KLF4 binding is involved in the organization and regulation of 3D enhancer networks during acquisition and maintenance of pluripotency.

Dafne Campigli Di Giammartino[1]*, Andreas Kloetgen[2]*, Alexander Polyzos[1]*, Yiyuan Liu[1], Daleum Kim[1], Dylan Murphy[1], Abderhman Abuhashem[1,3,4], Paola Cavaliere[5], Boaz Aronson[1], Veevek Shah[1], Noah Dephoure[5], Matthias Stadtfeld[6,9], Aristotelis Tsirigos[2,7,8]# and Effie Apostolou[1]#

[1]Sanford I. Weill Department of Medicine, Sandra and Edward Meyer Cancer Center, Weill Cornell Medicine, New York, NY 10021 USA

[2] Department of Pathology, NYU School of Medicine, New York, NY, 10016, USA

[3] Cell and Developmental Biology Program, Memorial Sloan Kettering Cancer Center, New York, New York, USA.

[4] Weill-Cornell/Rockefeller/Sloan Kettering Tri-Institutional MD-PhD program, New York, New York, USA.

[5] Department of Biochemistry, Sandra and Edward Meyer Cancer Center, Weill Cornell Medical College, New York, NY, USA

[6]Skirball Institute of Biomolecular Medicine, Department of Cell Biology and Helen L. and Martin S. Kimmel Center for Biology and Medicine, NYU School of Medicine, New York, NY 10016, USA

[7]Laura and Isaac Perlmutter Cancer Center and Helen L. and Martin S. Kimmel Center for Stem Cell Biology, NYU School of Medicine, New York, NY, 10016, USA

[8]Applied Bioinformatics Laboratories, NYU School of Medicine, New York, NY, 10016, USA

[9] Sanford I. Weill Department of Medicine, Weill Cornell Medicine, New York, NY 10021 USA

* These authors contributed equally

#Corresponding Authors

Lead Contact: efa2001@med.cornell.edu

SUMMARY

Cell fate transitions are accompanied by global transcriptional, epigenetic and topological changes driven by transcription factors (TFs), as is strikingly exemplified by reprogramming somatic cells to pluripotent stem cells (PSCs) via expression of OCT4, KLF4, SOX2 and cMYC. How TFs orchestrate the complex molecular changes around their target gene loci in a temporal manner remains incompletely understood. Here, using KLF4 as a paradigm, we provide the first TF-centric view of chromatin reorganization and its association to 3D enhancer rewiring and transcriptional changes of linked genes during reprogramming of mouse embryonic fibroblasts (MEFs) to PSCs. Inducible depletion of KLF factors in PSCs caused a genome-wide decrease in the connectivity of enhancers, while disruption of individual KLF4 binding sites from PSC-specific enhancers was sufficient to impair enhancer-promoter contacts and reduce expression of associated genes. Our study provides an integrative view of the complex activities of a lineage-specifying TF during a controlled cell fate transition and offers novel insights into the order and nature of molecular events that follow TF binding.

INTRODUCTION

The identity of each cell type is determined by a unique gene expression program as well as a characteristic epigenetic landscape and three-dimensional (3D) chromatin topology. All of these features are under the control and constant supervision of a network of critical transcription factors (TFs), known as master regulators of cell identity[1, 2]. Although the ability of master regulators to maintain or change cell identity is well accepted, the underlying mechanisms remain poorly understood.

Somatic cell reprogramming into induced pluripotent stem cells (iPSCs) by the so-called Yamanaka factors OCT4, KLF4, SOX2 and cMYC (OKSM) offers a tractable system to study the molecular mechanisms of cell fate determination and the roles and activities of each reprogramming TF[3, 4]. Research over the last decade started dissecting on a genome-wide level the transcriptional and epigenetic changes that result in successful erasure of somatic identity and establishment of pluripotency[5-7]. Distinct or synergistic roles of the reprogramming TFs as well as specific direct and indirect mechanisms for coordinating these molecular changes have been proposed [8-14]. In addition to the transcriptional and epigenetic changes, recent studies utilizing targeted or global chromatin conformation capture techniques revealed that the 3D chromatin topology differs between somatic and pluripotent stem cells (PSCs) and is largely reset during reprogramming[15-21]. However, the principles of chromatin reorganization during iPSC generation, its association with enhancer and gene activity and the involvement of TFs in these processes have only started to be explored.

79 Current models regarding the role of reprogramming TFs in 3D chromatin
80 organization are mostly based on computational analyses of 4C or HiC datasets,
81 which reveal a strong enrichment of OKS binding around long-range interactions
82 in PSCs and during reprogramming [15-17, 21]. For KLF4, an architectural function is
83 also supported by experimental evidence. In fact, KLF4 depletion abrogates
84 loops at specific genomic loci such as the *Pou5f1* (*Oct4*) locus in the context of
85 mouse PSCs[18], and the *HOPX* gene in human epidermal keratinocytes[22]. In
86 addition, depletion of the related factor KLF1 disrupts selected long-range
87 interactions in the context of erythropoiesis[23, 24]. These findings establish a link
88 between TF binding and chromatin architecture and suggest that OKS, and
89 particularly KLF4, may actively orchestrate long-range chromatin interactions
90 during reprogramming in order to establish and maintain the pluripotent
91 transcriptional program. To directly test this possibility in a genome-wide manner,
92 we captured the dynamic KLF4-centric topological reorganization during the
93 course of reprogramming and determined the relationships with epigenetic and
94 transcriptional changes. To do so, we used a well-characterized system to
95 reprogram mouse embryonic fibroblasts (MEFs) to iPSCs [14, 25] and applied
96 genome-wide assays that map KLF4 binding (ChIP-seq), chromatin accessibility
97 (ATAC-seq), enhancer and gene activity (H3K27ac ChIP-seq and RNA-seq),
98 enhancer connectivity (H3K27ac HiChIP) as well as KLF4-centric chromatin
99 looping (KLF4 HiChIP) at different stages during acquisition of pluripotency
100 (Fig.1a, top panel). Integrative analysis of our results generated a reference map
101 of stage-specific chromatin changes around KLF4 bound loci and established
102 strong links with enhancer rewiring and concordant transcriptional changes.

103 Inducible depletion of KLF factors in PSCs or genetic disruption of KLF4 binding
104 sites within specific PSC enhancers further supported the ability of KLF4 to
105 function both as a transcriptional regulator and a chromatin organizer.
106

107 **RESULTS**

108 **KLF4 binding during reprogramming induces chromatin opening and**
109 **precedes enhancer and gene activation**

110 To determine the molecular changes around KLF4 targets during iPSC formation,
111 we first mapped its genome-wide binding at different stages of reprogramming
112 using "reprogrammable" MEFs (Rosa26-M2rtTA/Col1a1-OKSM) induced with
113 doxycycline (dox)[25] in the presence of ascorbic acid (Fig.1a, bottom). Under
114 these conditions the resulting iPSCs are molecularly and functionally
115 indistinguishable from embryonic stem cells (ESCs) [26, 27] and, thus, the term
116 PSCs will be used throughout the text to describe either cell type. For our earliest
117 time point, we collected bulk populations on day 3 after dox treatment, whereas
118 at later stages, on day 6 and day 9, we sorted SSEA1^{+} cells to enrich for cells
119 that are on the right trajectory towards induced pluripotency[14, 25, 28, 29]
120 (Supplementary Fig.1a). Finally, we used isogenic ESCs and iPSCs[26, 27] as
121 reference points for established pluripotency. ChIP-seq analysis showed a highly
122 dynamic pattern of KLF4 occupancy during reprogramming with two major
123 categories of binding sites: (i) enriched during intermediate reprogramming
124 stages, but weakly detected in PSCs (Transient KLF4 targets) and (ii) PSC
125 targets, which represent the actual KLF4 binding repertoire once stem cell
126 identity is acquired (Fig.1b and Supplementary Table 1). Among the PSC KLF4

127 binding sites, 30% were already bound on day 3 (Early KLF4 targets), while the
128 rest were either gradually established during reprogramming (Mid KLF4 targets)
129 or enriched only in established PSCs (Late KLF4 targets). To gain insights into
130 the nature and potential function of each category of KLF4 targets, we performed
131 genomic annotation based on their chromatin state classification introduced by
132 Chronis et al[8] as well as Gene Ontology (GO) analysis using the GREAT tool[30]
133 (Fig.1c and Supplementary Fig.1b). Early KLF4 targets mostly enriched for
134 promoters of genes involved in regulation of metabolic processes and cell
135 junction organization, in agreement with the previously reported early role of
136 KLF4 in regulating these processes[14]. On the other hand, Mid and Late KLF4
137 targets included an increasing number of pluripotency-associated enhancers and
138 enriched for stem cell maintenance genes, including many master regulators of
139 pluripotency, such as *Sox2*, *Nanog*, *Esrrb* and *Klf4*. Finally, Transient KLF4
140 targets enriched for enhancers previously detected in partially reprogrammed
141 cells[8] (Supplementary Fig.1b) as well as genes involved in negative regulation of
142 cell cycle, apoptosis and various signaling pathways associated with
143 differentiation, such as TGF-beta signaling (Fig.1c). Therefore, transient KLF4
144 binding might be associated with unsuccessful reprogramming and alternative
145 fates induced by OKSM expression, as reported in other studies [31-33].
146 The differential kinetics of KLF4 binding prompted us to investigate the
147 epigenetic features of KLF4 targets in MEFs and during reprogramming (Fig.1d,
148 Fig.1e and Supplementary Fig.1c). Integration of ATAC-seq and KLF4 ChIP-seq
149 datasets revealed that ~60% of the Early KLF4 binding sites were already open
150 in MEFs, suggesting that preexisting chromatin accessibility could partly explain

151 the early binding of KLF4 on these targets (Fig.1d and Fig. 1e). In contrast, the
152 majority (>70%) of Mid and Late KLF4 targets were characterized by closed
153 chromatin configuration in MEFs (Fig.1d and Fig.1e) and higher DNA methylation
154 levels compared to early targets[12] (Supplementary Fig.1d). These genomic
155 regions gained accessibility concomitantly with KLF4 binding at later timepoints,
156 suggesting the requirement of additional factors for epigenetic remodeling.
157 However, we also observed a large number of inaccessible regions in MEFs that
158 became occupied by KLF4 on day 3 (~40% of early and ~75% of transient
159 targets, Fig.1e), indicating that the ability of this TF to access "closed" sites is
160 context-dependent (Fig.1d-e). In agreement, motif enrichment analysis revealed
161 distinct classes of candidate TFs that may synergize with KLF4 earlier or later in
162 reprogramming to promote its stage-specific binding (Supplementary Fig.1e).

163 KLF4 has been proposed to function both as an activator and repressor of
164 gene expression [11, 14]. To assess the impact of KLF4 binding on enhancer activity,
165 we performed ChIP-seq for H3K27 acetylation (H3K27ac) in MEFs, PSCs and
166 intermediate reprogramming stages and observed evidence for drastic changes
167 in enhancer usage during iPSC generation (Fig.1f and Supplementary Fig.1f).
168 Less than 5% of decommisioned MEF enhancers (regions that lost H3K27
169 acetylation between MEFs and day 3) were targeted by KLF4, whereas about
170 35% of the total acquired PSC enhancers and almost the entirety of (so-called)
171 super enhancers (SE)[34] were bound by KLF4 concomitantly or prior to H3K27
172 acetylation (Fig.1g and Supplementary Table 2). Moreover, RNA-seq analysis
173 (Supplementary Fig.1g and Supplementary Table 3) of genes linked to Early, Mid,
174 Late or Transient KLF4 ChIP-seq peaks showed a strong trend for upregulation,

175 rather than downregulation, at the respective stages of reprogramming
176 (Supplementary Fig.1h). These results suggest that KLF4 binding predominantly
177 results in enhancer and gene activation rather than repression during
178 reprogramming. Representative examples for each category of KLF4 target loci
179 are shown in Figure 1h.

180 In conclusion, our data document stage-specific KLF4 binding with
181 progressive targeting of PSC-associated enhancers, while genes related to
182 "failed" reprogramming trajectories, such as apoptosis or other somatic lineages,
183 were transiently occupied. Globally, the kinetics of KLF4 binding was partly
184 dependent on preexisting chromatin accessibility and DNA methylation levels
185 and either coincided with or preceded enhancer and gene activation.

186

187 **Enhancer interactions are extensively rewired between MEFs and PSCs in**
188 **concordance with epigenetic and transcriptional changes**

189 Previous studies utilizing targeted (4C-seq) or global (HiC) chromatin
190 conformation assays have demostrated that chromatin topology around specific
191 genomic loci and globally at the scale of compartments and Topologically
192 Associated Domains (TADs)[35], are drastically reorganized during
193 reprogramming[15, 17-21]. However, cell type-specific regulatory loops, such as
194 enhancer-promoter interactions, were under-represented in these studies likely
195 due to technical limitations. Here, we performed H3K27ac HiChIP[36] in MEFs and
196 PSCs in order to generate high-resolution contact maps around active enhancers
197 and promoters and characterize the degree of architectural reorganization during
198 reprogramming. We called statistically significant interactions at 10kb resolution

199 and within a maximum range of 2MB using Mango[37] (see Methods) to specifically
200 detect local interactions mediated by H3K27ac. We further refined our set of
201 candidate interactions by considering only loops that overlapped with H3K27ac
202 ChIP-seq peaks in at least one anchor (Supplementary Fig.2a). Differential
203 looping analyses between normalized read-counts (counts-per-million; CPM) of
204 the union of all significant loops called (pvalue<0.1 and Log Fold Change
205 (LogFC) >2 or <-2) revealed about 40,000 contacts that were enriched either in
206 MEFs or in PSCs (Fig.2a and Supplementary Table 4). By applying stringent
207 statistical criteria (pvalue>0.5, logFC<0.5 & logFC>-0.5) , we also identified a
208 group of ~8,000 H3K27ac contacts that show constant interaction strength
209 between MEFs and PSCs. Integration of RNA-seq data showed a significant
210 positive correlation of MEF-specific or PSC-specific H3K27ac loops with
211 increased expression of associated genes in the respective cell type (Fig.2b).
212 These findings demonstrate that H3K27ac HiChIP enables mapping of cell-type
213 specific regulatory contacts and assignment of active enhancers to target genes.
214 Examples of cell-type specific enhancer-promoter contacts are shown in
215 Figure 2c, which illustrates normalized H3K27 HiChIP signals in the format of
216 virtual 4C around *Mycn* and *Ets1*. The promoters of these genes establish high-
217 frequency contacts with distal enhancers (>100kb) in a cell-type specific manner.
218 The position and patterns of the detected chromatin loops are in high
219 concordance with acquisition or loss of H3K27ac marks and the respective
220 transcriptional changes during reprogramming (Fig.2d-e). Importantly, high-
221 resolution 4C-seq analysis around *Mycn* enhancer and *Ets1* promoter (Fig.2f)

222 showed a remarkable similarity with the HiChIP results, validating the cell-type
223 specific nature of HiChIP-detected interactions regardless of H3K27ac.

224 To determine in a global fashion the degree to which differential HiChIP
225 contacts reflect actual chromatin conformation changes, rather than a technical
226 bias due to acquisition or loss of the H3K27ac mark from loop anchors, we
227 performed HiC analysis in MEFs and PSCs. First, we observed that only ~50% of
228 the HiChIP contacts were also detected in HiC of similar sequencing depth (~100
229 million accepted reads per replicate) and using the same loop-calling pipeline
230 (Supplementary Fig.2b). This percentage increased to ~80% when published
231 ultra-resolution HiC data was used[38] (~400 million accepted reads in one
232 replicate, Supplementary Fig.2b), suggesting that sequencing depth is a limiting
233 factor for the ability of HiC to detect HiChIP-enriched loops. Higher local
234 background in HiC might be another limiting factor, as shown by comparing
235 virtual 4C plots of HiChIP and HiC signals around the Tbx3 locus (Supplementary
236 Fig.2c). Examples of contact heatmaps, using HiChIP (Fig. 2g, top) or HiC (Fig.
237 2g, bottom) data, further illustrate this point: although both depict a cell-type
238 specific configuration around select loci (dotted squares around *Jag1* and *Sox2*
239 genes), there are several cell-type specific loops (circled), which are strongly
240 detected by HiChIP, while are weakly detected or fully absent in HiC. Importantly,
241 when we focused on loops that are detected by both approaches, we observed
242 that MEF-specific or PSC-specific HiChIP loops showed significantly stronger
243 HiC signals in the respective cell type, confirming topological reorganization
244 around these regions (Supplementary Fig.2d). These results, in agreement with

245 previous reports[36, 39], highlight the increased sensitivity of HiChIP compared to
246 HiC to detect cell-type specific loops.
247
248 **Complex 3D connectomes in PSCs are associated with strong enhancer**
249 **activity**
250 In addition to simple Enhancer-Enhancer, Enhancer-Promoter and Promoter-
251 Promoter interactions, we observed that many genomic regions were involved in
252 more than one loop. The degree of connectivity, as detected by H3K27ac HiChIP,
253 was significantly higher among PSC-specific loops compared to MEF-specific or
254 constant loops, with hundreds of genomic anchors found to be connected with 10
255 or more (up to 33) distant genes and/or enhancers (Fig.3a and Supplemental Fig.
256 3a). Analysis of HiC data validated the higher connectivity degree of PSCs
257 compared to MEFs (Supplemental Figure 3b), possibly reflecting the more open
258 and plastic chromatin configuration of this cell type[5, 7, 40].
259 Among the highest connected regions in PSCs were critical stem cell regulators,
260 including *Mycn*, *Esrrb*, and *mir290 (*Fig.3a). PSC superenhancers (SE) were also
261 found to be more interactive than typical enhancers (TE)[34] and transcription start
262 sites (TSS) (Fig.3b). Enrichment analysis of HiChIP anchors based on their
263 connectivity degree (low=1 contact vs high = >4 contacts) showed that highly-
264 connected anchors preferentially associate with binding of Pol II, pluripotency
265 TFs, including KLF4, Mediator complex and transcriptional coactivators (Fig.3c)
266 and connect to highly-transcribed genes (Fig.3d). Cohesin subunits and YY1,
267 which was recently decribed to mediate enhancer-promoter loops[41, 42], were also
268 preferentially enriched in highly connected anchors, while the classic

269 architectural factor CTCF[43, 44] was not (Fig.3c). These results suggest that SEs
270 and highly-expressed genes engage in a higher number of chromatin interactions.
271 Importantly, the number of contacts around each enhancer showed poor
272 correlation with the strength of H3K27ac signal (Supplemental Fig.3c),
273 suggesting that our observations are not driven by the biased nature of the
274 HiChIP approach.
275

276 **3D-organized enhancer hubs are associated with coordinated cell-type**
277 **specific gene expression**

278 To gain insights into the biological role of complex enhancer-promoter
279 interactions, we decided to focus on enhancers that establish connections with
280 multiple gene promoters, potentially forming what we refer to as 3D regulatory
281 hubs (or simply enhancer hubs). Genes found within enhancer hubs were
282 enriched for "stem cell maintenance" categories, including many known
283 pluripotency-associated regulators (e.g. *Zic2*, *Etv2*, *Lin28a*, *Dnmt3l*)
284 (Supplemental Fig.4a) and showed significantly higher expression levels
285 compared to genes with a single-connected enhancer (non-hub genes) or all
286 PSC-expressed genes (Supplemental Fig.4b). Many SE that had been initially
287 assigned to a single gene[34, 45] were found to either contact individual novel distal
288 target genes or to form hubs with two or more genes of stem cell relevance (e.g.
289 *Utf1*, *Otx2* and *Nacc1*) and high expression levels (Supplemental Fig.4a-b).
290 These results expand the previous pool of candidate genes that are regulated by
291 superenhancers in PSCs[34, 45, 46]. In addition, they raise the possibility that 3D
292 enhancer hubs may coordinate robust expression of stem cell regulatory genes.

293 To test this hypothesis, we selected all protein-coding genes that participate in
294 hubs (2 or more genes contacting the same enhancer) and are differentially
295 expressed between MEFs and PSCs (FC>2, p-adj<0.01) (Supplementary Table
296 5). We then performed pair-wise comparisons among genes within hubs to
297 calculate the percentage of coregulation (both up- or both down-regulated in
298 PSCs compared to MEFs) or anti-correlation. For control groups we used
299 random gene pairs either of similar linear distance with our test group (global
300 random) or within the same TADs[35] (TAD-matched random). This approach
301 demonstrated a significant overrepresentation of coregulated gene pairs within
302 enhancer hubs compared to all control groups (Fig.4a) and revealed 311 gene
303 pairs that reside within PSC-enhancer hubs and become concordantly
304 upregulated during reprogramming.

305 To experimentally validate transcriptional coregulation within enhancer
306 hubs, we decided to modulate specific hubs and test transcriptional effects. For
307 this, we focused on an enhancer hub that contacts two proximal non-coding
308 genes (*Aw549542* and *Gm16063)* and the distal (~90kb) *Tbx3* gene in a PSC-
309 specific manner (Fig.4b). The PSC-specific nature of the HiChIP-detected
310 contacts was validated by 4C-seq (Fig.4c). H3K27ac ChIP-seq and RNA-seq
311 data showed that all connected genes and enhancers within this hub were
312 inactive in MEFs and reprogramming intermediates and became activated only in
313 PSCs, supporting coordinated activation within the hub (Fig.4d-e). Of note, this is
314 not the case for a gene outside the hub (~800kb), Med13l (Supplemental Fig.4c).
315 Using CRISPR/Cas9 technology[47] we deleted the distal *Tbx3* enhancer in PSCs,
316 using a deletion of a previously characterized proximal enhancer[48] – which is

317 also part of the same hub – as a reference (Fig. 4f and Supplementary Fig.4d).
318 RT-qPCR analysis of homozygous knock-out (KO) clones demonstrated that the
319 transcriptional levels of *Tbx3* were severely impaired upon disruption of either
320 enhancer (Dis-KO and Pro-KO), with the distal enhancer showing a stronger
321 effect (Fig.4f). Interestingly, the RNA levels of the other hub-connected genes
322 (*Gm1603* and *Aw549542)* were also reduced, while expression of *Med13l* was
323 unaffected (Fig.4f). Furthermore, we used dCas9-KRAB[49] to target a different
324 enhancer that contacts *Zic2* and *Zic5* genes (Supplemental Fig.4e), which are
325 also coactivated during reprogramming (Supplemental Fig.4f and 4g). CRISPRi-
326 mediated silencing of this enhancer (Supplemental Fig.4h-i) resulted in significant
327 downregulation of both genes, while non-hub genes in linear proximity were only
328 modestly affected (Supplemental Fig.4j).
329

330 **KLF4-centered chromatin reorganization during reprogramming associates**
331 **with enhancer rewiring and transcriptional changes of target genes**

332 Integration of H3K27ac HiChIP results with KLF4 ChIP-seq demonstrated that
333 Early, Mid and Late KLF4 targets (see Fig. 1b) were enriched in PSC-specific
334 H3K27ac interactions, while MEF-specific contacts enriched for transient KLF4
335 binding (Fig.5a). These results raise the possibility that KLF4 binding is involved
336 in 3D enhancer reorganization during reprogramming. To directly capture the
337 topological changes around KLF4-occupied sites during iPSC formation, we
338 performed KLF4 HiChIP in early (day 3) and mid (day 6) stages of
339 reprogramming as shown in Fig.1a and in PSC. Principle component analysis
340 (PCA) on all statistically-significant interactions called by Mango[37] distinguished

341 KLF4-bound loops from H3K27ac-marked loops (Supplementary Fig.5a),
342 demonstrating the different nature of chromatin contacts that each antibody
343 captures. Differential looping analysis generated four clusters of dynamic KLF4-
344 centered interactions (Fig. 5b and Supplementary Table 6): two clusters of
345 gained loops, detected either in mid or late reprogramming stages and two
346 clusters of lost loops detected only in early or mid stages.

