KLF4 binding during reprogramming is involved in 3D architectural rewiring and transcriptional regulation of enhancer hubs

Dafne Campigli Di Giammartino[1]*, Yiyuan Liu[1]*, Andreas Kloetgen[2]*, Alexander Polyzos[1], Daleum Kim[1], Matthias Stadtfeld[3], Aristotelis Tsirigos[2,4,5]# 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]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

[4]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

[5]Applied Bioinformatics Laboratories, NYU School of Medicine, New York, NY, 10016, 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 in somatic cell reprogramming to pluripotent stem cells (PSCs) by OCT4, KLF4, SOX2 and cMYC. How TFs orchestrate the complex molecular changes around their targets in a temporal manner remains largely elusive. Here, using KLF4 as a paradigm, we provide the first TF-centric view of chromatin reorganization during reprogramming and its association to enhancer rewiring and gene regulation. We captured the enhancer connectomes in fibroblasts and PSCs by H3K27ac HiChIP and identified complex 3D enhancer hubs that were strongly correlated with cell-type specific gene expression and coregulation. KLF4-centric conformational analysis at different reprogramming stages revealed that KLF4 is involved in the dissociation of fibroblast-specific and the establishment of PSC-specific enhancer loops concomitantly with repression or activation of linked genes. Moreover, KLF4 occupancy was significantly enriched within highly connected enhancers, suggesting a role in the formation and regulation of complex enhancer hubs. Indeed, disruption of a single KLF4 binding site from a PSC-specific enhancer was sufficient to reduce expression of multiple genes within the enhancer hub, partly by impairing long-range contact. Our study provides an integrative view of the intricate activities of a master regulator 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 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 (TF), known as master regulators of cell identity[1,2]. Although, the ability of master regulators to maintain or change cell identity is well appreciated, 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]. Distinct or synergistic roles of the reprogramming TFs as well as specific direct and indirect mechanisms for coordinating these molecular changes have been proposed[6-10]. 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[11-16]. However, the principles of chromatin reorganization during iPSC generation, the association with enhancer and gene activity and the involvement of TFs in these processes have only started to be explored.

75 The potential role of reprogramming TFs in 3D chromatin organization is
76 currently based mostly on computational analyses of 4C or Hi-C datasets, which
77 reveal a strong enrichment of OKS binding around long-range interactions in
78 PSCs or during reprogramming[11-13,15]. The architectural function of KLF4 is
79 further supported by experimental evidence, as it has been shown that KLF4
80 depletion abrogates specific loops around target loci in the context of PSCs [14],
81 and also in human epidermal keratinocytes[17], while depletion of the related factor
82 KLF1 disrupts selected long-range interactions in the context of erythropoiesis
83 [18,19]. These findings establish a link between reprogramming TF binding and
84 chromatin architecture and suggest that OKS, and particularly KLF4, may
85 actively orchestrate long-range chromatin interactions during reprogramming in
86 order to establish and maintain the pluripotent transcription program. To test this
87 possibility in a direct, genome-wide and dynamic manner during the course of
88 reprogramming, we used KLF4 as a paradigm to capture the TF-centric
89 topological reorganization and determine the relationships with epigenetic and
90 transcriptional changes. To do so, we used a well-characterized reprogramming
91 system[20,21] to apply genome-wide assays that map KLF4 binding (ChIP-seq),
92 chromatin accessibility (ATAC-seq), enhancer and gene activity (H3K27ac ChIP-
93 seq and RNA-seq), enhancer connectivity (H3K27ac HiChIP) as well as KLF4-
94 centric chromatin looping (KLF4 HiChIP) at different stages during acquisition of
95 pluripotency (Fig.1a top). Integrative analysis of our results generated a
96 reference map of stage-specific chromatin changes around KLF4 bound loci and
97 established strong links with enhancer rewiring and concordant transcriptional
98 changes. Genetic manipulation of KLF4 binding from a PSC enhancer further

supported the ability of KLF4 to function both as a transcriptional regulator and a chromatin organizer. In conclusion, our study offers novel insights into the intricate roles of a master regulator during cell fate transition.

Results

KLF4 binding induces chromatin opening and precedes enhancer and gene activation

To determine the molecular changes around KLF4 targets during iPSC formation, we first mapped its genome-wide binding at different stages of reprogramming using “reprogrammable” mouse embryonic fibroblasts (MEFs) (Rosa26-M2rtTA/Col1a1-OKSM) induced with doxycycline (dox)[20,21] in the presence of ascorbic acid. Under these conditions the resulting iPSCs are molecularly and functionally indistinguishable from embryonic stem cells[22,23] and, thus, the term PSCs will be used throughout the text to describe either cell type. For our earliest time point, we collected bulk populations on day 3 after dox treatment, whereas at later stages, on day 6 and day 9, we sorted SSEA1 positive cells to enrich for cells that are on the right trajectory towards induced pluripotency[21,24] (Fig.1a bottom and Supplementary Fig.1a). Finally, we used PSCs as a reference point for established pluripotency. ChIP-seq analysis showed a highly dynamic pattern of KLF4 occupancy during reprogramming with two major categories of binding sites: (i) transiently detected in reprogramming intermediate populations, but absent in PSCs (Transient KLF4 targets) and (ii) PSC targets, which represent the actual KLF4 binding repertoire once stem cell identity is acquired (Fig.1b and Supplementary Table 1). Among the PSC KLF4 binding sites, 30% were already

present on day 3 (Early KLF4 targets), while the rest were either gradually established during reprogramming (Mid KLF4 targets) or detected only in established PSCs (Late KLF4 targets). To gain insights into the nature and potential function of each category of KLF4 targets we performed genomic annotation based on their chromatin state classification introduced by Chronis et al [6] as well as Gene Ontology (GO) analysis using the GREAT tool [25] (Fig.1c). Early KLF4-targets mostly enriched for promoters of genes involved in regulation of metabolic processes and cell junction organization, in agreement with the previously reported early role of KLF4 in regulating these processes[21]. On the other hand, Mid and Late KLF4 targets included an increasing number of pluripotency-associated enhancers and enriched for stem cell maintenance genes, including many master regulators of pluripotency, such as *Sox2*, *Nanog*, *Esrrb* and *Klf4*. Finally, Transient KLF4 targets enriched for transient and partial-iPSC enhancers of genes involved in negative regulation of cell cycle, apoptosis and various signaling pathways associated with differentiation, such as TGF-beta signaling (Fig.1c). The different nature of stage-specific KLF4 binding sites was also supported by motif analysis, which revealed distinct classes of TFs enriched among Early (e.g. cMYC and YY1), Mid (e.g.ELF5 and RFX), Late (e.g. ESRRB and NR5A2) and Transient (e.g.FOSL1 and JUNB) KLF4 genomic targets (Supplementary Fig.1b).

The differential ability of KLF4 to bind Early or Late targets prompted us to investigate the chromatin accessibility in MEFs and its dynamic changes during reprogramming (Fig.1d and Supplemental Fig.1c). Integration of ATAC-seq and KLF4 ChIP-seq datasets revealed that ~60% of the Early KLF4 binding sites

147 were already open in MEFs, suggesting that preexisting accessibility could partly
148 explain the differential binding behavior of KLF4 during reprogramming (Fig.1d
149 and Supplementary Fig.1d) . In contrast, the majority (>70%) of mid, late and
150 transient KLF4 targets were “closed” in MEFs and gained accessibility at later
151 reprogramming stages (Fig.1d and Supplementary Fig.1d) in concordance with
152 the kinetics of KLF4 binding. This result is in agreement with the proposed
153 pioneer activity of KLF4[7,26] and/or with synergistic binding with the other
154 reprogramming factors[6].

155 KLF4 has been proposed to function both as an activator and repressor of
156 gene expression[9,21]. To assess the impact of KLF4 binding on enhancer and
157 gene activity, we first performed ChIP-seq for H3K27 acetylation in MEFs, PSCs
158 and intermediate reprogramming stages to map potential changes in enhancer
159 usage during iPSC generation (Supplementary Fig.1e). Our analysis
160 demostrated an acquisition of H3K27 acetylation concurrently with or
161 subsequently to KLF4 binding (Fig.1e and Supplementary Fig.1f), suggesting that
162 KLF4 might predominantly function as an activator. In further support of this
163 notion, less than 5% of decommisioned MEF enhancers were targeted by KLF4,
164 whereas about 35% of the total acquired PSC enhancers and almost the entirety
165 of (so-called) superenhancers (SE)[27] were bound by KLF4 concomitantly or prior
166 to H3K27 acetylation (Fig.1f, Fig.1g and Supplementary Table 2). Moreover,
167 RNA-seq analysis (Supplementary Fig.1g and Supplementary Table 3) of genes
168 that are linked to Early, Mid, Late or Transient KLF4 ChIP-seq peaks showed a
169 strong trend for upregulation, rather than downregulation, at the respective
170 stages of reprogramming (Supplementary Fig.1h). Representative examples of

171 ChIP-seq and ATAC-seq tracks as well as RNA-seq results for each category of
172 KLF4 targets are shown in Figure 1h.
173 In conclusion, our data revealed a stage-specific KLF4 binding during
174 reprogramming, which progressively targets PSC-associated enhancers and
175 eliminates transient, PSC-unrelated occupancy. KLF4 binding appeared to
176 depend only minimally on preexisting chromatin accessibility, had strong effects
177 on chromatin opening and accordingly on enhancer and gene activation.
178
179 **Enhancers are extensively rewired during reprogramming and form**
180 **complex interactomes in established pluripotent stem cells.**
181 Previous studies utilizing targeted (4C-seq) or global (Hi-C) chromatin
182 conformation assays have demonstrated that chromatin topology around specific
183 genomic loci or in large scale, such as compartments and Topologically
184 Associated Domains (TADs), is largely cell-type specific and drastically
185 reorganized during reprogramming[11,13-16,28]. However, the small-scale, cell type-
186 specific regulatory loops, such as enhancer-promoter interactions, were
187 insufficiently captured in these studies due to technical limitations. Here, we
188 performed H3K27ac HiChIP[29] in MEFs and PSCs in order to generate high-
189 resolution contact maps around active enhancers and promoters and
190 characterize the degree of architectural reorganization during reprogramming.
191 After calling high-confidence interactions using HiC-bench tool[30] (see methods),
192 we further refined our set by considering only loops that overlapped with
193 H3K27ac ChIP-seq peaks in at least one anchor (Supplementary Fig.2a).
194 Differential looping analyses (pvalue<0.01. LogFC>2) revealed about 50,000

contacts that were enriched either in MEFs or in PSCs (Fig.2a, Supplementary Fig.2b and Supplementary Table 4). By applying stringent statistic criteria (pvalue>0.5, logFC<0.5) , we also identified a group of ~6,000 constant H3K27ac contacts that show similar interaction strength in both MEFs and PSCs. Enrichment analysis using the LOLA tool[31], revealed that MEF-specific loops were enriched for binding of somatic TFs, such as Runx2 and Junb, whereas PSC-specific loops were strongly associated with pluripotency TFs, including OKS and NANOG as well as transcriptional co-activators, such as CBP and Mediator subunits (Supplementary Fig.2c).

In agreement with previous publications[32], H3K27ac HiChIP detected a large number of Enhancer-Enhancer (E-E), Enhancer-Promoter (E-P) and Promoter-Promoter (P-P) contacts (Supplementary Fig.2d). Interestingly, we observed that genomic regions that were detected as anchors in PSC-specific H3K27ac loops were involved in significantly higher number of long-range contacts (higher connectivity) than the ones detected in MEF-specific and constant loops (Fig. 2b). Together, these results document the dramatic 3D genomic reorganization of enhancers between MEFs and PSCs and reveal that cell type-specific enhancer connectomes are characterized by different degrees of complexity.

3D-organized enhancer cliques and enhancer hubs are associated with cell-type specific gene expression and coregulation

The presence of complex enhancer-promoter interactions, such as “enhancer cliques” (multiple enhancers contacting the same promoter) or “enhancer hubs”

219 (multiple promoters communicating with the same enhancer) (Supplementary
220 Table 5) prompted us to examine their potential cell-type specific function in
221 regulating gene expression. Integration of RNA-seq data overall showed a
222 significant correlation of MEF-specific or PSC-specific H3K27ac loops with
223 increased expression of associated genes in the respective cell type (Fig. 2c and
224 Supplementary Fig.2e). Interestingly, the strength of this correlation and the
225 levels of cell type-specific expression were higher when we focused on genes
226 within enhancer cliques (at least two enhancer contacts) and increased with the
227 number of enhancer contacts per gene (Fig.2c). These results support that
228 multiple enhancers connected to the same gene in a cell-type specific manner
229 are more likely to ensure robust and tight transcriptional activation.

