Trans-regulation of heterochromatin underlies genetic variation in 3D genome contacts

Haley J Fortin[1,2], Anna Z Struba[1], Christopher L Baker[1,2*]

1 The Jackson Laboratory, Bar Harbor, ME 04609, USA

2 Graduate School of Biomedical Sciences, Tufts University, Boston, MA 02111, USA

* Corresponding author. Tel: +1 802 288 6365; E-mail: christopher.baker@jax.org

Authors' contributions

CLB and HJF conceptualized the study and contributed to methodology. CLB provided resources. HJF and AZS investigated the study. HJF and CLB performed data curation and formal analysis. HJF and CLB performed visualization. HJF and CLB wrote the manuscript. HJF and CLB revised and edited the manuscript. CLB performed supervision. CLB performed funding acquisition.

Competing interests

The authors declare that they have no competing interests.

Keywords

3D genome organization, embryonic stem cells, epigenetics, genetics, genomics

Abstract

Genetic variation drives phenotypic diversity and disease susceptibility. *Trans*-acting genetic variation coordinates genome-wide chromatin changes, yet the molecular mechanisms underlying this regulation remain unclear. Here, we use the power of mouse genetics to investigate how genetic variation at *trans*-acting loci regulates 3D chromatin interactions.

Using HiChIP to map H3K27ac-associated regulatory elements in C57BL/6J and DBA/2J embryonic stem cells (ESCs), we identified 4,962 strain-differential interactions, 71% of which overlapped chromatin accessibility quantitative trait loci (QTL), establishing chromatin interaction variation is predominantly heritable. These differential interactions showed coordinated changes in chromatin state and gene expression, with stronger interactions associated with increased accessibility and transcription. Notably, loci regulated in *trans* exhibited a unique chromatin signature where weaker interactions were enriched for H3K9me3-marked heterochromatin. Analysis of F1 hybrids revealed dominant repressive effects, consistent with heterochromatin-mediated *trans*-regulation. To causally test this mechanism, we generated reciprocal congenic mouse strains carrying a Chr13 *trans*-QTL region. Integrated multiomic profiling of congenic ESC lines demonstrated that this single locus coordinates changes in H3K9me3, H3K27ac, chromatin accessibility, and 3D contact frequency at hundreds of distal targets. Consistent with our expectations, 73-83% of differential interactions at Chr13 *trans*-QTL targets changed in the predicted direction, demonstrating that heterochromatin-mediated *trans*-regulation coordinately regulates hundreds of regulatory loci.

This work establishes heterochromatin formation as a mechanism by which genetic variation at *trans*-acting loci coordinates changes across chromatin accessibility, histone modifications, and 3D genome organization, providing a framework for understanding how early developmental chromatin states could generate phenotypic variation while preserving essential developmental programs.

Background

In humans, genetic variation has a profound impact on normal health and disease. The majority of loci associated with variation in complex traits are concentrated in *cis* regulatory elements (CREs), such as promoters, enhancers, and insulators, rather than coding regions [1–3]. Of interest, many of these regulatory elements are active during early development [1,4], suggesting potential developmental origins of adult-onset diseases. The activity of regulatory elements is modulated through DNA and histone modifications, chromatin accessibility, and three-dimensional (3D) genome architecture [5–8]. Given that 99% of the human genome is non-coding, understanding how genetic variation influences CREs is critical for deciphering genome function [9,10], normal development, and, when these processes goes awry, the etiology of developmental disorders and disease.

Sequence variation can alter gene expression locally in *cis*, i.e., on the same chromosome (Chr), or distally, on another chromosome, in *trans*. *Cis*-regulatory variants affect nearby genes through changes in local CREs, whereas *trans*-regulatory effects are mediated by diffusible factors, often transcription factors or chromatin regulators, that can act on many CREs throughout the genome [11]. There is a growing interest in addressing how *trans*-regulation impacts complex traits and disease [12–18]. While effect sizes, tissue-specificity, and multiple testing burden similarly challenge both human and mouse mapping studies [15,16], the reduced power to detect *trans* quantitative trail loci (QTL) in human cohorts is largely driven by allele frequency. Rare variants often exhibit larger effects but require very large sample sizes for detection, whereas common variants, that typically exhibit weaker effects, can be mapped with smaller cohorts [19–21]. In contrast, mapping populations such as the mouse Diversity Outbred (DO) and BXD recombinant inbred panel effectively make rare variants common, enabling more powerful mapping of *trans*-QTL [22–25]. Using the BXD recombinant inbred panel, whose parental strains C57BL/6J (B6) and DBA/2J (D2) differ at over 5,000,000 single nucleotide variants (SNVs) [26], we previously identified both *cis* and *trans* quantitative trait loci (QTL) regulating chromatin accessibility, histone modifications, and gene expression in mouse embryonic stem cells (mESCs) and male germ cells [27,28]. In ESCs, we mapped several *trans*-QTL hotspots on Chrs 4, 5, 7, 12, and 13 that regulate coordinated changes in both chromatin accessibility and gene expression, impacting gene networks important for pluripotency and cell fate determination [28]. Interestingly, genes associated with development and disease have also been independently mapped to closely overlapping intervals within the Chr13 region [27,29–33]. This

