From Department of Medical Biochemistry and Biophysics
Karolinska Institutet, Stockholm, Sweden

QUANTITATIVE METHODS FOR PROFILING DYNAMIC CHROMATIN FEATURES

Banushree Kumar

Karolinska Institutet

Stockholm 2022

Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2022

ISBN 978-91-8016-576-1
Cover illustration: created with BioRender.com

Quantitative methods for profiling dynamic chromatin features

THESIS FOR DOCTORAL DEGREE (Ph.D.)

To be defended in public at **Lecture Hall CMB,**
Karolinska Institutet, Berzelius väg 21, Solna

On Friday **June 03rd,2022** at **09:30** AM.

By

Banushree Kumar

Principal Supervisor:
Associate Prof. Simon Elsässer
Karolinska Institutet
Department of Medical Biochemistry and Biophysics
Division of Genome Biology

Co-supervisor(s):
Assistant Prof. Fredrik Lanner
Karolinska Institutet
Department of Clinical Science, Intervention and Technology
Division of Obstetrics and Gynecology

Prof. Gonçalo Castelo-Branco
Karolinska Institutet
Department of Medical Biochemistry and Biophysics
Division of Molecular Neurobiology

Assistant Prof. Laura Baranello
Karolinska Institutet
Department of Cell and Molecular Biology

Opponent:
Dr. Aydan Bulut-Karslioglu
Max Planck Institute for Molecular Genetics
Department of Genome Regulation

Examination Board:
Prof. Mattias Mannervik
Stockholm University
Department of Molecular Biosciences

Prof. Joakim Lundeberg
KTH Royal Institute of Technology
Division of Gene Technology

Prof. Ola Söderberg
Uppsala University
Department of Pharmaceutical Biosciences

"Nothing whets the intelligence more than a passionate suspicion; nothing develops all the faculties of an immature mind more than a trail running away into the dark."

-Stefan Zweig

-To patti

ABSTRACT

Living systems, from entire organisms down to the single cells constituting them are dynamic entities that continuously adapt and respond to their local environment. Cells achieve this through gene expression programs derived from static information encoded in the DNA made dynamic through chemical modifications at the chromatin level, collectively termed the epigenome. Numerous epigenetic regulators have been implicated in early embryonic developmental transitions and pluripotency. *Ex vivo*, the different states of pluripotency can be recapitulated by embryonic stem cells (ESCs) grown in defined media conditions. Many developmental gene promoters in ESCs display co-occurrence of the activating histone H3 lysine 4 trimethylation (H3K4me3) mark and the repressive H3K27me3 mark. This distinctive 'bivalent' signature is considered to poise expression, allowing timely resolution to an active or inactive state depending on the signal.

The distribution of histone modifications and chromatin-associated factors across the genome can be mapped using chromatin immunoprecipitation followed by next-generation sequencing (ChIP-seq). However, traditional ChIP-seq methods fail to quantitatively profile the nuanced global and local epigenetic rewiring that takes place in key developmental stages. This thesis addresses this limitation through the development of a quantitative multiplexed ChIP-seq technology: MINUTE (multiplexed indexed unique molecule T7 amplification end to end sequencing) ChIP. Across the three papers included in this thesis, we reveal the underpinnings of chromatin state dynamics in early mouse and human embryonic development by employing MINUTE ChIP.

In **Paper I**, we first show that MINUTE ChIP enables accurate quantitative comparisons over a wide linear range. By employing it to characterize mouse ESCs grown in 2i and serum conditions, we find that the 2i naïve state is characterized by high global levels of H3K27me3 and low H3K4me3. At bivalent promoters, we observe that while H3K27me3 levels are stably maintained between serum and 2i, H3K4me3 levels are higher in the serum condition.
Through quantitative epigenome profiling, in **Paper II** we find that naïve human ESCs also have broad global gain of Polycomb repressive complex 2 (PRC2)-mediated H3K27me3 and define a previously unrecognized, naïve-specific set of bivalent promoters. Bulk and single-cell transcriptomics confirmed that naïve bivalency maintains key trophectoderm and mesoderm transcription factors in a transcriptionally poised state which is resolved to an active state upon depletion of H3K27me3. Therefore, we discovered that PRC2-mediated repression provides a highly adaptive mechanism to restrict lineage potential during early human development.
In **paper III** we show how quantitative RNA polymerase II occupancy profiles generated by MINUTE ChIP can be integrated with transient transcriptomics data to unravel genome wide transcriptional kinetics in three mESCs pluripotent states: naïve, ground and paused.

Taken together, this thesis provides compelling evidence for a broad H3K27me3 hypermethylation of the genome in both naïve mouse and human ESCs and the basis for substantially revising the model for bivalency during embryonic development.

LIST OF SCIENTIFIC PAPERS

I. *Banushree Kumar, Simon J Elsässer*
Quantitative multiplexed ChIP reveals global alterations that shape promoter bivalency in ground state embryonic stem cells.
Cell Reports 2019, Vol. 28, 3274-3284.e5

II. *Banushree Kumar*, Carmen Navarro*, Nerges Winblad*, John P Schell, Cheng Zhao, Jere Weltner, Laura Baqué-Vidal, Angelo Salazar Mantero, Sophie Petropoulos, Fredrik Lanner, Simon J Elsässer*
Polycomb Repressive Complex 2 shields naïve human pluripotent cells from trophectoderm differentiation.
Manuscript accepted at Nature Cell Biology 2022

III. *Rui Shao, Banushree Kumar, Katja Lidschreiber, Michael Lidschreiber, Patrick Cramer, Simon J Elsässe*r
Distinct transcription kinetics of pluripotent cell states
Molecular Systems Biology 2022, Vol.18, e10407

CONTENTS

LIST OF ABBREVIATIONS

ChIC-seq	chromatin immunnocleavage sequencing
ChIP-seq	chromatin immunoprecipitation followed by sequencing
co-ChIP	combinatorial ChIP
COMPASS	complex proteins associated with SET1
CUT&TAG	cleavage under targets and tagmentation
DamID	DNA adenine methyltransferase identification
DNA	deoxyribonucleic acid
DNMT	DNA methyltransferase
EpiSCs	epiblast stem cells
ESCs	embryonic stem cells
EZH2	enhancer of Zeste 2
EED	embryonic ectoderm development
SUZ12	suppressor of Zeste 12
FBS	fetal bovine serum
FGF	fibroblast growth factor
GSK3	glycogen synthase kinase 3
H2Aub	histone H2A lysine 119 monoubiquitination
H3K27me3	histone H3 lysine 27 trimethylation
H3K4me3	histone H3 lysine 4 trimethylation
HATs	histone acetyltransferases
HDAC	histone deacetylase
hESCs	human embryonic stem cells
hPTMs	histone post-translational modifications
ICM	inner cell mass
IVT	*in vitro* transcription
JAK-STAT3	janus kinase–signal transducer and activator of transcription 3
lncRNA	long noncoding RNA
MEFs	mouse embryonic fibroblasts
LIF	leukemia inhibitory factor
MEK	mitogen-activated protein kinase kinase

mESCs	mouse embryonic stem cells
Mint-ChIP	multiplexed, indexed T7 ChIP
MINUTE ChIP	multiplexed indexed unique barcoded T7 paired-end sequencing ChIP
MLL	myeloid-lineage leukemia
MNase	Micrococcal nuclease
mNET-seq	mammalian native elongating transcript sequencing
mTOR	mechanistic target of rapamycin
ncRNA	non-coding RNA
PcG	Polycomb group
PCR	polymerase chain reaction
Pol II	RNA polymerase II
PRC	Polycomb repressive complex
RNA	ribonucleic acid
RNA-seq	RNA sequencing
RT-PCR	reverse transcription PCR
SAM	S-adenosylmethionine
scRNA-seq	single cell RNA sequencing
TADs	topologically associated domains
TE	trophectoderm
TET	ten-eleven translocation
TT-seq	transient transcriptome sequencing
UMI	unique molecular identifier
XCI	X chromosome inactivation
ZF	zinc finger
ZGA	zygotic genome activation
2i	medium with MEK and GSK3 inhibitors

1 INTRODUCTION

The developmental cascade best embodies the capacity of a cell to modulate its internal functions based on external cues. This cellular plasticity is a function of gene expression programmes that are rendered dynamic by a diverse set of chromatin features (nucleosome occupancy, positioning, composition, chemical modifications as well as DNA methylation) collectively termed the epigenome[1–4]. This complex layer of regulatory information is shaped by the action of transcription factors, chromatin remodellers, histone chaperones, DNA/histone-modifying enzymes and non-coding RNAs that operate over several orders of magnitude of space and time[5,6].

A visual metaphor, termed Waddington's landscape[7] intuitively captures this process: if a ball symbolizing the cell were to roll down a declining slope representing the course of development, then it would be canalized into specific trajectories or end states by the valleys of the surface (Figure 1A). A network of strings depicting gene products arising from pegs representing genes support this landscape and the contours are established by tensions in the strings representing epigenetic factors (Figure 1B). Following the different trajectories, cells give rise to new lineages, tissues and ultimately an entire multicellular organism. The more we unravel about this complicated journey, the better we understand profiles of physiological cellular states and are better equipped to apply this knowledge to pathological contexts.

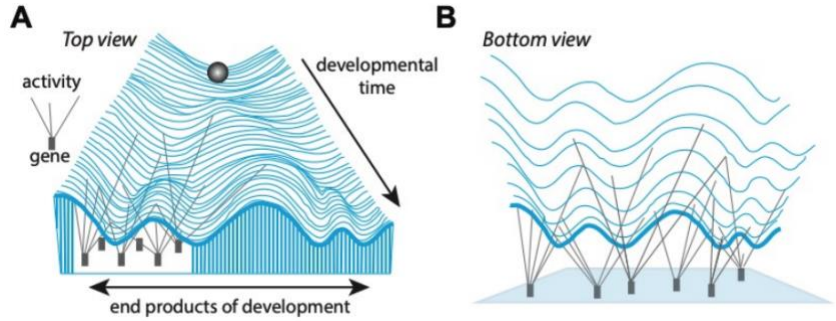

Fig.1: Waddington plot. Adapted from Elewa 2020.[8]

The introductory section of this thesis is structured to first traverse this landscape longitudinally: looking at cell types in early mouse and human embryonic development and later understand how the landscape is established and maintained, in particular by epigenetic factors. And finally, to discuss the method of choice for profiling chromatin marks (refers to histone post-translational modifications (hPTMs) in the context of this thesis).

1.1 EMBRYONIC DEVELOPMENT

Mammalian embryogenesis begins with the fertilization of the maternal (oocyte) and paternal (sperm) gametes to form the diploid zygote. Cells of the zygote then undergo rounds of cellular division to form the blastocyst, whose inner cell mass (ICM) gives rise to all the three germ layers (ectoderm, mesoderm, and endoderm) of the developing embryo, a

feature defined as pluripotency. The period from zygote to blastocyst formation can be broadly divided into pre-implantation and post-implantation (Figure 2); where implantation marks the attachment of the embryo to the maternal uterus. The first lineage segregation event is marked by the discrimination of the ICM from the trophectoderm (TE), the outer envelope of the pre-implantation blastocyst. Upon implantation, the TE gives rise to placental tissues while the ICM progresses via the epiblast stage to form the foetus[9,10].

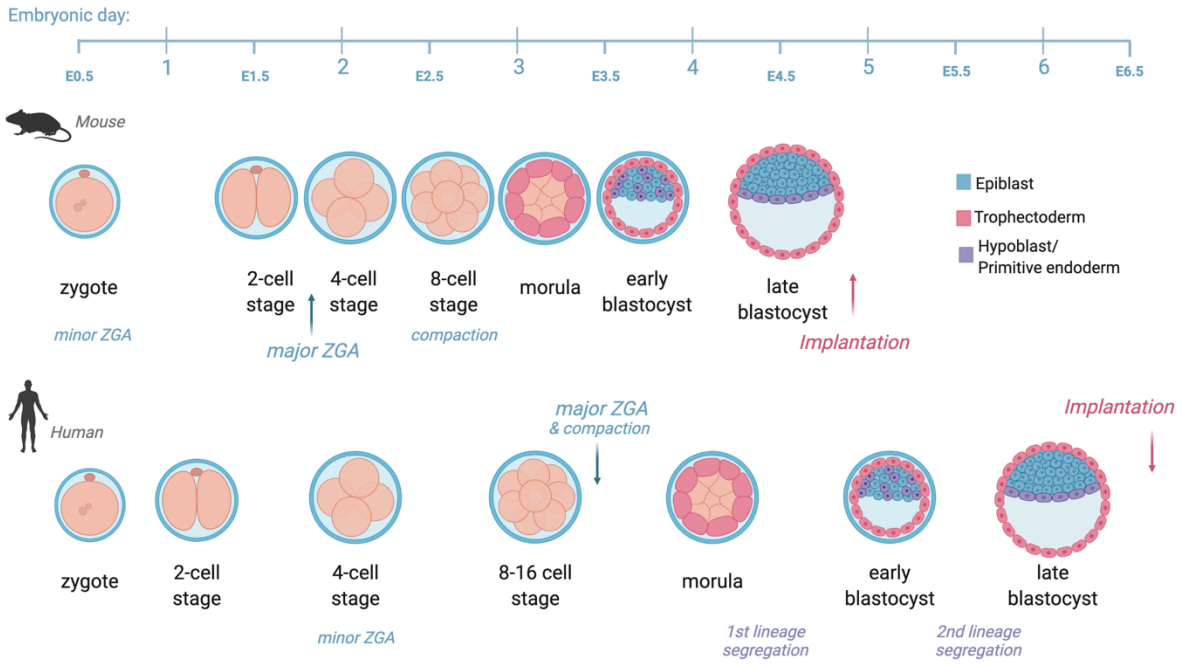

Fig.2: Early embryonic development (fom zygote to late blastocyst) in mouse vs human (created with BioRender.com based on Molè et al. 2020[11]). Mouse embryogenesis starts with formation of the zygote at fertilization, followed by cleavage divisions. The major wave of zygotic genome activation (ZGA) occurs at the two-cell stage. Compaction and polarization take place at the eight-cell stage. The first lineage segregation between outer cells (future trophectoderm) and inner cell mass (epiblast and primitive endoderm precursors) occurs at embryonic day 3.5. At embryonic day 3.75-4.0, the mouse embryo matures into a blastocyst. The second lineage segregation between epiblast and primitive endoderm is fully completed before initiation of implantation (embryonic day 4.75). Human embryogenesis starts with formation of the zygote at fertilization, followed by a series of cleavage divisions. The major wave of ZGA occurs at the eight-cell stage (embryonic day 3) alongside initiation of compaction and polarization. At embryonic day 4.5-5.0, the outer cells, precursors of the future trophectoderm segregate from the inner cell mass, precursors of the epiblast and hypoblast lineages, to form a blastocyst[11].

Morphological similarities up until the early blastocyst justified the initial use of mouse as model organism to study mammalian development. However, comparative studies in the later years have revealed that there are crucial differences in the embryogenesis of mice and humans. These include the timing and components of zygotic genome activation (ZGA), the timing and process of X chromosome inactivation, the timing and molecular pathways utilized for blastocyst lineage specification, and the presence or absence of a mechanism to promote embryo diapause, to name a few[11,12].

Deconstructing the above development events and elucidating the mechanisms governing them has been made possible due to *in vitro* culture systems for both mice and humans[13].

1.2 EMBRYONIC STEM CELLS

Embryonic stem cells (ESCs) can be derived from different stages of early embryonic development (most often from the ICM of the pre-implantation blastocysts) and maintained indefinitely *ex vivo* by supplementing exogenous factors. The salient characteristics of ESCs are multilineage differentiation potential or pluripotency, capacity of self-renewal and primary chimera formation. However, it is important to stress that self-renewal is only a transient feature during early development *in utero*.

Based on molecular characteristics, ESCs can be further classified into **naïve pluripotent**: a state that resembles the pre-implantation embryonic configurations or **primed pluripotent**: a state that resembles the post-implantation embryonic configurations, which have a more restricted lineage potential. These pluripotent state identities are primarily dictated by the derivation growth conditions rather than the source of the cells and the following section provides a chronological overview of the development of culture conditions for mouse cells, followed by human cells.

1.2.1 Growth condition for Mouse ESCs

Naïve Mouse embryonic stem cells (mESCs) were first derived from the ICM of the 129 mouse strain and cultured using mitotically inactive mouse embryonic fibroblasts (MEFs) as feeder cells and fetal bovine serum (FBS)[14,15]. Subsequently, leukaemia inhibitory factor (LIF) was identified as a key ingredient to maintain feeder free mESCs proliferation *in vitro*[16,17] by activating the Janus kinase–signal transducer and activator of transcription 3 (JAK-STAT3) pathway. Though this serum/LIF culture condition allowed cells to conserve their hallmark pre-implantation or naïve pluripotency features, subpopulations still underwent spontaneous differentiation deeming this state to be metastable. Stability and derivation efficiency of mESCs was further improved by serum- and feeder-free defined media containing two inhibitors for MEK (mitogen-activated protein kinase kinase) and GSK3 (glycogen synthase kinase 3) signaling, along with LIF (called 2i/LIF)[18–20]. While both serum/LIF and 2i/LIF states are considered 'naïve', the states do exhibit substantial differences in metabolism, transcriptome and epigenome (Figure 3). The development of the 3i condition (inhibitors for MEK, GSK3 and FGF (fibroblast growth factor)) further delineated the minimal requirements for naïve mESCs self-renewal *ex vivo*[21].

Primed In parallel, cells derived from the post-implantation epiblast of mice were cultured in media containing recombinant FGF2 and Activin A cytokines to recapitulate primed pluripotency and were termed epiblast stem cells (EpiSCs)[22,23]. Thereafter, alternative growth conditions including GSK3i/IWR1 (small-molecule tankyrase inhibitor)[24] or FGF2/IWR1[25] have been described to generate EpiSCs. Notably, studies have indicated that cells grown in these conditions represent different *in vivo* stages of post-implantation embryonic development[25–27].

Paused In nature, mice display a facultative reproductive strategy to combat unfavourable conditions termed diapause: a reversible suspended state of embryo development featuring proliferation arrest, decreased transcription and protein synthesis[28]. A study showed that the

deletion or pharmacological inhibition of the transcription factor *Myc* in the 2i/LIF grown mESCs could mimic this paused state *in vitro*[29]. Soon after it was reported that mESCs grown in either serum/LIF or 2i/LIF along with an inhibitor of mTOR (mechanistic target of rapamycin) recapitulated the transcriptome of the diapaused epiblast more consistently[30].

Collectively, evolution of culture conditions spanning three decades has provided a platform for the precise description and dissection of mouse pluripotent states and paved the way to decipher human pluripotency.