230 We then focused on enhancers that establish connections with multiple
231 gene promoters hypothesizing that they may coregulate these genes in the
232 context of 3D regulatory hubs (enhancer hubs). To test this hypothesis, we
233 selected all protein-coding genes that participate in such hubs (2 or more genes
234 contacting the same enhancer) and are expressed (>1TPM) in either MEFs or
235 PSCs and performed pair-wise comparisons among genes within hubs to
236 calculate the percentage of co-regulation or anti-regulation. For control groups
237 we used random gene pairs (from the same pool of genes as the test group)
238 either of similar linear distance with our test group (global random) or within the
239 same TADs[33](TAD-matched random). The results show a significant
240 overrepresentation of coregulated gene pairs within enhancer hubs compared to
241 our control groups (Fig.2d). This analysis revealed 217 gene pairs that reside
242 within PSC-enhancer hubs and become concordantly upregulated during

243 reprogramming (Supplementary Table 5). As an example, we found that the
244 previously described Zfp42 SE also contacts and potentially co-regulates the
245 distal *Triml2* gene (Fig.2e). *Triml2* , similarly to *Zfp42,* is activated specifically in
246 PSCs, suggesting a potential function in the maintenance of stem cell identity. A
247 more complex hub was revealed around a SE enhancer located 60kb
248 downstream of *Nanog,* which appeared to form PSC-specific contacts with
249 multiple pluripotency-associated genes, including the proximal (~25kb) Slc2a3
250 and the distal (~450kb) Phc1 (Supplementary Fig.2f). Therefore, our analyses
251 revealed that many PSC SE and typical enhancers (TE) that were previously
252 assigned to only one gene[27,34], formed hubs with multiple genes and not
253 necessarily the most proximal ones (enhancer skipping) (Supplementary Fig.2g)
254 in a PSC-specific manner. Importantly, Gene ontology analysis of the newly
255 discovered SE and TE target genes enriched for stem cell maintenance and cell
256 cycle categories, supporting their functional relevance (Supplementary Fig.2g).
257 To better understand the biological basis of low or high enhancer
258 connectivity, we ranked PSC H3K27ac anchors based on the number of contacts
259 they establish and looked for features that may contribute to their distinct looping
260 behaviour. Using available ChIP-seq datasets, we compared binding intensities
261 at regions with the lowest (only one contact, n=3,594) and highest connectivity
262 (equal or more than four contacts, n=1434) and found that binding of pluripotency
263 TFs, including KLF4, was significantly stronger ($p<10^{-16}$) in highly connected
264 enhancers, similarly to Mediator subunits Med1 and Med12 (Fig.2f). However,
265 well-described architectural proteins, such as CTCF and cohesin[35] or the recently
266 described enhancer-promoter mediator YY1[36,37], did not show significantly

267 increased binding within highly connected enhancers (Fig. 2f). The involvement
268 of KLF4 in PSC-specific H3K27ac interactions was further corroborated by the
269 fact that about 80% of the H3K27ac PSC-specific loops were occupied by KLF4
270 with almost 50% being Early or Mid KLF4 targets during reprogramming (Fig.2g
271 and Supplementary Fig.2h). These results establish a link between KLF4 binding
272 and enhancer-promoter looping in established PSCs and during reprogramming,
273 supporting a potential architectural function for KLF4 in addition to its
274 transcriptional and pioneering activity.
275
276 **Dynamic KLF4-centered chromatin reorganization during reprogramming**
277 **associates with enhancer rewiring and transcriptional changes of target**
278 **genes.**
279 To capture the KLF4-centric chromatin reorganization during iPSC formation, we
280 performed HiChIP in early (day 3) and mid (day 6) stages of reprogramming as
281 well as in PSC using KLF4 antibody (Supplementary Fig.3a). Principle
282 component analysis (PCA) on all high-confidence interactions distinguished
283 primarily KLF4 from H3K27ac HiChIP data (Supplementary Fig.3b) regardless of
284 cell type or stage, demonstrating the different nature of chromatin contacts that
285 each antibody captures. PCA2 and PCA3 showed a clear separation of the
286 different reprogramming stages, suggesting dynamic reorganization of KLF4-
287 bound contacts (Fig.3a). Differential looping analysis generated 4 clusters of
288 differential KLF4-centered interactions (Fig.3b and Supplementary Table 6): two
289 clusters of gained loops, detected either in mid or late reprogramming stages
290 and two clusters of lost loops detected only in early or mid stages. Integration of

291 KLF4 ChIP-seq data revealed a strong correlation of stage-specifc KLF4-looping
292 with the expected stage-specific KLF4 binding (Fig.3c). Specifically, the majority
293 of gained loops overlapped with mid and late KLF4 binding, while lost loops had
294 predominantly transient KLF4 occupancy. Interestingly, in about 25-30% of KLF4
295 loops, KLF4 binding was detected prior to loop formation or persisted after loop
296 loss (Fig.3c), suggesting that KLF4 binding may be causal but not always
297 sufficient for looping formation or detection.

298 To gain insights into the role and nature of the different KLF4-centered
299 loop clusters we investigated the expression changes of associated genes during
300 reprogramming. Fig.3d shows that the relative percentages of up- and down-
301 regulated genes are different among the distinct loop clusters. Lost KLF4
302 contacts mostly associate with gene repression, while gained KLF4 loops with
303 gene activation during reprogramming (Fig.3d). Furthermore, integration of
304 H3K27ac HiChIP data from MEFs and PSCs, showed that most of the gained
305 KLF4 loops were also detected in PSC-enriched H3K27ac HiChIP contacts, while
306 lost KLF4 loops mostly enriched for MEF enhancer contacts (Fig.3e). Examples
307 of such loops and genes are shown in Figure 3f, which illustrates the normalized
308 KLF4 and H3K27 HiChIP signal around *Mycn* and *Ets1* in the format of virtual 4C
309 along with the respective ChIP-seq tracks during reprogramming (Fig.3g). *Mycn*
310 is a pluripotency-associated gene activated during reprogramming (Fig.3h),
311 concomitantly with the establishment of KLF4-mediated loops with a distal
312 enhancer. In contrast, *Ets1* is downregulated during reprogramming (Fig.3h) in
313 concordance with the gradual loss of a pre-existing H3K27ac-loop in MEFs that
314 was abolished in PSCs (Fig. 3f). Importantly, comparison of published Hi-C

315 datasets with our H3K27ac HiChIP results validated the cell-type specific nature
316 of these loops (Supplementary Fig.3c). Moreover, independent 3C-qPCR
317 experiments validated that the *Mycn*-enhancer loop was gradually established
318 during reprogramming, while the *Ets1*-enhancer contact frequency was
319 significantly decreased by day 6 (Fig.3i). Together, these observations suggest
320 that stage-specific KLF4 binding is involved both in the formation/activation of
321 PSC enhancer loops and abrogation/repression of prexisting somatic enhancer
322 loops.
323
324 **KLF4 binding is involved in different types of chromatin loops**
325 The previous results establish strong links between KLF4-centered chromatin
326 reorganization during reprogramming with enhancer looping as well as gene
327 expression changes, but they do not explain all observed long-range contacts. As
328 shown in Fig.3d, a high percentage (~50%) of genes within dynamic KLF4 loops
329 did not significantly change their expression during reprogramming. In
330 accordance, a large proportion of both gained and lost KLF4-centered
331 interactions appeared to be either constitutive enhancer loops (similar H3K27ac
332 HiChIP signal in both MEF and PSC) or independent of H3K27ac looping
333 (Fig.3e). These data demonstrate that looping around KLF4-bound loci might not
334 always result in changes in enhancer and/or gene activity.
335 The fact that KLF4 participates in H3K27ac-dependent and independent
336 chromatin contacts prompted us to investigate the potentially distinct looping
337 mechanisms and architectural partners of KLF4 for each loop category. To do so,
338 we first clustered all PSC KLF4 binding sites based on their presence within

339 H3K27ac-dependent or independent loops (KLF4 loops that were also detected
340 or not in H3K27ac HiChIP) (Supplementary Table 7). To increase stringency,
341 KLF4 binding sites present in both anchors of a loop (bi-anchored sites) were
342 considered. As a control group, KLF4 target sites that were detected neither in
343 KLF4 nor in H3K27ac HiChIP contacts were used (non-anchored sites).
344 Importantly, KLF4 binding strength was similar in all categories (Supplementary
345 Fig.3d), suggesting that the differential looping probability around KLF4 target
346 sites may depend on the presence or absence of other factors. Enrichment
347 analysis using published ChIP-seq datasets showed that KLF4 binding sites
348 within H3K27ac-dependent loops enriched for active enhancer features such as
349 binding of pluripotency TFs (ESRRB, NANOG, SOX2 and POU5F1), YY1 as well
350 as POLR2A and Mediator subunits (Supplementary Fig.3e and Supplementary
351 Table 8). In addition, H3K27ac-dependent KLF4 loops were associated with
352 occupancy/activity of specific histone modulators such as BRD4, p300 and
353 KDM2B or KDM6B, suggesting a potential architectural function for such
354 complexes. On the other hand, H3K27ac-independent KLF4 anchors
355 preferentially enriched for binding of classic architectural factors, such as
356 Cohesin and CTCF (Supplementary Fig.3e), but also of PRC1 and PRC2
357 components, which have been reported to mediate looping among repressed or
358 bivalent genes in PSCs[13,38]. Given that many of the enriched factors directly
359 interact with KLF4 protein, these data suggest that KLF4 might establish
360 chromatin loops with distinct properties and functions by recruiting different
361 architectural cofactors.
362

363 **Disruption of KLF4 binding site from a pluripotency-associated enhancer**
364 **interferes with looping and decreases expression of multiple genes within**
365 **the hub.**
366 To examine the functional significance of KLF4 in long-range chromatin
367 interactions and associated gene expression, we decided to focus on the *Tbx3*
368 locus, which was detected to contact multiple enhancers specifically in PSCs by
369 both KLF4 and H3K27ac HiChIP and validated using published Hi-C data (Fig.4a
370 and Supplementary Fig.4a). 3C-qPCR experiments during reprogramming
371 confirmed the establishment of a contact between *Tbx3* promoter with a distal (-
372 90kb) enhancer at late stages of reprogramming and concomitantly with gene
373 activation (Fig.4b and 4c). To experimentally test the relevance of this putative
374 enhancer for *Tbx3* gene regulation, we utilized CRISPR/Cas9 technology to
375 genetically delete it in PSCs using a previous characterized proximal enhancer
376 as a reference[39] (Fig.4d, Supplementary Fig.4b and Supplementary Table 9). RT-
377 pPCR analysis of homozygous knock-out (KO) clones (Supplementary Fig.4c)
378 demostrated that the transcriptional levels of *Tbx3* were severly impaired upon
379 disruption of either enhancer (Dis-KO and Pro-KO), with the distal enhancer
380 showing a stronger effect (Fig.4e). Interestingly, the RNA levels of the
381 surrounding non-coding genes, *Gm1603* and *Aw549542,* which were part of the
382 same enhancer hub with *Tbx3* (Fig.4a and Supplementary Fig.4d), were also
383 reduced (Fig 4f). Importantly, mutation of just one KLF4 binding motif within the
384 distal enhancer (DisKLF4^{mut}) (Fig 4d and Supplementary Fig.4b and e), also
385 resulted in a significant downregulation of *Tbx3*, *Gm1603* and *Aw549542* RNA
386 levels (Fig. 4e), supporting a critical role of KLF4 binding in the regulatory activity

387 of this enhancer. In addition, and consistently with the transcriptional
388 downregulation, the long-range contact between *Tbx3* and the distal enhancer
389 was specifically and significantly reduced upon mutation of the KLF4 motif as
390 shown by 3C-qPCR (Fig.4f), while the interaction of *Tbx3* with the proximal
391 enhancer or a KLF4-independent contact in chromosome 7 remained unaffected
392 (Fig.4f). These results provide direct evidence for the role of KLF4 both as
393 transcriptional regulator and chromatin organizer within enhancer hubs.
394

395 **Discussion**

396 Here, we describe the dynamic binding of KLF4 and its effects on chromatin
397 accessibility, enhancer activity and connectivity, gene expression and 3D
398 chromatin organization during iPSC reprogramming. Our data demostrated that
399 KLF4 initial binding mostly occured in previously inaccessible chromatin regions
400 and coincided with chromatin opening. This binding behavior is likely the result of
401 its documented pioneering function[7,26] as well as cooperativity with other
402 reprogramming TFs[6]. KLF4 binding was predominantly linked to gene
403 upregulation rather than downregulation, suggesting that a previously reported
404 repressive activity of this TF[6,9,21] may be manifested at earlier timepoints[6] or by
405 indirect mechanisms, for example through activation of the suppressor SAP30,
406 as it has been recently shown in a different reprogramming system[40]. Finally,
407 KLF4 binding often occurred prior to enhancer and gene activation (early KLF4
408 binding-late H3K27 acetylation), suggesting a causality, but also the requirement
409 of additional factors, such as ESRRB or NANOG[6], which are reactivated only
410 later in reprogramming (not shown).