347 To gain insights into the role and nature of the different KLF4-centered
348 loop clusters we investigated the expression changes of associated genes during
349 reprogramming. We found that lost KLF4 HiChIP contacts mostly associate with
350 gene repression, while gained KLF4 loops correlate with gene activation during
351 reprogramming (Fig.5c). Accordingly, comparison with H3K27ac HiChIP data
352 showed that >40% of the lost KLF4 contacts were actually MEF enhancer loops,
353 while >50% of gained KLF4 loops overlapped with PSC enhancer interactions
354 (Fig.5d). Together, these observations support a role of KLF4 binding in the
355 formation/activation of PSC enhancer loops and abrogation/repression of pre-
356 existing somatic loops.

357 To better understand the relative effect of KLF4 binding and/or looping on
358 gene activation, we focused on enhancer-promoter loops detected by both KLF4
359 and H3K27ac HiChIP in PSCs and clustered them as: (i) early bound by KLF4
360 and early formed loops during reprogramming (day 3), (ii) early bound, but late
361 formed loops and (iii) late bound and late formed loops (Supplementary Fig. 5b,
362 left panel). Genes within the first category were robustly upregulated early during
363 reprogramming, while genes in the other two categories were activated only at
364 the late reprogramming stages (Supplementary Fig.5b, right panel). These

365 results indicate that looping coincides with gene activation while KLF4 binding
366 *per se* is not always sufficient to establish promoter-enhancer contacts and
367 activate transcription.
368
369 **KLF4 binding engages in both activating and repressive loops in PSCs**
370 Our analysis showed that about 30% of dynamic KLF4-centered loops did not
371 associate with any expression changes and did not overlap with enhancer
372 contacts (Fig.5c,d). Among all KLF4-centered loops in PSCs, 74% overlaps with
373 H3K27ac HiChIP contacts (H3K27ac-dependent), while 26% are H3K27ac-
374 independent (Supplementary Fig.5c). Enrichment analysis using LOLA showed
375 that KLF4 binding sites within H3K27ac-dependent loops are enriched for active
376 enhancer features such as binding of pluripotency TFs (ESRRB, NANOG, SOX2
377 and POU5F1), YY1 as well as RNA Pol II, co-activators, Cohesin and Mediator
378 subunits (Fig.5e). In contrast, H3K27ac-independent KLF4 anchors are enriched
379 for Polycomb repressive Complex 1 and 2 (PRC1 and PRC2) components, which
380 have been reported to mediate looping among repressed or bivalent genes in
381 PSCs[17, 50, 51]. Genes within H3K27ac-independent KLF4 loops were expressed at
382 significantly lower levels compared to the genes in H3K27ac-dependent loops
383 (Supplemental Fig.5d) and enriched for Gene Ontology categories associated
384 with development and lineage specification (Supplemental Fig.5e). These
385 findings raises the possibility that KLF4 is engaged in chromatin loops with
386 distinct properties and functions, possibly by interacting with different
387 architectural cofactors.

388 To test the chromatin co-occurrence of KLF4 with computationally-predicted
389 cofactors, we performed RIME[52] (Rapid Immunoprecipitation Mass spectrometry
390 of Endogenous proteins) in PSCs using either a KLF4 antibody or IgG as control
391 (Fig.5f). This identified 228 high-confidence (FC>1.5 over IgG and p-value<0.05)
392 protein partners (Supplementary Table 7). In addition to novel candidates, RIME
393 detected several of the predicted cofactors, including components of the Cohesin
394 complex, PRC1 and PRC2 as well as co-activators, such as BRD4.
395 Immunoprecipitation using PSC extracts followed by Western blot analysis
396 validated interaction of KLF4 with selected candidates (Fig.5g). These results
397 support the notion that KLF4 participates in different categories of loops in PSCs
398 (Supplemental Fig.5f): (i) activating chromatin loops that are enriched in Cohesin,
399 coactivators and other pluripotency TFs and engage highly-expressed genes
400 involved in cell cycle and stemness (e.g. *Nodal, Mycn, Pou5f1, Dppa2*); (ii)
401 repressive loops mediated by PRC1 and PRC2 components that involve genes
402 related to cell differentiation and development (e.g. *Hoxd10, Bmp4, Serpine3,*
403 *Fgf9*).
404

405 **Depletion of KLF factors in PSCs disrupts a subset of enhancer loops and**
406 **expression of linked genes**

407 To dissect the role of KLF4 in the 3D enhancer connectome of pluripotent cells,
408 we generated an ESC line that enables dox-inducible targeting of *Klf4* by
409 CRISPR-Cas9. Although KLF4 protein levels were successfully reduced 48 hours
410 after dox addition (Supplementary Fig.6a), we noticed that transcriptional levels
411 of *Klf2* and *Klf5*, encoding TFs with partially redundant function to KLF4[53], were

412 upregulated in these cells, suggesting compensatory mechanisms
413 (Supplementary Fig.6b). We therefore targeted all three KLF factors using the
414 same conditional system. Shortly after dox induction (24 hours), when the levels
415 of KLF proteins were successfully reduced but before other pluripotency factors
416 such as NANOG were affected (Supplementary Fig.6c), we performed H3K27ac
417 HiChIP and ChIP-seq as well as RNA-seq (Supplementary Table 8). Comparison
418 of enhancer connectomes in uninduced (WT) and induced (triple KO, TKO) cells,
419 revealed 7024 contacts which were consistently reduced (lost) in all TKO
420 replicates and 3488 newly established loops (Fig.6a). The observation that the
421 majority of contacts remained unaffected might be due to residual KLF protein
422 levels (Supplemental Fig.6c) during the intentionally short treatment with dox
423 and/or indicate the presence of additional factors that maintain enhancer
424 architecture and activity. More than 60% of lost loops were bound by KLF4
425 (ChIP-seq) on one or both anchors, indicating that disruption of these loops is
426 likely a direct effect of KLF factors downregulation (Fig.6b). Of note,
427 multiconnected hubs and superenhancers were preferentially affected compared
428 to typical enhancers, showing a significant reduction in the number of interactions
429 (Supplemental Fig.6d).
430 Integration of RNA-seq data showed that genes within lost or gained loops were
431 significantly down- or up-regulated, respectively, in TKO compared to WT cells
432 (Fig.6c). The relatively moderate transcriptional changes may reflect the short
433 dox-treatment and/or RNA stability. Examples of lost loops, represented as a
434 virtual 4C of H3K27ac HiChIP data in WT and TKO cells, along with the
435 respective KLF4 and H3K27ac ChIP-seq tracks are shown in Figure 6d and 6e.

436 The reduced mRNA levels of *Klf8*, *Fgf17* and *Eif2s2* genes and the disruption of
437 the respective gene-enhancer contacts in TKO cells were independently
438 validated by RT-qPCR and 3C-qPCR, respectively (Fig.6f and 6g). These results
439 demonstrate that depletion of KLF factors in PSCs results in abrogation of
440 thousands of enhancer contacts genome-wide and concordant dysregulation of
441 connected genes.
442

443 **Disruption of KLF4 binding sites interferes with enhancer looping and**
444 **transcriptional activation**

445 To ascertain whether KLF4 binding is critical for maintenance of 3D enhancer
446 contacts in PSCs, we targeted KLF4 binding sites within selected enhancer hubs
447 and examined local topological and transcriptional effects. We initally chose the
448 distal *Tbx3* enhancer, deletion of which resulted in downregulation of all three
449 hub-connected genes (Fig.4f). The multiple contacts of this enhancer with the
450 surrounding genes were detected both by H3K27ac and KLF4 HiChIP only in
451 PSCs but not in MEFs or reprogramming intermediates (Fig.7a). This is in
452 concordance with the late binding of KLF4 to this enhancer (Fig.7b) and the late
453 transcriptional activation of the entire locus (Fig.4e). We utilized CRISPR/Cas9
454 technology to disrupt the strongest KLF4 binding motif within this enhancer hub
455 (Fig.7c and Supplemental Fig.7a-c). Four different homozygous mutant clones
456 were validated for impaired KLF4 binding by ChIP-qPCR (Supplemental Fig.7d)
457 and used for further characterization. RT-qPCR analysis demostrated that the
458 transcriptional levels of all hub-connected genes *(Aw549542, Gm1603* and *Tbx3*)
459 were significantly reduced, whereas the expression of a gene outside the hub

460 was not affected (*Med13l*) (Fig.7d). Consistent with transcriptional
461 downregulation, the long-range contacts between the enhancer hub and its target
462 genes were significantly weakened in mutant clones as shown by 3C-qPCR
463 (Fig.7e), while the interaction of *Tbx3* with the proximal enhancer or a KLF4-
464 independent contact in a different genomic region remained unaffected (Fig.7e).
465 Using a similar approach, we mutated a strong KLF4 binding site within the
466 previously described *Zfp42* superenhancer[54], which contacts both *Zfp42* and the
467 distal (~150kb) *Triml2* gene in a PSC specific manner (Fig.7f-g). Homozygous
468 mutant ESCs showed significant downregulation of *Zfp42* expression and a
469 concordant reduction of enhancer-*Zfp42* promoter contact frequency *(*Fig.h-j*)*.
470 Intringuingly, the expression levels of *Triml2* remained unaffected in the mutant
471 clones and the connection with the enhancer appeared even stronger (Fig.7i-j),
472 suggesting that KLF-dependent and independent mechanisms may regulate
473 looping and activity of the same enhancer on different genes. Taken together,
474 these results provide evidence for a dual role of KLF4 as a transcriptional
475 regulator and chromatin organizer in PSCs.
476

477 **DISCUSSION**

478 Here, we describe the genome-wide dynamics of KLF4 binding and probe its
479 effects on chromatin accessibility, enhancer activity, gene expression and 3D
480 enhancer organization during iPSC reprogramming and in established PSCs. Our
481 data suggest that the kinetics of KLF4 binding and the temporal relationship with
482 gene and enhancer activity is partly dependent on preexisting chromatin
483 accessibility, the presence of epigenetic barriers such as DNA methylation and/or

484 the availability of additional TFs and cofactors, such as ESRRB or NANOG[8, 12, 13].
485 Nevertheless, KLF4 also binds to chromatin regions that are inaccessible and
486 highly methylated in somatic cells, which is in agreement with its documented
487 ability to act as a pioneer factor and induce chromatin opening and DNA
488 demethylation[9, 55, 56] and/or its cooperative binding with other reprogramming
489 TFs[8].

490 Previous studies utilizing 4C or HiC have characterized dynamic 3D
491 architectural changes during reprogramming either at a small-scale, around
492 specific genomic sites[15, 18], or at a large-scale, mostly at the levels of
493 compartments and domains[21]. These studies offered important insights into the
494 principles of topological reorganization during cell fate transitions, but they did
495 not capture the dynamic assembly and disassembly of cell-type enhancer
496 contacts. Here, we chose to apply H3K27ac HiChIP analysis, which was reported
497 to have significantly higher discovery rate for cell-type specific loops compared to
498 HiC and Capture HiC methods[36, 39]. Indeed, our data revealed dramatically
499 rewired enhancer connectomes between MEFs and PSCs generating a
500 reference map of cell-type specific regulatory loops. Independent 4C-seq and
501 HiC experiments largely validated the cell-type specific nature of the detected
502 HiChIP interactions, but also revealed technical biases and limitations for each
503 approach, highlighting the need for a deeper and systematic comparison of
504 different 3C assays and analytical tools. Our H3K27ac HiChIP analysis
505 uncovered a set of highly-connected enhancers, which communicate with
506 strongly expressed cell-type specific genes, supporting that high interactivity
507 might be an inherent characteristic of critical regulatory elements for cell identity,

508 as it has been suggested in previous studies[57, 58]. Moreover, we identified a
509 number of cell-type specific enhancers, including many SE, which frequently
510 interact with two or more coregulated genes, supporting a potential role for such
511 hubs in coordinating target gene activation, as previously shown in different
512 contexts[59]. In further support, deletion or inactivation of enhancer hubs resulted
513 in coordinated downregulation of all connected genes without affecting
514 neighboring non-hub genes. Recently developed technologies that capture
515 multiway interactions[60-63] will enable dissecting to what extent these enhancer
516 hubs represent multiple contacts occuring in the same cell and allele or highly
517 dynamic contacts with one gene at a time. In either case, our results provide
518 genome-wide evidence for the role of selected enhancers in coordinating gene
519 regulation during acquisition and maintenance of pluripotency and demonstrate
520 the potential of this approach to identify novel candidate genes and enhancers
521 critical for specific cellular identities.

522 There is increasing evidence that TFs are involved in mediating chromatin
523 contacts in different cellular contexts[21, 38, 39, 41, 42, 64-69], although the underlying
524 mechanisms and the temporal relationships between TF binding and topological
525 and transcriptional changes remain elusive. Encouraged by previous studies
526 reporting potential architectural functions for various KLF protein members[18, 23, 24],
527 we went on to capture for the first time in a direct and genome-wide manner the
528 dynamic chromatin reorganization around KLF4-binding sites during iPSC
529 formation by KLF4 HiChIP. This approach revealed that KLF4 binding associated
530 with *de novo* establishment of enhancer loops during reprogramming, promoting
531 transcriptional upregulation of linked genes. We also observed that KLF4 binding

532 was not always sufficient for looping formation and gene activation, suggesting
533 the requirement of additional architectural factors and coregulators. In support of
534 this notion, our computational and proteomics analyses revealed distinct sets of
535 candidate cofactors that interact with KLF4 protein either in the context of
536 activating enhancer loops or repressive/poised loops in PSCs. How these
537 proteins work together to form 3D chromatin contacts remains to be shown.
538 Recruitment of architectural cofactors capable to physically tether distal DNA
539 elements is a plausible scenario and is supported by the fact that KLF4 directly
540 interacts with cohesin subunits[43]. Another possibility is that formation of activating
541 or repressive topological assemblies, such as 3D enhancer hubs or polycomb
542 bodies[17, 50, 57, 62, 70, 71], is the result of "self-organization" through multiprotein
543 condensation. In support of this model, KLF4 and validated cofactors, such as
544 Mediator and BRD4, are charaterized by extensive intrinsically disordered
545 regions (IDRs), which have been shown to promote multivalent interactions and
546 formation of subnuclear condensates[72-75].

547 In contrast with previous studies that described the involvement of KLF4 in
548 the maintenance of selected chromatin loops[18, 22], our study provides evidence
549 for a functional role in the organization and regulation of 3D enhancer contacts
550 and hubs in PSCs at a genome-wide scale. In addition to the global topological
551 effects induced by KLF protein depletion, we showed that targeting individual
552 KLF4 binding sites within specific enhancer hubs was -in some cases- sufficient
553 to disrupt enhancer-promoter contacts and induce downregulation of associated
554 genes. Systematic functional interrogation of KLF4-bound enhancer hubs as
555 identified by HiChIP may enable a deeper understanding of KLF4-dependent and

556 independent mechanisms of topological organization and the establishment of
557 new criteria for identification and functional prioritization of critical regulatory
558 nodes for PSC identity.
559

560 **ACKNOWLEDGEMENTS**

561 We are grateful to Ari Melnick and members of the Apostolou, Tsirigos and
562 Stadtfeld lab for critical reading of the manuscript. We also want to thank Luke
563 Dow for sharing the CRISPR/Cas9 vectors. DCG was supported by the New
564 York Stem Cell Foundation and the Family-Friendly Postdoctoral Initiative at
565 Weill Cornell Medicine. AA is supported by a Medical Scientist Training Program
566 grant from the National Institute of General Medical Sciences of the National
567 Institutes of Health (NIH) under award number T32GM007739 to the Weill
568 Cornell/Rockefeller/Sloan Kettering Tri-Institutional MD-PhD Program. This work
569 was funded by the NIH Director's New Innovator Award (DP2DA043813) and the
570 Tri-Institutional Stem Cell Initiative by the Starr Foundation.

571 **AUTHOR CONTRIBUTIONS**

572 EA conceived, designed, supervised the study and wrote the manuscript together
573 with DCDG with help from all authors. DCDG performed all experiments with help
574 from DK and VS. AK and APP performed all HiChIP and HiC analyses and
575 integrative computational analysis under the guidance of AT. YL performed initial
576 ChIP-seq, RNA-seq and ATAC-seq analysis. DM performed HiC and CRISPRi
577 experiments using a stable dCas9-KRAB ESC line generated by BA. AA
578 performed RIME experiments and iPSC ChIP-seq. PC and ND run and analyzed

579 RIME results. MS provided reprogrammable cells and guidance on
580 reprogramming experiments.
581
582 **DATA AVAILABILITY STATEMENT**
583 All data (RNA-seq, ChIP-seq, ATAC-seq and HiChIP) were submitted to GEO
584 under the accession code GSE113431.
585
586 **COMPETING FINANCIAL INTERESTS**
587 The authors declare no competing financial interests.
588
589 **METHODS**

590 **Cell lines, culture conditions and reprogramming experiments**

591 Mouse ES V6.5 were cultured on irradiated feeder cells in KO-DMEM media
592 (Invitrogen) supplemented with 15% heat-inactivated fetal bovine serum,
593 GlutaMAX, penicillin-streptomycin, non-essential amino acids, β-mercaptoethanol
594 and 1000 U/ml LIF, with or without the presence of 2i (1uM MEKinhibitor
595 (Stemgent 04-0006) and 3uM GSK3 inhibitor (Stemgent 04-0004)).

596 Mouse embryonic fibroblasts (MEFs) were isolated from a "reprogrammable"
597 mouse harboring a polycystronic OKSM cassette in the *Col1a1* locus and M2rtTA
598 in the *Rosa26* locus[25]. Cells were reprogrammed in the presence of 1ug/ml
599 doxycycline and 50ug/ml ascorbic acid and cultured in ES medium as described
600 above. Cells were collected at the indicated time points.

601 **Lentiviral production and infection**

602 293T cells were transfected with overexpression constructs along with packaging
603 vectors VSV-G and Delta8.9 using PEI reagent (PEI MAX, Polyscience #24765-
604 2). Supernatant was collected after 48hrs and 72hrs and virus was concentrated
605 using Polyethylglycol (PEG, Sigma # P4338). V6.5 cells were infected in medium
606 containing 5ug/ml polybrene (Millipore, TR-1003-G) followed by centrifugation at
607 2100rpm for 90 mins at 32°C.

608 **MACS and FACS**

609 For isolating the SSEA1 positive cells from reprogramming intermediates at day6
610 and day9 we used magnetic microbeads conjugated to anti-SSEA1 antibody
611 (MACS Miltenyi Biotec #130-094-530) as per manufacture instructions. SSEA
612 positive and negative fractions were then stained for FACS analysis with an anti-
613 Thy1 antibody conjugated to pacific blue fluorophore (ebioscience # 48-0902-82)
614 and anti-SSEA antibody conjugated to APC fluorophore (biolegend #125608).

615 **Generation, selection and validation of KO cell lines**

616 gRNAs were cloned into the px458 vector (Addgene #48138) using the BbsI
617 restriction enzyme. 0.3 million ESC cells (V6.5) were transfected using 2ug of
618 Left-*Tbx3*-plasmid and 2ug of Right-Tbx3-plasmid (for *Tbx3* enhancer deletions)
619 or 4ug of *Tbx3*-KLF4mut-vector (mutation of KLF4 binding site within *Tbx3* distal
620 enhancer) or 4ug KLF4-Zfp42mut (mutation of KLF4 binding site within *Zfp42*
621 enhancer). DNA was pre-mixed with 50ul media with no additions, and in a
622 separate tube 10ul of Lipofectamine 2000 (Invitrogen #11668019) was pre-mixed
623 with 50ul media with no addition. After 5 minutes the two tubes were combined

624 and incubated at room temperature for 20 more minutes. Cells were then added
625 to the solution and plated on a gelatinized 12 well plate. 48hrs post-transfection,
626 GFP positive single cells were sorted by FACS into 96 well plates. Genotyping of
627 the single cell colonies was performed using a three-primer strategy (for
628 deletions) or by surveyor with T7 (for in-del mutation). Four (*Tbx3* hub) or five
629 (*Zfp42* hub) colonies with homozygous mutations (or w.t. colonies as control)
630 were expanded and used for RT-qPCR and 3C experiments. All gRNA, 3C and
631 RT-qPCR primers are described in Supplementary Table 9.

632 **CRISPRi of Zic2/5 enhancer**

633 V6.5 cells were infected with lentiviruses harboring the pHR-SFFV-dCas9-BFP-
634 KRAB vector (Addgene,46911) in which the SFFV promoter was replaced with
635 an Ef1a promoter. BFP expressing cells were selected by three rounds of FACS
636 sorting. The resulting V6.5, stably expressing the KRAB-dCas9, were then
637 infected with a lentivirus harboring the pLKO5.GRNA.EFS.PAC vector (Addgene,
638 57825) containing two gRNAs targeting the Zic2/5 enhancer. Cells were selected
639 with Puromycin (LifeTech K210015) for two days and subsequently collected for
640 RT-qPCR. gRNA and RT-qPCR primers are described in Supplementary Table 9.

641 **Generation of TKO cell line**

642 V6.5 cells were infected using lentiviruses harboring the c3GIC9 vector[77]
643 (TRE3G-Cas9-P2A-GFP-PGK-Puro-IRES-rtta) containing gRNA/s targeting
644 either *KLF4* only or *KLF2, KLF4* and *KLF5* in tandem. Following infection cells
645 were selected using Puromycin (LifeTech K210015) and clonal populations were

646 manually picked. Expression of CRISPR-Cas9 from these stable cell lines was
647 induced by addition of Doxycycline for 72hrs (1:1000 dilution of 2mg/ml stock)
648 and KO efficiency in each clonal population was verified by WB: KLF4 (R&D,
649 AF3158) KLF5 (R&D AF3758) KLF2 (Novus biologicals, NBP6181) ESRRB
650 (PPMX, PPH6705) NANOG (Bethyl laboratories, A300-397A) ACTIN (abcam,
651 ab49900). Successful KO clones were then used for subsequent experiments
652 (qPCR, ChIP-seq, 3C and HiChIP) after induction with doxycycline for the
653 indicated time points.

654 **3C-qPCR**

655 For each sample 1 to 2 million cells were lysed in 300ul of lysis buffer (10mM
656 Tris-HCl pH 8.0, 10mM NaCl, 0.2% Igepal CA630 with protease inhibitors) and
657 incubated on ice for 20 mins. Cells were centrifuged 2500g for 5min at 4°C and
658 pellet washed once in lysis buffer. Pellets were resuspended in 50ul of 0.5% SDS
659 and incubated at 65°C for 10 mins. 145ul of water and 25ul of 10% triton were
660 added to the samples and incubated 15mins at 37°C. 100 Units of MboI
661 restriction enzyme and 25ul of NEB buffer 2 were added and incubated over
662 night at 37°C with rotation. Next day the enzyme was inactivated at 65°C for 20
663 mins. The ligation reaction was carried out over night at 16°C by adding 120ul of
664 NEB T4 ligase buffer, 100ul of 10% Triton, 6ul of 20mg/l BSA, 100ul of 10mM
665 ATP and 5ul of T4 ligase (NEB #M0202). The following day, 50ul of 20mg/ml
666 proteinase K and 120ul of 10% SDS were added and the samples were
667 incubated over night at 65°C. Lastly, 10ul of 10mg/ml RNAse was added and
668 samples incubated 1 hour at 37°C. Following phenol chloroform purification, the

669 DNA was precipitated using 1.6 volumes of 100% ethanol and 0.1 volume of 3M
670 sodium acetate. After incubation at -80°C for 1 hour samples were spun for
671 15mins at 4°C at 16000rpm. Pellets were washed twice with 70% ethanol and
672 dissolved in 100ul of 10mM Tris pH8. Qbit was used to measure sample
673 concentrations and 100ng of material was used to amplify the desired regions by
674 qPCR. All primer sequences can be found in Supplementary table 9.