66 Chr13 region contains a cluster of KRAB zinc-finger proteins (KZFPs), known transcriptional repressors that act
67 in *trans* to establish heterochromatin. Despite success in identifying *trans*-QTL in both mice and humans
68 [27,28,34–36], their molecular mechanisms remain poorly understood.

69 Chromosome conformation capture (3C) technologies [37–42] have revealed regulatory roles for 3D
70 genome organization in gene regulation [43–48], from regulatory loops connecting CREs to their target genes
71 [49–52] to topologically associated domains [53–55]. Local genetic variation alters 3D genome structure in *cis*,
72 helping determine how SNVs are associated with disease risk [56] and how structural variants (SVs) impact
73 development of disorders ranging from rheumatoid arthritis and sex reversal to limb malformations [57–59].
74 Systems genetic approaches have also identified how genetic variation among human lymphoblastoid cell lines
75 influences diverse features of chromatin state and 3D genome architecture [60]. While these studies have
76 established how *cis*-acting genetic variants disrupt 3D chromatin organization, the mechanisms by which *trans*-
77 acting genetic variation influences 3D genome architecture remain largely unexplored. Even when *trans*-QTL
78 are detected, the molecular mechanisms by which single genetic loci coordinate chromatin state changes at
79 hundreds of distal genomic regions are largely unknown. Understanding these mechanisms is essential for
80 interpreting how genetic variation influences complex traits and disease susceptibility, and the contribution of
81 variation to early developmental regulation.

82 Here, we investigated molecular mechanisms underlying *trans*-regulation of 3D genome organization.
83 Using HiChIP to map H3K27ac-associated chromatin interactions in B6 and D2 mouse embryonic stem cells,
84 we identified that contact anchors of differential interactions substantially overlapped QTL associated with
85 chromatin accessibility (caQTL) and gene expression (eQTL). Integrated epigenomic profiling revealed that
86 heterochromatin formation distinguished *trans* from *cis* regulatory mechanisms. To causally test this mechanism,
87 we performed reciprocal genetic perturbations by exchanging a *trans*-QTL region between otherwise isogenic
88 B6 and D2 genetic backgrounds, enabling independent validation of predicted chromatin changes in both
89 directions. Multiomic profiling of these congenic mESCs validated that this *trans*-acting locus coordinates
90 chromatin state at hundreds of distal regulatory loci with high concordance. Together, these findings establish
91 *trans*-regulation through heterochromatin formation as a key mechanism that coordinates changes in the
92 regulatory landscape.

Results

Genetic variation alters chromatin interactions and coordinates gene expression

To investigate how genetic variation impacts 3D genome architecture at CREs, we performed HiChIP [39] on mESC lines derived from B6 and D2 embryos. HiChIP was performed using the histone modification H3K27ac to enrich for 3D interactions at active enhancers. In total, this approach identified a combined set of 90,098 high-confidence interactions across both genetic backgrounds, defined as those spanning greater than 5 kb, a false discovery rate (FDR) < 0.01, and supported by at least four paired-end tag (PET) counts in at least two samples. Independently derived replicate lines from the same strain showed high concordance, while between-strain comparisons revealed reduced concordance, indicating that genetic background, rather than line-specific variation, drives chromatin organization differences (Fig. 1a and Additional file 1: Fig. S1a). These interactions predominantly connected enhancers to promoters (e-p, 84%), or enhancers to other enhancers (e-e, 14%). To understand the functional impact of 3D contacts, we integrated RNA-seq data previously collected from the same ESCs [28]. We observed that gene expression at connected promoters increases with increasing number of chromatin interactions (All pairwise comparisons Tukey p < 2e-16; Fig. 1b), consistent with previous findings [61,62]. Furthermore, genes annotated for “stem cell population maintenance” (MGI GO:0019827) are enriched among the upper quartile of connected genes (Fisher’s Exact Test p < 0.032), consistent with the observation that highly connected genes represent cell type–specific function [63,64]. Among the high-confidence interactions, 4,962 (5.5%) were significantly different between B6 and D2 ESCs (Benjamini-Hochberg corrected p < 0.2; Fig. 1c). To validate HiChIP results, we used 3C-qPCR [65] on three differential interactions (DIs) and confirmed the directional trends of strain-specific interaction patterns (Additional file 1: Fig. S2-4).