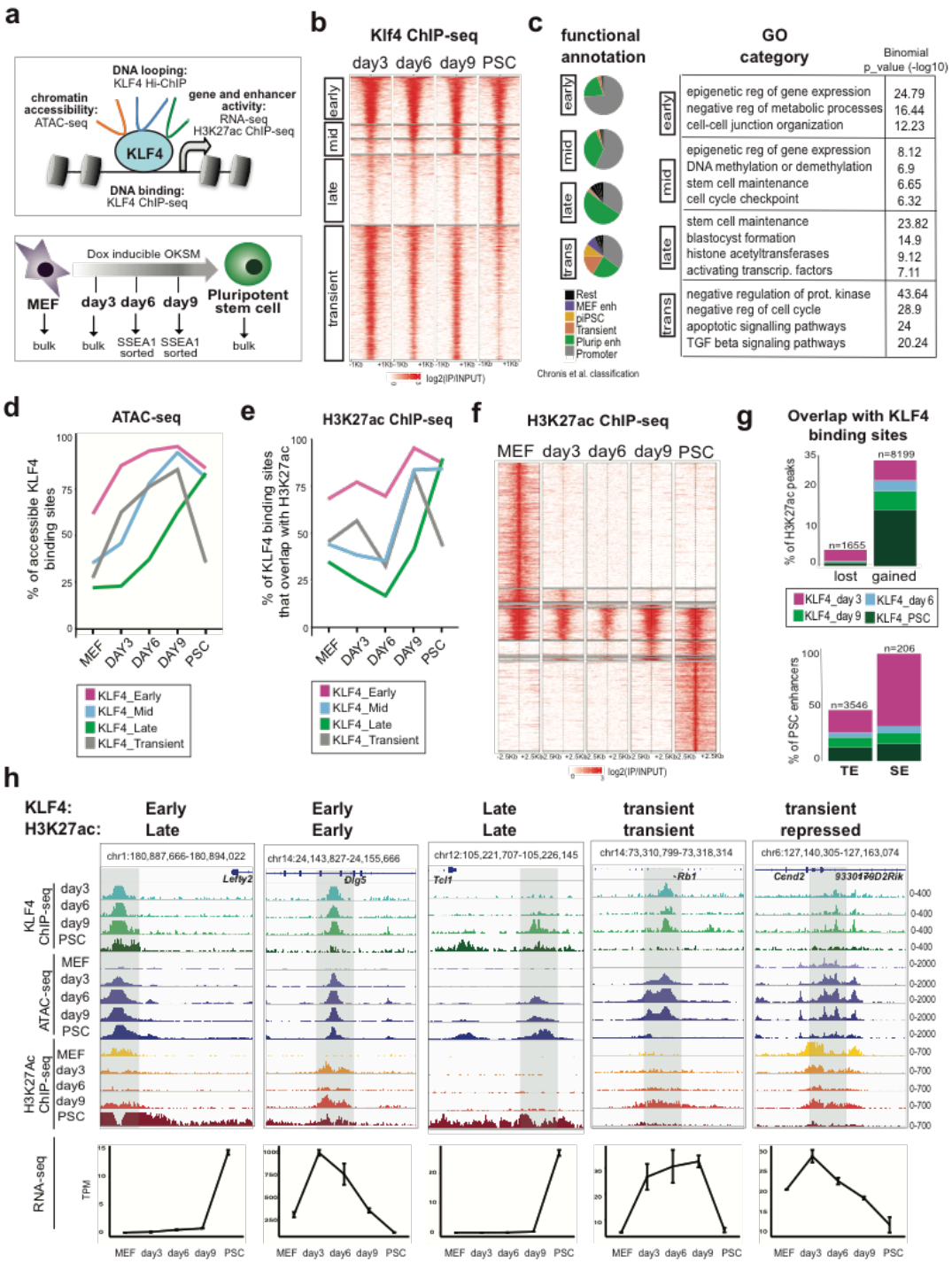
411 Previous studies utilizing 4C, 5C or Hi-C have reported that 3D genomic
412 architecture is drastically reorganized between somatic and pluripotent cells[11,13-
413 16]. However, the principles and mechanisms of such reorganization remain
414 largely elusive. A recent study performed Hi-C analysis at different stages of B
415 cell reprogramming and offered important insights into the kinetics and degree of
416 chromatin reorganization on the level of chromosome compartments and TADs[28].
417 However, only a minor fraction of cell-type specific loops were detected. Here,
418 we chose to apply H3K27ac HiChIP analysis, which has been recently shown to
419 have significantly higher discovery rate for cell-type specific loops, compared to
420 Hi-C and Capture Hi-C methods[29,32]. Indeed, our data revealed a dramatically
421 rewired enhancer connectome between MEFs and PSCs generating a reference
422 map of conformational changes, relevant for cell identity. Our analyses unraveled
423 cell-type specific H3K27ac interactomes that included simple and complex P-P,
424 E-E and E-P contacts. Interestingly, genes that communicated with multiple
425 enhancers (enhancer cliques[32]) in a cell-type specific manner appeared to be
426 more strongly upregulated in the respective cell-type, which supports the
427 possibility that genes critical for cell identity are under the control of more than
428 one enhancers ensuring increased specificity and robustness. In parallel, our
429 analysis identified a number of cell-type specific 3D enhancer hubs, including
430 many SE-hubs, which contained two or more coregulated genes. This provides
431 genome-wide evidence that complex 3D enhancer hubs may play a functional
432 role in coordinating gene regulation during cell fate transitions and may therefore
433 enable identification of novel candidate genes and enhancers critical for the
434 maintenace or establishment of specific cellular identities.

435 Encouraged by previous studies reporting potential architectural functions
436 for various KLF protein members[14,18,19], we went on to capture for the first time
437 the KLF4-centric chromatin reorganization around KLF4-binding sites during
438 iPSC formation. Integrative computational analysis established strong links
439 between KLF4-bound stage-specific contacts with cell-type specific enhancer
440 looping and concordant gene expression changes. Specifically, our analyses
441 showed that KLF4 binding associated with *de novo* formation of PSC-specific
442 enhancer loops during reprogramming promoting activation of linked genes. On
443 the other hand, transient KLF4 binding on preexisting enhancer loops in somatic
444 cells was linked to their gradual abrogation and silencing. This phenomenon of
445 TF "hijacking" on preestablished loops has been previously reported during
446 differentiation and reprogramming[28,41], but mostly as a trigger for looping
447 stablization and gene activation and less for loop destabilization and
448 transcriptional repression.
449 Finally, our results revealed that KLF4 binding occurs within distinct
450 categories of chromatin contacts: (i) loops enriched in active enhancer features,
451 such as H3K27ac, Mediator and pluripotency TFs, (ii) structural H3K27ac-
452 independent loops, which were established by strong architecural factors such as
453 Cohesin and CTCF and (iii) potentially repressive or paused loops that we
454 associated with PRC1/2 components. The role of KLF4 binding in the formation,
455 stabilization or functionality of these distinct loop categories remains to be shown.
456 In support of a critical role of KLF4 in maintaining enhancer-promoter
457 communication and activity, we showed that disruption of a single KLF4 binding

458 site from a newly identified PSC-specific enhancer, decreased expression of
459 multiple genes within the enhancer hub, partly by reducing looping frequency.
460 In conclusion, our study provided the first TF-centric chromatin
461 reorganization analysis during a cell fate transition and established mechanistic
462 and functional links with enhancer rewiring and gene regulation. The integrative
463 analysis of KLF4 occupancy and looping data with changes in chromatin
464 accessibility and enhancer and gene activity offered novel insights on the
465 multiple activities of a master regulator during somatic cell reprogramming. In
466 addition, our approach identified complex PSC-specific 3D enhancer hubs and
467 provided evidence for their role in gene coregulation during reprogramming.
468 Importantly, KLF4 binding was strongly enriched in high connectivity hubs,
469 suggesting a potential regulatory and architectural fuction for this TF, which was
470 further supported by genetic engineering experiments. Given the reported role of
471 KLF4 in multiple cell fate decisions, such as keratinocyte and monocyte
472 differentiation, and its context-dependent function as a tumor-suppressor or
473 oncogene[42], it will be interesting to identify and interrogate the significance of
474 KLF4-mediated 3D hubs in different cellular contexts. Future work will be critical
475 to (i) delineate whether chromatin looping around TF-binding sites occurs by
476 active tethering through recruited architectural cofactors or by “self-organization”
477 through liquid-phase separation as a result of local epigenetic alterations and (ii)
478 determine the functional significance of TF-mediated 3D chromatin hubs in cell
479 fate transitions.
480
481 **FIGURES AND FIGURE LEGENDS**

482

Figure 1



483

484 **Figure 1. Dynamic KLF4 binding during reprogramming and association**
485 **with chromatin accessibility and enhancer activity. (a)** Schematic illustration
486 of the experimental system and strategy. (**b)** Tornado plots of KLF4 ChIP-seq
487 signals at different reprogramming stages clustered in four different categories:
488 Early, Mid, Late and Transient KLF4 binding. ChIP-seq signals were centered
489 1kb upstream/downstream of peak centers. **(c)** Left: Pie charts of functional
490 classification of KLF4 Early, Mid, Late and Transient peaks (based on[6]) (piPSC,
491 partial iPSCs); Right: GREAT gene ontology analysis of Early, Mid, Late and
492 Transient KLF4 target sites. (**d-e)** Line plots showing the percentages of KLF4
493 Early, Mid, Late and Transient targets that overlapped with (d) ATAC-seq peaks
494 or (e) H3K27ac ChIP-seq peaks at each reprogramming stage (**f)** Tornado plot of
495 H3K27ac ChIP-seq signal at each reprogramming stage 2.5Kb upstream or
496 downstream of peak centers. **g)** Bar plots showing overlap of KLF4 binding with
497 either lost (MEF and d3) or gained (d9 and PSC) H3K27ac peaks (top) or with
498 typical PSC enhancers (TE) versus superenhancers (SE) as characterized by
499 [27](bottom). **h)** Examples of specific KLF4 targets with respective IGV tracks for
500 KLF4 ChIP-seq, ATAC-seq, H3K27ac ChIP-seq at each reprogramming stage.
501 Genomic coordinates are shown in parenthesis and the signal values on the right
502 side. The transcriptional changes of the depicted genes during reprogramming
503 are shown at the bottom, expressed as transcripts per million (TPM).

