Loss of MLL3/4 decouples enhancer H3K4 monomethylation, H3K27 acetylation, and gene activation during ESC differentiation

Ryan M. Boileau[1,2,3], Kevin X. Chen[1,2,3], Robert Blelloch[1,2,3,#]

[1]The Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, Center for Reproductive Sciences, University of California San Francisco, San Francisco, USA

[2]Department of Urology, University of California San Francisco, San Francisco, USA

[3]Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA

[#]Corresponding Author

Contact Information
Ryan M. Boileau, Ryan.Boileau@ucsf.edu
Kevin X. Chen, Kevin.Chen4@ucsf.edu
Robert Blelloch, Robert.Blelloch@ucsf.edu

Abstract

Enhancers are essential in defining cell fates through the control of cell type specific gene expression. Enhancer activation is a multi-step process involving chromatin remodelers and histone modifiers including the monomethylation of H3K4 (H3K4me1) by MLL3 (KMT2C) and MLL4 (KMT2D). MLL3/4 are thought to be critical for enhancer activation and cognate gene expression including through the recruitment of acetyltransferases for H3K27. Here we test this model by evaluating the impact of MLL3/4 loss on chromatin and transcription during early embryonic stem cell differentiation. We find that MLL3/4 activity is required at most if not all sites that gain or lose H3K4me1 but is largely dispensable at sites that remain stably methylated during this transition. This requirement extends to H3K27 acetylation (H3K27ac) at most transitional sites. However, many sites gain H3K27ac independent of MLL3/4 or H3K4me1 including enhancers regulating key factors in early differentiation. Furthermore, despite the failure to gain active histone marks at thousands of enhancers, transcriptional activation of nearby genes is largely unaffected, thus uncoupling the regulation of these chromatin events from transcriptional changes during this transition. These data challenge current models of enhancer activation and imply distinct mechanisms between stable and dynamically changing enhancers. Collectively, our study highlights gaps in knowledge about the steps and epistatic relationships of enzymes necessary for enhancer activation and cognate gene transcription.

Keywords

Enhancer, epistasis, pluripotency, MLL3, MLL4, H3K4me1, P300, H3K27ac, transcription, differentiation

83 **Background**

84 Gene regulatory networks that drive cell fate are regulated spatiotemporally by cell type
85 specific transcription factors (TFs). The critical functions of TFs in development are coupled to
86 their target genes through TF binding of *cis*-regulatory elements such as enhancers. Target gene
87 regulation is mediated by general chromatin regulators which are recruited to enhancers bound
88 by TFs. As such, chromatin regulators are an essential part of enhancer function and their
89 dysregulation has severe consequences in development and disease. For example, MLL3 and
90 MLL4 are functionally redundant histone methyltransferases within the COMPASS complex which
91 deposit the histone modification H3K4me1 primarily at enhancers(1,2). Moreover, mutations in
92 MLL3 and MLL4 cause developmental defects such as Kabuki Disease and are in the top ten of
93 frequently mutated genes in cancer(3–5). Despite the importance of chromatin regulators to
94 human health, we lack fundamental insight into how general chromatin regulators such as MLL3/4
95 coordinate enhancer function.

96 The interdependence of different histone modifications and their depositing enzymes at
97 enhancers has been widely studied, although their mechanistic roles remain controversial. A
98 prevailing model of enhancer activation is that DNA bound TFs recruit MLL3/4 which deposit
99 H3K4me1(6–9). MLL3/4 in turn recruits the acetyltransferases P300/CBP which deposit
100 H3K27ac. An activated enhancer then promotes expression of its cognate gene, most often one
101 that is a nearby neighbor(10,11). However, it is unclear how generalizable this model is. Indeed,
102 recent studies have begun to challenge the importance of histone modifications and even
103 chromatin regulators in gene regulation by enhancers(12–16).

104 Therefore, we decided to revisit the prevailing model including the epistatic relationship
105 between MLL3/4 to H3K4me1, H3K27ac, and gene transcription in the context of ESCs
106 transitioning from the naive to formative pluripotent states. The naive to formative transition is a
107 rapid, homogenous, and well-characterized transition that faithfully recapitulates the early
108 embryonic cell fate transition in epiblast cells of the mouse between E4.5 and E5.5 at the

109 transcriptional and epigenomic level(Fig.1A)(17–20). Importantly, this developmental window
110 immediately precedes previously described cell fate specification and migrational phenotypes
111 seen in gastrulation with the loss of MLL3/4 *in vivo*(6,21). Therefore, the naive to formative
112 transition provides an ideal model to study relationships in enhancer mechanics and
113 transcriptional regulation.

114 Using state-of-the-art epigenomics techniques to profile genetic deletion of both MLL3 and
115 MLL4 in the naive to formative transition we make several surprising discoveries. First, MLL3/4 is
116 only required at a small subset of sites that maintain stable H3K4me1 levels during the transition
117 but is essential for all sites that gain or lose H3K4me1. This demonstrates the existence of distinct
118 mechanisms of H3K4me1 deposition at most stable vs dynamic sites. Second, loss of MLL3/4
119 reduces H3K27ac levels at most dynamic sites, but often does so independent of underlying
120 H3K4me1 levels. Finally, despite dramatic loss of active histone modifications at many enhancers
121 in the formative state, few loci can be linked to changes in nearby gene expression with these
122 changes being relatively minor. Taken together, these results demonstrate that the regulation of
123 the deposition of H3K4me1 and H3K27ac can be uncoupled from each other and also from
124 changes in gene expression at most enhancer sites during early ESC differentiation. Therefore,
125 our data challenge the prevailing model of enhancer activation and suggests that the functional
126 role of MLL3/4 in cell fate transitions is more context dependent than previously appreciated.

127

128 **Results**

129 **MLL3/4 is dispensable for transcriptional activation of much of the formative program**

130 To study the role of MLL3 and MLL4 in stem cell self-renewal and differentiation we
131 acquired MLL3$^{-/-}$; MLL4$^{fl/fl}$ ESCs (hereafter called MLL3KO)(22). We integrated Cre-ERT2 into the
132 *Rosa26* locus and induced Cre-recombination with tamoxifen to generate MLL3$^{-/-}$; MLL4$^{-/-}$ double
133 knockout clones (DKO)(Fig.S1A). Western blotting confirmed loss of MLL3 protein alone in the
134 parental line and both MLL3 and MLL4 in the DKO cells compared to a wildtype (WT) control

135 (Fig.1B). Under naive pluripotency culture conditions (Serum+LIF+2i), the DKO cells formed
136 colonies that were less compact and grew at a slightly slower rate compared to their WT or
137 MLL3KO counterparts (Fig.1C, Fig.S1B). Next, we removed LIF+2i from the media to induce the
138 formative state after ~63 hours in WT cells (23). As expected, WT ESC colonies flattened out
139 during this transition. Similar phenotypic changes were seen in MLL3KO and DKO, although in
140 both cases cells appeared more dispersed. All cell lines expressed similar levels of OCT4 and
141 NANOG proteins in naive culture conditions and similarly downregulated NANOG in formative
142 conditions as indicated by Western blot (Fig.S1C). These data show that while there are some
143 morphological and proliferative differences, the knockout cells self-renew under naive culture
144 conditions and show evidence of a cell state transition in formative conditions.

145 To determine the transcriptional effects of loss of MLL3/4 we performed RNA-seq on WT,
146 MLL3KO, and DKO lines in naive and formative conditions. Differential gene expression analysis
147 between WT naive and formative samples uncovered 887 downregulated genes (naive enriched
148 genes) and 1602 upregulated genes (formative enriched genes) (Fig.1D). After subsetting for all
149 naive and formative enriched genes we conducted UMAP analysis of both the WT and mutant
150 cells under both conditions and found samples separated predominately by developmental state
151 rather than by genotype (Fig.1E). Correlation analysis of the gene expression changes seen in
152 MLL3KO vs WT and DKO vs. WT showed Pearson correlations of 0.8 and 0.6 respectively
153 suggesting that gene regulation is largely maintained in the absence of MLL3/4 (Fig.1F,G). There
154 was a slight dampening of gene expression changes in the MLL3/4 DKO as evidenced by a
155 reduced slope in the linear model. In particular, some naive genes appeared to be incompletely
156 downregulated. qPCR analysis validated the successful upregulation of known formative markers
157 (Fgf5, Otx2, Dnmt3b) and the incomplete downregulation of two out of three naive markers
158 (Zfp42/Rex1 and Klf4 vs. Klf2) (Fig.S1D, E). These results suggested a relatively normal overall
159 transition in gene expression changes between naive and formative even in the absence of both
160 MLL3 and MLL4.

161 In contrast, direct comparison of MLL3/4 DKO vs. WT identified hundreds of misregulated
162 genes in both the naive and formative states. In the naive state, DKO cells had 783 upregulated
163 genes and 929 downregulated genes (Fig.1H). We categorized these genes that were up or down
164 by whether they are normally naive enriched, formative enriched, or unchanged between naive
165 and formative (Fig.1I). This analysis showed a similar proportion of formative genes to be
166 prematurely expressed (275 of 783, 35%) as naive enriched genes that were reduced in the naive
167 state (348 of 928, 39%). In the formative state we found 646 genes were upregulated and 461
168 genes were downregulated in DKO cells compared to WT(Fig.1J). Surprisingly, fewer genes were
169 down in the DKOs in the formative state than the naive state (461 vs. 929). Focusing on genes
170 that normally change during the transition, 83 naive enriched genes were abnormally up and 251
171 formative enriched genes were abnormally down in the formative state (Fig.1K). These numbers
172 were overall low relative to normal gene expression changes seen with naive to formative
173 transition (887 down, 1602 up) showing that MLL3/4 only regulates a minor subset of genes
174 among those that are normally gained with the transition from naive to formative.
175
176 **MLL3/4 is required for all dynamic H3K4me1 deposition during pluripotent transition**
177 Though expression of a small number of genes were impacted it was unclear if this was
178 due to direct or indirect effects of MLL3/4 loss. To gain a better understanding of the direct effects
179 of MLL3/4 loss on transcription changes during the naive to formative transition we evaluated the
180 monomethylation of H3K4, the primary known role of MLL3/4 and a surrogate marker of their
181 binding. We performed CUT&RUN(24) to profile H3K4me1 in naive and formative samples for
182 WT, MLL3KO, and DKO cells. Differential peak analysis using DiffBind(25) on WT naive and
183 formative samples with peaks called by SEACR(26) uncovered approximately ~13000 peaks
184 each that significantly gained (formative H3K4me1 peaks) or lost H3K4me1 (naive H3K4me1
185 peaks) during the normal pluripotent transition (FDR < 0.01, Log2 FC > 1). An additional ~94000
186 peaks did not significantly change between cell states (shared H3K4me1 peaks, FDR > 0.1, Log2

187 FC < 0.7)(Fig.2A,B). Naïve and formative enriched peaks were almost completely lost in the DKO
188 cells, but only slightly reduced in MLL3KO cells (Fig.2C,D). Reanalysis by Diffbind uncovered 0
189 total peaks that showed a significant change between the two states in the DKO background while
190 we found 29088 significantly changing peaks in MLL3KO compared to 27458 total peaks in WT
191 (Fig.2E, Fig.S1A). In contrast, shared H3K4me1 peaks (shown in gray) were not obviously
192 affected. To test whether the loss of dynamic H3K4me1 was a direct effect of MLL3/4 enzymatic
193 activity, we performed and evaluated RNA-seq and H3K4me1 CUT&RUN data using ESCs with
194 point mutations in the SET domains of both MLL3 and 4 that render them catalytically dead (dCD
195 cells)(12). Consistent with previous reports(12,27), we found few transcriptional differences
196 between dCD and WT cells in the naive and formative states (Fig.S2B,C,D). Moreover, H3K4me1
197 CUT&RUN on dCD cells largely recapitulated the H3K4me1 defects observed in DKOs (Fig.2F).
198 Diffbind uncovered only 272 total differential peaks between naive and formative in the catalytic
199 mutant background (Fig.2G).

200 The majority of H3K4me1 sites, whether shared, naive-enriched, or formative-enriched
201 were located at distal elements (intergenic or intronic) and not promoters (Fig.2H). To quantify
202 dependency of MLL3/4 H3K4me1 at the various sites, we set cutoffs for dependent (Log2 FC
203 naive DKO/WT < -1 or Log2 FC formative DKO/WT < -1) and independent H3K4me1 (Log2 FC
204 naive DKO/WT > 0.7 & Log2 FC form DKO/WT < 0.7). Of the distal H3K4me1 sites, only 15% of
205 shared sites (8632 of 64494) showed dependency on MLL3/4 for H3K4me1, while all naive and
206 formative specific sites were dependent. Heatmap visualization validated these categories and
207 similar behavior at dynamic sites during the transition was observed with analysis of published
208 naive and formative H3K4me1 ChIP-seq datasets (Fig.2J and Fig.S2E)(23,28). Furthermore,
209 evaluation of published naive MLL3/4 ChIP-seq data showed enrichment of the proteins at
210 H3K4me1 sites whether naïve-enriched or shared (12). Together, these data show that MLL3/4
211 are essential for a subset of H3K4me1 peaks including all dynamic sites and a small fraction of
212 shared sites.

213 Next, we asked about the impact of MLL3/4 loss on the transcription of genes closest to
214 the sites representing the different MLL3/4 dependent categories (shared, naive, and formative
215 H3K4me1 sites) (Fig.2K). Genes near MLL3/4 dependent shared H3K4me1 sites showed a clear
216 upregulation in expression in the transition from naive to formative in WT cells, suggesting these
217 enhancers may be primed in the naive state and became activated in the formative state. The
218 loss of MLL3/4 had no impact on the upregulation of the genes. Genes nearby MLL3/4 dependent
219 naive-specific H3K4me1 sites showed a small decrease in expression between naive and
220 formative in WT cells, which was not seen in the MLL3/4 DKO. Surprisingly, while genes nearby
221 MLL3/4 dependent formative-specific H3K4me1 showed a strong gain in expression during the
222 naive to formative transition, this gain was not diminished by the loss of MLL3/4. Together, these
223 results show that while MLL3/4 is required for all dynamic H3K4me1 and a subset of shared
224 H3K4me1 between the naive and formative state, this requirement has minimal impact on nearby
225 gene transcription.
226
227 **MLL3/4 dependent and independent distal H3K27ac deposition**
228 MLL3/4 is thought to be important for the recruitment of histone acetyltransferase
229 P300/CBP and thus histone acetylation including at H3K27 at enhancers. To test this general
230 model at the genomic level, we conducted CUT&TAG(29) for H3K27ac in WT and DKO cells in
231 the naive and formative states. We performed Diffbind at all peaks called by SEACR and
232 intersected these with ATAC-seq peaks from naive and formative WT cells to further enrich for
233 active enhancers. A total of 12802 H3K27ac peaks were significantly down (naive H3K27ac
234 peaks) and 8924 peaks were significantly up (formative H3K27ac peaks) during the transition
235 (Fig.3A,B). Another 12205 peaks did not show significant change in H3K27ac between the two
236 cell states (shared H3K27ac peaks). Loss of MLL3/4 resulted in a dramatic decrease in the
237 dynamic changes in H3K27ac normally seen in the WT cells (Fig.3C). However, unlike with
238 H3K4me1, many sites still showed some change including sites that appeared to be unique to

239 each pluripotent state. Diffbind using DKO samples uncovered 4295 naive-enriched, 1935
240 formative-enriched, and 23454 shared peaks, a reduction compared to WT, but still suggestive of
241 cell state dependent dynamic acetylation (Fig.3D). In both WT or DKO cells, the majority of
242 H3K27ac was located at distal sites, including both shared and naive/formative-enriched peaks.
243 Most of the cell-state specific H3K27ac sites were dependent on MLL3/4 (Log2 FC DKO/WT < -
244 1 or Log2 FC DKO/WT > -0.7 respectively) with MLL3/4 dependent sites accounting for 22% of
245 shared (2661 of 12205), 57% at naive-enriched (5717 of 9963), and 62% at formative-enriched
246 sites (4495 of 7205) (Fig.3F). Heatmap visualization confirmed these findings (Fig.3G,H,
247 Fig.S3A). Interestingly though, metagene analysis identified a low-level persistent signal and
248 premature H3K27ac signal at MLL3/4 independent naive and formative distal sites respectively.
249 Analysis of published ChIP-seq for H3K27ac and the acetyltransferase P300 in naive and
250 formative cells, showed commensurate binding at the sites with correlated increases and
251 decreases in signal with H3K27ac signal in WT cells (Fig.S3B)(23,28). These data uncover both
252 MLL3/4 dependent and independent mechanisms of P300 acetylation of H3K27 at cell-type
253 specific enhancer sites.