Differential chromatin interactions between B6 and D2 ESCs correspond with coordinated changes in chromatin accessibility and gene expression. The *Zfp951* locus provides an example of coordinated regulation. Specifically, B6 shows higher PET counts at both e-e (Interaction 1) and e-p (Interaction 2) interactions compared to D2 (Fig. 1d, e). Increased 3D contact in B6 is consistent with higher epigenetic marks associated with increased CRE activity (e.g., chromatin accessibility and H3K4me3 enrichment) in the loop anchor region (Fig. 1D yellow highlight) and higher *Zfp951* expression in B6 (Fig. 1f). This pattern suggests a shared regulatory mechanism coordinating changes in CRE activity, 3D contacts, and gene expression. Genome-wide, DIs showed coordinated strain-specific changes in both chromatin accessibility and interaction strength (Fig. 1g). The strain-

121 specific bias seen for *Zfp951* expression generally holds for all DIs (Fig. 1h); chromatin interactions that are
122 stronger in either B6 or D2 coincide with higher expression of anchor-connected genes. In contrast, genes
123 connected to non-DIs, such as the housekeeping gene *Gapdh*, showed minimal expression bias ($\log_2$(B6/D2) =
124 0.05416) (Fig. 1h, Additional file 1: Fig. S1b, c). Together, these data demonstrate that genetic variation
125 coordinately modulates CRE activity, 3D interaction strength, and expression of interacting genes, but leaves
126 open the question of the mechanism of regulation.

127 **Genetic variation regulates chromatin interactions through both *cis*- and *trans*-acting mechanisms**

128 To test whether coordinated changes in chromatin state and gene expression reflect shared genetic
129 control, we determined if DI anchors are enriched for genetically-regulated CREs by examining overlap with
130 previously identified caQTL targets [28]. For instance, both gene expression of *Zfp951* and accessibility of an
131 upstream CRE (Chr5:104,727,903 bp) were mapped to the same Chr5 locus (Fig. 2a, b, c), with B6 haplotypes
132 conferring higher activity. In total, five interaction anchors contacting *Zfp951* overlap eight *cis*-caQTL all showing
133 higher B6 accessibility [28], demonstrating that individual anchors can harbor multiple caQTL. This pattern
134 extends genome-wide; 71.7% of DI anchors are enriched for caQTL targets at one (42.6%) or both (29.1%)
135 anchors (QTL at one and both anchors $p < 0.001$, Fig. 2d; Additional file 1: Fig. S1d, e), indicating that contact
136 frequency at anchors for most DIs are under genetic regulation. Further, these data suggest that genetic
137 regulation coordinates both chromatin accessibility and 3D genome structure.

138 Given QTL enrichment at DI anchors, we tested whether parental interaction strength in B6 and D2
139 correlates with QTL effect sizes measured across the BXD panel [28]. Interaction strength showed moderate but
140 highly significant positive correlation with chromatin accessibility at *cis* (Pearson r = 0.41, p = 1.4e-105) and *trans*
141 (r = 0.45, p = 7.1e-100) caQTL targets (Fig. 2e, f). We further explored the link between 3D contacts and gene
142 regulation (Fig. 2g, h) and identified a similar correlation at *cis* (r = 0.34, p = 2.5e-30) and *trans* (r = 0.15, p =
143 4.1e-3) eQTL targets. Beyond correlation magnitude, directional concordance provides stronger evidence for
144 coordinated regulation; loci where the B6 allele shows higher interaction strength might be expected to exhibit
145 B6-biased molecular phenotypes. Indeed, at the categorical level, we found significant allele concordance
146 between interaction strength and chromatin accessibility or gene expression at both *cis*- and *trans*-QTL targets
147 (Fig. 2i, Pearson's Chi-squared test all $p < 6.7e-6$). The high allelic concordance between 3D architecture,

148 chromatin accessibility, and gene expression at both *cis*- and *trans*-QTL suggests shared regulatory mechanisms
149 coordinate these molecular layers. Moreover, the presence of *trans*-acting effects indicates that diffusible
150 chromatin regulators mediate long-range genetic control of chromatin organization.

