DSG+PFA provides higher signal and better signal to background ratio, both in interphase and in mitosis. In agreement with immunostaining and live imaging, Nanog binding is globally lost, both in PFA and in DSG+PFA (Fig. 1D). From this analysis, we conclude that the global coating and the interaction of Esrrb with specific sites are two distinct phenomena. While mitotic Esrrb bookmarking (i.e. binding to specific sites) can be revealed by PFA and by DSG+PFA, global coating is only visible with DSG (or Glyoxal). If the former is a result of the DNA binding activity of Esrrb **(Festuccia et al. 2016)**, the molecular interactions underpinning the latter remain enigmatic. We can exclude, however, that DNA-independent Esrrb binding occurs exclusively at the periphery of the chromosomes, a

proteinaceous compartment that includes TFs **(Booth and Earnshaw 2017)**. Indeed, Esrrb is detected covering the entire area delimited by Ki-67 (Fig. 1B and Fig. S1). This indicates the global Esrrb coating of the chromatids detected by microscopy reflects aspecific interactions of this TF with the chromatin.

DSG versus PFA comparisons reveal different behaviours of other proposed BFs. In addition to Esrrb, other pluripotency TFs have been proposed to act as BFs in ES cells **(Deluz et al. 2016; Teves et al. 2016; Liu et al. 2017)**, although evidence is contradictory. Sox2 has been consistently shown to globally associate with the chromosomes in three independent studies **(Deluz et al. 2016; Teves et al. 2016; Liu et al. 2017)**. In contrast, while one study **(Deluz et al. 2016)** reported that Sox2 binds with poor efficiency to a few dozen regions in mitosis (compared to thousands of sites in interphase), another study claimed that Sox2 and Oct4 remain bound to virtually all their interphase targets **(Liu et al. 2017)**. In addition, Sox2 and Oct4 were shown to be phosphorylated by Aurora kinases, which inhibits DNA binding in mitotic cells **(Qi et al. 2016; Shin et al. 2016)**. In these studies, ChIP was performed after PFA fixation, which leads to an apparent depletion from mitotic chromosomes of both Sox2 and Oct4 (Fig. 2A, C). In contrast, we found that Sox2 displays bright signal all over the chromosomal arms, within the Ki-67 delimited region, by immunostaining after DSG and Glyoxal fixation (Fig. 2A and Fig. S1). We thus extended our ChIP-seq analysis based on DSG+PFA fixation to Sox2. Whereas DSG+PFA dramatically increases ChIP efficiency of Sox2 compared to PFA, the profiles in mitosis are very similar for both fixatives, with little evidence for mitotic bookmarking activity (Fig. 2C and Fig. 4). Therefore, while displaying a macroscopic behaviour similar to Esrrb, mitotic Sox2 does not appear to be an efficient BF. Next, we analysed Oct4 binding. By immunofluorescence, we observed a nearly complete depletion from the chromosomal arms in mitosis, both after DSG and Glyoxal (Fig. 2B, C and Fig. S1). In agreement, ChIP-seq analysis also showed almost complete loss of Oct4 binding at its interphase targets (Fig. 2D). Our results are in agreement with a number of other studies **(Deluz et al. 2016; Shin et al. 2016; Kim et al. 2018)**. However, even using DSG+PFA we could not reproduce recent results showing mitotic bookmarking by Sox2 and Oct4 **(Liu et al. 2017)**. The use of inhibitors of MEK/GSK3b in the conflicting publication, which leads to a reinforcement of the pluripotency network's activity, cannot account for these differences (Fig. S2). We conclude from our data that neither Sox2 nor Oct4 can be considered as potent BFs. Residual signal can be detected at some regions (Fig. 2B, Fig. 4 and Fig. S2), but not to levels comparable to Esrrb. In line with previous results, this further argues for the existence of two components driving the interaction of TFs with mitotic chromosomes: DNA-independent enrichment on the chromatin is detected for many, but not all, TFs; site-specific interactions with regulatory elements are a property of selected bookmarking factors like Esrrb.

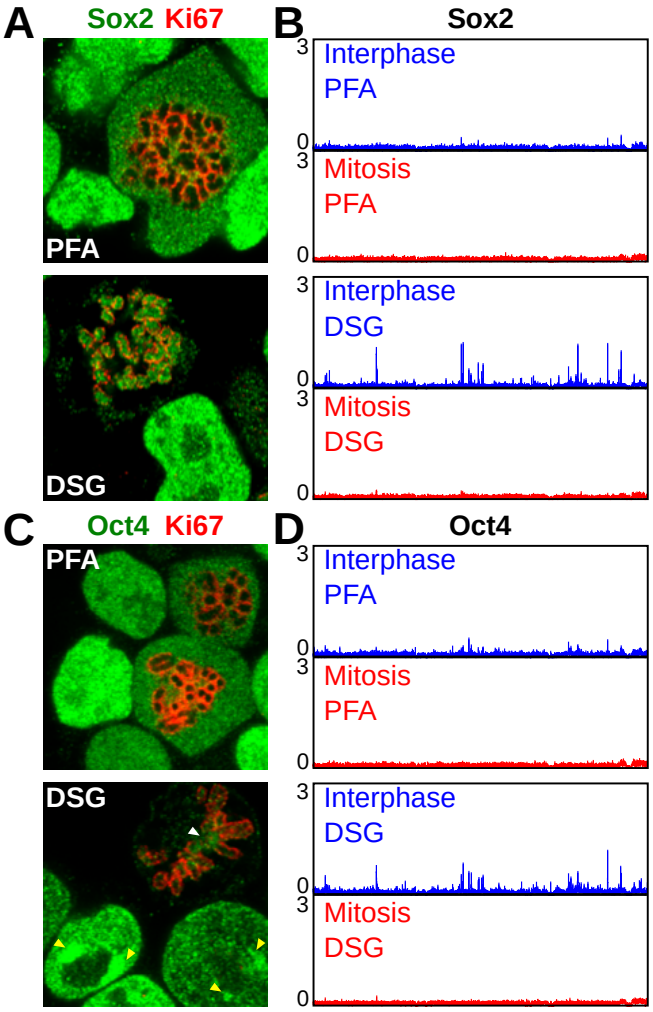


Fig. 2. Sox2 and Oct4 do not bind at regulatory regions in mitosis. (A) Sox2 immunofluorescence (green), after fixation with either PFA (top) or DSG+PFA (bottom; annotated as DSG). The mitotic chromosome periphery is identified by Ki67 (red). **(B)** Representative binding profiles of Sox2 presented as in Fig. 1D. **(C, D)** Results of the same analyses described in (A) and (B) are shown for Oct4. Arrowheads indicate peri-centric heterochromatin foci (PCH) in interphase (yellow) and centromeres in mitosis (white).

DSG enables capturing transient interactions at different chromatin compartments. Careful examination of the Oct4 stainings after DSG and Glyoxal, but not PFA fixation, unmasked a previously unnoticed accumulation of this TF at DAPI-rich regions, the chromocenters **(Saksouk et al. 2015)**, where several centromeres cluster together to form the peri-centric heterochromatin (PCH; yellow arrowheads in Fig. 2C and in Fig. S1). Moreover, in mitotic cells we could also observe focal enrichment of Oct4 at centromeric regions (white arrowheads in Fig. 2C and in Fig. S1). This characteristic pattern of co-localisation with the PCH was further validated by live imaging using ectopic Oct4-GFP and endogenously expressed Oct4-RFP fusion proteins (Fig. 3A and Fig. S3A). Same results were obtained in cells cultured in regular conditions (Fig. 3A) or with inhibitors of MEK/GSK3b (Fig. S3A). In the latter conditions, PCH shifts from H3K9me3 to H3K27me3 **(Tosolini et al. 2018)**, indicating that the PCH association of Oct4 is independent of the presence of specific heterochromatic marks. Remarkably, Oct4 staining is

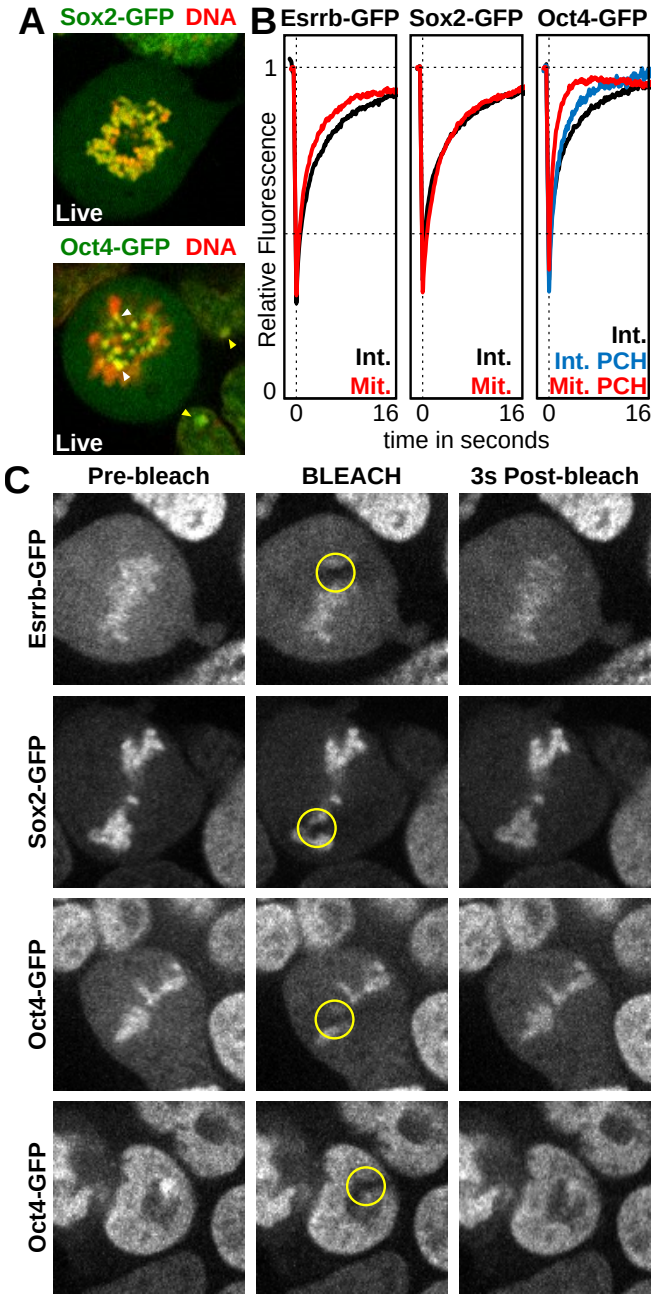


Fig. 3. The interactions captured by DSG are dynamic. **(A)** Localisation of Sox2-GFP (top) or Oct4-GFP (bottom) fusion proteins (green) in live cells cultured with Hoechst 33342 (red). Arrowheads indicate peri-centric heterochromatin foci (PCH) in interphase (yellow) and centromeres in mitosis (white). **(B)** Quantifications of FRAP experiments in interphase (black) and mitosis (red) performed in cells expressing Esrrb-GFP or Sox2-GFP. For cells expressing Oct-4-GFP, recovery of fluorescence at (blue), or outside of (black), peri-centric heterochromatin foci (PCH) is displayed for interphase. Recovery at PCH is displayed for mitosis (red). The Y axis shows the mean percentage of fluorescence relative to pre-bleach levels detected in multiple independent experiments; the X axis shows the time after bleaching. **(C)** Representative examples of Esrrb-GFP, Sox2-GFP and Oct4-GFP signal on mitotic chromosomes before and after bleaching, at the indicated time (seconds). For Oct4-GFP the recovery of signal at PCH is also shown for cells in interphase (bottom).

similar to that of Aurora kinase b (Fig. S3B), which has been shown to phosphorylate Oct4 in mitotic cells to inhibit DNA binding. In agreement, in the presence of the Aurora kinase inhibitor Hesperadin a slight increase of Oct4 coating throughout the chromosomal arms could be observed (Fig. S3C). Hence, using alternative fixatives to PFA not only enables the visualisation of the genuine mitotic localisation of TFs, but may also reveal additional activities in interphase. We then asked whether the interactions of Sox2 and Oct4 unmasked by DSG and Glyoxal are indeed dynamic, as generally reported **(Teves et al. 2016)**. We observed highly dynamic interactions, both in interphase and in mitosis, for all three factors fused to GFP and analysed in parallel experiments (Fig. 3B, 3C, and Fig. S4). Esrrb and Oct4 displayed faster fluorescent recovery after photobleaching (FRAP) in mitosis (Fig. 3B, C). This is particularly true for the interaction of Oct4 with PCH, which are already very dynamic in interphase (Fig. 3C). In reciprocal experiments, we assessed fluorescence loss in photobleaching (FLIP; Fig. S4). We could not identify any significant remnant signal on mitotic chromatids after one minute of continuous bleaching of the freely diffusing TF molecules. Hence, DSG (and Glyoxal) are capable of capturing the highly dynamic interactions established by Esrrb/Sox2 on the chromosomal arms, and by Oct4 in PCH, both in interphase and in mitosis. Altogether, we conclude that while Esrrb exhibits robust mitotic bookmarking activity, other factors are largely evicted during mitosis, irrespectively of their DNA-independent localisation to the chromatin.

Esrrb is the only prominent BF among Esrrb, Sox2, Oct4 and Nanog. Using the collection of datasets generated for Esrrb, Sox2, Oct4 and Nanog in interphase and in mitosis, we sought to comprehensively identify regions subject to mitotic bookmarking. To this end, we first identified the binding regions of individual TFs (Table S2) and confirmed that only Esrrb displays clear and frequent binding in mitosis (Fig. 4A); for Oct4, Sox2 and Nanog, only the regions displaying very high levels of binding in interphase show residual ChIP signal in mitosis, especially in DSG+PFA, where the number of detected peaks is increased. Peaks that were called only in DSG+PFA, and neither in our PFA samples nor in other publicly available datasets **(Chen et al. 2008; Marson et al. 2008; Aksoy et al. 2013; Whyte et al. 2013)**, tend to be smaller (Fig. S5). Nevertheless, their signal is clearly above background in all the analysed datasets of interphase cells fixed with PFA (Fig. S5). Hence, DSG helps capturing regions displaying low levels of binding and increases the overall efficiency of the ChIP. Nonetheless, it does not specifically unmask binding in mitosis. We then used a statistical differential occupancy approach to define regions as bookmarked or lost (see Methods for details and Table S2). We found 10144 regions bookmarked by Esrrb, representing 29.9% of its interphase sites. All other factors displayed a drastic contraction in binding in mitosis: 574 regions for Sox2 (2% of interphase targets); 102 regions for Oct4 (0.6%); 18 regions for Nanog (0.07%). Strong Esrrb binding motifs were identified at the vast majority of Esrrb bookmarked regions (73.4%, score > 12), but only at a smaller subset of the regions losing binding in mitosis (34.9%, score > 12). In contrast, regions losing Esrrb binding displayed an increased occurrence of Oct4/Sox2 composite motifs (Fisher p<7e-45, Oct4/Sox2 Motif score > 12). Remarkably, we observed

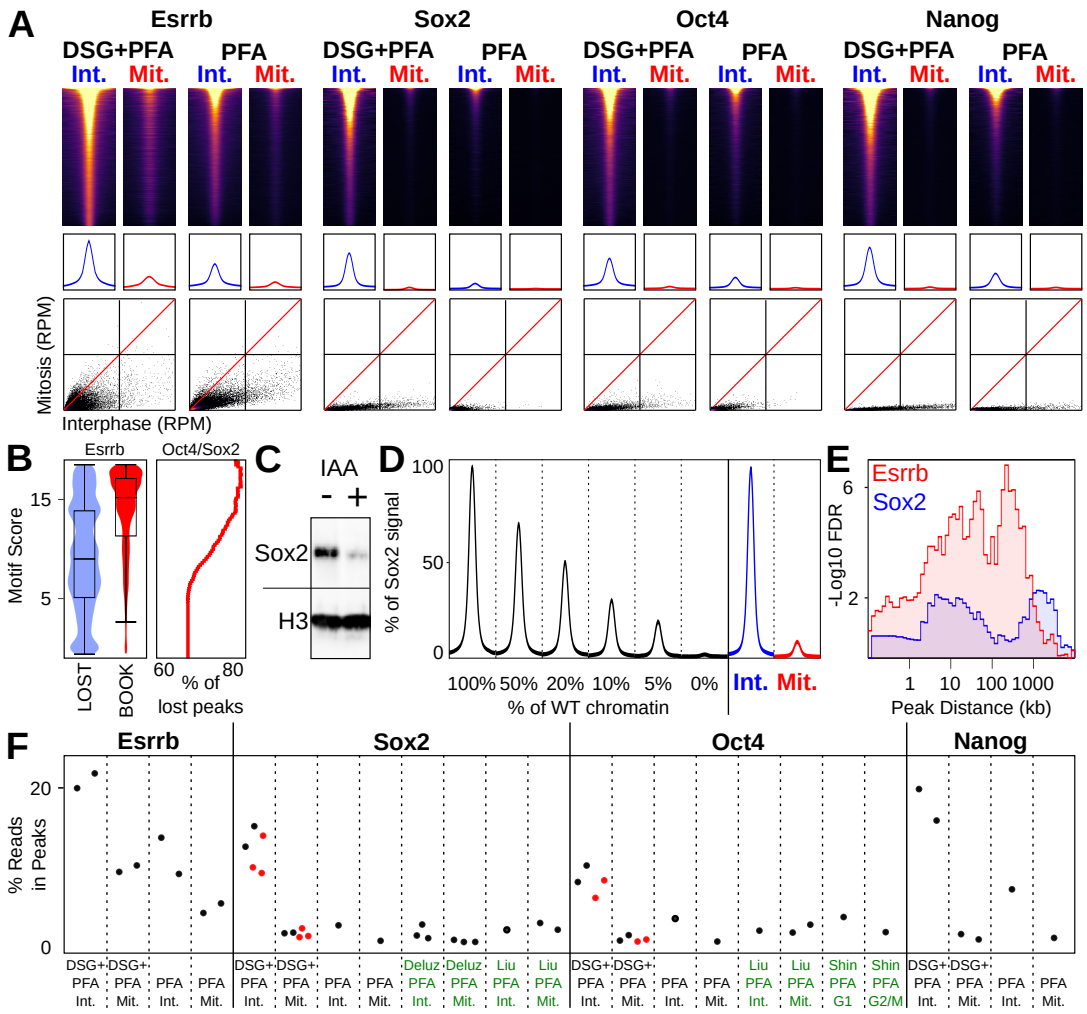


Fig. 4. Comparative analysis of Esrrb, Sox2, Oct4 and Nanog binding in interphase and in mitosis. (A) Top: heatmaps of ChIP-seq signal at Esrrb, Sox2, Oct4 and Nanog binding regions identified in Interphase (Int.) and mitosis (Mit.) for DSG+PFA and PFA alone (see Methods for details). Middle: average binding profile of the regions shown in the heatmaps. Heatmaps and average binding profiles are scaled to the mean occupancy of each factor measured in interphase after DSG+PFA fixation. Bottom: scatter plots of ChIP signal (RPM) at above regions for interphase and mitosis (DSG+PFA scale 0-40 RPM; PFA scale 0-20 RPM). **(B)** Violin plots (left) depicting the FIMO-called best motif score per Esrrb peak in sites losing binding in mitosis (LOST) or retaining binding (BOOK). Right: percentage of Lost peaks with a composite Oct4/Sox2 motif of at least a given quality score. **(C)** Levels of Sox2-AID fusion protein in cells cultured in the absence (-) or presence (+) of the Auxin analogue IAA for 2 hours; H3 is shown as control. **(D)** Percent of the Sox2 ChIP signal detected at binding regions after spiking increasing amounts of WT chromatin into chromatin prepared from Sox2-depleted cells shown alongside the average Sox2 binding profile at potentially bookmarked regions in WT cells in interphase and mitosis. **(E)** Enrichment of genes responsive to Esrrb (red) and Sox2 (blue) in early G1 with proximity to Esrrb or Sox2 bookmarked regions, displayed as -log10 Fisher FDR for genes within given distance of a bookmarked peak. **(F)** Percentage of ChIP-seq reads in identified binding sites for Esrrb, Sox2, Oct4 and Nanog, in both interphase (Int.) and mitosis (Mit.) and DSG+PFA or PFA fixation in our data (black labels) and public datasets (green labels). The red dots correspond to the samples that were added to our study to further corroborate our results.

