Transcriptional networks are dynamically regulated during cell cycle progression in human Pluripotent Stem Cells

Anna Osnato[1, 2], Ludovic Vallier[1, 2, *]

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

[2]Department of Surgery, University of Cambridge, Cambridge, UK

[*]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 progression [1]. Loops and Topological-Associated Domains (TADs) insulation emerge very early in the G1 phase, but A/B compartments reappear only during the S phase [2]. Global chromatin de-condense in the G1 phase, and refold during S/G2/M [3]. 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 TGFß signalling pathway [4]. 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 pipeline [6]. 88,479

49 reproducible peaks were identified in EG1, 117,283 in LG1, 99,109 in G1/S, 80,398 in S/G2/M
50 (Supplementary Information Fig. 1c), suggesting that chromatin shows high degree of
51 accessibility in all the cell cycle phases in hESCs especially when compared to definitive
52 Endoderm cells (dEN, Vallier lab, unpublished). Interestingly, the number of open regions was
53 higher in the early cell cycle while decreasing during the S/G2/M phase (Fig. 1b), thereby
54 confirming that chromatin could progressively close during progression towards cell division.
55 In total, we could identify 153,007 unique open regions. However, only ~33% of these regions
56 were opened in all the phases (50,461) whereas ~67% of these regions were accessible in either
57 a single phase (~34%) or more than one, but not all (~33%) (Fig. 1c). Conservative pairwise
58 comparisons [7] detected about 1,000 highly dynamic sites (|FC| >1.5 and corrected p-value p-
59 adj < 1E-6), mostly changing through the EG1 phase: 715 regions from EG1 to LG1, 195 from
60 S/G2/M to EG1 (Fig. 1d, e). These observations suggest that chromatin accessibility
61 dynamically changes during cell cycle progression. Interestingly, regions opening in EG1
62 phase were enriched for genes associated with transcriptional regulation, whereas regions
63 closing from S/G2/M to EG1 were enriched for genes related to cell cycle (Supplementary
64 Information Fig. 1d). Thus, regulation of transcriptional networks could occur in the early
65 phase of cell cycle, whereas mechanisms controlling cell division are induced later on. Further
66 analyses showed that chromatin around promoters rapidly open and reorganises after division
67 (Supplementary Information Fig. 1e) thereby suggesting that EG1 phase could be the most
68 dynamic in term of transcriptional regulation. In order to understand which transcription factors
69 are regulating such changes, we performed digital footprinting analyses on the regions showing
70 dynamic chromatin accessibility during cell cycle progression. Combining FLR [8] and
71 Wellington [9] pipelines, we generated a conservative list of 1,807 footprints. Of these, 74%
72 (1,330) were found in EG1, 8% (147) in LG1, 11% (203) in G1/S, and 7% (127) in S/G2/M
73 (Supplementary Information Fig. 1f). In all the phases, apart from the G1/S transition, the top-

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

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

84 We next decided to assess whether changes in chromatin accessibility correlate with gene
85 expression. Therefore, RNA-seq was performed on FUCCI-hESCs sorted in the different cell
86 cycle phases (EG1, LG1, S/G2/M; Fig. 2 and Supplementary Information Fig. 2a, b) and
87 DESeq2 [11] analysis revealed 394 differentially expressed genes (Fig. 2b) in agreement with
88 previous reports [12]. Gene Ontology and pathway analyses showed enrichment for cell cycle
89 terms especially regulators of cell division (Fig. 2c). Hierarchical clustering identified three
90 distinct patterns of expression (Fig. 2b, d). Cluster 1 included genes mostly induced during the
91 S/G2/M phase such as CDK1, CDC25C, AURKA (Supplementary Information Fig. 2c).
92 Cluster 3 included a list of metabolic genes downregulated from EG1 to S/G2/M, suggesting
93 that metabolic activity might decrease during cell cycle progression. In agreement with our
94 finding of highly dynamic chromatin around the EG1 phase, this cluster included several
95 histone genes clusters, and pathway analysis showed enrichment for chromatin rearrangement
96 terms such as promoter opening and histones acetylation/deacetylation (Supplementary
97 Information Fig. 2d). Cluster 2 includes genes upregulated in LG1, some of which are related
98 to endoderm formation and developmental processes, including EOMES, MIXL1, SOX17,

99 FOXA2, GATA4 and GATA6. qPCR analyses confirm the expression of these genes in LG1
100 (Supplementary Information Fig. 2e). However, their expression level remains low when
101 compared to induction observed during endoderm differentiation, suggesting priming of
102 expression rather than full induction (data not shown). In order to assess whether this
103 transcriptional background correlates with changes in chromatin accessibility, we overlapped
104 all of the 13,478 expressed transcripts ($log_2RPM > 1$) with all the associated features of the
105 ATAC-seq peaks: ~85% of expressed genes overlap with accessible regions (Fig. 2e),
106 confirming that most active genes are indeed located in regions of open chromatin. We then
107 overlapped the list of differentially expressed genes with the list of dynamic ATAC regions
108 (from Fig. 1d). Strikingly, only ~3% of differentially expressed genes overlapped with dynamic
109 regions thereby indicating that the majority of cell cycle regulated transcripts originate from
110 constantly open chromatin regions (Fig. 2f). Taken together, these results suggest that
111 differentiation markers are expressed at low level in the G1 phase, potentially marking the
112 priming of hESCs for differentiation. However, this transcriptional "leakiness" originates
113 mostly in constantly open chromatin regions suggesting that it could be induced by differential
114 binding of transcription factors or epigenetic modifiers during cell cycle progression.

115 **Transcription factors binding changes during cell cycle progression in hESCs**

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

124 these results show that CTCF genome wide occupancy changes considerably during cell cycle
125 progression. The increased binding of CTCF in EG1 supports the hypothesis that CTCF could
126 drive chromatin restructuring to regulate transcription [13] and genome integrity [14] during the
127 subsequent cell cycle.
128 In order to check whether this binding pattern could be shared with other transcription factors,
129 we performed ChIP-seq on sorted FUCCI-H9 for the core pluripotency factor OCT4, NANOG
130 and SOX2. For OCT4, we found a total of 4,843 binding sites distributed as follow: 611 in
131 EG1, 1,378 in LG1, and 2,854 in S/G2/M (Supplementary Information Fig. 3c) showing that
132 the number of genomic regions bound by this key pluripotency factor increases during cell
133 cycle progression (Fig. 3e, f). The same binding pattern was observed for NANOG and SOX2
134 with both factors binding the genome more frequently in S/G2/M phase (Fig. 3g,
135 Supplementary Information Fig. 3d). Intrigued by these findings, we asked whether additional
136 factors known to be linked to the core transcriptional network might behave similarly. For that,
137 we performed ChIP-seq for RING1B a chromatin modifier shown to partially bind chromatin
138 in an OCT4-dependent manner [15]. These analyses revealed a unique binding pattern with high
139 occupancy in the EG1 phase, a decrease in the LG1, and an increase in the S/G2/M (Fig. 3h).
140 Interestingly, RING1B is known to have several roles from chromatin compaction to regulation
141 of gene expression [16]. Thus, our observations could imply that these divergent roles might have
142 a temporal distribution during cell cycle progression. Indeed, regions bound by RING1B in
143 EG1 were enriched for CTCF binding sites together with SMC3, a core subunit of the cohesin
144 complex (Supplementary Information Fig. 3e), suggesting that RING1B might have a
145 structural role in opening chromatin at the start of the cell cycle. Interestingly, we also observed
146 a stronger binding of OCT4 at binding sites shared with RING1B in S/G2/M (Supplementary
147 Information Fig. 3f). Taken together, these findings show that transcription factors have