675 **ChIP-seq**

676 ChIP-seq was performed as previously described[78]. Specifically, cells were
677 crosslinked in 1% formaldehyde at RT for 10 minutes and quenched with 125mM
678 glycine for 5 mins at RT. 50 million cells were used for KLF4 ChIPs and 10
679 million for H3K27acetylation ChIP. Cell pellets were washed twice in PBS and
680 resuspended in 400ul lysis buffer (10mM Tris pH8, 1mM EDTA, 0.5% SDS) per
681 20 million cells. Cells were sonicated in a bioruptor device (30 cycles 30sec
682 on/off, high setting) and spun down 10 minutes at 4°C at maximum speed.
683 Supernatants were diluted 5 times with dilution buffer (0.01%SDS, 1.1%
684 triton,1.2mM EDTA,16.7mM Tris pH8, 167mM NaCl) and incubated with the
685 respective antibody (2-3ug/10M cells) (KLF4 R&D #3158, H3K27ac ab4729) O/N
686 with rotation at 4°C. Next day, protein G Dynabeads (ThermoScientific)
687 preblocked with BSA protein (100ng per 10ul Dynabeads) were added (10ul
688 blocked Dynabeads per 10 million cells) and incubated for 2-3 hours at 4°C.
689 Beads were immobilized on a magnet and washed twice in low salt buffer (0.1%
690 SDS,1% triton, 2mM EDTA, 150mM NaCl, 20mM Tris pH8), twice in high salt
691 buffer (0.1% SDS,1% triton, 2mM EDTA, 500mM NaCl, 20mM Tris pH8), twice in

692 LiCl buffer (0.25M LiCl, 1% NP40, 1% deoxycholic acid (sodium salt), 1mM
693 EDTA, 10mM Tris pH8) and once in TE. DNA was then eluted from the beads by
694 incubating with 150ul elution buffer (1% SDS, 100mM NaHCO3) for 20 minutes
695 at 65°C (vortexing every 10min). Supernatants were collected and reverse-
696 crosslinked by incubation at 65°C O/N in presence of proteinase K. After RNase
697 A treatment for 1hr at 37°C, DNA was purified using the minElute kit (Qiagen). 6-
698 10ng of immunoprecipitated material was used for ChIP-seq library preparation
699 using the KAPA Hyper prep kit (KAPA Biosystems). Libraries were sequenced on
700 an Illumina HiSeq 2500 platform on SE50 mode.

701 **ATAC-seq**

702 ATAC-seq was performed as previously described[79]. In brief, a total of 50,000
703 cells were washed once with 50 µL of cold PBS and resuspended in 50 µL lysis
704 buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.2% (v/v) IGEPAL
705 CA-630). The suspension of nuclei was then centrifuged for 10 min at 800 g at
706 4°C, followed by the addition of 50 µL transposition reaction mix (25 µL TD buffer,
707 2.5 µL Tn5 transposase and 22.5 µL nuclease-free H_2O) using reagents from the
708 Nextera DNA library Preparation Kit (Illumina #FC-121-103). Samples were then
709 incubated at 37°C for 30min. DNA was isolated using a ZYMO Kit (#D4014).
710 ATAC-seq libraries were first subjected to 5 cycles of pre-amplification. To
711 determine the suitable number of cycles required for the second round of PCR
712 the library was assessed by quantitative PCR as described in Buenrostro et al [79]
713 and the library was then PCR amplified for the appropriate number of cycles
714 using Nextera primers. Samples were subject to a dual size selection (0.55x-

715 1.5x) using SPRI beads (Beckman Coulter #B23317). Finally, the ATAC libraries
716 were sequenced on a HiSeq 2500 platform on PE50 mode.

717 **RNA-seq**

718 Total RNA was prepared with TRIZOL (Life technologies #15596018) following
719 manufacturer's instructions. Libraries were generated by the Weill Cornell
720 Genomics core facility using the Illumina TruSeq stranded mRNA library
721 preparation kit (#20020594) and sequenced on an Illumina HiSeq4000 platform
722 on SE50 mode.

723 **HiChIP**

724 HiChIPs were performed as previously described[36] with some modifications. In
725 brief, up to 15 million crosslinked cells (for KLF4 HiChIPs two samples of 15
726 million cells were combined at the end, for each sample replicate) were
727 resuspended in 500 µL of ice-cold HiC lysis buffer (10 mM Tris-HCl pH 7.5, 10
728 mM NaCl, 0.2% NP-40, 1× protease inhibitors) and rotated at 4°C for 30 min.
729 Nuclei were pelleted and washed once with 500 µL of ice-cold HiC lysis buffer.
730 Pellet was then resuspended in 100 µL of 0.5% SDS and incubated at 62°C for
731 10 min. 285 µL of water and 50 µL of 10% Triton X-100 were added, and
732 samples were rotated at 37°C for 15 min. 50 µL of NEB Buffer 2 and 15 µL of
733 25 U/µL MboI restriction enzyme (NEB, R0147) were then added, and sample
734 was rotated at 37°C for 2 h. MboI was then heat inactivated at 62°C for 20 min.
735 We added 52 µL of incorporation master mix: 37.5 µL of 0.4 mM biotin–dATP
736 (Thermo Fisher, 19524016); 4.5 µL of a dCTP, dGTP, and dTTP mix at 10 mM

737 each; and 10 μL of 5 U/μL DNA Polymerase I, Large (Klenow) Fragment (NEB,
738 M0210). The reactions were then rotated at 37°C for 1 h. 948 μL of ligation
739 master mix was then added: 150 μL of 10× NEB T4 DNA ligase buffer with 10
740 mM ATP (NEB, B0202), 125 μL of 10% Triton X-100, 3 μL of 50 mg/mL BSA
741 (Thermo Fisher, AM2616), 10 μL of 400 U/μL T4 DNA Ligase (NEB, M0202),
742 and 660 μL of water. The reactions were then rotated at room temperature for
743 4 h. After proximity ligation, the nuclei were pelleted and the supernatant was
744 removed. The nuclear pellet was brought up to 880 μL in Nuclear Lysis Buffer
745 (50 mM Tris-HCl pH 7.5, 10 mM EDTA, 0.5% SDS, 1× Roche protease
746 inhibitors, 11697498001), and sonicated with a Bioruptor 300 (Diagenode) for 8
747 cycles of 30sec each, on a medium setting. Clarified samples were transferred
748 to Eppendorf tubes and diluted five times with ChIP Dilution Buffer (0.01% SDS,
749 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH 7.5, 167 mM NaCl).
750 Cells were precleared with 30 μL of Protein G dynabeads (Life technology
751 #10004D) in rotation at 4°C for 1 h. Supernatants were transferred into fresh
752 tubes and antibody was added (8 μg of KLF4 antibody or 3ug H3K27Ac
753 antibody for 15 million cells) and incubated overnight at 4°C. The next day 30
754 μL of Protein G dynabeads were added to samples and rotated at 4°C for 2 h.
755 After bead capture, beads were washed three times each with low-salt wash
756 buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 7.5, 150
757 mM NaCl), high-salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20
758 mM Tris-HCl pH 7.5, 500 mM NaCl), and LiCl wash buffer (10 mM Tris-HCl pH
759 7.5, 250 mM LiCl, 1% NP-40, 1% sodium deoxycholate, 1 mM EDTA, make
760 fresh). Samples were eluted with 150 μL of DNA elution buffer (50 mM sodium

761 bicarbonate pH 8.0, 1% SDS, freshly made) and incubated at 37°C for 30 min
762 with rotation. Supernatant was transferred to a fresh tube and elution repeated
763 with another 150 µL elution buffer. 5 µL of Proteinase K (20mg/ml) (Thermo
764 Fisher) were added to the 300 µL reaction and samples were incubated
765 overnight at 65°C. Samples were purified with DNA Clean and Concentrator
766 columns (Zymo Research) and eluted in 10 µL of water. Post-ChIP DNA was
767 quantified by Qubit (Thermo Fisher) to estimate the amount of Tn5 (Illumina)
768 needed to generate libraries at the correct size distribution (see below). 5 µL of
769 Streptavidin C-1 beads (Thermo Fisher) were washed with Tween Wash Buffer
770 (5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl, 0.05% Tween-20) then
771 resuspended in 10 µL of 2× biotin binding buffer (10 mM Tris-HCl pH 7.5, 1
772 mM EDTA, 2 M NaCl). Beads were added to the samples and incubated at
773 room temperature for 15 min with shaking. After capture, beads were washed
774 twice by adding 500 µL of Tween Wash Buffer and incubated at 55°C for 2 min
775 with shaking. Samples were then washed in 100 µL of 1× TD Buffer (2× TD
776 Buffer is 20 mM Tris-HCl pH 7.5, 10 mM magnesium chloride, 20%
777 dimethylformamide). After washes, beads were resuspended in 25 µL of 2× TD
778 Buffer, Tn5 (for 50 ng of post-ChIP DNA we used 2.5 µL of Tn5), and water to
779 50 µL. Tn5 amount was adjusted linearly for different amounts of post-ChIP
780 DNA, with a maximum amount of 4 µL of Tn5. Samples were incubated at
781 55°C with interval shaking for 10 min. After removing the supernatant 50 mM
782 EDTA was added to samples and incubated with interval shaking at 50°C for
783 30 min. Beads were then washed two times each in 50 mM EDTA then Tween
784 Wash Buffer at 55°C for 2 min. Lastly, beads were washed in 10 mM Tris

785 before PCR amplification. Beads were resuspended in 25 µL of Phusion HF 2×
786 (New England Biosciences), 1 µL of each Nextera Ad1_noMX and Nextera
787 Ad2.X at 12.5 µM, and 23 µL of water. The following PCR program was
788 performed: 72°C for 5 min, 98°C for 1 min, then cycle at 98°C for 15 s, 63°C
789 for 30 s, and 72°C for 1 min (cycle number was estimated based on the
790 amount of material from the post-ChIP Qubit (approximately 50 ng was run in
791 six cycles, while 25 ng was run in seven, 12.5 ng was run in eight, etc.). Size
792 selection was performed using two-sided size selection with the Ampure XP
793 beads. After PCR, libraries were placed on a magnet and eluted into new tubes.
794 25 µL of Ampure XP beads were added, and the supernatant was kept to
795 capture fragments less than 700 bp. Supernatant was transferred to a new
796 tube, and 15 µL of fresh beads was added to capture fragments greater than
797 300 bp. After size selection, libraries were quantified with Qbit and sent for
798 Bioanalyzer to check for the quality and final size of the library. Libraries were
799 sequenced on a HiSeq 2500 platform on PE75 mode.

800 **4C-seq**

801 For each sample 10 million cells were fixed following our ChIP-seq protocol (see
802 above). Cell pellets were lysed in 1ml Lysis Buffer (50 mM Tris-HCl pH 7.5, 150
803 mM NaCl, 5 mM EDTA; 1x complete protease inhibitor, 0.5% NP-40, 1% triton)
804 and incubated on ice for 15 min. The samples were centrifuged at 2500xG for
805 5min at 4°C and the pellet was then resuspended in 360µl milli-Q, 60µl 10X DpnII
806 restriction buffer and 15ul 10%SDS. After 1 hour incubation at 37ºC, 150ul of
807 10% Triton was added and samples were incubated again at 37ºC for 1 hour. 4ul

808 DpnII enzyme (#R0543M, NEB) were added and samples were incubated at
809 37ºC over night while shaking in a thermomixer (9000rpm). After confirming the
810 digestion efficiency, the enzyme was inactivated by adding 80ul 10% SDS and
811 incubating at 65 ºC for 30 mins. The digested samples were then diluted with
812 4860ul Milli-Q, 700ul ligation buffer (500mM Tris pH 7.5, 100mM DTT, 100mM
813 MgCl2,10mM ATP), and 750ul of Triton and incubated at 37ºC for 1 hour. Then
814 2ul Ligase (NEB M0202M) were added and samples were incubated over night
815 at 16 ºC. Next morning, after testing the ligation efficiency, we reversed the
816 crosslinks by adding 30ul of proteinase K (10mg/ml) and incubating over night at
817 65ºC. Subsequently the RNA was removed using 30ul of RNase A (10mg/ml) for
818 45mins at 37ºC. Extensive phenol/chloroform extraction was followed by EtOH
819 precipitation and two washes with 70% EtOH. The pellets were dissolved in
820 150ul 10mM Tris pH 7.5 by incubating at 37 ºC. We then added 50ul 10x buffer B
821 (Fermentas), 5ul Csp6I (Fermentas, ER0211) and 299ul milli-Q water and
822 samples were digested at 37ºC over night. After determining digestion efficiency,
823 the restriction enzyme was inactivated by incubating the tubes at 65ºC for 25
824 mins. Samples were diluted in 12ml milli-Q, 3ul ligase (NEB, M0202M) and 1.4ml
825 10X ligation buffer (500mM Tris pH7.5, 100mM DTT, 100mM MgCl2, 10mMATP)
826 and incubated over night at 65ºC. Following phenol/chloroform and EtOH
827 precipitation the pellets were dissolved in 300ul 10mM Tris pH7.5 and DNA was
828 further purified using 4 Zymo columns per sample (Zymo, D4014). Each sample
829 was eluted in 200ul total of 10mM Tris pH7.5. Finally, 150ng of DNA was used
830 per reaction, to PCR-amplify the libraries using the KAPA HiFi enzyme (KAPA
831 biosystem, 07958927001). All primer sequences can be found in Supplementary

832 table 9. Four PCR reactions were combined per sample, following column
833 purification using the ZYMO kit (Zymo, D4014). Samples were sent for QC on a
834 bioanalyzer and then sequenced on a HiSeq 4000 platform on SE50 mode.

835 **RIME**

836 RIME was performed in 3 replicates for KLF4 and 2 for IgG, as previously
837 described[52] with minor modifications. 50 million V6.5 cells grown in 2i conditions
838 were used for each replicate. Cells were fixed, lysed, sonicated and incubated
839 with the respective antibody-bound beads, using the same conditions that were
840 used for KLF4 ChIP-seq (see above). The samples were then washed ten times
841 in RIPA buffer (50 mM HEPES (pH 7.6), 1 mM EDTA, 0.7% (wt/vol) sodium
842 deoxycholate, 1% (vol/vol) NP-40 and 0.5M LiC) and five times in 100mM AMBIC
843 solution. Treatment for enzymatic digestion and peptide desalting was carried out
844 as in the original protocol.

845 **Co-IP and WB**

846 50 million V6.5 cells grown in 2i condition were collected for each Co-IP
847 experiment and resuspended in 0.5ml lysis buffer (50mM Tris pH7.5, 100mM
848 Nacl, 0.2% triton, 0.5% glycerol and protease inhibitors). Cells were incubated on
849 ice for 40 mins followed by 3 cycles of sonication in a bioruptor device (30sec
850 on/off, high setting) and spun down 10 minutes at 4°C at maximum speed.
851 Supernatants were diluted with additional lysis buffer in a final volume of 2ml.
852 Lysates were pre-cleared with 10ul of protein G Dynabeads (ThermoScientific)
853 for 30 mins in rotation at 4°C. The supernatant was then incubated with 8ug of

854 KLF4 antibody (R&D, AF3158) or IgG (Calbiochem, NI02) for 2.5 hours in
855 rotation at 4°C. 30ul of protein G Dynabeads that were pre-blocked with BSA
856 were added to the samples and incubated 1.5hours in rotation at 4°C. Two
857 washes were performed with lysis buffer followed by three washes with high salt
858 buffer (same as lysis buffer but with 250mM NaCl). Finally, the samples were
859 eluted in loading buffer by boiling 5 minutes and transferring the sup to a new
860 tube. WBs were performed with the following antibodies: BRD4 (Bethyl, A301-
861 985A50), MED1 (Bethyl, A300-793A), SMC1a (Bethyl, A300-055A), RING1b
862 (Bethyl, A302-869A), SUZ12 (Santa Cruz, sc46264), LSD1(Abcam, ab 17721).

863 **ATAC-seq data analysis**

864 *Mapping, peak calling and peak processing*. Paired-end reads were aligned to
865 mm10 (bowtie2 version 2.3.2, --no-unal --local --very-sensitive-local --no-
866 discordant --no-mixed --contain --overlap --dovetail -I 10 -X 2000), and
867 mitochondrial DNA alignments were excluded. Fragments marked as positional
868 duplicates (sambamba version 0.6.6) or overlapping with mouseENCODE
869 blacklisted genomic regions[80] (liftOver to mm10) were filtered out. Read ends
870 were adjusted for Tn5 transposase offsets. Peaks were called at $p<10^{-5}$ (MACS
871 version 2.1.1) per replicate, and only common peaks between two independent
872 biological replicates were retained for further analysis.

873 **ChIP-seq data analysis**

874 *Mapping, peak calling and peak processing*. Study and published ChIP-seq
875 reads were trimmed for adapters (cutadapt version 1.8.1), and low-quality ends

876 (sickle version 1.33), respectively. Alignment to the mouse reference genome
877 version mm10 (GRCm38.p4) was performed using standard parameters,
878 permitting a maximum of one mismatch in seed alignment (bowtie2 version 2.3.2).
879 Reads marked as positional duplicates (sambamba version 0.6.6) or overlapping
880 with mouseENCODE blacklisted genomic regions (liftOver to mm10) were filtered
881 out. Study ChIP-seq peaks (enrichment of signals over background determined
882 by input samples) were called at $p<0.01$ (MACS version 2.1.1) per biological
883 replicate, and peaks detected in more than half of biological replicates were
884 retained for further analysis. Published ChIP-seq replicates were merged, and
885 peaks were called at $p<10^{-5}$ using input samples where applicable.

886 **Overlap analysis of ChIP-seq peaks for chromatin states of reprogramming**
887 **cell types.** Chromatin states (1 kb resolution) during reprogramming were
888 retrieved from ref[8], and cis-regulatory elements were annotated from chromatin
889 states as in the original publication. The assignment of ChIP-seq peaks to cis-
890 regulatory elements was determined by the largest degree of overlap in bp.

891 **RNA-seq and ChIP-seq gene ontology (GO) enrichment analysis.** Spatial
892 proximity of ChIP-seq peaks to transcript start sites (TSSs) and enriched GOs
893 were uncovered utilizing the GREAT (version 3.0.0) web application. We
894 selected the "basal plus extension rule" for the association of gene ontology
895 annotations with regulatory domains (customized setting: 5 kb upstream and 1 kb
896 downstream of TSSs, and further extended both directions by 250 kb).
897 Enrichment of ontology annotations was assessed by the binomial test of ChIP-
898 seq peak-overlaps with annotated regulatory regions. For differentially expressed

899 genes and gene groups (Fig. S4e) DAVID knowledgebase[81] was used for
900 pathway and biological process enrichment analysis.

901 **PSC typical enhancers and super-enhancers.** Coordinates of typical- and
902 super-enhancers in mESCs and other cell lines or tissues were ascertained from
903 ref[45] and ref[34], lifted over from mm9 to mm10 with UCSC liftOver.

904 **Overlaps of KLF4 binding with early lost or late gained H3K27ac peaks**
905 **during reprogramming or at typical- and super-enhancers.** The maximum
906 permitted distance between KLF4 binding detected in day3, 6 and 9 with PSC
907 and H3K27ac peaks or enhancers in ref[45] was 250 bp. Where H3K27ac peaks or
908 enhancers overlapped with KLF4 sites of different stages, the earliest stage was
909 prioritized (Fig. 1g).

910 **Motif analysis**

911 For each KLF4 cluster we generated a random background (by shuffling the
912 peaks randomly throughout the genome) to test motif enrichment within each
913 cluster. Analysis of the KLF4 clusters was performed with the use of HOMER and
914 'findMotifsGenome.pl' command with the following parameters: '-bg random.bed -
915 size 200 -len 15'. Only motifs with p-value≤1e-5 were considered significant. Two
916 heatmaps with the z-transformed '-log10(p-value)' and z-transformed 'motif
917 frequency' of selected motifs for each cluster are presented in Supplementary
918 Fig.1e.

919 **PCA analysis for ATAC-seq, RNA-seq and ChIP-seq experiments**

920 We first merged all accessible regions / H3K27ac peak detected from ATAC-seq
921 / H3K27ac ChIP-seq in any reprogramming stage using bedtools v2.25.0. Then,
922 we calculated the coverage of reads for each merged accessible region and
923 H3K27ac peak for each replicate independently. For the RNA-seq data, we
924 calculated the coverage for each exon and only exons with at least 1 read
925 covering every single base of the exon were used for downstream analysis. PCA
926 analysis was performed with R and PCA plots were generated with 'ggplot2'
927 library. In each PCA plot, we present the variability captured by the first two PCs
928 (PC1 and PC2).

929 **RNA-seq data analysis**

930 Expression of genes was quantified in transcripts per kilobase million (TPM)
931 using quasi mapping (Salmon, version 0.8.2) to GENCODE (version M6, mm10)
932 reference gene annotation. Salmon provides alignment-free transcript
933 quantification information in a single step[82].

934 **Line plots for gene expression analysis**

935 We plotted the median expression levels of all protein coding genes with their
936 corresponding 95% confidence interval (CI) that are bound by KLF4 in a distance
937 less than 50 kb from their corresponding transcription start site (TSS). For each
938 KLF4 cluster we calculated the closest (≤50 kb) TSS to each KLF4 binding site
939 and plotted the median expression levels (TPM) of all genes annotated in each
940 KLF4 cluster with the use of R.

941 **Processing of HiChIP / HiC datasets**

942 HiChIP and HiC datasets were uniformly pre-processed with the HiC-bench
943 platform[83], which is outlined in short in the following. First, all paired-end
944 sequencing reads were aligned against the mouse genome version mm10 with
945 bowtie2 version 2.2.3[84] (specific settings: --very-sensitive-local --local). Read-
946 filtering was conducted by the GenomicTools[85] gtools-hic filter command
947 (integrated in HiC-bench), which discards multi-mapped reads ("multihit"), read-
948 pairs with only one mappable read ("single sided"), duplicated read-pairs
949 ("ds.duplicate"), read-pairs with a low mapping quality of MAPQ < 20, read-pairs
950 resulting from self-ligated fragments and short-range interactions resulting from
951 read-pairs aligning within 10kb (together called "ds.filtered"). The percentage of
952 accepted intra-chromosomal read-pairs ("ds.accecpted intra") was high across all
953 HiC and HiChIP replicates and conditions and was consistently above 35%. In
954 order to create counts-matrices per chromosome in a binned fashion, we set the
955 bin size to 10kb for all datasets. For all the HiChIP sample and chromosome
956 matrices, the trajectories of each matrix bin to both anchors were overlaid with
957 the ChIP-Seq signal of the respective matching sample, requiring a minimal
958 overlap of 1bp between a HiChIP-bin and a ChIP-peak. Only loops of which at
959 least one anchor was supported by a ChIP-peak were kept for further analyses.
960 Next, we applied sequencing-depth normalization (leading to read-counts per
961 million, or CPM) per replicate followed by a statistical approach to identify
962 significant loops. We have adapted the approach first described in Mango[37], by
963 performing a binomial test in each diagonal of the counts-matrix up to a
964 maximum distance of 2MB.

965 High-confidence HiChIP loops were identified by p-value < 0.1 and requiring a
966 CPM > 3 per loop across all replicates of a single condition in order to maintain a
967 signal that is replicable. For high-confidence HiC loops, we have adjusted those
968 thresholds in order to avoid too much noise, and have applied filters of p-value <
969 0.01 and CPM > 15 across all replicates of a single condition.