a scaling relationship: regions containing high quality Oct4/Sox2 motifs, exhibit a higher tendency to lose Esrrb binding in mitosis (Fig. 4B). Altogether, this indicates that at bookmarked regions, Esrrb occupancy is primarily driven by specific interactions with the cognate binding sequence for this TF, as we have previously shown **(Festuccia et al. 2016)**. At lost regions, Esrrb is instead likely recruited indirectly by other TFs that are not capable of binding in mitosis. Hence, the substantial reduction of Esrrb binding sites observed in mitosis represents a striking example of the effect of loss of cooperative TF binding. Previously, we used titration experiments to investigate whether the binding levels seen for Esrrb in mitosis could be explained by contamination from interphase **(Festuccia et al.2016)**; all our mitotic preparation have less than 5% of remnant interphase cells, and typically between 2 and 4%. We repeated this analysis for Sox2, given the relatively high number of low mitotic peaks that we detected in comparison to Oct4 and Nanog. To generate Sox2-depleted chromatin, we generated an ES cell line with (i) both endogenous Sox2 alleles tagged with an auxin-inducible degradation domain (Sox2-AID), and (ii) a constitutive transgene expressing the Tir1 protein inserted at the TIGRE locus **(Madisen et al. 2015)**. Upon treatment with the auxin analogue IAA for 2h, a significant reduction of Sox2 protein levels was observed (Fig. 4C). To further deplete Sox2, cells were differentiated in the presence of retinoic acid (RA) and IAA for 4 days. Gradually increasing amounts of WT chromatin were then spiked into chromatin prepared from IAA-RA treated cells, and ChIP-seq analysis performed. We found that as little of 5% of WT chromatin was sufficient to detect

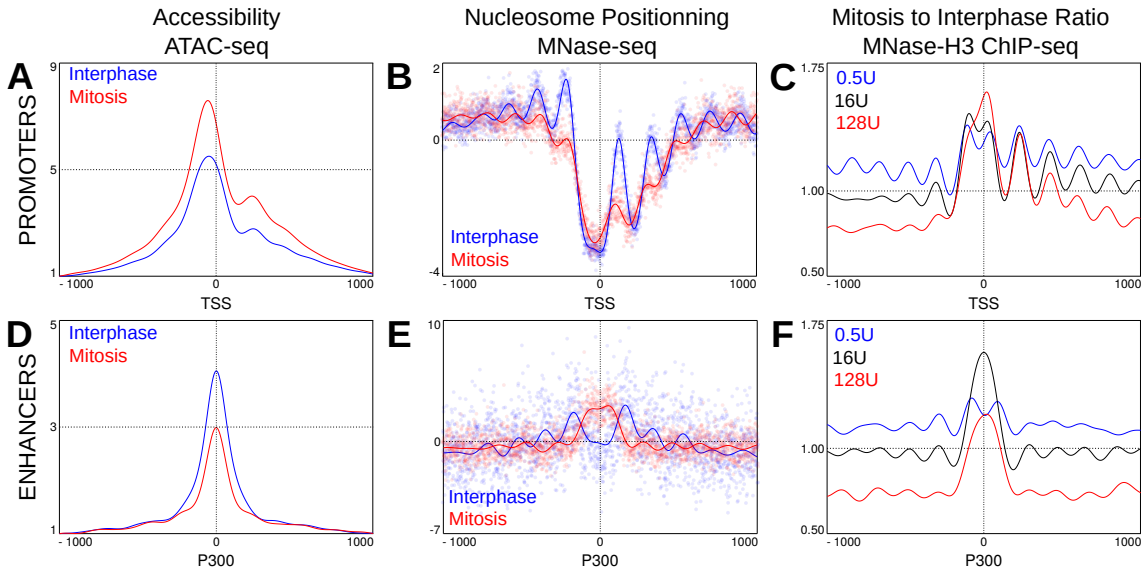


Fig. 5. Drastic reconfigurations of promoters and enhancers in mitotic cells. (A) Accessibility profiles measured by ATAC-seq in the region surrounding the TSS of active genes in interphase (blue) or mitosis (red). Signal is number of Tn5 cut sites for 0-100 bp fragments, normalised to minimum accessibility in +/- 1000 bp window. **(B)** Nucleosome positioning at the same set of promoters, established by MNase-seq. In this panel, the z-score of the number of midpoints of nucleosome-sized fragments (140-200bp) per base, after digestion with 16U of enzyme, are plotted. The lines represent a Gaussian process modelling nucleosome positioning (see Methods) in interphase (blue) and in mitosis (red). **(C)** Mitosis over interphase ratio of MNase H3 ChIP-seq signal for nucleosomal fragments (as assessed by Gaussian process regression; see Methods). Ratios shown for MNase digestions with 0.5U (blue), 16U (black) and 128U (red) of enzyme. **(D, E, F)** as (A, B, C) but for regions centred on summits of interphase P300 ChIP-seq peaks excluding promoters. Note that in (E) MNase-seq signal is from 128U digestions.

clear Sox2 peaks of reduced enrichment (Fig. 4D). Strikingly, the amount of signal observed by adding 5% of contaminant chromatin was higher, on average, to that seen in mitosis at the regions potentially bookmarked by Sox2 (Fig. 4E). Therefore, it is possible that a significant fraction of the regions seemingly bound by Sox2 in mitosis, as well as the absolute levels of enrichment in mitosis, results from the small percentage of contaminant interphase cells in our preparations. To further corroborate that Sox2 is not an efficient bookmarking factor we turned to a functional assay. Confirming our previously result, the set of Esrrb bookmarked regions identified here (Table S2) tend to be enriched in the vicinity of genes that are controlled by this TF in early G1 (Fig. 4E) **(Festuccia et al. 2016)**. We then introduced a GFP-Ccna cell-cycle reporter **(Festuccia et al. 2016)** into Sox2-AID cells, treated them with IAA for 2h and sorted early G1 cells to perform RNA-seq analyses. In comparison with Esrrb, we found a rather minor statistical association between the genes controlled by Sox2 in early G1 (Table S3) and the regions potentially bookmarked by Sox2 (Fig. 4E). We conclude that, whilst we cannot fully rule out that Sox2 may display minimal bookmarking activity, only Esrrb represents a potent and functionally relevant BF among the tested pluripotency factors. This conclusion is particularly well illustrated when the ChIP signal measured at each region is plotted in interphase versus mitosis (Fig. 4A; bottom panels), or when the proportion of reads on peaks are calculated for each TF (Fig. 4F). Why Sox2 and Oct4 have been previously found mitotically bound at most of their interphase targets **(Liu et al. 2017)** remains therefore unclear. This is particularly striking, taking into consideration that our DSG+PFA datasets clearly display improved ChIP efficiency compared to several other published profiles (Fig. 4F; Deluz et al. 2016; Shin et al. 2016; Liu et al. 2017). Despite our efforts, and the addition of 3 and 2 additional independent replicates for Sox2 and Oct4, respectively (red dots in Fig. 4F), we did not find strong evidence for Sox2 and Oct4 bookmarking.

Drastic changes in nucleosome organisation characterise regulatory elements in mitosis. Recently, mitotic chromatin has been shown to maintain surprisingly high levels of chromatin accessibility at virtually all regulatory elements that are active in interphase, in particular at promoters **(Hsiung et al. 2015; Teves et al. 2016)**. Accordingly, we observed that promoter accessibility in mitotic ES cells even surpasses the level observed in interphase, as evaluated by ATAC-seq (Fig. 5A). However, distinct nucleosome organisations might characterise accessible chromatin in these two phases of the cell cycle **(Kelly et al. 2010; Rhee et al. 2014; Mieczkowski et al. 2016; Voong et al. 2016; Mueller et al. 2017)**. To address this, we inferred nucleosome positioning and stability in interphase and in mitosis from a series of experiments based on MNase-seq and H3 ChIP-seq using chromatin digested with titrated MNase activity. In mitosis, we observed preserved nucleosome depleted regions (NDR) around the transcription start sites of promoters (TSS; Fig. 5B). Yet, the phasing of nucleosomes at both sides of the NDRs was drastically attenuated in mitotic cells, probably reflecting reduced transcriptional activity (Fig. 5B). Moreover, when we compared average H3 ChIP-seq signal between mitosis and interphase at different levels of MNase digestion (Fig. 5C), a clear asymmetry was revealed: upstream of the TSS, the sensitivity of the nucleosomes to MNase increased in mitotic cells (as shown by reduced signal with strong digestion); downstream, the +1 nucleosome displayed a similar stability than in interphase, while the following nucleosomes

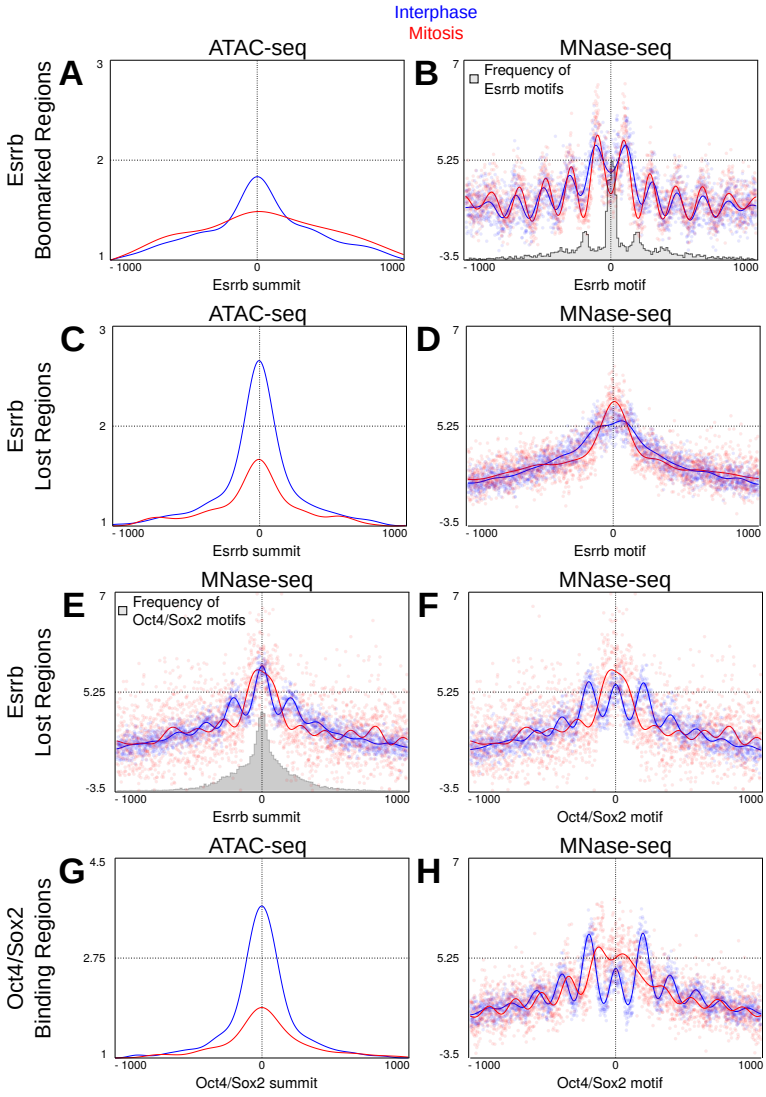


Fig. 6. Binding of Esrrb at its cognate motif drives nucleosome organisation in interphase and in mitosis. Accessibility **(A, C, G)** determined by ATAC-seq as in Fig. 5A, and nucleosome positioning **(B, D, E, F, H)** established as in Fig. 5B, at the regions indicated on the left and centred as shown on their corresponding X-axis, in interphase (blue) and mitosis (red). Histograms embedded in (B) and (E), depict rate of occurrence of the indicated binding motifs: (B) additional Esrrb motifs with FIMO score > 8; (E) top scoring Oct4/Sox2 composite motif.

acquired in mitosis increasing levels of fragility. At the minimal promoter region (TSS and 150bp upstream), we did not find evidence of a nucleosome displaying high occupancy either in interphase or in mitosis (Fig. 5B). Nonetheless, the H3 signal detected over the minimal promoter tend to increase in mitosis, irrespectively of the MNase conditions (Fig. 5C). These results indicate that, globally, the nucleosomes at promoters are more fragile **(Mieczkowski et al. 2016; Voong et al. 2016)** in mitosis, except at the minimal promoter region where they stabilise without increasing their overall occupancy. Moreover, the differential behaviour within and outside the transcription unit may potentially reflect the reduced transcriptional activity that has been recently detected in mitotic cells **(Palozola et al. 2017)**. Therefore, promoters are subject to drastic nucleosome reorganisation in mitotic cells. We then analysed enhancers (identified here as p300-bound elements, excluding TSSs and gene bodies). As previously shown **(Hsiung et al. 2015)**, we found enhancers to partially lose accessibility in mitosis (Fig. 5D). More strikingly, these elements display a profound reconfiguration in nucleosomal architecture (Fig. 5E): nucleosomes resistant to our most aggressive digestion conditions can be detected at the site of p300 recruitment exclusively in mitosis, and the phasing of the surrounding nucleosomes is altered (Fig. 5E). Moreover, titration of MNase activity followed by H3 ChIP-seq, revealed that both upstream and downstream of the stabilised nucleosome, increased fragility can be measured in mitotic cells (Fig. 5F). Therefore, even though promoters and enhancers maintain significant levels of accessibility in mitotic cells, the arrangement of their nucleosomes changes substantially.

Chromatin accessibility and nucleosome organisation as a function of Esrrb bookmarking. We then focused on the analysis of the regions bound by Esrrb (Fig. 6 and Fig. S6). While Esrrb-bookmarked regions partially lose accessibility (Fig. 6A), this reduction is significantly more pronounced at the regions where Esrrb binding is lost in mitosis

(Fig. 6C). Hence, there is a clear correlation between the ability of Esrrb to bind to certain targets in mitotic cells, and the partial maintenance of accessibility. Moreover, at bookmarked regions, we observed highly positioned nucleosomes both in interphase and mitosis: the Esrrb motif lies within a major NDR and phased nucleosomes spread both upstream and downstream the binding site (Fig. 6B). This pattern contrasts markedly with that seen at p300 enhancers (Fig. 5E), clearly establishing a strong correlation between Esrrb mitotic binding and the retention of well-structured nucleosome arrays. Moreover, in mitosis we observed a slight shrinking of the nucleosomal array converging towards the central Esrrb motif, leading to a modest change of position of the nucleosomes. Remarkably, when we calculated a frequency map of additional Esrrb motifs within these regions (grey histogram in Fig. 6B), we observed a small but clear enrichment precisely at the mitosis-specific inter-nucleosomal space between the -2/-1 and +1/+2 nucleosomes. This strongly indicates that in mitosis, the DNA binding activity of Esrrb becomes dominant in establishing nucleosome positioning. In contrast, at regions losing Esrrb binding in mitosis, the nucleosomal profiles were not found to be dramatically different in interphase and mitosis: in both cell-cycle phases the Esrrb motif is occupied by a nucleosome, which is more sharply positioned during division (Fig. 6D). However, these regions appeared barely organised compared to their bookmarked counterparts and lacked clear phasing at both sides of the Esrrb motif. Since high quality Esrrb motifs are not particularly prevalent at these regions (Fig. 4B), we reanalysed the data by re-centring on Esrrb summits. We noted that Oct4/Sox2 motifs are enriched in the vicinity of Esrrb summits (grey histogram in Fig. 6E), and therefore also re-centred these regions on these motifs (Fig. 6F). Both analyses unveiled a clear nucleosomal organisation in interphase that is highly modified in mitotic cells (Fig. 6E, F). This indicates that Esrrb may be recruited indirectly and play a minor role in establishing nucleosome positioning over these regions. In accord, the nucleosome pattern at regions centred on Esrrb summits was also highly similar to that seen at the bulk of Oct4/Sox2 binding sites (Fig. 6H). These regions show a consistent reduction in accessibility in mitosis (Fig. 6G) and major nucleosome repositioning, with signs of shifting in the nucleosomal array and invasion at both flanks of the Oct4/Sox2 motifs (Fig. 6H). At all these regions, a concomitant increase in occupancy by fragile nucleosomes could also be observed (Fig. S6A). Of note, local features like the ones we observed at TSSs and p300 summits could not be detected (Fig. 5E, F and Fig. S6A). Finally, at regions exhibiting low mitotic Sox2 ChIP-seq signal, we also observed major reorganisations of nucleosomes in mitosis. Nonetheless, the presence of a very narrow NDR could not be ruled out (Fig. S6B), possibly reflecting minimal bookmarking activity. From these analyses, we conclude that TF binding is likely required to maintain nucleosome positioning at regulatory elements during cell division. Esrrb acts as a major organiser of the chromatin in both phases of the cell cycle (Fig. 6B).