504

Figure 2

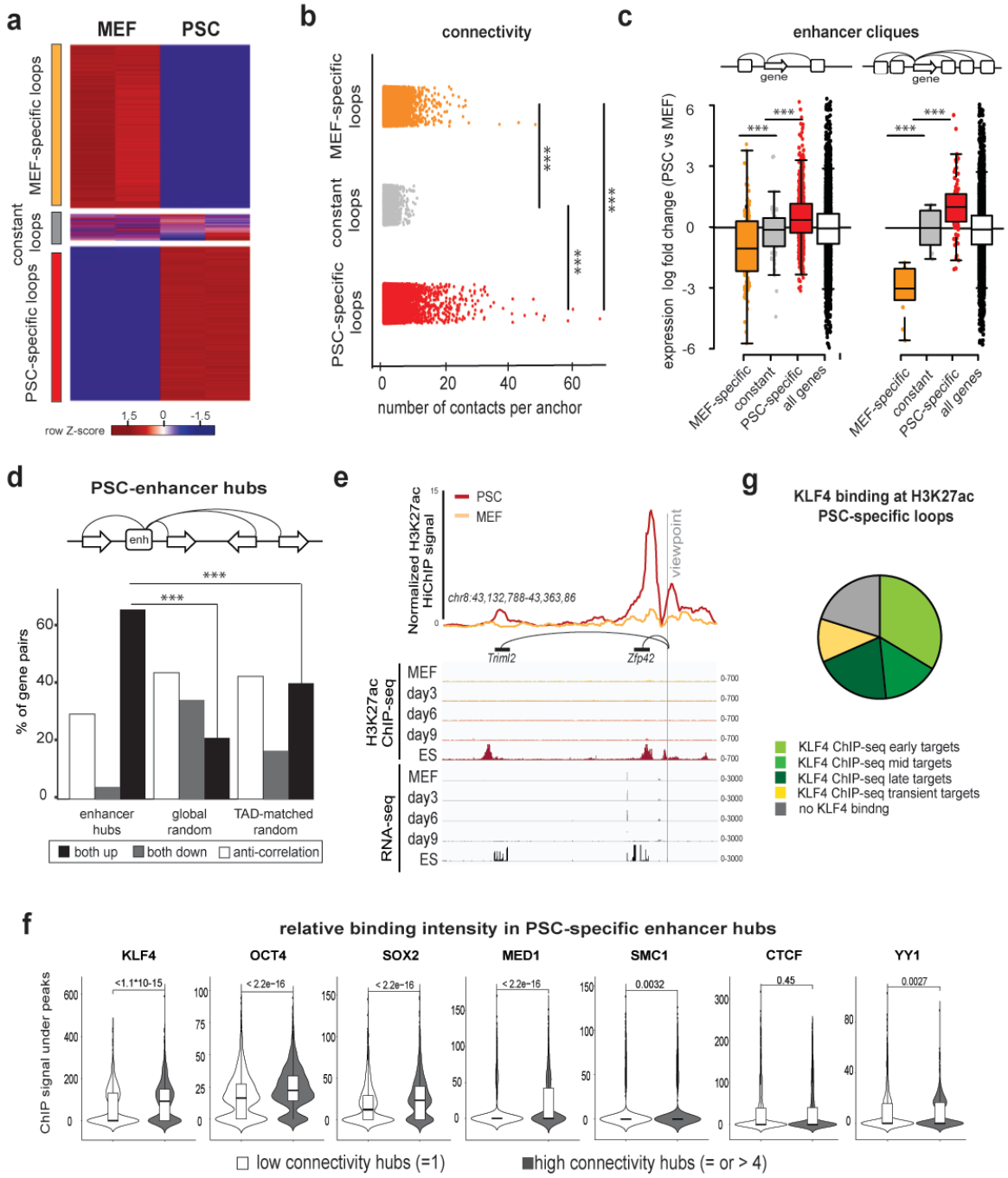


Figure 2. Detection and characterization of H3K27ac-associated loops in MEFs and PSCs by H3K27ac HiChIP analysis. (a) Heatmap of differential loops detected by H3K27ac HiChIP in MEF versus PSC. Differential loops were

509 called by average logFC > 2 and p-value < 0.01, constant loops were called by
510 average logFC > -0.5 & < 0.5 and p-value > 0.5. Heatmap shows row-wise z-
511 score of normalized interaction strength per replicate. **(b)** Connectivity of
512 H3K27ac differential or constant loops measured as number of high-confidence
513 contacts around each H3K27Ac HiChIP fragment. Asterisks indicate significant
514 difference with p<0.001 **(c)** RNA expression changes between MEFs and PSCs
515 of genes that were involved in at least two (left) or at least five (right) MEF-
516 specific, PSC-specific or constant H3K27ac loops. All protein-coding genes were
517 used as control. Asterisks indicate significant difference (p<0.001) as calculated
518 by paired t-test. **(d)** Barplot indicating the percentages of gene pairs within
519 enhancer hubs (enhancers contacting 2 or more gene promoters) that become
520 transcriptionally co-regulated (both up or both down) or anti-regulated (one up
521 one down) between MEFs and PSCs. Global Random or TAD-matched gene
522 pairs were used as controls (see also Methods). Genes that remain unchanged
523 (FC<1.5, p>0.5) were not considered in this analysis. Significance is indicated by
524 asterisks and was calculated by Fisher's exact test **(e)** Example of a newly
525 identified enhancer hub in PSCs. Normalized HiChIP signals around the
526 viewpoint are illustrated as virtual 4C plots. H3K27ac ChIP-seq and RNA-seq
527 IGV tracks for each reprogramming stage are shown below to highlight
528 concordance with H3K27ac HiChIP data and coordinated upregulation of target
529 genes. **(f)** Violin plots showing ChIP-seq intensity under peaks located within
530 PSC-specific enhancers with either low connectivity (only one contact, n=3594)
531 or high connectivity (more than 4 loops, n=1434). In addition to KLF4 (in-house
532 ChIP-seq data), we examined other pluripotency TFs as well as known

533 architectural factors using public ChIP-seq datasets in PSCs. Signal comparison
534 between low and high connectivity was done by Wilcoxon signed-rank test and
535 the p-values are reported. **(g)** Pie chart showing overlap of H3K27ac PSC-
536 specific loops with KLF4 Early, Mid, Late and Transient ChIP-seq peaks**.**

537

538

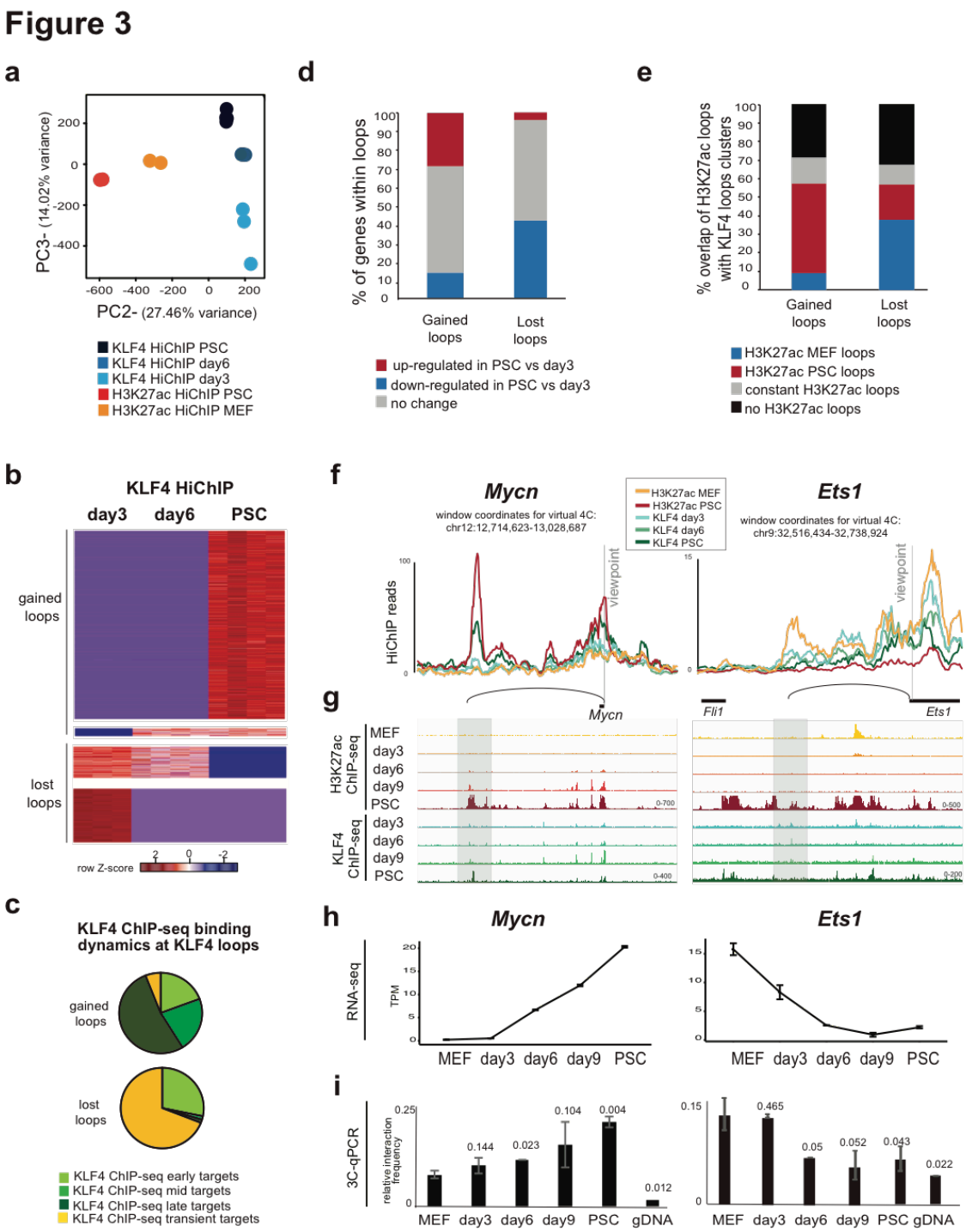


Figure 3. Capturing the KLF4-mediated chromatin reorganization during reprogramming and its association to enhancer and gene activity a) PCA analysis of high-confidence interactions of H3K27ac and KLF4 HiChIP samples

543 (see also Supplementary figures 3a and 3b). (**b)** Heatmap of differential KLF4
544 HiChIP analysis depicting 4 distinct clusters grouped into gained or lost loops.
545 Heatmap shows row-wise z-score of the normalized interaction strength for KLF4
546 HiChIP replicates. (**c)** Pie charts representing overlap between gained or lost
547 KLF4 HiChIP loops with Early, Mid, Late and Transient KLF4 ChIP-seq peaks.
548 (**d)** Stacked barplot indicating the percentages of genes within gained or lost
549 KLF4 loops that become upregulated or downregulated (logFC > 1.5, FDR < 0.01
550 in PSC vs day3) during reprogramming. **(e)** Stacked barplot showing the
551 percentage of gained or lost KLF4 loops that were also detected by H3K27ac
552 HiChIP analysis and were enriched in either MEFs or in PSC or in both (constant
553 loops). Note that about 25-30% of the loops were not detected by H3K27ac
554 (H3K27ac-independent KLF4 loops) (see also Supplementary figure 3d,e). **(f)**
555 Examples of KLF4-mediated loops that are either acquired (*Mycn*) or lost (*Ets1*)
556 during reprogramming. Normalized KLF4 and H3K27ac HiChIP signals are
557 illustrated in a virtual 4C format around the viewpoints (*Mycn* and *Ets1* gene
558 promoters). The respective ChIP-seq IGV tracks are shown in **(g)**, while the RNA
559 changes during reprogramming expressed as transcripts per million (TPM) are
560 shown in **(h)**. **(i)** 3C-qPCR validation of the relative interaction frequency between
561 Mycn and Ets1 promoters and the respective distal enhancers as depicted in (f)
562 at different reprogramming stages. Unpaired one-tailed Student's t-test was used
563 to determine significance relative to the MEF time point (value indicated on the
564 top of each bar). Error bars indicate standard deviation in *n=2* biological
565 replicates.