148 different binding patterns during cell cycle progression and that this regulation could define
149 their activity.
150 The dynamic binding of these transcription factors could be explained by a variety of
151 mechanisms including variation in pluripotency factors expression during cell cycle
152 progression. However, RNA-seq, qPCR and western blot validations showed that the
153 expression of these genes/proteins does not change between cell cycle phases (Supplementary
154 Information Fig. 3g, h). Therefore, variations in the quantity of transcription factors do not
155 explain their differential binding. We then hypothesised that post-translational modifications
156 could regulate the differential binding of pluripotency factors. Of particular interest,
157 phosphorylation at Serine 236 (p-OCT4) in hESCs is known to regulate OCT4 binding to
158 chromatin and its transcriptional activity [17], and equivalent to mouse Serine 229, sterically
159 hinders DNA binding [18]. To confirm that such mechanism could control OCT4 binding in EG1,
160 we performed immunostaining for p-OCT4 (Ser236) in FUCCI-hESCs. These analyses showed
161 that mitotic cells have enriched phosphorylation signal (Fig. 3i), thereby confirming findings
162 in mouse Embryonic Stem Cell (mESCs) [19]. However, hESCs in EG1 still partially retain p-
163 OCT4 (Fig. 3i), suggesting that complete de-phosphorylation only happens during the LG1
164 phase. Taken together, these results suggest that phosphorylation could regulate the genomic
165 binding of OCT4 during cell cycle progression in hESCs. Importantly, EG1 has been shown to
166 be a key phase during cell cycle for cell fate decision in hESCs [4]. For this reason, we
167 hypothesised that the reduced capacity of pluripotency factors to bind the genome in EG1 could
168 be necessary for induction of differentiation.

169 **Cell cycle specific OCT4 binding is required for induction of differentiation**

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

173 tamoxifen (4-OHT) inducible form of OCT4 (OCT4-ERT2), that relies on the fusion of OCT4
174 with a modified fragment of the Estrogen receptor [20] (Supplementary Information Fig. 4a, b).
175 Immunostaining confirmed correct nuclear translocation of OCT4 upon addition of 4-OHT in
176 as little as 4h in both pluripotent and neuroectoderm cells (Fig. 4b). We then sorted OCT4-
177 ERT2-FUCCI-hESCs in EG1 and re-plated them in culture condition inducing differentiation
178 [21,22] for 24h in the presence or absence of 4-OHT. Interestingly, OCT4 induction decreases the
179 expression of the key early endoderm markers such as T and MIXL1 (Fig. 4c) suggesting a
180 decreased differentiation. In parallel, we generated an inducible hESCs for the knockdown of
181 OCT4 expression (iKD-OCT4-FUCCI-hESCs), taking advantage of the Tetracycline (TET)
182 based system OptiKD [23] (Supplementary Information Fig. 4c, d). iKD-OCT4-FUCCI-hESCs
183 sorted in S/G2/M were re-plated in culture condition inducing endoderm differentiation for 24h
184 (Fig. 4d), in presence or absence of TET. Interestingly, decrease of OCT4 in the S/G2/M phase
185 promoted the expression of late endoderm (SOX17) and mesoderm (HAND1) markers (Fig.
186 4e) thereby suggesting a more efficient differentiation process. Taken together, these results
187 show that variation in OCT4 binding during cell cycle is required for optimal cell fate decision
188 in hESCs or more broadly, that cell cycle specific regulation of transcriptional networks could
189 control capacity of differentiation in hESCs.
190

191 **Discussion**

192 Our study describes for the first time the dynamic mechanisms that coordinate chromatin
193 accessibility, gene expression, and transcriptional networks during cell cycle progression in
194 hESCs. In the EG1 phase, immediately after cell division, chromatin is highly dynamic and
195 CTCF appear to regulate changes in DNA accessibility. Indeed, CTCF might be the driver for
196 reorganising chromatin after chromosomes condensation during cell division [13] working
197 together with other chromatin modifiers such as RING1B [24]. In parallel, the core pluripotency

198 factors OCT4, SOX2 and NANOG show low occupancy in the G1 phase allowing hESCs to
199 rapidly respond to differentiation cues. In the S/G2/M phase, they fully bind chromatin to re-
200 establish the pluripotency network, sustain self-renewal and block differentiation. These results
201 reinforce previous studies showing that this phase actively promote the pluripotent state [25].
202 This change in binding might be regulated via post-translational modifications such as the
203 phosphorylation preventing OCT4 to bind DNA in EG1 [17]. Importantly, these changes could
204 define the function of RING1B which might help CTCF in restructuring chromatin in EG1
205 while acting as transcriptional regulator in S/G2/M to maintain pluripotency [26]. Ultimately
206 these cell cycle specific interplays are necessary to coordinate induction of differentiation and
207 exit of pluripotency with cell division.
208 Taken together, our results demonstrate that key transcription factors are cell cycle regulated
209 and dynamically change their genomic location. This binding pattern suggests that transcription
210 factors activity and specificity are coordinated with cell cycle progression, and this process
211 could represent a key mechanism that allows stem cells to not only maintain their capacity to
212 change of cellular identity but also for transcription factors to have different functions. These
213 findings highlight the importance of studying transcriptional and epigenetic mechanisms in the
214 context of the cell cycle. More generally, most ChIP-seq studies are currently capturing
215 transcription factors' binding on the most extended cell cycle phase which can vary from cell
216 to cell. Thus, we would like to propose that genomic location of transcription factors should be
217 investigated in a cell cycle dependant manner, since phase-specific analyses could reveal new
218 molecular function even for commonly studied regulators. This novel concept could apply to
219 any cell types and thus have major impact on our understanding of transcription factors'
220 functions.
221
222

Acknowledgements

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

Author contributions

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

Competing Interests

The authors declare no competing interests.

239 # Figure legends

240

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

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

254

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

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

264 score. **(e)** Venn diagram summarising overlap of expressed genes (RNA-seq, log_2RPM > 1)
265 and accessible chromatin (ATAC-seq). 85% represents the percentage of expressed genes
266 present in open chromatin regions (overlapping ATAC peaks). **(f)** Venn diagram reporting
267 overlap of differentially expressed genes (RNA-seq DESeq2) and regions changing
268 accessibility (ATAC-seq DiffRep). 3% represents the percentage of differentially expressed
269 genes to which correspond changes in chromatin accessibility. Combined score is calculated
270 by multiplying unadjusted p-value with z-score.
271
272 **Figure 3. Transcription factors binding changes during cell cycle progression in hESCs.**
273 (**a**) Experimental approach. FUCCI hESCs were sorted in 3 different cell cycle phases: Early
274 G1 (EG1), Late G1 (LG1), and S/G2/M. ChIP-seq was performed following sorting on
275 biological duplicates for CTCF, OCT4, NANOG, and SOX2. **(b)** Heat-maps displaying CTCF
276 binding in the different cell cycle phases EG1, LG1, S/G2/M, compared to INPUT. **(c)**
277 Normalised ChIP-seq signal -/+ 1kb around all peaks in EG1, LG1, S/G2/M, compared to
278 INPUT. **(d)** Venn diagram reporting overlap of CTCF binding sites during cell cycle **(e)** Heat-
279 maps displaying OCT4 binding in the different cell cycle phases EG1, LG1, S/G2/M, compared
280 to INPUT. **(f)** Normalised ChIP-seq signal -/+ 1kb around all peaks in EG1, LG1, S/G2/M,
281 compared to INPUT. **(g)** Heat-maps displaying NANOG and SOX2 binding in the different
282 cell cycle phases EG1, LG1, S/G2/M, compared to INPUT. **(h)** Heat-maps displaying RING1B
283 binding in the different cell cycle phases EG1, LG1, S/G2/M, compared to INPUT. **(i)**
284 Immunofluorescence for phospho-OCT4 Ser236 (pOCT4, magenta) and nuclear staining
285 (DAPI, blue) in FUCCI hESCs (S/G2/M cells in green-mAG, LG1 cells in red-mKO2, EG1
286 cells no colour). Scale bar 20μm. For all the heat-maps, regions are centred on the merged
287 peaks for the two replicates -/+ 1kb in all the phases and compared to INPUT.
288