254 Given that H3K27ac is thought to be a marker of active enhancers, we next asked how
255 changes in H3K27ac associated with MLL3/4 impacted transcription at neighboring genes. To do
256 so, we identified the nearest neighbor transcription start site (TSS) to all distal sites where
257 H3K27ac was dependent on MLL3/4. This included H3K27ac peaks normally shared, naive-
258 enriched, or formative-enriched. We then evaluated expression of genes associated with the
259 TSSs (Fig.3I). Shared peaks were associated with genes that normally increased in expression
260 during the transition; this increase was not impacted by MLL3/4 loss. Naive-enriched peaks were
261 associated with genes that show a slight reduction during the transition and that reduction
262 occurred prematurely in the MLL3/4 DKO cells. Strikingly, there was a strong upregulation in gene
263 expression nearby sites that normally show MLL3/4 dependent increases in H3K27ac, but this
264 upregulation was not impacted by MLL3/4 loss. These data show that the upregulation of gene

265 expression associated with a gain of H3K27ac at nearby enhancer sites is largely independent of
266 that acetylation, while maintenance of gene expression associated with naive enriched H3K27ac
267 sites is at least in part dependent on maintenance of that acetylation.
268
269 **Enhancer Activation Can Occur Independently of MLL3/4**
270 The data above suggested similar trends in terms of the effects of MLL3/4 on H3K4me1,
271 H3K27ac and nearby gene expression, consistent with a partial dependency on MLL3/4
272 recruitment for H3K27ac deposition. To investigate this dependency, we focused on all distal sites
273 that show changes in H3K27ac in the context of their H3K4me1 status. We first separated all
274 chromatin accessible H3K4me1 sites (ATAC+) into those that lose, gain, or have shared
275 methylation during the naive to formative transition (Fig.S4A). Next, we subdivided the H3K4me1
276 categories into ones that lose, gain, maintain, or never have coincident H3K27ac (Fig.4A). In
277 general, most H3K4me1 sites never showed H3K27ac. However, among the 5960 sites that lose
278 H3K4me1 during the transition, 1370 also lose H3K27ac, and 913 maintain acetylation. Among
279 5583 sites that gain H3K4me1, 1172 sites also gain H3K27ac. We identified few sites (<150) for
280 the remaining categories of H3K27ac at sites that either gain or lose H3K4me1. Among 37589
281 sites maintaining H3K4me1, 2264 sites gain acetylation, 3498 sites lose acetylation, 4353 sites
282 show maintained acetylation. This suggests, although sites with dynamic H3K4me1 tended to
283 have dynamic H3K27ac in the same direction (gain or loss), most H3K27ac dynamics during the
284 naive to formative transition typically occur at sites with preexisting H3K4me1.
285 To understand how these subgroups were impacted by the loss of MLL3/4, we visualized
286 the data as heatmaps and filtered for MLL3/4 independent or dependent H3K27ac (Log2FC
287 DKO/WT > -0.7 or Log2FC DKO/WT < -1 respectively). Several important findings were
288 uncovered by these efforts. Among the sites that are enriched for both H3K4me1 and H3K27ac
289 (H3K4me1+/H3K27ac+) in the naive state, the loss of MLL3/4 resulted in a consistent loss of
290 monomethylation as seen across all naive enriched H3K4me1 sites (Fig.4B). However, only 61%

291 of these sites lost H3K27ac (891 of 1452 sites). Among sites that gain H3K4me1 and H3K27ac
292 during the transition from naive to formative, again all H3K4me1 is dependent on MLL3/4, while
293 77% also lost H3K27ac (901 out of 1172 sites)(Fig.4C). To confirm that the naive and formative
294 H3K4me1+/H3K27ac+ sites were direct MLL3/4 targets, we analyzed published naive ChIP-seq
295 of MLL3/4 and generated CUT&RUN for MLL4 in the formative state (Fig.4B,C). Naive MLL3/4
296 ChIP-seq signal was enriched at naive sites and MLL4 CUT&RUN was enriched specifically at
297 formative sites. Together these analyses show H3K27ac maintenance and de novo deposition
298 can occur at enhancer sites independent of MLL3/4 including at sites requiring MLL3/4 for
299 H3K4me1 deposition.

300 Next, we analyzed the sites that normally retain monomethylation during the transition
301 (H3K4me1 shared sites), which included a mix of sites that did or did not require MLL3/4 for
302 H3K4me1 (Fig.2J). Among the MLL3/4 independent H3K4me1 sites, 2039 sites normally gained
303 H3K27ac, 2357 sites lost H3K27ac, and 3611 sites had H3K27ac that did not change. An
304 additional 18135 sites never showed H3K27ac. For sites that gained H3K27ac, 48% required
305 MLL3/4 to do so even though the same enzymes were not required for the monomethylation at
306 those sites (Fig.4D). The remaining 52% showed strong H3K27ac in formative cells independent
307 of MLL3/4 and most showed premature H3K27ac deposition in naive DKO cells. MLL3/4 signal
308 was low at all these sites consistent with the unaffected H3K4me1 which suggested the loss of
309 acetylation at these sites was due to indirect effects. Similar to the formative H3K27ac sites, naive
310 H3K27ac at shared H3K4me1 sites also demonstrated mixed dependency on MLL3/4 with only
311 844 of the sites losing H3K27ac (Fig.S4B). Among the remaining 1320 sites, which retained high
312 levels of naive H3K27ac, the reduction in acetylation upon differentiation was incomplete in DKOs.
313 In contrast to the formative sites, MLL3/4 signal was strong at the naive dependent sites. In
314 MLL3/4 dependent shared H3K4me1 sites (2113 sites) few sites gained or had shared H3K27ac
315 (225 and 742, respectively). Despite losing H3K4me1 these sites showed mixed dependency of
316 H3K27ac on MLL3/4. Notably, sites with naive enriched H3K27ac comprised the largest group

317 (1141 sites) and most of these sites (973 of 1141 sites, 85%) were dependent on MLL3/4 for
318 H3K27ac. Finally, H3K4me1 independent shared sites showed no appreciable MLL3/4 binding
319 despite a loss of H3K27ac in DKO cells at 992 of 3297 sites, implying indirect effects (Fig.S4D).
320 Taken together, these data show that H3K27ac deposition at cell-type specific sites is not
321 necessarily coupled with binding or activity of MLL3/4. Indeed, we identified almost as many
322 exceptions as examples for the canonical model coupling MLL3/4 and H3K27ac deposition at
323 enhancers.

324 Next, we asked how the loss of MLL3/4 impacted H3K4me1 and H3K27ac at well-known
325 functionally validated enhancers regulating genes important for pluripotent cell identity including
326 Sox2, Nanog, Fgf5, and Oct6(28,30–34). Sox2 and Nanog expression normally decrease, while
327 Fgf5 and Oct6 expression increase during the naive to formative transition. In the MLL3/4 DKOs,
328 Sox2 expression failed to decrease during the transition, while the other genes showed normal
329 patterns of expression (Fig.1G, Fig.S4E). However, the impact of MLL3/4 loss on H3K4me1 and
330 H3K27ac at associated enhancers for each of these genes varied (Fig.4E). The Sox2 enhancer
331 cluster showed stable H3K4me1 and naive specific H3K27ac, both of which were dependent on
332 MLL3/4. Nanog has at least two known enhancers, all normally showing stable H3K4me1 and a
333 loss of H3K27ac during the transition; none of which were impacted by MLL3/4 loss. Fgf5 has 4
334 known enhancers, all normally showing increases in H3K4me1 and H3K27ac during the
335 transition. In the DKO cells, all four enhancers showed reduced H3K4me1, with two showing
336 normal upregulation of H3K27ac and two showing reduced H3K27ac in the formative state. Oct6
337 has one known enhancer, which normally shows a gain in H3K4me1 and H3K27ac during the
338 transition; H3K4me1 was reduced, but H3K27ac was unchanged in DKO cells. These examples
339 of well-known enhancers demonstrate the variable nature of dependency on MLL3/4 for
340 H3K4me1 and H3K27ac, including a surprising dispensability for de novo activation of enhancers
341 for Fgf5 and Oct6 expression.

342 We hypothesized that dependency of H3K4me1/H3K27ac on MLL3/4 might vary based
343 on the TFs bound at each enhancer. Therefore, we performed differential motif analysis using
344 Gimmemotifs(35) to determine TF binding motifs enriched at enhancer sites dependent on
345 MLL3/4 for H3K4me1 and/or H3K27ac (Fig.S5A). This analysis identified strong enrichment for
346 the binding motifs for TCF7L2, GRHL, and ZIC1 at sites requiring MLL3/4 for the gain of both
347 H3K4me1 and H3K27ac during the naive to formative transition. Expression of each of these TFs
348 is maintained or increased during the transition, and GRHL2 is essential for the gain of H3K4me1
349 and H3K27ac at hundreds of formative specific enhancers(23). At MLL3/4 dependent naive
350 peaks, we identified a strong enrichment for multiple TF binding motifs including RXRA, ZBTB20,
351 ZNF257, MYC, ESR2, and NR6A1, all of which are either expressed or have closely related TFs
352 that are expressed in naive cells (Fig.S5B). Interestingly, the motifs for the critical TF regulators
353 of the naive and/or formative states such as Oct4, Sox2, Nanog, KLFs, Oct6, and Otx2, were
354 generally not uncovered. This analysis suggests that the activity of only a small number of TFs is
355 dependent on MLL3/4 for maintaining and/or establishing cell state specific H3K4me1
356 independent of H3K27ac.
357

358 **Distal H3K4me1 and H3K27ac are not functionally coupled with formative transcriptional**
359 **activation**

360 We next asked whether focusing on sites of H3K27ac change in the context of dynamic
361 H3K4me1 peaks might enrich for sites near genes whose expression was impacted by MLL3/4.
362 Similar to analysis of H3K27ac alone, genes nearby naive H3K4me1+/H3K27ac+ sites with
363 MLL3/4 dependent H3K27ac showed reduced expression in naive cells (Fig.4B,Fig.S6A) while
364 genes nearby MLL3/4 independent H3K27ac were unchanged, supporting a role for acetylation
365 in maintaining expression at these sites. In contrast, in the formative state there was little impact
366 on transcription of genes nearby either MLL3/4 independent or dependent H3K27ac

367 (Fig.4C,Fig.S6B). This suggested gain of both H3K4me1 and H3K27ac at formative sites was not
368 associated with formative transcriptional activation.

369 To expand on this surprising finding, we used ChromHMM as an unbiased approach to
370 look at all potential combinations of H3K4me1, H3K27ac in WT and DKO cells in the naive and
371 formative states. We modeled 16 chromatin states to segment all possible combinations of these
372 states at distal WT ATAC peaks separately in naive and formative cells (Fig.5A,C). The sites were
373 then mapped to nearby TSS, and the impact of MLL3/4 on expression of these genes was
374 measured (Fig.5B,D). In the naive state, genes near sites that lost H3K27ac showed highly
375 significant downregulation whether or not they also lost H3K4me1 (States 1,15). These sites also
376 showed strong MLL3/4 signal supporting a direct role for the enzymes in coordinating both
377 acetylation and transcription. In the formative state, only genes near distal sites that lost both
378 H3K4me1 and H3K27ac showed a significant reduction in gene expression (State 2) (Fig.5D).
379 This state represented a relatively small number of enhancers (1040 sites) linked to only 154
380 genes. Therefore, this unbiased approach at defining chromatin states again showed that even
381 though MLL3/4 loss dramatically alters the landscape of enhancer activity states, the impact on
382 gene expression changes that occur during the naive to formative transition is relatively minor.

383 Genes can be regulated by multiple enhancers(36,37). Thus, we hypothesized that
384 MLL3/4 independent enhancers may compensate for loss of MLL3/4 dependent enhancers
385 explaining the underwhelming impact on expression. To test this hypothesis, we developed a
386 metric where for every gene we multiplied the number of nearby H3K4me1+/H3K27ac+ enhancer
387 sites by the number of sites that lost H3K27ac in the MLL3/4 DKO cells. In naive cells, this metric
388 showed a correlation with the change in nearby gene expression (Fig.5E). In formative cells, only
389 a very small effect was seen and only for genes with the greatest numbers of impacted enhancers.
390 The contrast between naive and formative was even more striking when only considering sites
391 within 10kb of the TSS (Fig.5F). We obtained similar results by analyzing either the proportion of
392 enhancers that were lost per gene in each pluripotent state or the number of enhancers lost per

393 gene (Fig.S6C,D). Therefore, compensation by non-MLL3/4 dependent enhancers does not
394 appear to be a major factor rescuing gene expression in formative cells.
395
396 **Gene-centric analysis reveals a subset of distal loci associate with MLL3/4 dependent**
397 **formative genes**
398 Our data showed a requirement for MLL3/4 for H3K4 monomethylation at all sites that normally
399 gain or lose the mark during the naive to formative transition (Fig.2D,J,Fig.6A). They also showed
400 a requirement for MLL3/4 for H3K27ac at 66% of sites that gain and 66% that lose acetylation
401 during the transition (Fig3G,H, Fig.6A, Fig.S7A). However, only 16% of formative genes that
402 normally increase in expression in the transition and only 39% naive genes that are normally
403 downregulated were impacted by loss of MLL3/4. Among the naive genes that were impacted,
404 there was a strong enrichment for cytokine pathways (Fig.S7B). Among the formative genes that
405 were impacted, there was an enrichment for development related pathways including
406 regionalization, migration, and morphogenesis (Fig.6B). This later finding suggested a role for
407 initiating transcription of genes involved in a later stage of development, specifically gastrulation.
408 Our previous analyses utilized an enhancer-centric approach to identify changes in nearby
409 genes. We next asked if impacted genes could be connected to nearby enhancers that showed
410 loss of H3K27ac. To take this gene-centric approach, we categorized all genes that were normally
411 upregulated during the transition to the formative state into those that were not impacted
412 (independent) vs. impacted (dependent) by MLL3/4 loss. Among formative genes, this resulted in
413 820 independent formative genes and 131 dependent formative genes which could be linked to
414 2693 and 402 nearby H3K4me1/H3K27ac+ enhancers respectively. As a background control, we
415 linked 12506 expressed genes from our RNA-seq to an associated 37362 H3K4me1/H3K27ac+
416 peaks. Both the dependent and independent formative genes normally showed an increase in
417 H3K27ac at nearby enhancer sites during the naive to formative transition (Fig.6C). This increase
418 in acetylation was mostly lost in DKO cells, whether or not the nearby genes failed to increase

419 gene expression. Similarly, both dependent and independent naive genes (normally
420 downregulated during transition) showed a decrease in nearby enhancer H3K27ac during the
421 naive to transition. However, H3K27ac at enhancers nearby dependent naive genes was more
422 impacted than those nearby independent genes in the DKO cells (Fig.S7C). Therefore, the gene-
423 centric approach was consistent with enhancer-centric approach showing a more prevalent role
424 of MLL3/4 regulated chromatin changes on the maintenance of naïve gene expression than the
425 gain of formative gene expression.

426 To look deeper, we investigated the chromatin states nearby the small number of genes
427 that were dependent on MLL3/4 for upregulation of transcript expression during the naive to
428 formative transition by creating genome tracks (Fig.6D). These dependent genes included
429 important gastrulation genes such as *T*/*Brachyury*, *Fgf8*, and *Lefty1*, the latter of which is a
430 previously identified target of MLL3/4 (Wang et al. 2016). All three of these genes had one or
431 more nearby enhancers that showed a loss of H3K4me1 and H3K27ac in MLL3/4 DKO formative
432 cells. Thus, the expression of these key regulators relies on enhancers which require MLL3/4 for
433 proper spatiotemporal regulation. However, they represent a very small subset of the thousands
434 of enhancers that lose active marks upon the loss of MLL3/4. All together, these data show that
435 from both a gene-centric or enhancer-centric perspective, H3K4me1 and/or H3K27ac at distal
436 sites are largely uncoupled from target gene activation during the naïve to formative transition
437 with the loss of MLL3/4.

438

439 **Discussion**

440 Enhancers are thought to play a central role in gene regulation during cell fate transitions.
441 Yet, the molecular epistasis of enhancer activation is incompletely understood. The prevailing
442 model proposes that H3K4me1 deposited by MLL3/4 precedes H3K27ac deposition by P300/CBP
443 and these events stimulate transcription. By combining state-of-the-art CUT&RUN/CUT&TAG
444 technologies and knockouts of both MLL3 and MLL4, we evaluated this model during the well-

445 defined transition from naive to formative pluripotency, where hundreds of genes are both
446 silenced and activated. Surprisingly, the majority of these gene expression changes including de
447 novo expression of many genes were unaffected by MLL3/4 loss. In contrast, loss of MLL3/4 led
448 to large effects on the active histone modification signatures at thousands of enhancers. These
449 results suggest that MLL3/4 and its orchestration of post-translational histone marks, play a
450 relatively minor role in gene regulation during early ESC differentiation in contrast to other
451 regulatory mechanisms.

452 MLL3/4 are thought to be the major H3K4 monomethyltransferases in mammals(1,2).
453 However, we find that the majority of H3K4me1 is unaffected by MLL3/4 loss in both the naive
454 and formative states. Consistent with our findings, similar results have been previously described
455 in naive ESCs(12). However, investigating the transition to the formative state, we were able to
456 make several novel discoveries. We discovered that all distal sites that normally show dynamic
457 changes in H3K4me1, either gained or lost, were fully dependent on MLL3/4 for monomethylation.
458 In contrast, only a small fraction of the sites that normally do not change during the transition
459 showed any requirement for enzymes in maintaining monomethylation levels. Therefore, there
460 appear to be at least two distinct types of cis-regulatory elements with regard to MLL3/4
461 dependency. This raises important follow-up questions such as what enzymes are responsible
462 for the predominant group of H3K4me1 sites that do not require MLL3/4 for maintenance? MLL2
463 has been suggested as an alternative monomethyltransferase in naive ESCs, but only at a small
464 number of sites that co-exist with H3K27m3, a repressive mark(12,38). MLL1 represents another
465 potential candidate. Interestingly, MLL1 has been shown to inhibit the reprogramming of primed
466 pluripotent (EpiSCs) to naive ESCs(39). However, its role in regulating H3K4me1 in naive cells
467 during differentiation has not been evaluated. Other candidates include SMYD1 and SMYD2
468 which are SET domain containing proteins that have previously been shown to be required for
469 H3K4me1 and or H3K4m2 at de novo enhancers during macrophage activation(40). Additionally,
470 whether there are distinct roles for the MLL3/4 dependent vs. independent distal sites remains

471 unclear. Given most enhancer activation in the absence of MLL3/4 occurs at these independent
472 H3K4me1 sites they may represent a collection of primed sites distinct from MLL3/4. Surprisingly
473 though, previous studies have shown that at least in the case of MLL3/4, the proteins, but not
474 their enzymatic activity are important in CBP/P300 recruitment and H3K27ac(12). Therefore, it
475 will be important to understand the link between the factors resulting in methylation of these
476 MLL3/4 independent sites and those that deposit H3K27ac at these sites.