Figure 4

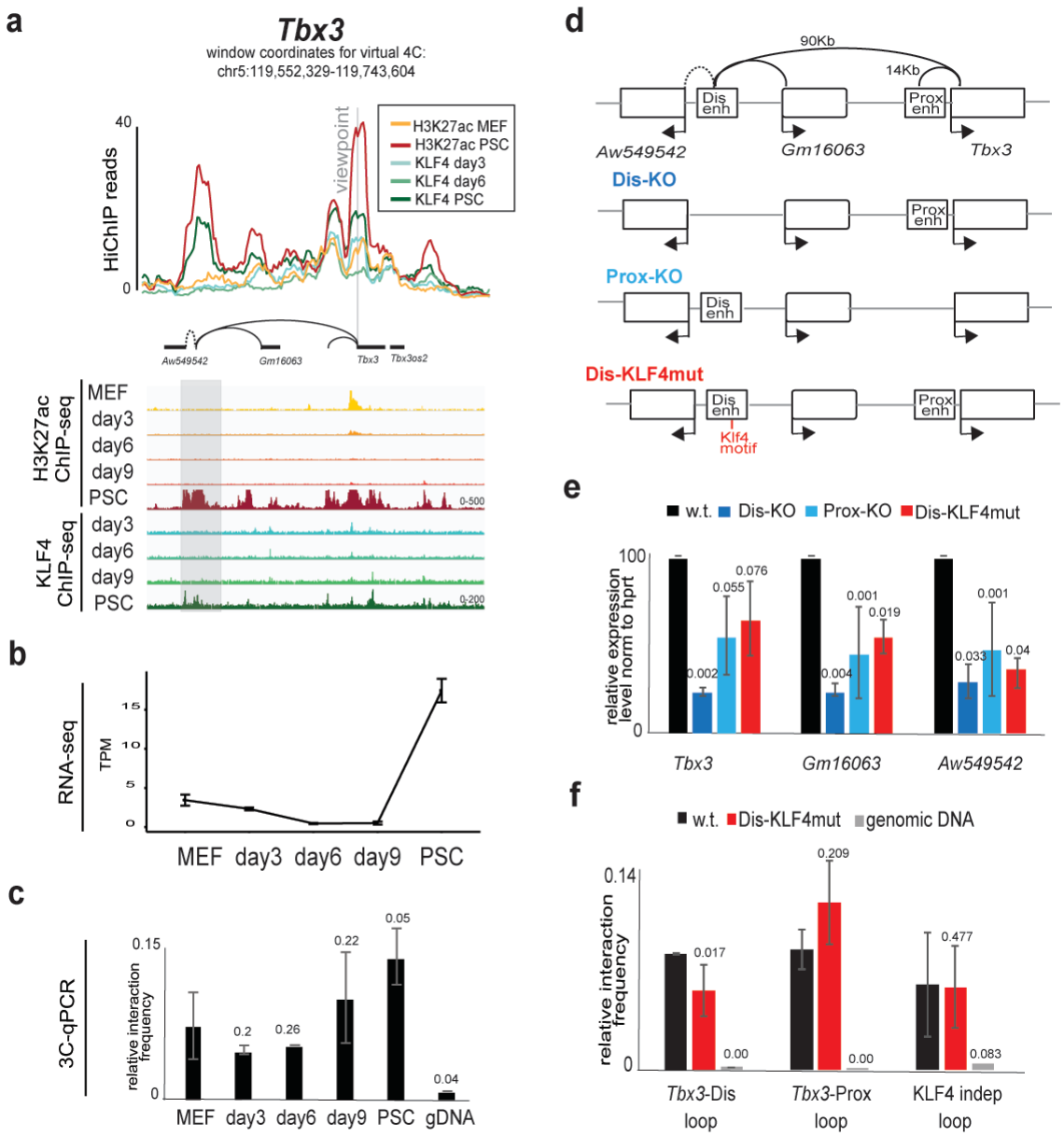


Figure 4. Disruption of KLF4 binding site from a distal *Tbx3* enhancer abrogates looping and decreases *Tbx3* expression in PSCs. (a) Normalized KLF4 and H3K27ac HiChIP signals are illustrated as virtual 4C line plots around the *Tbx3* promoter. The respective ChIP-seq IGV tracks are shown at the bottom. **(b)** *Tbx3* RNA changes during reprogramming expressed as transcripts per

572 million (TPM) **(c)** 3C-qPCR validation of the relative interaction frequency
573 between the *Tbx3* promoter and the most distal enhancer as depicted in (a) at
574 different reprogramming stages. Unpaired one-tailed Student's t-test was used to
575 determine significance relative to MEF (value indicated on the top of each bar).
576 Error bars indicate standard deviation in *n=2* biological replicates. **(d)** Schematic
577 diagram showing CRISPR-mediated manipulations of the *Tbx3* distal (Dis) and
578 proximal (Prox) enhancers. KO: knockout, KLF4mut: Mutation or indel of a KLF4
579 binding motif within the distal enhancer. **(e)** RT-qPCR showing expression
580 changes of *Tbx3* and nearby genes in either wild-type (w.t) PSCs or in CRISPR-
581 Cas9 engineered PSC carrying homozygous deletions of the distal *Tbx3*
582 enhancer (Dis-KO), the proximal (Prox-KO) or the KLF4mut. Unpaired one-tailed
583 Student's t-test was used to determine significance relative to w.t. (value
584 indicated on the top of each bar). Error bars indicate standard deviation from n=2
585 biological replicates **(f)** 3C-qPCR analysis to detect changes in looping formation
586 between the *Tbx3* promoter and either the distal or proximal enhancer following
587 mutation of the KLF4 binding site within the distal enhancer. Unpaired one-tailed
588 Student's t-test was used to determine significance relative to w.t. (value
589 indicated on the top of each bar). Error bars indicate standard deviation from n=2
590 biological replicates.

591

592 **METHODS**

593 **Cell lines, culture conditions and reprogramming experiments**
594 Mouse ES V6.5 were cultured on irradiated feeder cells in KO-DMEM media
595 (Invitrogen) supplemented with 15% heat-inactivated fetal bovine serum,
596 GlutaMAX, penicillin-streptomycin, non-essential amino acids, β-mercaptoethanol

597 and 1000 U/ml LIF, with or without the presence of 2i (1uM MEKinhibitor
598 (Stemgent 04-0006) and 3uM GSK3 inhibitor (Stemgent 04-0004)).
599 Mouse embryonic fibroblasts (MEFs) were isolated from a "reprogrammable"
600 mouse harboring a polycystronic OKSM cassette in the Col1a1 locus and M2rtTA
601 in the Rosa26 locus[20]. Cells were reprogrammed in the presence of 1ug/ml
602 doxycycline and 50ug/ml ascorbic acid and cultured in ES medium as described
603 above. Cells were collected at the indicated time points.

604 **MACS and FACS**
605 For isolating the SSEA1 positive cells from reprogramming intermediates at day6
606 and day9 we used magnetic microbeads conjugated to anti-SSEA1 antibody
607 (MACS Miltenyi Biotec #130-094-530) as per manufacture instructions. SSEA
608 positive and negative fractions were then stained for FACS analysis with an anti-
609 Thy1 antibody conjugated to pacific blue fluorophore (ebioscience # 48-0902-82)
610 and anti-SSEA antibody conjugated to APC fluorophore (biolegend #125608).
611
612 **Generation, selection and validation of KO**
613 gRNAs were cloned into the px458 vector using the BbsI restriction enzyme[43].
614 0.3 million ESC cells (V6.5) were transfected using 2ug of Left-*Tbx3*-plasmid and
615 2ug of Right-Tbx3-plasmid (for *Tbx3* enhancer deletions) or 4ug of *Tbx3*-
616 KLF4mut-vector (mutation of KLF4 binding site within *Tbx3* distal enhancer).
617 DNA was pre-mixed with 50ul media with no additions, and in a separate tube
618 10ul of Lipofectamine 2000 (Invitrogen #11668019) was pre-mixed with 50ul
619 media with no addition. After 5 minutes the two tubes were combined and
620 incubated at room temperature for 20 more minutes. Cells were then added to
621 the solution and plated on a gelatinized 12 well. 48hrs post-transfection, GFP
622 positive single cells were sorted by FACS into 96 well plates. Genotyping of the
623 single cell colonies was performed using a three-primer strategy (for deletions) or
624 by surveyor with T7 (for in-del mutation). Two colonies with homozygous
625 mutations (or w.t. colonies as control) were expanded and used for RT-qPCR
626 and 3C experiments. All gRNA and primers are described in Supplementary
627 Table 9.
628
629 **3C-qPCR**
630 For each sample 1 to 2 million cells were lysed in 300ul of lysis buffer (10mM
631 Tris-HCl pH 8.0, 10mM NaCl, 0.2% Igepal CA630 with protease inhibitors) and
632 incubated on ice for 20 mins. Cells were centrifuged 2500g for 5min at 4°C and
633 pellet washed once in lysis buffer. Pellets were resuspended in 50ul of 0.5% SDS
634 and incubated at 65°C for 10 mins. 145ul of water and 25ul of 10% triton were
635 added to the samples and incubated 15mins at 37°C. 100 Units of MboI
636 restriction enzyme and 25ul of NEB buffer 2 were added and incubated over
637 night at 37°C with rotation. Next day the enzyme was inactivated at 65°C for 20
638 mins. The ligation reaction was carried out over night at 16°C by adding 120ul of
639 NEB T4 ligase buffer, 100ul of 10% Triton, 6ul of 20mg/l BSA, 100ul of 10mM
640 ATP and 5ul of T4 ligase (NEB #M0202).The day after, 50ul of 20mg/ml
641 proteinase K and 120ul of 10% SDS were added and the samples were
642 incubated over night at 65°C. Lastly, 10ul of 10mg/ml RNAse was added and

643 samples incubated 1 hour at 37°C. Following phenol chloroform purification the
644 DNA was precipitated using 1.6 volumes of 100% ethanol and 0.1 volume of 3M
645 sodium acetate. After incubation at -80C for 1 hour samples were spun for
646 15mins at 4°C at 16000rpm. Pellets were washed twice with 70% ethanol and
647 dissolved in 100ul of 10mM Tris pH8. Qbit was used to measure sample
648 concentrations and 100ng of material was used to amplify the desired regions by
649 qPCR. All primer sequences can be found in Supplementary table 9.
650
651 **ChIP-seq**
652 ChIP-seq was performend as previously described[44]. Specifically, cells were
653 crosslinked in 1% formaldehyde at RT for 10 minutes and quenched with 125mM
654 glycine for 5 mins at RT. 50 million cells were used for KLF4 ChIPs and 10
655 million for H3K27acetylation ChIP. Cell pellets were washed twice in PBS and
656 resuspended in 400ul lysis buffer (10mM Tris pH8, 1mM EDTA, 0.5% SDS) per
657 20 million cells. Cells were sonicated in a bioruptor device (30 cycles 30sec
658 on/off, high setting) and spun down 10 minutes at 4°C at maximum speed.
659 Supernatants were diluted 5 times with dilution buffer (0.01%SDS, 1.1%
660 triton,1.2mM EDTA,16.7mM Tris pH8, 167mM NaCl) and incubated with the
661 respective antibody (2-3ug/10M cells) (KLF4 R&D #3158, H3K27ac ab4729)
662 O/N with rotation at 4°C. Next day, protein G Dynabeads (ThermoScientific)
663 preblocked with BSA protein (100ng per 10ul Dynabeads) were added (10ul
664 blocked Dynabeads per 10 million cells) and incubated for 2-3 hours at 4°C.
665 Beads were immobilized on a magnet and washed twice in low salt buffer (0.1%
666 SDS,1% triton, 2mM EDTA, 150mM NaCl, 20mM Tris pH8), twice in high salt
667 buffer (0.1% SDS,1% triton, 2mM EDTA, 500mM NaCl, 20mM Tris pH8), twice in
668 LiCl buffer (0.25M LiCl, 1% NP40, 1% deoxycholic acid (sodium salt), 1mM
669 EDTA, 10mM Tris pH8) and once in TE. DNA was then eluted from the beads by
670 incubating with 150ul elution buffer (1% SDS, 100mM NaHCO3) for 20 minutes
671 at 65°C (vortexing every 10min). Supernatants were collected and reverse-
672 crosslinked by incubation at 65°C O/N in presence of proteinase K. After RNase
673 A treatment for 1hr at 37°C, DNA was purified using the minElute kit (Qiagen). 6-
674 10ng of immunoprecipitated material was used for ChIP-seq library preparation
675 using the KAPA Hyper prep kit (KAPA Biosystems). Libraries were sequenced on
676 an Illumina HiSeq 2500 platform on SE50 mode.
677
678 **ATAC-seq**
679 ATAC-seq was performed as previously described[45]. In brief, a total of 50,000
680 cells were washed once with 50 µL of cold PBS and resuspended in 50 µL lysis
681 buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.2% (v/v) IGEPAL
682 CA-630). The suspension of nuclei was then centrifuged for 10 min at 800 g at
683 4°C, followed by the addition of 50 µL transposition reaction mix (25 µL TD buffer,
684 2.5 µL Tn5 transposase and 22.5 µL nuclease-free H_2O) using reagents from the
685 Nextera DNA library Preparation Kit (Illumina #FC-121-103). Samples were then
686 incubated at 37°C for 30min. DNA was isolated using a ZYMO Kit (#D4014).
687 ATAC-seq libraries were first subjected to 5 cycles of pre-amplification. To
688 determine the suitable number of cycles required for the second round of PCR
689 the library was assessed by quantitative PCR as described in Buenrostro et al