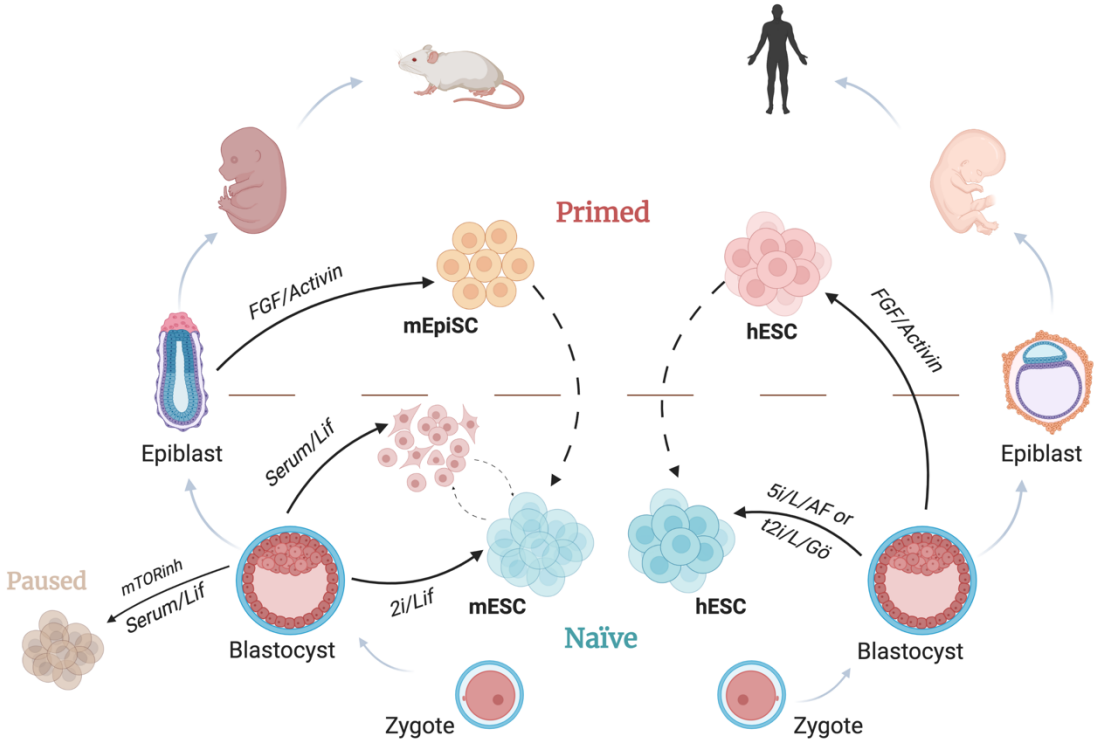

Fig.3: Naïve vs primed pluripotency in mouse and human (created with BioRender.com based on Weinberger et al. 2016[13]). ESCs can be shuttled between the above depicted pluripotent states by transitioning them into the appropriate media conditions.

1.2.2 Growth conditions for Human ESCs

Primed Human embryonic stem cells (hESCs) were first derived from the pre-implantation blastocysts and stabilized *in vitro* in media containing FGF2 and Activin A[31]. Comparative analyses with the relatively well-characterized mESC systems suggested that conventional hESCs resembled the murine EpiSCs and its features clustered closer to the post-implantation epiblast, emanating the name: primed hESCs[32] (Figure 3).

Naïve Drawing inspiration from the interconvertibility of pluripotent states in mESCs, numerous transgenic and chemical approaches to obtain naïve hESCs from pre-existing primed hESC lines have been reported[13,33]. Initial studies reported that overexpression of transcription factors (*KLF4* and *OCT4* or *KLF2*) associated with naïve pluripotency in 2i/LIF conditions could induce naïve pluripotency in hESCs[34,35]. Thereafter, transgene-independent conditions such as naïve human stem cell medium (NSHM)[36] and 5i/L/AF (2i + three kinase inhibitors targeting BRAF, SRC, and ROCK together with LIF, Activin, and

FGF2)[37] also enabled the generation of naïve hESCs directly from primed hESCs. Concurrent with the development of 5i/L/AF, it was shown that overexpression of KLF2 and NANOG in primed cells followed by culture in titrated 2i/L conditions with ROCK and a protein kinase C (PKC) inhibitor (t2i/L/Gö) generated naïve hESCs[38]. Both these conditions also allowed for stabilizing naïve hESCs derived directly from the human ICM[39,40]. Most of the above culture conditions require MEF feeder for maintaining the naïve hESC colonies but feeder-free alternatives that facilitate primed-to-naïve resetting using histone deacetylase (HDAC) and WNT inhibitors have also emerged[41]. Lastly, it has been observed that extended passaging of hESCs in naïve growth conditions could alter their stemness, delay responsiveness to differentiation signals and make them susceptible to genomic instability[42,43].

The repertoire of *in vitro* systems outlined above cue the possibility of species-specific variations during embryo development but also support the contention that a few core principles of pluripotency may be conserved between rodents and primates. The following section highlights some key developmental identities of the naïve and primed states.

1.2.3 *Distinguishing features of the pluripotent states*

DNA methylation: Naïve ESCs retain global levels of hypomethylation in both promoters and gene bodies, as observed in the ICM, while methylation accumulates over the genome in primed ESCs[44,45]. Modes of deposition and function of this chromatin feature are discussed in more detail in the following section.

X chromosome inactivation (XCI) status: One X-chromosome in female mammals is inactivated during early embryogenesis, to allow for proper dosage compensation of X-linked genes relative to males[46]. The process of XCI in both mouse and humans is mediated epigenetically by the long noncoding RNA (lncRNA) *XIST*, however the timing of this event *in vivo* varies between the species[47–49]. Naïve ESCs exhibit transcriptional activity of both X-chromosomes as seen in the ICM, however in humans this is accompanied by the expression of *XIST*. Primed ESCs capturing later stages in development, display XCI with monoallelic gene expression (XaXi). Of note, upon prolonged culture, primed hESCs have been shown to display a phenomenon termed X chromosome erosion (Xe) wherein the previously silenced X chromosome (Xi) is reactivated[50].

Transposable elements: Transposable elements are mobile genetic entities that constitute significant chunks of the mammalian genome, and have served as essential drivers of evolution. The expression of these elements during embryonic development is tightly regulated by species-specific *trans*-acting epigenetic factors[51]. Upon ZGA, distinct families of transposons become active in humans (SVAs, LTR5_Hs[40,52]) and mice (LINES, MERV-L[53–55])[12]. Naïve ESCs of the respective species display the matched transposon expression profiles but not primed.

Metabolic pathways: Naïve ESCs display both anaerobic and aerobic metabolism implying that they use both glycolytic and oxidative phosphorylation for their energy production, while primed ESCs are almost exclusively glycolytic. The two states also have significantly

different tryptophan metabolism[38,56–59]. It has also been implicated that these metabolomic differences dictate cell fate by regulating enzymes that sculpt the epigenetic landscape[60]. Moreover, examining expression of *N-Myc* related metabolic genes and glucose consumption levels in human blastocysts suggest these changes actually resemble the *in vivo* situation[58].

Cell cycle alterations: Cells of the pre-implantation epiblast have the capacity to proliferate at unusually rapid rates. Naïve ESCs representing this stage display a reduced G1 phase (much shorter in mouse as compared to humans) allowing the cell to rapidly shuttle between cell division (M phase) and DNA synthesis (S phase)[61–64]. Primed ESCs in comparison spend a longer time in the G1 phase and it is speculated that this lengthening provides a window of opportunity for differentiation cues[65–67].

Dissecting such features of pluripotency, in addition to furthering our knowledge of developmental biology, may lead to improvements in techniques for *in vitro* fertilization and induced pluripotent stem cells based regenerative medicine. Furthermore, stem cells share similarities (such as DNA hypomethylation) to self-renewing cells in tumours, called cancer stem cells. Consequently, a better understanding of stem cell maintenance would pave the way to new therapeutic approaches for human malignancies.Thus, several genomic, transcriptomic, proteomic and epigenomic analyses have been conducted to further understand the molecular mechanisms governing distinct pluripotent states[13,40,43,68–70].

1.3 MAINTENANCE OF PLURIPOTENCY

The molecular and functional characteristics of a pluripotent state *in vivo* and *in vitro* represent the net outcome of its transcriptional program which is controlled by signaling pathways, transcription factors and the cell's epigenetic landscape.

1.3.1 Signaling pathways

The overarching theme of the development cascade is intercellular communication in the growing embryo via signaling pathways[71]. Key signaling pathways (FGF, Hedgehog (HH), WNT, TGFß, Notch to name a few) can elicit diverse cellular responses depending on the species, cell type, context, and developmental timing[72]. This highlights the pleiotropic nature of signaling pathways, wherein they may simultaneously have positive and negative effects on pluripotency.

The function of LIF via the STAT3 and FGF/ERK signaling pathways in maintaining pluripotency has been extensively studied and has contributed to the development of mESC culture conditions (discussed above) to mirror ICM conditions, where these signaling cascades are crucial for blastocyst development and activation[73]. In the mouse embryo, the HIPPO signaling pathway also contributes to blastocyst formation[74] however the involvement of this pathway in human ICM vs TE formation has not been confirmed. Similarly, evidence for the importance of FGF/ERK signaling for establishing cell fate in the human ICM is lacking[12,75].

Recently other signaling factors such as activin-NODAL, platelet-derived growth factor (PDGF), and TGFβ (relevant in primed mESCs) have been suggested to be involved in lineage development in human embryos as well[13,76–78]. Evolutionarily conserved signaling pathways such as HH, WNT and bone-morphogenetic-protein (BMP) have been shown to regulate embryonic tissue patterning during gastrulation and implantation[9,10]. Intriguing data also suggest that hESCs, but not mESCs, can be induced to differentiate into extraembryonic lineages such as trophectoderm cells by targeting the BMP pathway[79,80].

Therefore, as evident from the composition of growth conditions for ESCs, signaling pathways are critical to fine-tune the network for pluripotency maintenance and often converge on controlling the expression of transcription factors.

1.3.2 Transcription factors

Initial pursuits to uncover the molecular mechanisms that determine decisions of ESCs resulted in the widely accepted notion that transcription factors rule pluripotency via a hierarchical, interconnected network[81,82]. The "triumvirate" of *OCT4*, *SOX2* and *NANOG* function as master regulators of ESC identity by regulating their own expression as well as other factors throughout development[83]. The balance in levels has however been found to be delicate since either knockout or overexpression of any one factor prompts differentiation of ESCs to specific lineages. These observations also motivated a revision in the ESC model, now stating that: "pluripotency is inherently ephemeral or metastable" meaning that lineage specifiers in ESCs are vying for dominance among one another to instruct cell fate commitment[84]. Importantly, it has been proposed that extrinsic signals must be continually applied to sustain undifferentiated self-renewal and to ensure that no lineage specifying pluripotency factor becomes dominant.

Whole-embryo and single-cell transcriptome analyses[85–87] have defined development stage specific transcriptional programs. Of note, these studies reported that only 40% of ZGA genes are shared between humans and mouse[12]. Human ICM epiblast cells and naïve hESCs do not express genes such as *KLF2* and *ESRRB*, that are thought to be important naïve pluripotency factors in mice. Instead, *KLF17* is predicted to have a human-specific role in pre-implantation development. The temporal and spatial pattern of *CDX2* and *OCT4* expression is also distinct among the species. The first lineage segregation event in mice is accompanied by expression of *CDX2* in the TE while *OCT4* is restricted to the ICM[88]. In humans, specification of TE is independent of *CDX2* and *OCT4* is not restricted to the ICM until the late blastocyst stage[89]. Another example is *TFAP2C*, which serves as a key regulatory gene for naïve human pluripotency but plays a different role during mouse embryogenesis[90,91]. Single-cell studies have also been able to identify key transcription factors that are characteristic of the various lineages during later stages of development[10,92,93]. In the future, genome-wide approaches including CRISPR-based genetic screens and computational models will be instructive to expand our understanding of the transcription factor circuitry governing pluripotency[94].

1.3.3 Epigenetic features

The act of relaying external signals into the cell through transcription factors is made more versatile by modifications of epigenetic nature on the chromatin, meaning that they are reversible, yet inheritable. They impact DNA accessibility and in turn modulate gene expression by shaping the landscape of transcription factor binding sites. Technological advancements in methods such as ChIP-seq and RNA sequencing (RNA-seq), have allowed us to surf the waves of epigenetic currents through development; zygotic gene activation to blastocyst formation to post-implantation embryo (including primordial germ cells) and up to mature somatic cell types[95,96]. Results of *in vivo* and *in vitro* studies have emphasized the reversible, heritable, and functional nature of chromatin annotations and defined the following layers of epigenetic information:

Nuclear organization: The genome is organized into three-dimensional territories called topologically associated domains (TADs), which serve as the structural and functional units of chromatin. Pre-implantation embryos have a rather relaxed higher-order chromatin organization and TADs get stabilized by CTCF/cohesion complex in later stages of development[97]. The distinction of *OCT4* enhancer usage in naïve (distal enhancer) vs primed (proximal enhancer) ESCs, does suggest that differences in such long-range interactions might contribute to the local genome organization[43,98,99]. Moreover, epigenetic modifications (discussed in the following sections) in addition to influencing the local transcriptional climate, have been shown to reshape the 3D genome organization.

Non-coding RNA (ncRNA): Yet another level of controlling the gene expression epigenetically is with ncRNA. Both long (lncRNA) and short ncRNA (miRNA, siRNA, piRNA) have been found to have a role in the alteration of compaction level of chromatin, hence affecting the gene expression. As discussed earlier, the most extensively studied of these in the context of ESCs, is the lncRNA *XIST*, which mediates XCI by spreading along the X chromosome in cis and interacting with protein complexes that deposit repressive epigenetic marks.

DNA methylation: The chromatin is decorated with methylation of cytosine at CpG dinucleotides. This distribution is reversible and mosaic, with GC-rich regions at the 5'end of many promoters (CpG islands) being unmethylated. Its occurrence and removal are mediated by the DNA methyltransferase (DNMT) family that methylate cytosine and the Ten-eleven translocation (TET) family of enzymes orchestrating active DNA demethylation through the sequential oxidation of intermediates. The notable role of DNA methylation at promoters is gene silencing, mediated by methyl CpG binding domain (MBD)-containing transcriptional repressors[100]. However, certain methylated DNA motifs can also be recognized by transcription activators involved in cell-type transitions (ex: *KLF4, OCT4, HOBX13, NKX*) and facilitate gene expression from otherwise refractory chromatin domains[101,102]. CpG domains have also been shown to guide gene activity by interacting with histone modifying enzymes (discussed in following sections).

During early embryonic development there are dynamic alterations to the global DNA methylation patterns that are deemed crucial for imprinting[103] (Figure 4). A first wave

constitutes active demethylation of the paternal genome (quicker in humans as compared to mice) followed by further DNA hypomethylation of the zygotic genome. Bisulphite sequencing based maps have shown that replication during early divisions predominantly contributes to the dramatic DNA demethylation in the pre-implantation blastocyst[104,105]. Upon implantation, there is an increase in *de novo* methylation activity and the methylome of the developing fetus is (re-)established[104,105] alongside a second wave of demethylation of primordial germ cells (PGCs) that have entered the gonads. Once specified during development, the methylated CpG landscape is relatively static across somatic tissues and faithfully inherited through cell division.

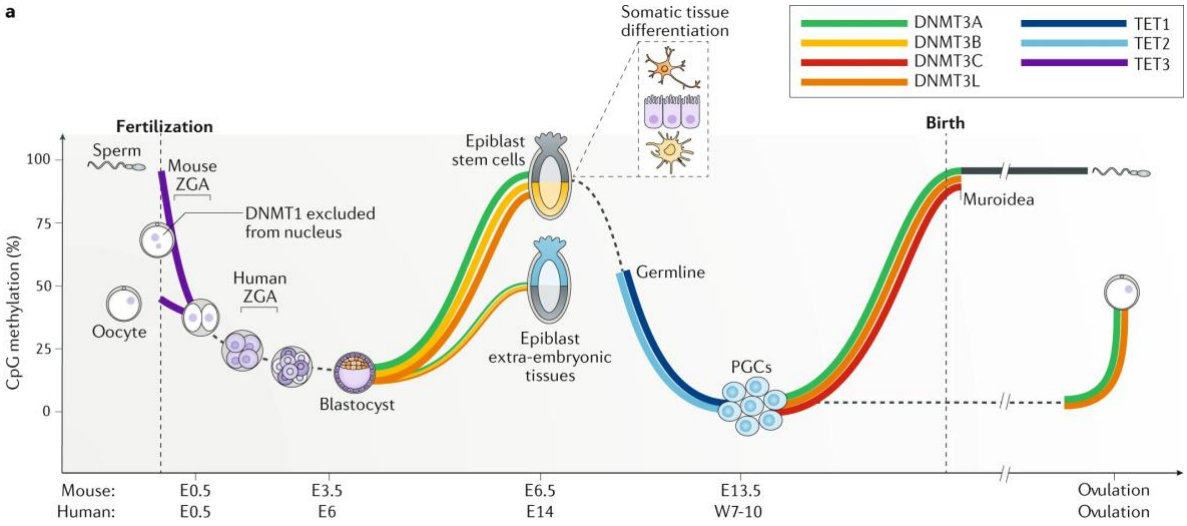

Fig.4: Reprogramming of DNA methylation during embryonic development. Adapted from Greenberg and Bourc'his 2019[106].

These reprogramming events are coupled to histone exchange and other chromatin regulation incidents and serve as a buffer to stabilize and regulate transcription factor-based decisions[106,107]. Naïve and primed cultures of ESCs effectively recapture the equivalent DNA methylation states *in vitro*. Importantly, via novel transgenic systems it has been shown that erasure of DNA methylation does not affect ESC self-renewal or its molecular signatures of pluripotency but causes defects during lineage commitment and differentiation[108,109].

Histone modifications: Nucleosomes, the basic unit of chromatin, are composed of histone octamers (two copies of core histone proteins H2A, H2B, H3 and H4), wrapped by about two turns (150 bp) of DNA[110]. Amino acid residues in the globular domain or N-terminal tails of each histone protein can be post-translationally modified. Specific enzymes, 'writers', deposit histone modifications of phosphorylation, acetylation, ubiquitylation and methylation, amongst others using metabolites such as ATP, acetyl-CoA, ubiquitin or S-adenosylmethionine (SAM), respectively[111]. The best studied of these writers are the Polycomb repressive complexes (PRC1 and PRC2) and the myeloid-lineage leukemia complexes (MLL/SET), which are described in more detail below. Removal of histone modifications is done by 'eraser' enzymes, while 'readers' bind to these modifications, serving as effector proteins. Readers are often a part of or recruit larger protein complexes with multiple subunits to bring about distinct downstream events[112] (Figure 5). Using biochemical and genetic studies, the functional role of these histone modifications in gene

expression is postulated in the "histone code" hypothesis: repressive and activating marks, with a caveat on heritability[113].