477 MLL3/4 are thought to be important recruiters of CBP/P300, whose catalytic byproduct
478 H3K27ac is a key marker of enhancer activation. In our data, we identify that more than half of
479 H3K27ac distal sites lose acetylation with MLL3/4 loss, with a greater proportion at dynamic sites
480 than unchanging sites. Many sites that lose H3K27ac do so independent of H3K4me1 loss.
481 Whether this represents an indirect role or a direct role for MLL3/4 in the recruitment of CBP/P300
482 at sites that also retain H3K4me1 is unclear. However, the presence of MLL3/4 binding at many
483 of these sites supports a direct role. In addition to the MLL3/4 dependent sites, there is also a
484 large number of MLL3/4 independent H3K27ac sites, including many occurring at regions that do
485 require MLL3/4 for H3K4me1. This indicates multiple mechanisms for CBP/P300 recruitment to
486 enhancers that are differentially used depending on the specific enhancer and chromatin state.
487 How different enhancers recruit CBP/P300 using distinct mechanisms is unclear but is likely in
488 part due to the underlying transcription factors binding to those enhancers. Enhancer
489 subcategories consisting of varying types of TFs have recently been shown to require specific
490 chromatin complexes for enhancer activity(41). Consistent with the role for different TFs
491 underlying these requirements, TF motif enrichment analysis uncovered several TFs whose
492 motifs were highly enriched at each distinct category of MLL3/4 dependent and independent
493 enhancers. CBP/P300 could either be directly recruited by these TFs and/or be recruited through
494 other factors compensating for loss of MLL3/4. Future studies are necessary to fully elucidate the
495 requirements for recruitment of specific epigenetic complexes by distinct TFs.

496 Ultimately, it is thought that enhancer activation as defined by the gain of H3K27ac drives
497 increased expression of its target gene. Surprisingly though, while we observe major impacts on
498 the chromatin state of sites with dynamic H3K4me1 and H3K27ac during the naive to formative
499 transition, there are relatively modest changes in the gain and loss of gene expression. Indeed,
500 there is a greater impact of MLL3/4 loss in the silencing of genes rather than the activation of
501 genes during the transition. The activation of the formative transcriptional program is largely
502 unaffected, with only a few exceptions such as Lefty1 and Fgf8. The key markers of the formative
503 state including Otx2, Oct6, and Fgf5 are all normally upregulated. This unexpected finding raised
504 the question whether there was any specific MLL3/4 dependent enhancer phenotype in formative
505 cells that could be linked to a failure to activate gene expression. Analysis of genes either nearby
506 MLL3/4 dependent formative H3K4me1 sites or MLL3/4 dependent formative H3K27ac sites
507 during the transition did not identify a reduction in gene activation. Most surprising, interrogation
508 of genes nearby enhancers that fail to gain both H3K4me1 and H3K27ac also showed normal
509 increases in gene expression. Only when considering the small number of sites where both
510 H3K4me1 and H3K27ac were depleted, independent of their dynamics, was there a small but
511 significant effect for a handful of genes. Consistent with our data, recent work mutating H3 and/or
512 H3.3 Lysine 27 sites to Arginine, which cannot be acetylated did not impact either naive gene
513 expression or the activation of the formative program in differentiating ESCs(14,15). However,
514 these findings are interpreted as a lack of a requirement for H3K27ac rather than an inability to
515 recruit enhancer CBP/P300 at large. Either way, it is unclear whether many of the sites dependent
516 on MLL3/4 for active histone marks are dispensable for gene expression changes and or these
517 sites are functional enhancers without an active histone signature.

518 MLL4 KO embryos die shortly following gastrulation but show defects starting in early
519 gastrulation(6,21). In contrast, MLL3 KO embryos die perinatally due to developmental defects in
520 the lung. MLL3/4 double KO embryos phenocopy the MLL4 KO consistent with the MLL4 being
521 able to compensate for MLL3 loss in early development. The direct molecular basis for these

522 phenotypes is poorly understood. However, it is important to note that even though MLL3/4 loss
523 had little impact on gene activation during the ESC transition from naive to formative pluripotency,
524 many genes were dysregulated in steady state naive ESCs and remained so in the formative
525 state. In the case of persisting naive genes like Zfp42/Rex1 and Klf4, this is consistent with
526 previous reports that MLL4 KOs fail to repress naive genes in LIF cultured cells after withdrawal
527 of 2i from naive (LIF+2i) culture conditions(42). Some misregulated genes are likely direct targets
528 given the downregulation of genes nearby MLL3/4 dependent enhancers upon loss of the
529 enzymes. In addition, there are likely indirect effects potentially including additional unknown non-
530 transcription related functions for MLL3/4. It is also important to note that other studies have
531 shown important roles for MLL3/4 in the upregulation of gastrulation markers in the setting of
532 embryoid body differentiation, a model for a later stage in early development than studied
533 here(22). Even later in development, MLL3/4 is important for developmental programs in a variety
534 of tissues including cardiac, neural crest, muscle, and adipose specific programs(6,7,43,44).
535 Therefore, it may be that transcriptional roles of MLL3/4 are limited to determinants of
536 differentiation pathways post-pluripotency and is most important in specifying somatic cell fates.
537 If that is the case, it will be important to understand the fundamental differences between
538 transcriptional control of somatic cells and pluripotent cells or between lineage determining genes
539 and genes defining other developmental processes.
540

541 **Conclusions**

542 In sum, our data demonstrates that cell-type specific enhancer activation is not uniquely
543 performed by MLL3 and MLL4. Instead, gene regulation appears to be partitioned during the naive
544 to formative transition. MLL3/4 is required for a small subset of genes critical for gastrulation and
545 later developmental processes, but formative transcriptional activation is mediated by other
546 mechanisms including MLL3/4 independent enhancers (Fig.6E). By identifying rules and
547 exceptions to the current model of enhancer activation we highlight current gaps in knowledge

548 concerning the molecular steps regulating these processes. Further investigation into the
549 molecular interplay and steps governing enhancer function will advance our understanding of
550 gene regulation in health and gene dysregulation in disease.
551

552 **Methods**

553 **Cell Culture**

554 Mouse ESCs were cultured in Knockout DMEM (Thermo Fisher, CAT#10829018) supplemented
555 with 15% FBS, L-Glutamine, Penicillin/Streptamycin, NEAA, LIF (1000U/mL), and 2i (1uM MEK
556 inhibitor PD0325901 and 3uM GSK3 inhibitor CHIR99021). Formative cells were generated by
557 removal of LIF and 2i. Briefly, 5e4 ESCs were plated per well of a 6 well plate on day −1 in LIF+2i
558 media. To initiate differentiation, LIF and 2i were removed 24 hours after seeding (Day 0).
559 Formative cells were collected on day 3 of differentiation, 63 hours after removal of LIF and 2i. To
560 overcome proliferation defects 7.5e4 MLL3/4 DKO cells were plated per well of a 6 well. Naive
561 cells were passaged and staged appropriately for simultaneous harvest. The expected amino acid
562 substitutions in MLL3/4 dCD cells were validated using Sanger sequencing on amplicons
563 generated from genomic PCR. To generate DKOs from MLL3KO/MLL fl/fl ESCs we transfected
564 with Rosa26-CRE-ERT2 plasmid linearized by EcoRI, selected with puromycin, and genotyped
565 using genomic PCR. We then added tamoxifen for 3 days and picked clones to validate knockout
566 of MLL4 floxed allele. Primers for genotyping listed in the supplemental table.
567

568 **Crystal Violet Assay**

569 To assess proliferation cells were plated in 24 well plates at approximately 12.5e4 cells per well.
570 After 24, 48, and 72 hours, cells were then washed with PBS and 200ul Crystal Violet (0.2%
571 Crystal Violet, 2% ethanol in dH2O) was added to the plate for 10 minutes at room temperature.
572 Cells were then washed twice by gently submerging plate in tap water before 600ul 1% SDS was

573 added to the well to solubilize the stain. Plate was placed on a shaker until color was uniform in
574 the well and 200ul was transferred to a 96 well plate for reading absorbance at 570nm with a plate
575 reader (SpectraMax M5). Absorbance for each time point was normalized to a blank well.
576
577 **qPCR and analysis**
578 To perform qPCR we first extracted RNA by adding Trizol directly to plates. After adding
579 chloroform, an isopropanol precipitation with GlycoBlue was performed followed by ethanol
580 washes. RNA pellets were resuspended in RNAse-free water and quantified using a NanoDrop.
581 200ug of RNA was used for cDNA synthesis using Maxima First Strand Synthesis Kit (Thermo
582 Fisher, CAT#K1672) with half reactions according to the manufacturer's protocol. qPCR on cDNA
583 was performed using SYBRgreen master mix (Applied Biosystems, CAT#A25742) using 6ul final
584 volume on a QuantStudio5 qPCR machine (Applied Biosystems). qPCR primers for targets are
585 listed in the supplemental table. Target Ct values were normalized to GAPDH internally for each
586 sample and then set relative to the naive WT negative control.
587
588 **Protein extraction and Westerns**
589 To harvest protein for western assays, cells were trypsinized, washed once with ice cold PBS
590 before adding RIPA buffer with protease inhibitors (Sigma CAT#P8340). After 15 minutes on ice,
591 lysed cells were centrifuged at 16000g for 10 minutes and the supernatants representing whole
592 cell fractions were collected and snap frozen in liquid nitrogen. Protein quantification was
593 conducted using a Micro BCA protein assay kit (Thermo CAT#23235). 40ug of protein was loaded
594 per well of SDS PAGE gels for western.
595
596 Westerns were typically conducted using a Bio-Rad system with Tris-Glycine gels purchased from
597 Bio-Rad and transferred to methanol activated PVDF membranes. Membranes were blocked, and

598 stained with primaries and secondaries using Licor Odyssey Blocking Buffer (Li-Cor, CAT#927-
599 60001) mixed 1:1 with TBS. Primary antibodies for western were Rabbit anti-MLL3 (provided by
600 Kai Ge), Rabbit anti-MLL4 (provided by Kai Ge), Mouse anti-OCT4 (BD Biosciences,
601 CAT#611202), and Rabbit anti-NANOG (Cell Signaling, CAT#61419)
602
603 High molecular weight westerns for MLL3 and MLL4 were conducted using a NuPage SDS-PAGE
604 and Transfer system with an XCell electrophoresis unit according to manufacturer's protocols with
605 several modifications (Thermo Fisher, CAT#E10002). In brief, 40ug of whole cell protein lysate
606 was incubated with LDS loading buffer (Thermo Fisher, CAT#NP0007, added BME to 1% final
607 concentration) and incubated at 70C for 10min. Samples were loaded into 3-8% Tris-acetate gels
608 (Thermo Fisher, CAT#EA03752BOX) and ran in sample running buffer (Thermo Fisher,
609 CAT#LA0041) supplemented with NuPage Antioxidant (Thermo Fisher, CAT#NP0005). Samples
610 were run at 80V for 30min and then 120V for 120min. Samples were then transferred to PVDF
611 membranes using NuPage Transfer Buffer (Thermo Fisher, CAT#NP0006, with added 10%
612 Methanol, NuPage Antioxidant, 0.01% SDS) at 30V in a cold room for either 2 hours or overnight.
613
614 **RNA sequencing**
615 Total RNA was extracted and purified from cells using Trizol followed by ethanol precipitation.
616 RNA-seq libraries were generated using the QuantSeq 3' mRNA-Seq Library Prep Kit FWD for
617 Illumina (Lexogen, CAT#A01172) according to their protocol using 200ng of total RNA for input.
618 We utilized the PCR Add-on kit for Illumina (Lexogen, CAT#M02096) to determine an appropriate
619 number of PCR cycles to amplify libraries. Amplified libraries were quantified using Agilent
620 Tapestation 4200. Libraries were pooled and sequenced using a HiSeq 4000 to obtain single end
621 50bp reads. At least 10 million mapped reads or more per sample were obtained.
622

623 **RNA sequencing processing and analysis**

624 Adapters for sample FastQ files were trimmed using cutadapt v2.5 followed by alignment to the
625 mm10 genome using STAR align 2.7.2b. Featurecounts from Subread v1.6.4 was used to
626 generate a counts matrix of reads per gene. FastQC v0.11.8 and MultiQC were used to validate
627 quality sequencing and mapping.

628

629 The gene count matrix was converted to CPM and filtered for genes greater than 1 cpm in at least
630 two total samples. Samples were normalized using TMM. Next, Log2 CPM averages were
631 calculated for replicates of each sample for scatterplot visualization and nearest neighbor TSS
632 analysis. All transcriptomics analyses were conducted using CPM values from TMM normalization
633 of all samples except for nearest neighbor analysis where WT and DKO were TMM normalized
634 together. To conduct differential gene expression we performed DESeq2 v1.34.0 analysis using
635 the raw gene count matrix as input for each desired comparison. Gene ontology was performed
636 using ClusterProfiler 4.2.2. Custom R code for other downstream transcriptomics analyses and
637 visualization provided on Github.

638

639 **CUT&RUN sequencing and processing**

640 CUT&RUN was conducted using the protocol from Skene et al. 2018 with the following
641 modifications: Freshly trypsinized cells were bound to activated Concanavalin A beads (Bang
642 Laboratories, #BP531) at a ratio of 2e5 cells/10ul beads in CR Wash buffer (20mM HEPES,
643 150mM NaCl, 0.5mM Spermidine with protease inhibitors added) at room temp. An input of 2e5
644 cells were used per target. Bead-bound cells were then incubated rotating overnight at 4C in CR
645 Antibody buffer (CR Wash with 0.05% Digitonin, 2mM EDTA) containing primary antibody. We
646 used the following antibodies for CUT&RUN: 1:100 Rabbit anti-H3K4me1 (Abcam, ab4729),
647 Rabbit anti-MLL4 1:100 (provided by Kai Ge), and 1:100 Rabbit IgG isotype control (Abcam,

648 ab171870). After primary, we washed 3 times 5 minutes each with cold CR Dig-wash buffer (CR
649 Wash with 0.05% Digitonin) and incubated with pA-MNase (1:100 of 143ug/mL provided by Steve
650 Henikoff) for 1 hour rotating at 4C. After MNase binding, we washed 3 times 5 each with cold CR
651 Dig-wash buffer, and chilled cells down to 0C using a metal tube rack partially submerged in an
652 ice water slurry. MNase digestion was induced by adding CaCl2 at a final concentration of 2mM.
653 After 30 minutes of digestion, the reaction was quenched using Stop Buffer containing 340mM
654 NaCl, 20mM EDTA, 4mM EGTA, 0.05% Digitonin, 100ug/mL RNAse A, 50ug/mL Glycogen, and
655 approximately 2pg/mL Yeast spike-in DNA (provided by Steve Henikoff). The digested fragments
656 for each sample were then extracted using a phenol chloroform extraction. Library preparation on
657 samples was conducted using manufacturer's protocols for NEBNext Ultra II Dna Library Prep Kit
658 (New England BioLabs, CAT#E7645) and NEB Multiplex Dual Index oligos (New England
659 BioLabs, CAT#E7600, #E7780) with the following modifications. We input approximately 10ng of
660 sample for half reactions, we diluted the NEBNext Illumina adaptor 1:25, we used the following
661 PCR cycling conditions: 1 cycle of Initial Denaturation at 98C for 30 seconds, 12+ cycles of
662 Denaturation at 98C for 10 seconds then Annealing/Extension at 65C for 10 seconds, and 1 final
663 cycle of extension at 65C for 5 minutes. Following library preparation, double size selection was
664 performed using Ampure beads and quality and concentration of libraries were determined by an
665 Agilent 4200 Tapestation with High-Sensitivity D1000 reagents before pooling for sequencing.
666
667 Fastq files for CUT&RUN samples were processed using Nextflow(45) and the nf-core CUT&RUN
668 pipeline v1.0.0 beta. In brief, adapters were trimmed using Trim Galore. Paired-end alignment
669 was performed using Bowtie2 and peaks were called using SEACR with a peak threshold of 0.05
670 using spike in calibration performed using the E.coli genome K12. However, all downstream
671 analysis was performed using CPM normalized samples.
672