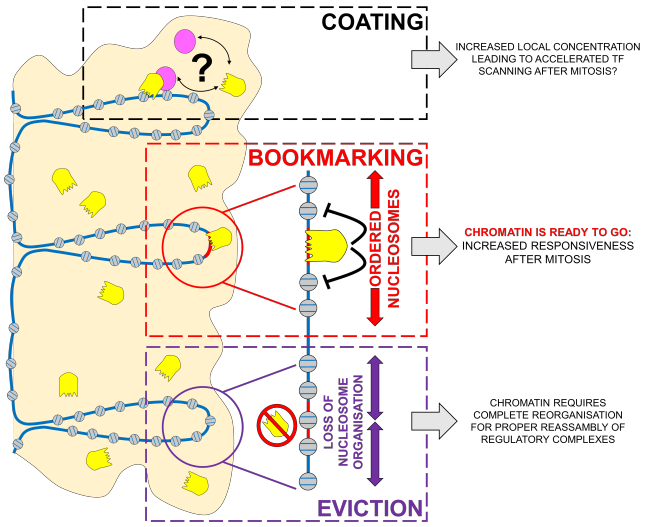


Fig. 7. Model summarising distinct behaviours of TFs in mitotic cells and their relationships to nucleosome organisation and post-mitotic gene regulation. Many TFs show global localisation on the chromosomes in mitosis. This localisation is likely driven by sequence-independent interactions with DNA or other components of the chromatin or the mitotic chromosomes, and might serve a function in increasing the local concentration of TFs in proximity of their targets, in turn facilitating binding in G1. In contrast, during division only few TFs remain dynamically bound to a subset of the sites they occupy in interphase. At bookmarked sites, the continued activity of these TFs maintains an ordered chromatin configuration, possibly limiting the extent of chromatin remodelling required to re-establish functional regulatory architectures in the following cell-cycle. At sites losing TF binding, nucleosome positioning is vastly disorganised, and increased occupancy by nucleosomes is detected at binding motifs. Although these sites do not become fully inaccessible, profound chromatin rearrangements are expected to be needed in early G1 to reinstate proper function.

Discussion

Proposed around 20 years ago **(Michelotti et al. 1997)**, the idea that certain TFs mitotically propagate gene regulatory information had been until recently only sporadically explored. Instead, over the last few years, several publications have revealed a continuously growing number of candidate mitotic bookmarking TFs **(Festuccia et al. 2017)**. Considering that PFA, arguably the most used cross-linker, leads to an artificial depletion of TFs from the mitotic chromosomes, as visualised by microscopy, many more TFs than those currently described are probably able to associate with the chromatin during division **(Pallier et al. 2003; Teves et al. 2016)**. However, whether all these TFs are engaged in site-specific interactions and therefore act as mitotic bookmarking factors remains unclear **(Festuccia et al. 2017)**. Here, we identify cross-linkers that preserve the global mitotic localisation of several TFs, providing a simple experimental method to study the behaviour of new transcriptional regulators during division and, more generally, visualise spatial organisations deriving from transient and fast binding events. Conversely, our results impose caution: we show that localisation of a TF to the chromatin does not necessarily imply sequence specific binding in mitosis (Fig. 7). This is exemplified by Sox2 and, as shown by others, by CTCF **(Oomen et al. 2018)**: while these TFs are both macroscopically retained, they are largely evicted from the sites occupied in interphase. The functional consequences of this distinction are major: we

failed to identify a strong relationship between the proximity of the few regions exhibiting Sox2 binding, albeit at low levels, and the transcriptional effects of Sox2 in early G1. Conversely, the functional relevance of site-specific mitotic binding (Fig. 7) has been documented for several canonical bookmarking factors, including Gata1 **(Kadauke et al. 2012)**, FoxA1 **(Caravaca et al. 2013)** and Esrrb **(Festuccia et al. 2016)**. Therefore, the emerging idea of a widespread mitotic bookmarking activity needs to be carefully considered and evaluated. At the same time, the potential function of a global chromosomal retention cannot be ignored and requires dedicated experimental setups. In this regard, our comparative analysis of fixatives reveals that distinct molecular mechanisms likely contribute to the overall mitotic localisation of TFs (Fig. 7). Esrrb displays highly correlated binding profiles by ChIP when the chromatin is fixed with PFA or with DSG. In contrast, only DSG captures global Esrrb enrichment on the chromatin. Given the ability of DSG to efficiently fix transient interactions, and in light of the results of FRAP and single molecule tracking studies **(Caravaca et al. 2013; Deluz et al. 2016; Teves et al. 2016)**, this reveals that most likely the bulk of the molecules for a given TF bound to the chromatids during mitosis are not engaged in sequence specific interactions with DNA. Somehow surprisingly, however, we showed previously that mutating 3 amino-acids of the Esrrb DNA biding domain that are engaged in base-specific contacts with the binding motif dramatically decreases the global decoration of the mitotic chromosomes **(Festuccia et al. 2016)**. It is possible that these amino acids of the Esrrb zinc-finger domain are also required for Esrrb to scan the DNA in search of its biding sites. Alternatively, these mutations may more generally alter the structure of Esrrb, preventing interactions with other proteins enriched at the mitotic chromosomes. Notably, the bifunctional cross-linkers that we have used, DSG and Glyoxal, are expected to increase the efficiency of fixation within large protein complexes, opening the possibility that the interactions driving the global enrichment of TFs on the chromatids are based on protein-protein rather than protein-DNA contacts. Thus, we propose that the model previously proposed for FoxA1 regarding the existence of at least two distinct phenomena underlying the behaviour of TFs in mitotic cells could be extended, and applied generally to BFs: on the one hand, both DNA scanning and the ability to interact with other proteins of the chromatin sustains the bulk localisation of TFs to the chromatids; on the other, bona-fide bookmarking, understood here as the capacity to mediate site-specific binding, drives functionally relevant accumulation of TFs at regulatory elements **(Festuccia et al. 2017)**. While FoxA1 is capable of binding nucleosomes directly **(Cirillo et al. 1998)**, by virtue of its inherent structural properties **(Clark et al. 1993; Ramakrishnan et al. 1993)**, the mitotic partners for the protein-protein interaction of other TFs decorating mitotic chromosomes may be more diverse (Fig. 7). These protein could be part of the chromatin or restricted to the chromosomal periphery **(Booth and Earnshaw 2017)**. While such restricted localisation can be excluded for Esrrb, Sox2 and Oct4, it may apply to other TFs. Indeed, a multitude of determinant of TF localisation seem to exist. This is the case of Oct4, that we report here as focally enriched within (peri-)centric regions, both in interphase and in mitosis. Extending beyond mitosis, given the complexity revealed by the used of multiple cross-linking agents, this study directly calls for a general reassessment of TF localisation and function as inferred from fixed samples.

Distinguishing TFs as enriched or depleted from mitotic chromosomes, and as binding or not at specific regulatory regions, will eventually allow us to establish a hierarchy of their contributions to the re-establishment of transcription after mitosis (Fig. 7). This will be particularly important in highly proliferative cells undergoing progressive implementation of new cell identities during development **(Festuccia et al. 2017)**. To gain a full understanding of the importance of mitotic bookmarking, it is also crucial to elucidate the molecular mechanisms mediating its function. Different lines of evidence point to the lack of permanent TF binding at mitotic chromosomes. Even in the extreme case of the general TF Tbp, the residence time on the mitotic chromatin is below 2 minutes **(Teves et al. 2018)**. Therefore, occupancy by single molecules of mitotic bookmarking factors do not physically transfer regulatory information from mother to daughter cells; to be functional, BFs may instead induce specific modifications around their mitotic target sites. However, regardless of their mitotic bookmarking status, most if not all active regulatory regions remain at least partially accessible in mitotic cells **(Hsiung et al. 2015; Teves et al. 2016)**. This has been now shown analysing the bookmarking sites of several TFs, including Gata1 **(Kadauke et al. 2012)**, and, here, Esrrb, Sox2, Oct4 and Nanog. Therefore, even if many other BFs remain to be identified, the general loss of TF binding characterising mitosis is unlikely to completely abolish chromatin accessibility. In general, the presence of destabilised nucleosomes at regulatory elements could suffice to maintain these regions less refractory to the binding of transcriptional regulators. Nevertheless, TF binding might still contribute towards maintaining comparatively high accessibility at selected loci. This was originally proposed for the bookmarking factor Foxl1 **(Yan et al. 2006)** and is further supported by our observation that the regions bookmarked by Esrrb display a milder reduction of ATAC signal compared to those where Esrrb is evicted. More strikingly, our nucleosome mapping studies indicate that Esrrb bookmarking plays a major role in preserving the fine patterns of nucleosome organisation, rather than mere accessibility, at regulatory elements (Fig. 7). Indeed, at regions bookmarked by Esrrb, binding motifs are strongly associated with a nucleosome depleted region, and are flanked by well organised and phased nucleosomes. This configuration is detected in interphase, but is significantly clearer in mitosis where even neighbouring inter-nucleosomal spaces correlate with the presence of additional Esrrb motifs. We believe this reflects the loss of counteracting effect from binding of other TFs in mitosis, and the consequent dominance of Esrrb over the organisation of the nucleosomes at these sites. In this light, mitosis might

represent a context of simplified interactions of TF with the chromatin, where few fundamental activities are maintained. In contrast, in the complete absence of mitotic TF binding, nucleosomal arrays are largely reconfigured. This is true at enhancers marked by p300, at regions losing Oct4/Sox2 binding, as well as at CTCF biding sites **(Oomen et al. 2018)**. Remarkably, at regions losing Esrrb in mitosis a clear nucleosomal organisation is only appreciated when regions are aligned relative to the Esrrb peak summit or the binding motifs for other TFs. Hence, at these regions, Esrrb might be recruited indirectly and the nucleosomal organisation of these regions, therefore, is not imposed by Esrrb. Together these observations clearly indicate that mitotic bookmarking by Esrrb is essentially driven by sequence-specific DNA interactions through which this factor imposes specific constraints on nucleosomal organisation. Therefore, the nucleosomal landscape around TF binding sites in mitosis may be used as a proxy for mitotic bookmarking activity, further indicating that neither Sox2 nor Oct4 are efficient bookmarking factors.

The recent observation of widespread chromatin accessibility in mitotic cells suggested that many TFs would act as bookmarking factors. In contrast, our analysis of TF binding, chromatin accessibility, nucleosome positioning and stability in mitotic ES cells, rather indicates that mitotic bookmarking can only be mediated by selected TFs, such as Esrrb in ES cells. Indeed, the stereotypical behaviour of enhancers that we observe here indicates that a robust nucleosome is positioned at p300 recruitment sites, with more fragile nucleosomes occupying the vicinities. These destabilised nucleosomes may explain the apparent accessibility of these regions. At promoters, we also observe a loss of phasing, and a relative stabilisation of the nucleosomes lying just upstream of the TSS as compared to those more distally located, which appear to be more fragile. While the molecular players destabilising these nucleosomes requires further investigation, our data indicate that Esrrb, and potentially other bookmarking factors, may generally act by locally preserving specific nucleosome architectures. These configurations in turn favour the re-establishment of functional regulatory complexes early after mitosis. We propose this mechanism to represent the molecular basis of the transmission of regulatory information by sequence-specific mitotic bookmarking factors (Fig. 7).

Acknowledgements. The authors acknowledge the Imagopole France–BioImaging infrastructure, supported by the French National Research Agency (ANR 10-INSB-04-01, Investments for the Future), for advice and access to the UltraVIEW VOX system. We also acknowledge the Transcriptome and EpiGenome, BioMics, Center for Innovation and Technological Research of the Institut Pasteur for NGS. 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 Agence Nationale de la Recherche (ANR 16 CE12 0004 01 MITMAT), and the Ligue contre le Cancer (LNCC EL2018 NAVARRO). N.F was supported by a Marie Curie IEF fellowship and a Pasteur-Cantarini Fellowship program. N.O. is supported by Revive.

Author contributions. N.F. performed or supervised all the experimental work. N.O. analysed sequencing data. N.F., N.O. and P.N. analysed and interpreted the results and wrote the manuscript. All other authors provided experimental help.

Declaration of interests. The authors declare no competing interests.

References

Aksoy I, Jauch R, Chen J, Dyla M, Divakar U, Bogu GK, Teo R, Leng Ng CK, Herath W, Lili S et al. 2013. Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. **EMBO J** 32: 938-953.

Booth DG, Earnshaw WC. 2017. Ki-67 and the Chromosome Periphery Compartment in Mitosis. **Trends Cell Biol** 27: 906-916.

Caravaca JM, Donahue G, Becker JS, He X, Vinson C, Zaret KS. 2013. Bookmarking by specific and nonspecific binding of FoxA1 pioneer factor to mitotic chromosomes. **Genes Dev** 27: 251-260.

Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, 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.

Cirillo LA, McPherson CE, Bossard P, Stevens K, Cherian S, Shim EY, Clark KL, Burley SK, Zaret KS. 1998. Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. **EMBO J** 17: 244-254.

Clark KL, Halay ED, Lai E, Burley SK. 1993. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. **Nature** 364: 412-420.

de Castro IJ, Gokhan E, Vagnarelli P. 2016. Resetting a functional G1 nucleus after mitosis. **Chromosoma** 125: 607-619.

Deluz C, Friman ET, Strebinger D, Benke A, Raccaud M, Callegari A, Leleu M, Manley S, Suter DM. 2016. A role for mitotic bookmarking of SOX2 in pluripotency and differentiation. **Genes Dev** 30: 2538-2550.

Festuccia N, Dubois A, Vandormael-Pournin S, Gallego Tejeda E, Mouren A, Bessonnard S, Mueller F, Proux C, Cohen-Tannoudji M, Navarro P. 2016. Mitotic binding of Esrrb marks key regulatory regions of the pluripotency network. **Nat Cell Biol** 18: 1139-1148.

Festuccia N, Gonzalez I, Owens N, Navarro P. 2017. Mitotic bookmarking in development and stem cells. **Development** 144: 3633-3645.

Hsiung CC, Morrissey CS, Udugama M, Frank CL, Keller CA, Baek S, Giardine B, Crawford GE, Sung MH, Hardison RC et al. 2015. Genome accessibility is widely preserved and locally modulated during mitosis. **Genome Res** 25: 213-225.

Kadauke S, Udugama MI, Pawlicki JM, Achtman JC, Jain DP, Cheng Y, Hardison RC, Blobel GA. 2012. Tissue-specific mitotic bookmarking by hematopoietic transcription factor GATA1. **Cell** 150: 725-737.

Kelly TK, Miranda TB, Liang G, Berman BP, Lin JC, Tanay A, Jones PA. 2010. H2A.Z maintenance during mitosis reveals nucleosome shifting on mitotically silenced genes. **Mol Cell** 39: 901-911.

Kim HJ, Shin J, Lee S, Kim TW, Jang H, Suh MY, Kim JH, Hwang IY, Hwang DS, Cho EJ et al. 2018. Cyclin-dependent kinase 1 activity coordinates the chromatin associated state of Oct4 during cell cycle in embryonic stem cells. **Nucleic Acids Res** 46: 6544-6560.

Liu Y, Pelham-Webb B, Di Giammartino DC, Li J, Kim D, Kita K, Saiz N, Garg V, Doane A, Giannakakou P et al. 2017. Widespread Mitotic Bookmarking by Histone Marks and Transcription Factors in Pluripotent Stem Cells. **Cell** Rep 19: 1283-1293.

Ma Y, Kanakousaki K, Buttitta L. 2015. How the cell cycle impacts chromatin architecture and influences cell fate. **Front Genet** 6: 19.

Madisen L, Garner AR, Shimaoka D, Chuong AS, Klapoetke NC, Li L, van der Bourg A, Niino Y, Egolf L, Monetti C et al. 2015. Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance. **Neuron** 85: 942-958.

Marson A, Levine SS, Cole MF, Frampton GM, Brambrink T, Johnstone S, Guenther MG, Johnston WK, Wernig M, Newman J et al. 2008. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. **Cell** 134: 521-533.

Michelotti EF, Sanford S, Levens D. 1997. Marking of active genes on mitotic chromosomes. **Nature** 388: 895-899.

Mieczkowski J, Cook A, Bowman SK, Mueller B, Alver BH, Kundu S, Deaton AM, Urban JA, Larschan E, Park PJ et al. 2016. MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. **Nat Commun** 7: 11485.

Mueller B, Mieczkowski J, Kundu S, Wang P, Sadreyev R, Tolstorukov MY, Kingston RE. 2017. Widespread changes in nucleosome accessibility without changes in nucleosome occupancy during a rapid transcriptional induction. **Genes Dev** doi:10.1101/gad.293118.116.