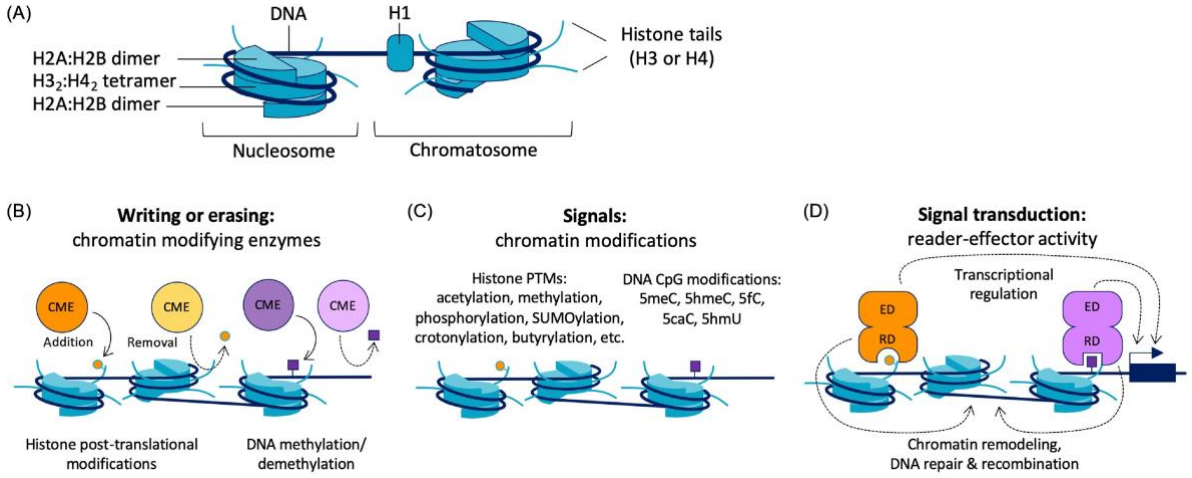

Fig.5: Relay of molecular information at the chromatin level by histone modifications. (a) nucleosome composition. (b) chromatin modifying enzymes (CMEs) write (add) or erase (remove) histone post-translational modifications (hPTMs). (c-d) chromatin modifications (signals) are read (bind) by protein complexes via a reader domain (RD) and drive processes through an effector domain (ED). Adapted from Franklin et al. 2022[112].

1.3.4 Polycomb Repressive complex (PRC): H3K27me3 & H2Aub

Polycomb proteins were initially discovered in *Drosophila melanogaster*, and associated with body plan specification by repression of homeotic (*Hox*) genes[114]. Polycomb genes were subsequently shown to have mammalian orthologues that have important roles in gene repression during development. Further, biochemical analyses have shown that Polycomb proteins assemble into two discrete multiprotein complexes (Polycomb repressive complex 1 (PRC1) and PRC2) that post-translationally modify specific histones[115].

Components

The Polycomb complexes harbor defined catalytic cores which then interact with auxiliary proteins to create distinct PRC1 (1 canonical/cPRC1 and 4 variant/vPRC1)[116,117] and PRC2 (.1 and .2) modules[118] (Figure 6). RING1B or its paralogue RING1A and one of six Polycomb group RING finger (PCGF1-6) proteins comprise the catalytic subunit of PRC1 that monoubiquitylates histone H2A at Lys119 (H2AK119ub1/H2Aub)[119,120]. The Polycomb repressive deubiquitinase (PR-DUB) complex catalyzes the deubiquitination[121]. Enhancer of Zeste 2 (EZH2) or its paralogue EZH1 along with embryonic ectoderm development (EED) and Suppressor of Zeste 12 (SUZ12) forms the catalytic lobe of PRC2[122] which is responsible for the mono-, di-, and tri-methylation of histone H3 at Lys27 (H3K27me1, H3K27me2, and H3K27me3 respectively). The histone demethylase JMJD3/UTX can remove the H3K27me3 mark[123].

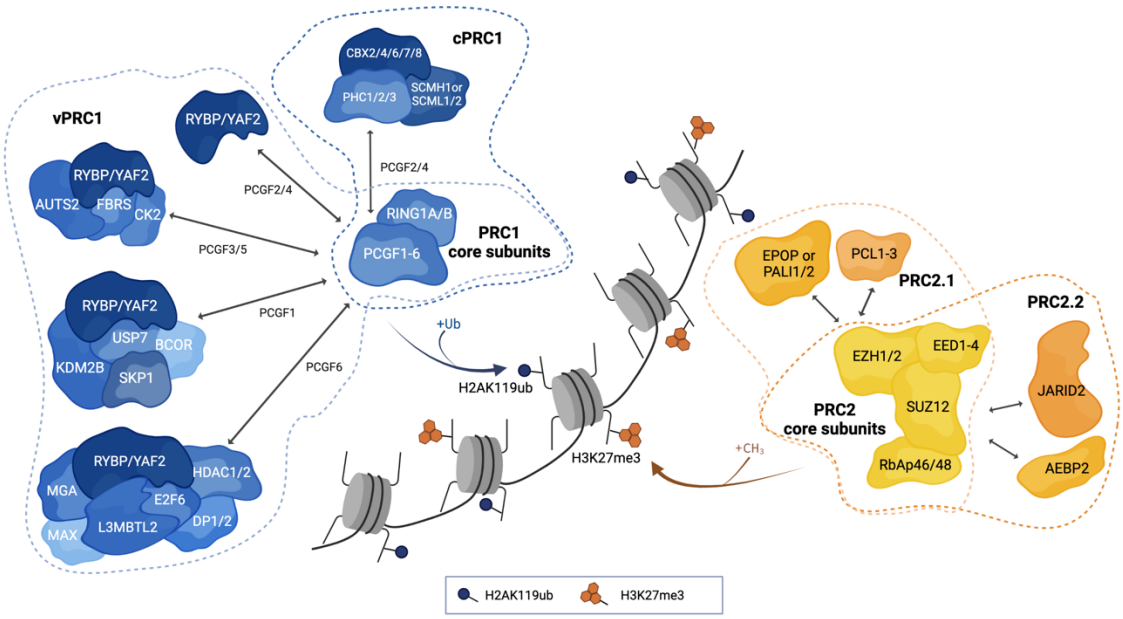

Fig.6: The Polycomb repressive complexes PRC1 and PRC2 (created with BioRender.com)

Occupancy

Spatially, PRC1 and PRC2 occupancy tends to converge on the same sites in the genome to form Polycomb chromatin domains, which are uniquely enriched in H2Aub and H3K27me3[124,125]. Targeting of Polycomb complexes to these sites in specific contexts has been attributed to one of the following mechanisms: sequence-specific DNA-binding factors (vPRC1 to E-box and T-box motifs); long noncoding RNAs associated to chromatin (by *XIST* on the inactive X chromosome); proteins that bind CpG islands[126]. Importantly, primary targeting of PRCs is followed by feedback mechanisms that enable the stabilization and spreading of Polycomb domains. Biochemical studies have shown that PRC1 and PRC2 have the capacity to recognize each other's histone modifications, thereby coupling their activities[127]. In contrast to the 'hierarchical' recruitment mechanism (*de novo* PRC2 function succeeded by PRC1 engagement and activity) described in *Drosophila*[128], the vertebrate Polycomb recruitment cascade begins with the binding of vPRC1 complexes to the chromatin and catalyzing the formation of H2Aub[129]. PRC2.2 is then recruited by the H2Aub modified nucleosome[130] and completes the Polycomb domain by depositing H3K27me3[120]. At more compact chromatin, PRC2 can also independently bind to regions and aids in the spreading of H3K27me3. cPRC1 complexes, albeit contributing minimally to H2Aub deposition, are stabilized on the chromatin by H3K27me3 and function in the 3D spatial organization of Polycomb domains in the nucleus[131–133].

Function

PRCs and their associated marks are commonly linked to transcriptional repression. However, growing evidence[134,135] suggests that their role in this context is responsive rather than instructive. PRCs, via their capacity to bind CpG islands at gene promoters, dynamically select potential target sites to establish Polycomb domains. Susceptibility of genes to this scanning is proposed to be controlled by their transcriptional activity, whereby at inactive genes the Polycomb system would function to restrict spurious transcription and maintain gene repression[136] (Figure 7). Neverthless, the exact mechanism(s) by which

PRCs and their associated marks affect transcription in different cell types and developmental stages are yet to be determined and is the focus of ongoing research in the field.

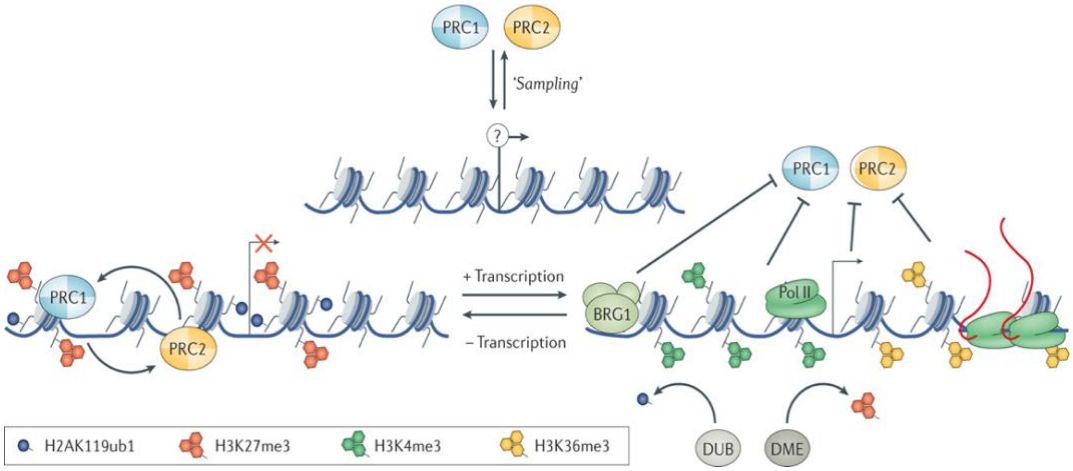

Fig.7: Model for Polycomb mediated gene repression. Adapted from Blackledge and Klose 2021[137]*. At lowly transcribed or untranscribed genes (left), these complexes respond to the minimal transcription by H2Aub and H3K27me3 to initiate the formation and spreading of Polycomb chromatin domains, which could help counteract inappropriate transcription and maintain an inactive chromatin state to protect cell identity. However, at expressed genes (right), transcription-associated features, including BRG1-mediated chromatin remodeling, and deubiquitylase (DUB) and demethylase (DME) activities, counteract Polycomb chromatin domain formation and limit Polycomb function at these genes.*

Relevance to pluripotency

Genome wide studies have shown that in addition to defined Polycomb domains, H2Aub and H3K27me3 pervasively blanket the genome of ESCs[138,139]. PRC regulated genes generally belong to development hierarchies such as organogenesis, morphogenesis, cell differentiation, and cell-fate commitment, among others[125,140]. Accordingly, PRC-null mice display embryonic lethality at the gastrulation stage[141]. In the ESC models, PRC2 inactivation leads to mild gene expression changes and does not affect pluripotency[142]. In contrast, PRC1 removal in mESCs causes more dramatic derepression of genes and Ring1a/b double knockout results in spontaneous differentiation[143,144]. Additionally, it also remains unclear to what extent coupling between PRC1 and PRC2 underpins Polycomb-dependent gene regulation in hESCs. Taken together, like DNA methylation, PRCs seem crucial during pluripotency exit and the differentiation programs that follow[145].

1.3.5 COMPASS-like complexes: H3K4me3

Complex proteins associated with SET1 (COMPASS) complexes were initially discovered in *Saccharomyces cerevisiae*[146,147]. They have an array of mammalian counterparts termed as COMPASS-like complexes that function in transcriptional activation[148]. These complexes contain histone methyltransferase (HMT) enzymes which deposit mono-, di- or tri- methyl groups at Lys4 on histone H3 (H3K4me1,H3K4me2, H3K4me3 respectively).

Components and Occupancy

Vertebrates have six COMPASS-like complexes with different HMTs which display unique genome-context specific functionalities. The shared and conserved protein core required for function includes the proteins WDR5, ASH2, RBBP5, and DPY30. Based on homology, the catalytic units are split into three groups: SET1A/B, MLL (mixed-lineage leukemia) 1/2 and MLL3/4[149–151] (Figure 8). Like the functional organization of their *Drosophila* orthologues, SET1A/B and MLL1/2 containing complexes are responsible for H3K4me3 enrichment at actively transcribing gene promoters. They are also responsible for the H3K4me2 distribution in regions flanking the 5'end peaks of H3K4me3. MLL3/4 containing complexes on the other hand, are typically associated with H3K4me1 deposition at enhancers[115,152].

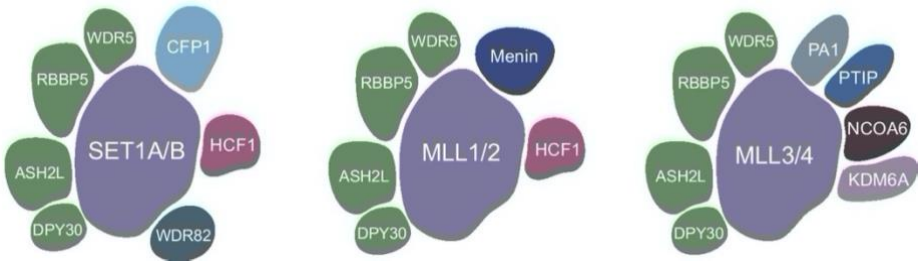

Fig.8: Components of the mammalian COMPASS-like complexes. Adapted from Hughes et al. 2020[153].

A unifying feature of all these target sites is the presence of non-methylated CpG islands in the region. H3K4 HMT subunits contain zinc finger (ZF)-CXXC domains that can specifically recognize non-methylated CpG islands, thereby facilitating recruitment[154,155]. In addition, these complexes also contain Plant Homeo Domain (PHD), that can guide them to preexisting H3K4me3 marks[156]. Multiple histone demethylases, including members of the KDM1/LSD, KDM2A/FBXL11, and KDM5/JARID families, are involved in the removal of methyl groups on H3K4[157].

Function

H3K4me3 is broadly associated with gene activation and transcriptional activity. Studies have identified several facets of the relationship between H3K4me3 and transcriptionally permissive chromatin (Figure 9). Presence of H3K4me3 at CpG islands of genes has been proposed to maintain a permissive state by counteracting the *de novo* methylation of CpG by DNMTs[158]. At certain loci, H3K4me3 could promote the removal of repressive marks (notably H3K27me3 and H3K9me3) by directly stabilizing demethylases and indirectly by preventing the binding of deacetylase compressor complex (NuRD), which in turn reduces the association of PRC2 with chromatin[159–162]. Conversely, H3K4me3 has been shown to favor the accumulation of active histone acetylation marks and increase the accessibility of chromatin at gene promoters by binding histone acetyltransferases (HATs) and chromatin remodelers[163,164]. Finally, there is emerging evidence for the direct interaction of H3K4me3 with RNA Polymerase II to mediate transcription initiation[165,166]. Despite these associations, there is little evidence that gain of H3K4me3 mediates gene activation[153].

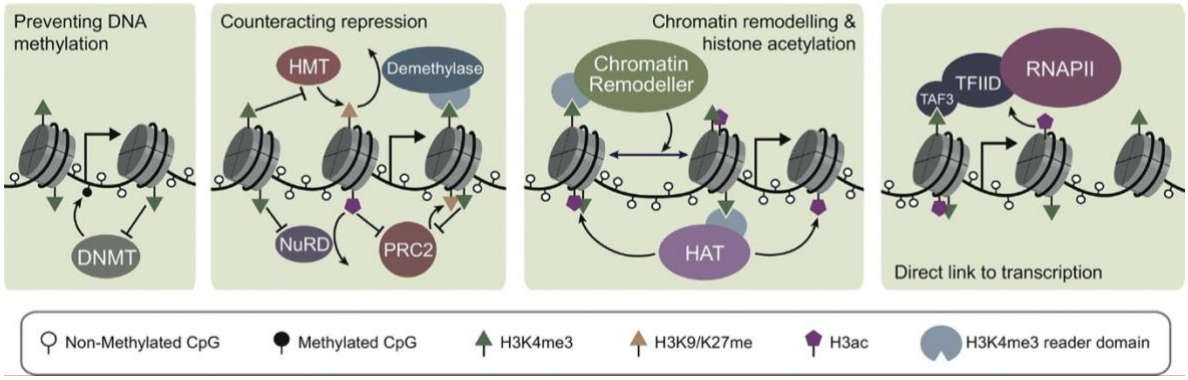

Fig.9: Mechanisms of regulating gene expression by H3K4me3. Adapted from Hughes et al. 2020[153]*. DNA methyltransferases (DNMT); histone methyltransferases (HMT); histone acetyltransferases (HAT).*

Relevance to pluripotency

In ESCs, the MLL2 containing COMPASS-like complex deposit H3K4me3 at the bulk of promoters[167], whereas a subset of genes, including HOX genes, harbor H3K4me3 deposited by MLL1[168], emphasizing the non-redundant function of these two HMTs in the context of development. H3K4me1 is mainly deposited by MLL3/MLL4 complexes as discussed before. The removal of MLL2 in ESCs only reduces H3K4me3 levels at inactive or lowly transcribed genes and does not affect expression programs during differentiation[169]. The loss of MLL3 and MLL4 function does not affect the self-renewal of mESCs either but has severe effects on their differentiation potential[170]. Importantly, it has been concluded that H3K4me3 regulation is less dynamic than other histone marks, owing to its relatively stable genome-wide levels across ESC states[171–173].

1.3.6 Bivalent domains

A hallmark of epigenetic crosstalk in ESCs is the colocalization of the activating mark H3K4me3 and repressive mark H3K27me3 at CpG islands of developmental gene promoters[174,175]. How the correct balance between PRC2 and MLL2 occupancy and activity is maintained at these 'bivalent' regions remains unclear. In ESCs, bivalent domains are transcribed at very low levels and are proposed to poise genes for either activation or repression, depending on the developmental stimuli[176,177]. Recent data however indicate that bivalency is not exclusively a stem cell feature and that committed somatic cells could also acquire *de novo* bivalency[178,179]. In either case, there is currently little evidence for the role of bivalent chromatin in modulating promoter activity[180].

Genome wide ChIP-seq studies of ESCs have shown a striking reduction in the H3K27me3 peaks at bivalent domains in the naïve state as compared to primed, both in mice and humans[40,171,181] making it an accepted criterion to differentiate the pluripotency states. These findings along with loss of function assays have led to the conclusion that bivalency does not serve as the master epigenetic blueprint for pluripotency. Moreover, maintenance of transcriptional fluidity of these developmental loci in naïve pluripotency has now been attributed to promoter proximal pausing of RNA polymerase II, providing a new perspective on gene expression control during development[171].

1.4 RNA POLYMERASE II IN DEVELOPMENT

The eukaryotic transcription cycle involving the multi-subunit containing RNA Polymerase II (Pol II) enzyme begins with the assembly of the pre-initiation complex, followed by subsequent DNA melting to form a transient open complex which is then converted to the initially transcribing complex where RNA synthesis begins in the presence of nucleoside triphosphates. After a critical length of the nascent RNA (6nt) has been achieved, Pol II forms the stable elongation complex, which traverses through the gene body synthesizing mRNA and then finally dissociates from the DNA template marking the termination of a cycle[182]. Rpb 1, the largest subunit of Pol II harbors the C-terminal repeats domain (CTD) consisting of heptapeptide (YSPTSPS) consensus repeats, which undergo phosphorylation changes as Pol II progresses through the transcription cycle. Functional studies with antibodies that recognize these phosphorylation changes have elucidated their role in facilitating the interaction of different factors with Pol II and in functional organization of transcription in the nucleus[183]. Moreover, profiling of genome-wide activities of Pol II with spatial and temporal resolution has provided information on global histone modification patterns regulating transcription as well as presence of subpopulations of Pol II that are stalled shortly after the gene promoters (20-60 bp downstream)[184,185]. It was seen that most of these promoter-proximal locked Pol II molecules were competent to resume RNA-synthesis and confirmed the presence of a paused yet poised state of Pol II[186].