151 **Heterochromatin formation regulates *trans*-acting chromatin interactions**

152 While both *cis* and *trans*-QTL showed coordinated effects on chromatin accessibility and 3D interactions,
153 the molecular mechanism distinguishing these regulatory modes remained unclear. We hypothesized that
154 distinct mechanisms underlie *trans* and *cis* regulation. To test this, we examined enrichment of DNA-binding
155 factors at interaction anchors that overlap *trans*-regulated loci (i.e., *trans*-targets). Supporting our previous
156 observations [28], we identified TRIM28 (KAP1) as the most highly enriched factor at *trans*-targets (Fig. 3a).
157 Given TRIM28's role in establishing heterochromatin through H3K9me3 deposition [66,67], we hypothesized that
158 *trans*-targets would show enrichment for repressive chromatin marks. Specifically, at loci with DIs, we predicted
159 H3K9me3 enrichment in the strain with weaker contacts and H3K27ac enrichment in the strain with stronger
160 contacts. To test this, we performed H3K9me3 and H3K27ac ChIP-seq in independently derived mESC lines
161 from three B6 and three D2 mice (Additional file 1: Fig. S5). At a putative CRE downstream from *Sox13* (Fig. 3b,
162 c, d), D2 shows weaker chromatin interactions than B6 and increased H3K9me3 compared to B6 at the anchor
163 overlapping a *trans*-target (Fig. 3b, yellow highlight). Conversely, the same *trans*-target showed increased
164 H3K27ac in B6 compared to D2. To broadly capture this trend, we quantified the level of both H3K27ac and
165 H3K9me3 at both *cis* and *trans*-targets and correlated strain-specific changes in histone modifications with
166 changes in chromatin accessibility (Fig. 3e, f, g, h). When considering H3K27ac, we identified a strong positive
167 correlation at *cis*- (Pearson $r = 0.83$, $p < 2.2e-16$) and *trans*- ($r = 0.91$, $p < 2.2e-16$) caQTL targets, indicating that
168 both signify active chromatin. The key distinction in chromatin state is with H3K9me3. While *cis*-targets showed
169 little correlation between H3K9me3 and accessibility (Fig. 3g), *trans*-targets showed a strong negative correlation
170 ($r = -0.66$, $p = 5.8e-143$; Fig. 3h), indicating that heterochromatin formation specifically characterizes *trans*-
171 regulation. Critically, *trans*-targets were significantly more likely than *cis*-targets to show discordant chromatin
172 states – defined as opposing changes in accessibility and H3K9me3 enrichment ($p < 2.2e-16$, odds ratio = 20.49,
173 logistic regression; Fig. 3i), establishing heterochromatin formation as a defining feature of *trans*-regulation.
174 Together, these data support a molecular mechanism in which genetic variation in *trans*-acting factors drives

175 TRIM28-mediated heterochromatin formation via H3K9me3 deposition, restricting chromatin accessibility and
176 weakening 3D contacts at distal loci.

177 **Chr13 *trans*-QTL coordinates chromatin accessibility and gene expression in an allele-specific manner**

178 To determine the effect of *trans*-regulation on coordinated chromatin regulation and gene expression, we
179 focused on a compound *trans*-QTL on Chr 13 that regulated the expression of 59 genes (eGenes) and 2,033
180 putative CREs (*trans*-targets) with LOD $\geq$ 5 [28]. Spatial coordination of *trans*-targets and eGenes was assessed
181 for proximity enrichment using a modified Region Associated Differentially expressed gene approach [68], testing
182 whether chromatin accessibility peaks cluster near transcriptionally regulated genes. Chr13 *trans*-targets
183 showed higher enrichment near Chr13 eGenes compared to all other *trans*-QTL eGenes (Fig. 4a), indicating
184 distal *trans*-regulation drives coordinated changes in chromatin accessibility and gene expression at co-localized
185 target loci. Critically, this spatial coordination was directionally concordant with QTL effects; chromatin peaks
186 with higher accessibility associated with the B6 allele at the QTL are specifically enriched near eGenes with
187 higher expression when the genotype at the QTL is B6 (Fig. 4b, left), and the reciprocal pattern held for D2 (Fig.
188 4b, right). These analyses demonstrate that independently measured chromatin accessibility and gene
189 expression are spatially coordinated in an allele-specific manner, supporting *cis*-propagation of *trans*-QTL
190 through localized chromatin remodeling.