Oomen ME, Hansen A, Liu Y, Darzacq X, Dekker J. 2018. CTCF sites display cell cycle dependent dynamics in factor binding and nucleosome positioning. **bioRxiv** doi:10.1101/365866.

Pallier C, Scaffidi P, Chopineau-Proust S, Agresti A, Nordmann P, Bianchi ME, Marechal V. 2003. Association of chromatin proteins high mobility group box (HMGB) 1 and HMGB2 with mitotic chromosomes. **Mol Biol Cell** 14: 3414-3426.

Palozola KC, Donahue G, Liu H, Grant GR, Becker JS, Cote A, Yu H, Raj A, Zaret KS. 2017. Mitotic transcription and waves of gene reactivation during mitotic exit. **Science** 358: 119-122.

Qi D, Wang Q, Yu M, Lan R, Li S, Lu F. 2016. Mitotic phosphorylation of SOX2 mediated by Aurora kinase A is critical for the stem-cell like cell maintenance in PA-1 cells. **Cell Cycle** 15: 2009-2018.

Ramakrishnan V, Finch JT, Graziano V, Lee PL, Sweet RM. 1993. Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. **Nature** 362: 219-223.

Rhee HS, Bataille AR, Zhang L, Pugh BF. 2014. Subnucleosomal structures and nucleosome asymmetry across a genome. **Cell** 159: 1377-1388.

Richter KN, Revelo NH, Seitz KJ, Helm MS, Sarkar D, Saleeb RS, D'Este E, Eberle J, Wagner E, Vogl C et al. 2018. Glyoxal as an alternative fixative to formaldehyde in immunostaining and super-resolution microscopy. **EMBO J** 37: 139-159.

Rizkallah R, Alexander KE, Hurt MM. 2011. Global mitotic phosphorylation of C2H2 zinc finger protein linker peptides. **Cell Cycle** 10: 3327-3336.

Rizkallah R, Hurt MM. 2009. Regulation of the transcription factor YY1 in mitosis through phosphorylation of its DNA-binding domain. **Mol Biol Cell** 20: 4766-4776.

Saksouk N, Simboeck E, Dejardin J. 2015. Constitutive heterochromatin formation and transcription in mammals. **Epigenetics Chromatin** 8: 3.

Shin J, Kim TW, Kim H, Kim HJ, Suh MY, Lee S, Lee HT, Kwak S, Lee SE, Lee JH et al. 2016. Aurkb/PP1-mediated resetting of Oct4 during the cell cycle determines the identity of embryonic stem cells. **Elife** 5: e10877.

Teves SS, An L, Bhargava-Shah A, Xie L, Darzacq X, Tjian R. 2018. A stable mode of bookmarking by TBP recruits RNA polymerase II to mitotic chromosomes. **Elife** 7.

Teves SS, An L, Hansen AS, Xie L, Darzacq X, Tjian R. 2016. A dynamic mode of mitotic bookmarking by transcription factors. **Elife** 5.

Tian B, Yang J, Brasier AR. 2012. Two-step cross-linking for analysis of protein-chromatin interactions. **Methods Mol Biol** 809: 105-120.

Tosolini M, Brochard V, Adenot P, Chebrout M, Grillo G, Navia V, Beaujean N, Francastel C, Bonnet-Garnier A, Jouneau A. 2018. Contrasting epigenetic states of heterochromatin in the different types of mouse pluripotent stem cells. **Sci Rep** 8: 5776.

Voong LN, Xi L, Sebeson AC, Xiong B, Wang JP, Wang X. 2016. Insights into Nucleosome Organization in Mouse Embryonic Stem Cells through Chemical Mapping. **Cell** 167: 1555-1570 e1515.

Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. **Cell** 153: 307-319.

Yan J, Xu L, Crawford G, Wang Z, Burgess SM. 2006. The forkhead transcription factor FoxI1 remains bound to condensed mitotic chromosomes and stably remodels chromatin structure. **Mol Cell Biol** 26: 155-168.

SUPPLEMENTARY INFORMATION

Six supplementary figures accompany this manuscript:

Fig.S1: Comparative analyses of different fixations.

Fig.S2: Oct4 and Sox2 binding in 2i-treated ES cells.

Fig.S3: Extended analysis of Oct4 localisation.

Fig.S4: Example of FLIP imaging.

Fig.S5: Analysis of peaks specifically detected in DSG+PFA.

Fig.S6: Additional information of nucleosome organisation at Esrrb, Oct4 and Sox2 sites.

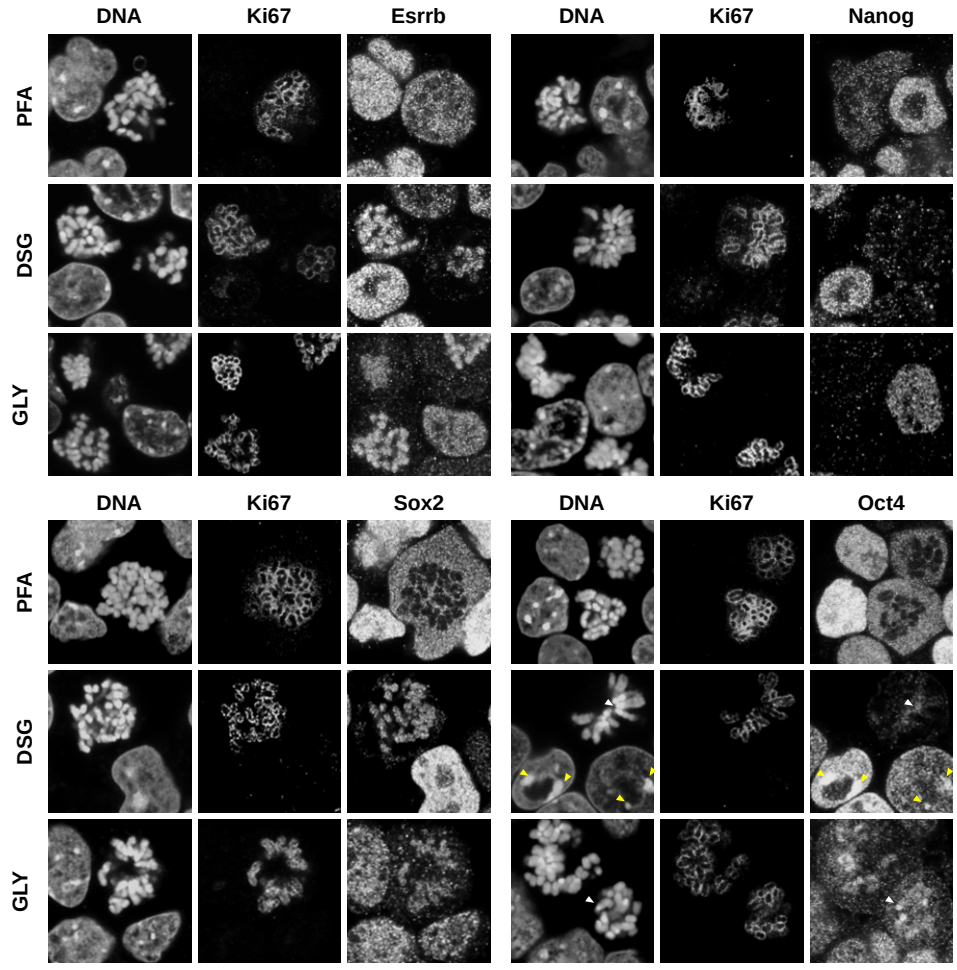
They can be found at the end of this document.

Three Supplementary Tables are available online:

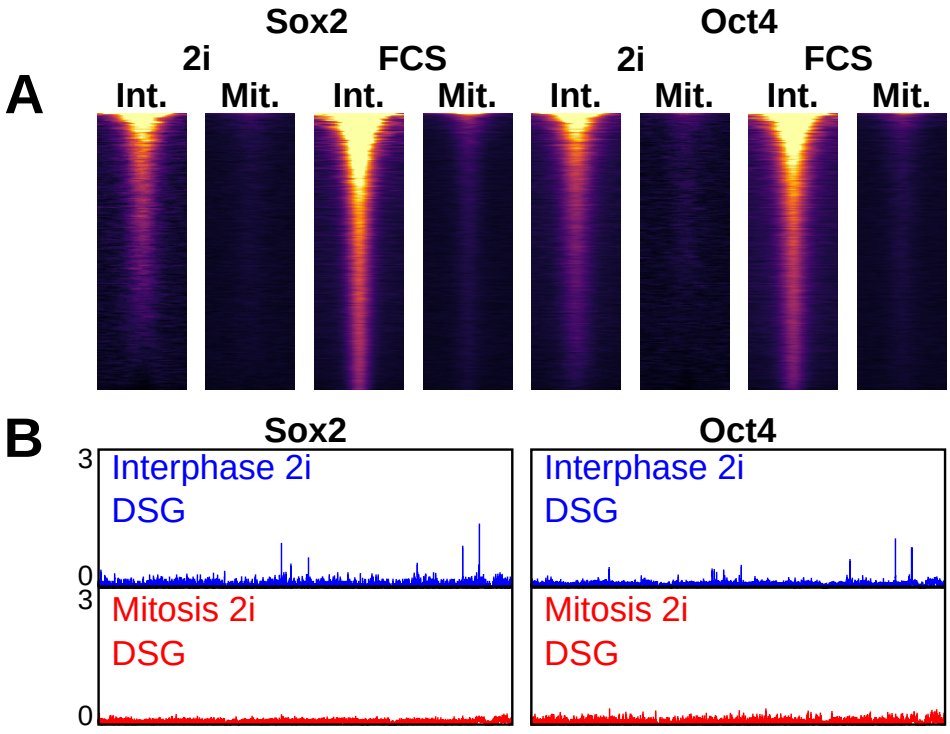
Table S1: Overview of ChIP-seq, MNase-Seq, MNase H3 ChIP-seq, ATAC-seq, RNA-seq libraries sequenced in this study.

Table S2: Peaks and bookmarking calls for Esrrb, Oct4, Sox2, Nanog

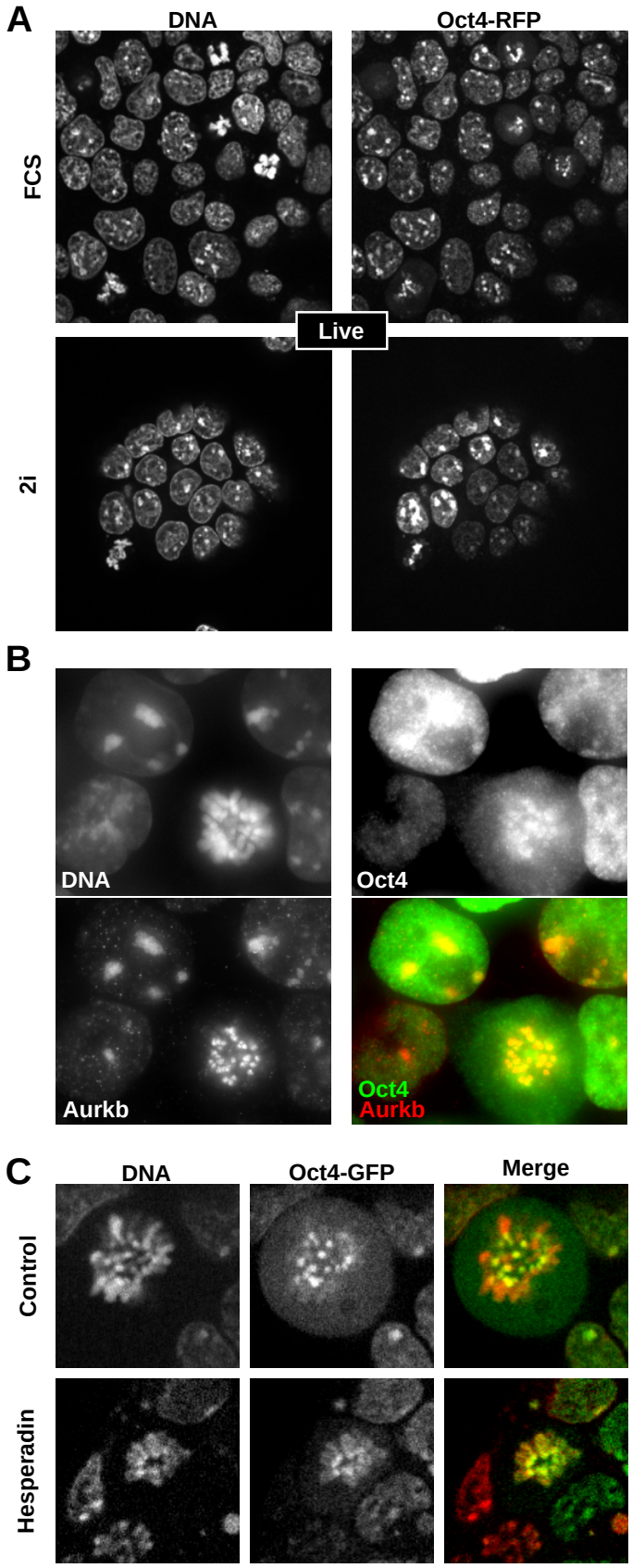
Table S3: Differential expression of genes responsive to Sox2 depletion in EG1.



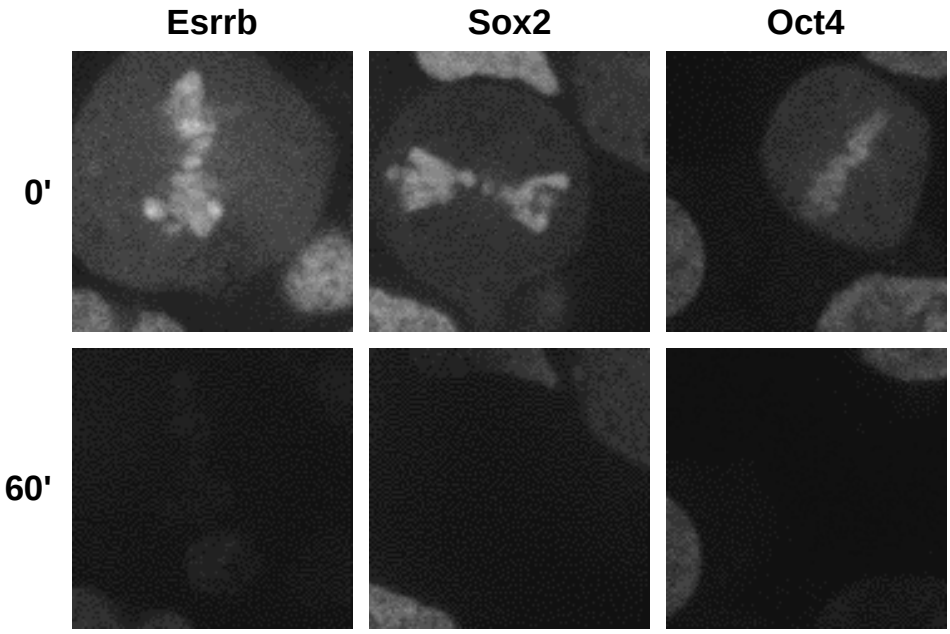
Supplementary Information, Fig. S 1. Comparative analyses of different fixations. Immunofluorescence for Esrrb (top left), Nanog (top right), Sox2 (bottom left) and Oct4 (bottom right), after fixation with either PFA, DSG+PFA (labelled as DSG only), or glyoxal. DNA was counterstained with DAPI. The mitotic chromosome periphery is identified by Ki67 staining. In the Oct4 staining, the arrowheads indicate peri-centric heterochromatin foci (PCH) in interphase (yellow) and centromeres in mitosis (white). Note that Sox2 immunofluorescene required a PFA post-fixation after Glyoxal.



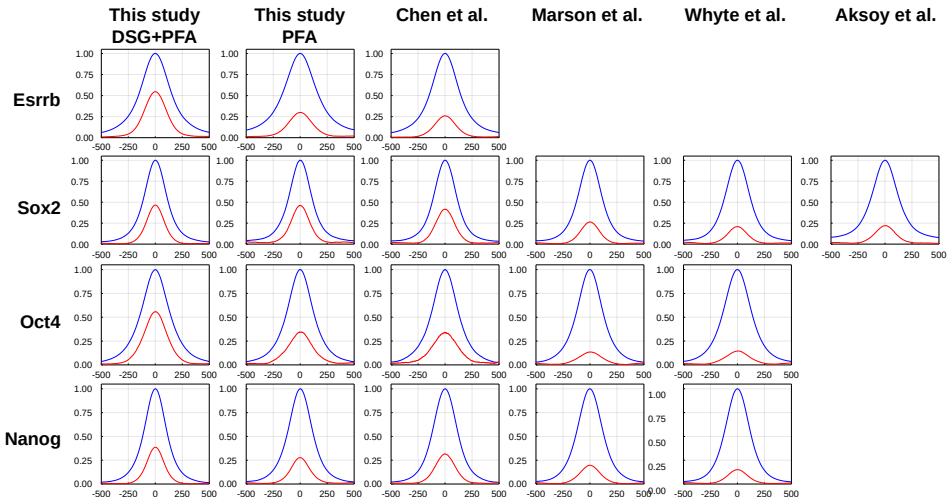
Supplementary Information, Fig. S 2. Oct4 and Sox2 binding in 2i-treated ES cells. (A) Heatmaps of ChIP-seq signal in interphase and mitosis for Sox2 and Oct4, in cells cultured in FCS/LIF with and without 2i. Binding regions are the union of peaks identified in both conditions. **(B)** Representative binding profile for Sox2 (left) or Oct4 (right) in interphase (blue) or mitosis (red), obtained after fixation with DSG+PFA in 2i treated cells; vertical scale RPM. The region corresponds exactly to that shown in Figs. 1 and 2.