As discussed earlier, transcriptional regulation during development is a function of several factors and efforts to understand how these factors interact with Pol II have been undertaken[187]. For example, the transcription factor *c-Myc* influences Pol II pause release at its target genes, which regulate cellular proliferation[188]. Studies have also highlighted clear functional roles for distinct histone marks such as H3K4me3, H3K9ac, H3K36me3, and other lysine marks in different stages of the Pol II transcription cycle[166,189–192]. Moreover, it has been observed that ESCs in the naïve state employ Pol II pausing more broadly than primed ESCs. The cause of this difference is currently unresolved because while some studies[193–196] show that there is an interplay between Pol II promoter-proximal pausing and Polycomb repression at bivalent loci, others suggest that the prevalence of such poised Pol II is at genes regulating cell cycle and signal transduction, rather than at developmental genes[197].

Discovery and characterization of the above-mentioned regulatory mechanisms of transcription necessitated a better understanding of how they influence global transcription kinetics. Several methods[198,201] used the growing RNA as a measure of Pol II elongation rates. Despite providing detailed maps of Pol II transcription in mammals, results from these studies could not i) establish the relationship between Pol II CTD modifications and the nascent RNA and ii) distinguish between RNA that remain attached to Pol II and those that are released. Recently, these two limitations have been addressed by mammalian native elongating transcript sequencing (mNET-seq)[199] and transient transcriptome sequencing (TT-seq)[200], respectively. To obtain a more nuanced framework for chromatin-based transcription regulation and its consequences on cell fate and survival, present day research is increasingly directed towards deriving dynamic parameters such as elongation rates and pause duration, among others.

In summary, developmental events are a consequence of the crosstalk between several mechanisms which are complex, dynamic, and highly regulated. An overall decrease in chromatin accessibility, increase in repressive marks and hypermethylation of the DNA have been noted in the post-implantation state *in vivo* or in the primed pluripotent state *in vivo*. The opposite trend is found to be true for pre-implantation development or naïve state pluripotency.

1.5 CHROMATIN IMMUNOPRECIPITATION SEQUENCING

Since its inception in the mid-1980's, chromatin immunoprecipitation (ChIP) has been the method of choice for detecting protein-DNA interactions by specifically pulling down the protein-DNA complex (native/fixed) using antibodies[201,202]. Coupling ChIP to suitable readout techniques has been assiduously researched over the years. Starting with qPCR, the field moved on to microarray-based ChIP-Chip experiments in the 1990s[203,204]. These methods were crippled by limitations of total coverage for large genomes, poor resolution, low signal-to-noise ratios, and reproducibility[205]. At the turn of the century, ChIP-PET was proposed to address some of these issues, albeit with its own procedural complications[206]. With the advent of 'second-generation' sequencing technologies such as Helicos, ABI SOLiD and Illumina, high throughput short-read sequencing was coalesced with ChIP to provide ChIP-seq around 2007[207–211]. In short, ChIP-seq involves preparation of the immunoprecipitated DNA into suitable libraries for sequencing to obtain reads that can be aligned to a reference genome. The count of occurrence of DNA reads/tags across the genome provides a numerical map of DNA interaction positions for a given protein (Figure 10).

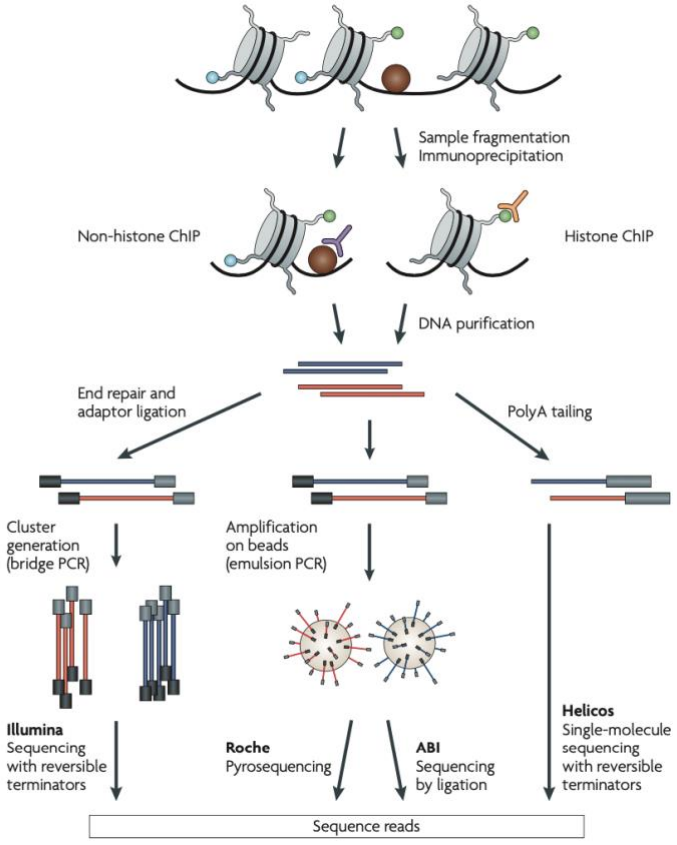

Fig.10 : Overview of a ChIP-seq experiment. Adapted from Park 2009[212].

ChIP-seq has revolutionized the way we look at genome wide distribution of transcription factors, histone modifications, and chromatin remodelers in terms of resolution, accuracy, throughput and most importantly cost, making it an indispensable tool for functional genomics. Over the decade, this method has broadened our horizons on mechanisms of gene expression, allowing us to appreciate the complexity of the system while raising nuanced questions on protein-DNA dynamics. Efforts to answer these questions have led to refined versions of the 'conventional' ChIP-seq protocol, each addressing a different attribute.

The first limitation to be tackled was mapping precision, by protocols such as ChIP-exo and ChIP nexus[213,214], where trimming of the ChIP DNA by an exonuclease at a defined distance from the site of protein-DNA crosslinking allowed to detect highly localized signals (such as transcription factors) at nearly base-pair resolution[215]. The next hurdle of using minimal starting material (low cell number input) was overcome by increasing the efficiency of immunoprecipitation itself and the subsequent library preparation[216]. Adoption of optimized molecular biology reagents and streamlined methods has made the aforementioned steps more robust and has resulted in several variants of the approach such as lobCHIP[217] (library on beads), FARP-ChIP[218] (favored amplification recovered via protection) to name a few. Mapping the co-occurrence of multiple proteins at the same genomic loci has also been addressed recently by re-ChIP or its improved version co-ChIP[96,181] (combinatorial). Moreover, to bring ChIP-seq up to date with recent advances in single cell genomic studies, a microfluidics system has been applied to the procedure and is termed DROP-ChIP[219]. Alongside these advancements, orthogonal techniques such as DamID-seq (DNA adenine methyltransferase identification), ChIC-seq[220] (chromatin immunnocleavage), CUT&RUN[221] (cleavage under targets and release using nuclease) and CUT&TAG[222] (cleavage under targets and tagmentation), relying on enzymatic methods, have also been developed to overcome systemic biases of ChIP-seq and validate previous findings[223].

Methodological advancements in ChIP-seq have been in tandem with creation of computational tools that can process the vast amount of data generated, robustly[224]. The analysis pipeline of a ChIP-seq experiment deals with different aspects of data assessment and management such as identifying artefacts, filtering low quality reads, estimating library complexity, sophisticated read alignment and peak calling algorithms, and downstream analysis tools. The choice of tool for each step of the pipeline depends on the underlying biological question and specifics of the data such as sequencing technology, sequencing depth and read length to name a few[225].

Integrative analysis of multiple ChIP-seq data sets can provide a more holistic perspective on biological phenomena[226]. However, the interpretation of results from such analysis requires caution since variation in experimental conditions might render a quantitative direct comparison of such raw datasets unfeasible. In this sense, a previous normalization step is necessary. One strategy for quantitatively comparing ChIP-seq signal intensities has been incorporation of exogenous DNA or synthetic histone spike-in controls. Recently, another method called parallel-factor ChIP[227] has been described that provides an internal control for quantitative analysis by utilizing a second antibody against the target chromatin.

Though this technique eliminates the need for an exogenous spike-in, its dependency on antibodies could cause concerns regarding reproducibility. The alternative strategy of multiplexing or combined sample processing in ChIP-seq allows for concurrent profiling of multiple DNA-protein interactions across the genome, in addition to providing quantitative consistency. This has been implemented by chromatin barcoding: performing adapter ligation to the DNA before (iChIP: indexing first[228], BarChIP: barcoded high-throughout[229], Mint-ChIP: multiplexed, indexed T7[230]) or during the IP (ChIPmentation: Tn5 based tagmentation [231]).

In conclusion, future research into gene regulation should be augmented with advancements made in methods such as ChIP-seq to refine our understanding of how a cell functions.

2 RESEARCH AIMS

The overall aim of this thesis was to develop a quantitative ChIP-seq methodology based on barcode-and-pool strategy (MINUTE ChIP) for assessing dynamic genome-wide epigenetic alterations.

The specific aims of the individual projects were:

I. To implement MINUTE ChIP to compare epigenomic profiles between mESCs grown in 2i and serum conditions.

II. To quantitatively profile the epigenome of naïve and primed hESCs, and functionally characterize the role of PRC2 in the two stages.

III. To assay global transcriptional kinetics in three pluripotent mESC states using TT-seq and MINUTE ChIP.

3 THE METHOD: MINUTE CHIP

This section provides an overview of the MINUTE ChIP workflow including a brief description of the analysis pipeline and the calibration experiment conducted to benchmark the protocol (included in Paper I). For a detailed description of all the other methods used, please refer to the methods section of the corresponding papers.

As previously discussed, the need for quantitative ChIP-seq has been widely appreciated, and several quantitative methods have been proposed. Mint-ChIP developed by the Bernstein lab[230], uses the barcoding based pool and split strategy which allows for multiplexing, while effectively removing technical variability between the samples. After a traditional DNA fragmentation by Micrococcal nuclease (MNase) on a fixed number of cells, the protocol features a direct ligation of barcoded adaptors to the DNA fragments, allowing samples to be distinguished during *in silico* analysis. Next, samples from the conditions to be compared are pooled and all subsequent steps including immunoprecipitation, library preparation and next-generation sequencing are performed on the barcoded pool. As a result, the read counts obtained reflect relative differences between samples and can be compared in a quantitative manner without the need for exogenous spike-in controls[232,233] or semisynthetic standards[234].

The Mint-ChIP library preparation includes an *in vitro* transcription reaction (by T7 polymerase using a promoter included in the ligated adaptor) to amplify the ChIP DNA. This implies that only one end of the chromatin fragment needs to be adapter-ligated for amplification, which increases efficiency. The resulting RNA is then reverse transcribed (RT), and prepared into an Illumina-compatible library using a low-cycle polymerase chain reaction (PCR) step, where a second barcode is introduced to identify the corresponding ChIP assay.

In comparison to the published protocol, which was performed in human cancer cell lines, buffer conditions for lysis and digestion of ESCs were optimized to contain lower salt concentrations to preserve native chromatin interactions. MNase concentration was titrated to obtain predominantly mono-nucleosomal fragments for cell numbers ranging from 10^4 to 10^6, while ensuring equal representation of all samples. The amplification steps during library preparation favor certain sequences and could induce duplication bias. To account for this, we introduced a unique molecular identifier sequence (UMI: six randomized nucleotides) in the adapters, enabling us to provide a more quantitative account for all genomic sequences including repetitive regions. Moreover, since mapping efficiency can be further improved by generating paired end reads, we additionally ligated a short sequence to the 3' end of the RNA before the RT step, to prime the cDNA. The above stated refinements to the original protocol motivated us to coin a new name for the method, multiplexed indexed unique barcoded T7 paired-end sequencing ChIP: MINUTE ChIP (Figure 11). In our method, quantitative normalization is achieved by relating the barcode representation after the ChIP back to the corresponding quantities in the input pool.

3.1 WORKFLOW

Lysis and chromatin fragmentation

$1x10^6$ cells were harvested for each growth condition, washed twice with PBS and cell pellets were flash frozen at -80°C prior to use. Cells were resuspended in 50 µl PBS, lysed and digested to mono- to tri-nucleosomes fragments by adding 50 µl of 2x Lysis buffer (*100 mM Tris-HCL [pH 8.0], 0.2% Triton X-100, 0.1% sodium deoxycholate, 10 mM $CaCl_2$ and 1x PIC*) containing 2U/µl of Micrococcal nuclease (New England BioLabs, M0247S) and incubating on ice for 20 min and then 37°C for 10 min.

Barcoding and pooling

For each sample, 40 µl of the whole cell lysate containing the digested chromatin was taken forward into an overnight blunt end ligation reaction (End-It DNA repair kit and Fast-Link DNA ligation kit, Epicentre) with double stranded DNA adapters at 16°C. Double-stranded DNA adaptors for barcoding and T7 amplification were generated by slow annealing of complementary single-stranded oligos. As in the original Mint-ChIP design, adaptors carried a partial SBS3 for Illumina sequencing flanked by a T7 RNA Polymerase for linear amplification. Between the SBS3 sequence and the 8bp sample barcode at the 3' end, a 6bp randomized sequence was introduced, serving as a unique molecular identifier (UMI). UMI and sample barcode are ligated 5' to the chromatin fragment and would therefore constitute the first 14 bases of read 1. The 4096 possible UMIs provide sufficient diversity to distinguish if two reads mapping to the exact same genomic location arose from a PCR amplification artifact or are indeed unique molecules. The adapter concentration was optimized to 2.5 µM / reaction to reduce adapter dimers and enable sequencing of the input. The ligation reaction was terminated with a lysis dilution buffer *(50 mM Tris-HCl [pH 8.0], 150 mM NaCl, 1% Triton X-100, 50 mM EGTA, 50 mM EDTA, 0.1% sodium deoxycholate and 1x PIC)* and barcoded samples were combined into a single pool, spun down at 24,000 r.p.m. for 10 min at 4°C.

Immunoprecipitation

50µl Protein A/G magnetic beads (BioRad) were washed twice with PBS-T (PBS+ 0.1% Tween 20) and coupled to one of the following antibodies in the same buffer for 1 hr at room temperature with rotation: 3 µl H3 (Active motif 39763), 5 µl H3K4me3 (Millipore 04-745), 5 µl H3K27me1 (Cell signaling 5326), 5 µl H3K27me2 (Cell signaling 9728), 5 µl H3K27me3 (Cell signaling 9733 or Diagenode C15410195 or Millipore 07-449). Beads were then washed quickly with RIPA buffer. 200-400 µl of the cleared lysate pool was added to the pre-coupled magnetic beads and parallel ChIP assays were incubated further for 4 h at 4°C with rotation. 5% of the above volume was saved as the input pool and processed through the remaining protocol in a manner similar to the IPs. Next, the beads were washed (RIPA, RIPA high salt, LiCl and TE buffer) resuspended in ChIP elution *buffer (10 mM Tris-HCl [pH 8.0], 1 mM EDTA, 0.1% SDS, and 300 mM NaCl)* containing 0.25 mg/mL Proteinase K (Thermo Fisher Scientific, 25530015) and eluted at 63°C for 1 h.

Linear amplification and library preparation

The native ChIP DNA (fragments longer than 100bp) was isolated using 1x SPRI beads (Beckman Coulter) and set up in an overnight *in vitro* transcription reaction (HiScribe T7 Quick High Yield RNA Synthesis kit, New England BioLabs). The resulting RNA was purified using Silane beads (Life Technologies) and an RNA 3' adapter was ligated onto it using T4 RNA ligase, truncated (New England BioLabs) for 1 h at 25°C. The mixture was subsequently supplemented with components of the reverse transcription reaction (SuperScript III First-Strand Synthesis SuperMix, Life Technologies) to produce cDNA, primed using the ligated 3' adapter. Final libraries for each ChIP were produced using 150-200 ng of purified cDNA in a PCR reaction (High-Fidelity 2x master mix, New England BioLabs) for 8 cycles with 0.2 µM primers that carried a second 8bp barcode sequence. Quality assessment and concentration estimation of the purified DNA was done using Qubit (Life Technologies) and BioAnalyzer (Agilent). Each library was then diluted to 4 nM and combined into a single pool before sequencing on the Illumina NextSeq500 platform. Sequencing was performed using 50:8:34 cycles (Read1:Index1:Read2)

3.2 ANALYSIS PIPELINE

Data processing was automated through the **minute** pipeline, developed by my colleague Carmen Navarro[235]. The pipeline functions to transform a set of multiplexed FASTQ files into properly scaled, directly comparable per-sample bigWig files that can be visualized on a genome browser tool and serve as basis for downstream analysis (Figure 12). Additionally, relevant quality control steps are performed and reports are generated throughout the workflow. Downstream genomic analysis was performed using existing bigWig suites and R packages, particular details of which can be found in the respective papers.

Input

The main input files needed for **minute** execution are paired-end multiplexed FASTQ files. Additionally, two tables are required: libraries.tsv specifies the barcode used in each of the multiplexed libraries, and groups.tsv specifies the scaling configuration and the reference to which each sample should be mapped. An optional BED file containing regions to be filtered out can also be provided. The reference genomes to be used must also be provided as FASTA files with corresponding Bowtie 2[236] indexes.

Quality control

QC metrics are generated at several steps of the pipeline and their outputs together with each step result are summarized into a final report using MultiQC[237]. FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) is used to process FASTQ files. Insert size metrics are gathered using Picard and mapping statistics are generated with SAMtools[238] (https://github.com/broadinstitute/picard). Known contaminant sequences (e.g. MINUTE adaptor concatemers or poly-G second reads from failure of cluster reversal in the Illumina flow cell) are removed explicitly using Cutadapt[239] with an error tolerance of 15%.

Demultiplexing

FASTQ files are demultiplexed with Cutadapt using a sample-specific barcode of 8 nucleotides at the 5' end of read 1 (first read in pair). By default, one mismatch is allowed for assigning each read to its corresponding sample since barcode pools should usually keep a minimum Hamming distance of 3. The allowed number of mismatches can be configured to better fine-tune performance to different barcode pool distributions.

Read mapping and Deduplication

Reads are mapped to the specified reference genome using Bowtie 2 with parameters --reorder (for reproducibility) and --fast. When pooled replicates are specified in the groups table, the generated BAM files are merged into a single one using SAMtools. Duplicate reads are found with je markdupes (https://github.com/gbcs-embl/Je/), which marks reads as duplicate if both their mapping locations and UMIs match. The program is run on a version of the BAM file that only contains the read 1 (first read in pair) sequences since T7-based amplification does not guarantee amplification of the full-length template fragment. Duplicate read pairs are then removed from the original BAM files by a custom script that uses pysam.

Input-normalized scaling factors

Scaling factors are calculated at this step to generate a set of genomic profiles that are quantitatively comparable. The rationale for calculating scaling factors is as follows: in a theoretical ideal scenario, all sample barcodes in the input pool are represented exactly equiproportional. Following from the biochemical binding equilibrium, a library prepared from the ChIPed DNA will then contain a barcode distribution that reflects proportionally the relative abundance of the immunoprecipitated epitope in each of the pooled samples. In practice, barcoding efficiency varies between samples due to experimental imperfection (small variations in cell number, volumes, adaptor oligo amount and reaction conditions). Hence, the input barcode distribution must be separately assessed from an Input library, and the read counts for each barcode in the ChIP library are normalized to their respective counts in the input library, yielding input-normalized ratios (INRs). In an equilibrium binding, INRs will reflect the relative abundance of the immunoprecipitated epitopes independent of the input variability.