970 **Principal component analysis for HiChIP samples**

971 Principal component analysis (PCA) as shown in Figures S5a was performed on
972 all available replicates on the high-confidence loops. Therefore, for each
973 detected high-confidence loop from any sample, the per replicate normalized
974 CPM was extracted before filtering for significant loops in order to also integrate
975 lowly detected interactions in the analysis. PCA was performed using the prcomp
976 function of R (version 3.3.0; scale=TRUE, center=TRUE).

977 **Differential loop analysis**

978 Differential looping analysis was performed on each significant loop
979 independently by applying an unpaired two-sided t-test on the normalized counts
980 (CPM) calculated before identifying significant loops between any pairwise
981 comparisons: PSC-KLF4 vs d3-KLF4, PSC-KLF4 vs d6-KLF4, d3-KLF4 vs d6-
982 KLF4, PSC-H3K27ac vs MEF-H3K27ac, TKO-0h vs TKO-24h. In order to
983 estimate the change in loop strength, we calculated the log2 fold-change (logFC)
984 between the average CPM per condition for the same pairwise comparisons after
985 adding a pseudo-count of 1 to each replicate and loop. For constant H3K27ac
986 loops in either MEF vs PSC or TKO-0h vs TKO-24h, we selected loops with p-

987 value > 0.5 and absolute logFC < 0.5 for the respective pairwise comparison.
988 MEF/PSC-specific H3K27ac loops were selected by p-value < 0.1 and logFC > 2
989 / logFC < -2 taken from the PSC H3K27ac vs MEF H3K27ac comparison,
990 respectively. TKO-0h/TKO-24h specific loops were selected by p-value < 0.05
991 and logFC > 0.58 / logFC < -0.58 taken from the TKO-0h vs TKO-24h
992 comparison. Mid and late established KLF4 loops were selected by applying p-
993 value < 0.01 and logFC > 2 in the pairwise comparisons of PSC-KLF4 vs d3-
994 KLF4 and d6-KLF4 vs d3-KLF4 (mid) and PSC-KLF4 vs d3-KLF4 and PSC-KLF4
995 vs d6-KLF4 (late). Early-lost and mid-lost KLF4 loops were selected by applying
996 a p-value < 0.01 and logFC < -2 in the pairwise comparisons of PSC-KLF4 vs d6-
997 KLF4 and PSC-KLF4 vs d3-KLF4 (early-lost) and PSC-KLF4 vs d3-KLF4 and d6-
998 KLF4 vs d3-KLF4 (mid-lost). For differential comparison of significant HiC loops,
999 we have applied a distance-normalization as previously described[86] before
1000 calculating significance and fold-changes between PSC and MEF HiC loops.
1001 Then, differential HiC loops were selected by applying a p-value < 0.1 and logFC
1002 < -0.32 or logFC > 0.32 (equivalent to a fold-change of 1.25) in the pairwise
1003 comparison of PSC-HiC vs MEF-HiC. All calculations were performed in R
1004 version 3.3.0, using the native t.test function (unpaired, two-sided).

1005 **Annotation of H3K27ac HiChIP loop anchors as promoters or enhancers**.

1006 H3K27ac HiChIP loop anchors were overlapped with transcription start sites
1007 (TSSs) of GENCODE (version M6) protein coding genes. Presence of one or
1008 more TSSs was considered a promoter HiChIP anchor, and the absence of any
1009 TSS but presence of at least one H3K27ac ChIP-seq constitutes an enhancer

1010 HiChIP anchor. In estimating connectivity, all HiChIP anchors, either promoter,
1011 enhancer or otherwise desolate, were considered.

1012 **RNA expression integration with differential HiChIP loops**

1013 For RNA expression integration, we overlapped all canonical TSSs of protein-
1014 coding genes (transcript support level/TSL = 1) downloaded from Ensembl
1015 Genes V85 for the mouse genome mm10 with all loop anchors. Because the
1016 TSS is a 1bp position in the genome, each gene was uniquely assigned to one
1017 bin, however, multiple TSSs per gene with a TSL=1 mapping to different bins are
1018 possible. Next, we filtered genes by occurrence of differential loop clusters that
1019 were obtained from the HiChIP experiments and have TPM > 1 in at least one
1020 reprogramming stage, and analyzed the expression patterns of such genes
1021 throughout reprogramming. For H3K27ac HiChIP data integration, we assigned
1022 genes to MEF/PSC-specific loops if their TSSs were found in >= 1 MEF/PSC-
1023 specific loops but in none of the other (Figure 2b). Genes contained in anchors of
1024 constant loops were filtered by having at least 1 or 3 constant loop anchors but
1025 no MEF/PSC-specific loop. To further validate expression changes based on
1026 differential looping, we applied an unpaired, one-sided t-test between genes
1027 logFCs of constant H3K27ac loops versus genes with MEF/PSC-specific loops,
1028 following the hypothesis of a positive correlation between looping and expression
1029 changes. As a negative control, we compared logFCs of genes with constant
1030 loops versus all annotated protein-coding genes. We have followed the same
1031 approach for the integration of expression data with differential loops obtained
1032 from Klf-TKO H3K27ac HiChIP experiments. In short, we assigned genes to

1033 TKO-0h/TKO-24h specific loops if their TSSs were found in >= 1 differential loops
1034 but not in the other differential loop category. We have compared logFCs of
1035 expression of TKO-0h and TKO-24h between TKO-0h/TKO-24h specific and
1036 constant loops.

1037 **Co-regulation of gene expression by H3K27ac HiChIP (enhancer hubs***).* In
1038 this analysis, promoter anchors of enhancer-mediated loops were filtered for
1039 protein-coding genes that have an expression TPM > 1 in PSC. Enhancer
1040 fragments that contact two to ten promoter fragments in PSC specific H3K27ac
1041 loops were selected. Genes were paired across different promoter fragments
1042 connecting to the same enhancer anchor (later on called hub), and repeated
1043 gene pairs were removed from the overall pool. Gene pairs were considered co-
1044 expressed, if both genes were up-regulated in PSC compared to MEF (p-
1045 adjusted<10^{-2} and fold change threshold of 2) or vice versa for down-regulation.
1046 Or otherwise, at least one gene in a pair unchanged between MEF and PSC
1047 constitute unchanged gene pairs. In order to test if the enrichment of the co-
1048 regulated gene pairs in the original hubs was significant we performed Fisher's
1049 exact test. The background probability was calculating by using an equal number
1050 of random gene pairs (protein-coding genes that have an expression TPM > 1 in
1051 PSC) either of similar linear distance with our test group (global random) or within
1052 the same TADs. TADs were called from normalized corrected HiC matrices in
1053 PSCs processed at 10kb resolution using a recently published software[87] with the
1054 use of the following parameters '--minDepth 120000 --maxDepth 420000 --
1055 thresholdComparison 0.001 --delta 0.01 --correctForMultipleTesting fdr' .

1056 **Overlap between H3K27ac loop clusters and KLF4 clusters**

1057 Overlap between any of the KLF4 peaks with any of the HiChIP anchors
1058 (H3K27aC or KLF4 loops) was performed with the use of bedtools v2.25.0. Odds
1059 ratio and significance of the overlap between the 2 groups was performed with
1060 the use of Fisher's exact test.

1061 **ChIP-seq feature enrichment at lowly or highly connected H3K27ac PSCs**
1062 **specific enhancer anchors**.

1063 H3K27ac HiChIP enhancer anchors were selected for low (N = 1,183) or high
1064 connectivity (contacting four or more anchors; N = 1,014). LOLA analysis was
1065 performed in these two groups of ChIP-seq peaks in order to identify which TFs
1066 participate in the formation of low vs high connected hubs (Fig.3g).

1067 **KLF4 looping involved in RNA expression changes**

1068 To estimate the effect of KLF4 associated looping on changes in RNA expression,
1069 we followed a similar approach as before for the H3K27ac HiChIP integration.
1070 After selecting expressed genes within anchors of each KLF4 loop cluster, we
1071 further filtered for differentially expressed genes between PSC and day3 (FDR <
1072 0.01; logFC > 1.0 (upregulated) or logFC < -1.0 (downregulated)). Information on
1073 differential expression was derived by DESeq with subsequent multiple testing
1074 correction as mentioned before. Genes determined as 'no change' were selected
1075 by applying FDR > 0.5 and absolute logFC < 0.25. All remaining genes were
1076 discarded from the analysis.

1077 **LOLA enrichment analysis**

1078 The identified differential loops were subjected for an enrichment analysis of
1079 further transcription factor bindings and other DNA binding proteins. First, the
1080 anchors of each differential loop were mapped back to the original ChIP or
1081 ATAC-peaks, because the 10kb stretches of the bins would give too many false
1082 positive findings. Each anchor that was overlapping an actual ChIP or ATAC-
1083 peak by at least 1bp was subjected for further analysis. Since two anchors can
1084 theoretically overlap with a single ChIP-peak using this approach, the resulting
1085 list was collapsed and only unique ChIP or ATAC-peaks were kept. Next, we
1086 applied LOLA version 1.8.0[88] against a database of analyzed ChIP-Seq datasets
1087 taken from LOLA Region Databases (regionDB) for mm10 (for Figure S3e we
1088 used Codex and encode TFBSmm10 databases). We excluded all ChIP-Seq
1089 datasets that were marked as treated with any agent and had less than 3000
1090 peaks in total. When multiple ChIP-seq data for the same antibody were
1091 significantly enriched in one of our tested regions we selected the one with the
1092 highest number of peaks. In addition to the ChIP-seq peaks provided by the
1093 LOLA database we manually constructed a database containing ChIP-seq from
1094 the following studies GSE22557, GSE90893, GSE99519, GSE22562 and our
1095 own ChIP-seq data. Data from these studies were re-analyzed with the same
1096 pipelines that were used for our ChIP-seq data. As a universe for LOLA, we used
1097 only unique ChIP or ATAC-peaks from the union of all ChIP or ATAC-Seq peaks
1098 for the respective antibody across all reprogramming stages.

1099 **Virtual 4C**

1100 Virtual 4C was performed to identify interaction signals of gene promoters or
1101 enhancers with their genomic vicinity. For this approach, we used the filtered
1102 HiChIP read-pairs as described above before binning and normalization of each
1103 replicate. We extracted all read pairs for which a read mate maps within +/- 10kb
1104 around the virtual viewpoint. Next, we defined successive overlapping windows
1105 for each chromosome at a 10kb resolution, and all adjacent windows are
1106 overlapping by 95% of their length (i.e. 9.5kb, or a shift of 500bp between
1107 adjacent windows). We then counted the second mapped read mate in all
1108 overlapping bins. Thus, each read-pair accounts for +1 in exactly 19 bins,
1109 however, the overlap of bins achieves a smoothed signal. Read counts for all
1110 bins were normalized to total sequencing depth of the respective replicate by
1111 edgeR version 3.14.0 to calculate counts-per-million (CPM) per bin. Significant
1112 differences between any condition (TKO-0h vs TKO-24h H3K27ac HiChIP or
1113 MEF vs ES H3K27ac HiChIP) was calculated using edgeR function glmQLFTest
1114 (we have not corrected for multiple testing, because the requirement of
1115 independent data-points for multiple testing correction is not given for the
1116 overlapping windows). For visualization, the average of the normalized virtual
1117 4C-signal across replicates of a single condition was calculated.

1118 **Analysis of 4C-seq data**

1119 The 4C-seq data was analyzed in a similar fashion as recently described[89, 90].
1120 Firstly, viewpoint primers were trimmed off from all sequencing reads using seqtk
1121 (version 1.3.0). Next, the remaining read-sequence was aligned using bowtie
1122 v1.0.0 against a reduced genome that consists only of reference genome

1123 sequences adjacent to DpnII cut-sites which was used during the 4C protocol
1124 (following the 4C-ker pipeline[89]). By aligning against the reduced genome, only
1125 reads matching the adjacent sequence of an actual digestion fragment are
1126 allowed, and the remaining reads are automatically discarded. Next, the genome
1127 was binned into 10kb bins shifted by 500bp (thus overlapping by 95% with
1128 adjacent bins), similar as the virtual 4C approach described above. Reads were
1129 counted by unique alignment position per bin, thus accounting for +1 read in 19
1130 adjacent bins to achieve a smoothed signal. Read counts per bin were
1131 normalized by sequencing depth per replicate using edgeR (version 3.14.0),
1132 resulting in counts per million (CPM). The visualization shows the average CPM
1133 signal across all replicates of a single condition.

1134 **RIME analysis**

1135 Summed 'signal to noise' intensity per protein from 3 KLF4 and 2 IgG samples
1136 was used to calculate significant enrichment of KLF4 protein complexes with the
1137 use of Welch's t-test. Only proteins with a p-value <0.05 and fold enrichment
1138 greater than 1.5 over IgG were considered significantly enriched in our samples.

1139 **REFERENCES**

1140 1. Young, R.A. Control of the embryonic stem cell state. *Cell* **144**, 940-954 (2011).
1141 2. Natoli, G. Maintaining cell identity through global control of genomic
1142 organization. *Immunity* **33**, 12-24 (2010).
1143 3. Graf, T. Historical origins of transdifferentiation and reprogramming. *Cell Stem*
1144 *Cell* **9**, 504-516 (2011).
1145 4. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse
1146 embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676
1147 (2006).
1148 5. Apostolou, E. & Hochedlinger, K. Chromatin dynamics during cellular
1149 reprogramming. *Nature* **502**, 462-471 (2013).

1150 6. Apostolou, E. & Stadtfeld, M. Cellular trajectories and molecular mechanisms of
1151 iPSC reprogramming. *Curr Opin Genet Dev* **52**, 77-85 (2018).
1152 7. Di Giammartino, D.C. & Apostolou, E. The Chromatin Signature of Pluripotency:
1153 Establishment and Maintenance. *Curr Stem Cell Rep* **2**, 255-262 (2016).
1154 8. Chronis, C. *et al.* Cooperative Binding of Transcription Factors Orchestrates
1155 Reprogramming. *Cell* **168**, 442-459 e420 (2017).
1156 9. Soufi, A., Donahue, G. & Zaret, K.S. Facilitators and impediments of the
1157 pluripotency reprogramming factors' initial engagement with the genome. *Cell*
1158 **151**, 994-1004 (2012).
1159 10. Chen, J. *et al.* Hierarchical Oct4 Binding in Concert with Primed Epigenetic
1160 Rearrangements during Somatic Cell Reprogramming. *Cell Rep* **14**, 1540-1554
1161 (2016).
1162 11. Sridharan, R. *et al.* Role of the murine reprogramming factors in the induction of
1163 pluripotency. *Cell* **136**, 364-377 (2009).
1164 12. Knaupp, A.S. *et al.* Transient and Permanent Reconfiguration of Chromatin and
1165 Transcription Factor Occupancy Drive Reprogramming. *Cell Stem Cell* **21**, 834-845
1166 e836 (2017).
1167 13. Li, D. *et al.* Chromatin Accessibility Dynamics during iPSC Reprogramming. *Cell*
1168 *Stem Cell* **21**, 819-833 e816 (2017).
1169 14. Polo, J.M. *et al.* A molecular roadmap of reprogramming somatic cells into iPS
1170 cells. *Cell* **151**, 1617-1632 (2012).
1171 15. Apostolou, E. *et al.* Genome-wide Chromatin Interactions of the Nanog Locus in
1172 Pluripotency, Differentiation, and Reprogramming. *Cell stem cell* **12**, 699-712
1173 (2013).
1174 16. de Wit, E. *et al.* The pluripotent genome in three dimensions is shaped around
1175 pluripotency factors. *Nature* **501**, 227-231 (2013).
1176 17. Denholtz, M. *et al.* Long-range chromatin contacts in embryonic stem cells reveal
1177 a role for pluripotency factors and polycomb proteins in genome organization.
1178 *Cell Stem Cell* **13**, 602-616 (2013).
1179 18. Wei, Z. *et al.* Klf4 Organizes Long-Range Chromosomal Interactions with the Oct4
1180 Locus in Reprogramming and Pluripotency. *Cell stem cell* (2013).
1181 19. Krijger, P.H. *et al.* Cell-of-Origin-Specific 3D Genome Structure Acquired during
1182 Somatic Cell Reprogramming. *Cell Stem Cell* **18**, 597-610 (2016).
1183 20. Beagan, J.A. *et al.* Local Genome Topology Can Exhibit an Incompletely Rewired
1184 3D-Folding State during Somatic Cell Reprogramming. *Cell Stem Cell* **18**, 611-624
1185 (2016).
1186 21. Stadhouders, R. *et al.* Transcription factors orchestrate dynamic interplay
1187 between genome topology and gene regulation during cell reprogramming. *Nat*
1188 *Genet* (2018).
1189 22. Rubin, A.J. *et al.* Lineage-specific dynamic and pre-established enhancer-
1190 promoter contacts cooperate in terminal differentiation. *Nat Genet* **49**, 1522-
1191 1528 (2017).
1192 23. Drissen, R. *et al.* The active spatial organization of the beta-globin locus requires
1193 the transcription factor EKLF. *Genes Dev* **18**, 2485-2490 (2004).

1194 24. Schoenfelder, S. *et al.* Preferential associations between co-regulated genes
1195 reveal a transcriptional interactome in erythroid cells. *Nat Genet* **42**, 53-61
1196 (2010).
1197 25. Stadtfeld, M., Maherali, N., Borkent, M. & Hochedlinger, K. A reprogrammable
1198 mouse strain from gene-targeted embryonic stem cells. *Nat Methods* **7**, 53-55
1199 (2010).
1200 26. Stadtfeld, M. *et al.* Ascorbic acid prevents loss of Dlk1-Dio3 imprinting and
1201 facilitates generation of all-iPS cell mice from terminally differentiated B cells.
1202 *Nat Genet* **44**, 398-405, S391-392 (2012).
1203 27. Stadtfeld, M. *et al.* Aberrant silencing of imprinted genes on chromosome 12qF1
1204 in mouse induced pluripotent stem cells. *Nature* **465**, 175-181 (2010).
1205 28. Zunder, E.R., Lujan, E., Goltsev, Y., Wernig, M. & Nolan, G.P. A continuous
1206 molecular roadmap to iPSC reprogramming through progression analysis of
1207 single-cell mass cytometry. *Cell Stem Cell* **16**, 323-337 (2015).
1208 29. Stadtfeld, M., Maherali, N., Breault, D.T. & Hochedlinger, K. Defining molecular
1209 cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell Stem Cell*
1210 **2**, 230-240 (2008).
1211 30. McLean, C.Y. *et al.* GREAT improves functional interpretation of cis-regulatory
1212 regions. *Nat Biotechnol* **28**, 495-501 (2010).
1213 31. Guo, L. *et al.* Resolving Cell Fate Decisions during Somatic Cell Reprogramming by
1214 Single-Cell RNA-Seq. *Mol Cell* **73**, 815-829 e817 (2019).
1215 32. Allison, T.F. *et al.* Identification and Single-Cell Functional Characterization of an
1216 Endodermally Biased Pluripotent Substate in Human Embryonic Stem Cells. *Stem*
1217 *Cell Reports* **10**, 1895-1907 (2018).
1218 33. Schiebinger, G. *et al.* Optimal-Transport Analysis of Single-Cell Gene Expression
1219 Identifies Developmental Trajectories in Reprogramming. *Cell* **176**, 928-943 e922
1220 (2019).
1221 34. Whyte, W.A. *et al.* Master transcription factors and mediator establish super-
1222 enhancers at key cell identity genes. *Cell* **153**, 307-319 (2013).
1223 35. Dixon, J.R. *et al.* Topological domains in mammalian genomes identified by
1224 analysis of chromatin interactions. *Nature* **485**, 376-380 (2012).
1225 36. Mumbach, M.R. *et al.* HiChIP: efficient and sensitive analysis of protein-directed
1226 genome architecture. *Nat Methods* **13**, 919-922 (2016).
1227 37. Phanstiel, D.H., Boyle, A.P., Heidari, N. & Snyder, M.P. Mango: a bias-correcting
1228 ChIA-PET analysis pipeline. *Bioinformatics* **31**, 3092-3098 (2015).
1229 38. Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Neural
1230 Development. *Cell* **171**, 557-572 e524 (2017).
1231 39. Mumbach, M.R. *et al.* Enhancer connectome in primary human cells identifies
1232 target genes of disease-associated DNA elements. *Nat Genet* **49**, 1602-1612
1233 (2017).
1234 40. Orkin, S.H. & Hochedlinger, K. Chromatin connections to pluripotency and
1235 cellular reprogramming. *Cell* **145**, 835-850 (2011).
1236 41. Weintraub, A.S. *et al.* YY1 Is a Structural Regulator of Enhancer-Promoter Loops.
1237 *Cell* **171**, 1573-1588 e1528 (2017).

1238 42. Beagan, J.A. *et al.* YY1 and CTCF orchestrate a 3D chromatin looping switch
1239 during early neural lineage commitment. *Genome Res* **27**, 1139-1152 (2017).
1240 43. Gomez-Diaz, E. & Corces, V.G. Architectural proteins: regulators of 3D genome
1241 organization in cell fate. *Trends Cell Biol* **24**, 703-711 (2014).
1242 44. Ong, C.T. & Corces, V.G. CTCF: an architectural protein bridging genome topology
1243 and function. *Nat Rev Genet* **15**, 234-246 (2014).
1244 45. Dowen, J.M. *et al.* Control of cell identity genes occurs in insulated
1245 neighborhoods in mammalian chromosomes. *Cell* **159**, 374-387 (2014).
1246 46. Novo, C.L. *et al.* Long-Range Enhancer Interactions Are Prevalent in Mouse
1247 Embryonic Stem Cells and Are Reorganized upon Pluripotent State Transition.
1248 *Cell Rep* **22**, 2615-2627 (2018).
1249 47. Ran, F.A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**,
1250 2281-2308 (2013).
1251 48. Kearns, N.A. *et al.* Functional annotation of native enhancers with a Cas9-histone
1252 demethylase fusion. *Nat Methods* **12**, 401-403 (2015).
1253 49. Larson, M.H. *et al.* CRISPR interference (CRISPRi) for sequence-specific control of
1254 gene expression. *Nat Protoc* **8**, 2180-2196 (2013).
1255 50. Schoenfelder, S. *et al.* Polycomb repressive complex PRC1 spatially constrains the
1256 mouse embryonic stem cell genome. *Nat Genet* **47**, 1179-1186 (2015).
1257 51. Joshi, O. *et al.* Dynamic Reorganization of Extremely Long-Range Promoter-
1258 Promoter Interactions between Two States of Pluripotency. *Cell Stem Cell* **17**,
1259 748-757 (2015).
1260 52. Mohammed, H. *et al.* Rapid immunoprecipitation mass spectrometry of
1261 endogenous proteins (RIME) for analysis of chromatin complexes. *Nat Protoc* **11**,
1262 316-326 (2016).
1263 53. Jiang, J. *et al.* A core Klf circuitry regulates self-renewal of embryonic stem cells.
1264 *Nat Cell Biol* **10**, 353-360 (2008).
1265 54. Zhang, S. *et al.* Epigenetic regulation of REX1 expression and chromatin binding
1266 specificity by HMGNs. *Nucleic Acids Res* (2019).
1267 55. Soufi, A. *et al.* Pioneer transcription factors target partial DNA motifs on
1268 nucleosomes to initiate reprogramming. *Cell* **161**, 555-568 (2015).
1269 56. Sardina, J.L. *et al.* Transcription Factors Drive Tet2-Mediated Enhancer
1270 Demethylation to Reprogram Cell Fate. *Cell Stem Cell* **23**, 727-741 e729 (2018).
1271 57. Schmitt, A.D. *et al.* A Compendium of Chromatin Contact Maps Reveals Spatially
1272 Active Regions in the Human Genome. *Cell Rep* **17**, 2042-2059 (2016).
1273 58. Huang, J. *et al.* Dissecting super-enhancer hierarchy based on chromatin
1274 interactions. *Nat Commun* **9**, 943 (2018).
1275 59. Fukaya, T., Lim, B. & Levine, M. Enhancer Control of Transcriptional Bursting. *Cell*
1276 **166**, 358-368 (2016).
1277 60. Oudelaar, A.M. *et al.* Single-allele chromatin interactions identify regulatory hubs
1278 in dynamic compartmentalized domains. *Nat Genet* **50**, 1744-1751 (2018).
1279 61. Zheng, M. *et al.* Multiplex chromatin interactions with single-molecule precision.
1280 *Nature* **566**, 558-562 (2019).