690 and the library was then PCR amplified for the appropriate number of cycles
691 using Nextera primers. Samples were subject to a dual size selection (0.55x-
692 1.5x) using SPRI beads (Beckman Coulter #B23317). Finally, the ATAC libraries
693 were sequenced on a HiSeq 2500 platform on PE50 mode.
694
695 **RNA-seq**
696 Total RNA was prepared with TRIZOL (Life technologies #15596018) following
697 manufacturer's instructions. Libraries were generated by the Weill Cornell
698 Genomics core facility using the Illumina TruSeq stranded mRNA library
699 preparation kit (#20020594) and sequenced on an Illumina HiSeq4000 platform
700 on SE50 mode.
701
702 **HiChIP**
703 HiChIPs were performed as previously described[29] with some modifications. In
704 brief, up to 15 million crosslinked cells (for KLF4 HiChIPs two samples of 15
705 million cells were combined at the end, for each sample replicate) were
706 resuspended in 500 µL of ice-cold Hi-C lysis buffer (10 mM Tris-HCl pH 7.5, 10
707 mM NaCl, 0.2% NP-40, 1× protease inhibitors) and rotated at 4°C for 30 min.
708 Nuclei were pelleted and washed once with 500 µL of ice-cold Hi-C lysis buffer.
709 Pellet was then resuspended in 100 µL of 0.5% SDS and incubated at 62°C for
710 10 min. 285 µL of water and 50 µL of 10% Triton X-100 were added, and
711 samples were rotated at 37°C for 15 min. 50 µL of NEB Buffer 2 and 15 µL of
712 25 U/µL MboI restriction enzyme (NEB, R0147) were then added, and sample
713 was rotated at 37°C for 2 h. MboI was then heat inactivated at 62°C for 20 min.
714 We added 52 µL of incorporation master mix: 37.5 µL of 0.4 mM biotin–dATP
715 (Thermo Fisher, 19524016); 4.5 µL of a dCTP, dGTP, and dTTP mix at 10 mM
716 each; and 10 µL of 5 U/µL DNA Polymerase I, Large (Klenow) Fragment (NEB,
717 M0210). The reactions were then rotated at 37°C for 1 h. 948 µL of ligation
718 master mix was then added: 150 µL of 10× NEB T4 DNA ligase buffer with 10
719 mM ATP (NEB, B0202), 125 µL of 10% Triton X-100, 3 µL of 50 mg/mL BSA
720 (Thermo Fisher, AM2616), 10 µL of 400 U/µL T4 DNA Ligase (NEB, M0202),
721 and 660 µL of water. The reactions were then rotated at room temperature for
722 4 h. After proximity ligation, the nuclei were pelleted and the supernatant was
723 removed. The nuclear pellet was brought up to 880 µL in Nuclear Lysis Buffer
724 (50 mM Tris-HCl pH 7.5, 10 mM EDTA, 0.5% SDS, 1× Roche protease
725 inhibitors, 11697498001), and sonicated with a Bioruptor 300 (Diagenode) for 8
726 cycles of 30sec each, on a medium setting. Clarified samples were transferred
727 to Eppendorf tubes and diluted five times with ChIP Dilution Buffer (0.01% SDS,
728 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH 7.5, 167 mM NaCl).
729 Cells were precleared with 30 µL of Protein G dynabeads (Life technology
730 #10004D) in rotation at 4°C for 1 h. Supernatants were transferred into fresh
731 tubes and antibody was added (8 µg of KLF4 antibody or 3ug H3K27Ac
732 antibody for 15 million cells) and incubated overnight at 4°C. The next day 30
733 µL of Protein G dynabeads were added to samples and rotated at 4°C for 2 h.
734 After bead capture, beads were washed three times each with low-salt wash
735 buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 7.5, 150
736 mM NaCl), high-salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20

737 mM Tris-HCl pH 7.5, 500 mM NaCl), and LiCl wash buffer (10 mM Tris-HCl pH
738 7.5, 250 mM LiCl, 1% NP-40, 1% sodium deoxycholate, 1 mM EDTA, make
739 fresh). Samples were eluted with 150 µL of DNA elution buffer (50 mM sodium
740 bicarbonate pH 8.0, 1% SDS, freshly made) and incubated at 37°C for 30 min
741 with rotation. Supernatant was transferred to a fresh tube and elution repeated
742 with another 150 µL elution buffer. 5 µL of Proteinase K (20mg/ml) (Thermo
743 Fisher) were added to the 300 µL reaction and samples were incubated
744 overnight in 65°C. Samples were purified with DNA Clean and Concentrator
745 columns (Zymo Research) and eluted in 10 µL of water. Post-ChIP DNA was
746 quantified by Qubit (Thermo Fisher) to estimate the amount of Tn5 (Illumina)
747 needed to generate libraries at the correct size distribution (see below). 5 µL of
748 Streptavidin C-1 beads (Thermo Fisher) were washed with Tween Wash Buffer
749 (5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl, 0.05% Tween-20) then
750 resuspended in 10 µL of 2× biotin binding buffer (10 mM Tris-HCl pH 7.5, 1
751 mM EDTA, 2 M NaCl). Beads were added to the samples and incubated at
752 room temperature for 15 min with shaking. After capture, beads were washed
753 twice by adding 500 µL of Tween Wash Buffer and incubated at 55°C for 2 min
754 with shaking. Samples were then washed in 100 µL of 1× TD Buffer (2× TD
755 Buffer is 20 mM Tris-HCl pH 7.5, 10 mM magnesium chloride, 20%
756 dimethylformamide). After washes, beads were resuspended in 25 µL of 2× TD
757 Buffer, Tn5 (for 50 ng of post-ChIP DNA we used 2.5 µL of Tn5), and water to
758 50 µL. Tn5 amount was adjusted linearly for different amounts of post-ChIP
759 DNA, with a maximum amount of 4 µL of Tn5. Samples were incubated at
760 55°C with interval shaking for 10 min. After removing the supernatant 50 mM
761 EDTA was added to samples and incubated with interval shaking at 50°C for
762 30 min. Beads were then washed two times each in 50 mM EDTA then Tween
763 Wash Buffer at 55°C for 2 min. Lastly, beads were washed in 10 mM Tris
764 before PCR amplification. Beads were resuspended in 25 µL of Phusion HF 2×
765 (New England Biosciences), 1 µL of each Nextera Ad1_noMX and Nextera
766 Ad2.X at 12.5 µM, and 23 µL of water. The following PCR program was
767 performed: 72 °C for 5 min, 98 °C for 1 min, then cycle at 98 °C for 15 s, 63 °C
768 for 30 s, and 72 °C for 1 min (cycle number was estimated based on the
769 amount of material from the post-ChIP Qubit (approximately 50 ng was run in
770 six cycles, while 25 ng was run in seven, 12.5 ng was run in eight, etc.). Size
771 selection was performed using two-sided size selection with the Ampure XP
772 beads. After PCR, libraries were placed on a magnet and eluted into new tubes.
773 25 µL of Ampure XP beads were added, and the supernatant was kept to
774 capture fragments less than 700 bp. Supernatant was transferred to a new
775 tube, and 15 µL of fresh beads was added to capture fragments greater than
776 300 bp. After size selection, libraries were quantified with Qbit and sent for
777 Bioanalyzer to check for the quality and final size of the library. Libraries were
778 sequenced on a HiSeq 2500 platform on PE75 mode.
779
780 **ATAC-seq data analysis**
781 *Mapping, peak calling and peak processing*. Paired-end reads were aligned to
782 mm10 (bowtie2 version 2.3.2, --no-unal --local --very-sensitive-local --no-
783 discordant --no-mixed --contain --overlap --dovetail -I 10 -X 2000), and

784 mitochondrial DNA alignments were excluded. Fragments marked as positional
785 duplicates (sambamba version 0.6.6) or overlapping with mouseENCODE
786 blacklisted genomic regions (liftOver to mm10;[46]) were filtered out. Read ends
787 were adjusted for Tn5 transposase offsets. Peaks were called at $p<10^{-5}$ (MACS
788 version 2.1.1) per replicate, and only common peaks between two independent
789 biological replicates were retained for further analysis.
790
791 **ChIP-seq data analysis**
792 *Mapping, peak calling and peak processing*. Study and published ChIP-seq
793 reads were trimmed for adapters with parameters of maximum error rate of 0.4
794 and minimum length of 15 bp (cutadapt version 1.8.1), and low-quality ends
795 (sickle version 1.33), respectively. Alignment to the mouse reference genome
796 version mm10 (GRCm38.p4) was performed using standard parameters,
797 permitting a maximum of one mismatch in seed alignment (bowtie2 version 2.3.2).
798 Reads marked as positional duplicates (sambamba version 0.6.6) or overlapping
799 with mouseENCODE blacklisted genomic regions (liftOver to mm10;[46]) were
800 filtered out. Study ChIP-seq peaks (enrichment of signals over background
801 determined by input samples) were called at $p<0.01$ (MACS version 2.1.1) per
802 biological replicate, and peaks detected in more than half of biological replicates
803 were retained for further analysis. Published ChIP-seq replicates were merged,
804 and peaks were called at $p<10^{-5}$ using input samples where applicable.
805
806 **Overlap analysis of ChIP-seq peaks for chromatin states of reprogramming**
807 **cell types.** Chromatin states (1 kb resolution) during reprogramming were
808 retrieved from ref [6], and cis-regulatory elements were annotated from chromatin
809 states as in the original publication. The assignment of ChIP-seq peaks to cis-
810 regulatory elements was determined by the largest degree of overlap in bp.
811
812 **ChIP-seq gene ontology (GO) enrichment analysis.** Spatial proximity of ChIP-
813 seq peaks to transcript start sites (TSSs) and enriched GOs were uncovered
814 utilizing the GREAT (version 3.0.0) web application. We selected the "basal plus
815 extension rule" for the association of gene ontology annotations with regulatory
816 domains (customized setting: 5 kb upstream and 1 kb downstream of TSSs, and
817 further extended both directions by 250 kb). Enrichment of ontology annotations
818 was assessed by the binomial test of ChIP-seq peak-overlaps with annotated
819 regulatory regions.
820
821 **PSC typical enhancers and super-enhancers.** Coordinates of typical- and
822 super-enhancers in mESCs were ascertained from ref[34], lifted over from mm9 to
823 mm10, and gene symbols featured in the GENCODE reference (version M6)
824 were retained.
825
826 **Overlaps of KLF4 binding with early lost or late gained H3K27ac peaks**
827 **during reprogramming or at typical- and super-enhancers.** The maximum
828 permitted distance between KLF4 binding detected in day3, 6 and 9 with PSC
829 and H3K27ac peaks or enhancers in ref[34] was 250 bp. Where H3K27ac peaks or

enhancers overlapped with KLF4 sites of different stages, the earliest stage was prioritized (Fig. 1g).

Motif analysis

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

PCA analysis for ATAC-seq, RNA-seq and ChIP-seq experiments:

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

RNA-seq data analysis

Expression of genes was quantified in transcripts per kilobase million (TPM) using quasi mapping (Salmon, version 0.8.2) to GENCODE (version M6, mm10) reference gene annotation. Salmon provides alignment-free transcript quantification information in a single step[47].

Line plots for gene expression analysis

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

Preprocessing of HiChIP / public Hi-C datasets

HiChIP datasets and public Hi-C datasets were uniformly pre-processed with the HiC-bench platform[30], which is outlined in short in the following. First, all paired-end sequencing reads were aligned against the mouse genome version mm10 with bowtie2 version 2.2.3[48] (specific settings: --very-sensitive-local --local). Read-filtering was conducted by the GenomicTools[49] gtools-hic filter command (integrated in HiC-bench), which discards multi-mapped reads ("multihit"), read-pairs with only one mappable read ("single sided"), duplicated read-pairs ("ds.duplicate"), read-pairs with a low mapping quality of MAPQ < 20, read-pairs

resulting from self-ligated fragments and short-range interactions resulting from read-pairs aligning within 10kb (together called “ds.filtered”). The percentage of accepted intra-chromosomal read-pairs (“ds.accecpted intra”) was high across all HiChIP replicates and conditions and was consistently above 35% (Figure S3a). In order to create counts-matrices per chromosome in a binned fashion, we set the bin size to 10kb for all datasets. For all the HiChIP sample and chromosome matrices, the trajectories of each matrix bin to both anchors were overlaid with the ChIP-Seq signal of the respective matching sample, requiring a minimal overlap of 1bp between a HiChIP-bin and a ChIP-peak. Only loops of which at least one anchor was supported by a ChIP-peak were kept for further analyses. Next, we applied sequencing-depth normalization (called “scaling”) and distance normalization as recently described[50].
High-confidence loops were identified by requiring at least 1 read-pair per loop across all replicates of a single condition in order to maintain a signal that is replicable. Next, all loops were ordered by average loop strength per cell-type and antibody and the top 50k / 100k interactions were used for downstream analyses for KLF4 and H3K27ac, respectively.