The scaling factor is calculated as follows:

Let $Reads_{ChIP,barcode}$ be the number of MINUTE ChIP mapped, deduplicated reads in a given condition (i.e. barcode) for a given ChIP, and $Reads_{input,barcode}$ the corresponding number of MINUTE ChIP input reads for the same condition. We define the Input Normalized Ratio (INR) as:

$$INR_{ChIP,condition} = Reads_{ChIP,condition} / Reads_{Input,condition}$$

To convert these raw INRs to an intuitively meaningful value, one condition (represented by a single barcode or replicate barcodes) is chosen as the reference condition. The reference condition receives a global mean of 1 (as commonly used in Reads per Genome Coverage, also termed "1x Genome Coverage" normalization methods). The remaining

conditions in the same ChIP pool are scaled relative to this reference condition. We define the minute-Scaled Ratio (MSR) for such ChIP, reference and condition as:

$$MSR_{ChIP,reference,condition} = INR_{ChIP,condition} / INR_{ChIP,reference}$$

Multiplying this ratio by a 1x genome coverage factor we can calculate the corresponding minute-Scaled Read Counts (MSRC):

$$MSRC_{ChIP,reference,condition} = MSR_{ChIP,reference,condition} * (G_{size} / F)$$

Note that $MSRC_{ChIP,reference,condition} = (G_{size}/F)$ when $reference = condition$, hence $MSRC_{ChIP,reference,condition}$ becomes a constant that represents 1x genome coverage for the corresponding reference genome and fragment length (RPGC).
A factor that scales the BAM file to the final quantitative track is calculated accordingly to generate scaled bigWig files in the next step. The scaling factor $K_{ChIP,reference,condition}$ is defined therefore as:

$$K_{ChIP,reference,condition} = MSRC_{ChIP,reference,condition} / Reads_{ChIP,condition}$$

K value is used as --scaleFactor parameter for deepTools when generating each bigWig file in the next step.

BigWig generation

minute-scaled bigWig files are produced using deepTools[240] using --scaleFactor K, calculated as described above. deepTools requires the effective genome size as another parameter and this is calculated as the number of non-N nucleotides in the reference genome provided as input. Unscaled bigWig files disregarding the quantitative scaling information are generated as well, each normalized individually to 1x GenomeCoverage.

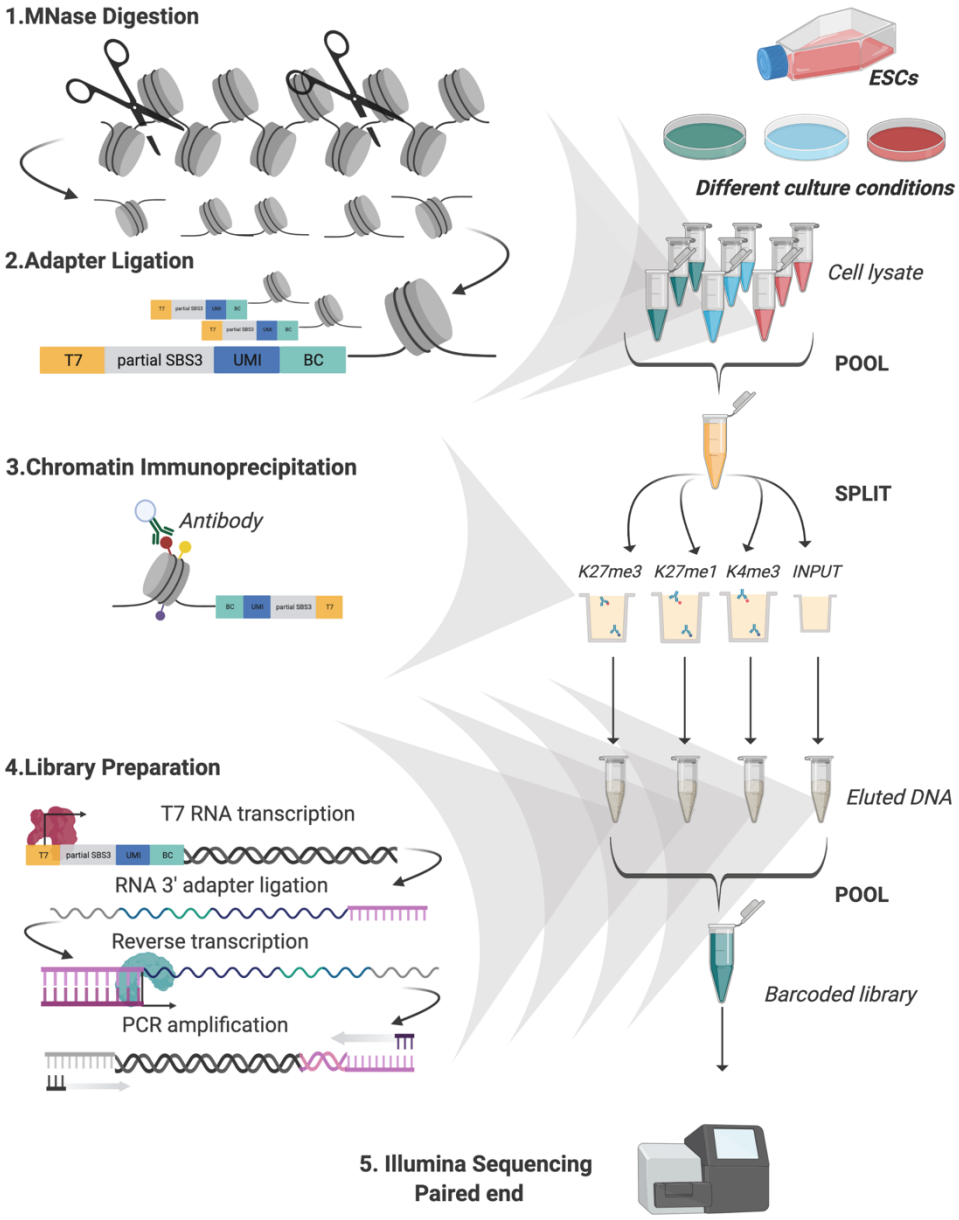

Fig.11: Scheme of the MINUTE ChIP workflow (created with BioRender.com). Triplicate cell pellets for each condition are lysed and the chromatin is enzymatically fragmented to mono- and di- nucleosomes and barcoded by ligating on a dsDNA adapter. Samples are then pooled, and the barcoded lysate is aliquoted to individual ChIP reactions (5% of the ChIP volume was reserved as input material and carried through protocol in a manner similar to the IPs) with magnetic beads pre-coupled with the respective antibodies. For constructing the final libraries, DNA from each IP is in vitro transcribed using the T7 promoter in the adapter. The resulting RNA is appended with a RNA 3' adapter (RA3) allowing for specific paired end sequencing. The RA3 in turn is used to prime the reverse transcription reaction, generating cDNA that serves as template for the final low-cycle library PCR. At this stage, in addition to the Illumina-compatible sequences, PCR primers also carry a second barcode sequence to serve as an identifier for the IP performed. Finally, libraries are pooled and sequenced on the Illumina platform.

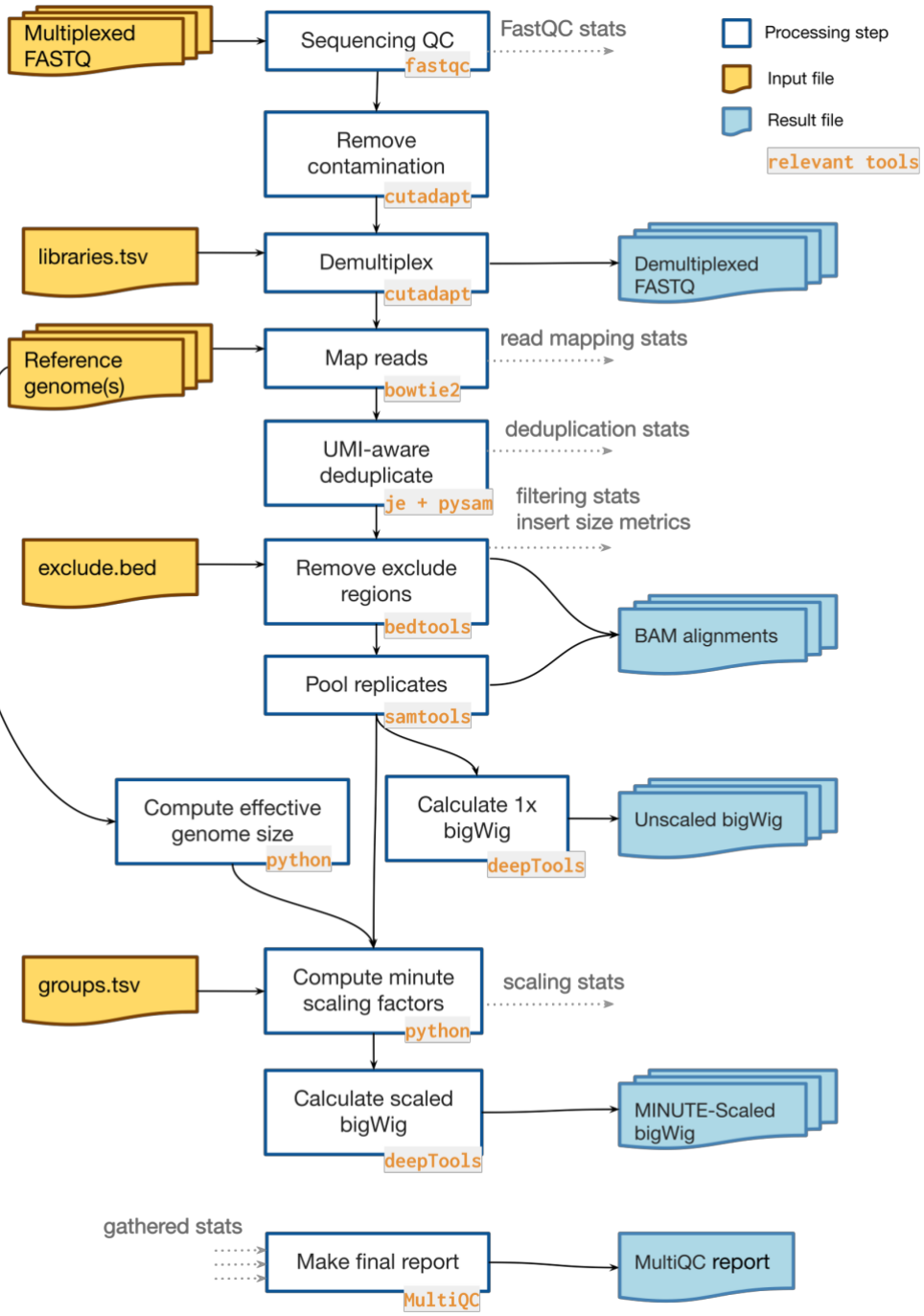

Fig.12 : minute pipeline workflow. Adapted from Navarro et al. 2022[235]*. Yellow files represent necessary inputs. Blue files represent final outputs. Each main processing step is annotated with its corresponding relevant tool. Quality control metrices extracted are shown as annotated grey arrows. Finally, a MultiQC report is generated gathering all QC stats.*

3.3 VALIDATION

The input-based normalization facilitated by the barcoding-pooling strategy confers MINUTE ChIP with the ability to report relative differences in occupancy of proteins. For these measurements to be truly quantitative they should reflect the true biological differences between samples, over a dynamic range. We therefore sought to benchmark MINUTE ChIP in terms of parameters such as linearity, limits and sensitivity using a calibration experiment.

We generated a dilution series of the histone modification H3K27me3 to simulate a standard curve-like assay, that would span the entire physiological range of the modification. For this purpose, mESCs were treated with inhibitors that target EZH2 of PRC2 (EPZ-6438) to reduce H3K27me3 below detectable levels, or demethylases JMJD3/UTX (GSK-J4) to increase H3K27me3 above physiologic levels. By mixing cells from these two treatments, H3K27me3 low (L) and H3K27me3 high (H), in defined ratios (Figure 13a) we generated a defined gradient of H3K27me3. Two replicates of each mixing ratio were uniquely barcoded and carried through the MINUTE ChIP protocol, and probed for H3K27me3, H3K27me2, H3K27me1, and total H3. The resulting raw data was processed through the analysis pipeline to obtain normalized read counts for each barcode. Indeed, the read counts were proportional to the mixing ratios ($R^2 = 0.97$), confirming that our measurements were effectively capturing the simulated differences (Figure 13b). For a locus specific context, we also assessed the distribution of H3K27me3 at Polycomb group (PcG) targets, where the highest enrichment for this modification is expected. Reassuringly, the average H3K27me3 signal across 2731 PcG binding sites followed the mixing ratios proportionally, over the entire gradient (Figure 13c). Importantly, the quantitative signal of H3 ChIP was essentially constant across conditions. Further, at genomic areas with much lower H3k27me3 signal (like active genes), the MINUTE ChIP quantification produced a proportional representation of the mixing ratios. Finally, we observed similar proportional representation in the H3K27me2 and H3K27me1 ChIPs and over a number of additional representative regions of the genome (data shown in Figure1 & S1 in Paper I[241]).

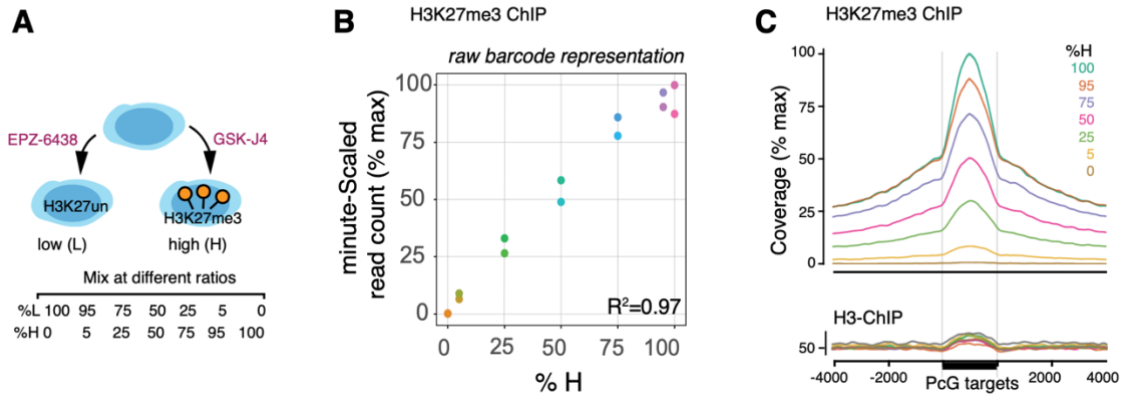

Fig.13: MINUTE ChIP accurately reflects global and local proportions[241]. (a) schematic of the calibration experiment setup. (b) ***minute*** *quantification after H3K27me3 immunoprecipitation: input normalised and scaled read counts for two technical replicates with different barcodes are plotted against mixing ratio. Coefficient of determination(R^2) for a linear regression is given. (c) H3K27me3 (average of two replicates) signal across 2,731 PcG target sites. For comparison, H3 signal is shown in the lower panel.*

4 SUMMARY OF RESEARCH PAPERS

4.1 MOUSE GROUND STATE PLURIPOTENCY CHARACTERIZED BY QUANTITATIVE CHIP-SEQ (PAPER I)

In this study, we implemented the quantitative MINUTE ChIP method to compare epigenetic profiles of mESCs grown in 2i (naïve), 2i/serum and serum conditions. We first demonstrated through a calibration experiment that MINUTE ChIP has a large linear dynamic range for quantifying relative differences of histone modifications across multiple pooled samples (as discussed in the previous section).

From three independent MINUTE ChIP experiments (including biological triplicates and different H3K27me3 antibodies), we observed a two-fold increase in H3K27me3 levels in the presence of 2i. This was reversible within three passages in serum-only media. Global H3K4me3 levels were decreased by 1.8-fold in (serum free) 2i condition. Taken together, high levels of H3K27me3 and low levels of H3K4me3 characterize the 2i ground state (Figure 14).

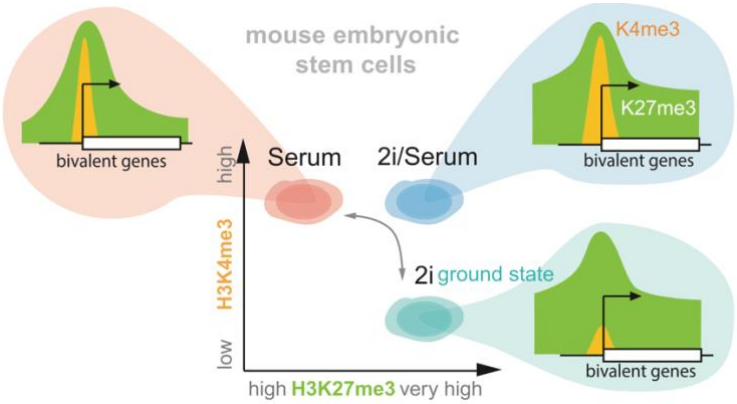

Fig.14: Murine pluripotency as a function of H3K4me3 and H3K27me3 levels globally and at bivalent promoters[241].

Our quantitative comparison revealed that the global reduction of H3K4me3 in 2i manifested in reduced peak sizes across all genes but most strikingly at bivalent promoters. In contrast, we identified a broad genome wide accumulation of H3K27me3 in 2i with H3K27me3 peaks being retained stably at PcG binding sites across conditions. Looking at bivalent genes (which overlapped largely with PcG targets) in 2i vs serum, we confirmed that there was no significant change in the H3K27me3 levels at ~2000 promoters. Moreover, the subset of bivalent promoters in our quantitative datasets that showed changes in H3K27me3 between 2i, and serum (171 lost and 126 gain) had matched gene expression changes (up and down regulation respectively).

Inhibitor based depletion of CpG methylation in the serum condition recapitulated the H3K27me3 hypermethylation seen in 2i, suggesting that the gain of H3K27me3 follows the loss of CpG methylation in naïve mESCs. Furthermore, inhibitor-based targeting of the writer (EZH2 of PRC2) and eraser (UTX/JMJD3 demethylases) of H3K27 methylation

revealed that these complexes function jointly to modulate the H3K27me3 levels seen in the different pluripotency states.

Discussion

A body of work has elucidated that embryonic developmental transitions are accompanied by epigenomic remodelling. This study, however, highlights the importance of quantitatively characterizing changes in histone modifications and demonstrates that changes in growth condition are accompanied by unexpected global shifts in H3K27me3 and H3K4me3 levels. Our quantitative data (coupled to pharmacological inhibitions) allowed us to better characterize bivalent promoter status in the different culture conditions and reconciles prior misconception of reduced bivalency in naïve mESCs. We provide an alternative explanation to the originally interpreted loss of H3K27me3[171]: bivalent domains start out in the ground state mainly covered by H3K27me3 and, only upon priming, accumulate H3K4me3 levels comparable to those at active genes. Essentially the dense lawn of H3K27me3 in naïve mESCs gets mowed upon transition to a primed state, revealing existing H3K27me3 peaks at all cognate PcG targets. These conclusions agree with the repressed state of bivalent genes in 2i mESCs. Thus, albeit dispensable for maintaining pluripotency, it remains to be elucidated whether the pervasive activity of PRC2 can be attributed to a specific function or rather represents a collateral effect of transcription factor networks overriding epigenomic control mechanisms to safeguard the pluripotent state.