1281 62. Allahyar, A. *et al.* Enhancer hubs and loop collisions identified from single-allele
1282 topologies. *Nat Genet* **50**, 1151-1160 (2018).
1283 63. Olivares-Chauvet, P. *et al.* Capturing pairwise and multi-way chromosomal
1284 conformations using chromosomal walks. *Nature* **540**, 296-300 (2016).
1285 64. Lyu, X., Rowley, M.J. & Corces, V.G. Architectural Proteins and Pluripotency
1286 Factors Cooperate to Orchestrate the Transcriptional Response of hESCs to
1287 Temperature Stress. *Mol Cell* **71**, 940-955 e947 (2018).
1288 65. Petrovic, J. *et al.* Oncogenic Notch Promotes Long-Range Regulatory Interactions
1289 within Hyperconnected 3D Cliques. *Mol Cell* **73**, 1174-1190 e1112 (2019).
1290 66. Pattison, J.M. *et al.* Retinoic acid and BMP4 cooperate with p63 to alter
1291 chromatin dynamics during surface epithelial commitment. *Nat Genet* **50**, 1658-
1292 1665 (2018).
1293 67. Schuijers, J. *et al.* Transcriptional Dysregulation of MYC Reveals Common
1294 Enhancer-Docking Mechanism. *Cell Rep* **23**, 349-360 (2018).
1295 68. Johanson, T.M. *et al.* Transcription-factor-mediated supervision of global
1296 genome architecture maintains B cell identity. *Nat Immunol* **19**, 1257-1264
1297 (2018).
1298 69. Nolis, I.K. *et al.* Transcription factors mediate long-range enhancer-promoter
1299 interactions. *Proc Natl Acad Sci U S A* **106**, 20222-20227 (2009).
1300 70. Mas, G. & Di Croce, L. The role of Polycomb in stem cell genome architecture.
1301 *Curr Opin Cell Biol* **43**, 87-95 (2016).
1302 71. Jiang, T. *et al.* Identification of multi-loci hubs from 4C-seq demonstrates the
1303 functional importance of simultaneous interactions. *Nucleic Acids Res* **44**, 8714-
1304 8725 (2016).
1305 72. Boija, A. *et al.* Transcription Factors Activate Genes through the Phase-
1306 Separation Capacity of Their Activation Domains. *Cell* **175**, 1842-1855 e1816
1307 (2018).
1308 73. Cho, W.K. *et al.* Mediator and RNA polymerase II clusters associate in
1309 transcription-dependent condensates. *Science* **361**, 412-415 (2018).
1310 74. Chong, S. *et al.* Imaging dynamic and selective low-complexity domain
1311 interactions that control gene transcription. *Science* **361** (2018).
1312 75. Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K. & Sharp, P.A. A Phase
1313 Separation Model for Transcriptional Control. *Cell* **169**, 13-23 (2017).
1314 76. Durand, N.C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-
1315 Resolution Hi-C Experiments. *Cell Syst* **3**, 95-98 (2016).
1316 77. Han, T. *et al.* R-Spondin chromosome rearrangements drive Wnt-dependent
1317 tumour initiation and maintenance in the intestine. *Nat Commun* **8**, 15945
1318 (2017).
1319 78. Liu, Y. *et al.* Widespread Mitotic Bookmarking by Histone Marks and
1320 Transcription Factors in Pluripotent Stem Cells. *Cell Rep* **19**, 1283-1293 (2017).
1321 79. Buenrostro, J.D., Wu, B., Chang, H.Y. & Greenleaf, W.J. ATAC-seq: A Method for
1322 Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* **109**, 21 29
1323 21-29 (2015).

1324 80. Consortium, E.P. An integrated encyclopedia of DNA elements in the human
1325 genome. *Nature* **489**, 57-74 (2012).
1326 81. Huang da, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis
1327 of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57
1328 (2009).
1329 82. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. & Kingsford, C. Salmon provides fast
1330 and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417-419
1331 (2017).
1332 83. Lazaris, C., Kelly, S., Ntziachristos, P., Aifantis, I. & Tsirigos, A. HiC-bench:
1333 comprehensive and reproducible Hi-C data analysis designed for parameter
1334 exploration and benchmarking. *BMC Genomics* **18**, 22 (2017).
1335 84. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat*
1336 *Methods* **9**, 357-359 (2012).
1337 85. Tsirigos, A., Haiminen, N., Bilal, E. & Utro, F. GenomicTools: a computational
1338 platform for developing high-throughput analytics in genomics. *Bioinformatics*
1339 **28**, 282-283 (2012).
1340 86. Gong, Y. *et al.* Stratification of TAD boundaries reveals preferential insulation of
1341 super-enhancers by strong boundaries. *Nat Commun* **9**, 542 (2018).
1342 87. Ramirez, F. *et al.* High-resolution TADs reveal DNA sequences underlying genome
1343 organization in flies. *Nat Commun* **9**, 189 (2018).
1344 88. Sheffield, N.C. & Bock, C. LOLA: enrichment analysis for genomic region sets and
1345 regulatory elements in R and Bioconductor. *Bioinformatics* **32**, 587-589 (2016).
1346 89. Raviram, R. *et al.* 4C-ker: A Method to Reproducibly Identify Genome-Wide
1347 Interactions Captured by 4C-Seq Experiments. *PLoS Comput Biol* **12**, e1004780
1348 (2016).
1349 90. Rocha, P.P. *et al.* A Damage-Independent Role for 53BP1 that Impacts Break
1350 Order and Igh Architecture during Class Switch Recombination. *Cell Rep* **16**, 48-
1351 55 (2016).

1352

Figure 1

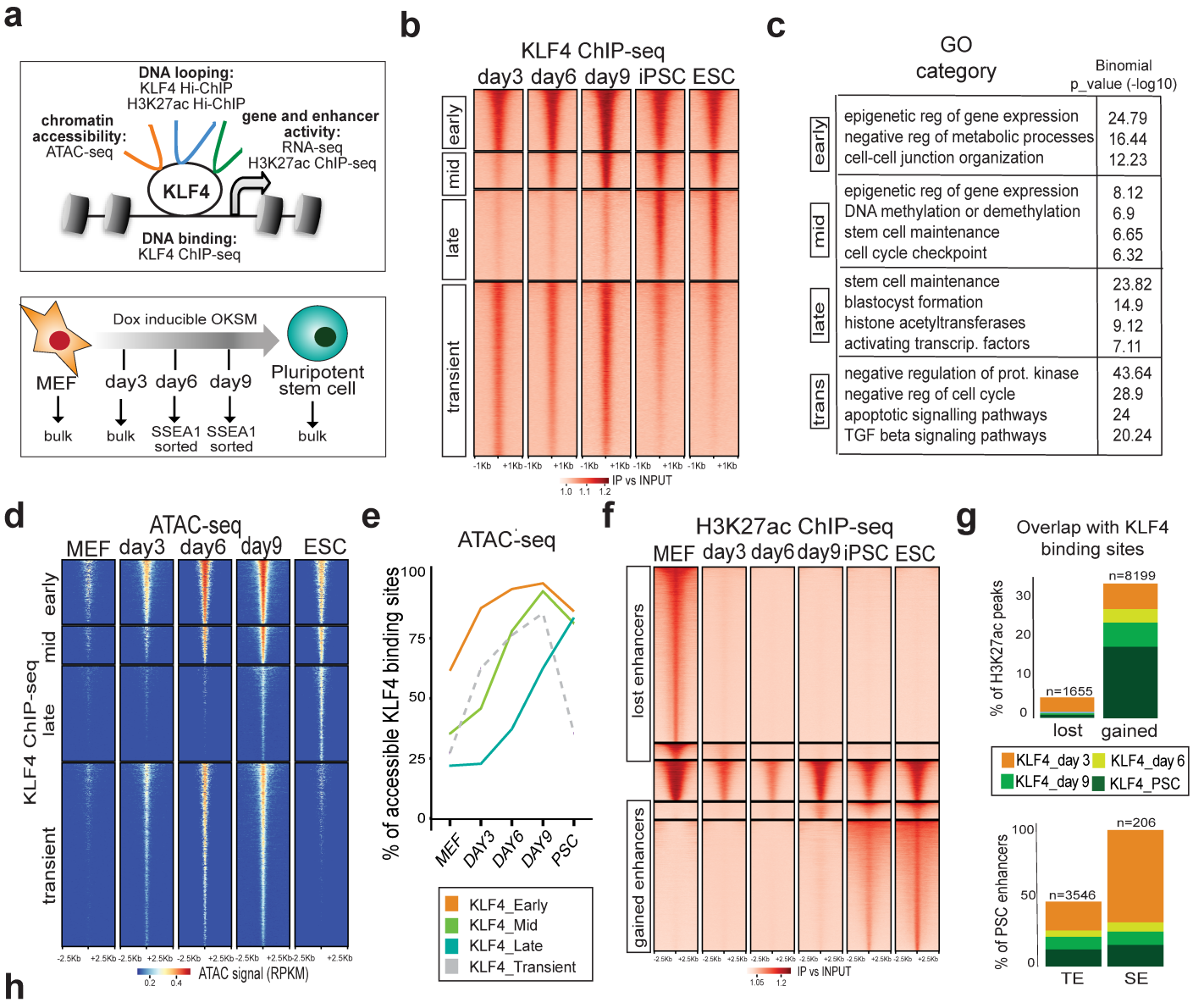

a
DNA looping:
KLF4 Hi-ChIP
H3K27ac Hi-ChIP
chromatin accessibility:
ATAC-seq
gene and enhancer activity:
RNA-seq
H3K27ac ChIP-seq
KLF4
DNA binding:
KLF4 ChIP-seq
Dox inducible OKSM
MEF
day3 day6 day9
Pluripotent stem cell
bulk
bulk
SSEA1 sorted
SSEA1 sorted
bulk
b
KLF4 ChIP-seq
day3 day6 day9 iPSC ESC
early
mid
late
transient
IP vs INPUT
c
GO category
Binomial p_value (-log10)
early
epigenetic reg of gene expression 24.79
negative reg of metabolic processes 16.44
cell-cell junction organization 12.23
mid
epigenetic reg of gene expression 8.12
DNA methylation or demethylation 6.9
stem cell maintenance 6.65
cell cycle checkpoint 6.32
late
stem cell maintenance 23.82
blastocyst formation 14.9
histone acetyltransferases 9.12
activating transcrip. factors 7.11
trans
negative regulation of prot. kinase 43.64
negative reg of cell cycle 28.9
apoptotic signalling pathways 24
TGF beta signaling pathways 20.24
d
ATAC-seq
MEF day3 day6 day9 ESC
KLF4 ChIP-seq
early
mid
late
transient
ATAC signal (RPKM)
e
ATAC-seq
% of accessible KLF4 binding sites
MEF DAY3 DAY6 DAY9 PSC
KLF4_Early
KLF4_Mid
KLF4_Late
KLF4_Transient
f
H3K27ac ChIP-seq
MEF day3 day6 day9 iPSC ESC
lost enhancers
gained enhancers
IP vs INPUT
g
Overlap with KLF4 binding sites
% of H3K27ac peaks
n=8199
n=1655
lost gained
KLF4_day 3
KLF4_day 6
KLF4_day 9
KLF4_PSC
% of PSC enhancers
n=206
n=3546
TE SE
h
KLF4:
H3K27ac:
Early Late
Early Late
Late Late
transient transient
transient repressed
chr1:180,887,483-180,893,839
Lefty2
chr14:99,293,255-99,311,973
Klf5
chr12:105,222,048-105,226,486
Tcl1
chr14:73,310,799-73,318,314
Rb1
chr6:127,140,305-127,163,074
Ccnd2
9330179D2Rik
KLF4 ChIP-seq
day3
day6
day9
PSC
ATAC-seq
MEF
day3
day6
day9
PSC
H3K27Ac ChIP-seq
MEF
day3
day6
day9
PSC
RNA-seq
TPM
MEF day3 day6 day9 PSC

Figure 1. Dynamic KLF4 binding during reprogramming and association with chromatin accessibility and enhancer activity. a Schematic illustration of the experimental system and strategy**. b** Tornado plots of KLF4 ChIP-seq signals at different reprogramming stages clustered in four different categories: Early, Mid, Late and Transient KLF4 binding. ChIP-seq signals (fold enrichment over input) are showing 1kb upstream/downstream of peak centers. **c,** GREAT gene ontology analysis of Early, Mid, Late and Transient KLF4 target sites. **d,** Tornado plot of ATAC-seq signal ar different reprogramming stages around KLF4 binding sites (Early, Mid, Late, Transient). ATAC-seq signals are showing 2.5kb upstream/downstream of peak centers. RPKM (Read Per Kilobase Million). **e,** Line plots showing the percentages of KLF4 Early, Mid, Late and Transient targets that overlapped with ATAC-seq peaks (accessible regions) at each reprogramming stage **f,** Tornado plot of H3K27ac ChIP-seq signal showing MEF peaks, PSC peaks and constant peaks at each reprogramming stage. ChIP-seq signals (fold enrichment over input) are showing 2.5kb upstream/downstream of peak centers **g,** Bar plots showing overlap of KLF4 binding with either lost (MEF) or gained (PSC) H3K27ac peaks (top) or with typical PSC enhancers (TE) versus superenhancers (SE) (as characterized by Whyte et al., 2013) (bottom). **h,** Examples of genomic regions (see genomic coordinates) that show different kinetics of KLF4 binding and H3K27ac occupancy during reprogramming. IGV tracks for KLF4 ChIP-seq, ATAC-seq, H3K27ac ChIP-seq at each reprogramming stage are shown and the signal values are indicated on the right. The transcriptional changes of the depicted genes during reprogramming are shown at the bottom, expressed as transcripts per million (TPM).

Figure 2

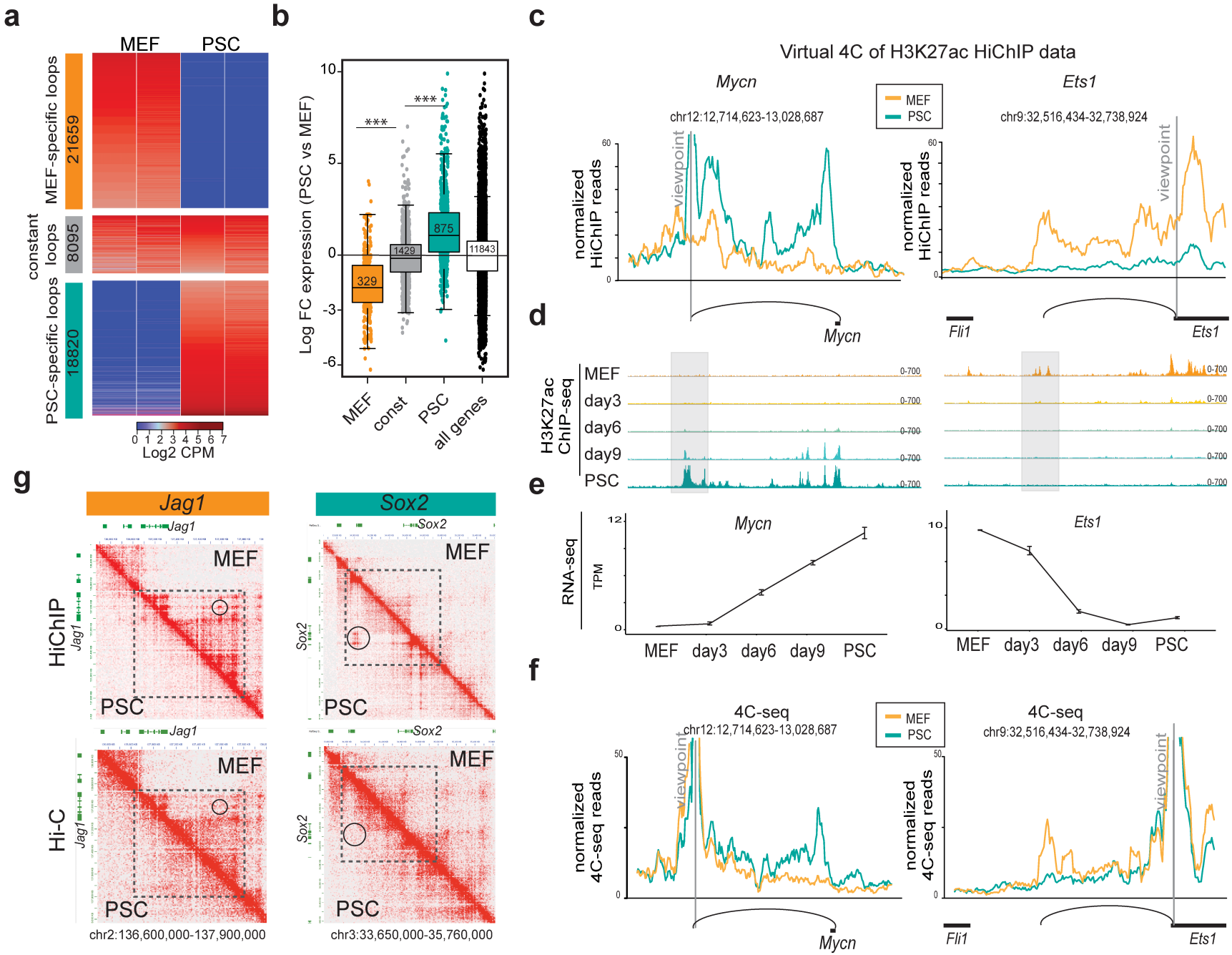

a
MEF
PSC
MEF-specific loops
21659
constant loops
8095
PSC-specific loops
18820
0 1 2 3 4 5 6 7
Log2 CPM
b
Log FC expression (PSC vs MEF)
10
5
0
-3
-6
***
***
329
1429
875
11843
MEF
const
PSC
all genes
c
Virtual 4C of H3K27ac HiChIP data
Mycn
chr12:12,714,623-13,028,687
Ets1
chr9:32,516,434-32,738,924
MEF
PSC
normalized HiChIP reads
viewpoint
Mycn
Fli1
Ets1
d
H3K27ac ChIP-seq
MEF
day3
day6
day9
PSC
0-700
e
RNA-seq
TPM
Mycn
Ets1
MEF
day3
day6
day9
PSC
f
4C-seq
chr12:12,714,623-13,028,687
4C-seq
chr9:32,516,434-32,738,924
normalized 4C-seq reads
viewpoint
Mycn
Fli1
Ets1
g
Jag1
Sox2
HiChIP
Hi-C
MEF
PSC
chr2:136,600,000-137,900,000
chr3:33,650,000-35,760,000

Figure 2. Characterization of 3D enhancer connectomes in MEFs and PSCs by H3K27ac HiChIP analysis. a, Heatmap of differential loops detected by H3K27ac HiChIP in MEF versus PSC. Differential loops were called by average logFC>2 or <-2 and p-value < 0.1, constant loops were called by average logFC >-0.5 & logFC <0.5 and p-value >0.5. Heatmap shows Log2 counts-per-million (CPM) per replicate. **b,** RNA expression changes between MEFs and PSCs of genes that were exclusively involved in at least one MEF-specific, PSC-specific or constant H3K27ac loops. All protein-coding genes were used as control. Asterisks indicate significant difference (p<0.001) as calculated by unpaired one-sided t-test. **c**, Virtual 4C representation of normalized H3K27ac HiChIP signals around selected viewpoints (*Mycn* enhancer and *Ets1* promoter) . The respective H3K27ac ChIP-seq IGV tracks are shown in **d,** while the RNA changes during reprogramming, expressed as transcripts per million (TPM), are shown in **e**. **f**, 4C-seq analysis around the same viewpoints as in (**c**) validate the presence and cell-type specificity of HiChIP-detected loops. 4C-seq signals normalized by sequencing depth and averaged across replicates are shown. **g**, HiChIP (top) and HiC (bottom) heatmaps generated by Juicebox[76] at 10Kb resolution around MEF-specific (*Jag1*) or PSC-specific (*Sox2*) contacts. Both PSC and MEF data are shown, separated by the diagonal. Signal indicates CPM normalized per matrix. Dotted squares indicate regions with cell-type specific configuration as detected by both HiC and HiChIP. Circles show examples of cell-type specific contacts that are detected in HiChIP and missed in HiC data.

Figure 3

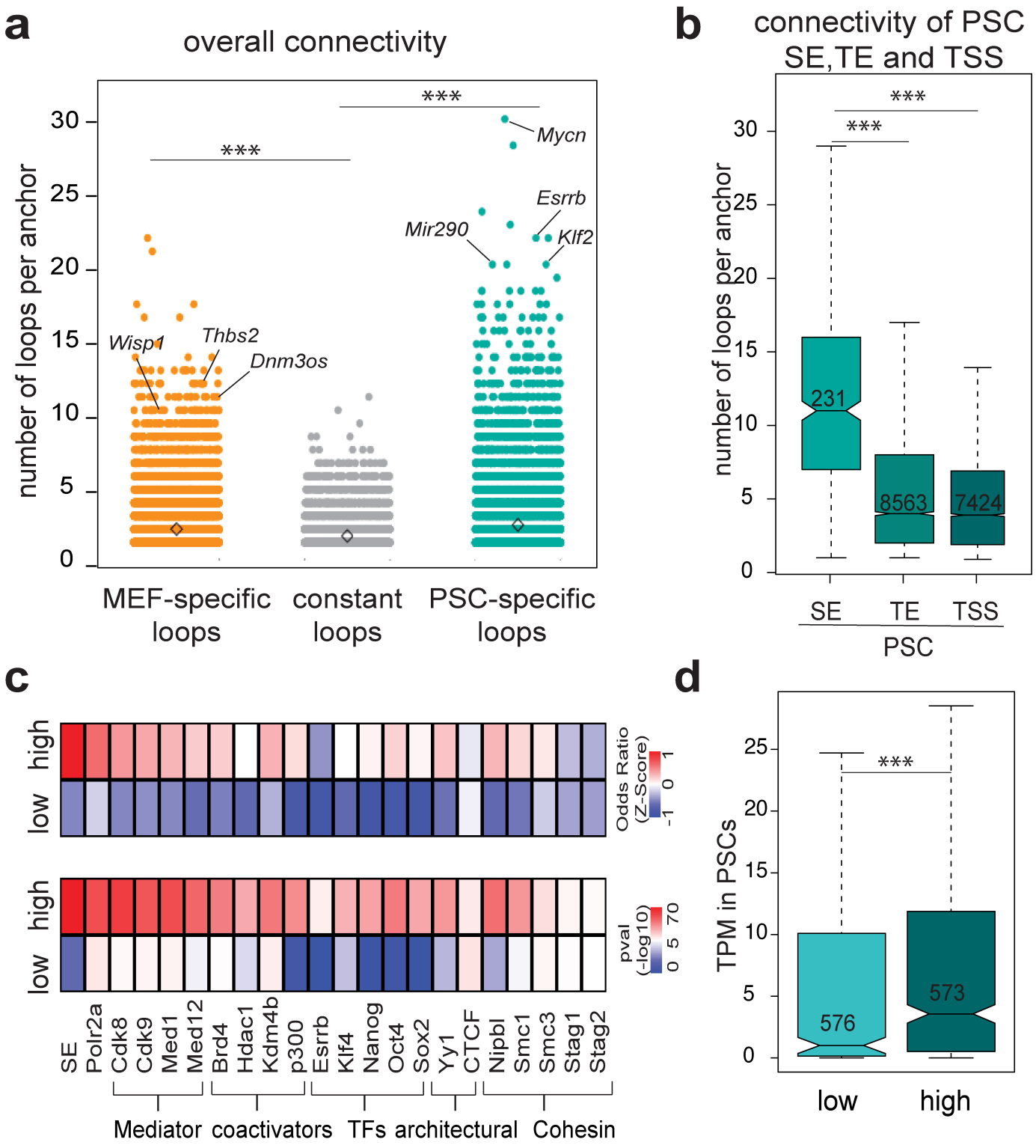

a
overall connectivity
number of loops per anchor
***
***
Mycn
Esrrb
Mir290
Klf2
Wisp1
Thbs2
Dnm3os
MEF-specific loops
constant loops
PSC-specific loops
b
connectivity of PSC SE,TE and TSS
number of loops per anchor
231
8563
7424
SE
TE
TSS
PSC
c
high
low
Odds Ratio (Z-Score)
pval (-log10)
SE
Polr2a
Cdk8
Cdk9
Med1
Med12
Brd4
Hdac1
Kdm4b
p300
Esrrb
Klf4
Nanog
Oct4
Sox2
Yy1
CTCF
Nipbl
Smc1
Smc3
Stag1
Stag2
Mediator
coactivators
TFs
architectural
Cohesin
d
TPM in PSCs
576
573
low
high

Figure 3. PSC enhancers are characterized by higher connectivity. a, Dot plot showing the number of high-confidence contacts (connectivity) around each H3K27ac HiChIP anchor. Asterisks indicate significant difference with $p<0.001$, as calculated by Wilcoxon rank sum test. **b,** Connectivity of HiChIP anchors containing PSC SE, TE or TSS in PSC. Asterisks indicate significant difference as in (**a) c,** LOLA enrichment analysis of enhancer anchors with low (n=1183) or high connectivity (n=1014) in PSCs using in-house and public ChIP-seq datasets from ESCs (see methods). Heatmaps represent either -log10 p-value (left) or z-score of odds ratio (right). **d**, Expression levels of genes found in low or high connected anchors (expressed in TPM).