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

309 # Methods

310

311

312 **hESCs culture and germ layers specification**

313 FUCCI-hESCs (H9/WA09 line; WiCell) were grown feeder- and serum-free as previously
314 described [27]. Briefly, cells were grown in Chemically Defined Medium (CDM, Supplementary
315 Table 2), supplemented with Bovine Serum Albumin (BSA), recombinant Activin-A (10
316 ng/ml) and Fibroblast growth factor 2 (FGF2, 12 ng/ml), both from Dr Marko Hyvonen
317 (Department of Biochemistry, University of Cambridge) on 0.1% Gelatin and MEF-medium
318 coated plates. Cells were passaged weekly using Collagenase IV and plated as clumps of 50-
319 100 cells dispensed at a density of 100-150 clumps/cm^2. All the FUCCI-hESCs lines used were
320 kept under selection with Puromycin (1mM) and Geneticin (G-418, 0.2 mg/ml), or alternatively
321 G-418 (0.2 mg/ml) and Blasticidin (1mM). Germ layer differentiation was initiated 24h after
322 passaging. Definitive Endoderm (dEN) was induced for 3 days in CDM-Polyvinyl alcohol
323 (PVA) without insulin supplemented with 20ng/ml FGF2, 10μM LY294002, 100ng/ml Activin
324 A, and 10ng/ml BMP4, changing media daily, as previously described [21,22]. Neuroectoderm
325 was induced for 6 days in CDM-BSA supplemented with 12ng/ml FGF2 and 10μM SB431542
326 (Activin/Nodal/TGFβ inhibitor; Tocris), as previously described [28]. hESCs were monthly
327 monitored for mycoplasma infection.

328

329 **Immunofluorescence (IF)**

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

334 Supplementary Table 3) in 1% Donkey Serum and 0.1% Triton X-100 PBS overnight at 4°C.
335 Cells were then washed three times with PBS and incubate with the appropriate secondary
336 antibodies (1:1,500; AlexaFluor donkey anti-goat/rabbit/mouse IgG, Invitrogen) in 1% Donkey
337 Serum and 0.1% Triton X-100 PBS 1h at room temperature protected from light. Cells were
338 then washed three times with PBS, for nuclei staining Hoechst (1:10,000, Sigma-Aldrich) was
339 added to the first wash. Images were acquired using Zeiss LSM 700 Confocal Microscope and
340 Zen2009 imaging software.
341
342 **Western Blotting**
343 Cells were harvested by scraping in PBS, centrifuged at 250g for 5 min at 4°C, and re-
344 suspended in ice-cold CellLyticM buffer (Sigma-Aldrich) supplemented with cOmplete
345 Protease Inhibitor (Roche). After 30 min incubation on ice, lysates were clarified by
346 centrifugation at 16,000g for 5 min at 4°C. The supernatants were transferred to new tubes, and
347 proteins in cellular extracts were quantified using Protein Quantification Kit-Rapid (Sigma)
348 according to manufacturer's instructions. 1% β-mercaptoethanol (Sigma-Aldrich) was added
349 to LDS NuPage Sample Buffer (Invitrogen NP0007) and protein denaturated at 95°C for 5 min.
350 Total protein (1-50µg depending on the relative abundance) was loaded on a precast NuPAGE
351 4-12% Bis-Tris protein gel (Invitrogen) and run for 45min, 200V, 130 mA. Proteins were
352 transferred from the gels to Immuno-Blot® PVDF Membranes (Bio-Rad) using the Bio-Rad
353 mini-Trans Blot system for 1h, 100V, 180mA. Membranes were incubated in blocking solution
354 (4% milk, 0.05% Tween-20 in PBS) for 30 min at room temperature, and blotted with the
355 appropriate primary antibody (1:1,000, Supplementary Table 3) overnight at 4°C. Afterwards,
356 membranes were washed three times for 5 min with 0.05% Tween-20 PBS, incubated for 1h at
357 room temperature with the appropriate secondary antibody (1:10,000, anti-mouse/rabbit/goat)

358 conjugated to Horseradish Peroxidase (HRP, Sigma-Aldrich), and developed using the Pierce
359 ECL detection kit (Thermo Scientific) according to manufacturer's instructions.
360
361 **RT-qPCR**
362 RNA was extracted using the GenElute Mammalian Total RNA Miniprep Kit and the On-
363 Column DNase I Digestion Set (both from Sigma-Aldrich) following manufacturer's
364 instructions. cDNA was synthesised using 500ng of RNA in a reaction with 250ng random
365 primers, 0.5mM deoxynucleotides (dNTPs), 20U RNaseOUT, and 25U of SuperScript II (all
366 Invitrogen) in a total of 20µl. cDNA was diluted 30-fold and 5µl were used for qPCR using
367 SensiMix SYBR low-ROX (Bioline) with 150nM Forward and Reverse primers (Sigma-
368 Aldrich; Supplementary Table 4 for primer sequences). Samples were run in technical
369 duplicates on 96-well plates on a Stratagene Mx-3005P (Agilent) or 384-well plate on a
370 QuantStudio Flex Real-Time PCR System, and results were analysed using the delta-delta
371 cycle threshold (ΔΔCt) approach using PBGD as housekeeping gene. Primers were designed
372 using PrimerBlast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/), and tested to produce a
373 single PCR product, with an efficiency between 80% and 120%.
374
375 **Cell Sorting**
376 FUCCI-hESCs were sorted as previously described [4]. Briefly, cells were washed with PBS and
377 collected by incubating with CDB for 10 min at 37°C. When needed, crosslinking of proteins
378 to DNA was performed prior to sorting, according to the specific experiment (see Chromatin
379 Immunoprecipitation). Cells were washed with 1% PBS/BSA and incubated in 1% PBS/BSA
380 with anti-TRA-1-60-antibody mouse monoclonal IgM (1:100, Santa Cruz Biotechnology, Inc.)
381 for 20 min on ice, together with the secondary antibody, Alexa Fluor® 647 goat anti-mouse
382 IgM (1:1,000, Thermo Fisher). Cells were washed with PBS, re-suspended in 6ml of PBS and