4.2 PRC2 ACTIVITY SHIELDS HUMAN NAÏVE PLURIPOTENCY (PAPER II)

In this study, we explored the function of promoter bivalency in human pluripotent stem cells by using epigenomic profiling, bulk, and single cell transcriptomics. Using MINUTE ChIP, we profiled three histone modifications H3K27me3, H3K4me3, and H2Aub in biological triplicates of naïve and primed hESCs. Quantitative maps revealed a diffuse increase in H3K27me3 in naïve hESCs as compared to primed (~3.3 fold higher), while genome wide H3K4me3 signal was comparable between the conditions. Like H3K27me3, global H2Aub levels were significantly higher in naïve cells (~2.1 fold) in line with the view of coregulation of PRC1/PRC2 activity. Intriguingly, genome-wide enrichment analysis across 10 kb windows and functionally annotated chromatin states revealed that H2Aub did not mirror the H3K27me3 trends, and its level remained mostly unchanged in H3K27me3-depleted naïve cells. Therefore, H3K27me3 serves as a distinguishing histone modification between naïve and primed hESCs. Chromosome wide analysis in naïve hESCs highlighted that the transcriptionally active X chromosomes were decorated with ~10% of all the H3K27me3 marks, more than any autosome. However, RNAseq showed that neither the natural accumulation nor PRC2 inhibitor-based removal of the repressive H3K27me3 mark had any global effect on the transcriptional output of the naïve X chromosomes.

Despite previous studies reporting an overall loss of H3K27me3 at bivalent promoters in the naïve state, we sought to revisit bivalency in light of our quantitative datasets and defined the following five categories of promoters based on the H3K27me3/H3K4me3

status in naïve and/or primed hESCs: primed-bivalent; naïve-bivalent; common bivalent, H3K4me3 only, and none. Next, we examined if H3K27me3 played a functional role at these promoters using RNA-seq. Expression of naïve-bivalent genes was indeed significantly lower in naïve hESCs, while primed-bivalent were predominantly lower expressed in primed. EZH2 inhibitor-based depletion of H3K27me3 resulted in increased gene expression in both states but we noted significant de-repression of only the state specific bivalent genes. Our data, therefore, disproves the dogma that promoter bivalency is lost in naïve hESCs and instead reveals that H3K27me3 mediates the repression of a distinct set of bivalent genes in each state, which include key transcription factors of naïve and primed pluripotency (Figure 15a).

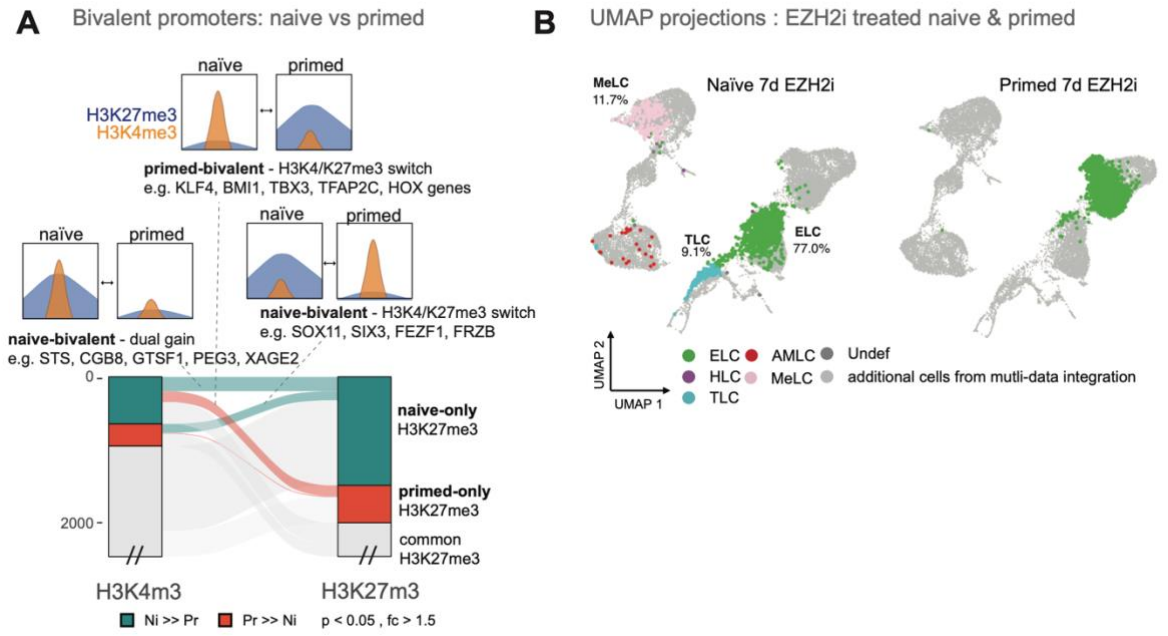

Fig.15: H3K27me3 represses a naïve-specific subset of bivalent genes. (a) Alluvial plot showing H3K27me3 and H3K4me3 gains and losses (DESeq2 p.adj < 0.05 and fold-change > 1.5 from three replicates) at bivalent promoters between naïve and primed hESCs. Select connections (each containing more than 100 genes) are annotated. (b) UMAP projection and cell identity annotations of single-cell transcriptome datasets obtained for EZH2i treated naïve and primed cells at indicated time point, showing Epiblast-like cells (ELC), Trophectoderm-like cells (TLC), Mesoderm-like cells (MeLC), Amnion-like cells (AMLC), hypoblast-like cells (HLC).

EZH2i-based depletion of H3K27me3 caused gene expression changes in both naïve and primed cells. However, the number of genes upregulated upon removal of the repressive mark was strikingly higher in the treated naïve hESCs and included several known developmental genes. This led us to assess expression changes of a comprehensive list of human lineage markers upon inhibition of PRC2, of which the extraembryonic (trophectoderm, amnion and mesoderm) lineage markers showed the strongest upregulation. Investigating the bivalent status of the induced genes directed us to GATA3, which is considered as the master transcription factor for trophectoderm development. Alternative strategies to target PRC2 coupled to immunostaining (two inhibitor EZH2i and EEDi and CRISPR based genetic knock out of EED) univocally confirmed the induction of GATA3 in a fraction of naïve hESCs upon removal of H3K27me3.

To quantitatively resolve the cellular heterogeneity resulting from the targeting of PRC2 in hESCs we performed single cell RNA sequencing (scRNA-seq) of naïve and primed cultures along with a EZH2i time course treatment (2, 4, and 7 days). Upon mapping the resulting data on to a comprehensive cellular reference, we observed that although 75% of EZH2i treated naïve cells remained in the epiblast-like cluster (ELC), a fraction of cells progressively shifted to trophectoderm-like cells (TLC) and mesoderm-like cells (MeLC) reaching 9.1% and 11.7% respectively after 7 days of treatment (Figure 15b). On the other hand, 99.96% of the EZH2i treated primed cells remained within the ELC cluster. Further, by performing trajectory-based differential expression analysis we demonstrate how naïve hESCs upon PRC2 inhibition, transition through a bifurcation point (activated ELC) before committing to a stable, lineage end state (MeLC or TLC). We also describe the transcription factors expressed in the activated, uncommitted precursor population and in the distinct lineage clusters. Taken together, these findings in addition to delineating our bulk RNA-seq observations, establish the role of PRC2 in shielding naïve human pluripotency from differentiation.

Discussion

Stabilizing the human naïve pluripotent state *in vitro* involves targeting a diverse set of factors and signaling pathways and has necessitated more definitive markers to characterize naïve hESCs. Epigenetic signatures, specifically hPTMs, have emerged as a criterion to evaluate cell state transitions. In this study, we couple MINUTE ChIP with transcriptome profiling to functionally characterize the epigenome of the human naïve and primed pluripotency stages.

Similar to the murine model (Paper I), our quantitative epigenomic profiling contradicts earlier reports of loss of H3K27me3 across bivalent promoters in naïve hESCs[33,40]. We observe an increased pervasive deposition of H3K27me3 and H2Aub throughout the genome of naïve hESCs. Surprisingly, erasure of H327me3 by inhibiting PRC2 catalytic activity had no drastic effect in the H2Aub pattern. Our observation reinforces an emerging theme in the Polycomb field: PRC1 and PRC2 might have independent gene regulation mechanisms despite their molecular coupling on the chromatin[137].

In addition, with quantitative profiles of H3K4me3 we discover previously undescribed trends at bivalent promoters, specific to the different pluripotent states. Despite the extensive overlap of target promoters between naïve and primed states, ablation of PRC2 activated independent gene sets, characteristic for each state. In naïve hESCs removal of H3K27me3 resulted in a fraction of cells to spontaneously express trophectoderm marker genes. A strikingly different transcriptional response of primed cells to EZH2i inhibition highlights the exquisitely context-specific function of PRC2 in repressing largely non-overlapping developmental regulators in closely related pluripotent states. Therefore, within the frame of culture conditions used, our findings are at odds with the notion that PRC2 is dispensable for maintenance of naïve pluripotency[242,243].

Moving forward, a central objective remains to uncover the detailed molecular mechanisms through which Polycomb complexes achieve transcriptional poising and if these observations hold true in the *in vivo* setting of early human development.

4.3 DISTINCT TRANSCRIPTION KINETICS OF MOUSE PLURIPOTENT CELL STATES (PAPER III)

In this study, we unravel RNA metabolism kinetics in three pluripotent mESC states (2i naïve, SL ground and SL mTORi paused) by combining transient transcriptome sequencing (TT seq) with MINUTE ChIP based profiling of RNA polymerase II (Pol II) occupancy. Metabolic labelling of newly synthesized RNA using 4-thiouridine (4sU), confers TT seq with the ability to demarcate transcription units (TUs) and measure transcriptional activity with high sensitivity and accuracy. We exploited this feature to *de novo* annotate coding (mRNA) and non-coding TUs (ncRNA) for each mESC state and stratified the latter further based on their relative location and direction to protein coding TUs.

Comparing TT-seq total and labelled RNA, scaled by spike-in RNAs revealed that 2i and mTORi states had a smaller labelled-to-total RNA ratio, suggesting a slowdown in RNA turnover and decreased RNA abundance upon treatment. Next, differential expression analysis showed that the SL to mTORi transition was associated with a uniform down-regulation of both mRNA and ncRNA synthesis. In contrast, converting cells to the 2i naïve state induced a highly variable response with both up- and down-regulation of individual mRNAs and ncRNAs. Interestingly, we found that expression trends were concerted between protein coding TUs and neighbouring (+/- 100 kb) intergenic TUs suggesting that transcriptional variations observed in pluripotent state transitions are modulated by genomic positioning.

A key determinant of transcriptional kinetics is elongation velocity, and in this study, we estimated this as the ratio of the number of polymerases released into elongation per unit time, as measured by TT-seq labelled RNA coverage, over the average density of polymerases on the DNA template, measured by MINUTE ChIP Pol II CTD serine 5 phosphorylation (S5p) coverage. mESCs grown in 2i and mTORi showed a steep decrease in median elongation velocities relative to SL (2.2- and 3.7- fold respectively). Intriguingly, pausing time, which corresponds to the time Pol II needs to travel through the promoter-proximal pausing interval, showed good correlation with elongation time in all three conditions (Figure 16). Moreover, in SL cells, we observed several associations between transcription velocity and epigenomic features (histone modifications, chromatin remodelers). Finally, by studying conditions with slower global elongation velocities, we reveal that elongation velocity influences the choice of transcription termination site to a great extent.

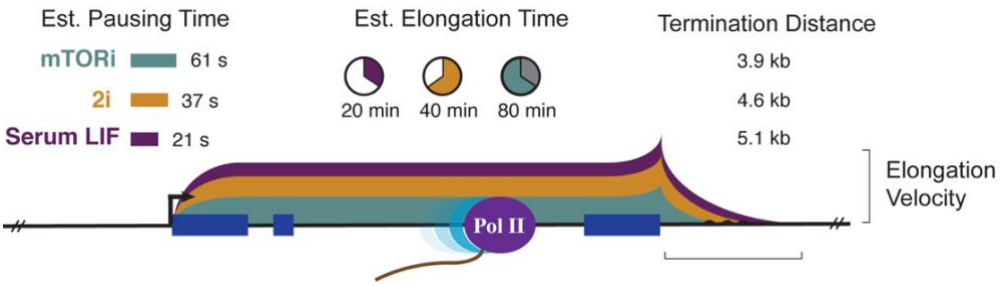

Fig.16: Global transcriptional kinetics in three pluripotency states of mESCs[244].

Discussion

Pluripotent state conversions of mESCs are accompanied by rewiring of the regulatory circuits and have characteristic gene expression signatures. In this study we investigate the extent to which global transcriptional activity itself contributes to transcript abundance in different pluripotent states. We demonstrate the strength of combining quantitative ChIP-seq profiles with transient transcriptome measures to robustly estimate elongation velocities that provide key global and local kinetic parameters across conditions. Our data support a model in which inhibition of MEK/GSK3β or mTOR signaling pathways decreases global transcriptional output, total RNA abundance, RNA turnover, RNA polymerase II elongation velocity, and termination distance. This is in line with a recent observation that transcription in ESCs was tuned to translational output via a positive feedback loop between chromatin and translation, whereby the transcriptionally permissive chromatin state both depends on and controls protein synthesis[245].

5 CONCLUDING REMARKS AND PERSPECTIVE

The work presented in this thesis highlights the merits of using a multiplexed, quantitative ChIP-seq methodology to study dynamic chromatin features during early embryonic development. As highlighted in the summary of the respective papers, our quantitative epigenetic profiles have allowed us to revise the bivalency dogma both in mouse and human naïve pluripotent states. Indeed, the discrepancies between our observations of broad H3K27me3 distribution and those in the literature can be attributed to the limitations of conventional ChIP-seq method to differentiate signal from noise, reliably and quantitatively.

MINUTE ChIP detects changes in histone marks (such as H3K27me3, H3K4me3, and H2Aub) with high sensitivity that can be attributed to a specific cell state or identity. But taken alone, this still only provides a descriptive view of the epigenetic landscape. Moreover, besides the correlation between active/repressive histone marks and transcription, there is little evidence as to how histone modifying systems cause gene expression changes. In the studies included in this thesis we infer function by pharmacological manipulation of PRC2 and coupling MINUTE ChIP to other technologies such as TT-seq, bulk and scRNA-seq.

In the context of mESCs, **Paper I** showed that naïve and primed states exhibit distinct states of bivalency which are most probably driven by transcription factor networks. In addition, **Paper III** elucidated how genome-wide changes in transcriptional kinetics that correlate with chromatin features invoke the various mouse pluripotent states. In the context of hESCs, **Paper II** revealed an extensive rewiring of promoter bivalency and the role of PRC2 in shielding naïve human pluripotency from trophectoderm differentiation. Returning to the analogy of the Waddington landscape, we speculate that PRC2-mediated H3K27me3 sets up a 'barrier' that dynamically follows and reinforces lineage choices. Importantly, the repressive function of H3K27me3 is distinctly different in the mouse and human naïve states, albeit being present pervasively across the genome of both species.

Future work on deciphering if epigenetic modifications play a directive, supportive or consequential role in gene expression will benefit from time-resolved studies (conditional knockouts/inducible degron) that capture the versatility of the system. In this regard, CRISPR based 'epigenome editing' will also serve as a valuable tool to determine causal functions of chromatin marks[246,247].

Furthermore, the conclusions drawn above pertain to a small developmental window modeled in stem cells that do not capture the full complexity of *in-utero* growth. Indeed, it is becoming evident that chromatin dynamics is context-specific and differs considerably across developmental stages. Though possible in mice, direct molecular analysis of the human post-implantation embryos is lacking due to ethical reasons. The alternative approach, fueled by synthetic biology is the use of ESCs to build simplified *in vitro* models of real embryos, called blastoids or synthetic embryos[248]. Blastoids resemble human blastocysts in terms of their morphology, composition and allocation of different cell

lineages making them an excellent system to validate our findings[249] and further our understanding of epigenetic control in multicellular development.

As a final point, ChIPs for histone modifications are routinely (including studies in this thesis) performed in native conditions but several protein-DNA interactions are transient, and efforts are underway to adapt MINUTE ChIP for fixed material arising from different sources and to reduce the amount of starting material required. Admittedly, our 'bulk-cell' method provides a population average measurement and does not capture the possible cell-to-cell heterogeneity. As discussed in the introduction, it is crucial to acknowledge the fallibility of extrinsic signaling required to maintain pluripotency *in vitro* and the resulting stochastic levels of various factors and epigenetic marks in an ESC population. Adoption of ChIP for single-cell applications has been attempted[219] but the efficiency of the immunoprecipitation step which is subject to abundance of epitope-antibody interactions, limits the data quality/complexity. Single-cell version of techniques such as scDamID[250], scChIC-seq[251], scCUT&TAG[222], that employ genetic or immunological targeting of enzymatic activity to specific genomic loci, provide the resolution required to resolve cell-state-specific enrichment of hPTMs[252]. However, normalization and consequent interpretation of data collected from such techniques gets convoluted by the multiple sources of noise in single cell measurements[6,253].

While having contributed to the Polycomb and development field, we envision that spatiotemporal analysis of protein DNA interactions in other biological processes would benefit from the multiplexing and quantitative essence of MINUTE ChIP.

6 ACKNOWLEDGEMENTS

If you are reading this, thank you for being part of my doctoral journey and know that your kindness in one form or another has truly impacted me!

*This section lacks a detailed account of your support owing to my aversion for writing texts but be rest assured that I will express my deepest gratitude to you, in person. Also, my sincere apologies to those whose names I might have forgotten to spell out here. *facepalm* Please hit me up and I will make it up to you!*

I would first like to thank my supervisor, **Simon,** for his belief in my abilities and work and for his constant support. I have truly enjoyed being exposed to your ingenious approach to problem solving and your contagious enthusiasm for research. I would like to thank my co-supervisors **Fredrik**, **Goncalo** and **Laura** for their encouragement and assistance. I would also like to acknowledge **Claudia**, my mentor for her valuable guidance and for being an amazing role model. I am forever grateful to my master thesis supervisors **Orsolya** and **Ola** for encouraging and preparing me for the PhD endeavour. Thanks are also in order to the HR team and the doctoral education committee (**Elias**, **Victoria, Linnea** and **Alessandra)** at MBB, for all the bureaucratic support I have received during my time at Karolinska.

I would like to thank my dissertation committee members **Mattias Mannervik**, **Joakim Lundeberg**, **Ola Söderberg**, my opponent **Aydan Bulut-Karslioglu**, and chairperson **Andreas Lennartsson** for finding the time to read my thesis and for lending their expertise to the defence proceedings.

I'm extremely grateful to the present (**Rozina, Birthe, Carmen, Lorenzo, Philip, Jing, Rui, Angelo, Dörte, Yuying, Chris, and Anna**) and former members (**Anna Maria, Rahul, Maria, Marianne, Lea, Daniel, Hannes, Ruth, Javi, Kyle, Laura, and Christian)** of the Elsässer lab, who have provided stimulating discussions as well as happy distractions to rest my mind outside of my research. Special thanks to Carmen and Angelo for proofreading this thesis at very short notice! I would also like to take this opportunity to thank the students (**Annika, Marieke, Jana, Daryl, Adrià, Hanna, Bruno, and Eva**) I have had the pleasure to work and learn with. The work amounting to this thesis would not have been possible without the contributions of co-authors and collaborators from the Lanner lab (**JP, Nerges, Chen, and Jere**).