Figure 4

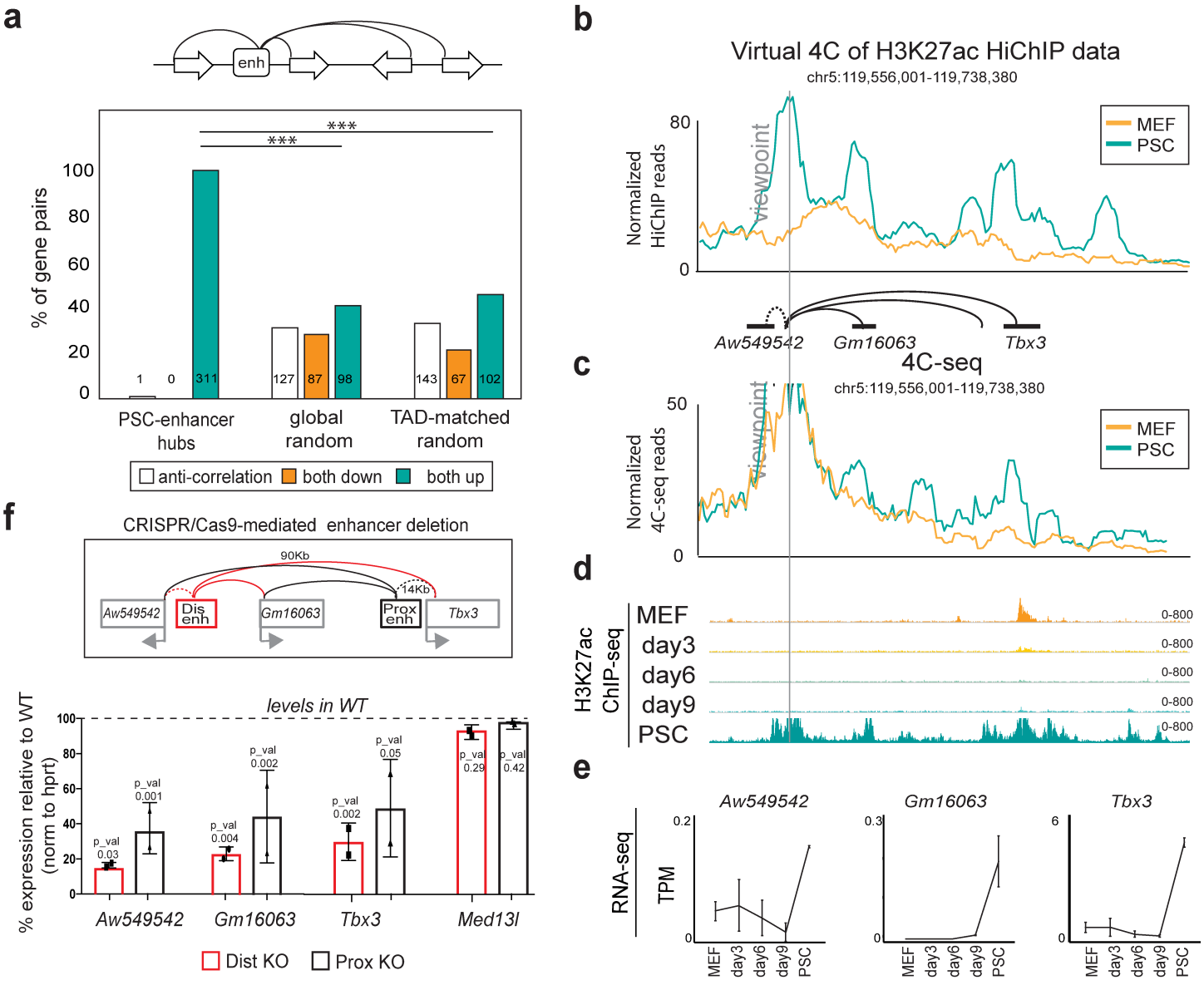

a
enh
% of gene pairs
***
***
1
0
311
127
87
98
143
67
102
PSC-enhancer hubs
global random
TAD-matched random
anti-correlation
both down
both up
b
Virtual 4C of H3K27ac HiChIP data
chr5:119,556,001-119,738,380
Normalized HiChIP reads
viewpoint
MEF
PSC
Aw549542
Gm16063
Tbx3
c
4C-seq
chr5:119,556,001-119,738,380
Normalized 4C-seq reads
viewpoint
MEF
PSC
d
H3K27ac ChIP-seq
MEF
day3
day6
day9
PSC
0-800
e
RNA-seq
TPM
Aw549542
Gm16063
Tbx3
MEF
day3
day6
day9
PSC
f
CRISPR/Cas9-mediated enhancer deletion
90Kb
14Kb
Aw549542
Dis enh
Gm16063
Prox enh
Tbx3
% expression relative to WT (norm to hprt)
levels in WT
p_val 0.03
p_val 0.001
p_val 0.004
p_val 0.002
p_val 0.002
p_val 0.05
p_val 0.29
p_val 0.42
Aw549542
Gm16063
Tbx3
Med13l
Dist KO
Prox KO

Figure 4. Co-regulation of genes within highly interacting enhancer hubs. a, Top: schematic representation of enhancer hubs interacting with two or more gene promoters. Bottom: Barplot indicating the percentage of gene pairs within enhancer hubs that become transcriptionally co-regulated (both up or both down with log2 fold change >=1 or <=-1 & p-adj<=0.01) or anti-regulated (one up and one down) between MEFs and PSCs. Global Random or TAD-matched gene pairs were used as controls (see also Methods). Non-differentially expressed genes were not considered in this analysis (n=487). Significance is indicated by asterisks and was calculated by Fisher's exact test. **b**, Example of a newly identified enhancer hub in PSCs. Normalized HiChIP signal around the viewpoint is illustrated as a virtual 4C plot. **c**, 4C-seq analysis around the same viewpoint as in (b). 4C-seq counts normalized per sequencing depth are plotted. **d**, H3K27ac ChIP-seq IGV tracks during reprogramming. **e**, RNA-seq signal (TPM) of genes within the hub are shown to highlight coordinated upregulation during reprogramming. **f**, Top: experimental strategy for CRISPR-Cas9-mediated deletions of the *Tbx3* distal (Dis) or proximal (Prox) enhancers within the hub indicated in panel (b). Bottom: RT-qPCR showing expression changes of *Tbx3*, *Gm16063*, *Aw54954* and a control gene outside the hub (*Med13l*) in CRISPR-Cas9 engineered PSC carrying homozygous deletions of the distal (Dis-KO) and proximal (Prox-KO) *Tbx3* enhancer calculated as percentage relative to wild=type (WT). P-values are calculated using unpaired one-sided t-test. Error bars indicate standard deviation from n=2 biological replicates. KO: knockout.

Figure 5

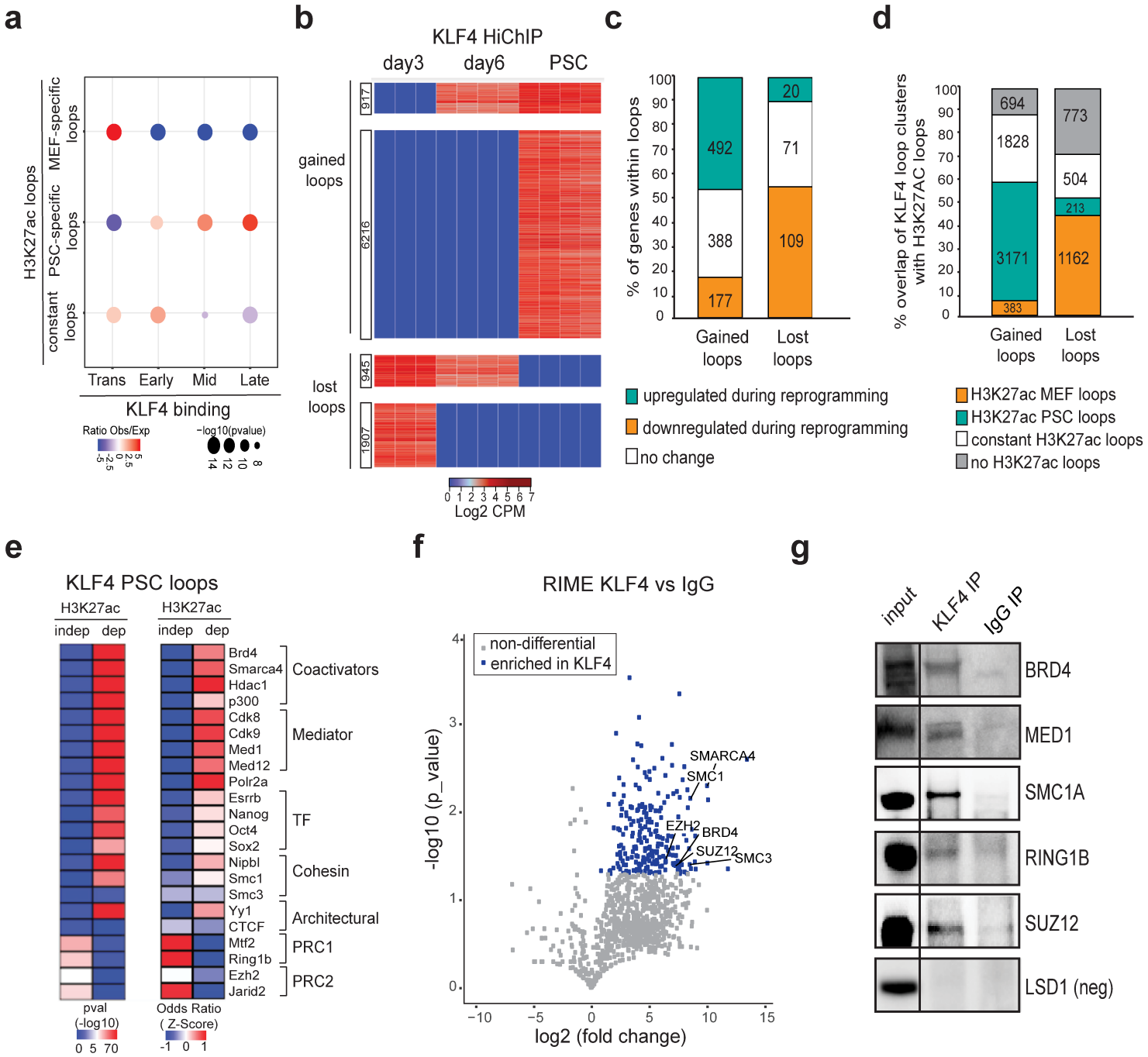

a
H3K27ac loops
MEF-specific loops
PSC-specific loops
constant loops
Trans
Early
Mid
Late
KLF4 binding
Ratio Obs/Exp
-log10(pvalue)
b
KLF4 HiChIP
day3
day6
PSC
gained loops
lost loops
917
6216
945
1907
Log2 CPM
c
% of genes within loops
492
388
177
20
71
109
Gained loops
Lost loops
upregulated during reprogramming
downregulated during reprogramming
no change
d
% overlap of KLF4 loop clusters with H3K27AC loops
694
1828
3171
383
773
504
213
1162
Gained loops
Lost loops
H3K27ac MEF loops
H3K27ac PSC loops
constant H3K27ac loops
no H3K27ac loops
e
KLF4 PSC loops
H3K27ac
indep
dep
Brd4
Smarca4
Hdac1
p300
Cdk8
Cdk9
Med1
Med12
Polr2a
Esrrb
Nanog
Oct4
Sox2
Nipbl
Smc1
Smc3
Yy1
CTCF
Mtf2
Ring1b
Ezh2
Jarid2
Coactivators
Mediator
TF
Cohesin
Architectural
PRC1
PRC2
pval (-log10)
Odds Ratio (Z-Score)
f
RIME KLF4 vs IgG
non-differential
enriched in KLF4
-log10 (p_value)
log2 (fold change)
SMARCA4
SMC1
EZH2
BRD4
SUZ12
SMC3
g
input
KLF4 IP
IgG IP
BRD4
MED1
SMC1A
RING1B
SUZ12
LSD1 (neg)

Figure 5. Chromatin reorganization around KLF4 binding sites during reprogramming associates with enhancer rewiring and requires additional cofactors. a, Dot plot showing overlap of MEF-specific loops, PSC-specific loops or constant loops as detected by H3K27ac HiChIP with KLF4 Early, Mid, Late and Transient ChIP-seq peaks**.** The size of the dot represents p-value (as calculated by Fisher's exact test), while the color indicates the ratio of observed (Obs) versus expected (Exp). **b**, Heatmap of differential KLF4 HiChIP analysis depicting 4 distinct clusters grouped into gained or lost loops. Differential loops were called by average logFC > 2/ or <-2 and p-value < 0.01 between specific pair-wise comparisons (see Methods). Heatmap shows Log2 CPM per replicate. **c**, Stacked barplot indicating the relative proportion of genes within gained or lost KLF4 loops that become upregulated or downregulated (logFC > 1.0, FDR < 0.01 in PSC vs day3) or remain unchanged (logFC > -0.25 & logFC < 0.25) during reprogramming. Numbers of genes per categopry are shown in the respective bars. **d.** Stacked barplot showing the percentage of gained or lost KLF4 loops that were also detected by H3K27ac HiChIP analysis in either MEFs or PSC or in both (constant loops). Note that among all the KLF4 PSC loops, 26% are H3K27ac independent (see Supplementary Figure 4c). **e**, LOLA enrichment analysis of KLF4 binding sites in PSCs that overlap either with H3K27ac-dependent loops (detected by both KLF4 and H3K27ac HiChIP) or -independent (detected only by KLF4 HiChIP). Selected factors that scored as significantly enriched over background are shown. Heatmaps represent either -log10 of p-value (left) or z-score of OddsRatio (right). **f**, Volcano plot showing relative enrichment of proteins that were co-immunoprecipitated with KLF4 versus IgG as identified by RIME (rapid immunoprecipitation mass spectrometry of endogenous protein). Significantly enriched proteins with a p-value< 0.05 and FC >1.5 are colored in blue. Selected co-factors are labeled. **g**, Immunoprecipitation using KLF4 antibody or IgG in PSC extracts followed by western blot analysis validated interaction with selected factors. LSD1 was used as negative control.

Figure 6

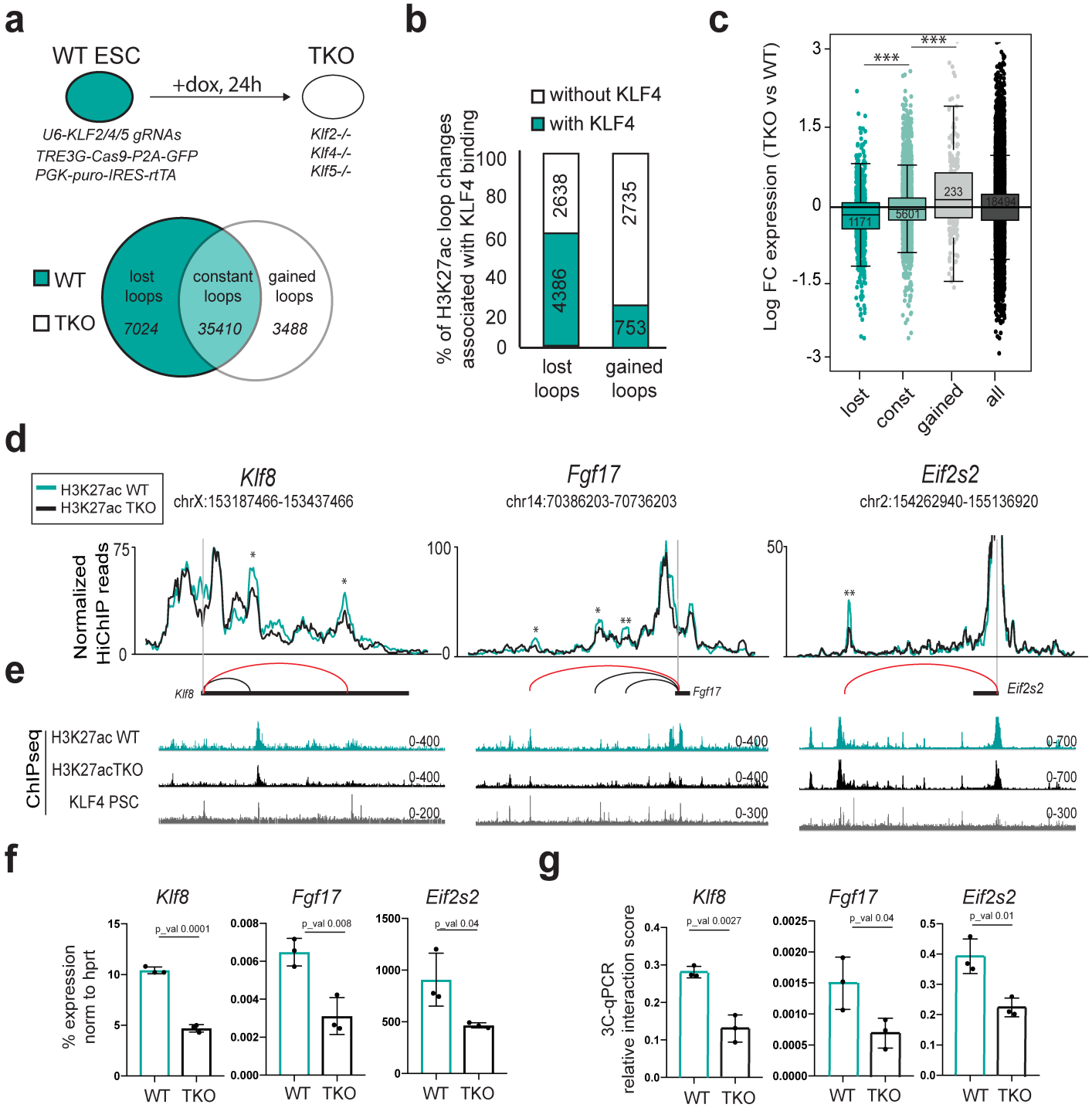

a
WT ESC
+dox, 24h
TKO
U6-KLF2/4/5 gRNAs
TRE3G-Cas9-P2A-GFP
PGK-puro-IRES-rtTA
Klf2-/-
Klf4-/-
Klf5-/-
WT
TKO
lost loops
constant loops
gained loops
7024
35410
3488
b
without KLF4
with KLF4
% of H3K27ac loop changes associated with KLF4 binding
2638
4386
2735
753
lost loops
gained loops
c
Log FC expression (TKO vs WT)
***
***
1171
5601
233
18494
lost
const
gained
all
d
H3K27ac WT
H3K27ac TKO
Klf8
chrX:153187466-153437466
Fgf17
chr14:70386203-70736203
Eif2s2
chr2:154262940-155136920
Normalized HiChIP reads
e
ChIPseq
H3K27ac WT
H3K27acTKO
KLF4 PSC
0-400
0-400
0-200
0-400
0-400
0-300
0-700
0-700
0-300
f
Klf8
Fgf17
Eif2s2
% expression norm to hprt
p_val 0.0001
p_val 0.008
p_val 0.04
WT
TKO
g
3C-qPCR relative interaction score
p_val 0.0027
p_val 0.04
p_val 0.01
WT
TKO

Figure 6. Inducible depletion of KLF proteins induces 3D enhancer reorganization and concordant transcriptional changes. a, Top: schematic diagram of the experimental approach used to knock out (KO) KLF2, KLF4 and KLF5 protein in ESCs using a doxycycline (dox)- inducible CRISPR-Cas9 construct. Bottom: Venn diagram showing the number of H3K27ac HiChIP loops that were gained or lost (p-value<0.05 and fold change >1.5 or <-1.5) or remained constant (logFC >-0.5 & <0.5 and p-value>0.5) in triple knock out (TKO) ESCs compared to uninduced (WT) ESCs. **b**, Stacked barplots showing the percentage of gained or lost H3K27ac HiChIP loops in TKO versus WT, whose anchors overlap or not with KLF4 ChIP-seq peaks in PSCs. Numbers represent the actual number of loops. **c**, RNA expression changes of genes within anchors of H3K27ac HiChIP loops (lost, constant or gained loops). All protein-coding genes were used as control. The respective numbers of genes are shown in the boxes. Asterisks indicate significant difference (p<0.001) as calculated by an unpaired one-sided t-test. **d**, Examples of H3K27ac lost loops in TKO vs WT ESC as identified by H3K27ac HiChIP. Normalized H3K27ac HiChIP signals are illustrated in a virtual 4C format around the viewpoints (*Klf8* promoter*, Fgf17* promoter*, Eif2s2* promoter*).* Asterisks mark the differential loops detected (* p<0.1, ** p<0.01). Statistics were calculated with the R-package edgeR (see Methods for more details). **e**, H3K27ac and KLF4 ChIP-seq tracks around each of the genomic regions indicated in (**d**). **f,** RT-qPCR showing expression changes of *Klf8, Fgf17 and Eif2s2* in WT and TKO PSC calculated as percentage relative to *Hprt* levels**.** P-values were calculated using an unpaired one-sided t-test. Error bars indicate standard deviation from n=3 biological replicates. **g**, 3C-qPCR analysis validating the reduced contact frequency between *Klf8, Fgf17 and Eif2s2* promoters and their respective distal enhancers (marked with a red line in panel (**d)**) in TKO compared to WT ESCs. Unpaired one-sided t-test was used to determine P-values. Error bars indicate standard deviation using n=3 biological replicates.