383 sorted into the different cell cycle phases using BD FACS AriaTM III, BD Fusion or BD Influx
384 sorters by the Cambridge NIHR BRC Cell Phenotyping Hub. Gating strategy set as following:
385 cells were then size selected for Forward (FS) and Side Scatter (SS), and doublets were
386 removed. TRA-1-60+ cells were then sorted in the four different cell cycle phases according
387 to the FUCCI reporter: Early G1 (EG1) for double negative (mKO2- / mAG-), Late G1 (LG1)
388 for red-mKO2+, G1/S transition for double positive (yellow-mKO2+ / mAG+), and S/G2/M
389 for green-mAG+. Sorted cells were then used for ATAC-seq, ChIP-seq, and RNA-seq. All the
390 experiments have been performed normalising for cell number, usually to the lowest sample,
391 depending on the downstream application.
392
393 **Chromatin Immunoprecipitation (ChIP)**
394 Crosslinking of protein to DNA was performed prior to sorting in 1% PBS/Formaldehyde,
395 rocking for 10 min at room temperature. All the experiments were performed with single
396 crosslinking and neutralised with Glycine (0.125M), for 5 min shaking at room temperature.
397 ChIP was then carried out using LowCell# ChIP kit (C0101007, Diagenode) and DNA
398 recovered using iPure kit (C03010012, Diagenode) according to manufacturer's instructions,
399 with the following modifications. 5µg of antibody were used per IP, and the quantity of
400 magnetic beads adjusted accordingly, based on their binding capacity (10µl of beads for ~3µg).
401 Elution was made in Buffer C from the iPure kit, compatible with qPCR and next generation
402 sequencing. For all the experiments, the same number of cells was used per IP after sorting,
403 normalised for the less abundant cell cycle phase, usually EG1 ~1M cells.
404
405 **Assay for transposase-accessible chromatin (ATAC)**
406 TRA-1-60+ hESCs were sorted in the different cell cycle phases (EG1, LG1, G1/S, S/G2/M).
407 Samples were generated from two independent experiments, 100,000 cells per sample, as

408 previously described [29] with some modifications. In brief, cells were washed with ice-cold PBS
409 and incubated with 500µl of ice-cold sucrose buffer (10mM Tris-Cl pH 7.5, 3mM CaCl2, 2mM
410 MgCl2 and 0.32M sucrose) for 12 min on ice. To collect the nuclei, Triton X-100 was added
411 to a final concentration of 0.5% and cells incubated for additional 6 min on ice after mixing.
412 Samples were then resuspended in 1ml of sucrose buffer and transferred to a fresh 1.5 ml
413 Eppendorf tube. The nuclei were centrifuged at 1500g for 3 min at 4°C and the supernatant
414 discarded. Nuclei pellets resuspended in 50µl of Nextera DNA Sample Preparation Kit
415 (Illumina FC-121-1030) for tagmentation, comprising 25µl 2x Tagment DNA buffer, 20µl
416 nuclease-free water and 5µl Tagment DNA Enzyme 1. The tagmentation reaction mixture was
417 transferred in a 1.5ml low-binding tube and incubated at 37°C for 30 min and stopped by the
418 addition of 250µl Buffer PB (Qiagen). The tagmented chromatin was purified using the
419 MinElute PCR purification kit (Qiagen 28004), according to the manufacturer's instructions,
420 eluting in 11.5µl of buffer EB (Qiagen). For PCR amplification, 10µl of the tagmented
421 chromatin was mixed with 2.5µl of Nextera PCR primer cocktail and 7.5µl of Nextera PCR
422 mastermix (Illumina FC-121-1030) in low-binding PCR tube. Primers were from the Nextera
423 Index kit (Illumina FC-121-1011), using 2.5µl of an i5 primer and 2.5µl of an i7 primer per
424 PCR, totalling 25µl. PCR amplification was performed as follow: 72°C for 3 min and 98°C for
425 30 sec, followed by 12 cycles at 98°C for 10 sec 63°C for 30 sec and 72°C for 3 min. The
426 sample volumes were then increased to 50µl by adding Qiagen EB buffer. The PCR primers
427 were removed with 1x 0.9:1 SPRI beads (Beckman Coulter, Cat no. A63880) according to
428 manufacturer's instruction and samples eluted in 20µl. Following amplification, excess of
429 primers was removed running the samples of 50ml 1% agarose TAE gel at 90V for 25 min,
430 and size selection obtained by cutting from 120bp to 1kb. MinElute Gel Extraction kit (Qiagen
431 28604) was used to extract DNA, eluted in 20µl of buffer EB.
432

433 **Transfection**

434 The generation of the inducible lines was performed using lipofection. Lipofection was
435 performed as previously described [30]. In brief, FUCCI-hESCs cells were seeded into 6 well
436 plates using medium without Penicillin-Streptomycin (Pen/Strep, Gibco), in low density. After
437 48-72h (depending on colony size), cells were transfected using Lipofectamine® 2000
438 Transfection Reagent (Thermo Fisher), according to manufacturer's instruction. In sum,
439 medium was aspirate and replaced with 1mL/well of Opti-MEM® Reduced Serum Media
440 (Thermo Fisher). 4µg of (each) plasmid was resuspended in 0.25ml of Opti-MEM
441 medium/well. 10µl of Lipofectamine 2000 were dissolved in 0.25ml of Opti-MEM
442 medium/well, and incubated for 5 min at room temperature. Lipofectamine 2000 and the
443 plasmid were mixed together and incubated for 20 min at room temperature. 0.5ml of the
444 DNA/Lipofectamine mix was added drop by drop to the well and incubate for 18-20h. Opti-
445 MEM was then replaced with regular media without antibiotics. Antibiotic selection was
446 started 4-5 days after the transfection (Puromycin 1 mg/ml, G-418 50µg/ml). After 6 days,
447 clones were picked into 24 wells plate. Each clone was passaged into 12 well plates, for
448 analysis of transgene expression by qPCR, immunostaining or western blot.

449

450 **Bioinformatics analyses and pipelines**

451 The following data processing pipelines have been used for all the bioinformatics analyses
452 according to ENCODE [6].

453 **ChIP-seq**

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

458 between 100bp and 400bp (average length of 200bp) were used to prepare barcoded sequencing
459 libraries using NEBNext Sample Prep Kit1 (NEB) following manufacturer's instructions. 10ng
460 of input material (as a pool of the different phases) for each experiment was processed in
461 parallel as control. To test whether pooled inputs were an acceptable control, one experiment
462 was run with one input per cell cycle phase. No differences were reported with > 90% peaks
463 shared with the pooled input (data not shown). Equimolar amounts of each library were pooled,
464 and 10 samples/lane were multiplexed on Illumina HiSeq 2000, 2 X 75bp paired-end reads. On
465 average, 400M reads per lane were generated, mapped to human genome GRCh38 reference
466 assembly and stored as cram files. Cram were converted in bam keeping only reads with
467 mapping score above 10 (q>10) using Samtools view [31]. From two independent replicates,
468 consistency between replicates was measured using the Irreproducibility Discovery Rate (IDR)
469 pipeline (https://sites.google.com/site/anshulkundaje/projects/idr). This approach is based on
470 an independent peak-calling for the two replicates with a relaxed cut-off (-p1e-1, --to-large)
471 using MACS [32]. This allows to sample enough noise in the experiment for efficient statistical
472 comparison of replicates in subsequent steps, meaning that many false positives are present in
473 this peak set and they are not considered as final peaks for the separate replicates. Peaks are
474 then ranked for p-value, and the same number is compared based on reproducibility (top
475 100,000 peaks). This is used to score rankings between peaks in each replicate to determine an
476 optimal cut-off for significance. The final peak calls (bed and bigBed format) are the set of
477 peak calls that pass IDR at a threshold of 5-10% max. This is the conservative peak set and it
478 is used for further analyses. If minimum number of recommended reads is not achieved,
479 replicates are pooled. For visualisation purposes, biological replicates were combined.
480 Normalised bedGraph format files were generated for each sample using BEDTools v2.24.0
481 [33]. The reads mapped at both DNA strands from 5' to 3' direction were extended to a length of
482 300bp, and the read-enrichment was normalized by million mapped reads and size of the