673 **CUT&TAG sequencing and processing**

674 CUT&TAG was conducted using the protocol from Kaya-Okur et al. 2020 with the following
675 modifications: freshly trypsinized cells were bound to Concanavalin A beads at a ratio of 2e5
676 cells/7ul beads in CR wash at room temp. We used 2e5 cells as input per sample. Bead-bound
677 cells were then incubated rotating overnight at 4C in CT Antibody buffer (CR Wash with 0.05%
678 Digitonin, 2mM EDTA, 1mg/mL BSA) containing primary antibody. We used the following primary
679 antibodies for CUT&TAG: 1:100 Rabbit anti-H3K27ac (Abcam, ab4729), 1:100 Rabbit anti-
680 H3K4m3 (Abcam, ab8580), 1:100 Rabbit IgG isotype control (Abcam, ab171870). After primary,
681 samples were washed 3 times for 5 minutes each using CR Dig-wash buffer and resuspended in
682 1:100 secondary antibody (Guinea pig anti-rabbit, Antibodies Online #ABIN101961) in CR Dig-
683 wash buffer at 4C for 1 hour rotating at 4C. Samples were then incubated for 1 hour at 4C with
684 50ul of approximately 25nM homemade pA-Tn5 in CT Dig300 wash buffer (20mM HEPES,
685 300mM NaCl, 0.01% Digitonin, 0.5mM Spermidine with Roche cOmplete protease inhibitors
686 added). Recombinant Tn5 was purified and loaded with adapters as previously described (Kaya-
687 Okur et al. 2019). After Tn5 incubation, samples were washed 3 times for 5 minutes each with CT
688 Dig300 wash buffer. Tagmentation was then initiated for 1hr at 37C in a thermocycler by adding
689 MgCl2 to 10mM final concentration in 50uL volume. The tagmentation reaction was quenched
690 immediately afterwards by adding 1.6ul of 0.5M EDTA, 1ul of 10mg/mL Proteinase K, and 1ul of
691 5% SDS. Samples were then incubated at 55C for 2 hours in a thermocycler to denature Tn5 and
692 solubilize tagmented chromatin. After incubation, samples were magnetized and the supernatant
693 was transferred to new wells where SPRI bead purification was performed using homemade
694 beads to select all DNA fragment lengths larger than 100bp. Samples were eluted in 0.1X TE and
695 approximately half of each sample was used for library preparation using NEBNext HIFI
696 Polymerase with custom indices synthesized by IDT. An appropriate number of cycles for each
697 target was chosen to prevent overamplification bias. After amplification, libraries were purified
698 with 1.2x homemade SPRI beads to select for fragments >250bp and eluted in 0.1X TE. Quality

699 and concentration of libraries were determined by an Agilent 4200 Tapestation with D1000
700 reagents before pooling for sequencing.
701
702 CUT&TAG samples were processed similarly to CUT&RUN samples using the same Nextflow
703 pipeline.
704
705 **ATAC-seq sequencing and processing**
706 ATAC-seq libraries were generated using the Active Motif commercial kit following the provided
707 protocol with kit components (Active Motif, CAT#53150): Specifically, freshly trypsinized cells
708 were washed with ice cold PBS using 75K cells as input for each sample. Cells were lysed using
709 ice cold ATAC Lysis Buffer and we added 50ul of Tagmentation Master Mix to each sample. We
710 incubated tagmentation reactions at 37C for 30min in a thermomixer set to 800 rpm. We
711 immediately transferred samples to a new tube and performed DNA column purification. After
712 purification we amplified the libraries using the following PCR conditions: 1 cycle at 72C for 5
713 minutes, 1 cycle at 98C for 30 seconds, and finally 10 cycles of 98C for 10 seconds, 63C for 30
714 seconds, 72C for 1 minute. Amplified libraries were purified using 1.2x SPRI bead solution. Quality
715 and concentration of libraries were determined by an Agilent 4200 Tapestation with D1000
716 reagents before pooling for sequencing.
717
718 Fastq files for ATAC-seq samples were processed using Nextflow and the nf-core ATAC-seq
719 pipeline v1.2.1. In brief, adapters were trimmed using Trim Galore. Paired-end alignment was
720 performed using BWA and peaks were called using MACS2 in broad peak mode with a cutoff
721 value of 0.1. BAM files were converted to bigWigs. Using read counts from consensus peaks of
722 WT and DKO samples we performed TMM normalization using EdgeR to generate scale factors
723 used during bigWig generation of samples.

724

725 **ChIP-seq processing**

726 Fastq files for ChIP-seq from previously published studies were processed using Nextflow and
727 the nf-core ChIP-seq pipeline v1.2.1. In brief, adapters were trimmed using Trim Galore. Paired-
728 end alignment was performed using BWA, duplicates were removed, and peaks were called using
729 MACS2 in broad peak mode with a cutoff value of 0.1.

730

731 **Peak analysis**

732 We performed differential signal enrichment analysis using Diffbind on SEACR peaks derived
733 from CUT&RUN or CUT&TAG samples. We then used the output from Diffbind, containing the
734 union of input SEACR peaks, for two packages: 1) We used Deeptools multiBigwigsummary to
735 count reads at each peak for the different targets/samples and 2) we used HOMER
736 annotatePeaks to identify nearest gene and annotate genomic features. These files were then
737 imported into R and joined together by peak coordinates to make a dataframe which we used to
738 filter for phenotypes and calculate fold changes. To calculate Log2CPM peak densities we took
739 counts for each peak from multibigwigsummary multiplied them by a constant (1e6), added 1, and
740 then performed Log2(x). Finally we divided by the width in basepairs of the peak. Fold changes
741 were conducted using these Log2CPM peak density values. H3K27ac CUT&TAG data was
742 subset by H3K27ac peaks that overlapped at least 75% of an WT ATAC-seq peak using bedops.
743 We also used bedops on ATAC subset H3K27ac sites to identify H3K27ac peaks that had
744 overlapping H3K4me1 peaks to generate our H3K4me1/H3K27ac+ peak lists. ChromHMM
745 segmentations were similarly filtered for chromatin states that overlapped at least 75% of an WT
746 ATAC-seq peak.

747

748 **Declarations**

749 **Ethics approval and consent to participate**

750 Not applicable

751

752 **Consent for publication**

753 Not applicable

754

755 **Availability of data and materials**

756 The data generated and their analysis in the current study are available in the GEO Database
757 using the identifier GSExxxxxx. Code generated for analysis will be hosted on Github or otherwise
758 available by request.

759

760 **Competing interests**

761 The authors declare that they have no competing interests.

762

763 **Funding**

764 This work was made possible by funding from the National Institute of General Medical Sciences
765 (grant no. R01GM122439). Additionally, R.M.B. was supported by the ARCS foundation and the
766 UCSF Discovery Fellowship.

767

768 **Author's Contributions**

769 Conceptualization: RMB, RB; Methodology: RMB; Validation: RMB; Formal Analysis: RMB;
770 Investigation: RMB, KC; Resources: RB; Data Curation: RMB; Writing - original draft: RMB;
771 Writing - review and editing: RMB, RB; Visualization: RMB; Supervision: RMB, RB; Project
772 Administration: RB; Funding Acquisition: RB.

773

774 **Acknowledgements**

775 We thank Kai Ge for providing the MLL3KO;MLL4 floxed ESC line and antibodies for MLL3 and
776 MLL4. We thank Joanna Wysocka for providing the MLL3/4 catalytic dead cell line. We thank
777 Steve Henikoff for sending purified pA-MNase. We would like to thank the Nextflow community
778 for helpful advice on processing of sequencing samples including Chris Cheshire and Charlotte
779 West for providing access to a beta for nf-core CUT&RUN pipeline. We would like to thank
780 Bryan Marsh, Deniz Gökbuget, and Brian Deveale for helpful discussions throughout the project
781 including proofreading the original manuscript. We would like to thank Dan Lim, Yin Shen, and
782 Elphege Nora for their thoughtful comments and constructive feedback on the manuscript.
783

784 **References**

785 1. Hu D, Gao X, Morgan MA, Herz HM, Smith ER, Shilatifard A. The MLL3/MLL4 branches of
786 the COMPASS family function as major histone H3K4 monomethylases at enhancers. Mol
787 Cell Biol. 2013 Dec;33(23):4745–54.

788 2. Sze CC, Shilatifard A. MLL3/MLL4/COMPASS Family on Epigenetic Regulation of Enhancer
789 Function and Cancer. Cold Spring Harb Perspect Med. 2016 Nov;6(11):a026427.

790 3. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, et al.
791 Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet.
792 2010 Sep;42(9):790–3.

793 4. Lavery WJ, Barski A, Wiley S, Schorry EK, Lindsley AW. KMT2C/D COMPASS complex-
794 associated diseases [KCDCOM-ADs]: an emerging class of congenital regulopathies. Clin
795 Epigenetics. 2020 Jan 10;12(1):10.

796 5. Mendiratta G, Ke E, Aziz M, Liarakos D, Tong M, Stites EC. Cancer gene mutation
797 frequencies for the U.S. population. Nat Commun. 2021 Oct 13;12(1):5961.

798 6. Lee JE, Wang C, Xu S, Cho YW, Wang L, Feng X, et al. H3K4 mono- and di-
799 methyltransferase MLL4 is required for enhancer activation during cell differentiation. eLife.
800 2013 Dec 24;2:e01503.

801 7. Lai B, Lee JE, Jang Y, Wang L, Peng W, Ge K. MLL3/MLL4 are required for CBP/p300
802 binding on enhancers and super-enhancer formation in brown adipogenesis. Nucleic Acids
803 Res. 2017 Jun 20;45(11):6388–403.

804 8. Wang SP, Tang Z, Chen CW, Shimada M, Koche RP, Wang LH, et al. A UTX-MLL4-p300
805 Transcriptional Regulatory Network Coordinately Shapes Active Enhancer Landscapes for
806 Eliciting Transcription. Mol Cell. 2017 Jul 20;67(2):308-321.e6.

807 9. Park YK, Lee JE, Yan Z, McKernan K, O'Haren T, Wang W, et al. Interplay of BAF and
808 MLL4 promotes cell type-specific enhancer activation. Nat Commun. 2021 Mar
809 12;12(1):1630.

810 10. Moore JE, Pratt HE, Purcaro MJ, Weng Z. A curated benchmark of enhancer-gene
811 interactions for evaluating enhancer-target gene prediction methods. Genome Biol. 2020
812 Jan 22;21(1):17.

813 11. Nasser J, Bergman DT, Fulco CP, Guckelberger P, Doughty BR, Patwardhan TA, et al.
814 Genome-wide enhancer maps link risk variants to disease genes. Nature. 2021
815 May;593(7858):238–43.

816 12. Dorighi KM, Swigut T, Henriques T, Bhanu NV, Scruggs BS, Nady N, et al. Mll3 and Mll4
817 facilitate enhancer RNA synthesis and transcription from promoters independently of H3K4
818 monomethylation. Mol Cell. 2017 May 18;66(4):568-576.e4.

819 13. Rickels R, Herz HM, Sze CC, Cao K, Morgan MA, Collings CK, et al. Histone H3K4
820 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is
821 dispensable for development and viability. Nat Genet. 2017 Nov;49(11):1647–53.

822 14. Zhang T, Zhang Z, Dong Q, Xiong J, Zhu B. Histone H3K27 acetylation is dispensable for
823 enhancer activity in mouse embryonic stem cells. Genome Biol. 2020 Feb 21;21(1):45.

824 15. Sankar A, Mohammad F, Sundaramurthy AK, Wang H, Lerdrup M, Tatar T, et al. Histone
825 editing elucidates the functional roles of H3K27 methylation and acetylation in mammals.
826 Nat Genet. 2022 Jun;54(6):754–60.

827 16. Martire S, Nguyen J, Sundaresan A, Banaszynski LA. Differential contribution of p300 and
828 CBP to regulatory element acetylation in mESCs. BMC Mol Cell Biol. 2020 Jul 20;21(1):55.

829 17. Smith A. Formative pluripotency: the executive phase in a developmental continuum. Dev
830 Camb Engl. 2017 Feb 1;144(3):365–73.

831 18. Gökbuget D, Blelloch R. Epigenetic control of transcriptional regulation in pluripotency and
832 early differentiation. Dev Camb Engl. 2019 Sep 25;146(19):dev164772.

833 19. Yang P, Humphrey SJ, Cinghu S, Pathania R, Oldfield AJ, Kumar D, et al. Multi-omic
834 Profiling Reveals Dynamics of the Phased Progression of Pluripotency. Cell Syst. 2019 May
835 22;8(5):427-445.e10.

836 20. Krishnakumar R, Chen AF, Pantovich MG, Danial M, Parchem RJ, Labosky PA, et al.
837 FOXD3 Regulates Pluripotent Stem Cell Potential by Simultaneously Initiating and
838 Repressing Enhancer Activity. Cell Stem Cell. 2016 Jan 7;18(1):104–17.

839 21. Ashokkumar D, Zhang Q, Much C, Bledau AS, Naumann R, Alexopoulou D, et al. MLL4 is
840 required after implantation, whereas MLL3 becomes essential during late gestation. Dev
841 Camb Engl. 2020 Jun 17;147(12):dev186999.

842 22. Wang C, Lee JE, Lai B, Macfarlan TS, Xu S, Zhuang L, et al. Enhancer priming by H3K4
843 methyltransferase MLL4 controls cell fate transition. Proc Natl Acad Sci U S A. 2016 Oct
844 18;113(42):11871–6.

845 23. Chen AF, Liu AJ, Krishnakumar R, Freimer JW, DeVeale B, Blelloch R. GRHL2-dependent
846 enhancer switching maintains a pluripotent stem cell transcriptional subnetwork after exit
847 from naïve pluripotency. Cell Stem Cell. 2018 Aug 2;23(2):226-238.e4.

848 24. Skene PJ, Henikoff JG, Henikoff S. Targeted in situ genome-wide profiling with high
849 efficiency for low cell numbers. Nat Protoc. 2018 May;13(5):1006–19.

850 25. Stark R, Brown G. DiffBind: Differential Binding Analysis of ChIP-Seq Peak Data [Internet].
851 Bioconductor version: Release (3.15); 2022 [cited 2022 Aug 24]. Available from:
852 https://bioconductor.org/packages/DiffBind/

853 26. Meers MP, Tenenbaum D, Henikoff S. Peak calling by Sparse Enrichment Analysis for
854 CUT&RUN chromatin profiling. Epigenetics Chromatin. 2019 Jul 12;12(1):42.

855 27. Bleckwehl T, Crispatzu G, Schaaf K, Respuela P, Bartusel M, Benson L, et al. Enhancer-
856 associated H3K4 methylation safeguards in vitro germline competence. Nat Commun. 2021
857 Oct 1;12:5771.

858 28. Buecker C, Srinivasan R, Wu Z, Calo E, Acampora D, Faial T, et al. Reorganization of
859 enhancer patterns in transition from naive to primed pluripotency. Cell Stem Cell. 2014 Jun
860 5;14(6):838–53.

861 29. Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, et al. CUT&Tag
862 for efficient epigenomic profiling of small samples and single cells. Nat Commun. 2019 Apr
863 29;10(1):1930.

864 30. Zhou HY, Katsman Y, Dhaliwal NK, Davidson S, Macpherson NN, Sakthidevi M, et al. A
865 Sox2 distal enhancer cluster regulates embryonic stem cell differentiation potential. Genes
866 Dev. 2014 Dec 15;28(24):2699–711.

867 31. Blinka S, Reimer MH, Pulakanti K, Rao S. Super-Enhancers at the Nanog Locus
868 Differentially Regulate Neighboring Pluripotency-Associated Genes. Cell Rep. 2016 Sep
869 27;17(1):19–28.

870 32. Thomas HF, Kotova E, Jayaram S, Pilz A, Romeike M, Lackner A, et al. Temporal
871 dissection of an enhancer cluster reveals distinct temporal and functional contributions of
872 individual elements. Mol Cell. 2021 Mar 4;81(5):969-982.e13.

873 33. Romeike M, Spach S, Huber M, Feng S, Vainorius G, Elling U, et al. Transient upregulation
874 of IRF1 during exit from naive pluripotency confers viral protection. EMBO Rep. 2022 Jul
875 19;e55375.

876 34. Agrawal P, Blinka S, Pulakanti K, Reimer MH, Stelloh C, Meyer AE, et al. Genome editing
877 demonstrates that the -5 kb Nanog enhancer regulates Nanog expression by modulating
878 RNAPII initiation and/or recruitment. J Biol Chem. 2021 Jun;296:100189.

879 35. Bruse N, Heeringen SJ van. GimmeMotifs: an analysis framework for transcription factor
880 motif analysis [Internet]. bioRxiv; 2018 [cited 2022 Aug 24]. p. 474403. Available from:
881 https://www.biorxiv.org/content/10.1101/474403v1

882 36. Moorthy SD, Davidson S, Shchuka VM, Singh G, Malek-Gilani N, Langroudi L, et al.
883 Enhancers and super-enhancers have an equivalent regulatory role in embryonic stem cells
884 through regulation of single or multiple genes. Genome Res. 2017 Feb;27(2):246–58.

885 37. Osterwalder M, Barozzi I, Tissières V, Fukuda-Yuzawa Y, Mannion BJ, Afzal SY, et al.
886 Enhancer redundancy provides phenotypic robustness in mammalian development. Nature.
887 2018 Feb 8;554(7691):239–43.

888 38. Mas G, Blanco E, Ballaré C, Sansó M, Spill YG, Hu D, et al. Promoter bivalency favors an
889 open chromatin architecture in embryonic stem cells. Nat Genet. 2018 Oct;50(10):1452–62.

890 39. Zhang H, Gayen S, Xiong J, Zhou B, Shanmugam AK, Sun Y, et al. MLL1 Inhibition
891 Reprograms Epiblast Stem Cells to Naïve Pluripotency. Cell Stem Cell. 2016 Apr
892 7;18(4):481–94.

893 40. Kaikkonen MU, Spann N, Heinz S, Romanoski CE, Allison KA, Stender JD, et al.
894 Remodeling of the enhancer landscape during macrophage activation is coupled to
895 enhancer transcription. Mol Cell. 2013 Aug 8;51(3):310–25.