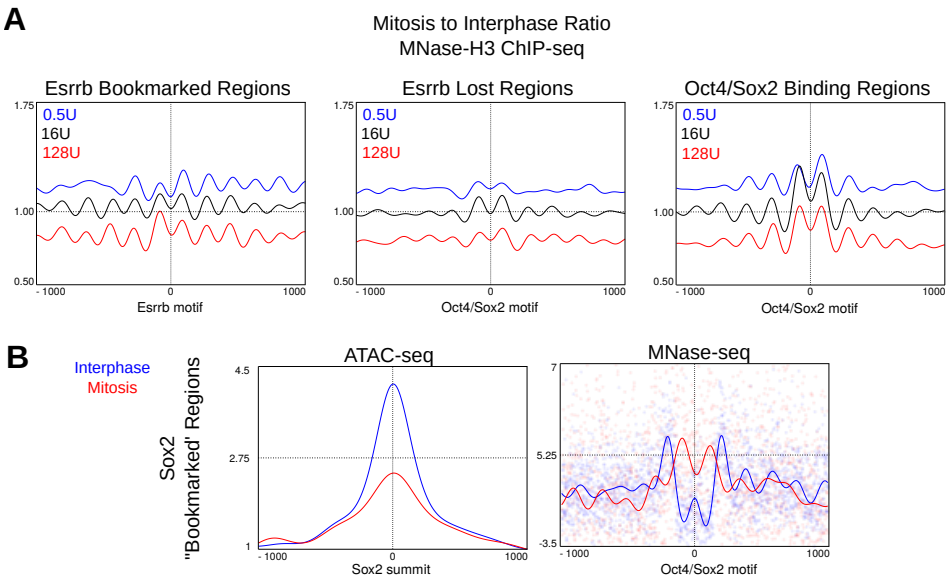
Supplementary Information, Fig. S 3. Extended analysis of Oct4 localisation. (A) Localisation of Oct4-RFP fusion proteins expressed from one of the two endogenous Oct4 alleles in live cells cultured in FCS/LIF medium (top) on in FCS-free 2i-containing medium (bottom). DNA is visualised by Hoechst 33342 (red). **(B)** Oct4 (green in the merge) and Aurkb (red in the merge) immunofluorescence after fixation with DSG. Note the large overlap at PCH and at centromeres in interphase and in mitosis. **(C)** Localisation of Oct4-GFP fusion proteins in live cells cultured in FCS/LIF medium supplemented (bottom) or not (top) with the Aurkb inhibitor Hesperadin. DNA is visualised by Hoechst 33342 (red). Note this image corresponds exactly to that shown in Fig. 3A.



Supplementary Information, Fig. S 4. Example of FLIP imaging. Representative examples of Esrrb-GFP, Sox2-GFP and Oct4-GFP signal on mitotic chromosomes before and after 60 second of continuous bleaching of freely diffusing molecules outside the chromatids.



Supplementary Information, Fig. S 5. Average binding profile in interphase of Esrrb, Sox2, Oct4 and Nanog, in this study and in public datasets. Blue line depicts all binding regions identified in this study, red depicts regions detected in DSG+PFA exclusively, i.e. regions with no significant peak in any of the indicated PFA dataset.



Supplementary Information, Fig. S 6. Additional information of nucleosome organisation at Esrrb, Oct4 and Sox2 sites. (A) Ratio of MNase H3 ChIP-seq nucleosomal size fragment signal of mitosis over interphase, as described in Fig 5, for 0.5U (blue), 16U (black) and 128U (red) MNase concentrations, at Esrrb bookmarked and lost regions and all Oct4+Sox2 binding regions, centred on the top Esrrb motif, top Oct4/Sox2 motif and top Oct4/Sox2 motif respectively. **(B)** Left: Accessibility measured by cut sites of 0-100 bp ATAC-seq fragments at Sox2 putative bookmarked sites in interphase and mitosis, centred on Sox2 peak summits. Right: Nucleosome positioning measured by MNase-seq nucleosomal size fragments (140-200 bp) after digestion with 16U at Sox2 putative bookmarked sites centred on Oct4/Sox2 motif. Vertical scale gives z-score.

IV. Heterochromatin organization and the pluripotency transcription factor OCT4

Recently, an interesting observation was made by a post-doctoral scientist in our lab as a result of an immunostaining for one of the core pluripotency TFs, Oct4. Surprisingly, it was revealed that Oct4 is locally enriched at chromocenters in interphase cells but also at centromeres in mitotic ES cells. This finding was also supported by live imaging of ES cell lines where endogenous Oct4 was fused to an RFP reporter. In line with our interest in the nuclear organization in ES cells, part of which pericentric heterochromatin is, the finding that Oct4 is focalized on centromeres prompted us to investigate whether there is a link between the organization of pericentric heterochromatin, the transcription of satellites repeats and the TF activity of Oct4. In order to address the question, we initially decided to look if Oct4 depletion would affect the transcriptional activity of the major and minor satellites, as well as the structure of the chromocenters. We performed Oct4 KD using RNA interference technology and measured the expression levels of major and minor satellites. We did not find their transcription to be affected in the absence of Oct4 (data not shown). In addition, we conducted immunostainings for the main factors of the heterochromatinization process: H3K9me3 and HP1a. We did not observe any difference in their distribution upon Oct4 depletion and the chromocenters, stained by DAPI and enriched in H3K9me3 and Hp1a signal, were comparable in the presence and absence of Oct4 protein (data not shown).

Some of the data that were obtained during these investigations are presented in the following paper. In addition, my personal contribution consisted in performing the live imaging of different ES cell lines where TFs were fused to fluorescent reporters by spinning-disk confocal microscopy. I conducted and analyzed immunostainings for TFs by confocal microscopy, and I performed Aurora kinase B inhibition assays (with the use of hesperadin) in order to assess the impact on the Oct4 protein in mitotic cells by immunostainings and western blots.

V. References

Abranches, E., Bekman, E., and Henrique, D. (2013). Generation and Characterization of a Novel Mouse Embryonic Stem Cell Line with a Dynamic Reporter of Nanog Expression. PLOS ONE *8*, e59928.

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. Dev. Camb. Engl. *140*, 43–55.

Acampora, D., Omodei, D., Petrosino, G., Garofalo, A., Savarese, M., Nigro, V., Di Giovannantonio, L.G., Mercadante, V., and Simeone, A. (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 Rep. *15*, 2651–2664.

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 Rep. *9*, 1642–1659.

Agrelo, R., Souabni, A., Novatchkova, M., Haslinger, C., Leeb, M., Komnenovic, V., Kishimoto, H., Gresh, L., Kohwi-Shigematsu, T., Kenner, L., et al. (2009). SATB1 Defines the Developmental Context for Gene Silencing by Xist in Lymphoma and Embryonic Cells. Dev. Cell *16*, 507–516.

Ahmed, K., Dehghani, H., Rugg-Gunn, P., Fussner, E., Rossant, J., and Bazett-Jones, D.P. (2010). Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo. PloS One *5*, e10531.

Amândio, A.R., Necsulea, A., Joye, E., Mascrez, B., and Duboule, D. (2016). Hotair Is Dispensible for Mouse Development. PLOS Genet. *12*, e1006232.

Amaral, P.P., Neyt, C., Wilkins, S.J., Askarian-Amiri, M.E., Sunkin, S.M., Perkins, A.C., and Mattick, J.S. (2009). Complex architecture and regulated expression of the Sox2ot locus during vertebrate development. RNA N. Y. N *15*, 2013–2027.

Anderson, D.M., Anderson, K.M., Chang, C.-L., Makarewich, C.A., Nelson, B.R., McAnally, J.R., Kasaragod, P., Shelton, J.M., Liou, J., Bassel-Duby, R., et al. (2015). A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. Cell *160*, 595–606.

Andersson, R. (2015). Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model. BioEssays *37*, 314–323.

Bain, G., Kitchens, D., Yao, M., Huettner, J.E., and Gottlieb, D.I. (1995). Embryonic Stem Cells Express Neuronal Properties in Vitro. Dev. Biol. *168*, 342–357.

Ballarino, M., Cipriano, A., Tita, R., Santini, T., Desideri, F., Morlando, M., Colantoni, A., Carrieri, C., Nicoletti, C., Musarò, A., et al. (2018). Deficiency in the nuclear long noncoding RNACharme causes myogenic defects and heart remodeling in mice. EMBO J. e99697.

Bartolomei, M.S., Zemel, S., and Tilghman, S.M. (1991). Parental imprinting of the mouse H19 gene. Nature *351*, 153–155.

Baubec, T., Colombo, D.F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A.R., Akalin, A., and Schübeler, D. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature *520*, 243–247.

Beagrie, R.A., Scialdone, A., Schueler, M., Kraemer, D.C.A., Chotalia, M., Xie, S.Q., Barbieri, M., Santiago, I. de, Lavitas, L.-M., Branco, M.R., et al. (2017). Complex multi-enhancer contacts captured by genome architecture mapping. Nature *543*, 519–524.

Bell, C.C., Amaral, P.P., Kalsbeek, A., Magor, G.W., Gillinder, K.R., Tangermann, P., di Lisio, L., Cheetham, S.W., Gruhl, F., Frith, J., et al. (2016). The Evx1/Evx1as gene locus regulates anterior-posterior patterning during gastrulation. Sci. Rep. *6*, 26657.

de Belle, I., Cai, S., and Kohwi-Shigematsu, T. (1998). The genomic sequences bound to special AT-rich sequence-binding protein 1 (SATB1) in vivo in Jurkat T cells are tightly associated with the nuclear matrix at the bases of the chromatin loops. J. Cell Biol. *141*, 335–348.

Bensaude, O. (2011). Inhibiting eukaryotic transcription. Transcription *2*, 103–108.

Berezney, R., and Coffey, D.S. (1974). Identification of a nuclear protein matrix. Biochem. Biophys. Res. Commun. *60*, 1410–1417.

Berezney, R., and Jeon, K.W. (1995). Nuclear Matrix: Structural and Functional Organization (Elsevier).

Bergmann, J.H., and Spector, D.L. (2014). Long non-coding RNAs: modulators of nuclear structure and function. Curr. Opin. Cell Biol. *26*, 10–18.

Bergmann, J.H., Li, J., Eckersley-Maslin, M.A., Rigo, F., Freier, S.M., and Spector, D.L. (2015a). Regulation of the ESC transcriptome by nuclear long noncoding RNAs. Genome Res. *25*, 1336–1346.

Bergmann, J.H., Li, J., Eckersley-Maslin, M.A., Rigo, F., Freier, S.M., and Spector, D.L. (2015b). Regulation of the ESC transcriptome by nuclear long noncoding RNAs. Genome Res.

Bernard, D., Prasanth, K.V., Tripathi, V., Colasse, S., Nakamura, T., Xuan, Z., Zhang, M.Q., Sedel, F., Jourdren, L., Coulpier, F., et al. (2010). A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. EMBO J. *29*, 3082–3093.

Bernstein, B.E., Humphrey, E.L., Erlich, R.L., Schneider, R., Bouman, P., Liu, J.S., Kouzarides, T., and Schreiber, S.L. (2002). Methylation of histone H3 Lys 4 in coding regions of active genes. Proc. Natl. Acad. Sci. *99*, 8695–8700.

Berry, J., Weber, S.C., Vaidya, N., Haataja, M., and Brangwynne, C.P. (2015). RNA transcription modulates phase transition-driven nuclear body assembly. Proc. Natl. Acad. Sci. *112*, E5237–E5245.

Bertone, P., Stolc, V., Royce, T.E., Rozowsky, J.S., Urban, A.E., Zhu, X., Rinn, J.L., Tongprasit, W., Samanta, M., Weissman, S., et al. (2004). Global Identification of Human Transcribed Sequences with Genome Tiling Arrays. Science *306*, 2242–2246.

Bogu, G.K., Vizán, P., Stanton, L.W., Beato, M., Di Croce, L., and Marti-Renom, M.A. (2016). Chromatin and RNA Maps Reveal Regulatory Long Noncoding RNAs in Mouse. Mol. Cell. Biol. *36*, 809–819.

Boisvert, F.-M., van Koningsbruggen, S., Navascués, J., and Lamond, A.I. (2007). The multifunctional nucleolus. Nat. Rev. Mol. Cell Biol. *8*, 574–585.

Bolzer, A., Kreth, G., Solovei, I., Koehler, D., Saracoglu, K., Fauth, C., Müller, S., Eils, R., Cremer, C., Speicher, M.R., et al. (2005). Three-Dimensional Maps of All Chromosomes in Human Male Fibroblast Nuclei and Prometaphase Rosettes. PLOS Biol. *3*, e157.

Bonasio, R., and Shiekhattar, R. (2014). Regulation of Transcription by Long Noncoding RNAs. Annu. Rev. Genet. *48*, 433–455.

Bond, C.S., and Fox, A.H. (2009). Paraspeckles: nuclear bodies built on long noncoding RNA. J. Cell Biol. *186*, 637–644.

Boroviak, T., Loos, R., Lombard, P., Okahara, J., Behr, R., Sasaki, E., Nichols, J., Smith, A., and Bertone, P. (2015). Lineage-Specific Profiling Delineates the Emergence and Progression of Naive Pluripotency in Mammalian Embryogenesis. Dev. Cell *35*, 366–382.

Bradley, A., Evans, M., Kaufman, M.H., and Robertson, E. (1984). Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. Nature *309*, 255–256.

Brannan, C.I., Dees, E.C., Ingram, R.S., and Tilghman, S.M. (1990). The product of the H19 gene may function as an RNA. Mol. Cell. Biol. *10*, 28–36.

Brockdorff, N., Ashworth, A., Kay, G.F., McCabe, V.M., Norris, D.P., Cooper, P.J., Swift, S., and Rastan, S. (1992). The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell *71*, 515–526.

Brons, I.G.M., Smithers, L.E., Trotter, M.W.B., Rugg-Gunn, P., Sun, B., Lopes, S.M.C. de S., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R.A., et al. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature *448*, 191–195.

Brown, C.J., Ballabio, A., Rupert, J.L., Lafreniere, R.G., Grompe, M., Tonlorenzi, R., and Willard, H.F. (1991). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature *349*, 38–44.

Brown, C.J., Hendrich, B.D., Rupert, J.L., Lafrenière, R.G., Xing, Y., Lawrence, J., and Willard, H.F. (1992). The human XIST gene: Analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell *71*, 527–542.

Burdon, T., Stracey, C., Chambers, I., Nichols, J., and Smith, A. (1999). Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. Dev. Biol. *210*, 30–43.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. *25*, 1915–1927.

Capco, D.G., Wan, K.M., and Penman, S. (1982). The nuclear matrix: Three-dimensional architecture and protein composition. Cell *29*, 847–858.

Carelli, F.N., Liechti, A., Halbert, J., Warnefors, M., and Kaessmann, H. (2018). Repurposing of promoters and enhancers during mammalian evolution. Nat. Commun. *9*, 4066.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., et al. (2005). The Transcriptional Landscape of the Mammalian Genome. Science *309*, 1559–1563.

Casanova, M., Pasternak, M., El Marjou, F., Le Baccon, P., Probst, A.V., and Almouzni, G. (2013). Heterochromatin Reorganization during Early Mouse Development Requires a Single-Stranded Noncoding Transcript. Cell Rep. *4*, 1156–1167.

Caudron-Herger, M., and Rippe, K. (2012). Nuclear architecture by RNA. Curr. Opin. Genet. Dev. *22*, 179–187.

Cerase, A., Armaos, A., Cid-Samper, F., Avner, P., and Tartaglia, G.G. (2018). Xist lncRNA forms silencing granules that induce heterochromatin formation and repressive complexes recruitment by phase separation.

Chakraborty, D., Paszkowski-Rogacz, M., Berger, N., Ding, L., Mircetic, J., Fu, J., Iesmantavicius, V., Choudhary, C., Anastassiadis, K., Stewart, A.F., et al. (2017). lncRNA Panct1 Maintains Mouse Embryonic Stem Cell Identity by Regulating TOBF1 Recruitment to Oct-Sox Sequences in Early G1. Cell Rep. *21*, 3012–3021.

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.

Chambeyron, S., Silva, N.R.D., Lawson, K.A., and Bickmore, W.A. (2005). Nuclear re-organisation of the Hoxb complex during mouse embryonic development. Development *132*, 2215–2223.

Chaumeil, J., Baccon, P.L., Wutz, A., and Heard, E. (2006). A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. Genes Dev. *20*, 2223–2237.

Chen, L.-L., and Carmichael, G.G. (2009). Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. Mol. Cell *35*, 467–478.

Chen, J., Shishkin, A.A., Zhu, X., Kadri, S., Maza, I., Guttman, M., Hanna, J.H., Regev, A., and Garber, M. (2016). Evolutionary analysis across mammals reveals distinct classes of long non-coding RNAs. Genome Biol. *17*, 19.

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.

Cheng, L., Ming, H., Zhu, M., and Wen, B. (2016). Long noncoding RNAs as Organizers of Nuclear Architecture. Sci. China Life Sci. *59*, 236–244.

Cho, W.-K., Spille, J.-H., Hecht, M., Lee, C., Li, C., Grube, V., and Cisse, I.I. (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science *361*, 412–415.

Choi, J., Clement, K., Huebner, A.J., Webster, J., Rose, C.M., Brumbaugh, J., Walsh, R.M., Lee, S., Savol, A., Etchegaray, J.-P., et al. (2017). DUSP9 Modulates DNA Hypomethylation in Female Mouse Pluripotent Stem Cells. Cell Stem Cell *20*, 706-719.e7.

Chow, J.C., Ciaudo, C., Fazzari, M.J., Mise, N., Servant, N., Glass, J.L., Attreed, M., Avner, P., Wutz, A., Barillot, E., et al. (2010). LINE-1 Activity in Facultative Heterochromatin Formation during X Chromosome Inactivation. Cell *141*, 956–969.

Chu, C., Qu, K., Zhong, F.L., Artandi, S.E., and Chang, H.Y. (2011). Genomic Maps of Long Noncoding RNA Occupancy Reveal Principles of RNA-Chromatin Interactions. Mol. Cell *44*, 667–678.