483 library. bedGraph files were converted to bigWig using UCSC tool bedGraphToBigWig, and
484 visualised on the Biodalliance genome viewer [34]. For the annotation of genomic intervals,
485 Heatmaps and Metaplot generation, the R packages Genomation and ChIPpeakAnno were used
486 [35,36]. GREAT (Genomic Regions Enrichment of Annotations Tool) was used to associate genes
487 with regulatory regions[37], combined with UCSC liftOver tool to convert bed coordinates from
488 Homo Sapiens GRCh38/hg38 assembly to GRCh37/hg37 assembly
489 (https://genome.ucsc.edu/cgi-bin/hgLiftOver).
490
491 **ATAC-seq**
492 TRA-1-60+ hESCs were sorted in the different cell cycle phases (EG1, LG1, G1/S, S/G2/M).
493 Samples generated from two independent experiments, 100,000 were used per sample, as
494 described above. Library preparation and sequencing were performed at the Wellcome Trust
495 Sanger Institute next-generation sequencing facility. 8 ATAC-seq libraries were prepared with
496 one of i5 and i7 Nextera tags combination and pooled equimolarly. Sequencing was performed
497 on Illumina HiSeq 2000, 2 X 75bp paired-end reads obtaining more than 60M mapped reads
498 per library. Reads were mapped to human genome GRCh38 reference assembly and stored as
499 cram files. Duplicates were removed with Samtools rmdup [31]. Cram were converted in bam
500 keeping only reads with mapping score above 5 (q>5) and from chromosome 1-22 and X (H9
501 is a female line) to remove mitochondrial contamination using Samtools view [31]. From two
502 independent replicates, consistency between replicates was measured using the
503 Irreproducibility Discovery Rate (IDR) pipeline, as described above. Conservative peak set
504 was used for further analyses. Diagnostic plots of fragment size distribution were generated
505 using the R package ATACseqQC [38]. For visualisation purposes, biological replicates were
506 combined as described above to generate normalised bigWig.

507 Differential accessibility was assessed using the G-test implemented in diffReps [7] by Dr Pedro
508 Madrigal. This tool takes into account the biological variations within a group of samples and
509 uses that information to enhance the statistical power. For this, we used a sliding window of
510 600bp, $p < 1E-6$. Regions with differential binding that did not overlap with previously detected
511 peaks through the IDR pipeline were removed. Genomic regions were ranked by their fold
512 change (FC), and reported as differential only if both the $|FC| > 1.5$ and corrected p-value padj
513 $< 1E-6$. Footprinting analyses were done by Dr Pedro Madrigal combining two methods:
514 Footprint Log-likelihood Ratio (FLR) [8] and Wellington [9], adjusted for ATAC-seq. To get high
515 quality reproducible footprints, motif matches with FLR ≥ 10.0 in each biological replicate were
516 classified as "bound". Transcription factor footprints that were present in both replicates, for a
517 motif with at least 50 footprints, were considered as final predicted transcription factor binding
518 sites. FLR mean of the replicates was calculated, only based on unique genomic regions (for
519 those palindromic sequences with a footprint reported in each strand, we obtained the mean
520 FLR). The quality of footprint was assessed calculating the Protection Score, removing scores
521 below 0 [39]. Wellington instead uses a "non-motif centric" approach, which considers the
522 imbalance in this strand-specific information to efficiently identify DNA footprints. For this, a
523 relaxed cut-off has been used (-p32 -A -fdr 0.1 --FDR_limit -4 --pv_cutoffs -4 -fp 4,30,1 -
524 fdriter 500 —one_dimension) on merged replicates (to increase sequencing depth), in order to
525 achieve a significant overlap with FLR, and a final list generated. A conservative list of
526 footprints was generated by overlap between FLR and Wellington, removing not expressed
527 genes according to RNA-seq data (Fragments Per Kilobase of transcript per Million mapped
528 reads, FPKM > 1.0).
529 For visualisation of footprints, bigWigs of ATAC-seq data containing Tn5 insertion
530 frequencies have been generated using NucleoATAC [40] from the bam files (--bam $bamFile -
531 -out $outputBasename --cores 16).

532

533 **RNA-seq**

534 TRA-1-60+ hESCs were sorted in the different cell cycle phases (EG1, LG1, S/G2/M). From
535 three independent experiments, RNA was extracted using GenEluteTM Mammalian Total
536 RNA Miniprep Kit and On-Column DNase I Digestion Set (both Sigma-Aldrich) and a total of
537 9 samples fulfilling the submission requirements (0.5-1.0µg total RNA in nuclease free dH2O
538 with A260/A280 and A60/A230 ratios above 1.8) were sent for library preparation and next-
539 generation sequencing at CGS (Cambridge Genomic Services). PolyA purified library
540 preparation kit was used, and samples were multiplexed in one HIGH Output NextSeq run
541 Cycle, 2 X 75bp, paired-end reads. A total of 462,569,268 reads evenly distributed across the
542 lanes were obtained (50,547,439 on average per sample). Reads were trimmed using Sickle [41],
543 an application for trimming FASTQ files based on quality values (Sanger), with a q=20 (quality
544 threshold) and l=30 (length threshold). For the alignment, the transcriptome was built with
545 TopHat2 [42] (v2.1.0), based on Bowtie2 [43] (v2.2.6) using as reference genome the human
546 GRCh38.p6 and GTF annotation from Ensembl (http://ftp.ensembl.org/pub/release-
547 83/gtf/homo_sapiens/). Alignment was performed using TopHat2 with standard parameters.
548 Using Samtools view[31], reads with mapping quality > 10 were kept for further analyses. Data
549 were quantitated and differential expression analysis was performed using SeqMonk [44]. In sum,
550 data were quantitated using the RNA-seq pipeline, and differential expression analysis was
551 performed using the Intensity Difference Test and the R package DESeq2 [11] (for this,
552 uncorrected raw read counts were used), generating a list of transcripts passing the filter, with
553 a p-value cutoff < 0.05. Hierarchial clustering plots (r > 0.7). Gene Ontology analyses were
554 performed using the Enrichment analysis tool by the Gene Ontology Consortium (Panther) [45],
555 and validated using Enrich [46].

556

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.

561 # Supplementary Information Figure legends

562

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

565 **(a)** Gating strategy for FUCCI hESCs: P1 cells are size selected for forward scatter area (FSC-
566 A) and side scatter area (SSC-A); P2, doublets are removed (FSC-A, FSC-W, width); P3 TRA-
567 1-60+ cells are selected and sorted in the four different cell cycle phases according to the
568 FUCCI reporter: EG1 for double negative (mKO2-/mAG-), LG1 for red-mKO2+, G1/S
569 transition for double positive (yellow-mKO2+/mAG+), and S/G2/M for green-mAG+. For all
570 events, average and standard deviation are reported with n=10. **(b)** ATAC-seq fragment size
571 density plot for the 4 sequenced libraries, EG1, LG1, G1/S, and S/G2/M sorted cells. Inset
572 histograms (log-transformed) show periodicity to four nucleosomes. **(c)** IDR (Irreproducible
573 Discovery Rate) plots reporting reproducible ATAC-seq peaks between two replicates in EG1,
574 LG1, G1/S, S/G2/M. IDR threshold < 5%. **(d)** Gene ontology for Biological Process analysis
575 of differentially accessible peaks in EG1vsLG1 down, EG1vsLG1 up, S/G2/MvsEG1 down,
576 S/G2/MvsEG1 up (diffReps, Fig. 1d). Combined score is calculated by multiplying the
577 unadjusted p-values with the z-score. **(e)** Feature distribution for the ATAC-seq peaks for
578 differentially open regions (diffReps) in EG1vsLG1 down, EG1vsLG1 up, S/G2/MvsEG1
579 down, S/G2/MvsEG1 up. **(f)** Number of predicted footprints overlapping with dynamic regions
580 (ATAC- diffReps) in the different cell cycle phases. Final footprints were obtained combining
581 FLR (Yardimici et al., 2014) and Wellington (Piper et al., 2013) pipelines (combining list with
582 Footprint Likelihood Ratio (FLR) ≥10.0, Protection Score > 0). **(g)** Pairwise comparisons of
583 all cell cycle phases, reporting percentages of footprints for specific transcription factors, log
584 scale axis.

