1 Transcriptional networks are dynamically regulated during cell cycle

2 progression in human Pluripotent Stem Cells

3

4 Anna Osnato^{1, 2}, Ludovic Vallier^{1, 2, *}

5 ¹Wellcome – MRC Cambridge Stem Cell Institute, Jeffrey Cheah Biomedical Centre, University of Cambridge, CB2 0AW,
6 Cambridge, UK

7 ²Department of Surgery, University of Cambridge, Cambridge, UK

8 ^{*}Corresponding author (lv225@cam.ac.uk)

Abstract

Cell cycle progression follows a precise sequence of events marked by different phases and check points which are associated with specific chromatin organisation. Whilst these changes have been extensively studied, their consequences on transcriptional networks remain to be fully uncovered, especially in dynamic model systems such as stem cells. Here, we take advantage of the FUCCI reporter system to show that chromatin accessibility, gene expression and key transcription factors binding change during cell cycle progression in human Embryonic Stem Cells (hESCs). These analyses reveal that core pluripotency factors such as OCT4, NANOG and SOX2 but also chromatin remodelers such as CTCF and RING1B bind the genome at specific phases of the cell cycle. Importantly, this binding pattern allows differentiation in the G1 phase while preserving pluripotency in the S/G2/M. Our results highlight the importance of studying transcriptional and epigenetic regulations in the dynamic context of the cell cycle.

Introduction

Cell cycle regulation is essential to control the number of cells generated during development and during organ homeostasis, while dysregulation of proliferation can result in disease. Importantly, recent studies have suggested that cell cycle regulation could go beyond its canonical function in controlling cell division initiation. Indeed, during cell cycle chromatin alternates between an invariant mitotic conformation and a cell-type specific interphase organisation during cell cycle progression¹. Loops and Topological-Associated Domains (TADs) insulation emerge very early in the G1 phase, but A/B compartments reappear only during the S phase². Global chromatin de-condense in the G1 phase, and refold during S/G2/M³. In addition, cell fate propensity has been shown to be determined by the cell cycle state of hESCs via the regulation of SMAD2/3, the main effector of the TGFβ signalling pathway⁴. However, the genome wide impact of these changes on transcriptional networks and the link with chromatin structure are still unknown. To address these questions, we used a FUCCI-hESC reporter line^{4,5} to perform ATAC-seq, RNA-seq and ChIP-seq on specific transcription factors during cell cycle progression to look at the dynamic changes in chromatin accessibility and transcription factors binding. We observed that the genomic location of several pluripotency factors and epigenetic modifiers is cell cycle regulated, and this allows induction of differentiation in the G1 phase.

Results

Chromatin accessibility changes during cell cycle progression in hESCs

FUCCI-hESCs were grown in chemically defined conditions maintaining pluripotency and then sorted based on their fluorescence (Early G1 (EG1), Late G1 (LG1), G1/S and S/G2/M) (Supplementary Information Fig. 1a). ATAC-seq (Fig. 1a, Supplementary Information Fig. 1b) was performed and a conservative set of peaks was generated using the IDR pipeline⁶. 88,479

reproducible peaks were identified in EG1, 117,283 in LG1, 99,109 in G1/S, 80,398 in S/G2/M (Supplementary Information Fig. 1c), suggesting that chromatin shows high degree of accessibility in all the cell cycle phases in hESCs especially when compared to definitive Endoderm cells (dEN, Vallier lab, unpublished). Interestingly, the number of open regions was higher in the early cell cycle while decreasing during the S/G2/M phase (Fig. 1b), thereby confirming that chromatin could progressively close during progression towards cell division. In total, we could identify 153,007 unique open regions. However, only ~33% of these regions were opened in all the phases (50,461) whereas ~67% of these regions were accessible in either a single phase (~34%) or more than one, but not all (~33%) (Fig. 1c). Conservative pairwise comparisons⁷ detected about 1,000 highly dynamic sites ($|FC| > 1.5$ and corrected p-value $p_{adj} < 1E-6$), mostly changing through the EG1 phase: 715 regions from EG1 to LG1, 195 from S/G2/M to EG1 (Fig. 1d, e). These observations suggest that chromatin accessibility dynamically changes during cell cycle progression. Interestingly, regions opening in EG1 phase were enriched for genes associated with transcriptional regulation, whereas regions closing from S/G2/M to EG1 were enriched for genes related to cell cycle (Supplementary Information Fig. 1d). Thus, regulation of transcriptional networks could occur in the early phase of cell cycle, whereas mechanisms controlling cell division are induced later on. Further analyses showed that chromatin around promoters rapidly open and reorganises after division (Supplementary Information Fig. 1e) thereby suggesting that EG1 phase could be the most dynamic in term of transcriptional regulation. In order to understand which transcription factors are regulating such changes, we performed digital footprinting analyses on the regions showing dynamic chromatin accessibility during cell cycle progression. Combining FLR⁸ and Wellington⁹ pipelines, we generated a conservative list of 1,807 footprints. Of these, 74% (1,330) were found in EG1, 8% (147) in LG1, 11% (203) in G1/S, and 7% (127) in S/G2/M (Supplementary Information Fig. 1f). In all the phases, apart from the G1/S transition, the top-

ranking footprint corresponded to the transcription factor CTCF (Fig. 1f, g; Supplementary Information Fig. 1g, Supplementary table 1). More specifically, CTCF represented ~50% of all the footprints in the EG1 phase, dropping to ~10% in the other phases. A wider distribution of factors' footprints could be detected in the rest of the cell cycle, with a gradual increase in FOX-protein footprints (FOXO3, FOXO4) which were previously described as pioneering factors able to bind condensed chromatin¹⁰. These data suggest that transcriptional networks are dynamically regulated during cell cycle progression in hESCs and that CTCF could play a major role in chromatin regulations especially at the start of the cell cycle.

Changes in chromatin accessibility during cell cycle are not associated with transcriptional variations

We next decided to assess whether changes in chromatin accessibility correlate with gene expression. Therefore, RNA-seq was performed on FUCCI-hESCs sorted in the different cell cycle phases (EG1, LG1, S/G2/M; Fig. 2 and Supplementary Information Fig. 2a, b) and DESeq2¹¹ analysis revealed 394 differentially expressed genes (Fig. 2b) in agreement with previous reports¹². Gene Ontology and pathway analyses showed enrichment for cell cycle terms especially regulators of cell division (Fig. 2c). Hierarchical clustering identified three distinct patterns of expression (Fig. 2b, d). Cluster 1 included genes mostly induced during the S/G2/M phase such as CDK1, CDC25C, AURKA (Supplementary Information Fig. 2c). Cluster 3 included a list of metabolic genes downregulated from EG1 to S/G2/M, suggesting that metabolic activity might decrease during cell cycle progression. In agreement with our finding of highly dynamic chromatin around the EG1 phase, this cluster included several histone genes clusters, and pathway analysis showed enrichment for chromatin rearrangement terms such as promoter opening and histones acetylation/deacetylation (Supplementary Information Fig. 2d). Cluster 2 includes genes upregulated in LG1, some of which are related to endoderm formation and developmental processes, including EOMES, MIXL1, SOX17,

FOXA2, GATA4 and GATA6. qPCR analyses confirm the expression of these genes in LG1 (Supplementary Information Fig. 2e). However, their expression level remains low when compared to induction observed during endoderm differentiation, suggesting priming of expression rather than full induction (data not shown). In order to assess whether this transcriptional background correlates with changes in chromatin accessibility, we overlapped all of the 13,478 expressed transcripts ($\log_2\text{RPM} > 1$) with all the associated features of the ATAC-seq peaks: ~85% of expressed genes overlap with accessible regions (Fig. 2e), confirming that most active genes are indeed located in regions of open chromatin. We then overlapped the list of differentially expressed genes with the list of dynamic ATAC regions (from Fig. 1d). Strikingly, only ~3% of differentially expressed genes overlapped with dynamic regions thereby indicating that the majority of cell cycle regulated transcripts originate from constantly open chromatin regions (Fig. 2f). Taken together, these results suggest that differentiation markers are expressed at low level in the G1 phase, potentially marking the priming of hESCs for differentiation. However, this transcriptional “leakiness” originates mostly in constantly open chromatin regions suggesting that it could be induced by differential binding of transcription factors or epigenetic modifiers during cell cycle progression.