191 To determine the nature of the QTL effect on chromatin accessibility, we performed ATAC-seq on three
192 independently derived mESCs from (B6xD2)F1 hybrids. Focusing on high-confidence caQTL (LOD > 8), *trans*-
193 caQTL showed dominant repression, where F1 hybrid mESCs resembled the parent with lower accessibility
194 rather than intermediate (additive) values. This is exemplified by the Chr13 *trans*-target at Chr14:122776397,
195 where the F1 showed the same (dominant) repressed accessibility as the D2 parent (Fig. 4c, left). In contrast, a
196 *cis*-QTL target at Chr5:104727903 showed additive inheritance with F1 chromatin accessibility intermediate
197 between the parent lines (Fig. 4c, right). This distinction between *trans* and *cis* effects extended genome-wide;
198 *trans*-caQTL showed dominant repression (median r = -0.95), while *cis*-caQTL showed near-additive inheritance
199 (median r = -0.04; Wilcoxon $p < 2.2e-16$, effect size r = 0.61; Fig. 4d). Chr13 *trans*-caQTL targets specifically
200 showed strong dominant repression (median = -1.04), with the D2 allele at the Chr13 QTL associated with
201 reduced accessibility at most targets (84%, N = 1,709) (Fig. 4e).

202 The dominant repressive inheritance pattern supports a model in which the Chr13 QTL encodes one or
203 more diffusible repressors that establish heterochromatin at *trans*-targets. To causally test whether this locus
204 coordinates heterochromatin formation and 3D architecture, we generated reciprocal congenic mouse strains
205 that exchange the Chr13 *trans*-QTL. The first strain, B6.D2-(rs30079041-rs51756099)/Bakr (hereafter referred
206 to as B6-13D2), was bred to contain the D2 Chr13 congenic region (Chr13:52,765,995-71,064,928) on an
207 otherwise B6 background. Conversely, D2.B6-(rs29224292-rs51756099)/Bakr (hereafter referred to as D2-
208 13B6) was bred to carry the B6 Chr13 congenic region (Chr13:47,066,720-71,064,928) on an otherwise D2
209 background (Fig. 4f). Together, these congenic lines provide ideal resources to investigate the impact of the
210 Chr13 *trans*-QTL on heterochromatin formation.

211 **Congenic strains confirm Chr13 *trans*-QTL drives heterochromatin formation at *trans*-targets**

212 To test whether the Chr13 QTL drives *trans*-regulation of heterochromatin, we tested predictions based
213 on QTL effects of BXDs on reciprocal mESCs from congenic strains. If Chr13 is causal, swapping the QTL
214 region should alter chromatin state according to the genotype introduced. Specifically, B6-13D2 should
215 resemble D2 at *trans*-targets, while D2-13B6 should resemble B6. We tested these predictions using
216 H3K9me3 and H3K27ac ChIP-seq, and ATAC-seq in B6, D2, B6-13D2, and D2-13B6 mESCs (Additional file 1:
217 Fig. S6).

218 If repression is mediated by H3K9me3-marked heterochromatin, we predicted *trans*-targets should show
219 increased H3K9me3 in B6-13D2 (with D2 Chr13) compared to B6. Strikingly, 77% of Chr13 *trans*-caQTL targets
220 showed the predicted H3K9me3 increase in B6-13D2 (Fig. 5a, g, left; exact binomial test $p < 2.2e-16$).
221 Reciprocally, 89% showed decreased H3K9me3 in D2-13B6 (with B6 Chr13) compared to D2 (Fig. 5b, left, g; p
222 $< 2.2e-16$). Of the remaining targets (324 loci predicted to be repressed when the allele at the QTL is B6), 69%
223 showed the inverse pattern where H3K9me3 increased in D2-13B6 and decreased in B6-13D2 (Fig. 5b, c, right,
224 g). Consistent with H3K9me3-mediated repression, active chromatin marks showed the reciprocal pattern as
225 predicted. Targets gaining H3K9me3 in B6-13D2 showed reduced chromatin accessibility (77% of predicted
226 decreases validated; Fig. 5c, g) and reduced H3K27ac (94% validated; Fig. 5e, g). Similarly, targets losing
227 H3K9me3 in D2-13B6 gained accessibility and H3K27ac (Fig. 5d, f, g). Further, *trans*-targets that changed in the
228 predicted direction showed significantly larger effect sizes than those that did not (Fig. 5h), indicating non-

229 validated targets likely reflect measurement noise rather than biological exceptions. These reciprocal epigenomic
230 changes, validated through two independent genetic perturbations across three chromatin features, directly
231 demonstrate that Chr13 causally regulates hundreds of distal *trans*-targets through heterochromatin formation.