Figure 7

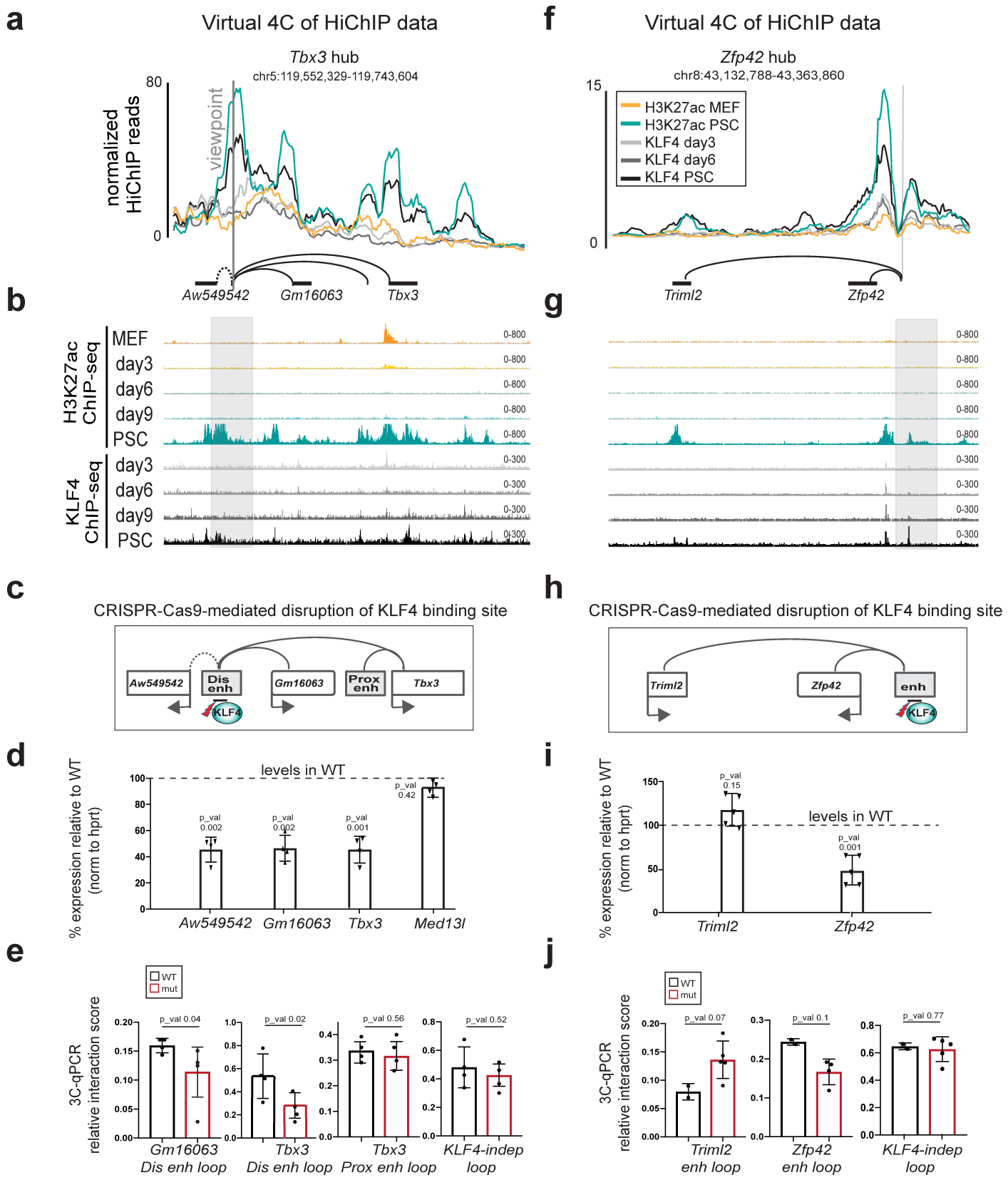
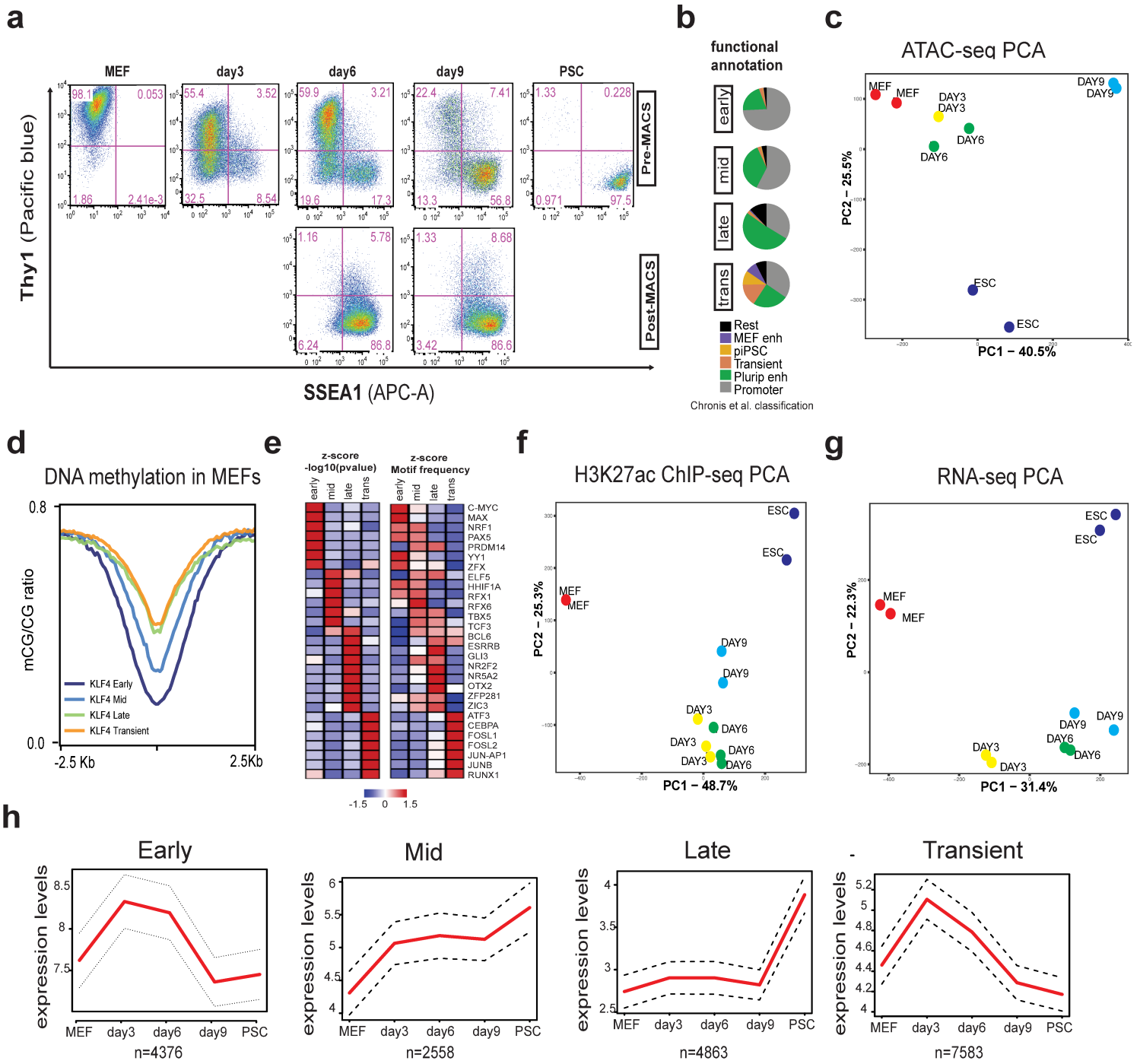

a
Virtual 4C of HiChIP data
Tbx3 hub
chr5:119,552,329-119,743,604
normalized HiChIP reads
viewpoint
Aw549542
Gm16063
Tbx3
b
H3K27ac ChIP-seq
MEF
day3
day6
day9
PSC
KLF4 ChIP-seq
c
CRISPR-Cas9-mediated disruption of KLF4 binding site
Dis enh
Prox enh
KLF4
d
% expression relative to WT (norm to hprt)
levels in WT
p_val 0.002
p_val 0.002
p_val 0.001
p_val 0.42
Med13l
e
WT
mut
3C-qPCR relative interaction score
p_val 0.04
p_val 0.02
p_val 0.56
p_val 0.52
Gm16063 Dis enh loop
Tbx3 Dis enh loop
Tbx3 Prox enh loop
KLF4-indep loop
f
Virtual 4C of HiChIP data
Zfp42 hub
chr8:43,132,788-43,363,860
H3K27ac MEF
H3K27ac PSC
KLF4 day3
KLF4 day6
KLF4 PSC
Triml2
Zfp42
g
h
CRISPR-Cas9-mediated disruption of KLF4 binding site
enh
i
p_val 0.15
p_val 0.001
j
p_val 0.07
p_val 0.1
p_val 0.77
Triml2 enh loop
Zfp42 enh loop

Figure 7. Disruption of KLF4 binding site within *Tbx3* and *Zfp42* enhancers induces looping abrogation and downregulation of linked genes in PSCs. a, Normalized KLF4 and H3K27ac HiChIP signals are illustrated as virtual 4C line plots around the *Tbx3* distal enhancer hub (see also Fig.4b-f). The respective ChIP-seq IGV tracks are shown in **b**. **c**, Schematic illustration of the CRISPR-Cas9 targeting strategy to generate mutated KLF4 binding motifs (mut) within the distal *Tbx3* enhancer. **d**, RT-qPCR showing expression changes of hub-associated genes (*Tbx3*, *Gm16063* and *Aw54954)*. *Med13l* is used as control gene outside the hub. Values were calculated as percentage relative to wild type (WT) after normalization relative to *Hprt* mRNA levels**.** Unpaired one-sided t-test was used to determine significance relative to WT (p-value is indicated on the top of each bar). Error bars indicate standard deviation from n=4 different PSC clones carrying homozygous mutations of KLF4 binding motif (mut). **e**, 3C-qPCR analysis showing the relative interaction frequency of *Tbx3* distal enhancer with the promoters of linked genes in WT and mutant (mut) clones. Error bars indicate standard deviation. *n*=2 for WT and n=4 for mut biological replicates. Unpaired one-tailed Student's t-test was used to determine significance relative to WT (the value is indicated on the top of each bar). **f-j,** Representation, analysis and functional validation of *Zfp42* enhancer hub similarly to panels (a-e) for *Tbx3* hub. The same normalizations and statistical tests were applied, with the only difference that n=5 mutant clones carrying homozygous mutations of KLF4 binding motif were used.

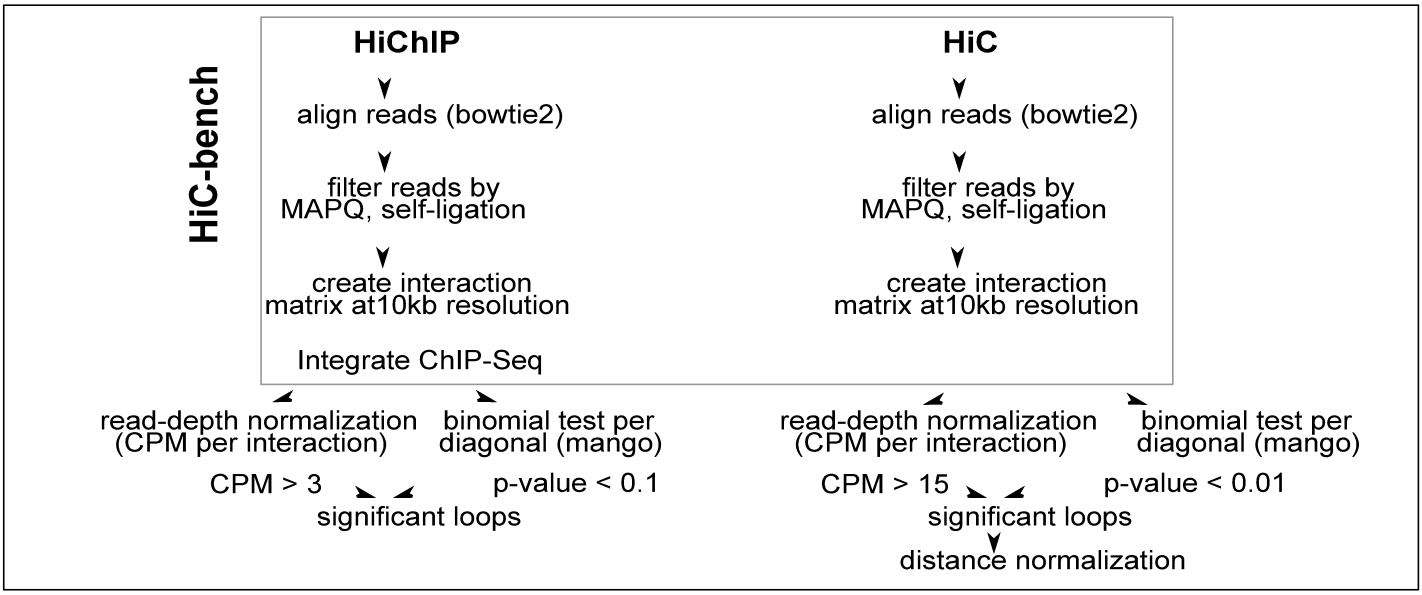
Figure S1


a
MEF
day3
day6
day9
PSC
Thy1 (Pacific blue)
SSEA1 (APC-A)
Pre-MACS
Post-MACS
b
functional annotation
early
mid
late
trans
Rest
MEF enh
piPSC
Transient
Plurip enh
Promoter
Chronis et al. classification
c
ATAC-seq PCA
PC2 – 25.5%
PC1 – 40.5%
d
DNA methylation in MEFs
mCG/CG ratio
KLF4 Early
KLF4 Mid
KLF4 Late
KLF4 Transient
-2.5 Kb
2.5Kb
e
z-score -log10(pvalue)
z-score Motif frequency
C-MYC
MAX
NRF1
PAX5
PRDM14
YY1
ZFX
ELF5
HHIF1A
RFX1
RFX6
TBX5
TCF3
BCL6
ESRRB
GLI3
NR2F2
NR5A2
OTX2
ZFP281
ZIC3
ATF3
CEBPA
FOSL1
FOSL2
JUN-AP1
JUNB
RUNX1
f
H3K27ac ChIP-seq PCA
PC2 – 25.3%
PC1 – 48.7%
g
RNA-seq PCA
PC2 – 22.3%
PC1 – 31.4%
h
Early
Mid
Late
Transient
expression levels
MEF
day3
day6
day9
PSC
n=4376
n=2558
n=4863
n=7583

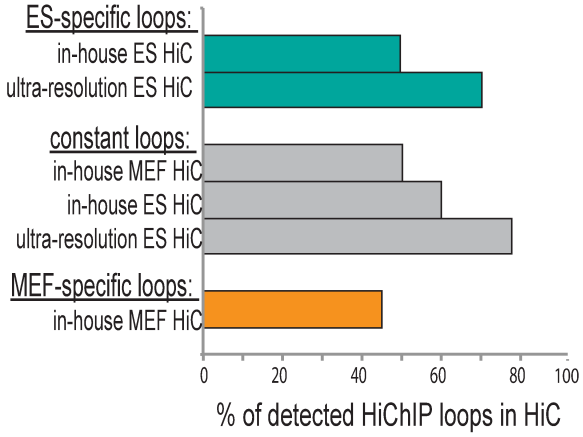
Supplementary Figure 1. a, FACS analysis plots showing expression of SSEA1 (early pluripotency marker) and Thy1 (somatic marker) at different stages of reprogramming, before and after SSEA1 enrichment by MACS isolation. **b**, Pie charts of functional classification of KLF4 Early, Mid, Late and Transient peaks (based on Chronis et al. 2017) (piPSC= partial iPSCs). **c**, PCA analysis of ATAC-seq peaks in MEF, PSC and different stages of reprogramming. **d**, Average line plot showing the methylated CG to non-methylated CG ratio from MEF data[12] centered (+/-2.5Kb) around different clusters of KLF4 binding sites (Early, Mid, Late or Transient KLF4 targets, Fig.2b)**.** **e**, Motif enrichment for Early, Mid, Late and Transient KLF4 binding sites. Selected factors are shown and their significance is expressed as Z-score of –log10(pvalue) (left) or z-score of motif frequency (right). **f**, PCA analysis of H3K27ac ChIP-seq peaks called in MEF, PSC and different stages of reprogramming **g**, PCA of RNA-seq in MEF, PSC and different stages of reprogramming. **h**, Line plots of the median expression (red line) of genes closest to Early, Mid, Late and Transient peaks, expressed as TPM (transcripts per million).

Figure S2

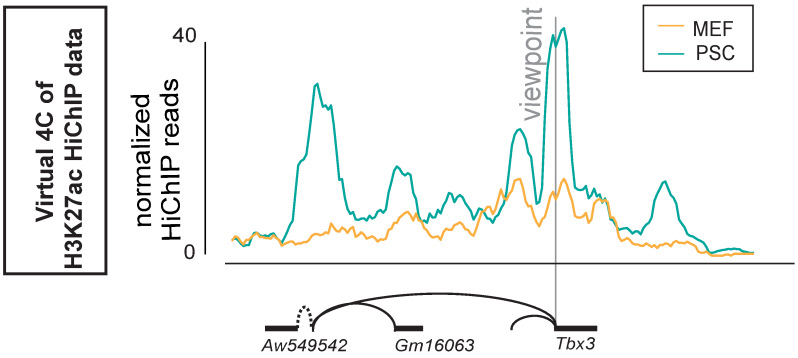
a

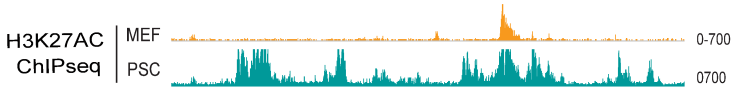


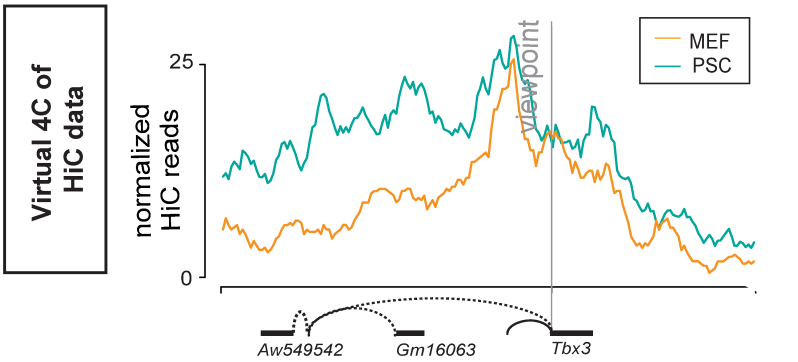
b



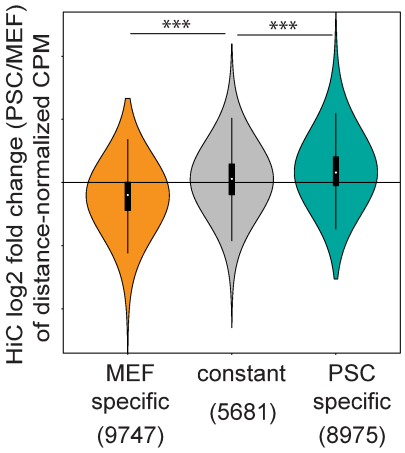
c





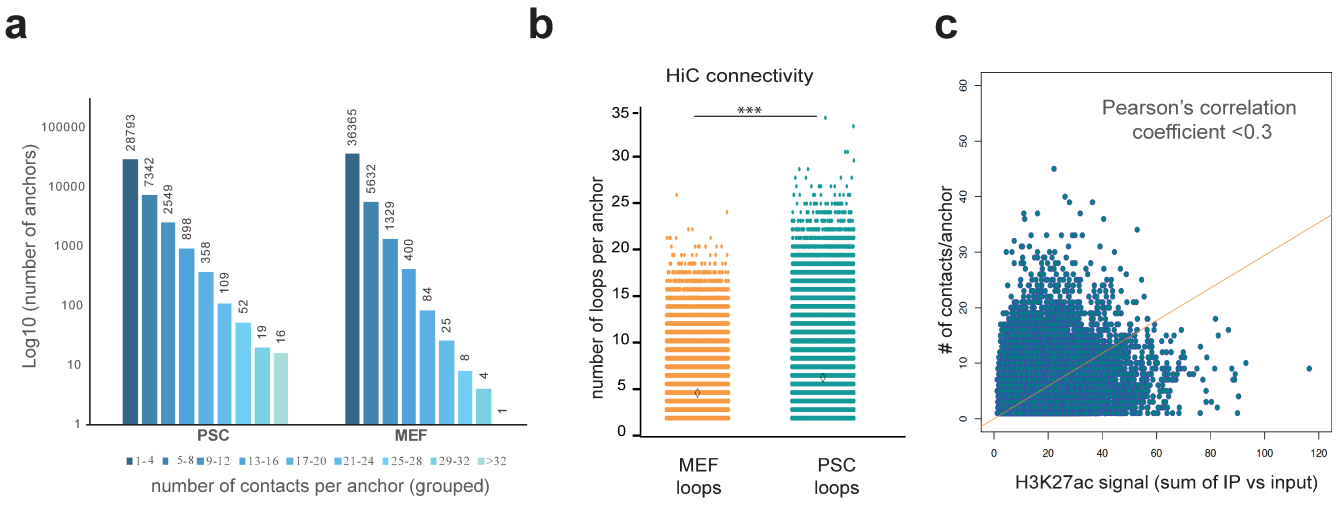


d



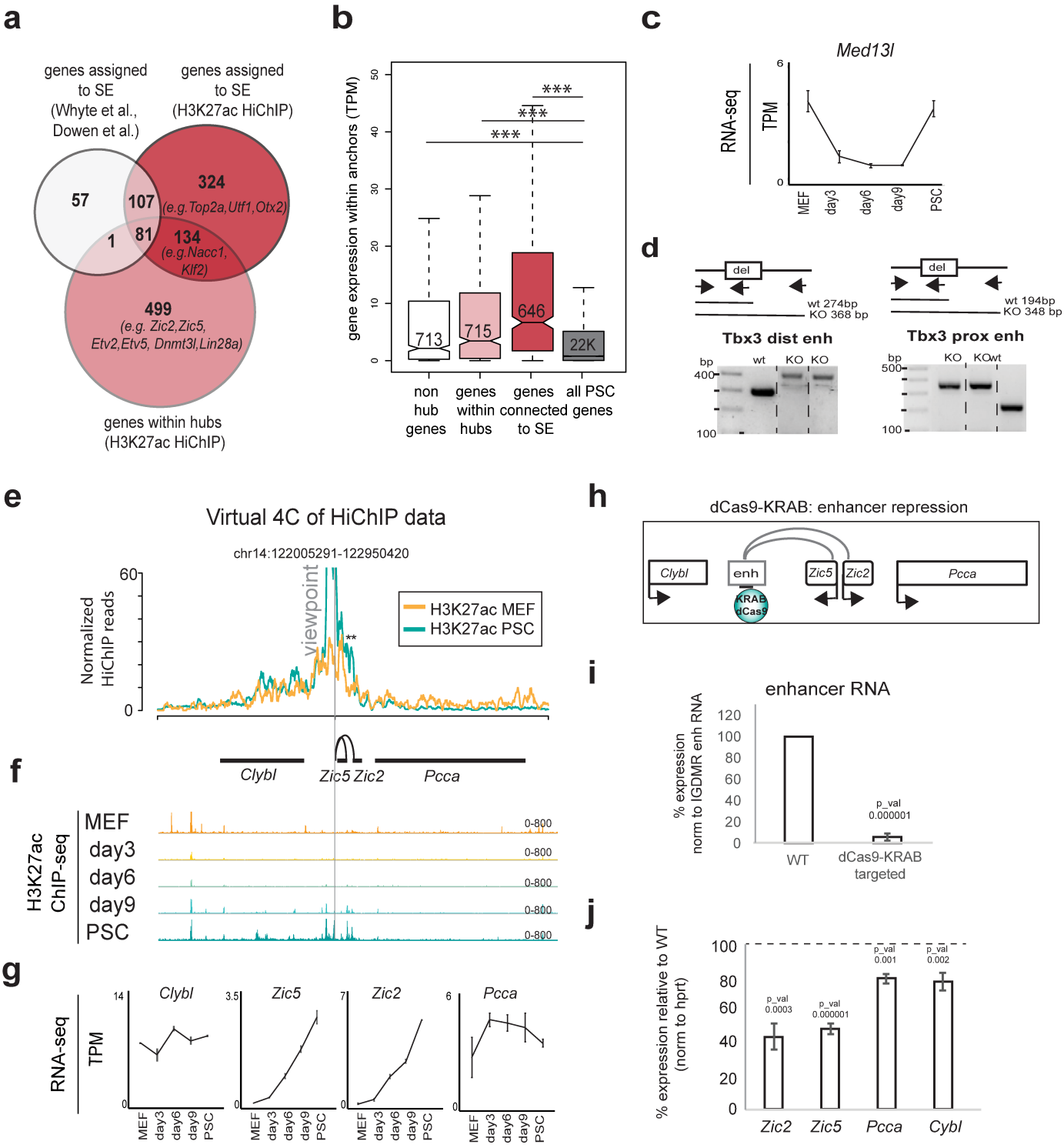
Supplementary Figure 2. a, Schematic work-flow for HiChIP and HiC analysis. **b**, Percentages of PSC-specific, constant or MEF-specific H3K27ac HiChIP loops that were detected in HiC experiments (either generated in-house or published ultra-resolution HiC in PSC[38]). **c**, Normalized HiChiP (top) and HiC (bottom) signals in MEF and PSC are illustrated in a virtual 4C format around the indicated viewpoint (*Tbx3* promoter). H3K27ac ChIP-seq tracks are shown in MEF and PSC. **d**, Violin plot representing log2 fold change of distance-normalized HiC signal in PSCs versus MEFs of MEF-specific, constant and PSC-specific loops as called by H3K27ac HiChIP. Only contacts that were detected as significant in HiC data are considered. Numbers of considered loops per category are shown in parenthesis. Unpaired two-sided t-test was used to determine the p-value.