Transcription factors binding changes during cell cycle progression in hESCs

In order to test whether transcription factors are driving such transcriptional changes, we performed several ChIP-seq in sorted FUCCI-H9 (Fig. 3a) starting with CTCF. Genome wide mapping revealed 5,606 unique binding sites, of which 4,401 were found in EG1 (Fig. 3b-d). Of the 708 CTCF footprints identified using ATAC-seq data ~60% overlapped with CTCF ChIP-seq confirming the robustness of our prediction (Supplementary Information Fig. 3a). Interestingly, 77% of CTCF binding sites were located in open chromatin regions, with a higher overlap in the EG1 phase (Supplementary Information Fig. 3b), thereby suggesting a specific role for CTCF in regulating chromatin accessibility in this phase of the cell cycle. All together,

these results show that CTCF genome wide occupancy changes considerably during cell cycle progression. The increased binding of CTCF in EG1 supports the hypothesis that CTCF could drive chromatin restructuring to regulate transcription¹³ and genome integrity¹⁴ during the subsequent cell cycle.

In order to check whether this binding pattern could be shared with other transcription factors, we performed ChIP-seq on sorted FUCCI-H9 for the core pluripotency factor OCT4, NANOG and SOX2. For OCT4, we found a total of 4,843 binding sites distributed as follow: 611 in EG1, 1,378 in LG1, and 2,854 in S/G2/M (Supplementary Information Fig. 3c) showing that the number of genomic regions bound by this key pluripotency factor increases during cell cycle progression (Fig. 3e, f). The same binding pattern was observed for NANOG and SOX2 with both factors binding the genome more frequently in S/G2/M phase (Fig. 3g, Supplementary Information Fig. 3d). Intrigued by these findings, we asked whether additional factors known to be linked to the core transcriptional network might behave similarly. For that, we performed ChIP-seq for RING1B a chromatin modifier shown to partially bind chromatin in an OCT4-dependent manner¹⁵. These analyses revealed a unique binding pattern with high occupancy in the EG1 phase, a decrease in the LG1, and an increase in the S/G2/M (Fig. 3h). Interestingly, RING1B is known to have several roles from chromatin compaction to regulation of gene expression¹⁶. Thus, our observations could imply that these divergent roles might have a temporal distribution during cell cycle progression. Indeed, regions bound by RING1B in EG1 were enriched for CTCF binding sites together with SMC3, a core subunit of the cohesin complex (Supplementary Information Fig. 3e), suggesting that RING1B might have a structural role in opening chromatin at the start of the cell cycle. Interestingly, we also observed a stronger binding of OCT4 at binding sites shared with RING1B in S/G2/M (Supplementary Information Fig. 3f). Taken together, these findings show that transcription factors have

different binding patterns during cell cycle progression and that this regulation could define their activity.

The dynamic binding of these transcription factors could be explained by a variety of mechanisms including variation in pluripotency factors expression during cell cycle progression. However, RNA-seq, qPCR and western blot validations showed that the expression of these genes/proteins does not change between cell cycle phases (Supplementary Information Fig. 3g, h). Therefore, variations in the quantity of transcription factors do not explain their differential binding. We then hypothesised that post-translational modifications could regulate the differential binding of pluripotency factors. Of particular interest, phosphorylation at Serine 236 (p-OCT4) in hESCs is known to regulate OCT4 binding to chromatin and its transcriptional activity¹⁷, and equivalent to mouse Serine 229, sterically hinders DNA binding¹⁸. To confirm that such mechanism could control OCT4 binding in EG1, we performed immunostaining for p-OCT4 (Ser236) in FUCCI-hESCs. These analyses showed that mitotic cells have enriched phosphorylation signal (Fig. 3i), thereby confirming findings in mouse Embryonic Stem Cell (mESCs)¹⁹. However, hESCs in EG1 still partially retain p-OCT4 (Fig. 3i), suggesting that complete de-phosphorylation only happens during the LG1 phase. Taken together, these results suggest that phosphorylation could regulate the genomic binding of OCT4 during cell cycle progression in hESCs. Importantly, EG1 has been shown to be a key phase during cell cycle for cell fate decision in hESCs⁴. For this reason, we hypothesised that the reduced capacity of pluripotency factors to bind the genome in EG1 could be necessary for induction of differentiation.

Cell cycle specific OCT4 binding is required for induction of differentiation

To test whether pluripotency factors binding is regulating initiation of differentiation, we decided to focus on OCT4 and develop a FUCCI-hESC lines to respectively overexpress or knock-down OCT4 in a cell cycle phase specific manner (Fig. 4a). For that, we developed a

tamoxifen (4-OHT) inducible form of OCT4 (OCT4-ERT2), that relies on the fusion of OCT4 with a modified fragment of the Estrogen receptor²⁰ (Supplementary Information Fig. 4a, b). Immunostaining confirmed correct nuclear translocation of OCT4 upon addition of 4-OHT in as little as 4h in both pluripotent and neuroectoderm cells (Fig. 4b). We then sorted OCT4-ERT2-FUCCI-hESCs in EG1 and re-plated them in culture condition inducing differentiation^{21,22} for 24h in the presence or absence of 4-OHT. Interestingly, OCT4 induction decreases the expression of the key early endoderm markers such as T and MIXL1 (Fig. 4c) suggesting a decreased differentiation. In parallel, we generated an inducible hESCs for the knockdown of OCT4 expression (iKD-OCT4-FUCCI-hESCs), taking advantage of the Tetracycline (TET) based system OptiKD²³ (Supplementary Information Fig. 4c, d). iKD-OCT4-FUCCI-hESCs sorted in S/G2/M were re-plated in culture condition inducing endoderm differentiation for 24h (Fig. 4d), in presence or absence of TET. Interestingly, decrease of OCT4 in the S/G2/M phase promoted the expression of late endoderm (SOX17) and mesoderm (HAND1) markers (Fig. 4e) thereby suggesting a more efficient differentiation process. Taken together, these results show that variation in OCT4 binding during cell cycle is required for optimal cell fate decision in hESCs or more broadly, that cell cycle specific regulation of transcriptional networks could control capacity of differentiation in hESCs.