232 **Heterochromatin formation alters 3D contacts at *trans*-QTL target sites**

233 To validate that Chr13-driven heterochromatin formation alters 3D interactions at *trans*-targets, we
234 repeated H3K27ac HiChIP in B6 and D2, while including B6-13D2 and D2-13B6 mESCs (two technical replicates
235 per strain; Additional file 1: Fig. S7). We hypothesized that introducing the D2 Chr13 region in B6-13D2 would
236 reduce chromatin interaction at *trans*-targets (compared to B6), by increasing heterochromatin. Further,
237 introducing the B6 Chr13 region into D2-13B6 would strengthen interactions (compared to D2), by reducing
238 heterochromatin. Consistent with our hypothesis, at the *Litaf* locus we observed increased interactions in D2-
239 13B6 compared to D2 (Fig. 6a, b, Interactions 1 and 2 $p < 0.027$). Reciprocally, at the same locus we observe
240 weaker interactions in B6-13D2 compared to B6 (Fig 6a, b; Interactions 1 and 2, $p < 1.51e-3$). This is particularly
241 noteworthy because *Litaf* is a Chr13 eQTL target, highlighting a potential functional link between chromatin
242 architecture and gene regulation.

243 To test if heterochromatin formation underlies these changes in chromatin interactions, we examined
244 H3K9me3 enrichment at the interaction anchors. Critically, H3K9me3 showed the reciprocal pattern to 3D
245 contacts. At *Litaf*, D2-13B6 exhibited significantly lower H3K9me3 than D2 at the Chr13 *trans*-target; and B6-
246 13D2 showed increased H3K9me3 compared to B6 (Fig. 6c; DD13B $p = 4.81e-8$, BB13D $p = 2.98e-15$). These
247 examples demonstrate that the Chr13 QTL drives heterochromatin formation at interaction anchors, coincident
248 with altered 3D interaction strength. Further, in agreement with a heterochromatin-mediated mechanism,
249 H3K27ac showed the opposite patterns, with D2-13B6 exhibiting higher levels at the same anchor compared to
250 D2 at *Litaf* (Fig. 6d; $p < 9.24-103$), while B6-13D2 showed lower H3K27ac than B6 (Fig 6d; BB13D $p = 4.54e-$
251 119). Consistent DI patterns and opposing H3K9me3/H3K27ac profiles are observed at additional *trans*-
252 regulated loci, including *Zfp516* and *Fam241a* (Additional file 1: Fig. S8-9). Together these data confirm that
253 Chr13 regulates heterochromatin formation and 3D organization at these *trans*-targets.

254 To assess whether these examples reflect genome-wide patterns, we tested whether genetic differences
255 identified between parental strains (B6 and D2) predict changes in congenic strains. Remarkably, 73-83% of

differential interactions changed in the predicted direction (Fig. 7a, b, c; Exact binomial test: all $p < 1.5e-8$) Furthermore, interactions that change contact frequency as predicted generally showed larger effect sizes than those that did not (Fig. 7d), indicating non-validated targets may reflect measurement noise. These validation rates, achieved across two independent genetic perturbations and testing reciprocal predictions, demonstrate that Chr13 causally regulates chromatin state at hundreds of distal loci. The consistent validation across H3K9me3, H3K27ac, chromatin accessibility, and 3D contact frequency establishes *trans*-regulation of heterochromatin as a mechanism for coordinated regulation of the epigenome (Fig. 7e).

263 **Discussion**

264 Here, we establish heterochromatin formation as the molecular mechanism distinguishing *trans* from *cis*
265 regulatory effects on 3D genome organization in mESCs. *Trans*-regulated regions of the genome show a 20-fold
266 enrichment for discordant H3K9me3 and chromatin accessibility, providing a unique epigenetic signature
267 compared to *cis*-QTL. This mechanism is supported genetically by the dominant repressive pattern observed in
268 F1 hybrids. Critically, reciprocal genetic perturbations demonstrated that a single locus causally coordinates
269 chromatin state at hundreds of distal regulatory elements with high concordance across molecular assays.
270 Validation rates ranged from 73-83% for 3D contact frequency to 89-94% for H3K9me3 and H3K27ac,
271 demonstrating that the coordinated mechanism is robust across the epigenome. Together, these findings explain
272 how *trans*-acting genetic variation achieves coordinated regulation across molecular layers through targeted
273 heterochromatin deposition.

274 Consistent with previous observations [60,62,69,70], we find that 3D genome interactions are positively
275 correlated with gene expression and active histone modifications, with contact frequency at individual loci
276 associated with increased transcription. Prior work in human induced pluripotent stem cells (iPSCs)
277 demonstrated that genetic variation affecting contact propensity results in coordinated changes in H3K27ac
278 [71,72]; however, these studies did not distinguish *cis* from *trans* regulation. Our findings extend beyond
279 correlations by establishing causal relationships through reciprocal genetic perturbations and identifying
280 heterochromatin formation as a defining feature of *trans*-regulation. Notably, the 20-25% of cases where
281 increased contacts did not correspond to increased expression suggest additional regulatory mechanisms
282 beyond direct connection between contact frequency and expression and warrant future investigation.