Chujo, T., Yamazaki, T., Kawaguchi, T., Kurosaka, S., Takumi, T., Nakagawa, S., and Hirose, T. (2017). Unusual semi-extractability as a hallmark of nuclear body-associated architectural noncoding RNAs. EMBO J. e201695848.

Clark, M.B., Johnston, R.L., Inostroza-Ponta, M., Fox, A.H., Fortini, E., Moscato, P., Dinger, M.E., and Mattick, J.S. (2012). Genome-wide analysis of long noncoding RNA stability. Genome Res. *22*, 885–898.

Clemson, C.M., McNeil, J.A., Willard, H.F., and Lawrence, J.B. (1996). XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. J. Cell Biol. *132*, 259–275.

Clemson, C.M., Hutchinson, J.N., Sara, S.A., Ensminger, A.W., Fox, A.H., Chess, A., and Lawrence, J.B. (2009). An Architectural Role for a Nuclear Noncoding RNA: NEAT1 RNA Is Essential for the Structure of Paraspeckles. Mol. Cell *33*, 717–726.

Cohen, S.M. (2014). Everything old is new again: (linc)RNAs make proteins! EMBO J. *33*, 937–938.

Cole, M.F., Johnstone, S.E., Newman, J.J., Kagey, M.H., and Young, R.A. (2008). Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. Genes Dev. *22*, 746–755.

Cremer, T., Cremer, M., Dietzel, S., Müller, S., Solovei, I., and Fakan, S. (2006). Chromosome territories--a functional nuclear landscape. Curr. Opin. Cell Biol. *18*, 307–316.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. *107*, 21931–21936.

Croft, J.A., Bridger, J.M., Boyle, S., Perry, P., Teague, P., and Bickmore, W.A. (1999). Differences in the localization and morphology of chromosomes in the human nucleus. J. Cell Biol. *145*, 1119–1131.

Dang-Nguyen, T.Q., and Torres-Padilla, M.-E. (2015). How cells build totipotency and pluripotency: nuclear, chromatin and transcriptional architecture. Curr. Opin. Cell Biol. *34*, 9–15.

Dekker, J., and Heard, E. (2015). Structural and functional diversity of Topologically Associating Domains. FEBS Lett. *589*, 2877–2884.

Dekker, J., and Mirny, L. (2016). The 3D Genome as Moderator of Chromosomal Communication. Cell *164*, 1110–1121.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306–1311.

Deng, Q., Ramsköld, D., Reinius, B., and Sandberg, R. (2014). Single-Cell RNA-Seq Reveals Dynamic, Random Monoallelic Gene Expression in Mammalian Cells. Science *343*, 193–196.

Dinger, M.E., Amaral, P.P., Mercer, T.R., Pang, K.C., Bruce, S.J., Gardiner, B.B., Askarian-Amiri, M.E., Ru, K., Soldà, G., Simons, C., et al. (2008). Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. Genome Res. *18*, 1433–1445.

Dinger, M.E., Amaral, P.P., Mercer, T.R., and Mattick, J.S. (2009). Pervasive transcription of the eukaryotic genome: functional indices and conceptual implications. Brief. Funct. Genomics *8*, 407–423.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. Nature *489*, 101–108.

Duffié, R., Ajjan, S., Greenberg, M.V., Zamudio, N., Escamilla del Arenal, M., Iranzo, J., Okamoto, I., Barbaux, S., Fauque, P., and Bourc'his, D. (2014). The Gpr1/Zdbf2 locus provides new paradigms for transient and dynamic genomic imprinting in mammals. Genes Dev. *28*, 463–478.

Dundr, M. (2012). Nuclear bodies: multifunctional companions of the genome. Curr. Opin. Cell Biol. *24*, 415–422.

Dundr, M., and Misteli, T. (2010). Biogenesis of Nuclear Bodies. Cold Spring Harb. Perspect. Biol. *2*, a000711.

Engelke, R., Riede, J., Hegermann, J., Wuerch, A., Eimer, S., Dengjel, J., and Mittler, G. (2014). The Quantitative Nuclear Matrix Proteome as a Biochemical Snapshot of Nuclear Organization. J. Proteome Res. *13*, 3940–3956.

Engreitz, J.M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Kadri, S., Xing, J., Goren, A., Lander, E.S., et al. (2013). The Xist lncRNA Exploits Three-Dimensional Genome Architecture to Spread Across the X Chromosome. Science *341*, 1237973.

Engreitz, J.M., Ollikainen, N., and Guttman, M. (2016). Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. Nat. Rev. Mol. Cell Biol. *17*, 756–770.

Eskiw, C.H., Rapp, A., Carter, D.R.F., and Cook, P.R. (2008). RNA polymerase II activity is located on the surface of protein-rich transcription factories. J. Cell Sci. *121*, 1999–2007.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154–156.

Factor, D.C., Corradin, O., Zentner, G.E., Saiakhova, A., Song, L., Chenoweth, J.G., McKay, R.D., Crawford, G.E., Scacheri, P.C., and Tesar, P.J. (2014). Epigenomic comparison reveals activation of "seed" enhancers during transition from naive to primed pluripotency. Cell Stem Cell *14*, 854–863.

Ferguson-Smith, A.C. (2011). Genomic imprinting: the emergence of an epigenetic paradigm. Nat. Rev. Genet. *12*, 565–575.

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.

Fiorini, A., Gouveia, F. de S., and Fernandez, M.A. (2006). Scaffold/Matrix Attachment Regions and intrinsic DNA curvature. Biochem. Biokhimiia *71*, 481–488.

Fox, A.H., Nakagawa, S., Hirose, T., and Bond, C.S. (2018). Paraspeckles: Where Long Noncoding RNA Meets Phase Separation. Trends Biochem. Sci. *43*, 124–135.

Fraichard, A., Chassande, O., Bilbaut, G., Dehay, C., Savatier, P., and Samarut, J. (1995). In vitro differentiation of embryonic stem cells into glial cells and functional neurons. J. Cell Sci. *108*, 3181–3188.

Francastel, C., Schübeler, D., Martin, D.I., and Groudine, M. (2000). Nuclear compartmentalization and gene activity. Nat. Rev. Mol. Cell Biol. *1*, 137–143.

Fulco, C.P., Munschauer, M., Anyoha, R., Munson, G., Grossman, S.R., Perez, E.M., Kane, M., Cleary, B., Lander, E.S., and Engreitz, J.M. (2016). Systematic mapping of functional enhancer-promoter connections with CRISPR interference. Science aag2445.

Fussner, E., Djuric, U., Strauss, M., Hotta, A., Perez-Iratxeta, C., Lanner, F., Dilworth, F.J., Ellis, J., and Bazett-Jones, D.P. (2011). Constitutive heterochromatin reorganization during somatic cell reprogramming. EMBO J. *30*, 1778–1789.

Gabory, A., Jammes, H., and Dandolo, L. (2010). The H19 locus: Role of an imprinted non-coding RNA in growth and development. BioEssays *32*, 473–480.

Georgomanolis, T., Sofiadis, K., and Papantonis, A. (2016). Cutting a Long Intron Short: Recursive Splicing and Its Implications. Front. Physiol. *7*.

Gibcus, J.H., and Dekker, J. (2013). The Hierarchy of the 3D Genome. Mol. Cell *49*, 773–782.

Gil, N., and Ulitsky, I. (2018). Production of spliced long noncoding RNAs specifies regions with increased enhancer activity. BioRxiv 319400.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. Cell *154*, 442–451.

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647–661.

Goldman, R.D., Gruenbaum, Y., Moir, R.D., Shumaker, D.K., and Spann, T.P. (2002). Nuclear lamins: building blocks of nuclear architecture. Genes Dev. *16*, 533–547.

Goudarzi, M., Berg, K., Pieper, L.M., and Schier, A.F. (2018). Long non-coding RNAs are largely dispensable for zebrafish embryogenesis, viability and fertility. BioRxiv 374702.

Grant, J., Mahadevaiah, S.K., Khil, P., Sangrithi, M.N., Royo, H., Duckworth, J., McCarrey, J.R., VandeBerg, J.L., Renfree, M.B., Taylor, W., et al. (2012). Rsx, a metatherian RNA with Xist-like properties. Nature *487*, 254–258.

Greenberg, M., Teissander, A., Walter, M., Noordermeer, D., and Bourc'his, D. (2018). Dynamic enhancer partitioning instructs activation of a growth regulator during exit from naïve pluripotency. BioRxiv 441824.

Greenberg, M.V.C., Glaser, J., Borsos, M., Marjou, F.E., Walter, M., Teissandier, A., and Bourc'his, D. (2017). Transient transcription in the early embryo sets an epigenetic state that programs postnatal growth. Nat. Genet. *49*, 110–118.

Gruenbaum, Y., Margalit, A., Goldman, R.D., Shumaker, D.K., and Wilson, K.L. (2005). The nuclear lamina comes of age. Nat. Rev. Mol. Cell Biol. *6*, 21–31.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W., et al. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature *453*, 948–951.

Guo, G., Huang, Y., Humphreys, P., Wang, X., and Smith, A. (2011). A PiggyBac-Based Recessive Screening Method to Identify Pluripotency Regulators. PLOS ONE *6*, e18189.

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y., et al. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. Cell *162*, 900–910.

Guttman, M. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature *458*, 223–227.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature *458*, 223–227.

Guttman, M., Garber, M., Levin, J.Z., Donaghey, J., Robinson, J., Adiconis, X., Fan, L., Koziol, M.J., Gnirke, A., Nusbaum, C., et al. (2010a). *Ab initio* reconstruction of cell type–specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat. Biotechnol. *28*, 503–510.

Guttman, M., Garber, M., Levin, J.Z., Donaghey, J., Robinson, J., Adiconis, X., Fan, L., Koziol, M.J., Gnirke, A., Nusbaum, C., et al. (2010b). Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat. Biotechnol. *28*, 503–510.

Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B., Ach, R., Bruhn, L., et al. (2011a). lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature *477*, 295–300.

Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B., Ach, R., Bruhn, L., et al. (2011b). lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature *477*, 295–300.

Hacisuleyman, E., Goff, L.A., Trapnell, C., Williams, A., Henao-Mejia, J., Sun, L., McClanahan, P., Hendrickson, D.G., Sauvageau, M., Kelley, D.R., et al. (2014). Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. Nat. Struct. Mol. Biol. *21*, 198–206.

Hacisuleyman, E., Shukla, C.J., Weiner, C.L., and Rinn, J.L. (2016). Function and evolution of local repeats in the *Firre* locus. Nat. Commun. *7*, 11021.

Hall, L.L., Smith, K.P., Byron, M., and Lawrence, J.B. (2006). Molecular anatomy of a speckle. Anat. Rec. A. Discov. Mol. Cell. Evol. Biol. *288A*, 664–675.

Hall, L.L., Carone, D.M., Gomez, A.V., Kolpa, H.J., Byron, M., Mehta, N., Fackelmayer, F.O., and Lawrence, J.B. (2014a). Stable C0T-1 Repeat RNA Is Abundant and Is Associated with Euchromatic Interphase Chromosomes. Cell *156*, 907–919.

Hall, L.L., Carone, D.M., Gomez, A.V., Kolpa, H.J., Byron, M., Mehta, N., Fackelmayer, F.O., and Lawrence, J.B. (2014b). Stable C0T-1 Repeat RNA Is Abundant and Is Associated with Euchromatic Interphase Chromosomes. Cell *156*, 907–919.

Han, X., Luo, S., Peng, G., Lu, J.Y., Cui, G., Liu, L., Yan, P., Yin, Y., Liu, W., Wang, R., et al. (2018). Mouse knockout models reveal largely dispensable but context-dependent functions of lncRNAs during development. J. Mol. Cell Biol. *10*, 175–178.

Han, Z., Feng, J., Hong, Z., Chen, L., Li, W., Liao, S., Wang, X., Ji, T., Wang, S., Ma, D., et al. (2013). Silencing of the STAT3 signaling pathway reverses the inherent and induced

chemoresistance of human ovarian cancer cells. Biochem. Biophys. Res. Commun. *435*, 188–194.

Handoko, L., Xu, H., Li, G., Ngan, C.Y., Chew, E., Schnapp, M., Lee, C.W.H., Ye, C., Ping, J.L.H., Mulawadi, F., et al. (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. Nat. Genet. *43*, 630–638.

Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K., and Kjems, J. (2013). Natural RNA circles function as efficient microRNA sponges. Nature *495*, 384–388.

Hasegawa, Y., Brockdorff, N., Kawano, S., Tsutui, K., Tsutui, K., and Nakagawa, S. (2010). The Matrix Protein hnRNP U Is Required for Chromosomal Localization of Xist RNA. Dev. Cell *19*, 469–476.

Hathaway, N.A., Bell, O., Hodges, C., Miller, E.L., Neel, D.S., and Crabtree, G.R. (2012). Dynamics and memory of heterochromatin in living cells. Cell *149*, 1447–1460.

Hayashi, K., de Sousa Lopes, S.M.C., and Surani, M.A. (2007). Germ cell specification in mice. Science *316*, 394–396.

Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., and Saitou, M. (2011). Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells. Cell *146*, 519–532.

Hayashi, T., Ozaki, H., Sasagawa, Y., Umeda, M., Danno, H., and Nikaido, I. (2018). Single-cell full-length total RNA sequencing uncovers dynamics of recursive splicing and enhancer RNAs. Nat. Commun. *9*, 619.

He, D.C., Nickerson, J.A., and Penman, S. (1990). Core filaments of the nuclear matrix. J. Cell Biol. *110*, 569–580.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Calcar, S.V., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318.

Herman, R., Weymouth, L., and Penman, S. (1978). Heterogeneous nuclear RNA-protein fibers in chromatin-depleted nuclei. J. Cell Biol. *78*, 663–674.

Heurtier, V., Owens, N., Gonzalez, I., Mueller, F., Proux, C., Mornico, D., Clerc, P., Dubois, A., and Navarro, P. (2018). The molecular logic of Nanog-induced self-renewal. BioRxiv 374371.

Hezroni, H., Koppstein, D., Schwartz, M.G., Avrutin, A., Bartel, D.P., and Ulitsky, I. (2015). Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. Cell Rep. *11*, 1110–1122.

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat. Biotechnol. *33*, 510–517.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-Enhancers in the Control of Cell Identity and Disease. Cell *155*, 934–947.

Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A Phase Separation Model for Transcriptional Control. Cell *169*, 13–23.

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. Nat. Biotechnol. *31*, 827–832.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and Applications of CRISPR-Cas9 for Genome Engineering. Cell *157*, 1262–1278.

Huang, Y., Osorno, R., Tsakiridis, A., and Wilson, V. (2012). In Vivo differentiation potential of epiblast stem cells revealed by chimeric embryo formation. Cell Rep. *2*, 1571–1578.

Hussein, S.M.I., Puri, M.C., Tonge, P.D., Benevento, M., Corso, A.J., Clancy, J.L., Mosbergen, R., Li, M., Lee, D.-S., Cloonan, N., et al. (2014). Genome-wide characterization of the routes to pluripotency. Nature *516*, 198–206.

Hutchinson, J.N., Ensminger, A.W., Clemson, C.M., Lynch, C.R., Lawrence, J.B., and Chess, A. (2007). A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. BMC Genomics *8*, 39.

Ip, J.Y., and Nakagawa, S. (2012). Long non-coding RNAs in nuclear bodies. Dev. Growth Differ. *54*, 44–54.

Ishizuka, A., Hasegawa, Y., Ishida, K., Yanaka, K., and Nakagawa, S. (2014). Formation of nuclear bodies by the lncRNA Gomafu-associating proteins Celf3 and SF1. Genes Cells Devoted Mol. Cell. Mech. *19*, 704–721.

Jiapaer, Z., Li, G., Ye, D., Bai, M., Li, J., Guo, X., Du, Y., Su, D., Jia, W., Chen, W., et al. (2018). LincU Preserves Naïve Pluripotency by Restricting ERK Activity in Embryonic Stem Cells. Stem Cell Rep.

Joung, J., Engreitz, J.M., Konermann, S., Abudayyeh, O.O., Verdine, V.K., Aguet, F., Gootenberg, J.S., Sanjana, N.E., Wright, J.B., Fulco, C.P., et al. (2017). Genome-scale activation screen identifies a lncRNA locus regulating a gene neighbourhood. Nature *548*, 343–346.

Kapusta, A., Kronenberg, Z., Lynch, V.J., Zhuo, X., Ramsay, L., Bourque, G., Yandell, M., and Feschotte, C. (2013). Transposable Elements Are Major Contributors to the Origin, Diversification, and Regulation of Vertebrate Long Noncoding RNAs. PLOS Genet. *9*, e1003470.

Kataoka, N., Dobashi, I., Hagiwara, M., and Ohno, M. (2013). hDbr1 is a nucleocytoplasmic shuttling protein with a protein phosphatase-like motif essential for debranching activity. Sci. Rep. *3*, 1090.

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., et al. (2005). Antisense Transcription in the Mammalian Transcriptome. Science *309*, 1564–1566.

Kelley, D., and Rinn, J. (2012). Transposable elements reveal a stem cell-specific class of long noncoding RNAs. Genome Biol. *13*, R107.

Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Morales, D.R., Thomas, K., Presser, A., Bernstein, B.E., Oudenaarden, A. van, et al. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. Proc. Natl. Acad. Sci. *106*, 11667–11672.

Kim, T.-K., and Shiekhattar, R. (2015). Architectural and Functional Commonalities between Enhancers and Promoters. Cell *162*, 948–959.

Kim, T.-K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010a). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182–187.

Kim, T.-K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010b). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182–187.

Kind, J., Pagie, L., de Vries, S.S., Nahidiazar, L., Dey, S.S., Bienko, M., Zhan, Y., Lajoie, B., de Graaf, C.A., Amendola, M., et al. (2015). Genome-wide maps of nuclear lamina interactions in single human cells. Cell *163*, 134–147.

Kino, T., Hurt, D.E., Ichijo, T., Nader, N., and Chrousos, G.P. (2010). Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. Sci. Signal. *3*, ra8.