Discussion

Our study describes for the first time the dynamic mechanisms that coordinate chromatin accessibility, gene expression, and transcriptional networks during cell cycle progression in hESCs. In the EG1 phase, immediately after cell division, chromatin is highly dynamic and CTCF appear to regulate changes in DNA accessibility. Indeed, CTCF might be the driver for reorganising chromatin after chromosomes condensation during cell division¹³ working together with other chromatin modifiers such as RING1B²⁴. In parallel, the core pluripotency

factors OCT4, SOX2 and NANOG show low occupancy in the G1 phase allowing hESCs to rapidly respond to differentiation cues. In the S/G2/M phase, they fully bind chromatin to re-establish the pluripotency network, sustain self-renewal and block differentiation. These results reinforce previous studies showing that this phase actively promote the pluripotent state²⁵. This change in binding might be regulated via post-translational modifications such as the phosphorylation preventing OCT4 to bind DNA in EG1¹⁷. Importantly, these changes could define the function of RING1B which might help CTCF in restructuring chromatin in EG1 while acting as transcriptional regulator in S/G2/M to maintain pluripotency²⁶. Ultimately these cell cycle specific interplays are necessary to coordinate induction of differentiation and exit of pluripotency with cell division.

Taken together, our results demonstrate that key transcription factors are cell cycle regulated and dynamically change their genomic location. This binding pattern suggests that transcription factors activity and specificity are coordinated with cell cycle progression, and this process could represent a key mechanism that allows stem cells to not only maintain their capacity to change of cellular identity but also for transcription factors to have different functions. These findings highlight the importance of studying transcriptional and epigenetic mechanisms in the context of the cell cycle. More generally, most ChIP-seq studies are currently capturing transcription factors' binding on the most extended cell cycle phase which can vary from cell to cell. Thus, we would like to propose that genomic location of transcription factors should be investigated in a cell cycle dependant manner, since phase-specific analyses could reveal new molecular function even for commonly studied regulators. This novel concept could apply to any cell types and thus have major impact on our understanding of transcription factors' functions.

223 **Acknowledgements**

224 The work was supported by the Wellcome PhD programme (A.O.), the European Research
225 Council starting grant “Relieve IMDs” (L.V.); and a core support grant from the Wellcome
226 Trust and Medical Research Council to the Wellcome Trust – Medical Research Council
227 Cambridge Stem Cell Institute. We thank Dr Pedro Madrigal for performing diffReps and
228 digital footprinting analysis and Dr Siim Pauklin for providing the ATAC-seq samples. We
229 thank the Cambridge NIHR BRC Cell Phenotyping Hub for help with cell sorting and the DNA
230 Pipelines Operations at the Wellcome Sanger Institute for next generation sequencing.

231

232 **Author contributions**

233 A.O. conceived the study, performed all the experiments, analysed data and wrote the
234 manuscript. L.V. conceived, supervised, and supported the study, wrote and provided final
235 approval of the manuscript.

236

237 **Competing Interests**

238 The authors declare no competing interests.

Figure legends

Figure 1. Chromatin accessibility changes during cell cycle progression in hESCs.

(a) Experimental approach. FUCCI hESCs were sorted in 4 different cell cycle phases: Early G1 (EG1), Late G1 (LG1), G1/S transition (G1/S), and S/G2/M. ATAC-seq was performed following sorting on biological duplicates. (b) Normalised ATAC-seq signal $\pm$ 1kb around nucleosome free regions in EG1, LG1, G1/S, S/G2/M, compared to definitive Endoderm (dEN). (c) Venn diagram showing ATAC-seq peaks overlap. (d) Number of differentially accessible peaks during cell cycle progression obtained by pair-wise comparison ($|FC| > 1.5$, corrected p-value (p-adj) $< 1E-6$). (e) Heat-maps displaying normalised accessibility signal of differentially accessible regions in (d) averaged using 50bp bins. Regions are centred on the merged peaks for the two replicates $\pm$ 1kb. K-means clustering correspond to k=2. (f) Footprints frequency (expressed as percentage) related to single transcription factors in EG1, LG1, G1/S, S/G2/M. (g) Average profile of CTCF footprint in the different cell cycle phases generated from Tn5 insertion frequencies.

Figure 2. Changes in chromatin accessibility during cell cycle are not associated with transcriptional variations.

(a) Experimental approach. FUCCI hESCs were sorted in 3 different cell cycle phases: Early G1 (EG1), Late G1 (LG1), G1/S transition (G1/S), and S/G2/M. RNA-seq was performed following sorting on biological triplicates. (b) Hierarchical clustering of 394 differentially expressed genes during cell cycle progression (p-value < 0.05 , per probe normalised). Three main clusters and relative representative genes are reported. (c) Gene Ontology (Biological Processes and Reactome) for the differentially expressed genes ranked by combined score. (d) Gene Ontology (Biological Processes) for the three main clusters in (b), ranked by combined

score. **(e)** Venn diagram summarising overlap of expressed genes (RNA-seq, $\log_2\text{RPM} > 1$) and accessible chromatin (ATAC-seq). 85% represents the percentage of expressed genes present in open chromatin regions (overlapping ATAC peaks). **(f)** Venn diagram reporting overlap of differentially expressed genes (RNA-seq DESeq2) and regions changing accessibility (ATAC-seq DiffRep). 3% represents the percentage of differentially expressed genes to which correspond changes in chromatin accessibility. Combined score is calculated by multiplying unadjusted p-value with z-score.

Figure 3. Transcription factors binding changes during cell cycle progression in hESCs.

(a) Experimental approach. FUCCI hESCs were sorted in 3 different cell cycle phases: Early G1 (EG1), Late G1 (LG1), and S/G2/M. ChIP-seq was performed following sorting on biological duplicates for CTCF, OCT4, NANOG, and SOX2. **(b)** Heat-maps displaying CTCF binding in the different cell cycle phases EG1, LG1, S/G2/M, compared to INPUT. **(c)** Normalised ChIP-seq signal $\pm$ 1kb around all peaks in EG1, LG1, S/G2/M, compared to INPUT. **(d)** Venn diagram reporting overlap of CTCF binding sites during cell cycle **(e)** Heat-maps displaying OCT4 binding in the different cell cycle phases EG1, LG1, S/G2/M, compared to INPUT. **(f)** Normalised ChIP-seq signal $\pm$ 1kb around all peaks in EG1, LG1, S/G2/M, compared to INPUT. **(g)** Heat-maps displaying NANOG and SOX2 binding in the different cell cycle phases EG1, LG1, S/G2/M, compared to INPUT. **(h)** Heat-maps displaying RING1B binding in the different cell cycle phases EG1, LG1, S/G2/M, compared to INPUT. **(i)** Immunofluorescence for phospho-OCT4 Ser236 (pOCT4, magenta) and nuclear staining (DAPI, blue) in FUCCI hESCs (S/G2/M cells in green-mAG, LG1 cells in red-mKO2, EG1 cells no colour). Scale bar 20 μm . For all the heat-maps, regions are centred on the merged peaks for the two replicates $\pm$ 1kb in all the phases and compared to INPUT.

Figure 4. Cell cycle specific OCT4 binding is required for induction of differentiation.