My work environment at SciLifeLab (Alpha4) has been so very wonderful the past few years that I hardly ever left it, courtesy present, and past colleagues of the BiCro lab (**Tomek**, **Maud, Michelle, Reza, GG, Su, Eleni, Britta, Quim, and Masahiro**), OFC lab (**Alba, Valeria, Maria, Xuexin, Louise and Dani**), JB lab (**Asimina, Dimitris, Ann Sophie, and Jaime**) and fellow Indians from the KTH wing (**Tharagan, Vamakshi, and Sharath**). In particular, thank you Tomek and Alba for being my dream team and support system! I'd also like to thank everyone else I have had the pleasure of interacting with at SciLife, especially the gene expression journal club members and SciLife pub crew (**Marcel, Yerma, Gustavo, and Christian**).

And then to all my friends outside of work, thank you for your support and for being truly accommodating. Thanks to my childhood friends **Sam**, **Arun** and **Sonic** for keeping me sane, cushioning my PhD salary and for thinking that I was curing cancer. I'd also like to acknowledge my partner, **Juan** for his unwavering support and for enduring me at my worst. Finally, I'd like to thank my parents for being my relentless cheerleaders for life!

A special shout out to the band **God is an astronaut**, for making music that has fueled the writing of this thesis.

7 REFERENCES

1. Zentner, G. E. & Henikoff, S. Regulation of nucleosome dynamics by histone modifications. *Nat. Struct. Mol. Biol.* **20**, 259–266 (2013).

2. Jiang, C. & Pugh, B. F. Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.* **10**, 161–172 (2009).

3. Henikoff, S. Nucleosome destabilization in the epigenetic regulation of gene expression. *Nat. Rev. Genet.* **9**, 15–26 (2008).

4. Talbert, P. B. & Henikoff, S. Histone variants--ancient wrap artists of the epigenome. *Nat. Rev. Mol. Cell Biol.* **11**, 264–275 (2010).

5. Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* **17**, 487–500 (2016).

6. Cuvier, O. & Fierz, B. Dynamic chromatin technologies: from individual molecules to epigenomic regulation in cells. *Nat. Rev. Genet.* **18**, 457–472 (2017).

7. Waddington, C. H. *The strategy of the genes*. (Routledge, 1957). doi:10.4324/9781315765471.

8. Elewa, A. A Waddington Epigenetic Landscape for the *C. elegans* embryo. *BioRxiv* (2020) doi:10.1101/2020.01.01.892414.

9. Wang, H. & Dey, S. K. Roadmap to embryo implantation: clues from mouse models. *Nat. Rev. Genet.* **7**, 185–199 (2006).

10. Rossant, J. & Tam, P. P. L. Early human embryonic development: Blastocyst formation to gastrulation. *Dev. Cell* **57**, 152–165 (2022).

11. Molè, M. A., Weberling, A. & Zernicka-Goetz, M. Comparative analysis of human and mouse development: From zygote to pre-gastrulation. *Curr. Top. Dev. Biol.* **136**, 113–138 (2020).

12. Rossant, J. & Tam, P. P. L. New Insights into Early Human Development: Lessons for Stem Cell Derivation and Differentiation. *Cell Stem Cell* **20**, 18–28 (2017).

13. Weinberger, L., Ayyash, M., Novershtern, N. & Hanna, J. H. Dynamic stem cell states: naive to primed pluripotency in rodents and humans. *Nat. Rev. Mol. Cell Biol.* **17**, 155–169 (2016).

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

15. Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci USA* **78**, 7634–7638 (1981).

16. Williams, R. L. *et al.* Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336**, 684–687 (1988).

17. Smith, A. G. *et al.* Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336**, 688–690 (1988).

18. Buehr, M. *et al.* Capture of authentic embryonic stem cells from rat blastocysts. *Cell* **135**, 1287–1298 (2008).

19. Li, P. *et al.* Germline competent embryonic stem cells derived from rat blastocysts. *Cell* **135**, 1299–1310 (2008).

20. Chen, H. *et al.* Erk signaling is indispensable for genomic stability and self-renewal of mouse embryonic stem cells. *Proc Natl Acad Sci USA* **112**, E5936-43 (2015).

21. Ying, Q.-L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519–523 (2008).

22. Tesar, P. J. *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**, 196–199 (2007).

23. Brons, I. G. M. *et al.* Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448**, 191–195 (2007).

24. Kim, H. *et al.* Modulation of β-catenin function maintains mouse epiblast stem cell and human embryonic stem cell self-renewal. *Nat. Commun.* **4**, 2403 (2013).

25. Wu, J. *et al.* An alternative pluripotent state confers interspecies chimaeric competency. *Nature* **521**, 316–321 (2015).

26. Kojima, Y. *et al.* The transcriptional and functional properties of mouse epiblast stem cells resemble the anterior primitive streak. *Cell Stem Cell* **14**, 107–120 (2014).

27. Han, D. W. *et al.* Epiblast stem cell subpopulations represent mouse embryos of distinct pregastrulation stages. *Cell* **143**, 617–627 (2010).

28. Fenelon, J. C., Banerjee, A. & Murphy, B. D. Embryonic diapause: development on hold. *Int. J. Dev. Biol.* **58**, 163–174 (2014).

29. Scognamiglio, R. *et al.* Myc depletion induces a pluripotent dormant state mimicking diapause. *Cell* **164**, 668–680 (2016).

30. Bulut-Karslioglu, A. *et al.* Inhibition of mTOR induces a paused pluripotent state. *Nature* **540**, 119–123 (2016).

31. Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).

32. Nichols, J. & Smith, A. Naive and primed pluripotent states. *Cell Stem Cell* **4**, 487–492 (2009).

33. Dong, C., Fischer, L. A. & Theunissen, T. W. Recent insights into the naïve state of human pluripotency and its applications. *Exp. Cell Res.* **385**, 111645 (2019).

34. Hanna, J. *et al.* Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. *Proc Natl Acad Sci USA* **107**, 9222–9227 (2010).

35. Wang, W. *et al.* Rapid and efficient reprogramming of somatic cells to induced pluripotent stem cells by retinoic acid receptor gamma and liver receptor homolog 1. *Proc Natl Acad Sci USA* **108**, 18283–18288 (2011).

36. Gafni, O. *et al.* Derivation of novel human ground state naive pluripotent stem cells. *Nature* **504**, 282–286 (2013).

37. Theunissen, T. W. *et al.* Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell* **15**, 471–487 (2014).

38. Takashima, Y. *et al.* Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell* **158**, 1254–1269 (2014).

39. Guo, G. *et al.* Naive Pluripotent Stem Cells Derived Directly from Isolated Cells of the Human Inner Cell Mass. *Stem Cell Reports* **6**, 437–446 (2016).

40. Theunissen, T. W. *et al.* Molecular criteria for defining the naive human pluripotent state. *Cell Stem Cell* **19**, 502–515 (2016).

41. Guo, G. *et al.* Epigenetic resetting of human pluripotency. *Development* **144**, 2748–2763 (2017).

42. Dong, C. *et al.* Derivation of trophoblast stem cells from naïve human pluripotent stem cells. *eLife* **9**, (2020).

43. Ware, C. B. Concise Review: Lessons from Naïve Human Pluripotent Cells. *Stem Cells* **35**, 35–41 (2017).

44. Leitch, H. G. *et al.* Naive pluripotency is associated with global DNA hypomethylation. *Nat. Struct. Mol. Biol.* **20**, 311–316 (2013).

45. Hackett, J. A. *et al.* Synergistic mechanisms of DNA demethylation during transition to ground-state pluripotency. *Stem Cell Reports* **1**, 518–531 (2013).

46. Lyon, M. F. Gene action in the X-chromosome of the mouse (Mus musculus L.). *Nature* **190**, 372–373 (1961).

47. Okamoto, I., Otte, A. P., Allis, C. D., Reinberg, D. & Heard, E. Epigenetic dynamics of imprinted X inactivation during early mouse development. *Science* **303**, 644–649 (2004).

48. Petropoulos, S. *et al.* Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell* **167**, 285 (2016).

49. Okamoto, I. *et al.* Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. *Nature* **472**, 370–374 (2011).

50. Vallot, C. *et al.* Erosion of X Chromosome Inactivation in Human Pluripotent Cells Initiates with XACT Coating and Depends on a Specific Heterochromatin Landscape. *Cell Stem Cell* **16**, 533–546 (2015).

51. Friedli, M. & Trono, D. The developmental control of transposable elements and the evolution of higher species. *Annu. Rev. Cell Dev. Biol.* **31**, 429–451 (2015).

52. Grow, E. J. *et al.* Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* **522**, 221–225 (2015).

53. He, J. *et al.* Transposable elements are regulated by context-specific patterns of chromatin marks in mouse embryonic stem cells. *Nat. Commun.* **10**, 34 (2019).

54. Macfarlan, T. S. *et al.* Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57–63 (2012).

55. Fu, H. *et al.* Elevated retrotransposon activity and genomic instability in primed pluripotent stem cells. *Genome Biol.* **22**, 201 (2021).

56. Yousefi, M., Marashi, S.-A., Sharifi-Zarchi, A. & Taleahmad, S. The metabolic network model of primed/naive human embryonic stem cells underlines the importance of oxidation-reduction potential and tryptophan metabolism in primed pluripotency. *Cell Biosci.* **9**, 71 (2019).

57. Zhou, W. *et al.* HIF1α induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition. *EMBO J.* **31**, 2103–2116 (2012).

58. Gu, W. *et al.* Glycolytic metabolism plays a functional role in regulating human pluripotent stem cell state. *Cell Stem Cell* **19**, 476–490 (2016).

59. Shyh-Chang, N. & Ng, H.-H. The metabolic programming of stem cells. *Genes Dev.* **31**, 336–346 (2017).

60. Sperber, H. *et al.* The metabolome regulates the epigenetic landscape during naive-to-primed human embryonic stem cell transition. *Nat. Cell Biol.* **17**, 1523–1535 (2015).

61. Coronado, D. *et al.* A short G1 phase is an intrinsic determinant of naïve embryonic stem cell pluripotency. *Stem Cell Res.* **10**, 118–131 (2013).

62. Roccio, M. *et al.* Predicting stem cell fate changes by differential cell cycle progression patterns. *Development* **140**, 459–470 (2013).

63. Savatier, P., Lapillonne, H., Jirmanova, L., Vitelli, L. & Samarut, J. Analysis of the cell cycle in mouse embryonic stem cells. *Methods Mol. Biol.* **185**, 27–33 (2002).

64. Orford, K. W. & Scadden, D. T. Deconstructing stem cell self-renewal: genetic insights into cell-cycle regulation. *Nat. Rev. Genet.* **9**, 115–128 (2008).

65. Pauklin, S. & Vallier, L. The cell-cycle state of stem cells determines cell fate propensity. *Cell* **155**, 135–147 (2013).

66. Waisman, A. *et al.* Cell cycle dynamics of mouse embryonic stem cells in the ground state and during transition to formative pluripotency. *Sci. Rep.* **9**, 8051 (2019).

67. Filipczyk, A. A., Laslett, A. L., Mummery, C. & Pera, M. F. Differentiation is coupled to changes in the cell cycle regulatory apparatus of human embryonic stem cells. *Stem Cell Res.* **1**, 45–60 (2007).

68. Collier, A. J. *et al.* Comprehensive cell surface protein profiling identifies specific markers of human naive and primed pluripotent states. *Cell Stem Cell* **20**, 874-890.e7 (2017).

69. De Clerck, L. *et al.* Untargeted histone profiling during naive conversion uncovers conserved modification markers between mouse and human. *Sci. Rep.* **9**, 17240 (2019).

70. Wojdyla, K. *et al.* Cell-Surface Proteomics Identifies Differences in Signaling and Adhesion Protein Expression between Naive and Primed Human Pluripotent Stem Cells. *Stem Cell Reports* **14**, 972–988 (2020).

71. Sonnen, K. F. & Janda, C. Y. Signalling dynamics in embryonic development. *Biochem. J.* **478**, 4045–4070 (2021).

72. Sanz-Ezquerro, J. J., Münsterberg, A. E. & Stricker, S. Editorial: signaling pathways in embryonic development. *Front. Cell Dev. Biol.* **5**, 76 (2017).

73. Wang, Y. *et al.* Entire mitogen activated protein kinase (MAPK) pathway is present in preimplantation mouse embryos. *Dev. Dyn.* **231**, 72–87 (2004).

74. Cockburn, K., Biechele, S., Garner, J. & Rossant, J. The Hippo pathway member Nf2 is required for inner cell mass specification. *Curr. Biol.* **23**, 1195–1201 (2013).

75. Roode, M. *et al.* Human hypoblast formation is not dependent on FGF signalling. *Dev. Biol.* **361**, 358–363 (2012).

76. Bertero, A. *et al.* Activin/nodal signaling and NANOG orchestrate human embryonic stem cell fate decisions by controlling the H3K4me3 chromatin mark. *Genes Dev.* **29**, 702–717 (2015).

77. Linneberg-Agerholm, M. *et al.* Naïve human pluripotent stem cells respond to Wnt, Nodal and LIF signalling to produce expandable naïve extra-embryonic endoderm. *Development* **146**, (2019).

78. Vallier, L., Alexander, M. & Pedersen, R. A. Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *J. Cell Sci.* **118**, 4495–4509 (2005).

79. Xu, R.-H. *et al.* BMP4 initiates human embryonic stem cell differentiation to trophoblast. *Nat. Biotechnol.* **20**, 1261–1264 (2002).

80. Amita, M. *et al.* Complete and unidirectional conversion of human embryonic stem cells to trophoblast by BMP4. *Proc Natl Acad Sci USA* **110**, E1212-21 (2013).

81. Silva, J. & Smith, A. Capturing pluripotency. *Cell* **132**, 532–536 (2008).

82. Li, M. & Izpisua Belmonte, J. C. Deconstructing the pluripotency gene regulatory network. *Nat. Cell Biol.* **20**, 382–392 (2018).

83. Morey, L., Santanach, A. & Di Croce, L. Pluripotency and epigenetic factors in mouse embryonic stem cell fate regulation. *Mol. Cell. Biol.* **35**, 2716–2728 (2015).

84. Loh, K. M. & Lim, B. A precarious balance: pluripotency factors as lineage specifiers. *Cell Stem Cell* **8**, 363–369 (2011).

85. Blakeley, P. *et al.* Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development* **142**, 3151–3165 (2015).

86. Petropoulos, S., Panula, S. P., Schell, J. P. & Lanner, F. Single-cell RNA sequencing: revealing human pre-implantation development, pluripotency and germline development. *J. Intern. Med.* **280**, 252–264 (2016).

87. Yan, L. *et al.* Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1131–1139 (2013).

88. Smith, A. The battlefield of pluripotency. *Cell* **123**, 757–760 (2005).

89. Niakan, K. K. & Eggan, K. Analysis of human embryos from zygote to blastocyst reveals distinct gene expression patterns relative to the mouse. *Dev. Biol.* **375**, 54–64 (2013).

90. Pastor, W. A. *et al.* TFAP2C regulates transcription in human naive pluripotency by opening enhancers. *Nat. Cell Biol.* **20**, 553–564 (2018).

91. Bai, Q. *et al.* Dissecting the first transcriptional divergence during human embryonic development. *Stem Cell Rev.* **8**, 150–162 (2012).

92. Zheng, Y. *et al.* Controlled modelling of human epiblast and amnion development using stem cells. *Nature* **573**, 421–425 (2019).

93. Tyser, R. C. V. *et al.* Single-cell transcriptomic characterization of a gastrulating human embryo. *Nature* **600**, 285–289 (2021).

94. Dunn, S. J., Martello, G., Yordanov, B., Emmott, S. & Smith, A. G. Defining an essential transcription factor program for naïve pluripotency. *Science* **344**, 1156–1160 (2014).

95. Takahashi, S., Kobayashi, S. & Hiratani, I. Epigenetic differences between naïve and primed pluripotent stem cells. *Cell. Mol. Life Sci.* **75**, 1191–1203 (2018).

96. Truax, A. D. & Greer, S. F. ChIP and Re-ChIP assays: investigating interactions between regulatory proteins, histone modifications, and the DNA sequences to which they bind. *Methods Mol. Biol.* **809**, 175–188 (2012).

97. Zheng, H. & Xie, W. The role of 3D genome organization in development and cell differentiation. *Nat. Rev. Mol. Cell Biol.* **20**, 535–550 (2019).

98. Yeom, Y. I. *et al.* Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* **122**, 881–894 (1996).

99. Factor, D. C. *et al.* Epigenomic comparison reveals activation of "seed" enhancers during transition from naive to primed pluripotency. *Cell Stem Cell* **14**, 854–863 (2014).

100. Hendrich, B. & Bird, A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.* **18**, 6538–6547 (1998).

101. Spruijt, C. G. *et al.* Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146–1159 (2013).

102. Yin, Y. *et al.* Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**, (2017).

103. Smith, Z. D. *et al.* A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484**, 339–344 (2012).

104. Reik, W., Dean, W. & Walter, J. Epigenetic reprogramming in mammalian development. *Science* **293**, 1089–1093 (2001).

105. Burton, A. & Torres-Padilla, M.-E. Epigenetic reprogramming and development: a unique heterochromatin organization in the preimplantation mouse embryo. *Brief. Funct. Genomics* **9**, 444–454 (2010).

106. Greenberg, M. V. C. & Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* **20**, 590–607 (2019).

107. Brinkman, A. B. *et al.* Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res.* **22**, 1128–1138 (2012).

108. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14**, 204–220 (2013).

109. Smith, Z. D. *et al.* DNA methylation dynamics of the human preimplantation embryo. *Nature* **511**, 611–615 (2014).

110. Luger, K. & Hansen, J. C. Nucleosome and chromatin fiber dynamics. *Curr. Opin. Struct. Biol.* **15**, 188–196 (2005).

111. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).

112. Franklin, K. A., Shields, C. E. & Haynes, K. A. Beyond the marks: reader-effectors as drivers of epigenetics and chromatin engineering. *Trends Biochem. Sci.* **47**, 417–432 (2022).

113. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).

114. Kassis, J. A., Kennison, J. A. & Tamkun, J. W. Polycomb and trithorax group genes in *Drosophila*. *Genetics* **206**, 1699–1725 (2017).

115. Schuettengruber, B., Bourbon, H.-M., Di Croce, L. & Cavalli, G. Genome regulation by polycomb and trithorax: 70 years and counting. *Cell* **171**, 34–57 (2017).

116. Healy, E. & Bracken, A. P. If You Like It Then You Shoulda Put Two "RINGs" on It: Delineating the Roles of vPRC1 and cPRC1. *Mol. Cell* **77**, 685–687 (2020).

117. Gao, Z. *et al.* PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol. Cell* **45**, 344–356 (2012).

118. van Mierlo, G., Veenstra, G. J. C., Vermeulen, M. & Marks, H. The complexity of PRC2 subcomplexes. *Trends Cell Biol.* **29**, 660–671 (2019).