283 The functional genomic, genetic, and positional evidence continue to point to variation in genes that
284 encode KRAB Zinc-Finger Proteins (KZFPs) as probable molecular basis for *trans*-regulation. Firstly, KZFPs
285 provide DNA-binding specificity to recruit TRIM28, and chromatin modifying enzymes, to form heterochromatin
286 [66,73–77]. Consistent with this, TRIM28 is enriched at *trans*-QTL targets and a cluster of KZFPs is found in the
287 Chr13 QTL [78]. Second, the unique association of *trans*-QTL with H3K9me3 and their dominant repressive
288 effect in F1 hybrids indicates a diffusible silencing factor [27,28]. Third, KZFPs are highly variable within species,
289 indicating a source of rapid genetic variation [78,79]. The Chr13 region represents a compound QTL containing
290 structurally variant KZFP genes, each encoding distinct DNA-binding specificities; this genetic architecture

291 explains how a single locus coordinates changes at hundreds of *trans*-targets. Fourth, KZFPs have been
292 demonstrated to impact transcriptional repression through heterochromatin spreading from their DNA binding
293 sites [80], supporting the regional *cis*-acting effects we see on enhancer-promoter contacts. Lastly, the epigenetic
294 engineering approach of CRISPR interference (CRISPRi) [81] leverages a nuclease-dead Cas9 fused to a KRAB
295 domain to induce heterochromatin and silence target genomic loci, directly demonstrating *trans*-regulation of
296 heterochromatin. Building on this, recent work used CRISPRi to show that targeted H3K9me3 deposition led to
297 a reduction in 3D contacts between enhancers and promoters [69], demonstrating direct effect of
298 heterochromatin on 3D contacts in an inducible system.

299 We propose that *trans*-regulation of heterochromatin formation represents an evolutionarily
300 advantageous mechanism for generating regulatory variation without disrupting core transcription factor (TF)
301 networks. Classical models of *trans*-regulation invoke changes in TF expression or sequence, which would
302 coordinately affect all targets genome-wide. However, developmental TFs and their core gene regulatory
303 networks are under strong purifying selection due to their essential roles in lineage specification; and large
304 variation in these factors likely results in severe phenotypic consequences. In contrast, KZFP DNA-binding
305 domains evolve rapidly through structural variation [82,83], and specifically at Chromosome 13 in mice [78],
306 potentially enabling population-level variation in which genomic regions are differentially targeted for
307 heterochromatin formation based on the repertoire of zinc-finger domains present in an individual. Recent cross-
308 species comparisons between human and macaque cells demonstrate that *trans*-regulatory divergence is
309 specifically enriched at transposable elements bearing KZFP footprints [84], providing evolutionary support that
310 heterochromatin variation drives regulatory divergence between populations while preserving core
311 developmental networks. Further, heterochromatin acts as a barrier to pioneer transcription factor binding during
312 cellular reprogramming [85,86], and depletion of a specific KZFP reduced this barrier and enhanced
313 reprograming efficiency [87]. The coordinated regulation we observe here reveals a powerful and
314 underappreciated layer of genome control where a single locus affects heterochromatin deposition and therefore
315 chromatin accessibility and 3D contacts at hundreds of regulatory elements. By allowing population-level
316 variation in heterochromatin targeting rather than disrupting essential TF function, this mechanism may provide
317 standing regulatory variation necessary for evolutionary adaptation while maintaining developmental robustness.

The developmental implications of heterochromatin-mediated *trans*-regulation established in early pluripotency may extend to adult phenotypes. Heterochromatin can act as a barrier to lineage specification [85]; and variation in early heterochromatin formation may be propagated, creating a persistent epigenetic state that influences lineage commitment and cellular function. Supporting functional consequences, the Chr13 region investigated here has been associated with multiple developmental phenotypes in mice including craniofacial development, limb malformation, lupus susceptibility, meiotic recombination frequency, and imprint stability [27,29–33]. The mechanism we establish in mouse ESCs appears conserved in human biology. Inter-individual variation in heterochromatin landscapes across diverse human iPSC lines correlates with differentiation propensity for early neural fates versus other lineages [88], consistent with the model that heterochromatin variation creates tunable barriers affecting lineage-specific transcription factor access. This suggests population-level variation in heterochromatin may contribute to phenotypic diversity in humans through the same mechanism. Further, KZFPs show dynamic expression beyond pluripotency, specifically upregulated in adult immune cells and neurons [83]. Consistent with functional relevance, genomic regions containing KZFPs in humans contain signatures of recent selection and are associated with adult disease [89]. Together, these observations suggest that early variation in heterochromatin establishment may act to constrain developmental outcomes and contribute to phenotypic variation in complex traits.