(a) Experimental approach. FUCCI OCT4-ERT2 were sorted in Early G1 (EG1) and re-plated in Endoderm inducing conditions (FLyAB). OCT4 nuclear translocation was induced upon 4-Hydroxytamoxifen addition (+4OHT). FUCCI iKD OCT4 were sorted in S/G2/M and re-plated in Endoderm inducing conditions (FLyAB). OCT4 knock-down was induced upon tetracycline addition (+TET). **(b)** Immunofluorescence for OCT4 (magenta) and nuclear staining DAPI (blue) in FUCCI OCT4-ERT2. In presence of 4-Hydroxytamoxifen (+4OHT), OCT4 is uniquely nuclear. Upon neuroectoderm induction (SB), endogenous OCT4 is strongly reduced and upon 4-Hydroxytamoxifen induction is translocated into the nucleus (SB +4OHT). Scale bar 20µm. **(c)** Gene expression qPCR of endoderm markers T and MIXL1 in presence or absence of 4-Hydroxytamoxifen (+4OHT) in EG1 sorted cells after 24h of differentiation. **(d)** Gene expression time-course during OCT4 knock-down in presence of tetracycline (+TET) compared to -TET control. **(e)** Gene expression qPCR of late endoderm markers SOX17, HAND1 in presence or absence of tetracycline in S/G2/M sorted cells after 24h of differentiation. **(f)** Schematic summary of the proposed working model. In the EG1 phase, immediately after cell division, chromatin is highly dynamic and CTCF binding is proposed to regulate changes in accessibility. Core pluripotency factors OCT4, SOX2 and NANOG (OSN) show low occupancy in the G1 phase, and this allows hESCs to respond to differentiation cues. In the S/G2/M phase, core pluripotency factors fully re-bind chromatin re-establishing the pluripotency network, sustaining self-renewal and blocking differentiation.

Methods

hESCs culture and germ layers specification

FUCCI-hESCs (H9/WA09 line; WiCell) were grown feeder- and serum-free as previously described²⁷. Briefly, cells were grown in Chemically Defined Medium (CDM, Supplementary Table 2), supplemented with Bovine Serum Albumin (BSA), recombinant Activin-A (10 ng/ml) and Fibroblast growth factor 2 (FGF2, 12 ng/ml), both from Dr Marko Hyvonen (Department of Biochemistry, University of Cambridge) on 0.1% Gelatin and MEF-medium coated plates. Cells were passaged weekly using Collagenase IV and plated as clumps of 50-100 cells dispensed at a density of 100-150 clumps/cm². All the FUCCI-hESCs lines used were kept under selection with Puromycin (1mM) and Geneticin (G-418, 0.2 mg/ml), or alternatively G-418 (0.2 mg/ml) and Blasticidin (1mM). Germ layer differentiation was initiated 24h after passaging. Definitive Endoderm (dEN) was induced for 3 days in CDM-Polyvinyl alcohol (PVA) without insulin supplemented with 20ng/ml FGF2, 10μM LY294002, 100ng/ml Activin A, and 10ng/ml BMP4, changing media daily, as previously described^{21,22}. Neuroectoderm was induced for 6 days in CDM-BSA supplemented with 12ng/ml FGF2 and 10μM SB431542 (Activin/Nodal/TGFβ inhibitor; Tocris), as previously described²⁸. hESCs were monthly monitored for mycoplasma infection.

Immunofluorescence (IF)

FUCCI hESCs were fixed for 20 min at 4°C in 4% paraformaldehyde (PFA)/Phosphate-Buffered Saline (PBS, Gibco) and washed three times with PBS. Non-specific bindings were blocked incubating cells with 10% Donkey Serum (Biorad) and 0.1% Triton X-100 (Sigma-Aldrich) PBS for 30 min. Cells were then incubated with the appropriate antibody (1:100,

Supplementary Table 3) in 1% Donkey Serum and 0.1% Triton X-100 PBS overnight at 4°C. Cells were then washed three times with PBS and incubate with the appropriate secondary antibodies (1:1,500; AlexaFluor donkey anti-goat/rabbit/mouse IgG, Invitrogen) in 1% Donkey Serum and 0.1% Triton X-100 PBS 1h at room temperature protected from light. Cells were then washed three times with PBS, for nuclei staining Hoechst (1:10,000, Sigma-Aldrich) was added to the first wash. Images were acquired using Zeiss LSM 700 Confocal Microscope and Zen2009 imaging software.

Western Blotting

Cells were harvested by scraping in PBS, centrifuged at 250g for 5 min at 4°C, and re-suspended in ice-cold CellLyticM buffer (Sigma-Aldrich) supplemented with cOmplete Protease Inhibitor (Roche). After 30 min incubation on ice, lysates were clarified by centrifugation at 16,000g for 5 min at 4°C. The supernatants were transferred to new tubes, and proteins in cellular extracts were quantified using Protein Quantification Kit-Rapid (Sigma) according to manufacturer's instructions. 1% β-mercaptoethanol (Sigma-Aldrich) was added to LDS NuPage Sample Buffer (Invitrogen NP0007) and protein denatured at 95°C for 5 min. Total protein (1-50μg depending on the relative abundance) was loaded on a precast NuPAGE 4-12% Bis-Tris protein gel (Invitrogen) and run for 45min, 200V, 130 mA. Proteins were transferred from the gels to Immuno-Blot® PVDF Membranes (Bio-Rad) using the Bio-Rad mini-Trans Blot system for 1h, 100V, 180mA. Membranes were incubated in blocking solution (4% milk, 0.05% Tween-20 in PBS) for 30 min at room temperature, and blotted with the appropriate primary antibody (1:1,000, Supplementary Table 3) overnight at 4°C. Afterwards, membranes were washed three times for 5 min with 0.05% Tween-20 PBS, incubated for 1h at room temperature with the appropriate secondary antibody (1:10,000, anti-mouse/rabbit/goat)

conjugated to Horseradish Peroxidase (HRP, Sigma-Aldrich), and developed using the Pierce ECL detection kit (Thermo Scientific) according to manufacturer's instructions.

RT-qPCR

RNA was extracted using the GenElute Mammalian Total RNA Miniprep Kit and the On-Column DNase I Digestion Set (both from Sigma-Aldrich) following manufacturer's instructions. cDNA was synthesised using 500ng of RNA in a reaction with 250ng random primers, 0.5mM deoxynucleotides (dNTPs), 20U RNaseOUT, and 25U of SuperScript II (all Invitrogen) in a total of 20µl. cDNA was diluted 30-fold and 5µl were used for qPCR using SensiMix SYBR low-ROX (Bioline) with 150nM Forward and Reverse primers (Sigma-Aldrich; Supplementary Table 4 for primer sequences). Samples were run in technical duplicates on 96-well plates on a Stratagene Mx-3005P (Agilent) or 384-well plate on a QuantStudio Flex Real-Time PCR System, and results were analysed using the delta-delta cycle threshold ($\Delta\Delta C_t$) approach using PBGD as housekeeping gene. Primers were designed using PrimerBlast (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>), and tested to produce a single PCR product, with an efficiency between 80% and 120%.