896 41. Neumayr C, Haberle V, Serebreni L, Karner K, Hendy O, Boija A, et al. Differential cofactor
897 dependencies define distinct types of human enhancers. Nature. 2022 Jun;606(7913):406–
898 13.

899 42. Cao K, Collings CK, Morgan MA, Marshall SA, Rendleman EJ, Ozark PA, et al. An
900 Mll4/COMPASS-Lsd1 epigenetic axis governs enhancer function and pluripotency transition
901 in embryonic stem cells. Sci Adv. 2018 Jan;4(1):eaap8747.

902 43. Ang SY, Uebersohn A, Spencer CI, Huang Y, Lee JE, Ge K, et al. KMT2D regulates specific
903 programs in heart development via histone H3 lysine 4 di-methylation. Dev Camb Engl.
904 2016 Mar 1;143(5):810–21.

905 44. Shpargel KB, Mangini CL, Xie G, Ge K, Magnuson T. The KMT2D Kabuki syndrome histone
906 methylase controls neural crest cell differentiation and facial morphology. Dev Camb Engl.
907 2020 Jul 17;147(21):dev187997.

908 45. Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, et al. The nf-core framework
909 for community-curated bioinformatics pipelines. Nat Biotechnol. 2020 Mar;38(3):276–8.

910

Main Figure Legends

Figure 1: MLL3/4 is dispensable for transcriptional activation of much of the formative program. A) Schematic of naïve to formative transition *in vivo* and *in vitro*. B) Western validation of MLL3 and MLL4 knockout cell lines. C) Brightfield microscopy of cell lines in naive and formative conditions, 20x, scale bar = 100um. D) Differential gene expression analysis (DGE) on WT naive and formative RNA-seq samples (significant genes colored, P.adj < 0.05, Log2 Foldchange (Log2FC) >1). E) UMAP analysis on all RNA-seq samples subset by all naive and formative genes from WT DGE. F,G) Fold change (formative/naive) of all genes for either MLL3KO and DKO compared to WT. WT formative and naive genes from WT DGE analysis colored. R, Pearson's coefficient. Dashed line and linear equation represent linear model of all genes. H) DGE analysis on WT naive and DKO naive samples (significant genes colored, P.adj < 0.05 and Log2FC > 1). I) Naive state DKO misexpressed genes categorized by whether they are also WT naive or formative genes. J,K) Same as H,I respectively but comparing within the formative state.

Figure 2: MLL3/4 is required for all dynamic H3K4me1 deposition during pluripotent transition. A) Differential signal enrichment between naive and formative WT H3K4me1 peaks identifies significant peaks (Black, FDR <0.05, Log2 Foldchange (Log2FC) > 1). B) Scatterplot of H3K4me1 peak intensities between naive and formative WT samples. (gray, shared peaks, FDR > 0.1 & Log2FC < 0.7, red or blue, FDR < 0.05 & Log2FC > 1). C,D) H3K4me1 signal in naive and formative MLL3KO or DKO samples at peaks called from WT. Colors correspond to peak categories derived from WT. E) Differential signal enrichment of naive and formative DKO samples. F) Same as C,D but with naive and formative dCD cells. G) Differential signal enrichment of naive and formative dCD samples. H) Feature annotation of WT peaks stratified by peak category. I) Distal H3K4me1 peak categories further stratified by MLL3/4 independent (DKO/WT Log2FC > -0.7) or dependent (DKO/WT Log2FC < -1). J) Heatmap of MLL3/4 independent and

dependent H3K4me1 peak categories, rows sorted on “S” columns. All heatmap values and range are in CPM. For metagene analysis the range in CPM is the same as shown in heatmap for each factor. K) nearest neighbor TSS analysis of expression levels in Log2CPM for each RNA-seq dataset near H3K4me1 peak categories. Multi-comparison paired Wilcoxon Rank-Sum Test, Benjamini-Hochberg corrected. *p<0.05 **p<0.01 ***p<0.001 ****p<0.0001

Figure 3: MLL3/4 dependent and independent distal H3K27ac deposition. A) Differential signal enrichment between naive and formative WT H3K27ac peaks identifies significant peaks (Black, FDR <0.05, Log2 Foldchange (Log2FC) > 1). B) Scatterplot of H3K27ac peak intensities between naive and formative WT samples. (gray, shared peaks, FDR > 0.1 & Log2FC < 0.7, red or blue, FDR < 0.05 & Log2FC > 1). C) H3K27ac signal in naive and formative DKO samples at peaks called from WT. Colors correspond to peak categories derived from WT. D) Feature annotation of WT or DKO peaks stratified by peak category. E) Distal H3K27ac peak categories further stratified by MLL3/4 independent (DKO/WT Log2FC > -0.7) or dependent (DKO/WT Log2FC < -1). G,H) Heatmap of MLL3/4 independent and dependent H3K27ac peak categories for naive or formative specific peaks, rows sorted on “S” columns. All heatmap values and range are in CPM. For metagene analysis the range in CPM is the same as shown in heatmap for each factor. I) nearest neighbor TSS analysis of expression levels Log2CPM near H3K27ac peak categories. Multi-comparison paired Wilcoxon Rank-Sum Test, Benjamini-Hochberg corrected. *p<0.05 **p<0.01 ***p<0.001 ****p<0.0001

Figure 4: Enhancer Activation Can Occur Independently of MLL3/4. A) Pie charts showing fraction of coincident changes in H3K27ac within the different categories of changing H3K4me1 following MLL3/4. B) Heatmap of sites that normally lose H3K4me1 and H3K27ac during naive to formative transition clustered by either MLL3/4 dependent or independent H3K27ac. Rows sorted

52 on “S” columns. Y-axis scale of metagene profiles above is same as signal ranges of heatmap,
53 values in CPM. C) Same as B but using sites that normally gain H3K4me1 and H3K27ac during
54 naive to formative transition clustered by either MLL3/4 dependent or independent H3K27ac. D)
55 Same as B but using sites that have unchanging H3K4me1 but gain H3K27ac during naive to
56 formative transition clustered by either MLL3/4 dependent or independent H3K27ac. All heatmap
57 values and range are in CPM. For metagene analysis the range in CPM is the same as shown in
58 heatmap for each factor. E) Naive and formative examples of functionally validated enhancers
59 that are either MLL3/4 dependent or independent. Naive MLL3/4 dependent Sox2 enhancer
60 cluster, Naive MLL3/4 independent Nanog enhancers, Formative independent and dependent
61 enhancers of Fgf5, formative independent enhancer of Pou3f1/Oct6.
62
63 **Figure 5: Distal H3K4me1 and H3K27ac are not functionally coupled with formative**
64 **transcriptional activation.** A) Naive ChromHMM emissions for 16 states yields all combinations
65 of H3K4me1 and H3K27ac in WT and DKO. Heatmaps are clustered by each chromatin state.
66 Black arrows denote states with MLL3/4 dependent H3K27ac. B) Gene expression changes of
67 nearest TSS for all sites in each given state. Monte Carlo permutation sampling of “all genes
68 control” to perform Mann-Whitney U Test for each state. Benjamini-Hochberg corrected. Selected
69 statistics shown, additional statistics provided in supplemental table. C) Same as A with Formative
70 samples. D) Same as B with Formative RNA-seq. All heatmap values and range are in CPM. For
71 metagene analysis the range in CPM is the same as shown in heatmap for each factor.
72 ChromHMM states without any H3K4me1 or H3K27ac emission probability are excluded. E) For
73 all H3K4me1/H3K27ac+ sites in either Naive (top panel) or Formative (bottom panel), weighted
74 enhancer loss (total associated peaks * MLL3/4 dependent H3K27ac peaks). Each dot represents
75 one gene. Blue line represents generalized linear model, gray 95% confidence interval. F) Same
76 as E but only considering sites within 10kb of any given TSS. *$p<0.05$ **$p<0.01$ ***$p<0.001$
77 ****$p<0.0001$

78

79 **Figure 6: Gene-centric analysis reveals a subset of distal loci associate with MLL3/4**
80 **dependent formative genes.** A) Pie charts showing fraction of normally gained H3K4me1,
81 H3K27ac, or gene expression that are dependent on MLL3/4 or not. B) Clusterprofile analysis of
82 Biological Processes Gene Ontology for 215 MLL3/4 dependent formative genes. C) Fold change
83 Log2CPM of H3K27ac density for H3K4me1/H3K27ac+ enhancers that are associated with
84 formative genes that gained expression during transition in an either MLL3/4 independent or
85 dependent fashion. D) Genome tracks for known and predicted formative targets of MLL3/4
86 Lefty1, Brachyury/T, and Fgf8. E) Model for the role of MLL3/4 independent and dependent
87 mechanisms in the regulation of gained gene expression during the naïve to formative transition.

88

89 **Supplementary Figure Legends**

90 **Figure S1: MLL3/4 is dispensable for transcriptional activation of much of the formative**
91 **program**. A) Confirmation of multiple DKO clones using genomic PCR following tamoxifen
92 treatment. B) Proliferation measurements using crystal violet assays. C) Westerns on whole cell
93 protein fractions of OCT4, NANOG and αTUBULIN loading control. D) qPCR of naive enriched
94 markers Klf2, Klf4, Zfp42. For each naive marker selected comparisons are shown resulting from
95 a Two-way ANOVA, Tukey's multiple comparison test for all samples. Mean and S.E.M. in black.
96 Additional statistics provided in supplemental table. E) Same as D with formative enriched
97 markers Fgf5, Otx2, Dnmt3b. *p<0.05 **p<0.01 ***p<0.001 ****p<0.0001

98

99 **Figure S2: MLL3/4 is required for all dynamic H3K4me1 deposition during pluripotent**
100 **transition.** A) Diffbind analysis of H3K4me1 signal in MLL3KO cells at WT peaks. B) mRNA
101 Log2CPM of naive MLL3/4 dCD cells vs. WT cells. Significant genes highlighted. C) same as B
102 with formative samples. D) Foldchange (formative/naive) of all genes for MLL3/4 dCD compared

103 to WT. WT formative and naive genes from WT DGE analysis colored. R, Pearson's coefficient.
104 Dashed line and linear equation represent linear model of all genes. E) Heatmap of ChIP-seq for
105 H3K4me1 in naive and formative cells from two published studies. All heatmap values and range
106 are in CPM. For metagene analysis the range in CPM is the same as shown in heatmap for each
107 factor.
108
109 **Figure S3: MLL3/4 dependent and independent distal H3K27ac deposition.** A) Heatmap of
110 shared H3K27ac sites clustered by MLL3/4 independent or dependent H3K27ac. B) Heatmaps of
111 ChIP-seq for H3K27ac in naive and formative cells from twopublished studies. All heatmap values
112 and range are in CPM. For metagene analysis the range in CPM is the same as shown in heatmap
113 for each factor.
114
115 **Figure S4: Enhancer Activation Can Occur Independently of MLL3/4.** A) Peak count of
116 H3K4me1 categories from Fig2. after filtering for distal H3K4me1 overlapping WT ATAC peaks.
117 B) Heatmaps for shared H3K4me1 that overlaps with naive enriched H3K27ac, clustered by
118 MLL3/4 independent or dependent H3K27ac. C) Heatmaps for MLL3/4 dependent H3K4me1
119 overlapped with sites that have shared, naive, or formative H3K27ac. Clustered additionally by
120 MLL3/4 independent or dependent H3K27ac. D) Heatmaps for shared H3K4me1 that overlaps
121 with shared H3K27ac, clustered by MLL3/4 independent or dependent H3K27ac. All heatmap
122 values and range are in CPM. For metagene analysis the range in CPM is the same as shown in
123 heatmap for each factor. E) Log2CPM of Sox2 and Nanog in naive and formative samples
124 comparing WT and DKO cells. Sox2 is significantly up in the formative state.
125
126 **Figure S5: Enhancer Activation Can Occur Independently of MLL3/4.** A) Clusters from
127 differential motif analysis using Gimmemotifs on formative enriched MLL3/4 independent and
128 dependent enhancers marked by both H3K4me1 and H3K27ac (Sites from Figure 4C and 4D).

129 Motifs derived from JASPAR 2020 database, k-means clustering used. B) Same as A but with
130 naive enriched independent and dependent enhancers (using sites from Figure 4B and Figure
131 S5B).
132
133 **Figure S6: Distal H3K4me1 and H3K27ac are not functionally coupled with formative**
134 **transcriptional activation.** A) Nearest neighbor TSS analysis of expression levels in Log2CPM
135 near naive H3K4me1 dependent sites that are either have MLL3/4 independent (left panel) or
136 dependent H3K27ac (right panel)(clusters from Fig.4B). Multi-comparison paired Wilcoxon Rank-
137 Sum Test, Benjamini-Hochberg corrected. B) Same as A but at formative H3K4me1 dependent
138 sites (clusters from Fig.4C). C) Boxplots of relative expression DKO/WT of RNA levels for all
139 genes associated with any H3K4me1/H3K27ac+ peak in either naive or formative state (far left,
140 middle left) or the same analysis considering only H3K4me1/H3K27ac+ peaks within 10kb of any
141 given TSS (middle right, far right). Genes are binned by the percentage of their associated
142 enhancer loss in DKOs. D) Same as C but plotting the number of dependent enhancers for each
143 gene in Log2.
144
145 **Figure S7: Gene-centric analysis reveals a subset of distal loci associate with MLL3/4**
146 **dependent formative genes.** A) A major percentage of naive enriched peaks fail to maintain
147 distal H3K4me1 and H3K27ac with moderate consequences on naive gene expression. B)
148 Clusterprofile analysis of Biological Processes Gene Ontology for 348 MLL3/4 dependent naive
149 genes. C) Fold change Log2CPM of H3K27ac density for H3K4me1/H3K27ac+ enhancers that
150 are associated with naive genes that gained expression during transition in an either MLL3/4
151 independent or dependent fashion.

Figure 1

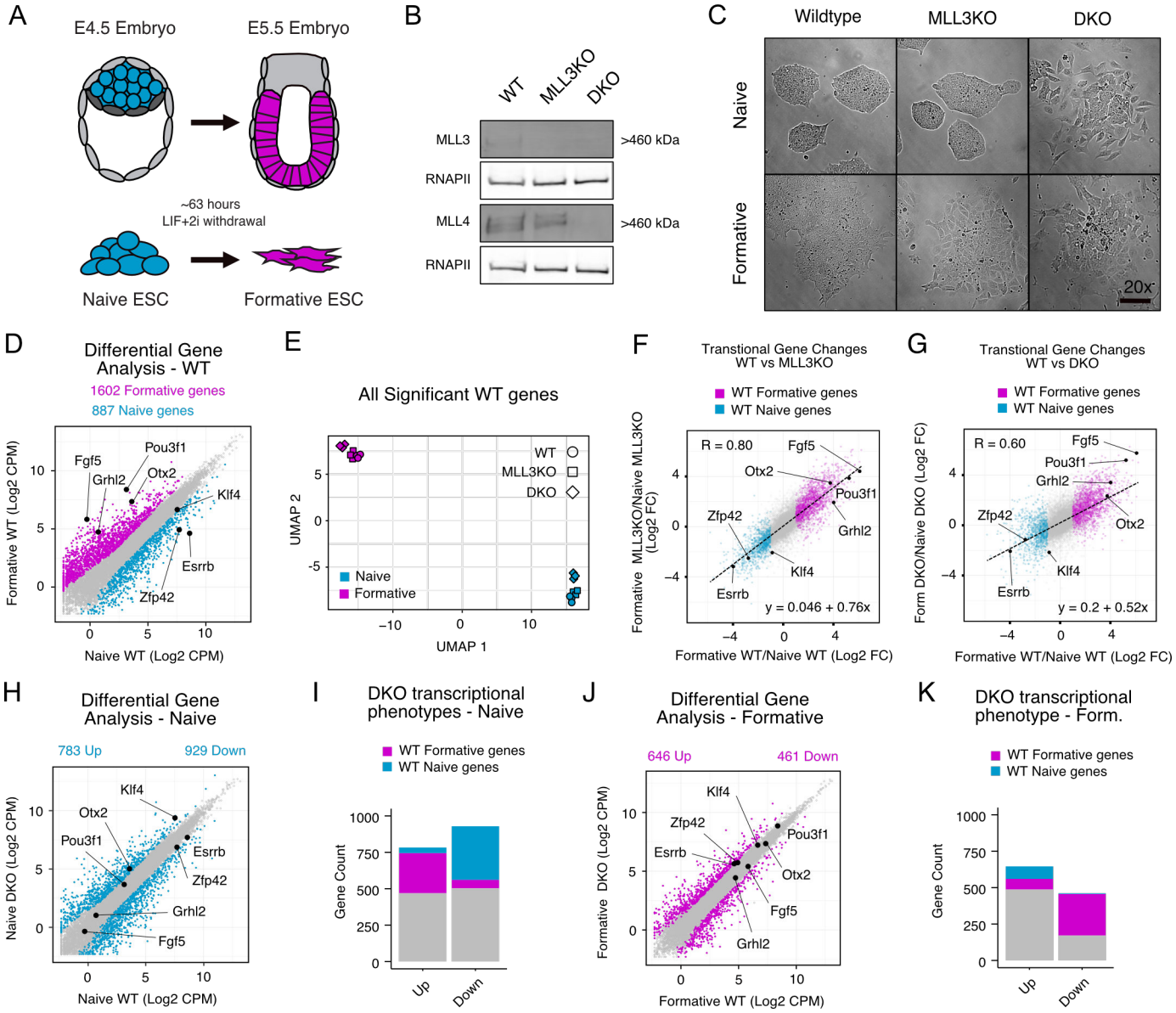

A
E4.5 Embryo
E5.5 Embryo
~63 hours
LIF+2i withdrawal
Naive ESC
Formative ESC
B
WT
MLL3KO
DKO
MLL3
>460 kDa
RNAPII
MLL4
>460 kDa
RNAPII
C
Wildtype
MLL3KO
DKO
Naive
Formative
20x
D
Differential Gene Analysis - WT
1602 Formative genes
887 Naive genes
Formative WT (Log2 CPM)
Naive WT (Log2 CPM)
Fgf5
Grhl2
Pou3f1
Otx2
Klf4
Esrrb
Zfp42
E
All Significant WT genes
WT
MLL3KO
DKO
Naive
Formative
UMAP 1
UMAP 2
F
Transtional Gene Changes WT vs MLL3KO
WT Formative genes
WT Naive genes
R = 0.80
y = 0.046 + 0.76x
Formative MLL3KO/Naive MLL3KO (Log2 FC)
Formative WT/Naive WT (Log2 FC)
G
Transtional Gene Changes WT vs DKO
WT Formative genes
WT Naive genes
R = 0.60
y = 0.2 + 0.52x
Form DKO/Naive DKO (Log2 FC)
Formative WT/Naive WT (Log2 FC)
H
Differential Gene Analysis - Naive
783 Up
929 Down
Naive DKO (Log2 CPM)
Naive WT (Log2 CPM)
I
DKO transcriptional phenotypes - Naive
WT Formative genes
WT Naive genes
Gene Count
Up
Down
J
Differential Gene Analysis - Formative
646 Up
461 Down
Formative DKO (Log2 CPM)
Formative WT (Log2 CPM)
K
DKO transcriptional phenotype - Form.
WT Formative genes
WT Naive genes
Gene Count
Up
Down