Principal component analysis for HiChIP samples

Principal component analysis (PCA) as shown in Figures 3a and S3b was performed on all available replicates on the high-confidence loops. Therefore, for each detected high-confidence loop from any sample, the per replicate normalized loop strength was extracted before filtering for top 50k / 100k loops in order to also integrate lowly detected interactions in the analysis. PCA was performed using the prcomp function of R (version 3.3.0; scale=TRUE, center=TRUE)

HiChIP/Hi-C heatmaps

All HiChIP and Hi-C heatmaps were generated with the image.plot function of R the package fields (version 8.10), using the normalized interaction matrices. Color scale was set the same across all conditions but separately for Hi-C and HiChIP samples.

Differential loop analysis

Differential looping analysis was performed on each loop independently by applying a two-sided t-test on the normalized counts between any pairwise comparisons: ES-KLF4 vs d3-KLF4, ES-KLF4 vs d6-KLF4, d3-KLF4 vs d6-KLF4 and ES-H3K27ac vs MEF-H3K27ac. In order to estimate the change in loop strength, we calculated the log fold-change (logFC) between the average normalized loop-strength per condition for the same pairwise comparisons after adding a pseudo-count of 1 to each replicate and loop. For constant H3K27ac loops, we selected loops with p-value > 0.5 and absolute logFC < 0.5 for the pairwise comparison of H3K27ac. MEF/ES-specific H3K27ac loops were selected by p-value < 0.01 and logFC > 2 / logFC < -2 taken from the ES H3K27ac vs MEF H3K27ac comparison, respectively. Mid and late established KLF4 loops were selected by applying p-value < 0.01 and logFC > 2 in the pairwise comparisons of ES-KLF4 vs d3-KLF4 and d6-KLF4 vs d3-KLF4 (mid)

924 and ES-KLF4 vs d3-KLF4 and ES-KLF4 vs d6-KLF4 (late). Transient and mid-
925 lost KLF4 loops were selected by applying a p-value < 0.01 and logFC < -2 in the
926 pairwise comparisons of PSC-KLF4 vs d6-KLF4 and PSC-KLF4 vs d3-KLF4
927 (transient) and PSC-KLF4 vs d3-KLF4 and d6-KLF4 vs d3-KLF4 (mid-lost). All
928 calculations were performed in R version 3.3.0, using the native t.test function.
929
930 **Annotation of H3K27ac HiChIP loop anchors as promoter or enhancers***.*
931 H3K27ac HiChIP loop anchors were overlapped with transcription start sites
932 (TSSs) of GENCODE (version M6) genes. Presence of one or more TSSs was
933 considered a promoter HiChIP anchor, and the absence of any TSS but
934 presence of at least one H3K27ac ChIP-seq constitutes an enhancer HiChIP
935 anchor. In estimating connectivity, all HiChIP anchors, either promoter, enhancer
936 or otherwise desolate, were considered.
937
938 **RNA expression integration (enhancer cliques)**
939 For RNA expression integration, we overlapped all canonical TSSs of protein-
940 coding genes (transcript support level/TSL = 1) downloaded from Ensembl
941 Genes V85 for the mouse genome mm10 with all loop anchors. Because the
942 TSS is a 1bp position in the genome, each gene was uniquely assigned to one
943 bin, however, multiple TSSs per gene with a TSL=1 mapping to different bins are
944 possible. Next, we filtered genes by occurrence of differential loop clusters that
945 were obtained from the HiChIP experiments and have TPM > 1 in at least one
946 reprogramming stage, and analyzed the expression patterns of such genes
947 throughout reprogramming. For H3K27ac HiChIP data integration, we assigned
948 genes to MEF/PSC-specific loops if their TSSs were found in >= 1, >= 2 or >= 5
949 MEF/PSC-specific loops but in none of the other (Figures 2c and S2e). Genes
950 contained in anchors of constant loops were solely filtered by having at least two
951 constant loop anchors. To further validate expression changes based on
952 differential looping, we applied a two-sided t-test between genes logFCs of
953 constant H3K27ac loops versus genes with MEF/PSC-specific loops. As a
954 negative control, we compared logFCs of genes with constant loops versus all
955 annotated protein-coding genes.
956
957 **Co-regulation of gene expression by H3K27ac HiChIP (enhancer hubs***).* In
958 this analysis, promoter anchors of enhancer-mediated loops were filtered for
959 protein-coding genes that have an expression TPM > 1 in MEF or PSC.
960 Enhancer fragments that contact two to ten promoter fragments in PSC specific
961 H3K27ac loops were selected. Genes were paired across different promoter
962 fragments connecting to the same enhancer anchor (later on called hub), and
963 repeated gene pairs were removed from the overall pool. Owing to lower
964 promoter connectivity of enhancer hubs in MEF-specific or constant H3K27ac
965 HiChIP loops, the constraint of promoter contacts was reduced to five or three for
966 enhancer hubs in MEF-specific or constant loops. Gene pairs were considered
967 co-expressed, if both genes were up-regulated in PSC compared to MEF ($p<10^{-2}$
968 and fold change threshold of 1.5 in R limma package version 3.32.10) or vice
969 versa for down-regulation. Or otherwise, at least one gene in a pair unchanged
970 between MEF and PSC constitute unchanged gene pairs. Enrichment of co-

regulated gene pairs in the genuine data compared to either gene pairs in random shuffled distance matched window or from within TADs was statistically assessed using Fisher's exact test (Hi-C data from Pelham-Webb et al, unpublished results). Topologically associated domains (TADs) were called from normalized Hi-C matrices in PSCs processed at 40kb resolution using a recently published software[51]. The insulation window parameter was set to 500 kb.

Overlap between H3K27ac PSC-specific loops and KLF4 clusters
Overlap between any of the KLF4 peaks with any of the HiChIP anchors (H3K27aC or KLF4 loops) was performed with the use of bedtools v2.25.0.

ChIP-seq feature enrichment at lowly or highly connected H3K27ac HiChIP enhancer anchors. H3K27ac HiChIP anchors were selected for low (contacting only one other anchor; N = 3,594) or high connectivity (contacting four or more anchors; N = 1,434), and the low connectivity anchors were sub-sampled to match numbers of the high connectivity class. Average signals of ChIP-seq features (reads per kilobase per million/RPKM) per base under peaks or the fraction of peaks that are within anchors were statistically (Wilcoxon signed-rank) tested for disparity between low and high connectivity anchors (Fig.2g). Publicly available ChIP-seq data was used for OCT4, SOX2, MED1, SMC1, CTCF and YY1:

	Study Accession	Sample Accession
OCT4	GSE44288	SRR713340
SOX2	GSE44288	SRR713341
MED1	GSE22557	SRR058987|SRR058988
SMC1	GSE22557	SRR058981|SRR058982|SRR058995|SRR058996
CTCF	https://www.encodeproject.org	https://www.encodeproject.org/files/ENCFF001LIY/@@download/ENCFF001LIY.fastq.gz
YY1	GSE99518	SRR5634080

KLF4 peaks involved in KLF4 and H3K27ac looping.
Anchors of high-confidence PSC KLF4 loops and of PSC-specific H3K27ac loops were stratified by KLF4 binding. . KLF4 peaks were overlapped with anchors of loops of KLF4 and H3K27ac, separately. In these separate analyses, the viewpoint was placed at this anchor to see if any of the partner anchors has KLF4 binding. If so, these were called "bi-anchor" for the viewpoint anchor. Thereby KLF4 peaks in ChIP-seq carried the identity of the viewpoint anchor. A small number of KLF4 peaks may have conflict identities, due to partial overlapping with multiple anchors, and thereby removed. A KLF4 peak is called 'no-anchor' if it does not overlap with any anchors in neither KLF4 nor H3K27ac HiChIP loops.

KLF4 looping involved in RNA expression changes
To estimate the effect of KLF4 associated looping on changes in RNA expression, we followed a similar approach as for 'enhancer cliques'. After selecting

1009 expressed genes within anchors of each KLF4 loop cluster, we further filtered for
1010 differentially expressed genes between PSC and day3 (FDR < 0.01; logFC > 1.5
1011 (upregulated) or logFC < -1.5 (downregulated)). Information on differential
1012 expression was derived by edgeR (version 3.14.0) with subsequent multiple
1013 testing correction. Genes determined as 'no change' were selected by applying
1014 FDR > 0.5 and absolute logFC < 0.5. All remaining genes were discarded from
1015 the analysis.
1016
1017 **LOLA enrichment analysis**
1018 The identified differential loops were subjected for an enrichment analysis of
1019 further transcription factor bindings and other DNA binding proteins. First, the
1020 anchors of each differential loop were mapped back to the original ChIP-peaks,
1021 because the 10kb stretches of the bins would give too many false positive
1022 findings. Each anchor that was overlapping an actual ChIP-peak by at least 1bp
1023 was subjected for further analysis. Since two anchors can theoretically overlap
1024 with a single ChIP-peak using this approach, the resulting list was collapsed and
1025 only unique ChIP-peaks were kept. Next, we applied LOLA version 1.4.0 [31]
1026 against a database of analyzed ChIP-Seq datasets taken from LOLA Region
1027 Databases (regionDB) for mm10 (for Figure S3e we used Codex and encode
1028 TFBSmm10 databases). We excluded all ChIP-Seq datasets that were marked
1029 as treated with any agent and had less than 3,000 peaks in total (5,000 peaks for
1030 Figure S3e). As a universe for LOLA, we used only unique ChIP-peaks of the
1031 union of all ChIP-Seq peaks for the respective antibody across all reprogramming
1032 stages.
1033
1034 **Virtual 4C**
1035 Virtual 4C was performed to identify interaction signals of gene promoters or
1036 enhancers with their genomic vicinity. For this approach, we used the filtered
1037 HiChIP read-pairs as described above before binning and normalization of each
1038 replicate. We extracted all read pairs for which a read mate maps within +/- 10kb
1039 around the virtual viewpoint. Next, we defined successive overlapping windows
1040 for each chromosome at a 10kb resolution, and all adjacent windows are
1041 overlapping by 95% of their length (i.e. 9.5kb, or a shift of 500bp between
1042 adjacent windows). We then counted the second mapped read mate in all
1043 overlapping bins. Thus, each read-pair accounts for +1 in exactly 19 bins,
1044 however, the overlap of bins achieves a smoothed signal. Read counts for all
1045 bins were normalized to total sequencing depth of the respective replicate by
1046 edgeR version 3.14.0 to calculate counts-per-million (CPM) per bin. For
1047 visualization, the average of the normalized virtual 4C-signal across replicates of
1048 a single condition was calculated. We have used the following viewpoints:
1049
1050 *Zfp42/Triml2* locus chr8:43,321,380-43,321,381
1051 *Mycn* locus chr12:12,941,173-12,941,174
1052 *Ets1* locus chr9:32,696,322-32,696,323
1053 *Tbx3* locus chr5:119,670,608-119,670,609 (promoter) / chr5:119,585,186-
1054 119,585,187 (Dist enhancer)
1055

ACKNOWLEDGEMENTS

We thank all members from the Apostolou and Tsirigos groups for critical reading of the manuscript. We would also like to acknowledge Olivier Elemento and Ashley Doane for input with initial computational analyses and WCM Epigenomics and Genomics facilities for support on next-generation sequencing. D.C.D.G is a New York Stem Cell Foundation postdoctoral fellow and also supported by the Family-Friendly Postdoctoral Initiative at Weill Cornell Medicine. This work was partly funded by the NIH Director's New Innovator Award (1DP2DA043813), the Edward Jr Mallinckrodt Foundation and the Tri-Institutional Stem Cell Initiative by the Starr Foundation.