Cell Sorting

FUCCI-hESCs were sorted as previously described⁴. Briefly, cells were washed with PBS and collected by incubating with CDB for 10 min at 37°C. When needed, crosslinking of proteins to DNA was performed prior to sorting, according to the specific experiment (see Chromatin Immunoprecipitation). Cells were washed with 1% PBS/BSA and incubated in 1% PBS/BSA with anti-TRA-1-60-antibody mouse monoclonal IgM (1:100, Santa Cruz Biotechnology, Inc.) for 20 min on ice, together with the secondary antibody, Alexa Fluor® 647 goat anti-mouse IgM (1:1,000, Thermo Fisher). Cells were washed with PBS, re-suspended in 6ml of PBS and

sorted into the different cell cycle phases using BD FACS Aria™ III, BD Fusion or BD Influx sorters by the Cambridge NIHR BRC Cell Phenotyping Hub. Gating strategy set as following: cells were then size selected for Forward (FS) and Side Scatter (SS), and doublets were removed. TRA-1-60+ cells were then sorted in the four different cell cycle phases according to the FUCCI reporter: Early G1 (EG1) for double negative (mKO2- / mAG-), Late G1 (LG1) for red-mKO2+, G1/S transition for double positive (yellow-mKO2+ / mAG+), and S/G2/M for green-mAG+. Sorted cells were then used for ATAC-seq, ChIP-seq, and RNA-seq. All the experiments have been performed normalising for cell number, usually to the lowest sample, depending on the downstream application.

Chromatin Immunoprecipitation (ChIP)

Crosslinking of protein to DNA was performed prior to sorting in 1% PBS/Formaldehyde, rocking for 10 min at room temperature. All the experiments were performed with single crosslinking and neutralised with Glycine (0.125M), for 5 min shaking at room temperature. ChIP was then carried out using LowCell# ChIP kit (C0101007, Diagenode) and DNA recovered using iPure kit (C03010012, Diagenode) according to manufacturer's instructions, with the following modifications. 5µg of antibody were used per IP, and the quantity of magnetic beads adjusted accordingly, based on their binding capacity (10µl of beads for ~3µg). Elution was made in Buffer C from the iPure kit, compatible with qPCR and next generation sequencing. For all the experiments, the same number of cells was used per IP after sorting, normalised for the less abundant cell cycle phase, usually EG1 ~1M cells.

Assay for transposase-accessible chromatin (ATAC)

TRA-1-60+ hESCs were sorted in the different cell cycle phases (EG1, LG1, G1/S, S/G2/M). Samples were generated from two independent experiments, 100,000 cells per sample, as

previously described²⁹ with some modifications. In brief, cells were washed with ice-cold PBS and incubated with 500µl of ice-cold sucrose buffer (10mM Tris-Cl pH 7.5, 3mM CaCl₂, 2mM MgCl₂ and 0.32M sucrose) for 12 min on ice. To collect the nuclei, Triton X-100 was added to a final concentration of 0.5% and cells incubated for additional 6 min on ice after mixing. Samples were then resuspended in 1ml of sucrose buffer and transferred to a fresh 1.5 ml Eppendorf tube. The nuclei were centrifuged at 1500g for 3 min at 4°C and the supernatant discarded. Nuclei pellets resuspended in 50µl of Nextera DNA Sample Preparation Kit (Illumina FC-121-1030) for tagmentation, comprising 25µl 2x Tagment DNA buffer, 20µl nuclease-free water and 5µl Tagment DNA Enzyme 1. The tagmentation reaction mixture was transferred in a 1.5ml low-binding tube and incubated at 37°C for 30 min and stopped by the addition of 250µl Buffer PB (Qiagen). The tagmented chromatin was purified using the MinElute PCR purification kit (Qiagen 28004), according to the manufacturer's instructions, eluting in 11.5µl of buffer EB (Qiagen). For PCR amplification, 10µl of the tagmented chromatin was mixed with 2.5µl of Nextera PCR primer cocktail and 7.5µl of Nextera PCR mastermix (Illumina FC-121-1030) in low-binding PCR tube. Primers were from the Nextera Index kit (Illumina FC-121-1011), using 2.5µl of an i5 primer and 2.5µl of an i7 primer per PCR, totalling 25µl. PCR amplification was performed as follow: 72°C for 3 min and 98°C for 30 sec, followed by 12 cycles at 98°C for 10 sec 63°C for 30 sec and 72°C for 3 min. The sample volumes were then increased to 50µl by adding Qiagen EB buffer. The PCR primers were removed with 1x 0.9:1 SPRI beads (Beckman Coulter, Cat no. A63880) according to manufacturer's instruction and samples eluted in 20µl. Following amplification, excess of primers was removed running the samples of 50ml 1% agarose TAE gel at 90V for 25 min, and size selection obtained by cutting from 120bp to 1kb. MinElute Gel Extraction kit (Qiagen 28604) was used to extract DNA, eluted in 20µl of buffer EB.

Transfection

The generation of the inducible lines was performed using lipofection. Lipofection was performed as previously described³⁰. In brief, FUCCI-hESCs cells were seeded into 6 well plates using medium without Penicillin-Streptomycin (Pen/Strep, Gibco), in low density. After 48-72h (depending on colony size), cells were transfected using Lipofectamine[®] 2000 Transfection Reagent (Thermo Fisher), according to manufacturer's instruction. In sum, medium was aspirate and replaced with 1mL/well of Opti-MEM[®] Reduced Serum Media (Thermo Fisher). 4µg of (each) plasmid was resuspended in 0.25ml of Opti-MEM medium/well. 10µl of Lipofectamine 2000 were dissolved in 0.25ml of Opti-MEM medium/well, and incubated for 5 min at room temperature. Lipofectamine 2000 and the plasmid were mixed together and incubated for 20 min at room temperature. 0.5ml of the DNA/Lipofectamine mix was added drop by drop to the well and incubate for 18-20h. Opti-MEM was then replaced with regular media without antibiotics. Antibiotic selection was started 4-5 days after the transfection (Puromycin 1 mg/ml, G-418 50µg/ml). After 6 days, clones were picked into 24 wells plate. Each clone was passaged into 12 well plates, for analysis of transgene expression by qPCR, immunostaining or western blot.

Bioinformatics analyses and pipelines

The following data processing pipelines have been used for all the bioinformatics analyses according to ENCODE⁶.

ChIP-seq

TRA-1-60+ hESCs were sorted in the different cell cycle phases (EG1, LG1, S/G2/M). From two independent experiments, chromatin immunoprecipitation was performed as described above. Library preparation and sequencing were performed at the Wellcome Trust Sanger Institute next-generation sequencing facility. Size selection was applied, and fragments

between 100bp and 400bp (average length of 200bp) were used to prepare barcoded sequencing libraries using NEBNext Sample Prep Kit1 (NEB) following manufacturer's instructions. 10ng of input material (as a pool of the different phases) for each experiment was processed in parallel as control. To test whether pooled inputs were an acceptable control, one experiment was run with one input per cell cycle phase. No differences were reported with > 90% peaks shared with the pooled input (data not shown). Equimolar amounts of each library were pooled, and 10 samples/lane were multiplexed on Illumina HiSeq 2000, 2 X 75bp paired-end reads. On average, 400M reads per lane were generated, mapped to human genome GRCh38 reference assembly and stored as cram files. Cram were converted in bam keeping only reads with mapping score above 10 ($q > 10$) using Samtools view³¹. From two independent replicates, consistency between replicates was measured using the Irreproducibility Discovery Rate (IDR) pipeline (<https://sites.google.com/site/anshulkundaje/projects/idr>). This approach is based on an independent peak-calling for the two replicates with a relaxed cut-off (-p1e-1, --to-large) using MACS³². This allows to sample enough noise in the experiment for efficient statistical comparison of replicates in subsequent steps, meaning that many false positives are present in this peak set and they are not considered as final peaks for the separate replicates. Peaks are then ranked for p-value, and the same number is compared based on reproducibility (top 100,000 peaks). This is used to score rankings between peaks in each replicate to determine an optimal cut-off for significance. The final peak calls (bed and bigBed format) are the set of peak calls that pass IDR at a threshold of 5-10% max. This is the conservative peak set and it is used for further analyses. If minimum number of recommended reads is not achieved, replicates are pooled. For visualisation purposes, biological replicates were combined. Normalised bedGraph format files were generated for each sample using BEDTools v2.24.0³³. The reads mapped at both DNA strands from 5' to 3' direction were extended to a length of 300bp, and the read-enrichment was normalized by million mapped reads and size of the

library. bedGraph files were converted to bigWig using UCSC tool bedGraphToBigWig, and visualised on the Biodalliance genome viewer³⁴. For the annotation of genomic intervals, Heatmaps and Metaplot generation, the R packages Genomation and ChIPpeakAnno were used^{35,36}. GREAT (Genomic Regions Enrichment of Annotations Tool) was used to associate genes with regulatory regions³⁷, combined with UCSC liftOver tool to convert bed coordinates from Homo Sapiens GRCh38/hg38 assembly to GRCh37/hg37 assembly (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>).