Figure 2

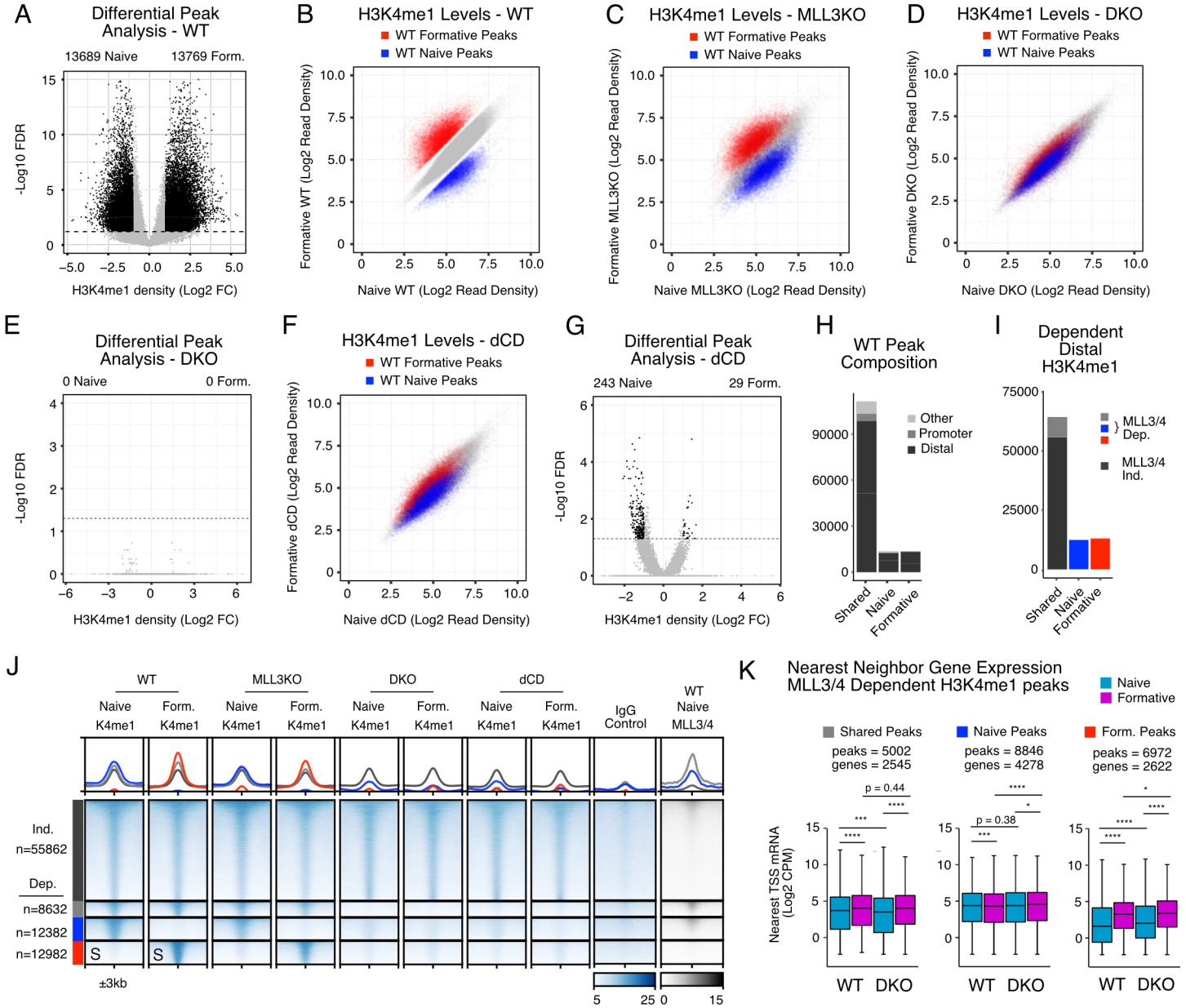

A
Differential Peak Analysis - WT
13689 Naive
13769 Form.
-Log10 FDR
H3K4me1 density (Log2 FC)
B
H3K4me1 Levels - WT
WT Formative Peaks
WT Naive Peaks
Formative WT (Log2 Read Density)
Naive WT (Log2 Read Density)
C
H3K4me1 Levels - MLL3KO
WT Formative Peaks
WT Naive Peaks
Formative MLL3KO (Log2 Read Density)
Naive MLL3KO (Log2 Read Density)
D
H3K4me1 Levels - DKO
WT Formative Peaks
WT Naive Peaks
Formative DKO (Log2 Read Density)
Naive DKO (Log2 Read Density)
E
Differential Peak Analysis - DKO
0 Naive
0 Form.
-Log10 FDR
H3K4me1 density (Log2 FC)
F
H3K4me1 Levels - dCD
WT Formative Peaks
WT Naive Peaks
Formative dCD (Log2 Read Density)
Naive dCD (Log2 Read Density)
G
Differential Peak Analysis - dCD
243 Naive
29 Form.
-Log10 FDR
H3K4me1 density (Log2 FC)
H
WT Peak Composition
Other
Promoter
Distal
Shared
Naive
Formative
I
Dependent Distal H3K4me1
MLL3/4 Dep.
MLL3/4 Ind.
Shared
Naive
Formative
J
WT
MLL3KO
DKO
dCD
Naive K4me1
Form. K4me1
IgG Control
WT Naive MLL3/4
Ind.
n=55862
Dep.
n=8632
n=12382
n=12982
S
±3kb
K
Nearest Neighbor Gene Expression
MLL3/4 Dependent H3K4me1 peaks
Naive
Formative
Shared Peaks
peaks = 5002
genes = 2545
Naive Peaks
peaks = 8846
genes = 4278
Form. Peaks
peaks = 6972
genes = 2622
p = 0.44
p = 0.38
Nearest TSS mRNA (Log2 CPM)
WT
DKO

Figure 3

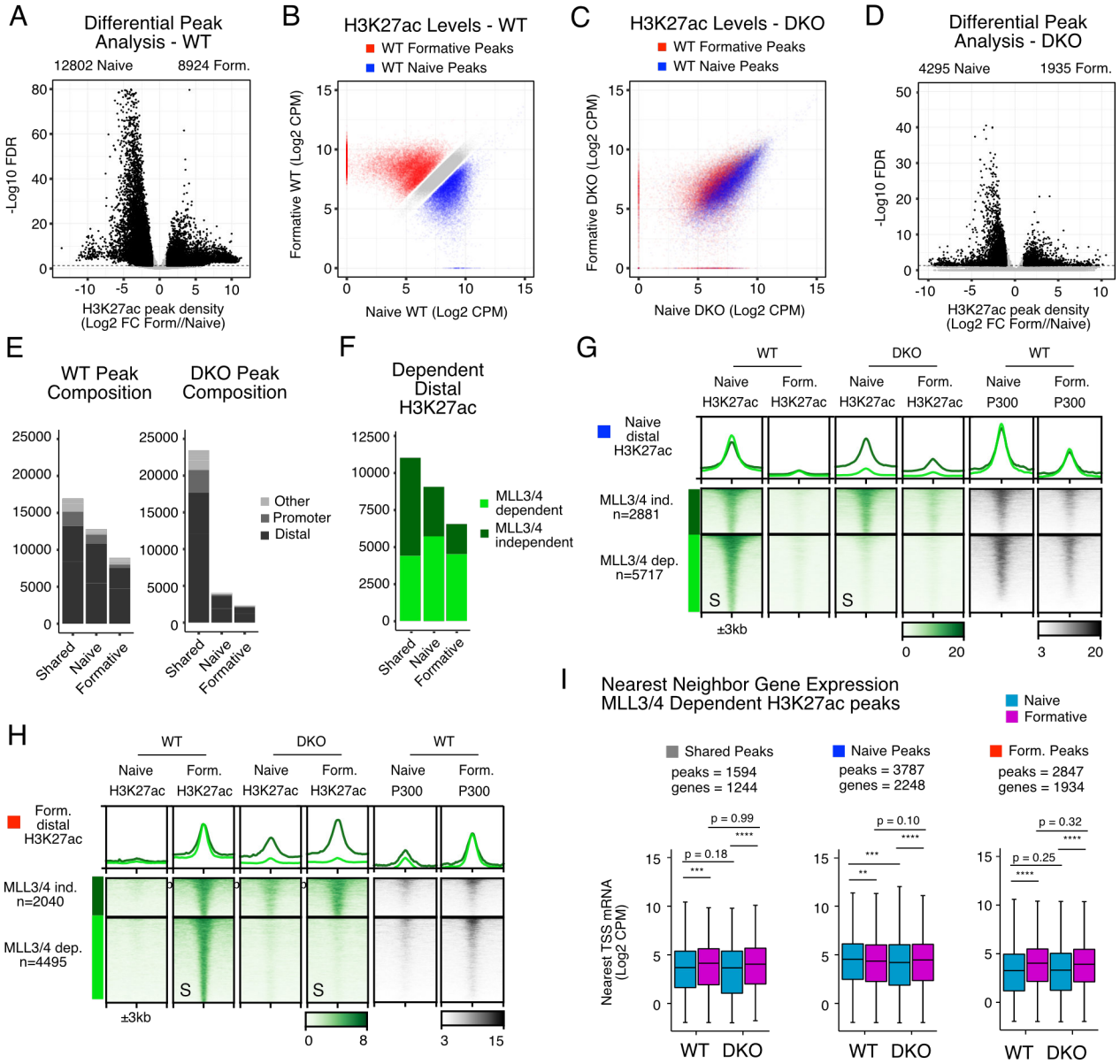

A
Differential Peak Analysis - WT
12802 Naive
8924 Form.
-Log10 FDR
H3K27ac peak density (Log2 FC Form//Naive)
B
H3K27ac Levels - WT
WT Formative Peaks
WT Naive Peaks
Formative WT (Log2 CPM)
Naive WT (Log2 CPM)
C
H3K27ac Levels - DKO
WT Formative Peaks
WT Naive Peaks
Formative DKO (Log2 CPM)
Naive DKO (Log2 CPM)
D
Differential Peak Analysis - DKO
4295 Naive
1935 Form.
-Log10 FDR
H3K27ac peak density (Log2 FC Form//Naive)
E
WT Peak Composition
DKO Peak Composition
Other
Promoter
Distal
Shared
Naive
Formative
F
Dependent Distal H3K27ac
MLL3/4 dependent
MLL3/4 independent
Shared
Naive
Formative
G
WT
DKO
WT
Naive H3K27ac
Form. H3K27ac
Naive H3K27ac
Form. H3K27ac
Naive P300
Form. P300
Naive distal H3K27ac
MLL3/4 ind. n=2881
MLL3/4 dep. n=5717
S
±3kb
H
WT
DKO
WT
Naive H3K27ac
Form. H3K27ac
Naive H3K27ac
Form. H3K27ac
Naive P300
Form. P300
Form. distal H3K27ac
MLL3/4 ind. n=2040
MLL3/4 dep. n=4495
S
±3kb
I
Nearest Neighbor Gene Expression
MLL3/4 Dependent H3K27ac peaks
Naive
Formative
Shared Peaks
peaks = 1594
genes = 1244
Naive Peaks
peaks = 3787
genes = 2248
Form. Peaks
peaks = 2847
genes = 1934
p = 0.99
p = 0.18
p = 0.10
p = 0.32
p = 0.25
Nearest TSS mRNA (Log2 CPM)
WT
DKO

Figure 4

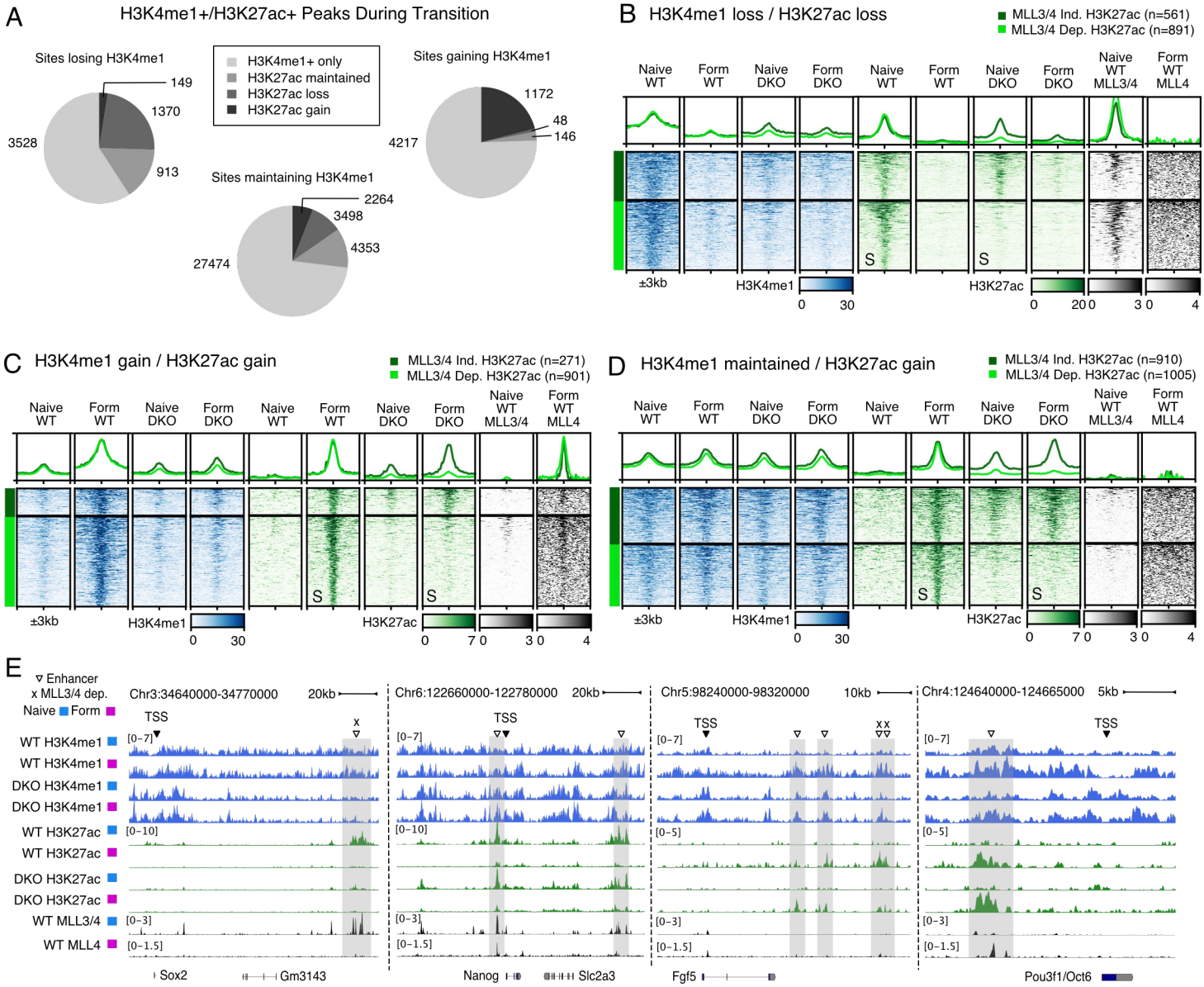

A
H3K4me1+/H3K27ac+ Peaks During Transition
Sites losing H3K4me1
149
1370
3528
913
H3K4me1+ only
H3K27ac maintained
H3K27ac loss
H3K27ac gain
Sites gaining H3K4me1
1172
48
146
4217
Sites maintaining H3K4me1
2264
3498
4353
27474
B
H3K4me1 loss / H3K27ac loss
MLL3/4 Ind. H3K27ac (n=561)
MLL3/4 Dep. H3K27ac (n=891)
Naive WT
Form WT
Naive DKO
Form DKO
Naive WT
Form WT
Naive DKO
Form DKO
Naive WT MLL3/4
Form WT MLL4
S
S
±3kb
H3K4me1
0
30
H3K27ac
0
20
0
3
0
4
C
H3K4me1 gain / H3K27ac gain
MLL3/4 Ind. H3K27ac (n=271)
MLL3/4 Dep. H3K27ac (n=901)
Naive WT
Form WT
Naive DKO
Form DKO
Naive WT
Form WT
Naive DKO
Form DKO
Naive WT MLL3/4
Form WT MLL4
S
S
±3kb
H3K4me1
0
30
H3K27ac
0
7
0
3
0
4
D
H3K4me1 maintained / H3K27ac gain
MLL3/4 Ind. H3K27ac (n=910)
MLL3/4 Dep. H3K27ac (n=1005)
Naive WT
Form WT
Naive DKO
Form DKO
Naive WT
Form WT
Naive DKO
Form DKO
Naive WT MLL3/4
Form WT MLL4
S
S
±3kb
H3K4me1
0
30
H3K27ac
0
7
0
3
0
4
E
Enhancer
x MLL3/4 dep.
Naive
Form
WT H3K4me1
WT H3K4me1
DKO H3K4me1
DKO H3K4me1
WT H3K27ac
WT H3K27ac
DKO H3K27ac
DKO H3K27ac
WT MLL3/4
WT MLL4
Chr3:34640000-34770000
20kb
TSS
x
[0-7]
[0-10]
[0-3]
[0-1.5]
Sox2
Gm3143
Chr6:122660000-122780000
20kb
TSS
[0-7]
[0-10]
[0-3]
[0-1.5]
Nanog
Slc2a3
Chr5:98240000-98320000
10kb
TSS
xx
[0-7]
[0-5]
[0-3]
[0-1.5]
Fgf5
Chr4:124640000-124665000
5kb
TSS
[0-7]
[0-5]
[0-3]
[0-1.5]
Pou3f1/Oct6