Kirk, J.M., Kim, S.O., Inoue, K., Smola, M.J., Lee, D.M., Schertzer, M.D., Wooten, J.S., Baker, A.R., Sprague, D., Collins, D.W., et al. (2018). Functional classification of long non-coding RNAs by k -mer content. Nat. Genet. 1.

Klattenhoff, C.A., Scheuermann, J.C., Surface, L.E., Bradley, R.K., Fields, P.A., Steinhauser, M.L., Ding, H., Butty, V.L., Torrey, L., Haas, S., et al. (2013). Braveheart, a Long Noncoding RNA Required for Cardiovascular Lineage Commitment. Cell *152*, 570–583.

Koch, F., and Andrau, J.-C. (2011). Initiating RNA Polymerase II and TIPs as hallmarks of enhancer activity and tissue-specificity. Transcription *2*, 263–268.

Kojima, Y., Kaufman-Francis, K., Studdert, J.B., Steiner, K.A., Power, M.D., Loebel, D.A.F., Jones, V., Hor, A., de Alencastro, G., Logan, G.J., et al. (2014). The transcriptional and functional properties of mouse epiblast stem cells resemble the anterior primitive streak. Cell Stem Cell *14*, 107–120.

Kopp, J.L., Ormsbee, B.D., Desler, M., and Rizzino, A. (2008). Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. Stem Cells Dayt. Ohio *26*, 903–911.

Kosak, S.T., Scalzo, D., Alworth, S.V., Li, F., Palmer, S., Enver, T., Lee, J.S.J., and Groudine, M. (2007). Coordinate gene regulation during hematopoiesis is related to genomic organization. PLoS Biol. *5*, e309.

Kowalczyk, M.S., Hughes, J.R., Garrick, D., Lynch, M.D., Sharpe, J.A., Sloane-Stanley, J.A., McGowan, S.J., De Gobbi, M., Hosseini, M., Vernimmen, D., et al. (2012). Intragenic Enhancers Act as Alternative Promoters. Mol. Cell *45*, 447–458.

Kubo, N., Ishii, H., Gorkin, D., Meitinger, F., Xiong, X., Fang, R., Liu, T., Ye, Z., Li, B., Dixon, J., et al. (2017). Preservation of Chromatin Organization after Acute Loss of CTCF in Mouse Embryonic Stem Cells. BioRxiv 118737.

Kunath, T., Saba-El-Leil, M.K., Almousailleakh, M., Wray, J., Meloche, S., and Smith, A. (2007). FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. Development *134*, 2895–2902.

Kutter, C., Watt, S., Stefflova, K., Wilson, M.D., Goncalves, A., Ponting, C.P., Odom, D.T., and Marques, A.C. (2012). Rapid turnover of long noncoding RNAs and the evolution of gene expression. PLoS Genet. *8*, e1002841.

Lai, F., Orom, U.A., Cesaroni, M., Beringer, M., Taatjes, D.J., Blobel, G.A., and Shiekhattar, R. (2013). Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature *494*, 497–501.

Lam, M.T., Li, W., Rosenfeld, M.G., and Glass, C.K. (2014). Enhancer RNAs and regulated transcriptional programs. Trends Biochem. Sci. *39*, 170–182.

Lamond, A.I., and Earnshaw, W.C. (1998). Structure and function in the nucleus. Science *280*, 547–553.

Langdon, E.M., Qiu, Y., Niaki, A.G., McLaughlin, G.A., Weidmann, C., Gerbich, T.M., Smith, J.A., Crutchley, J.M., Termini, C.M., Weeks, K.M., et al. (2018). MRNA structure determines specificity of a polyQ-driven phase separation. Science eaar7432.

Larson, A.G., Elnatan, D., Keenen, M.M., Trnka, M.J., Johnston, J.B., Burlingame, A.L., Agard, D.A., Redding, S., and Narlikar, G.J. (2017). Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. Nature *547*, 236–240.

Latos, P.A., Pauler, F.M., Koerner, M.V., Şenergin, H.B., Hudson, Q.J., Stocsits, R.R., Allhoff, W., Stricker, S.H., Klement, R.M., Warczok, K.E., et al. (2012). Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. Science *338*, 1469–1472.

Lee, H.J., Hore, T.A., and Reik, W. (2014). Reprogramming the Methylome: Erasing Memory and Creating Diversity. Cell Stem Cell *14*, 710–719.

Lee, S., Kopp, F., Chang, T.-C., Sataluri, A., Chen, B., Sivakumar, S., Yu, H., Xie, Y., and Mendell, J.T. (2016). Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. Cell *164*, 69–80.

Leeb, M., Dietmann, S., Paramor, M., Niwa, H., and Smith, A. (2014). Genetic Exploration of the Exit from Self-Renewal Using Haploid Embryonic Stem Cells. Cell Stem Cell *14*, 385–393.

Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X., et al. (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature *498*, 516–520.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289–293.

Lin, M.F., Jungreis, I., and Kellis, M. (2011). PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. Bioinforma. Oxf. Engl. *27*, i275-282.

Lin, N., Chang, K.-Y., Li, Z., Gates, K., Rana, Z.A., Dang, J., Zhang, D., Han, T., Yang, C.-S., Cunningham, T.J., et al. (2014). An evolutionarily conserved long noncoding RNA TUNA controls pluripotency and neural lineage commitment. Mol. Cell *53*, 1005–1019.

Lionnet, T., and Singer, R.H. (2012). Transcription goes digital. EMBO Rep. *13*, 313–321.

Liu, Z., Legant, W.R., Chen, B.-C., Li, L., Grimm, J.B., Lavis, L.D., Betzig, E., and Tjian, R. (2014). 3D imaging of Sox2 enhancer clusters in embryonic stem cells.

López-Maury, L., Marguerat, S., and Bähler, J. (2008). Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. Nat. Rev. Genet. *9*, 583–593.

Lubelsky, Y., and Ulitsky, I. (2018). Sequences enriched in Alu repeats drive nuclear localization of long RNAs in human cells. Nature *555*, 107–111.

Lv, J., Liu, H., Yu, S., Liu, H., Cui, W., Gao, Y., Zheng, T., Qin, G., Guo, J., Zeng, T., et al. (2015). Identification of 4438 novel lincRNAs involved in mouse pre-implantation embryonic development. Mol. Genet. Genomics MGG *290*, 685–697.

Ma, H., Siegel, A.J., and Berezney, R. (1999). Association of chromosome territories with the nuclear matrix. Disruption of human chromosome territories correlates with the release of a subset of nuclear matrix proteins. J. Cell Biol. *146*, 531–542.

Ma, W., Ay, F., Lee, C., Gulsoy, G., Deng, X., Cook, S., Hesson, J., Cavanaugh, C., Ware, C.B., Krumm, A., et al. (2015). Fine-scale chromatin interaction maps reveal the *cis*-regulatory landscape of human lincRNA genes. Nat. Methods *12*, 71–78.

Ma, Z., Swigut, T., Valouev, A., Rada-Iglesias, A., and Wysocka, J. (2011). Sequence-specific regulator Prdm14 safeguards mouse ESCs from entering extraembryonic endoderm fates. Nat. Struct. Mol. Biol. *18*, 120–127.

Maamar, H., Cabili, M.N., Rinn, J., and Raj, A. (2013). linc-HOXA1 is a noncoding RNA that represses Hoxa1 transcription in cis. Genes Dev. *27*, 1260–1271.

Maass, P.G., Barutcu, A.R., and Rinn, J.L. (2018). Interchromosomal interactions: A genomic love story of kissing chromosomes. J Cell Biol jcb.201806052.

Maharana, S., Wang, J., Papadopoulos, D.K., Richter, D., Pozniakovsky, A., Poser, I., Bickle, M., Rizk, S., Guillén-Boixet, J., Franzmann, T., et al. (2018). RNA buffers the phase separation behavior of prion-like RNA binding proteins. Science eaar7366.

Mangan, H., Gailín, M.Ó., and McStay, B. Integrating the genomic architecture of human nucleolar organizer regions with the biophysical properties of nucleoli. FEBS J. *284*, 3977–3985.

Mao, Y.S., Zhang, B., and Spector, D.L. (2011). Biogenesis and function of nuclear bodies. Trends Genet. *27*, 295–306.

Marks, H., Kalkan, T., Menafra, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Francis Stewart, A., Smith, A., et al. (2012). The Transcriptional and Epigenomic Foundations of Ground State Pluripotency. Cell *149*, 590–604.

Marques, A.C., Hughes, J., Graham, B., Kowalczyk, M.S., Higgs, D.R., and Ponting, C.P. (2013). Chromatin signatures at transcriptional start sites separate two equally populated yet distinct classes of intergenic long noncoding RNAs. Genome Biol. *14*, R131.

Marson, A., Foreman, R., Chevalier, B., Bilodeau, S., Kahn, M., Young, R.A., and Jaenisch, R. (2008). Wnt signaling promotes reprogramming of somatic cells to pluripotency. Cell Stem Cell *3*, 132–135.

Martello, G., Sugimoto, T., Diamanti, E., Joshi, A., Hannah, R., Ohtsuka, S., Göttgens, B., Niwa, H., and Smith, A. (2012). Esrrb Is a Pivotal Target of the Gsk3/Tcf3 Axis Regulating Embryonic Stem Cell Self-Renewal. Cell Stem Cell *11*, 491–504.

Martello, G., Bertone, P., and Smith, A. (2013). Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. EMBO J. *32*, 2561–2574.

Masaki, S., Yoshimoto, R., Kaida, D., Hata, A., Satoh, T., Ohno, M., and Kataoka, N. (2015). Identification of the specific interactors of the human lariat RNA debranching enzyme 1 protein. Int. J. Mol. Sci. *16*, 3705–3721.

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A.A., et al. (2007). Pluripotency governed by *Sox2* via regulation of *Oct3/4* expression in mouse embryonic stem cells. Nat. Cell Biol. *9*, 625–635.

Mayer, C., and Grummt, I. (2005). Cellular stress and nucleolar function. Cell Cycle Georget. Tex *4*, 1036–1038.

McHugh, C.A., Chen, C.-K., Chow, A., Surka, C.F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., et al. (2015). The *Xist* lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature *521*, 232–236.

Meister, P., Towbin, B.D., Pike, B.L., Ponti, A., and Gasser, S.M. (2010). The spatial dynamics of tissue-specific promoters during C. elegans development. Genes Dev. *24*, 766–782.

Melcer, S., and Meshorer, E. (2010). The silence of the LADs: dynamic genome-lamina interactions during ESC differentiation. Cell Stem Cell *6*, 495–497.

Melnik, S., Deng, B., Papantonis, A., Baboo, S., Carr, I.M., and Cook, P.R. (2011). The proteomes of transcription factories containing RNA polymerases I, II or III. Nat. Methods *8*, 963–968.

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. Nature *495*, 333–338.

Mercer, T.R., and Mattick, J.S. (2013). Structure and function of long noncoding RNAs in epigenetic regulation. Nat. Struct. Mol. Biol. *20*, 300–307.

Mercer, T.R., Dinger, M.E., Sunkin, S.M., Mehler, M.F., and Mattick, J.S. (2008). Specific expression of long noncoding RNAs in the mouse brain. Proc. Natl. Acad. Sci. U. S. A. *105*, 716–721.

Merkenschlager, M., and Nora, E.P. (2016). CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. Annu. Rev. Genomics Hum. Genet. *17*, 17–43.

Meshorer, E., and Misteli, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. Nat. Rev. Mol. Cell Biol. *7*, 540–546.

Meshorer, E., Yellajoshula, D., George, E., Scambler, P.J., Brown, D.T., and Misteli, T. (2006). Hyperdynamic Plasticity of Chromatin Proteins in Pluripotent Embryonic Stem Cells. Dev. Cell *10*, 105–116.

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.

Minajigi, A., Froberg, J.E., Wei, C., Sunwoo, H., Kesner, B., Colognori, D., Lessing, D., Payer, B., Boukhali, M., Haas, W., et al. (2015). A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. Science *349*, aab2276.

Mintz, P.J., Patterson, S.D., Neuwald, A.F., Spahr, C.S., and Spector, D.L. (1999). Purification and biochemical characterization of interchromatin granule clusters. EMBO J. *18*, 4308–4320.

Moorthy, S.D., Davidson, S., Shchuka, V.M., Singh, G., Malek-Gilani, N., Langroudi, L., Martchenko, A., So, V., Macpherson, N.N., and Mitchell, J.A. (2017). Enhancers and super-enhancers have an equivalent regulatory role in embryonic stem cells through regulation of single or multiple genes. Genome Res. *27*, 246–258.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods *5*, 621–628.

Murrell, A., Heeson, S., and Reik, W. (2004). Interaction between differentially methylated regions partitions the imprinted genes *Igf2* and *H19* into parent-specific chromatin loops. Nat. Genet. *36*, 889–893.

Nagano, T., Mitchell, J.A., Sanz, L.A., Pauler, F.M., Ferguson-Smith, A.C., Feil, R., and Fraser, P. (2008). The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science *322*, 1717–1720.

Nagano, T., Lubling, Y., Stevens, T.J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E.D., Tanay, A., and Fraser, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature *502*, 59–64.

Nakagawa, S., Naganuma, T., Shioi, G., and Hirose, T. (2011). Paraspeckles are subpopulation-specific nuclear bodies that are not essential in mice. J. Cell Biol. *193*, 31–39.

Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B.R., Mirny, L.A., and Dekker, J. (2013). Organization of the mitotic chromosome. Science *342*, 948–953.

Necsulea, A., Soumillon, M., Warnefors, M., Liechti, A., Daish, T., Zeller, U., Baker, J.C., Grützner, F., and Kaessmann, H. (2014). The evolution of lncRNA repertoires and expression patterns in tetrapods. Nature *505*, 635–640.

Nelson, B.R., Makarewich, C.A., Anderson, D.M., Winders, B.R., Troupes, C.D., Wu, F., Reese, A.L., McAnally, J.R., Chen, X., Kavalali, E.T., et al. (2016). A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. Science *351*, 271–275.

Nickerson, J. (2001). Experimental observations of a nuclear matrix. J. Cell Sci. *114*, 463–474.

Nickerson, J.A., Krochmalnic, G., Wan, K.M., and Penman, S. (1989). Chromatin architecture and nuclear RNA. Proc. Natl. Acad. Sci. *86*, 177–181.

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

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., Berkum, N.L. van, Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

Nora, E.P., Goloborodko, A., Valton, A.-L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell *169*, 930-944.e22.

Novo, C.L., Javierre, B.-M., Cairns, J., Segonds-Pichon, A., Wingett, S.W., Freire-Pritchett, P., Furlan-Magaril, M., Schoenfelder, S., Fraser, P., and Rugg-Gunn, P.J. (2018). Long-Range Enhancer Interactions Are Prevalent in Mouse Embryonic Stem Cells and Are Reorganized upon Pluripotent State Transition. Cell Rep. *22*, 2615–2627.

Ørom, U.A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q., et al. (2010). Long Noncoding RNAs with Enhancer-like Function in Human Cells. Cell *143*, 46–58.

Pai, A.A., Paggi, J.M., Yan, P., Adelman, K., and Burge, C.B. (2018). Numerous recursive sites contribute to accuracy of splicing in long introns in flies. PLOS Genet. *14*, e1007588.

Pandey, R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D., and Kanduri, C. (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol. Cell *32*, 232–246.

Paralkar, V.R., Taborda, C.C., Huang, P., Yao, Y., Kossenkov, A.V., Prasad, R., Luan, J., Davies, J.O.J., Hughes, J.R., Hardison, R.C., et al. (2016). Unlinking an lncRNA from Its Associated cis Element. Mol. Cell *62*, 104–110.

Pardo, M., Lang, B., Yu, L., Prosser, H., Bradley, A., Babu, M.M., and Choudhary, J. (2010). An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. Cell Stem Cell *6*, 382–395.

Pascual-Reguant, L., Blanco, E., Galan, S., Dily, F.L., Cuartero, Y., Serra-Bardenys, G., Carlo, V.D., Iturbide, A., Cebrià-Costa, J.P., Nonell, L., et al. (2018). Lamin B1 mapping reveals the existence of dynamic and functional euchromatin lamin B1 domains. Nat. Commun. *9*, 3420.

Pegueroles, C., and Gabaldón, T. (2016). Secondary structure impacts patterns of selection in human lncRNAs. BMC Biol. *14*.

Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S., and Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. Nature *379*, 131–137.

Percharde, M., Lin, C.-J., Yin, Y., Guan, J., Peixoto, G.A., Bulut-Karslioglu, A., Biechele, S., Huang, B., Shen, X., and Ramalho-Santos, M. (2018). A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. Cell *174*, 391-405.e19.

Pereira, L., Yi, F., and Merrill, B.J. (2006). Repression of Nanog Gene Transcription by Tcf3 Limits Embryonic Stem Cell Self-Renewal. Mol. Cell. Biol. *26*, 7479–7491.

Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W.M., Solovei, I., Brugman, W., Gräf, S., Flicek, P., Kerkhoven, R.M., Lohuizen, M. van, et al. (2010). Molecular Maps of the Reorganization of Genome-Nuclear Lamina Interactions during Differentiation. Mol. Cell *38*, 603–613.

Phair, R.D., and Misteli, T. (2001). Kinetic modelling approaches to in vivo imaging. Nat. Rev. Mol. Cell Biol. *2*, 898–907.

Phillips-Cremins, J.E., Sauria, M.E.G., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S.K., Ong, C.-T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. Cell *153*, 1281–1295.

Pienta, K.J., and Coffey, D.S. (1985). The Nuclear Matrix: an Organizing Structure for the Interphase Nucleus and Chromosome. In Structure and Function of the Genetic Apparatus, (Springer, Boston, MA), pp. 83–98.

Plath, K., Mlynarczyk-Evans, S., Nusinow, D.A., and Panning, B. (2002). Xist RNA and the mechanism of X chromosome inactivation. Annu. Rev. Genet. *36*, 233–278.