ATAC-seq

TRA-1-60+ hESCs were sorted in the different cell cycle phases (EG1, LG1, G1/S, S/G2/M). Samples generated from two independent experiments, 100,000 were used per sample, as described above. Library preparation and sequencing were performed at the Wellcome Trust Sanger Institute next-generation sequencing facility. 8 ATAC-seq libraries were prepared with one of i5 and i7 Nextera tags combination and pooled equimolarly. Sequencing was performed on Illumina HiSeq 2000, 2 X 75bp paired-end reads obtaining more than 60M mapped reads per library. Reads were mapped to human genome GRCh38 reference assembly and stored as cram files. Duplicates were removed with Samtools rmdup³¹. Cram were converted in bam keeping only reads with mapping score above 5 ($q > 5$) and from chromosome 1-22 and X (H9 is a female line) to remove mitochondrial contamination using Samtools view³¹. From two independent replicates, consistency between replicates was measured using the Irreproducibility Discovery Rate (IDR) pipeline, as described above. Conservative peak set was used for further analyses. Diagnostic plots of fragment size distribution were generated using the R package ATACseqQC³⁸. For visualisation purposes, biological replicates were combined as described above to generate normalised bigWig.

Differential accessibility was assessed using the G-test implemented in diffReps ⁷ by Dr Pedro Madrigal. This tool takes into account the biological variations within a group of samples and uses that information to enhance the statistical power. For this, we used a sliding window of 600bp, $p < 1E-6$. Regions with differential binding that did not overlap with previously detected peaks through the IDR pipeline were removed. Genomic regions were ranked by their fold change (FC), and reported as differential only if both the $|FC| > 1.5$ and corrected p-value $\text{padj} < 1E-6$. Footprinting analyses were done by Dr Pedro Madrigal combining two methods: Footprint Log-likelihood Ratio (FLR) ⁸ and Wellington ⁹, adjusted for ATAC-seq. To get high quality reproducible footprints, motif matches with $\text{FLR} \geq 10.0$ in each biological replicate were classified as “bound”. Transcription factor footprints that were present in both replicates, for a motif with at least 50 footprints, were considered as final predicted transcription factor binding sites. FLR mean of the replicates was calculated, only based on unique genomic regions (for those palindromic sequences with a footprint reported in each strand, we obtained the mean FLR). The quality of footprint was assessed calculating the Protection Score, removing scores below 0 ³⁹. Wellington instead uses a “non-motif centric” approach, which considers the imbalance in this strand-specific information to efficiently identify DNA footprints. For this, a relaxed cut-off has been used (`-p32 -A -fdr 0.1 --FDR_limit -4 --pv_cutoffs -4 -fp 4,30,1 -fdriter 500 —one_dimension`) on merged replicates (to increase sequencing depth), in order to achieve a significant overlap with FLR, and a final list generated. A conservative list of footprints was generated by overlap between FLR and Wellington, removing not expressed genes according to RNA-seq data (Fragments Per Kilobase of transcript per Million mapped reads, $\text{FPKM} > 1.0$).

For visualisation of footprints, bigWigs of ATAC-seq data containing Tn5 insertion frequencies have been generated using NucleoATAC ⁴⁰ from the bam files (`--bam $bamFile -out $outputBasename --cores 16`).

RNA-seq

TRA-1-60+ hESCs were sorted in the different cell cycle phases (EG1, LG1, S/G2/M). From three independent experiments, RNA was extracted using GenElute™ Mammalian Total RNA Miniprep Kit and On-Column DNase I Digestion Set (both Sigma-Aldrich) and a total of 9 samples fulfilling the submission requirements (0.5-1.0µg total RNA in nuclease free dH2O with A260/A280 and A60/A230 ratios above 1.8) were sent for library preparation and next-generation sequencing at CGS (Cambridge Genomic Services). PolyA purified library preparation kit was used, and samples were multiplexed in one HIGH Output NextSeq run Cycle, 2 X 75bp, paired-end reads. A total of 462,569,268 reads evenly distributed across the lanes were obtained (50,547,439 on average per sample). Reads were trimmed using Sickle⁴¹, an application for trimming FASTQ files based on quality values (Sanger), with a q=20 (quality threshold) and l=30 (length threshold). For the alignment, the transcriptome was built with TopHat2⁴² (v2.1.0), based on Bowtie2⁴³ (v2.2.6) using as reference genome the human GRCh38.p6 and GTF annotation from Ensembl (http://ftp.ensembl.org/pub/release-83/gtf/homo_sapiens/). Alignment was performed using TopHat2 with standard parameters. Using Samtools view³¹, reads with mapping quality > 10 were kept for further analyses. Data were quantitated and differential expression analysis was performed using SeqMonk⁴⁴. In sum, data were quantitated using the RNA-seq pipeline, and differential expression analysis was performed using the Intensity Difference Test and the R package DESeq2¹¹ (for this, uncorrected raw read counts were used), generating a list of transcripts passing the filter, with a p-value cutoff < 0.05. Hierarchical clustering plots (r > 0.7). Gene Ontology analyses were performed using the Enrichment analysis tool by the Gene Ontology Consortium (Panther)⁴⁵, and validated using Enrich⁴⁶.

557 **Data availability**

558 The datasets generated during this study are available at ArrayExpress under the following
559 accession numbers: E-MTAB-8347 (RNA-seq), E-MTAB-8348 (ATAC-seq), E-MTAB-8358
560 (ChIP-seq). Login details available upon request.

Supplementary Information Figure legends

Supplementary Information Figure 1. CTCF footprints overlaps with ATAC-seq changes during cell cycle progression.