585

586 **Supplementary Information Figure 2. Transcriptome changes during cell cycle**
587 **progression.**
588 **(a)** Datastore Tree clustering of different samples/replicates. (Pearson correlation). **(b)** PCA
589 clustering of FUCCI hESCs sorted in EG1, LG1, S/G2/M. Each dot represents a biological
590 replicate. **(c)** Examples of differentially expressed cell cycle regulator CDK1, CDC25C,
591 AURKA, up-regulated in the S/G2/M phase (p-value cut-off < 0.05). **(d)** Gene ontology for
592 Reactome Pathway analysis for Cluster 3 (Fig.2). Combined score is calculated by multiplying
593 the unadjusted p-values with the z-score. **(e)** RNA-seq log2RPM values for the genes SOX17,
594 EOMES, and MIXL1, validated by q-PCR (values relative to housekeeping gene PBGD).
595
596 **Supplementary Information Figure 3. Transcription factors binding dynamics during**
597 **cell cycle.**
598 **(a)** Venn diagram reporting the overlap of CTCF footprints predicted by ATAC-seq and CTCF
599 ChIP-seq peaks. 60% represent the percentage of footprints overlapping with ChIP-seq peaks.
600 **(b)** Venn diagram reporting the overlap of CTCF ChIP-seq peaks and ATAC-seq peaks. 77%
601 represent the percentage of CTCF ChIP-seq peaks with ATAC-seq peaks. **(c)** Venn diagram
602 showing peaks overlap for OCT4 ChIP-seq in EG1, LG1, S/G2/M. **(d)** Venn diagrams showing
603 peaks overlap for SOX2 and NANOG ChIP-seq in EG1, LG1, S/G2/M. **(e)** Gene ontology for
604 Biological Process and Encode ChEA consensus sequences of RING1b binding sites and
605 RING1B EG1 consensus sequences. Combined score is calculated by multiplying the
606 unadjusted p-values with the z-score. **(f)** qPCR for OCT4, NANOG and SOX2 in EG1, LG1,
607 and S/G2/M relative to housekeeping gene PBGD (two-ways ANOVA, n=3), compared to
608 RNA-seq data for OCT4, SOX2, and NANOG, in EG1, LG1, and S/G2/M, expressed in
609 log2RPM. **(g)** Western-blot for OCT4, SOX and NANOG in EG1, LG1, and S/G2/M, plus
610 unsorted (UNS) control (α-tubulin as loading control).

611

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

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

628

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-see 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. Dalliance: 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).

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

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

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

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

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

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

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

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

Figure 1

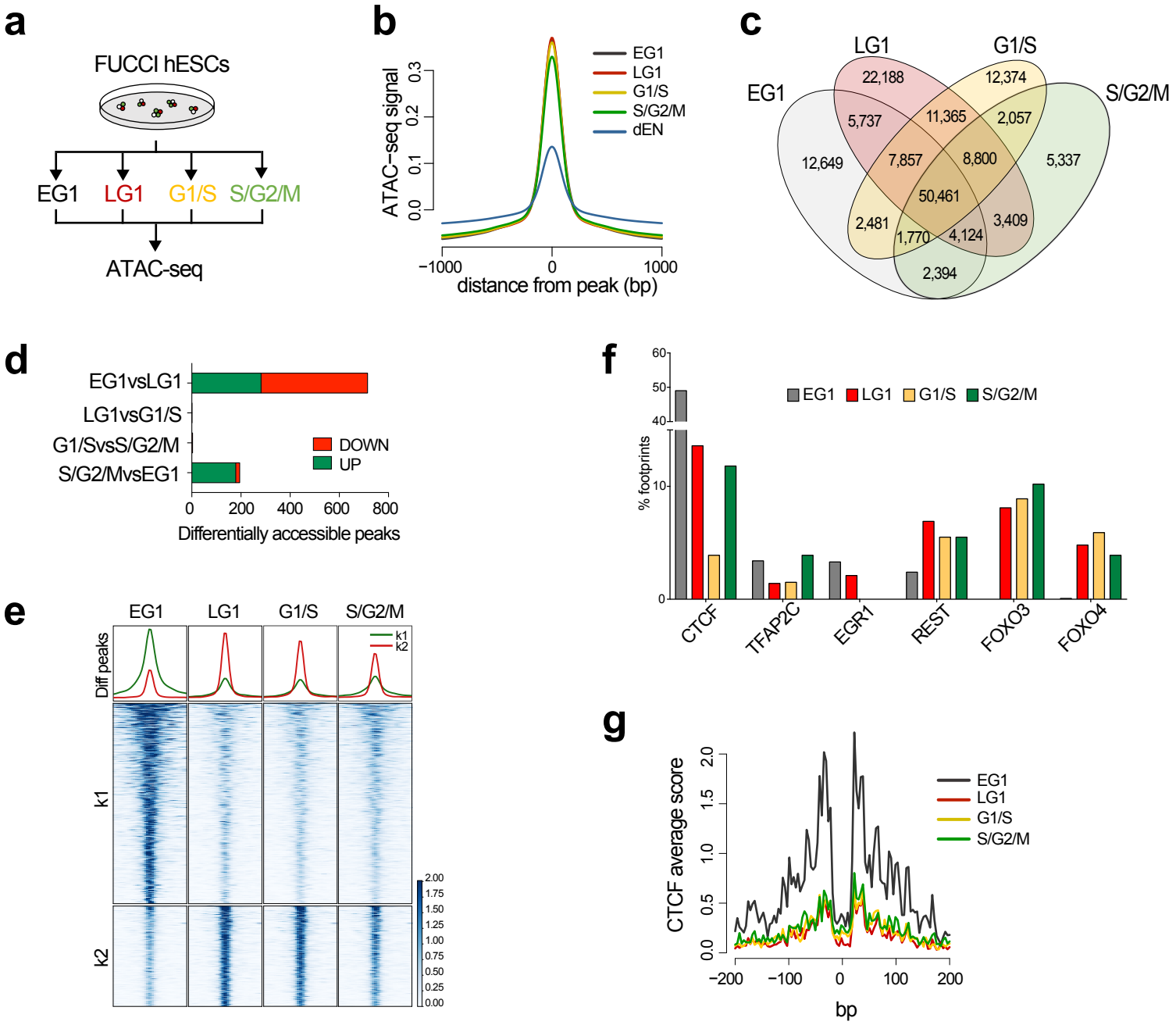

a
FUCCI hESCs
EG1
LG1
G1/S
S/G2/M
ATAC-seq
b
ATAC−seq signal
0.0
0.1
0.2
0.3
−1000
0
1000
distance from peak (bp)
EG1
LG1
G1/S
S/G2/M
dEN
c
EG1
LG1
G1/S
S/G2/M
22,188
12,374
5,737
11,365
2,057
12,649
7,857
8,800
5,337
50,461
2,481
3,409
1,770
4,124
2,394
d
EG1vsLG1
LG1vsG1/S
G1/SvsS/G2/M
S/G2/MvsEG1
DOWN
UP
0
200
400
600
800
Differentially accessible peaks
e
EG1
LG1
G1/S
S/G2/M
Diff peaks
k1
k2
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
f
% footprints
0
10
40
50
60
EG1
LG1
G1/S
S/G2/M
CTCF
TFAP2C
EGR1
REST
FOXO3
FOXO4
g
CTCF average score
0.0
0.5
1.0
1.5
2.0
−200
−100
0
100
200
bp
EG1
LG1
G1/S
S/G2/M