Figure 5

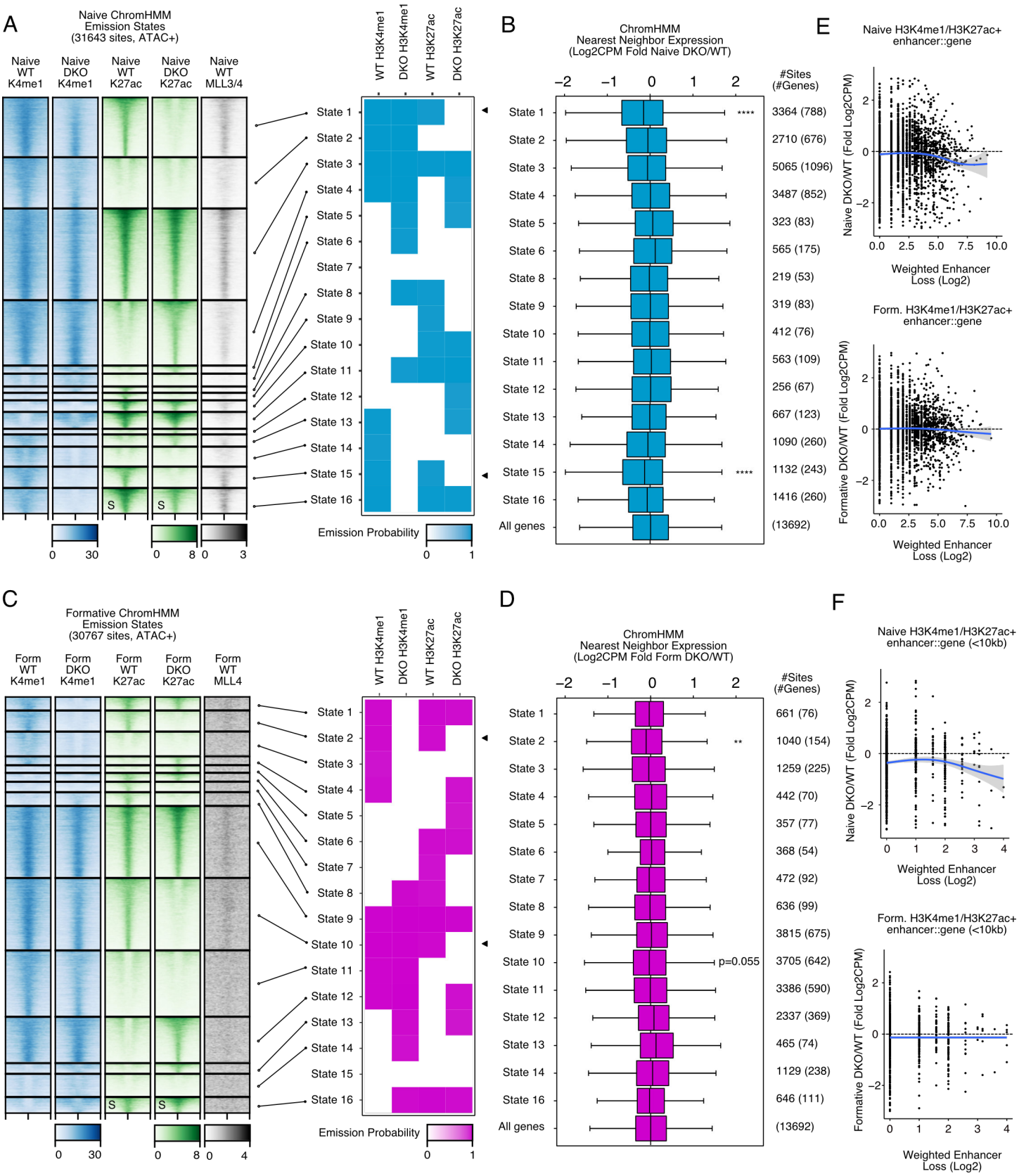
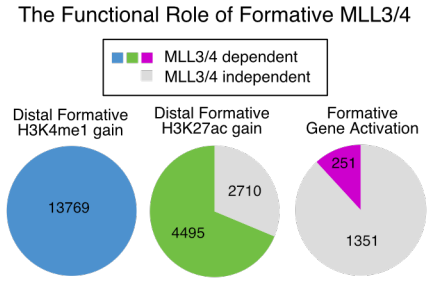
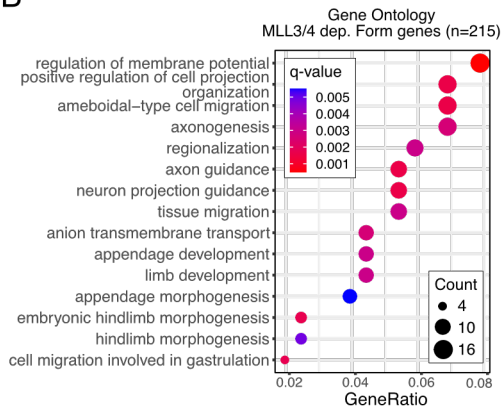

A
Naive ChromHMM
Emission States
(31643 sites, ATAC+)
Naive WT K4me1
Naive DKO K4me1
Naive WT K27ac
Naive DKO K27ac
Naive WT MLL3/4
WT H3K4me1
DKO H3K4me1
WT H3K27ac
DKO H3K27ac
State 1
State 2
State 3
State 4
State 5
State 6
State 7
State 8
State 9
State 10
State 11
State 12
State 13
State 14
State 15
State 16
Emission Probability
B
ChromHMM
Nearest Neighbor Expression
(Log2CPM Fold Naive DKO/WT)
#Sites (#Genes)
3364 (788)
2710 (676)
5065 (1096)
3487 (852)
323 (83)
565 (175)
219 (53)
319 (83)
412 (76)
563 (109)
256 (67)
667 (123)
1090 (260)
1132 (243)
1416 (260)
(13692)
All genes
E
Naive H3K4me1/H3K27ac+ enhancer::gene
Naive DKO/WT (Fold Log2CPM)
Weighted Enhancer Loss (Log2)
Form. H3K4me1/H3K27ac+ enhancer::gene
Formative DKO/WT (Fold Log2CPM)
C
Formative ChromHMM
Emission States
(30767 sites, ATAC+)
Form WT K4me1
Form DKO K4me1
Form WT K27ac
Form DKO K27ac
Form WT MLL4
D
ChromHMM
Nearest Neighbor Expression
(Log2CPM Fold Form DKO/WT)
p=0.055
661 (76)
1040 (154)
1259 (225)
442 (70)
357 (77)
368 (54)
472 (92)
636 (99)
3815 (675)
3705 (642)
3386 (590)
2337 (369)
465 (74)
1129 (238)
646 (111)
(13692)
F
Naive H3K4me1/H3K27ac+ enhancer::gene (<10kb)
Form. H3K4me1/H3K27ac+ enhancer::gene (<10kb)