(a) Gating strategy for FUCCI hESCs: P1 cells are size selected for forward scatter area (FSC-A) and side scatter area (SSC-A); P2, doublets are removed (FSC-A, FSC-W, width); P3 TRA-1-60+ cells are selected and sorted in the four different cell cycle phases according to the FUCCI reporter: EG1 for double negative (mKO2-/mAG-), LG1 for red-mKO2+, G1/S transition for double positive (yellow-mKO2+/mAG+), and S/G2/M for green-mAG+. For all events, average and standard deviation are reported with n=10. **(b)** ATAC-seq fragment size density plot for the 4 sequenced libraries, EG1, LG1, G1/S, and S/G2/M sorted cells. Inset histograms (log-transformed) show periodicity to four nucleosomes. **(c)** IDR (Irreproducible Discovery Rate) plots reporting reproducible ATAC-seq peaks between two replicates in EG1, LG1, G1/S, S/G2/M. IDR threshold < 5%. **(d)** Gene ontology for Biological Process analysis of differentially accessible peaks in EG1vsLG1 down, EG1vsLG1 up, S/G2/MvsEG1 down, S/G2/MvsEG1 up (diffReps, Fig. 1d). Combined score is calculated by multiplying the unadjusted p-values with the z-score. **(e)** Feature distribution for the ATAC-seq peaks for differentially open regions (diffReps) in EG1vsLG1 down, EG1vsLG1 up, S/G2/MvsEG1 down, S/G2/MvsEG1 up. **(f)** Number of predicted footprints overlapping with dynamic regions (ATAC- diffReps) in the different cell cycle phases. Final footprints were obtained combining FLR (Yardimici et al., 2014) and Wellington (Piper et al., 2013) pipelines (combining list with Footprint Likelihood Ratio (FLR) ≥ 10.0 , Protection Score > 0). **(g)** Pairwise comparisons of all cell cycle phases, reporting percentages of footprints for specific transcription factors, log scale axis.

Supplementary Information Figure 2. Transcriptome changes during cell cycle progression.

(a) Datastore Tree clustering of different samples/replicates. (Pearson correlation). **(b)** PCA clustering of FUCCI hESCs sorted in EG1, LG1, S/G2/M. Each dot represents a biological replicate. **(c)** Examples of differentially expressed cell cycle regulator CDK1, CDC25C, AURKA, up-regulated in the S/G2/M phase (p-value cut-off < 0.05). **(d)** Gene ontology for Reactome Pathway analysis for Cluster 3 (Fig.2). Combined score is calculated by multiplying the unadjusted p-values with the z-score. **(e)** RNA-seq log2RPM values for the genes SOX17, EOMES, and MIXL1, validated by q-PCR (values relative to housekeeping gene PBGD).

Supplementary Information Figure 3. Transcription factors binding dynamics during cell cycle.

(a) Venn diagram reporting the overlap of CTCF footprints predicted by ATAC-seq and CTCF ChIP-seq peaks. 60% represent the percentage of footprints overlapping with ChIP-seq peaks. **(b)** Venn diagram reporting the overlap of CTCF ChIP-seq peaks and ATAC-seq peaks. 77% represent the percentage of CTCF ChIP-seq peaks with ATAC-seq peaks. **(c)** Venn diagram showing peaks overlap for OCT4 ChIP-seq in EG1, LG1, S/G2/M. **(d)** Venn diagrams showing peaks overlap for SOX2 and NANOG ChIP-seq in EG1, LG1, S/G2/M. **(e)** Gene ontology for Biological Process and Encode ChEA consensus sequences of RING1b binding sites and RING1B EG1 consensus sequences. Combined score is calculated by multiplying the unadjusted p-values with the z-score. **(f)** qPCR for OCT4, NANOG and SOX2 in EG1, LG1, and S/G2/M relative to housekeeping gene PBGD (two-ways ANOVA, n=3), compared to RNA-seq data for OCT4, SOX2, and NANOG, in EG1, LG1, and S/G2/M, expressed in log2RPM. **(g)** Western-blot for OCT4, SOX and NANOG in EG1, LG1, and S/G2/M, plus unsorted (UNS) control (α -tubulin as loading control).

Supplementary Information Figure 4. Cell cycle specific OCT4 overexpression and knock-down.

(a) The modified fragment of the Estrogen receptor (ERT2) is fused to the OCT4 gene, which is retained in the cytoplasm. In presence of an Estrogen receptor antagonist, such as 4-Hydroxytamoxifen (4OHT), the fusion protein binds to it and relocates into the nucleus. CAAG: CMV early enhancer, chicken β -actin and rabbit β -globin hybrid promoter; IRES: internal ribosome entry site; PuroR: Puromycin resistance. **(b)** OCT4 WB for 2 control lines (lane 1-2) and 4 OCT4 ERT2 lines (lanes 3-6), wild-type OCT4 band at 39 kDa and the OCT4 ERT2 fusion-protein band at 74 kDa. α -tubulin as loading control. **(c)** Single step optimised inducible knock-down from Bertero et al. 2016. AAVS1 locus targeting, with OCT4 shRNA under the control of a tetracycline inducible promoter. ZFN: zinc-finger nucleases; 5'-HAR/3'-HAR: upstream/downstream homology arm; H1-TO: Tetracycline-inducible H1 Pol III promoter carrying one tet operon after the TATA box; CAAG: CMV early enhancer, chicken β -actin and rabbit β -globin hybrid promoter; TetR: Tetracycline-sensitive repressor protein. **(d)** OCT4 knock-down WB at 0h, 6h, 12h, 24h, 48h, 72h after TET induction (+TET), compared to -TET control. α -tubulin as loading control.

629 **References**

- 630 1 Naumova, N. *et al.* Organization of the mitotic chromosome. *Science* **342**, 948-953,
631 doi:10.1126/science.1236083 (2013).
- 632 2 Nagano, T. *et al.* Cell-cycle dynamics of chromosomal organization at single-cell
633 resolution. *Nature* **547**, 61-67, doi:10.1038/nature23001 (2017).
- 634 3 Chen, X. *et al.* ATAC-seq reveals the accessible genome by transposase-mediated
635 imaging and sequencing. *Nat Methods* **13**, 1013-1020, doi:10.1038/nmeth.4031
636 (2016).
- 637 4 Pauklin, S. & Vallier, L. The cell-cycle state of stem cells determines cell fate
638 propensity. *Cell* **155**, 135-147, doi:10.1016/j.cell.2013.08.031 (2013).
- 639 5 Sakaue-Sawano, A. *et al.* Visualizing spatiotemporal dynamics of multicellular cell-
640 cycle progression. *Cell* **132**, 487-498, doi:10.1016/j.cell.2007.12.033 (2008).
- 641 6 Landt, S. G. *et al.* ChIP-seq guidelines and practices of the ENCODE and
642 modENCODE consortia. *Genome Res* **22**, 1813-1831, doi:10.1101/gr.136184.111
643 (2012).
- 644 7 Shen, L. *et al.* diffReps: detecting differential chromatin modification sites from
645 ChIP-seq data with biological replicates. *PLoS One* **8**, e65598,
646 doi:10.1371/journal.pone.0065598 (2013).
- 647 8 Yardimci, G. G., Frank, C. L., Crawford, G. E. & Ohler, U. Explicit DNase sequence
648 bias modeling enables high-resolution transcription factor footprint detection. *Nucleic*
649 *Acids Res* **42**, 11865-11878, doi:10.1093/nar/gku810 (2014).
- 650 9 Piper, J. *et al.* Wellington: a novel method for the accurate identification of digital
651 genomic footprints from DNase-seq data. *Nucleic Acids Res* **41**, e201,
652 doi:10.1093/nar/gkt850 (2013).
- 653 10 Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: establishing competence for
654 gene expression. *Genes Dev* **25**, 2227-2241, doi:10.1101/gad.176826.111 (2011).
- 655 11 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and
656 dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550,
657 doi:10.1186/s13059-014-0550-8 (2014).
- 658 12 Singh, A. M. *et al.* Cell-cycle control of developmentally regulated transcription
659 factors accounts for heterogeneity in human pluripotent cells. *Stem Cell Reports* **1**,
660 532-544, doi:10.1016/j.stemcr.2013.10.009 (2013).
- 661 13 Sekiya, T., Kato, K., Kawaguchi, A. & Nagata, K. Involvement of CTCF in
662 transcription regulation of EGR1 at early G1 phase as an architecture factor. *Sci Rep*
663 **9**, 329, doi:10.1038/s41598-018-36753-x (2019).
- 664 14 Canela, A. *et al.* Genome Organization Drives Chromosome Fragility. *Cell* **170**, 507-
665 521 e518, doi:10.1016/j.cell.2017.06.034 (2017).
- 666 15 Endoh, M. *et al.* Polycomb group proteins Ring1A/B are functionally linked to the
667 core transcriptional regulatory circuitry to maintain ES cell identity. *Development*
668 **135**, 1513-1524, doi:10.1242/dev.014340 (2008).
- 669 16 Aranda, S., Mas, G. & Di Croce, L. Regulation of gene transcription by Polycomb
670 proteins. *Sci Adv* **1**, e1500737, doi:10.1126/sciadv.1500737 (2015).
- 671 17 Swaney, D. L., Wenger, C. D., Thomson, J. A. & Coon, J. J. Human embryonic stem
672 cell phosphoproteome revealed by electron transfer dissociation tandem mass
673 spectrometry. *Proc Natl Acad Sci U S A* **106**, 995-1000,
674 doi:10.1073/pnas.0811964106 (2009).
- 675 18 Saxe, J. P., Tomilin, A., Scholer, H. R., Plath, K. & Huang, J. Post-translational
676 regulation of Oct4 transcriptional activity. *PLoS One* **4**, e4467,
677 doi:10.1371/journal.pone.0004467 (2009).