Figure 2

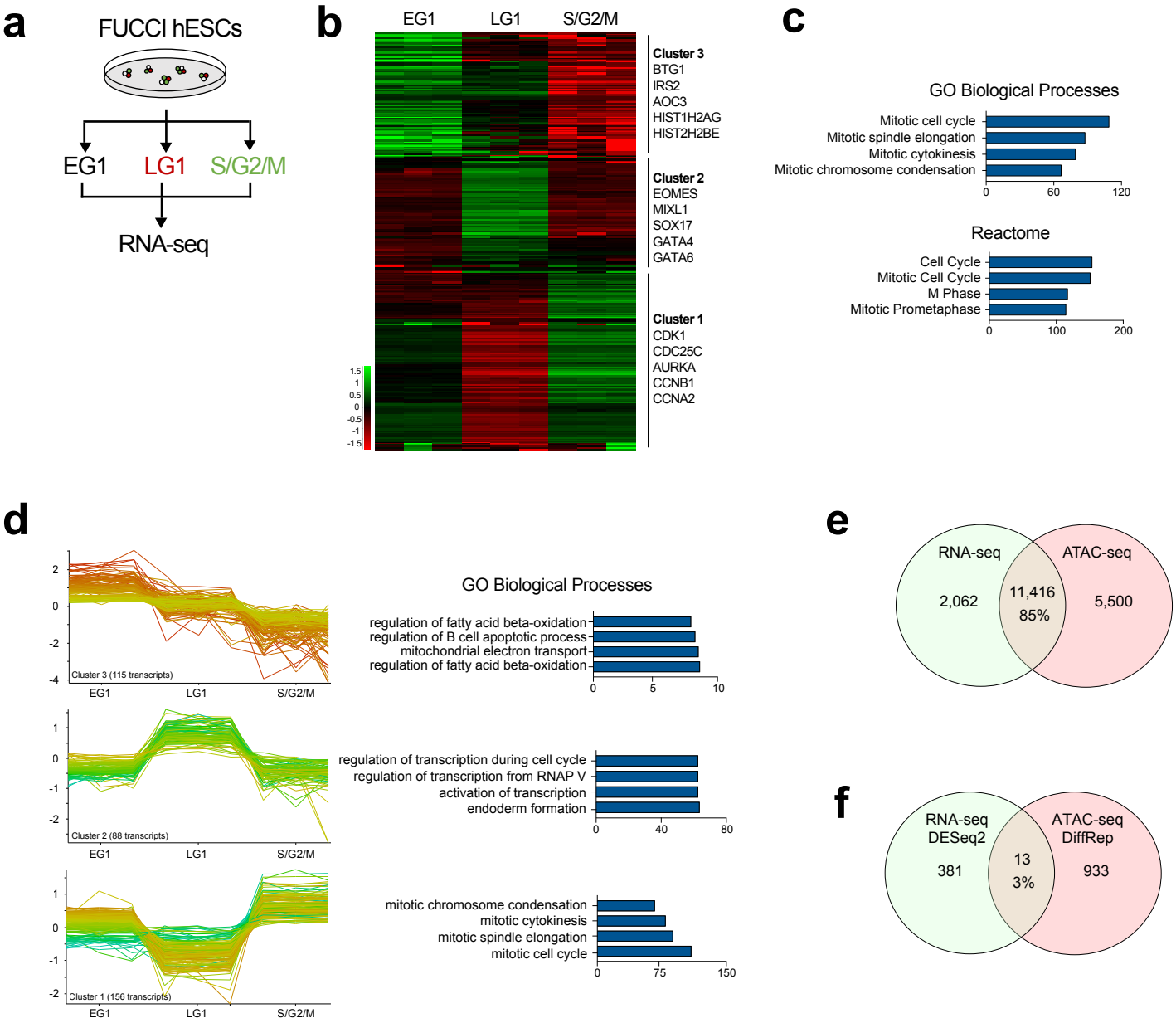

a
FUCCI hESCs
EG1
LG1
S/G2/M
RNA-seq
b
EG1
LG1
S/G2/M
Cluster 3
BTG1
IRS2
AOC3
HIST1H2AG
HIST2H2BE
Cluster 2
EOMES
MIXL1
SOX17
GATA4
GATA6
Cluster 1
CDK1
CDC25C
AURKA
CCNB1
CCNA2
c
GO Biological Processes
Mitotic cell cycle
Mitotic spindle elongation
Mitotic cytokinesis
Mitotic chromosome condensation
Reactome
Cell Cycle
Mitotic Cell Cycle
M Phase
Mitotic Prometaphase
d
Cluster 3 (115 transcripts)
Cluster 2 (88 transcripts)
Cluster 1 (156 transcripts)
GO Biological Processes
regulation of fatty acid beta-oxidation
regulation of B cell apoptotic process
mitochondrial electron transport
regulation of fatty acid beta-oxidation
regulation of transcription during cell cycle
regulation of transcription from RNAP V
activation of transcription
endoderm formation
mitotic chromosome condensation
mitotic cytokinesis
mitotic spindle elongation
mitotic cell cycle
e
RNA-seq
ATAC-seq
2,062
11,416
85%
5,500
f
RNA-seq
DESeq2
ATAC-seq
DiffRep
381
13
3%
933