119. Blackledge, N. P. *et al.* PRC1 catalytic activity is central to polycomb system function. *Mol. Cell* **77**, 857-874.e9 (2020).

120. Blackledge, N. P., Rose, N. R. & Klose, R. J. Targeting Polycomb systems to regulate gene expression: modifications to a complex story. *Nat. Rev. Mol. Cell Biol.* **16**, 643–649 (2015).

121. Scheuermann, J. C. *et al.* Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature* **465**, 243–247 (2010).

122. Margueron, R. *et al.* Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Mol. Cell* **32**, 503–518 (2008).

123. Agger, K. *et al.* UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* **449**, 731–734 (2007).

124. Bracken, A. P., Dietrich, N., Pasini, D., Hansen, K. H. & Helin, K. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev.* **20**, 1123–1136 (2006).

125. Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349–353 (2006).

126. Blackledge, N. P. & Klose, R. J. Getting under the skin of Polycomb-dependent gene regulation. *Genes Dev.* **35**, 301–303 (2021).

127. Cooper, S. *et al.* Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. *Cell Rep.* **7**, 1456–1470 (2014).

128. Wang, L. *et al.* Hierarchical recruitment of polycomb group silencing complexes. *Mol. Cell* **14**, 637–646 (2004).

129. Tavares, L. *et al.* RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell* **148**, 664–678 (2012).

130. Kalb, R. *et al.* Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat. Struct. Mol. Biol.* **21**, 569–571 (2014).

131. McLaughlin, K. *et al.* DNA Methylation Directs Polycomb-Dependent 3D Genome Re-organization in Naive Pluripotency. *Cell Rep.* **29**, 1974-1985.e6 (2019).

132. Denholtz, M. *et al.* Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. *Cell Stem Cell* **13**, 602–616 (2013).

133. Eskeland, R. *et al.* Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol. Cell* **38**, 452–464 (2010).

134. Riising, E. M. *et al.* Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. *Mol. Cell* **55**, 347–360 (2014).

135. Hosogane, M., Funayama, R., Nishida, Y., Nagashima, T. & Nakayama, K. Ras-induced changes in H3K27me3 occur after those in transcriptional activity. *PLoS Genet.* **9**, e1003698 (2013).

136. Klose, R. J., Cooper, S., Farcas, A. M., Blackledge, N. P. & Brockdorff, N. Chromatin sampling--an emerging perspective on targeting polycomb repressor proteins. *PLoS Genet.* **9**, e1003717 (2013).

137. Blackledge, N. P. & Klose, R. J. The molecular principles of gene regulation by Polycomb repressive complexes. *Nat. Rev. Mol. Cell Biol.* **22**, 815–833 (2021).

138. Zhao, X. D. *et al.* Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. *Cell Stem Cell* **1**, 286–298 (2007).

139. Leeb, M. *et al.* Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes Dev.* **24**, 265–276 (2010).

140. Lee, T. I. *et al.* Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301–313 (2006).

141. Aloia, L., Di Stefano, B. & Di Croce, L. Polycomb complexes in stem cells and embryonic development. *Development* **140**, 2525–2534 (2013).

142. Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343–349 (2011).

143. Endoh, M. *et al.* Histone H2A mono-ubiquitination is a crucial step to mediate PRC1-dependent repression of developmental genes to maintain ES cell identity. *PLoS Genet.* **8**, e1002774 (2012).

144. Endoh, M. *et al.* Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. *Development* **135**, 1513–1524 (2008).

145. Piunti, A. & Shilatifard, A. The roles of Polycomb repressive complexes in mammalian development and cancer. *Nat. Rev. Mol. Cell Biol.* **22**, 326–345 (2021).

146. Krogan, N. J. *et al.* COMPASS, a histone H3 (Lysine 4) methyltransferase required for telomeric silencing of gene expression. *J. Biol. Chem.* **277**, 10753–10755 (2002).

147. Miller, T. *et al.* COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proc Natl Acad Sci USA* **98**, 12902–12907 (2001).

148. Piunti, A. & Shilatifard, A. Epigenetic balance of gene expression by Polycomb and COMPASS families. *Science* **352**, aad9780 (2016).

149. FitzGerald, K. T. & Diaz, M. O. MLL2: A new mammalian member of the trx/MLL family of genes. *Genomics* **59**, 187–192 (1999).

150. Lee, J.-H., Tate, C. M., You, J.-S. & Skalnik, D. G. Identification and characterization of the human Set1B histone H3-Lys4 methyltransferase complex. *J. Biol. Chem.* **282**, 13419–13428 (2007).

151. Lee, J.-E. *et al.* H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. *eLife* **2**, e01503 (2013).

152. Hyun, K., Jeon, J., Park, K. & Kim, J. Writing, erasing and reading histone lysine methylations. *Exp. Mol. Med.* **49**, e324 (2017).

153. Hughes, A. L., Kelley, J. R. & Klose, R. J. Understanding the interplay between CpG island-associated gene promoters and H3K4 methylation. *Biochim. Biophys. Acta Gene Regul. Mech.* **1863**, 194567 (2020).

154. Voo, K. S., Carlone, D. L., Jacobsen, B. M., Flodin, A. & Skalnik, D. G. Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. *Mol. Cell. Biol.* **20**, 2108–2121 (2000).

155. Kusch, T. Histone H3 lysine 4 methylation revisited. *Transcription* **3**, 310–314 (2012).

156. Brown, D. A. *et al.* The SET1 Complex Selects Actively Transcribed Target Genes via Multivalent Interaction with CpG Island Chromatin. *Cell Rep.* **20**, 2313–2327 (2017).

157. Lee, M. G. *et al.* Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. *Science* **318**, 447–450 (2007).

158. Morselli, M. *et al.* In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. *eLife* **4**, e06205 (2015).

159. Binda, O. *et al.* Trimethylation of histone H3 lysine 4 impairs methylation of histone H3 lysine 9: regulation of lysine methyltransferases by physical interaction with their substrates. *Epigenetics* **5**, 767–775 (2010).

160. Lohse, B. *et al.* Posttranslational modifications of the histone 3 tail and their impact on the activity of histone lysine demethylases in vitro. *PLoS ONE* **8**, e67653 (2013).

161. Schmitges, F. W. *et al.* Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol. Cell* **42**, 330–341 (2011).

162. Zegerman, P., Canas, B., Pappin, D. & Kouzarides, T. Histone H3 lysine 4 methylation disrupts binding of nucleosome remodeling and deacetylase (NuRD) repressor complex. *J. Biol. Chem.* **277**, 11621–11624 (2002).

163. Koutelou, E., Hirsch, C. L. & Dent, S. Y. R. Multiple faces of the SAGA complex. *Curr. Opin. Cell Biol.* **22**, 374–382 (2010).

164. Wysocka, J. *et al.* A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442**, 86–90 (2006).

165. Vermeulen, M. *et al.* Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131**, 58–69 (2007).

166. Lauberth, S. M. *et al.* H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell* **152**, 1021–1036 (2013).

167. Denissov, S. *et al.* Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. *Development* **141**, 526–537 (2014).

168. Milne, T. A. *et al.* MLL targets SET domain methyltransferase activity to Hox gene promoters. *Mol. Cell* **10**, 1107–1117 (2002).

169. Hu, D. *et al.* Not all H3K4 methylations are created equal: mll2/compass dependency in primordial germ cell specification. *Mol. Cell* **65**, 460-475.e6 (2017).

170. Wang, C. *et al.* Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition. *Proc Natl Acad Sci USA* **113**, 11871–11876 (2016).

171. Marks, H. *et al.* The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**, 590–604 (2012).

172. Hawkins, R. D. *et al.* Dynamic chromatin states in human ES cells reveal potential regulatory sequences and genes involved in pluripotency. *Cell Res.* **21**, 1393–1409 (2011).

173. Atlasi, Y. & Stunnenberg, H. G. The interplay of epigenetic marks during stem cell differentiation and development. *Nat. Rev. Genet.* **18**, 643–658 (2017).

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

175. King, A. D. *et al.* Reversible regulation of promoter and enhancer histone landscape by DNA methylation in mouse embryonic stem cells. *Cell Rep.* **17**, 289–302 (2016).

176. Voigt, P., Tee, W.-W. & Reinberg, D. A double take on bivalent promoters. *Genes Dev.* **27**, 1318–1338 (2013).

177. Vastenhouw, N. L. & Schier, A. F. Bivalent histone modifications in early embryogenesis. *Curr. Opin. Cell Biol.* **24**, 374–386 (2012).

178. Jadhav, U. *et al.* Acquired Tissue-Specific Promoter Bivalency Is a Basis for PRC2 Necessity in Adult Cells. *Cell* **165**, 1389–1400 (2016).

179. Shah, R. N. *et al.* Re-evaluating the role of nucleosomal bivalency in early development. *BioRxiv* (2021) doi:10.1101/2021.09.09.458948.

180. Mas, G. *et al.* Promoter bivalency favors an open chromatin architecture in embryonic stem cells. *Nat. Genet.* **50**, 1452–1462 (2018).

181. Weiner, A. *et al.* Co-ChIP enables genome-wide mapping of histone mark co-occurrence at single-molecule resolution. *Nat. Biotechnol.* **34**, 953–961 (2016).

182. Shandilya, J. & Roberts, S. G. E. The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. *Biochim. Biophys. Acta* **1819**, 391–400 (2012).

183. Phatnani, H. P. & Greenleaf, A. L. Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.* **20**, 2922–2936 (2006).

184. Muse, G. W. *et al.* RNA polymerase is poised for activation across the genome. *Nat. Genet.* **39**, 1507–1511 (2007).

185. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848 (2008).

186. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13**, 720–731 (2012).

187. Gaertner, B. & Zeitlinger, J. RNA polymerase II pausing during development. *Development* **141**, 1179–1183 (2014).

188. Rahl, P. B. *et al.* c-Myc regulates transcriptional pause release. *Cell* **141**, 432–445 (2010).

189. Gates, L. A. *et al.* Acetylation on histone H3 lysine 9 mediates a switch from transcription initiation to elongation. *J. Biol. Chem.* **292**, 14456–14472 (2017).

190. Gates, L. A., Foulds, C. E. & O'Malley, B. W. Histone marks in the "driver's seat": functional roles in steering the transcription cycle. *Trends Biochem. Sci.* **42**, 977–989 (2017).

191. Wen, H. *et al.* ZMYND11 links histone H3.3K36me3 to transcription elongation and tumour suppression. *Nature* **508**, 263–268 (2014).

192. Neri, F. *et al.* Intragenic DNA methylation prevents spurious transcription initiation. *Nature* **543**, 72–77 (2017).

193. Ferrari, K. J. *et al.* Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. *Mol. Cell* **53**, 49–62 (2014).

194. Erickson, B., Sheridan, R. M., Cortazar, M. & Bentley, D. L. Dynamic turnover of paused Pol II complexes at human promoters. *Genes Dev.* **32**, 1215–1225 (2018).

195. Mantsoki, A., Devailly, G. & Joshi, A. Dynamics of promoter bivalency and RNAP II pausing in mouse stem and differentiated cells. *BMC Dev. Biol.* **18**, 2 (2018).

196. Liu, J., Wu, X., Zhang, H., Pfeifer, G. P. & Lu, Q. Dynamics of RNA Polymerase II Pausing and Bivalent Histone H3 Methylation during Neuronal Differentiation in Brain Development. *Cell Rep.* **20**, 1307–1318 (2017).

197. Williams, L. H. *et al.* Pausing of RNA polymerase II regulates mammalian developmental potential through control of signaling networks. *Mol. Cell* **58**, 311–322 (2015).

198. Kwak, H., Fuda, N. J., Core, L. J. & Lis, J. T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339**, 950–953 (2013).

199. Nojima, T. *et al.* Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* **161**, 526–540 (2015).

200. Schwalb, B. *et al.* TT-seq maps the human transient transcriptome. *Science* **352**, 1225–1228 (2016).

201. Gilmour, D. S. & Lis, J. T. Detecting protein-DNA interactions in vivo: distribution of RNA polymerase on specific bacterial genes. *Proc Natl Acad Sci USA* **81**, 4275–4279 (1984).

202. Solomon, M. J., Larsen, P. L. & Varshavsky, A. Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. *Cell* **53**, 937–947 (1988).

203. Hecht, A., Strahl-Bolsinger, S. & Grunstein, M. Spreading of transcriptional repressor SIR3 from telomeric heterochromatin. *Nature* **383**, 92–96 (1996).

204. Ren, B. *et al.* Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306–2309 (2000).

205. Marinov, G. K. A decade of ChIP-seq. *Brief. Funct. Genomics* **17**, 77–79 (2018).

206. Wei, C.-L. *et al.* A global map of p53 transcription-factor binding sites in the human genome. *Cell* **124**, 207–219 (2006).

207. McKernan, K. J. *et al.* Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res.* **19**, 1527–1541 (2009).

208. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).

209. Harris, T. D. *et al.* Single-molecule DNA sequencing of a viral genome. *Science* **320**, 106–109 (2008).

210. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).

211. Shendure, J. *et al.* Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**, 1728–1732 (2005).

212. Park, P. J. ChIP-seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.* **10**, 669–680 (2009).

213. He, Q., Johnston, J. & Zeitlinger, J. ChIP-nexus enables improved detection of in vivo transcription factor binding footprints. *Nat. Biotechnol.* **33**, 395–401 (2015).

214. Rhee, H. S. & Pugh, B. F. ChIP-exo method for identifying genomic location of DNA-binding proteins with near-single-nucleotide accuracy. *Curr. Protoc. Mol. Biol.* **Chapter 21**, Unit 21.24 (2012).

215. Venters, B. J. Insights from resolving protein-DNA interactions at near base-pair resolution. *Brief. Funct. Genomics* **17**, 80–88 (2018).

216. Dahl, J. A. & Gilfillan, G. D. How low can you go? Pushing the limits of low-input ChIP-seq. *Brief. Funct. Genomics* **17**, 89–95 (2018).

217. Wallerman, O., Nord, H., Bysani, M., Borghini, L. & Wadelius, C. lobChIP: from cells to sequencing ready ChIP libraries in a single day. *Epigenetics Chromatin* **8**, 25 (2015).

218. Zheng, X. *et al.* Low-Cell-Number Epigenome Profiling Aids the Study of Lens Aging and Hematopoiesis. *Cell Rep.* **13**, 1505–1518 (2015).

219. Rotem, A. *et al.* Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat. Biotechnol.* **33**, 1165–1172 (2015).

220. Schmid, M., Durussel, T. & Laemmli, U. K. ChIC and ChEC; genomic mapping of chromatin proteins. *Mol. Cell* **16**, 147–157 (2004).

221. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, (2017).

222. Kaya-Okur, H. S. *et al.* CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1930 (2019).

223. Policastro, R. A. & Zentner, G. E. Enzymatic methods for genome-wide profiling of protein binding sites. *Brief. Funct. Genomics* **17**, 138–145 (2018).

224. Pepke, S., Wold, B. & Mortazavi, A. Computation for ChIP-seq and RNA-seq studies. *Nat. Methods* **6**, S22-32 (2009).

225. de Boer, B. A. *et al.* OccuPeak: ChIP-Seq peak calling based on internal background modelling. *PLoS ONE* **9**, e99844 (2014).

226. Jiang, S. & Mortazavi, A. Integrating ChIP-seq with other functional genomics data. *Brief. Funct. Genomics* **17**, 104–115 (2018).

227. Guertin, M. J., Cullen, A. E., Markowetz, F. & Holding, A. N. Parallel factor ChIP provides essential internal control for quantitative differential ChIP-seq. *Nucleic Acids Res.* **46**, e75 (2018).

228. Lara-Astiaso, D. *et al.* Immunogenetics. Chromatin state dynamics during blood formation. *Science* **345**, 943–949 (2014).

229. Chabbert, C. D. *et al.* A high-throughput ChIP-Seq for large-scale chromatin studies. *Mol. Syst. Biol.* **11**, 777 (2015).

230. van Galen, P. *et al.* A multiplexed system for quantitative comparisons of chromatin landscapes. *Mol. Cell* **61**, 170–180 (2016).

231. Schmidl, C., Rendeiro, A. F., Sheffield, N. C. & Bock, C. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat. Methods* **12**, 963–965 (2015).

232. Bonhoure, N. *et al.* Quantifying ChIP-seq data: a spiking method providing an internal reference for sample-to-sample normalization. *Genome Res.* **24**, 1157–1168 (2014).

233. Orlando, D. A. *et al.* Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. *Cell Rep.* **9**, 1163–1170 (2014).

234. Grzybowski, A. T., Chen, Z. & Ruthenburg, A. J. Calibrating ChIP-Seq with Nucleosomal Internal Standards to Measure Histone Modification Density Genome Wide. *Mol. Cell* **58**, 886–899 (2015).

235. Navarro, C., Martin, M. & Elsässer, S. minute: A MINUTE-ChIP data analysis workflow. *BioRxiv* (2022) doi:10.1101/2022.03.14.484318.

236. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

237. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).

238. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

239. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j.* **17**, 10 (2011).

240. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187-91 (2014).

241. Kumar, B. & Elsässer, S. J. Quantitative Multiplexed ChIP Reveals Global Alterations that Shape Promoter Bivalency in Ground State Embryonic Stem Cells. *Cell Rep.* **28**, 3274-3284.e5 (2019).

242. Shan, Y. *et al.* PRC2 specifies ectoderm lineages and maintains pluripotency in primed but not naïve ESCs. *Nat. Commun.* **8**, 672 (2017).

243. Moody, J. D. *et al.* First critical repressive H3K27me3 marks in embryonic stem cells identified using designed protein inhibitor. *Proc Natl Acad Sci USA* **114**, 10125–10130 (2017).

244. Shao, R. *et al.* Distinct transcription kinetics of pluripotent cell states. *Mol. Syst. Biol.* **18**, e10407 (2022).

245. Bulut-Karslioglu, A. *et al.* The transcriptionally permissive chromatin state of embryonic stem cells is acutely tuned to translational output. *Cell Stem Cell* **22**, 369-383.e8 (2018).

246. Stricker, S. H., Köferle, A. & Beck, S. From profiles to function in epigenomics. *Nat. Rev. Genet.* **18**, 51–66 (2017).

247. Dominguez, A. A., Lim, W. A. & Qi, L. S. Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. *Nat. Rev. Mol. Cell Biol.* **17**, 5–15 (2016).

248. Shahbazi, M. N. & Zernicka-Goetz, M. Deconstructing and reconstructing the mouse and human early embryo. *Nat. Cell Biol.* **20**, 878–887 (2018).

249. De Santis, R. & Brivanlou, A. H. The treasure inside human naive pluripotency, generation of trophectoderm and blastoids. *Cell Stem Cell* **28**, 985–987 (2021).

250. Kind, J. *et al.* Genome-wide maps of nuclear lamina interactions in single human cells. *Cell* **163**, 134–147 (2015).

251. Ku, W. L. *et al.* Single-cell chromatin immunocleavage sequencing (scChIC-seq) to profile histone modification. *Nat. Methods* **16**, 323–325 (2019).

252. Carter, B. & Zhao, K. The epigenetic basis of cellular heterogeneity. *Nat. Rev. Genet.* **22**, 235–250 (2021).

253. Schwartzman, O. & Tanay, A. Single-cell epigenomics: techniques and emerging applications. *Nat. Rev. Genet.* **16**, 716–726 (2015).