Figure S3


a
Log10 (number of anchors)
100000
10000
1000
100
10
1
28793
7342
2549
898
358
109
52
19
16
36365
5632
1329
400
84
25
8
4
1
PSC
MEF
1-4
5-8
9-12
13-16
17-20
21-24
25-28
29-32
>32
number of contacts per anchor (grouped)
b
HiC connectivity
***
number of loops per anchor
35
30
25
20
15
10
5
0
MEF loops
PSC loops
c
Pearson's correlation coefficient <0.3
# of contacts/anchor
0
10
20
30
40
50
60
0
20
40
60
80
100
120
H3K27ac signal (sum of IP vs input)

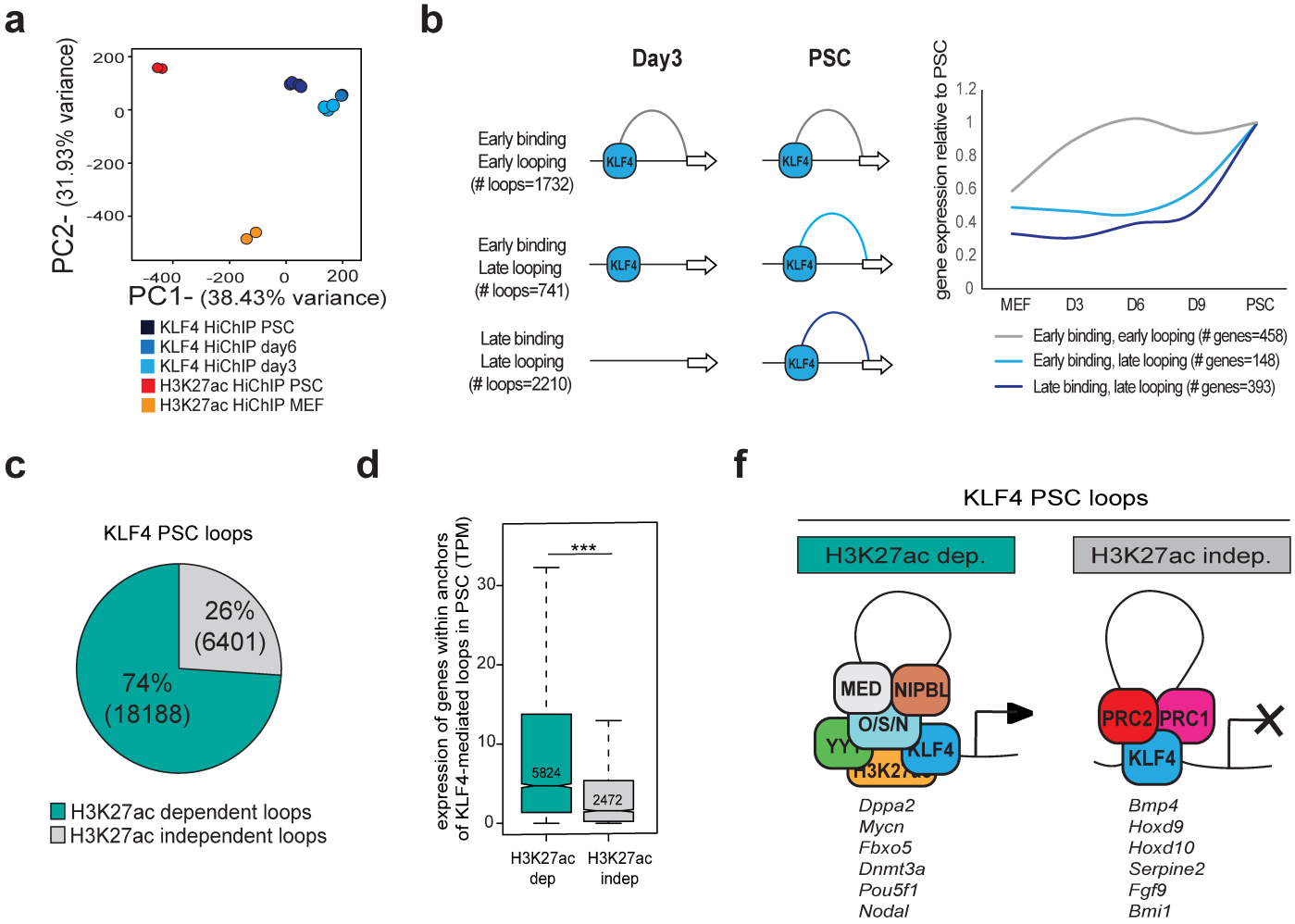
Supplementary Figure 3. a, Histogram of anchor connectivity based on H3K27ac MEF and PSC HiChIP called loops. The numbers of contacts per anchor are grouped as shown in the bottom and the actual number of anchors is depicted on top of each bar. **b,** Connectivity of MEF or PSC anchors based on HiC-called loops represented as number of high-confidence contacts around each 10kb anchor. Wilcoxon rank sum test was used to compare connectivity and asterisks indicate significant difference with $p<0.001$. **c**, Scatter plot showing the correlation of H3K27ac ChIP-seq strength (sum of H3K27ac ChIP/input of all peaks within the anchor) with the number of H3K27ac HiChIP contacts per anchor in PSCs.

Figure S4


a
genes assigned to SE (Whyte et al., Dowen et al.)
genes assigned to SE (H3K27ac HiChIP)
genes within hubs (H3K27ac HiChIP)
57
107
324
(e.g.Top2a,Utf1,Otx2)
1
81
134
(e.g.Nacc1, Klf2)
499
(e.g. Zic2,Zic5, Etv2,Etv5, Dnmt3l,Lin28a)
b
gene expression within anchors (TPM)
713
715
646
22K
non hub genes
genes within hubs
genes connected to SE
all PSC genes
c
Med13l
RNA-seq
TPM
MEF
day3
day6
day9
PSC
d
del
wt 274bp
KO 368 bp
wt 194bp
KO 348 bp
Tbx3 dist enh
Tbx3 prox enh
bp
wt
KO
e
Virtual 4C of HiChIP data
chr14:122005291-122950420
Normalized HiChIP reads
viewpoint
H3K27ac MEF
H3K27ac PSC
Clybl
Zic5 Zic2
Pcca
f
H3K27ac ChIP-seq
MEF
day3
day6
day9
PSC
0-800
g
RNA-seq
TPM
Clybl
Zic5
Zic2
Pcca
h
dCas9-KRAB: enhancer repression
Clybl
enh
KRAB dCas9
Zic5
Zic2
Pcca
i
enhancer RNA
% expression norm to IGDMR enh RNA
p_val 0.000001
WT
dCas9-KRAB targeted
j
% expression relative to WT (norm to hprt)
p_val 0.0003
p_val 0.000001
p_val 0.001
p_val 0.002
Zic2
Zic5
Pcca
Cybl

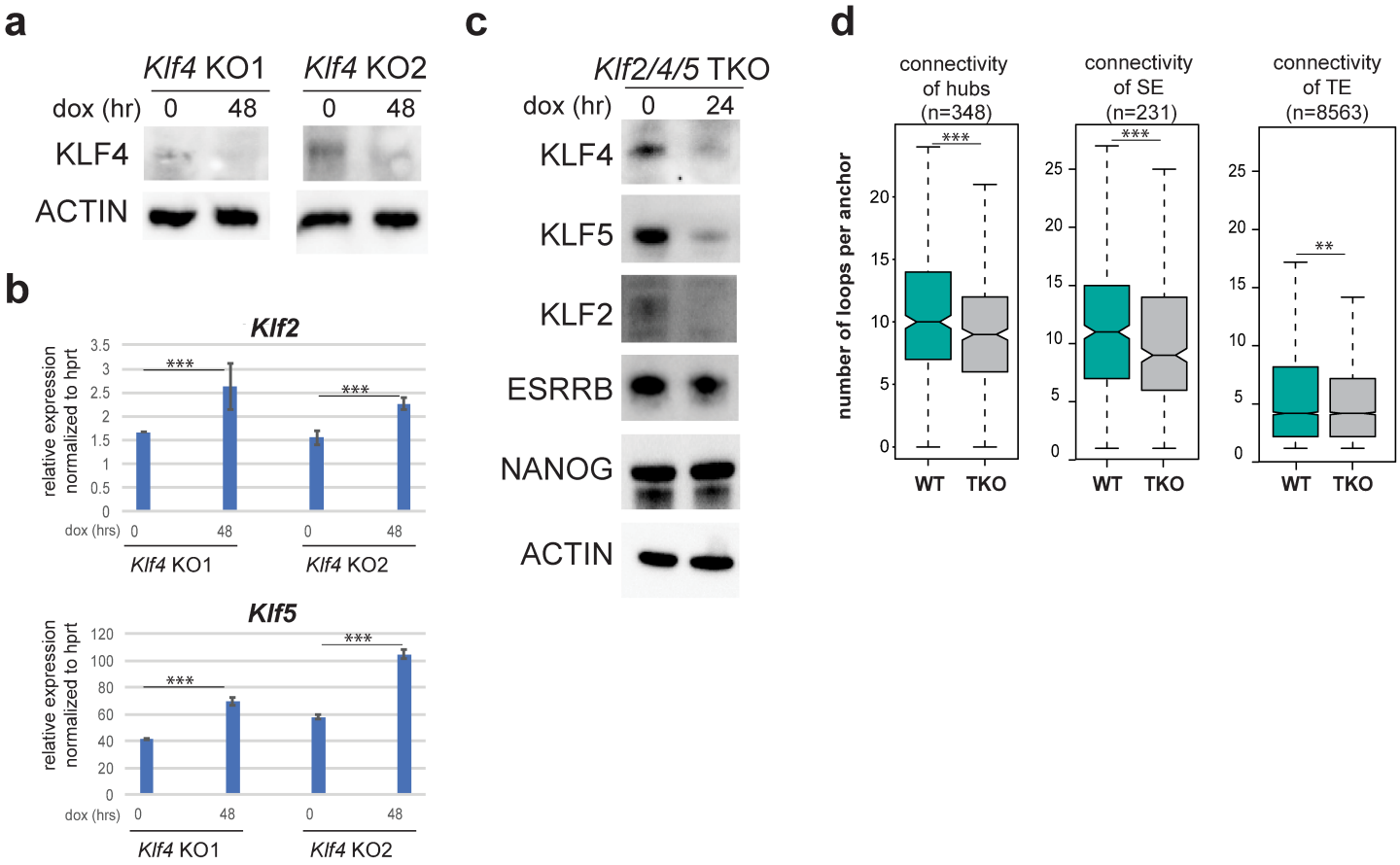
Supplementary Figure 4. a, Venn diagram showing overlap between previously assigned target genes for super-enhancers (SE), newly identified SE target genes based on H3K27ac HiChIP contacts in PSCs, and genes connected to PSC-specific enhancer hubs, which represent enhancers contacting more than one gene according H3K27ac HiChIP (see also Fig.4a). **b**, RNA levels of hub genes, non-hub genes or genes connected to SE in PSC samples as measured by RNA-seq and expressed as transcripts per million (TPM). All genes that are not connected to enhancer hubs, but are still detected within PSC-specific HiChIP loops were considered. Expression of all genes expressed in PSC (>1TPM) is shown as reference. **c**, RNA-seq signal (TPM) of *Med13l* -which is not part of the *Tbx3* enhancer hub (see Fig.4b)- during reprogramming **d**, Genotyping strategy and results confirming the homozygous deletion of the distal (left) or the proximal (right) *Tbx3* enhancers. **e**, Example of a newly identified enhancer hub in PSCs. Normalized HiChIP signal around the viewpoint is illustrated as a virtual 4C plot. **f**, H3K27ac ChIP-seq IGV tracks during reprogramming. **g**, RNA-seq signal of genes within the hub (*Zic2* and *Zic5*), or nearby genes (*Clybl* and *Pcca*), are shown for each reprogramming stage to highlight concordance with H3K27ac HiChIP data and coordinated upregulation of genes within the hub. **h**, Schematic illustration of the CRSIPRi (dCas9-KRAB) targeting strategy for inactivation of the *Zic2/Zic5* enhancer hub. **i**, RT-qPCR showing relative levels of the enhancer RNA (normalized to an unaffected enhancer RNA (IGDMR)) in wild type (WT) or dCas9-KRAB-targeted ESCs. P-values were calculated using unpaired one-tailed t-test. Error bars indicate standard deviation from n=2 biological replicates. **j**, RT-qPCR showing expression changes of genes within the hub (*Zic2* and *Zic5*) and nearby genes (*Clybl* and *Pcca),* calculated as percentage relative to WT after normalization to *Hprt* expression**.** P-values were calculated using unpaired one-tailed t-test. Error bars indicate standard deviation from n=2 biological replicates.

Figure S5


a
PC2- (31.93% variance)
PC1- (38.43% variance)
KLF4 HiChIP PSC
KLF4 HiChIP day6
KLF4 HiChIP day3
H3K27ac HiChIP PSC
H3K27ac HiChIP MEF
b
Day3
PSC
Early binding
Early looping
(# loops=1732)
Early binding
Late looping
(# loops=741)
Late binding
Late looping
(# loops=2210)
KLF4
gene expression relative to PSC
MEF
D3
D6
D9
PSC
Early binding, early looping (# genes=458)
Early binding, late looping (# genes=148)
Late binding, late looping (# genes=393)
c
KLF4 PSC loops
26%
(6401)
74%
(18188)
H3K27ac dependent loops
H3K27ac independent loops
d
expression of genes within anchors
of KLF4-mediated loops in PSC (TPM)
***
5824
2472
H3K27ac
dep
H3K27ac
indep
f
KLF4 PSC loops
H3K27ac dep.
H3K27ac indep.
MED
NIPBL
O/S/N
YY1
KLF4
PRC2
PRC1
Dppa2
Mycn
Fbxo5
Dnmt3a
Pou5f1
Nodal
Bmp4
Hoxd9
Hoxd10
Serpine2
Fgf9
Bmi1
e
GO of genes in anchors of H3K27ac-dep KLF4 loops
stem cell population maintenance
Signaling pathways regulating pluripotency
of stem cells
RNA splicing
DNA repair
cell-cell adhesion
cell cycle
transcription, DNA-templated
-log10 (p_value)
GO of genes in anchors of H3K27ac-indep KLF4 loops
axon guidance
cell differentiation
nervous system development
embryonic skeletal system development
brain development
transcription, DNA-templated
anterior/posterior pattern specification
multicellular organism development
-log10 (p_value)

Supplementary Figure 5. a, PCA analysis of loops called as significant by H3K27ac and KLF4 HiChIP in different samples. **b**, Left: Chromatin loops that were detected by both KLF4 and H3k27ac HiChIP in PSCs were clustered based on the timing of KLF4 binding and looping during reprogramming. Right: Line plot showing expression changes of genes that belong to each of the indicated loop categories during reprogramming (median values are plotted relative to PSC). **c**, Pie chart showing the percentage of KLF4 PSC loops that were also detected by H3K27ac HiChIP in PSCs (H3K27ac-dependent) or not (H3K27ac-independent). **d**, Boxplot showing expression of genes within all anchors of KLF4-mediated loops that are either H3K27ac-dependent or independent. **e**, Gene ontology for genes within anchors of H3K27ac-dependent or -independent KLF4 loops. **f**, Proposed model for different categories of chromatin loops mediated by KLF4 and cofactors. Example genes are reported for each category.

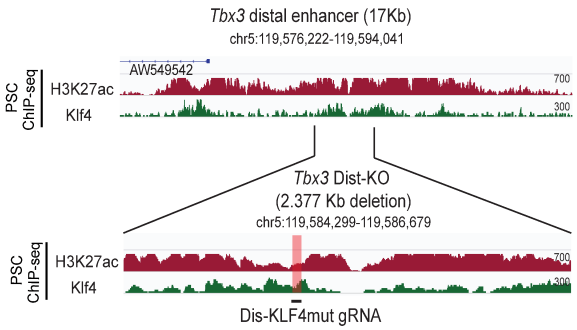
Figure S6


a
Klf4 KO1
Klf4 KO2
dox (hr)
0
48
0
48
KLF4
ACTIN
b
Klf2
relative expression normalized to hprt
3.5
3
2.5
2
1.5
1
0.5
0
***
***
dox (hrs)
0
48
0
48
Klf4 KO1
Klf4 KO2
Klf5
relative expression normalized to hprt
120
100
80
60
40
20
0
***
***
dox (hrs)
0
48
0
48
Klf4 KO1
Klf4 KO2
c
Klf2/4/5 TKO
dox (hr)
0
24
KLF4
KLF5
KLF2
ESRRB
NANOG
ACTIN
d
connectivity of hubs (n=348)
connectivity of SE (n=231)
connectivity of TE (n=8563)
number of loops per anchor
***
***
**
WT
TKO
WT
TKO
WT
TKO

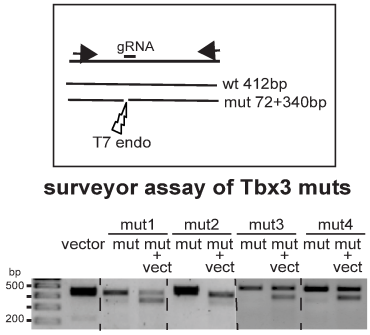
Supplementary Figure 6. a, Western blot analysis showing KLF4 protein levels before (0) and after (48hr) dox-induction in two ESC clones that harbor dox-inducible CRISPR-Cas9 and gRNAs that target the *Klf4* gene (KLF4 KO1 and KLF4 KO2). **b**, RT-qPCR showing elevated levels of *Klf2* and *Klf5* genes in dox-induced KLF4 KO ESCs. **c**, Western blot showing levels of indicated proteins in a clonal population of ESCs containing an inducible CRISPR-Cas9 construct and gRNAs that target the *Klf2*, *Klf5* and *Klf4* genes. Cells were either untreated (0, wild type or WT cells) or treated with dox for 24 hours (triple knock-out or TKO). **d**, Boxplot showing the connectivity of H3K27ac HiChIP anchors that contain hubs, supoerenhancers (SE) or typical enhancers (TE) in WT or TKO ESCs. Asterisks indicate significance as calculated using Wilcoxon rank sum test.

Figure S7

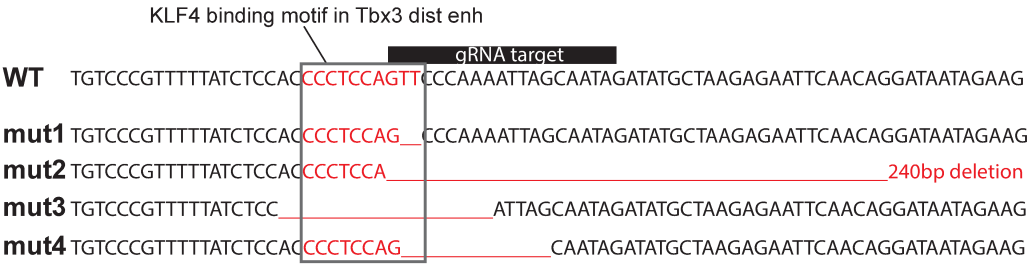
a



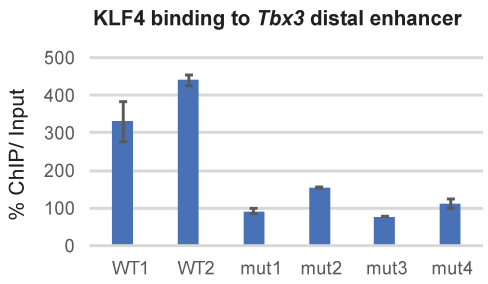
b

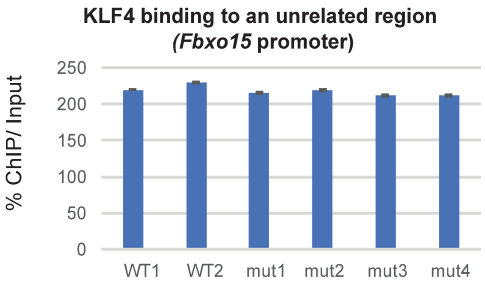


c



d





Supplementary Figure 7. a, IGV tracks of H3K27ac and KLF4 ChIP-seq in PSCs showing the whole *Tbx3* distal enhancer (top), the region that was deleted by CRISPR/Cas9 (Dist-KO, bottom, see Fig.4f) and the location of the gRNA used to mutate a specific KLF4 binding motif (Dis-KLF4mut gRNA). **b**, Genotyping strategy of the surveyor assay used to detect mutation/indel at the target KLF4 binding site within the distal *Tbx3* enhancer (Dis-KLF4mut*)*. The results for 4 homozygously mutated clones (mut1-4) are shown. **c**, Sequencing results of the four Mut clones compared to the wild type (WT). **d**, ChIP-qPCR showing the relative levels of KLF4 binding to *Tbx3* distal enhancer in two WT clones and four Mut clones (left panel). Values show percentage of ChIP signal over input. As control, binding of KLF4 to an unaffected region (*Fbxo15* promoter) was tested (right panel).