Figure 3

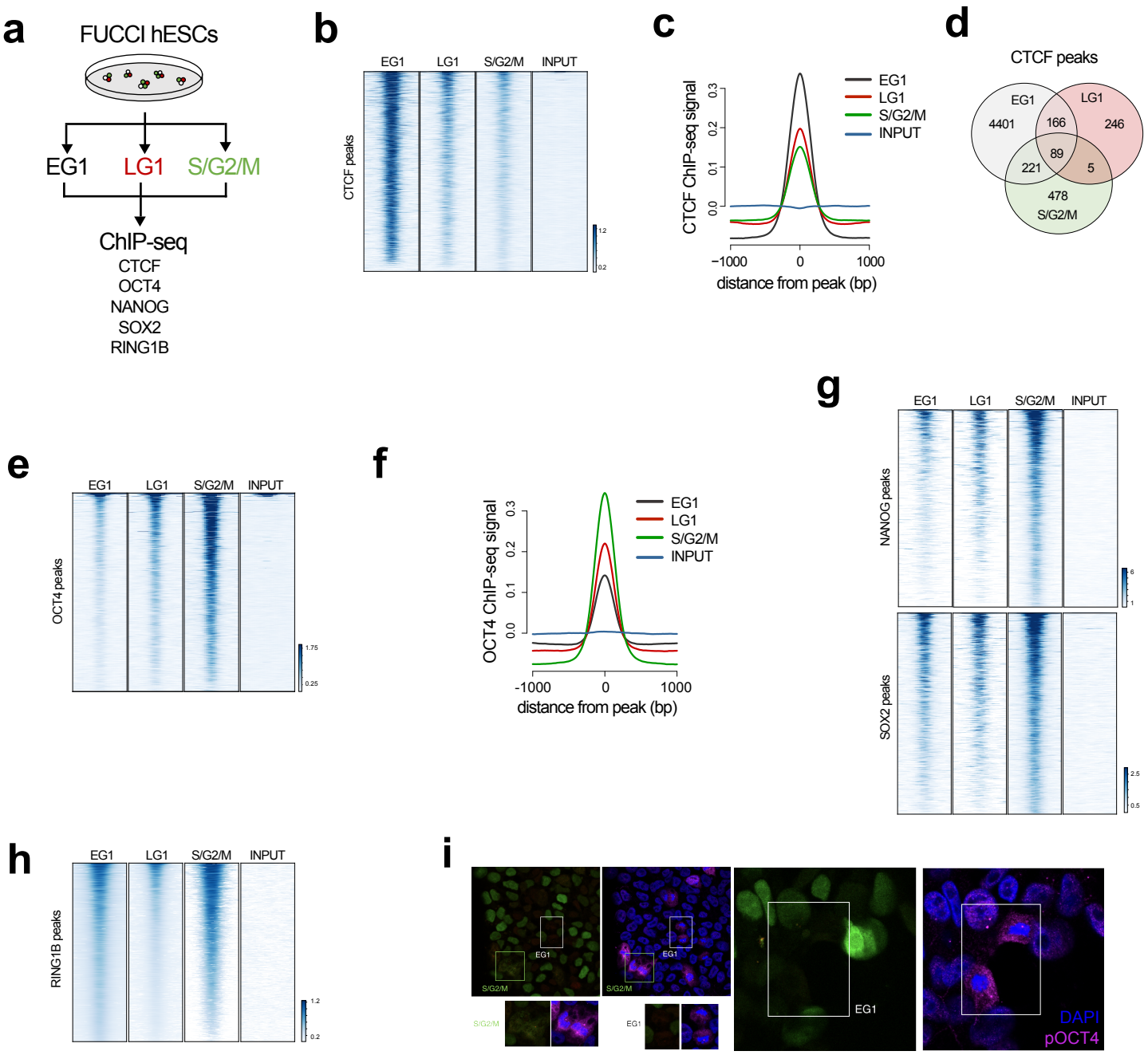
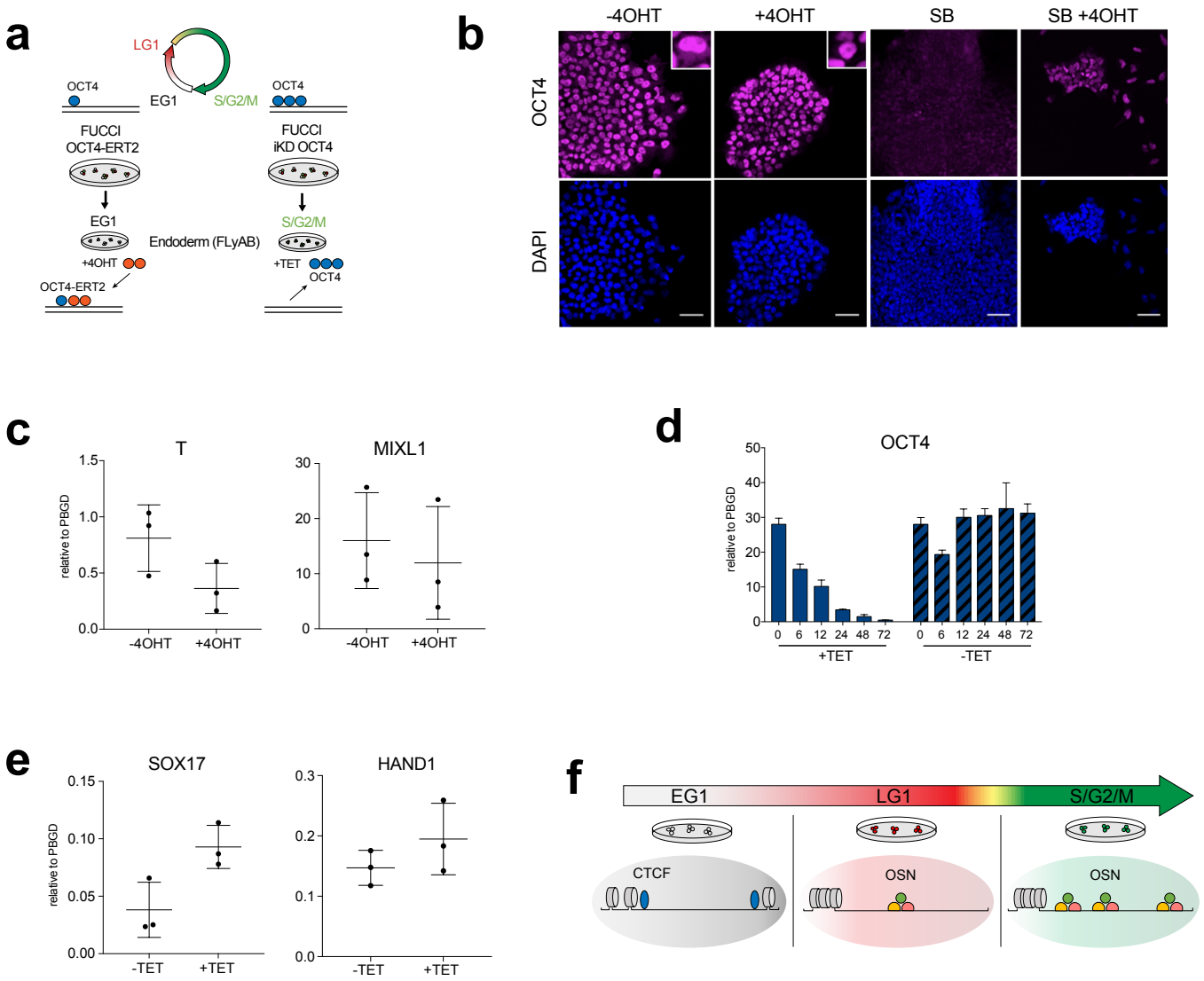

a
FUCCI hESCs
EG1
LG1
S/G2/M
ChIP-seq
CTCF
OCT4
NANOG
SOX2
RING1B
b
EG1
LG1
S/G2/M
INPUT
CTCF peaks
c
CTCF ChIP-seq signal
EG1
LG1
S/G2/M
INPUT
distance from peak (bp)
d
CTCF peaks
EG1
LG1
S/G2/M
4401
166
246
89
221
5
478
e
OCT4 peaks
f
OCT4 ChIP-seq signal
distance from peak (bp)
g
NANOG peaks
SOX2 peaks
h
RING1B peaks
i
DAPI
pOCT4

Figure 4


a
LG1
OCT4
EG1
S/G2/M
OCT4
FUCCI
OCT4-ERT2
FUCCI
iKD OCT4
EG1
S/G2/M
Endoderm (FLyAB)
+4OHT
+TET
OCT4
OCT4-ERT2
b
-4OHT
+4OHT
SB
SB +4OHT
OCT4
DAPI
c
T
MIXL1
relative to PBGD
-4OHT
+4OHT
d
OCT4
relative to PBGD
0 6 12 24 48 72
+TET
-TET
e
SOX17
HAND1
relative to PBGD
-TET
+TET
f
EG1
LG1
S/G2/M
CTCF
OSN
OSN