678 19 Shin, J. *et al.* Aurkb/PP1-mediated resetting of Oct4 during the cell cycle determines
679 the identity of embryonic stem cells. *Elife* **5**, doi:10.7554/eLife.10877 (2016).

680 20 Teo, A. K. K. *et al.* Pluripotency factors regulate definitive endoderm specification
681 through eomesodermin. *Gene Dev* **25**, 238-250, doi:10.1101/gad.607311 (2011).

682 21 Vallier, L. *et al.* Signaling pathways controlling pluripotency and early cell fate
683 decisions of human induced pluripotent stem cells. *Stem Cells* **27**, 2655-2666,
684 doi:10.1002/stem.199 (2009).

685 22 Touboul, T. *et al.* Generation of functional hepatocytes from human embryonic stem
686 cells under chemically defined conditions that recapitulate liver development.
687 *Hepatology* **51**, 1754-1765, doi:10.1002/hep.23506 (2010).

688 23 Bertero, A. *et al.* Optimized inducible shRNA and CRISPR/Cas9 platforms for in
689 vitro studies of human development using hPSCs. *Development* **143**, 4405-4418,
690 doi:10.1242/dev.138081 (2016).

691 24 Kundu, S. *et al.* Polycomb Repressive Complex 1 Generates Discrete Compacted
692 Domains that Change during Differentiation. *Mol Cell* **65**, 432-446 e435,
693 doi:10.1016/j.molcel.2017.01.009 (2017).

694 25 Gonzales, K. A. *et al.* Deterministic Restriction on Pluripotent State Dissolution by
695 Cell-Cycle Pathways. *Cell* **162**, 564-579, doi:10.1016/j.cell.2015.07.001 (2015).

696 26 Endoh, M. *et al.* Histone H2A mono-ubiquitination is a crucial step to mediate PRC1-
697 dependent repression of developmental genes to maintain ES cell identity. *PLoS*
698 *Genet* **8**, e1002774, doi:10.1371/journal.pgen.1002774 (2012).

699 27 Vallier, L. Serum-free and feeder-free culture conditions for human embryonic stem
700 cells. *Methods Mol Biol* **690**, 57-66, doi:10.1007/978-1-60761-962-8_3 (2011).

701 28 Vallier, L. *et al.* Early cell fate decisions of human embryonic stem cells and mouse
702 epiblast stem cells are controlled by the same signalling pathways. *PLoS One* **4**,
703 e6082, doi:10.1371/journal.pone.0006082 (2009).

704 29 Kumasaka, N., Knights, A. J. & Gaffney, D. J. Fine-mapping cellular QTLs with
705 RASQUAL and ATAC-seq. *Nat Genet* **48**, 206-213, doi:10.1038/ng.3467 (2016).

706 30 Vallier, L. *et al.* Enhancing and diminishing gene function in human embryonic stem
707 cells. *Stem Cells* **22**, 2-11, doi:10.1634/stemcells.22-1-2 (2004).

708 31 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**,
709 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

710 32 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137,
711 doi:10.1186/gb-2008-9-9-r137 (2008).

712 33 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing
713 genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033
714 (2010).

715 34 Down, T. A., Piipari, M. & Hubbard, T. J. Dalliace: interactive genome viewing on
716 the web. *Bioinformatics* **27**, 889-890, doi:10.1093/bioinformatics/btr020 (2011).

717 35 Akalin, A., Franke, V., Vlahovicek, K., Mason, C. E. & Schubeler, D. Genomation: a
718 toolkit to summarize, annotate and visualize genomic intervals. *Bioinformatics* **31**,
719 1127-1129, doi:10.1093/bioinformatics/btu775 (2015).

720 36 Zhu, L. J. *et al.* ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and
721 ChIP-chip data. *BMC Bioinformatics* **11**, 237, doi:10.1186/1471-2105-11-237 (2010).

722 37 McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory
723 regions. *Nat Biotechnol* **28**, 495-501, doi:10.1038/nbt.1630 (2010).

724 38 Ou J, Y. J., Kelliher M, Castilla L, Lawson N and Zhu LJ. ATACseqQC: ATAC-seq
725 Quality Control. R package version 1.0.5. (2017).

726 39 Gusmao, E. G., Allhoff, M., Zenke, M. & Costa, I. G. Analysis of computational
727 footprinting methods for DNase sequencing experiments. *Nat Methods* **13**, 303-309,
728 doi:10.1038/nmeth.3772 (2016).

729 40 Schep, A. N. *et al.* Structured nucleosome fingerprints enable high-resolution
730 mapping of chromatin architecture within regulatory regions. *Genome Res* **25**, 1757-
731 1770, doi:10.1101/gr.192294.115 (2015).

732 41 Sickel: A sliding-window, adaptive, quality-based trimming tool for FastQ files
733 Software. Available at <https://github.com/najoshi/sickle>. v. 1.33 (2011).

734 42 Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of
735 insertions, deletions and gene fusions. *Genome Biol* **14**, R36, doi:10.1186/gb-2013-
736 14-4-r36 (2013).

737 43 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat*
738 *Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).

739 44 SeqMonk: A tool to visualise and analyse high throughput mapped sequence data.
740 <http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/> (2007).

741 45 Thomas, P. D. *et al.* PANTHER: a library of protein families and subfamilies indexed
742 by function. *Genome Res* **13**, 2129-2141, doi:10.1101/gr.772403 (2003).

743 46 Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment
744 analysis tool. *BMC Bioinformatics* **14**, 128, doi:10.1186/1471-2105-14-128 (2013).
745

Figure 1

Figure 2

Figure 3

Figure 4