AUTHOR CONTRIBUTIONS

E.A conceived, designed and supervised the study with input from A.T and M.S. A.T. supervised the computational analyses. D.C.D.G designed and performed most of the experiments with help from D.K and E.A. Y.L analyzed all ChIP-seq, RNA-seq and ATAC-seq datasets. A.K analyzed all HiChIP and Hi-C data. Y.L, A.K and A.P.P performed integrative analyses. D.K contributed to the CRISPR/Cas9 experiments. E.A and D.C.G wrote the manuscript with input from all authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

References

1 Young, R. A. Control of the embryonic stem cell state. *Cell* **144**, 940-954, doi:10.1016/j.cell.2011.01.032 (2011).

2 Natoli, G. Maintaining cell identity through global control of genomic organization. *Immunity* **33**, 12-24, doi:10.1016/j.immuni.2010.07.006 (2010).

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

4 Graf, T. Historical origins of transdifferentiation and reprogramming. *Cell Stem Cell* **9**, 504-516, doi:10.1016/j.stem.2011.11.012 (2011).

5 Apostolou, E. & Hochedlinger, K. Chromatin dynamics during cellular reprogramming. *Nature* **502**, 462-471, doi:10.1038/nature12749 (2013).

6 Chronis, C. *et al.* Cooperative Binding of Transcription Factors Orchestrates Reprogramming. *Cell* **168**, 442-459 e420, doi:10.1016/j.cell.2016.12.016 (2017).

7 Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151**, 994-1004, doi:10.1016/j.cell.2012.09.045 (2012).

1095 8 Chen, J. *et al.* Hierarchical Oct4 Binding in Concert with Primed
1096 Epigenetic Rearrangements during Somatic Cell Reprogramming. *Cell Rep* **14**,
1097 1540-1554, doi:10.1016/j.celrep.2016.01.013 (2016).
1098 9 Sridharan, R. *et al.* Role of the murine reprogramming factors in the
1099 induction of pluripotency. *Cell* **136**, 364-377, doi:S0092-8674(09)00007-5 [pii]
1100 10.1016/j.cell.2009.01.001 (2009).
1101 10 Knaupp, A. S. *et al.* Transient and Permanent Reconfiguration of
1102 Chromatin and Transcription Factor Occupancy Drive Reprogramming. *Cell Stem*
1103 *Cell* **21**, 834-845 e836, doi:10.1016/j.stem.2017.11.007 (2017).
1104 11 Apostolou, E. *et al.* Genome-wide Chromatin Interactions of the Nanog
1105 Locus in Pluripotency, Differentiation, and Reprogramming. *Cell stem cell* **12**,
1106 699-712, doi:S1934-5909(13)00149-5 [pii]10.1016/j.stem.2013.04.013 (2013).
1107 12 de Wit, E. *et al.* The pluripotent genome in three dimensions is shaped
1108 around pluripotency factors. *Nature* **501**, 227-231, doi:10.1038/nature12420
1109 (2013).
1110 13 Denholtz, M. *et al.* Long-range chromatin contacts in embryonic stem cells
1111 reveal a role for pluripotency factors and polycomb proteins in genome
1112 organization. *Cell Stem Cell* **13**, 602-616, doi:10.1016/j.stem.2013.08.013 (2013).
1113 14 Wei, Z. *et al.* Klf4 Organizes Long-Range Chromosomal Interactions with
1114 the Oct4 Locus in Reprogramming and Pluripotency. *Cell stem cell*, doi:S1934-
1115 5909(13)00203-8 [pii]10.1016/j.stem.2013.05.010 (2013).
1116 15 Krijger, P. H. *et al.* Cell-of-Origin-Specific 3D Genome Structure Acquired
1117 during Somatic Cell Reprogramming. *Cell Stem Cell* **18**, 597-610,
1118 doi:10.1016/j.stem.2016.01.007 (2016).
1119 16 Beagan, J. A. *et al.* Local Genome Topology Can Exhibit an Incompletely
1120 Rewired 3D-Folding State during Somatic Cell Reprogramming. *Cell Stem Cell*
1121 **18**, 611-624, doi:10.1016/j.stem.2016.04.004 (2016).
1122 17 Rubin, A. J. *et al.* Lineage-specific dynamic and pre-established enhancer-
1123 promoter contacts cooperate in terminal differentiation. *Nat Genet* **49**, 1522-1528,
1124 doi:10.1038/ng.3935 (2017).
1125 18 Drissen, R. *et al.* The active spatial organization of the beta-globin locus
1126 requires the transcription factor EKLF. *Genes Dev* **18**, 2485-2490,
1127 doi:10.1101/gad.317004 (2004).
1128 19 Schoenfelder, S. *et al.* Preferential associations between co-regulated
1129 genes reveal a transcriptional interactome in erythroid cells. *Nat Genet* **42**, 53-61,
1130 doi:10.1038/ng.496 (2010).
1131 20 Stadtfeld, M., Maherali, N., Borkent, M. & Hochedlinger, K. A
1132 reprogrammable mouse strain from gene-targeted embryonic stem cells. *Nat*
1133 *Methods* **7**, 53-55, doi:10.1038/nmeth.1409 (2010).
1134 21 Polo, J. M. *et al.* A molecular roadmap of reprogramming somatic cells
1135 into iPS cells. *Cell* **151**, 1617-1632, doi:10.1016/j.cell.2012.11.039 (2012).
1136 22 Stadtfeld, M. *et al.* Ascorbic acid prevents loss of Dlk1-Dio3 imprinting and
1137 facilitates generation of all-iPS cell mice from terminally differentiated B cells. *Nat*
1138 *Genet* **44**, 398-405, S391-392, doi:10.1038/ng.1110 (2012).
1139 23 Stadtfeld, M. *et al.* Aberrant silencing of imprinted genes on chromosome
1140 12qF1 in mouse induced pluripotent stem cells. *Nature* **465**, 175-181,
1141 doi:10.1038/nature09017 (2010).

1142 24 Stadtfeld, M., Maherali, N., Breault, D. T. & Hochedlinger, K. Defining
1143 molecular cornerstones during fibroblast to iPS cell reprogramming in mouse.
1144 *Cell Stem Cell* **2**, 230-240, doi:10.1016/j.stem.2008.02.001 (2008).
1145 25 McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-
1146 regulatory regions. *Nat Biotechnol* **28**, 495-501, doi:10.1038/nbt.1630 (2010).
1147 26 Soufi, A. *et al.* Pioneer transcription factors target partial DNA motifs on
1148 nucleosomes to initiate reprogramming. *Cell* **161**, 555-568,
1149 doi:10.1016/j.cell.2015.03.017 (2015).
1150 27 Whyte, W. A. *et al.* Master transcription factors and mediator establish
1151 super-enhancers at key cell identity genes. *Cell* **153**, 307-319,
1152 doi:10.1016/j.cell.2013.03.035 (2013).
1153 28 Stadhouders, R. *et al.* Transcription factors orchestrate dynamic interplay
1154 between genome topology and gene regulation during cell reprogramming. *Nat*
1155 *Genet*, doi:10.1038/s41588-017-0030-7 (2018).
1156 29 Mumbach, M. R. *et al.* HiChIP: efficient and sensitive analysis of protein-
1157 directed genome architecture. *Nat Methods* **13**, 919-922,
1158 doi:10.1038/nmeth.3999 (2016).
1159 30 Lazaris, C., Kelly, S., Ntziachristos, P., Aifantis, I. & Tsirigos, A. HiC-
1160 bench: comprehensive and reproducible Hi-C data analysis designed for
1161 parameter exploration and benchmarking. *BMC Genomics* **18**, 22,
1162 doi:10.1186/s12864-016-3387-6 (2017).
1163 31 Sheffield, N. C. & Bock, C. LOLA: enrichment analysis for genomic region
1164 sets and regulatory elements in R and Bioconductor. *Bioinformatics* **32**, 587-589,
1165 doi:10.1093/bioinformatics/btv612 (2016).
1166 32 Mumbach, M. R. *et al.* Enhancer connectome in primary human cells
1167 identifies target genes of disease-associated DNA elements. *Nat Genet* **49**,
1168 1602-1612, doi:10.1038/ng.3963 (2017).
1169 33 Dixon, J. R. *et al.* Topological domains in mammalian genomes identified
1170 by analysis of chromatin interactions. *Nature* **485**, 376-380,
1171 doi:10.1038/nature11082 (2012).
1172 34 Dowen, J. M. *et al.* Control of cell identity genes occurs in insulated
1173 neighborhoods in mammalian chromosomes. *Cell* **159**, 374-387,
1174 doi:10.1016/j.cell.2014.09.030 (2014).
1175 35 Gomez-Diaz, E. & Corces, V. G. Architectural proteins: regulators of 3D
1176 genome organization in cell fate. *Trends Cell Biol* **24**, 703-711,
1177 doi:10.1016/j.tcb.2014.08.003 (2014).
1178 36 Weintraub, A. S. *et al.* YY1 Is a Structural Regulator of Enhancer-
1179 Promoter Loops. *Cell* **171**, 1573-1588 e1528, doi:10.1016/j.cell.2017.11.008
1180 (2017).
1181 37 Beagan, J. A. *et al.* YY1 and CTCF orchestrate a 3D chromatin looping
1182 switch during early neural lineage commitment. *Genome Res* **27**, 1139-1152,
1183 doi:10.1101/gr.215160.116 (2017).
1184 38 Schoenfelder, S. *et al.* Polycomb repressive complex PRC1 spatially
1185 constrains the mouse embryonic stem cell genome. *Nat Genet* **47**, 1179-1186,
1186 doi:10.1038/ng.3393 (2015).

1187 39 Kearns, N. A. *et al.* Functional annotation of native enhancers with a
1188 Cas9-histone demethylase fusion. *Nat Methods* **12**, 401-403,
1189 doi:10.1038/nmeth.3325 (2015).
1190 40 Li, D. *et al.* Chromatin Accessibility Dynamics during iPSC
1191 Reprogramming. *Cell Stem Cell* **21**, 819-833 e816,
1192 doi:10.1016/j.stem.2017.10.012 (2017).
1193 41 Samstein, R. M. *et al.* Foxp3 exploits a pre-existent enhancer landscape
1194 for regulatory T cell lineage specification. *Cell* **151**, 153-166,
1195 doi:10.1016/j.cell.2012.06.053 (2012).
1196 42 Ghaleb, A. M. & Yang, V. W. Kruppel-like factor 4 (KLF4): What we
1197 currently know. *Gene* **611**, 27-37, doi:10.1016/j.gene.2017.02.025 (2017).
1198 43 Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system.
1199 *Nat Protoc* **8**, 2281-2308, doi:10.1038/nprot.2013.143 (2013).
1200 44 Liu, Y. *et al.* Widespread Mitotic Bookmarking by Histone Marks and
1201 Transcription Factors in Pluripotent Stem Cells. *Cell Rep* **19**, 1283-1293,
1202 doi:10.1016/j.celrep.2017.04.067 (2017).
1203 45 Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A
1204 Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol*
1205 **109**, 21 29 21-29, doi:10.1002/0471142727.mb2129s109 (2015).
1206 46 Consortium, E. P. An integrated encyclopedia of DNA elements in the
1207 human genome. *Nature* **489**, 57-74, doi:10.1038/nature11247 (2012).
1208 47 Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon
1209 provides fast and bias-aware quantification of transcript expression. *Nat Methods*
1210 **14**, 417-419, doi:10.1038/nmeth.4197 (2017).
1211 48 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2.
1212 *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
1213 49 Tsirigos, A., Haiminen, N., Bilal, E. & Utro, F. GenomicTools: a
1214 computational platform for developing high-throughput analytics in genomics.
1215 *Bioinformatics* **28**, 282-283, doi:10.1093/bioinformatics/btr646 (2012).
1216 50 Gong, Y. *et al.* Stratification of TAD boundaries reveals preferential
1217 insulation of super-enhancers by strong boundaries. *Nat Commun* **9**, 542,
1218 doi:10.1038/s41467-018-03017-1 (2018).
1219 51 Crane, E. *et al.* Condensin-driven remodelling of X chromosome topology
1220 during dosage compensation. *Nature* **523**, 240-244, doi:10.1038/nature14450
1221 (2015).
1222

1223