Figure 6

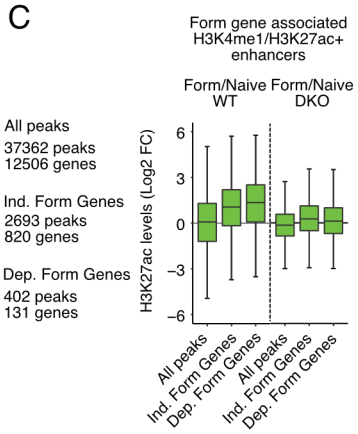
A



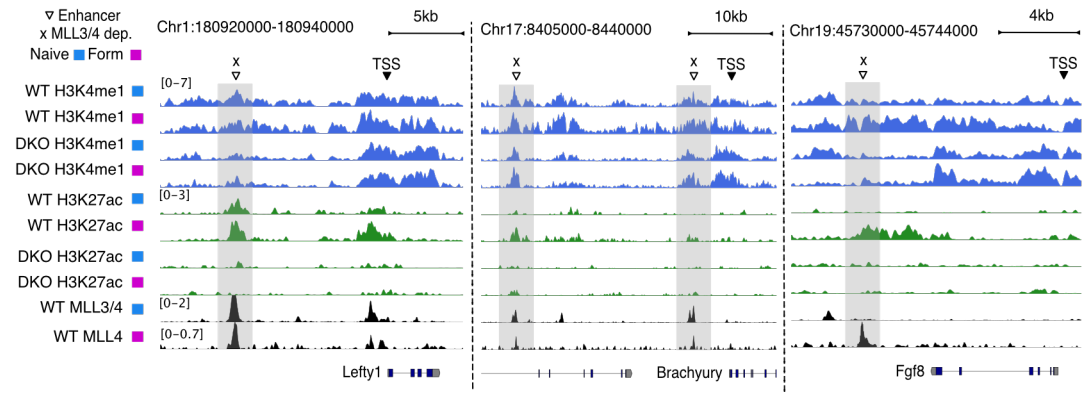
B



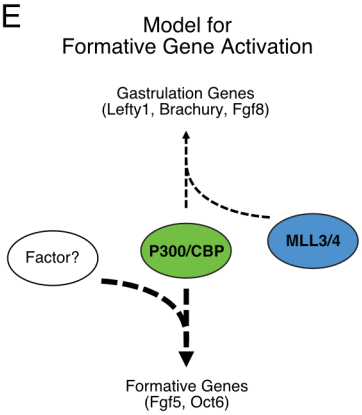
C



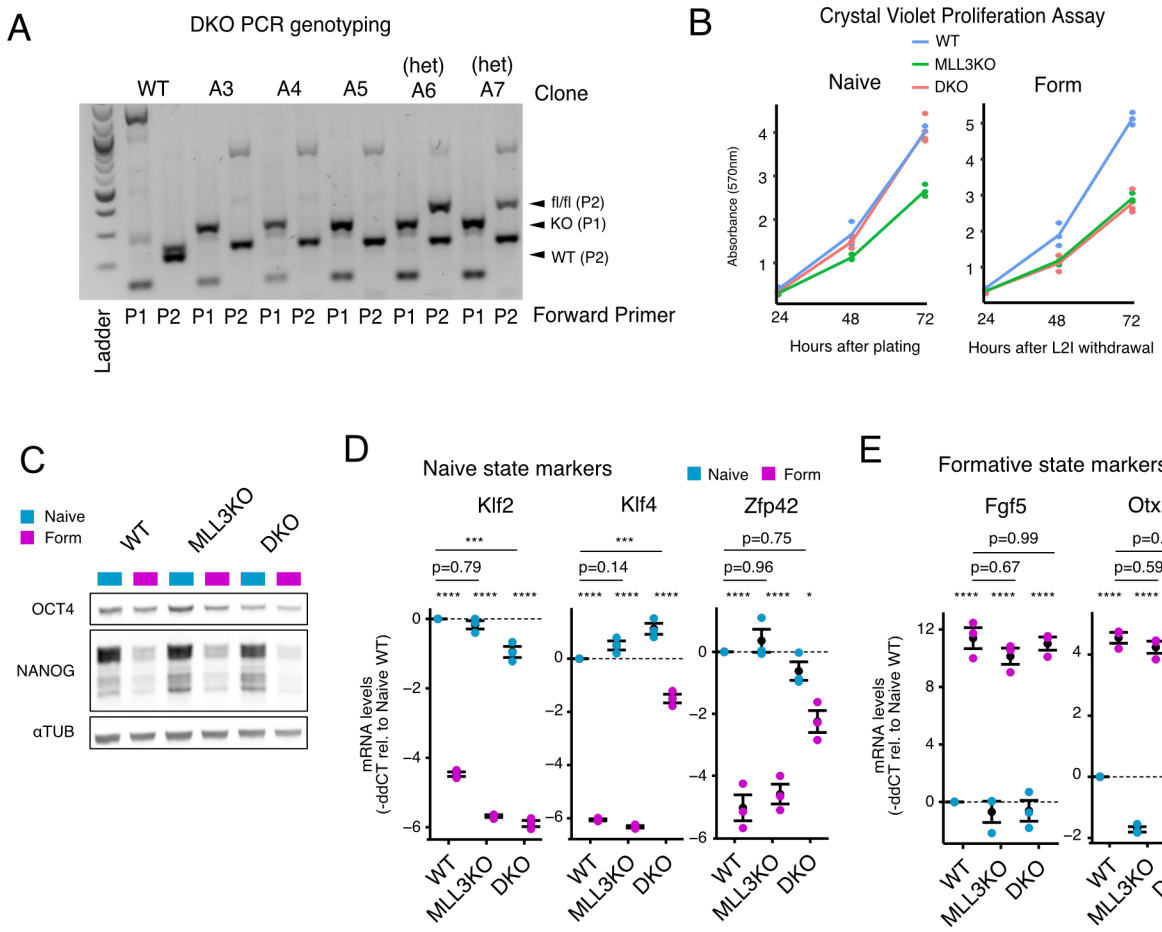
D



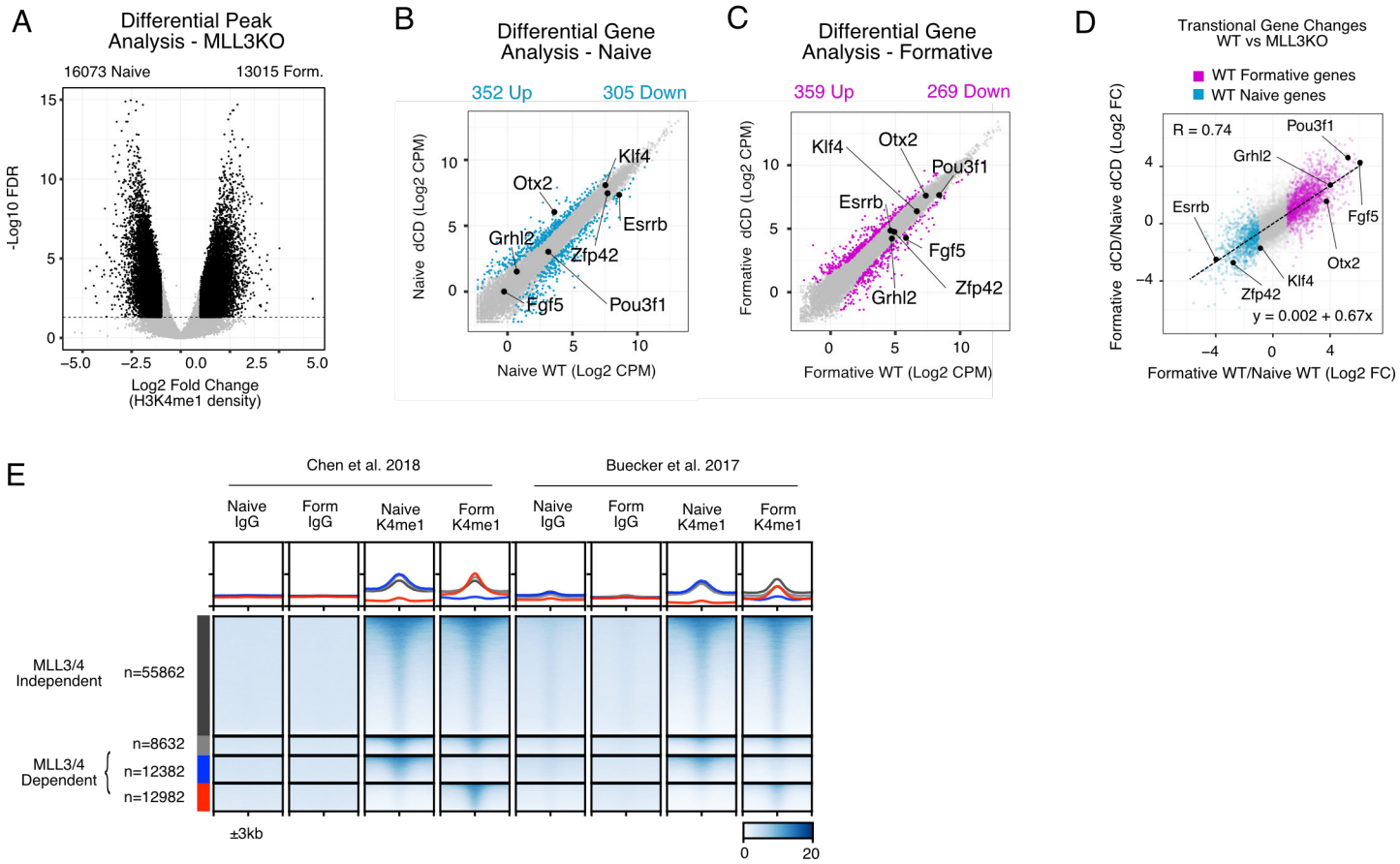
E



Supplemental Figure 1

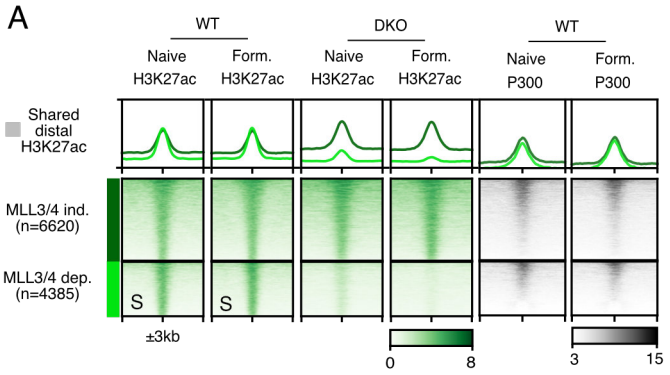


Supplemental Figure 2

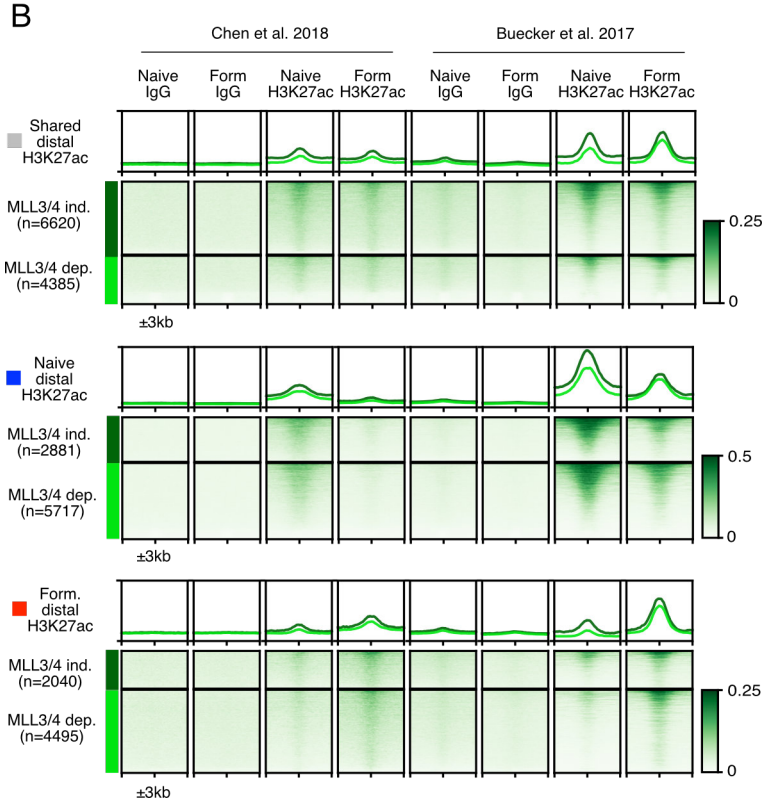


Supplemental Figure 3

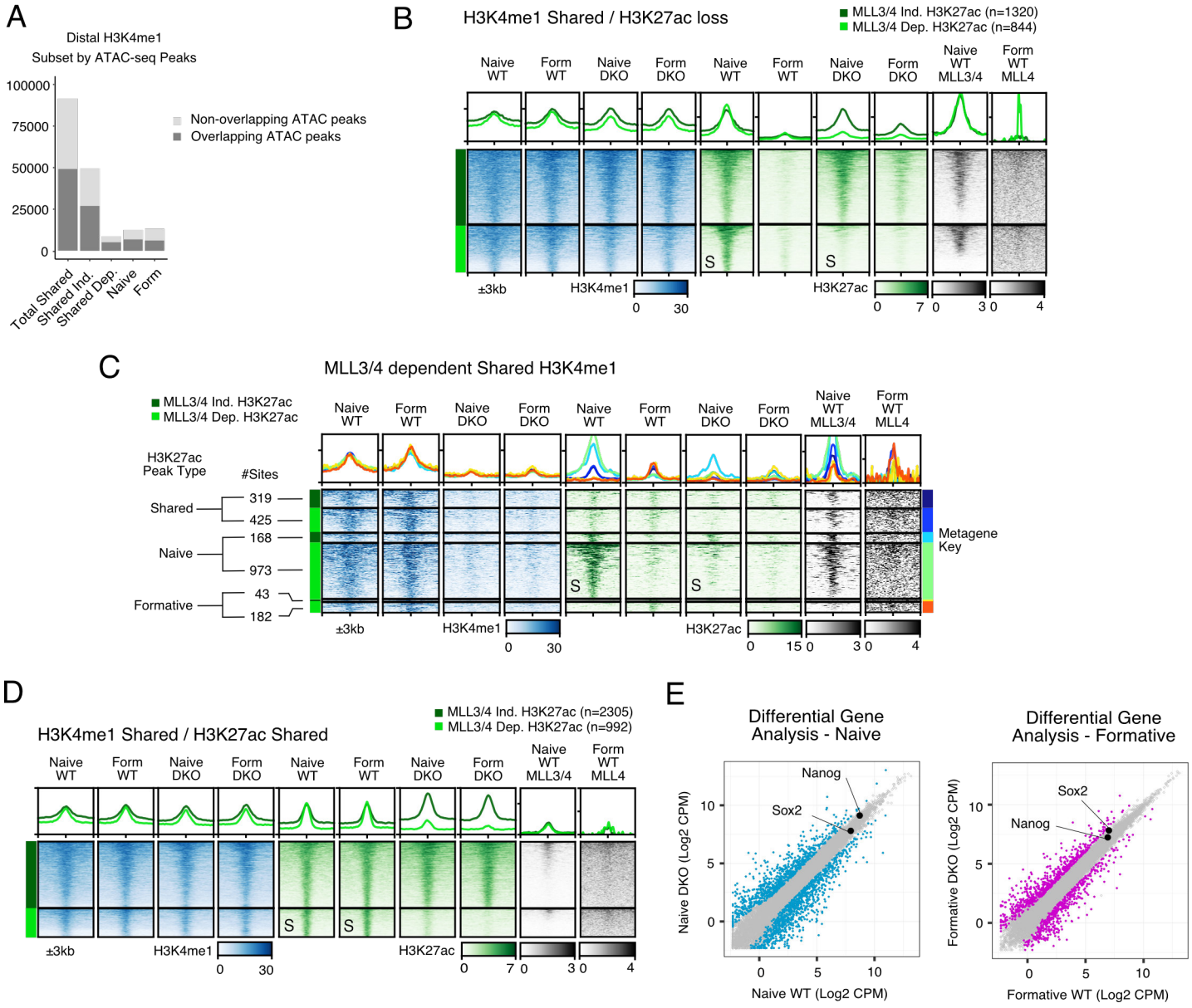
A



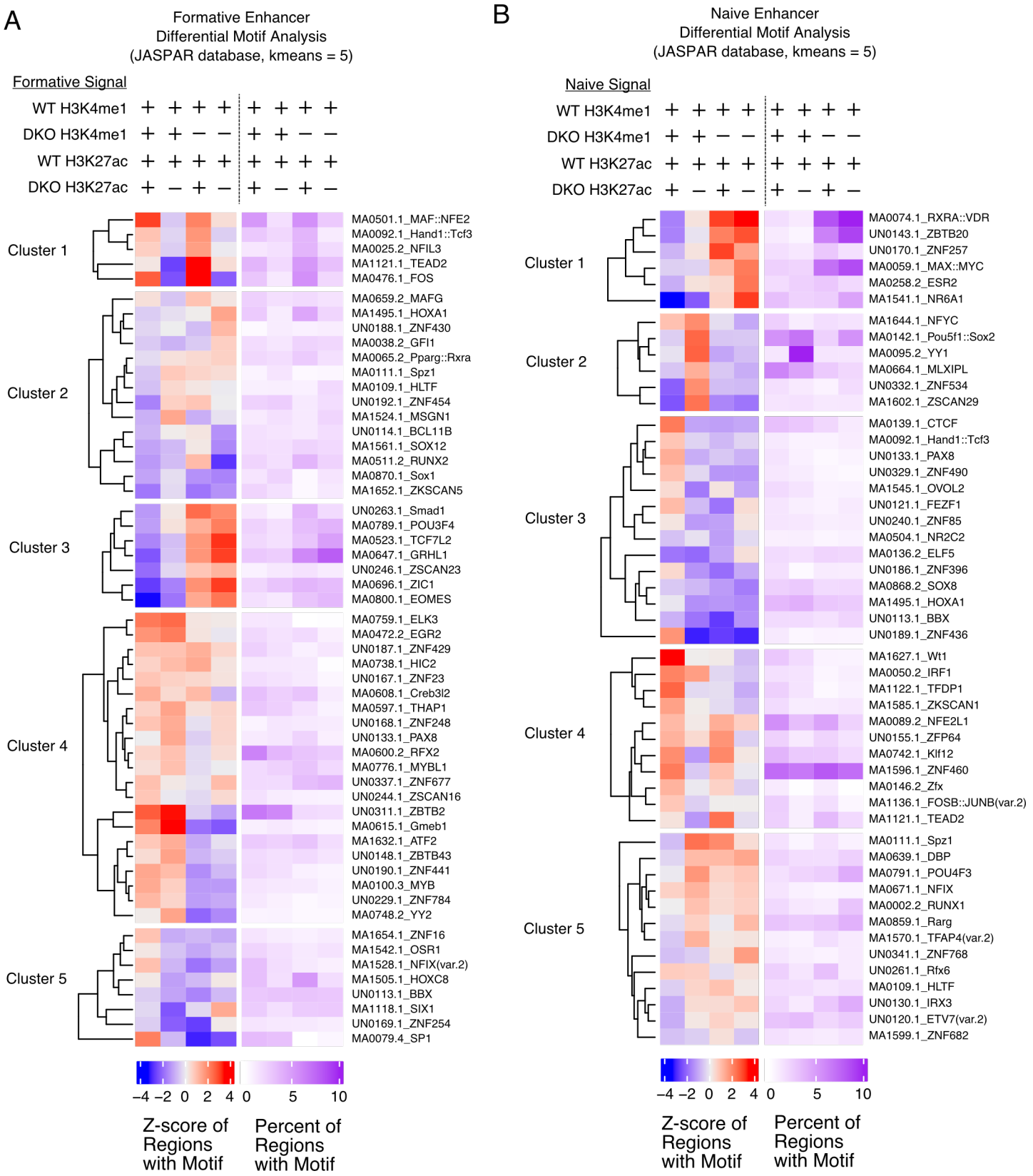
B



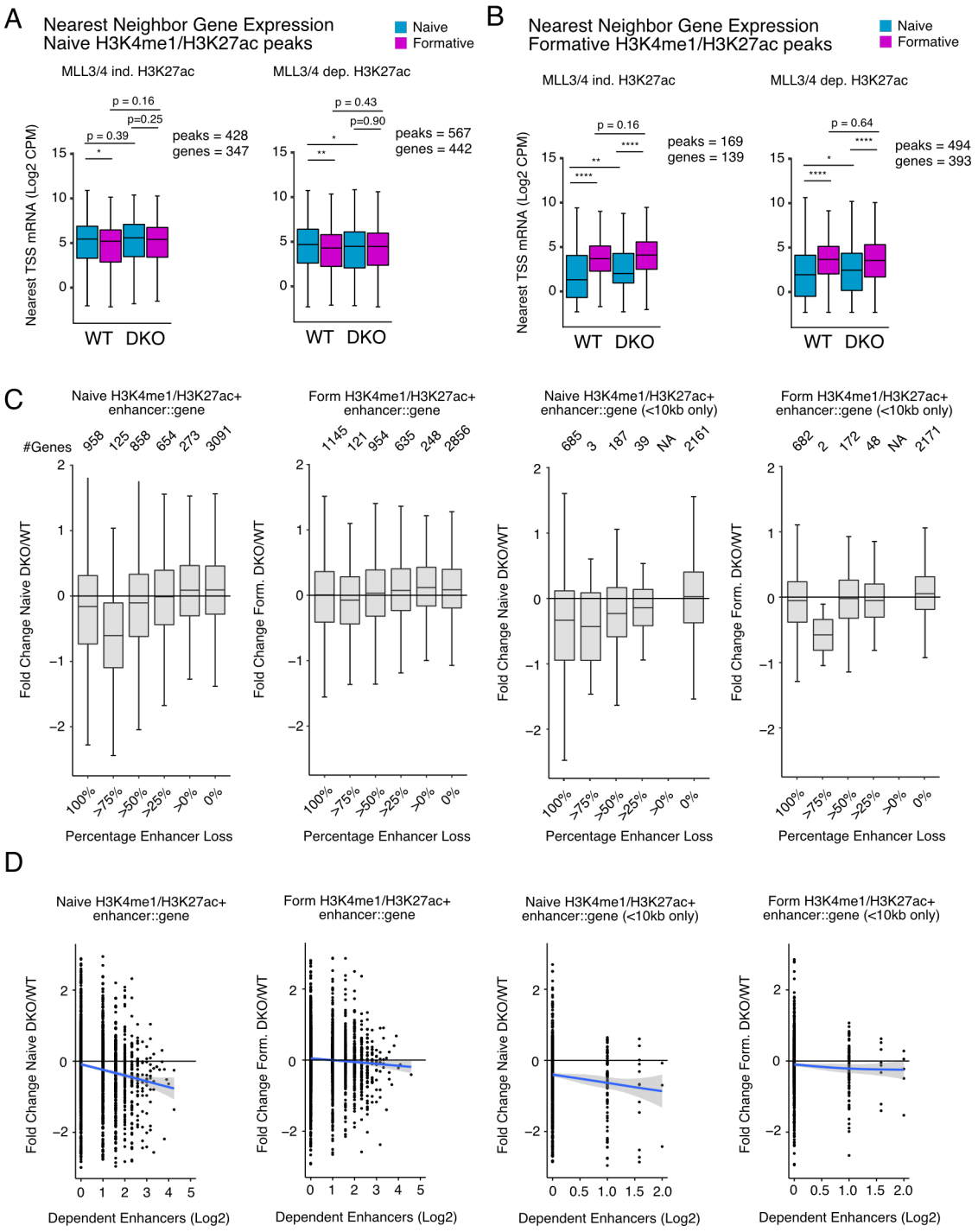
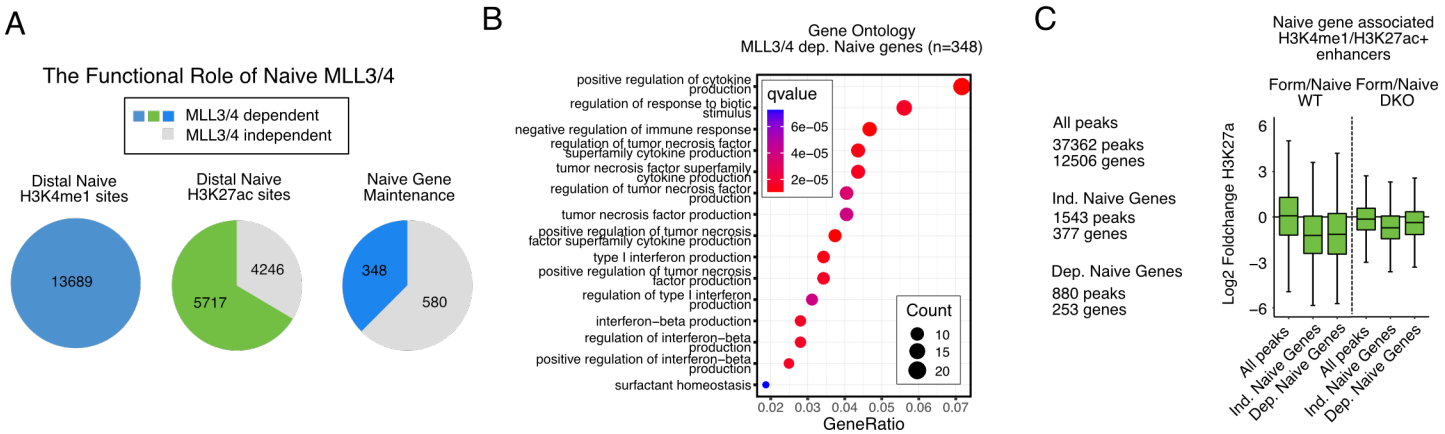
Supplemental Figure 4



Supplemental Figure 5



Supplemental Figure 6



Supplemental Figure 7