Probst, A.V., Okamoto, I., Casanova, M., El Marjou, F., Le Baccon, P., and Almouzni, G. (2010). A Strand-Specific Burst in Transcription of Pericentric Satellites Is Required for Chromocenter Formation and Early Mouse Development. Dev. Cell *19*, 625–638.

Pueschel, R., Coraggio, F., and Meister, P. (2016). From single genes to entire genomes: the search for a function of nuclear organization. Development *143*, 910–923.

Quinn, J.J., and Chang, H.Y. (2016). Unique features of long non-coding RNA biogenesis and function. Nat. Rev. Genet. *17*, 47–62.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell *159*, 1665–1680.

Rao, S.S.P., Huang, S.-C., Glenn St Hilaire, B., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.-R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D., et al. (2017). Cohesin Loss Eliminates All Loop Domains. Cell *171*, 305-320.e24.

Razin, S.V., Chernokhvostov, V.V., Roodyn, A.V., Zbarsky, I.B., and Georgiev, G.P. (1981). Proteins tightly bound to DNA in the regions of DNA attachment to the skeletal structures of interphase nuclei and metaphase chromosomes. Cell *27*, 65–73.

Rinn, J., and Guttman, M. (2014). RNA and dynamic nuclear organization. Science *345*, 1240–1241.

Rinn, J.L., and Chang, H.Y. (2012). Genome Regulation by Long Noncoding RNAs. Annu. Rev. Biochem. *81*, 145–166.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., et al. (2007). Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. Cell *129*, 1311–1323.

Romig, H., Fackelmayer, F.O., Renz, A., Ramsperger, U., and Richter, A. (1992). Characterization of SAF-A, a novel nuclear DNA binding protein from HeLa cells with high affinity for nuclear matrix/scaffold attachment DNA elements. EMBO J. *11*, 3431–3440.

Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A.V., Manteiga, J.C., et al. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. Science eaar3958.

Santa, F.D., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.-L., and Natoli, G. (2010). A Large Fraction of Extragenic RNA Pol II Transcription Sites Overlap Enhancers. PLOS Biol. *8*, e1000384.

Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C.T., Schreiber, S.L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature *419*, 407–411.

Sasaki, Y.T.F., Ideue, T., Sano, M., Mituyama, T., and Hirose, T. (2009). MENε/β noncoding RNAs are essential for structural integrity of nuclear paraspeckles. Proc. Natl. Acad. Sci. *106*, 2525–2530.

Sato, N., Meijer, L., Skaltsounis, L., Greengard, P., and Brivanlou, A.H. (2004). Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. Nat. Med. *10*, 55–63.

Sawyer, I.A., Bartek, J., and Dundr, M. (2018). Phase separated microenvironments inside the cell nucleus are linked to disease and regulate epigenetic state, transcription and RNA processing. Semin. Cell Dev. Biol.

Schulz, E.G., Meisig, J., Nakamura, T., Okamoto, I., Sieber, A., Picard, C., Borensztein, M., Saitou, M., Blüthgen, N., and Heard, E. (2014). The Two Active X Chromosomes in Female ESCs Block Exit from the Pluripotent State by Modulating the ESC Signaling Network. Cell Stem Cell *14*, 203–216.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. Cell *148*, 458–472.

Sharova, L.V., Sharov, A.A., Piao, Y., Shaik, N., Sullivan, T., Stewart, C.L., Hogan, B.L.M., and Ko, M.S.H. (2007). Global gene expression profiling reveals similarities and differences among mouse pluripotent stem cells of different origins and strains. Dev. Biol. *307*, 446–459.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V., et al. (2012). A map of the *cis*-regulatory sequences in the mouse genome. Nature *488*, 116–120.

Shy, B.R., Wu, C.-I., Khramtsova, G.F., Zhang, J.Y., Olopade, O.I., Goss, K.H., and Merrill, B.J. (2013). Regulation of Tcf7l1 DNA Binding and Protein Stability as Principal Mechanisms of Wnt/β-Catenin Signaling. Cell Rep. *4*, 1–9.

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.

Simon, M.D., Wang, C.I., Kharchenko, P.V., West, J.A., Chapman, B.A., Alekseyenko, A.A., Borowsky, M.L., Kuroda, M.I., and Kingston, R.E. (2011). The genomic binding sites of a noncoding RNA. Proc. Natl. Acad. Sci. *108*, 20497–20502.

Smith, A. (2017). Formative pluripotency: the executive phase in a developmental continuum. Development *144*, 365–373.

Smith, A.G., Heath, J.K., Donaldson, D.D., Wong, G.G., Moreau, J., Stahl, M., and Rogers, D. (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. Nature *336*, 688–690.

Sofueva, S., Yaffe, E., Chan, W.-C., Georgopoulou, D., Vietri Rudan, M., Mira-Bontenbal, H., Pollard, S.M., Schroth, G.P., Tanay, A., and Hadjur, S. (2013). Cohesin-mediated interactions organize chromosomal domain architecture. EMBO J. *32*, 3119–3129.

Sone, M., Hayashi, T., Tarui, H., Agata, K., Takeichi, M., and Nakagawa, S. (2007a). The mRNA-like noncoding RNA Gomafu constitutes a novel nuclear domain in a subset of neurons. J Cell Sci *120*, 2498–2506.

Sone, M., Hayashi, T., Tarui, H., Agata, K., Takeichi, M., and Nakagawa, S. (2007b). The mRNA-like noncoding RNA Gomafu constitutes a novel nuclear domain in a subset of neurons. J. Cell Sci. *120*, 2498–2506.

Spector, D.L., and Lamond, A.I. (2011). Nuclear speckles. Cold Spring Harb. Perspect. Biol. *3*.

Stavridis, M.P., Lunn, J.S., Collins, B.J., and Storey, K.G. (2007). A discrete period of FGF-induced Erk1/2 signalling is required for vertebrate neural specification. Development *134*, 2889–2894.

van Steensel, B., and Henikoff, S. (2000). Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. Nat. Biotechnol. *18*, 424–428.

Strübing, C., Ahnert-Hilger, G., Shan, J., Wiedenmann, B., Hescheler, J., and Wobus, A.M. (1995). Differentiation of pluripotent embryonic stem cells into the neuronal lineage in vitro gives rise to mature inhibitory and excitatory neurons. Mech. Dev. *53*, 275–287.

Sultan, M., Schulz, M.H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Seifert, M., Borodina, T., Soldatov, A., Parkhomchuk, D., et al. (2008). A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science *321*, 956–960.

Sundaram, V., Choudhary, M.N.K., Pehrsson, E., Xing, X., Fiore, C., Pandey, M., Maricque, B., Udawatta, M., Ngo, D., Chen, Y., et al. (2017). Functional cis-regulatory modules encoded by mouse-specific endogenous retrovirus. Nat. Commun. *8*.

Sunwoo, H., Dinger, M.E., Wilusz, J.E., Amaral, P.P., Mattick, J.S., and Spector, D.L. (2009). MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. Genome Res. *19*, 347–359.

Suter, D.M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., and Naef, F. (2011). Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics. Science *332*, 472–474.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Takizawa, T., Gudla, P.R., Guo, L., Lockett, S., and Misteli, T. (2008). Allele-specific nuclear positioning of the monoallelically expressed astrocyte marker GFAP. Genes Dev. *22*, 489–498.

Tan, J.Y., Biasini, A., Young, R.S., and Marques, A. (2018). An unexpected contribution of lincRNA splicing to enhancer function. BioRxiv 287706.

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.

Tang, Z., Luo, O.J., Li, X., Zheng, M., Zhu, J.J., Szalaj, P., Trzaskoma, P., Magalska, A., Wlodarczyk, J., Ruszczycki, B., et al. (2015). CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. Cell *163*, 1611–1627.

Tani, H., Mizutani, R., Salam, K.A., Tano, K., Ijiri, K., Wakamatsu, A., Isogai, T., Suzuki, Y., and Akimitsu, N. (2012). Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. Genome Res. *22*, 947–956.

Tark-Dame, M., Jerabek, H., Manders, E.M.M., van der Wateren, I.M., Heermann, D.W., and van Driel, R. (2014). Depletion of the chromatin looping proteins CTCF and cohesin causes chromatin compaction: insight into chromatin folding by polymer modelling. PLoS Comput. Biol. *10*, e1003877.

Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D.G. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature *448*, 196–199.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science *282*, 1145–1147.

Tichon, A., Gil, N., Lubelsky, Y., Havkin Solomon, T., Lemze, D., Itzkovitz, S., Stern-Ginossar, N., and Ulitsky, I. (2016). A conserved abundant cytoplasmic long noncoding RNA modulates repression by Pumilio proteins in human cells. Nat. Commun. *7*, 12209.

Tosolini, M., Brochard, V., Adenot, P., Chebrout, M., Grillo, G., Navia, V., Beaujean, N., Francastel, C., Bonnet-Garnier, A., and Jouneau, A. (2018). Contrasting epigenetic states of heterochromatin in the different types of mouse pluripotent stem cells. Sci. Rep. *8*, 5776.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinforma. Oxf. Engl. *25*, 1105–1111.

Ulitsky, I. (2016). Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. Nat. Rev. Genet. *17*, 601–614.

Ulitsky, I., and Bartel, D.P. (2013a). lincRNAs: genomics, evolution, and mechanisms. Cell *154*, 26–46.

Ulitsky, I., and Bartel, D.P. (2013b). lincRNAs: Genomics, Evolution, and Mechanisms. Cell *154*, 26–46.

Ulitsky, I., Shkumatava, A., Jan, C.H., Sive, H., and Bartel, D.P. (2011). Conserved Function of lincRNAs in Vertebrate Embryonic Development despite Rapid Sequence Evolution. Cell *147*, 1537–1550.

Velazquez Camacho, O., Galan, C., Swist-Rosowska, K., Ching, R., Gamalinda, M., Karabiber, F., De La Rosa-Velazquez, I., Engist, B., Koschorz, B., Shukeir, N., et al. (2017). Major satellite repeat RNA stabilize heterochromatin retention of Suv39h enzymes by RNA-nucleosome association and RNA:DNA hybrid formation. ELife *6*.

Vizlin-Hodzic, D., Johansson, H., Ryme, J., Simonsson, T., and Simonsson, S. (2011). SAF-A Has a Role in Transcriptional Regulation of Oct4 in ES Cells Through Promoter Binding. Cell. Reprogramming *13*, 13–27.

Walter, J., Schermelleh, L., Cremer, M., Tashiro, S., and Cremer, T. (2003). Chromosome order in HeLa cells changes during mitosis and early G1, but is stably maintained during subsequent interphase stages. J. Cell Biol. *160*, 685–697.

Wang, K.C., Yang, Y.W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B.R., Protacio, A., Flynn, R.A., Gupta, R.A., et al. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature *472*, 120–124.

Wang, L., Park, H.J., Dasari, S., Wang, S., Kocher, J.-P., and Li, W. (2013a). CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res. *41*, e74.

Wang, Y., Xu, Z., Jiang, J., Xu, C., Kang, J., Xiao, L., Wu, M., Xiong, J., Guo, X., and Liu, H. (2013b). Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. Dev. Cell *25*, 69–80.

Washietl, S., Hofacker, I.L., Lukasser, M., Hüttenhofer, A., and Stadler, P.F. (2005). Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. Nat. Biotechnol. *23*, 1383–1390.

Wettstein, R., Bodak, M., and Ciaudo, C. (2016). Generation of a Knockout Mouse Embryonic Stem Cell Line Using a Paired CRISPR/Cas9 Genome Engineering Tool. In Embryonic Stem Cell Protocols, K. Turksen, ed. (New York, NY: Springer New York), pp. 321–343.

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.

Wijchers, P.J., Geeven, G., Eyres, M., Bergsma, A.J., Janssen, M., Verstegen, M., Zhu, Y., Schell, Y., Vermeulen, C., de Wit, E., et al. (2015). Characterization and dynamics of pericentromere-associated domains in mice. Genome Res. *25*, 958–969.

Williams, R.L., Hilton, D.J., Pease, S., Willson, T.A., Stewart, C.L., Gearing, D.P., Wagner, E.F., Metcalf, D., Nicola, N.A., and Gough, N.M. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. Nature *336*, 684–687.

Williams, R.R.E., Azuara, V., Perry, P., Sauer, S., Dvorkina, M., Jørgensen, H., Roix, J., McQueen, P., Misteli, T., Merkenschlager, M., et al. (2006). Neural induction promotes large-scale chromatin reorganisation of the Mash1 locus. J. Cell Sci. *119*, 132–140.

Williamson, C.M., Ball, S.T., Dawson, C., Mehta, S., Beechey, C.V., Fray, M., Teboul, L., Dear, T.N., Kelsey, G., and Peters, J. (2011). Uncoupling antisense-mediated silencing and DNA methylation in the imprinted Gnas cluster. PLoS Genet. *7*, e1001347.

de Wit, E., Vos, E.S.M., Holwerda, S.J.B., Valdes-Quezada, C., Verstegen, M.J.A.M., Teunissen, H., Splinter, E., Wijchers, P.J., Krijger, P.H.L., and de Laat, W. (2015). CTCF Binding Polarity Determines Chromatin Looping. Mol. Cell *60*, 676–684.

Wray, J., Kalkan, T., and Smith, A.G. (2010). The ground state of pluripotency. Biochem. Soc. Trans. *38*, 1027–1032.

Wu, C.-I., Hoffman, J.A., Shy, B.R., Ford, E.M., Fuchs, E., Nguyen, H., and Merrill, B.J. (2012). Function of Wnt/β-catenin in counteracting Tcf3 repression through the Tcf3-β-catenin interaction. Dev. Camb. Engl. *139*, 2118–2129.

Yang, F., Deng, X., Ma, W., Berletch, J.B., Rabaia, N., Wei, G., Moore, J.M., Filippova, G.N., Xu, J., Liu, Y., et al. (2015). The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. Genome Biol. *16*, 52.

Yang, L., Lin, C., Jin, C., Yang, J.C., Tanasa, B., Li, W., Merkurjev, D., Ohgi, K.A., Meng, D., Zhang, J., et al. (2013). lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. Nature *500*, 598–602.

Yassour, M., Kaplan, T., Fraser, H.B., Levin, J.Z., Pfiffner, J., Adiconis, X., Schroth, G., Luo, S., Khrebtukova, I., Gnirke, A., et al. (2009). Ab initio construction of a eukaryotic transcriptome by massively parallel mRNA sequencing. Proc. Natl. Acad. Sci. U. S. A. *106*, 3264–3269.

Yin, Y., Yan, P., Lu, J., Song, G., Zhu, Y., Li, Z., Zhao, Y., Shen, B., Huang, X., Zhu, H., et al. (2015). Opposing Roles for the lncRNA Haunt and Its Genomic Locus in Regulating HOXA Gene Activation during Embryonic Stem Cell Differentiation. Cell Stem Cell *16*, 504–516.

Ying, Q.L., Nichols, J., Chambers, I., and Smith, A. (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell *115*, 281–292.

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.

Yu, P., Xiao, S., Xin, X., Song, C.-X., Huang, W., McDee, D., Tanaka, T., Wang, T., He, C., and Zhong, S. (2013). Spatiotemporal clustering of the epigenome reveals rules of dynamic gene regulation. Genome Res. *23*, 352–364.

Zappulla, D.C., and Cech, T.R. (2006). RNA as a flexible scaffold for proteins: yeast telomerase and beyond. Cold Spring Harb. Symp. Quant. Biol. *71*, 217–224.

Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C.G., et al. (2018). Ensembl 2018. Nucleic Acids Res. *46*, D754–D761.

Zhang, B., Arun, G., Mao, Y.S., Lazar, Z., Hung, G., Bhattacharjee, G., Xiao, X., Booth, C.J., Wu, J., Zhang, C., et al. (2012). The lncRNA Malat1 Is Dispensable for Mouse Development but Its Transcription Plays a cis-Regulatory Role in the Adult. Cell Rep. *2*, 111–123.

Zhang, J., Gao, Y., Yu, M., Wu, H., Ai, Z., Wu, Y., Liu, H., Du, J., Guo, Z., and Zhang, Y. (2015). Retinoic Acid Induces Embryonic Stem Cell Differentiation by Altering Both Encoding RNA and microRNA Expression. PLoS ONE *10*.

Zhang, X., Rice, K., Wang, Y., Chen, W., Zhong, Y., Nakayama, Y., Zhou, Y., and Klibanski, A. (2010). Maternally expressed gene 3 (MEG3) noncoding ribonucleic acid: isoform structure, expression, and functions. Endocrinology *151*, 939–947.

Zhao, S., Nichols, J., Smith, A.G., and Li, M. (2004). SoxB transcription factors specify neuroectodermal lineage choice in ES cells. Mol. Cell. Neurosci. *27*, 332–342.

Zhao, Y., Li, H., Fang, S., Kang, Y., wu, W., Hao, Y., Li, Z., Bu, D., Sun, N., Zhang, M.Q., et al. (2016). NONCODE 2016: an informative and valuable data source of long non-coding RNAs. Nucleic Acids Res. *44*, D203–D208.

Zhao, Z., Tavoosidana, G., Sjölinder, M., Göndör, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K.S., Singh, U., et al. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat. Genet. *38*, 1341–1347.

Zheng, L.-L., Li, J.-H., Wu, J., Sun, W.-J., Liu, S., Wang, Z.-L., Zhou, H., Yang, J.-H., and Qu, L.-H. (2016). deepBase v2.0: identification, expression, evolution and function of small RNAs, LncRNAs and circular RNAs from deep-sequencing data. Nucleic Acids Res. *44*, D196-202.

Zheng, X., Hu, J., Yue, S., Kristiani, L., Kim, M., Sauria, M., Taylor, J., Kim, Y., and Zheng, Y. (2018). Lamins Organize the Global Three-Dimensional Genome from the Nuclear Periphery. Mol. Cell *71*, 802-815.e7.