Our study provides strong genetic and epigenetic evidence that heterochromatin formation mediates *trans*-regulation of 3D genome organization and chromatin state, but several limitations remain. First, all experiments were performed in mESCs, and whether similar mechanisms operate in differentiated tissues, other species, or in vivo is unknown. Second, the Chr13 congenic interval spans a large region and contains multiple KZFPs, making it unclear whether *trans*-regulation reflects the combined effects of this region or a single causal gene. Fine-mapping, and targeted perturbations, such as overexpression or CRISPR-based manipulation, will be needed to identify the causal gene(s). Finally, while reciprocal congenic experiments validate chromatin changes, functional consequences for pluripotency, lineage specification, and organismal phenotypes remain to be tested; and extension to human systems will be essential to establish relevance to complex traits and disease.

Conclusions

We establish heterochromatin formation as a defining chromatin signature distinguishing *trans* from *cis* regulatory mechanisms in mouse embryonic stem cells. Through reciprocal genetic perturbations, we

346 demonstrate that genetic variation at a single *trans*-acting locus regulates chromatin accessibility, histone
347 modifications, and 3D interactions at hundreds of distal regulatory elements. This work provides a framework for
348 understanding how *trans*-regulation operates mechanistically through targeted heterochromatin formation to
349 reshape regulatory landscapes and suggests that genetic variation in early developmental chromatin states may
350 contribute to phenotypic diversity in complex traits and disease susceptibility.

Methods

Mice

All mice were obtained from The Jackson Laboratory (Bar Harbor, ME) including C57BL/6J (JR# 000664) and DBA/2J (JR# 000671) parental inbred strains. B6 and D2 parents were crossed to the F3 generation prior to marker-based genotyping and subsequent backcrossing for a minimum of ten generations for form the B6.D2-(rs30079041-rs51756099)/Bakr (JR# 37111) and D2.B6-(rs29224292-rs51756099)/Bakr (JR# 37112) congenic lines. Congenic genotypes were confirmed using MiniMUGA array [90] (Neogen) for which the background analysis report is available for a representative mouse (B6-13D2: Additional file 2; D2-13B6: Additional file 3). All animal experiments were approved by the Animal Care and Use Committee of The Jackson Laboratory (Summary #04008). Congenic lines are available upon request.

Derivation and maintenance of mouse embryonic stem cells

Derivation of mouse embryonic stem cells for B6-13D2 and D2-13B6 strains were performed, after backcrossing for a minimum of 10 generations, using protocols outlined in Czechanski et al. [91]. For all cultures, mouse ESCs were seeded with irradiated mouse embryonic fibroblasts (MEFs) from B6 animals and cultured in DMEM high glucose base medium supplemented with 15% fetal bovine serum (FBS, Lonza, cat. no. 14-501F lot no. 0000217266), 1X Pen/Strep, 2 mM GlutaMAX, 1 mM sodium pyruvate, 0.1 mM MEM-NEAA, 0.1 mM 2-mercaptoethanol, 103 IU LIF, 1 μM PD0325901, and 3 μM CHIR99021. Mouse ESCs were expanded for 2–3 days to reach ~70% confluency for molecular assays. For functional genomic assays, mouse ESCs were enzymatically disassociated into single cell suspensions using trypsin and MEFs were depleted by incubation on gelatine coated plates for 20 minutes to allow to settle. All cell lines tested negative for mycoplasma and are available upon request. Genotyping was performed to determine zygosity of congenic region and sex of cell lines. Primers for KASP genotyping used SNVs within the congenic region to distinguish between B6 and D2 backgrounds. All primers used for genotyping are listed in Additional file 4, Table S1.

ATAC-seq sample preparation and sequencing

ATAC-seq experiments were performed, using two or three independent cultures from the same cell line per strain, following the FAST-ATAC method as described [92], using 100,000 cells. DNA quality and quantity was determined using High Sensitivity D5000 ScreenTape (Agilent Technologies). Libraries were sequenced to a read depth of about 40 million reads per sample on an Illumina NovaSeq X Plus with a 150 bp read length.

379 **Chromatin Immunoprecipitation (ChIP) sample preparation and sequencing**

380 ChIP-seq experiments performed with B6 and D2 used three independently derived cell lines from each strain.
381 Validation ChIP-seq experiments performed with B6, D2, B6-13D2, and D2-13B6 used two or three
382 independent cultures from the same cell line for each strain. For each ESC ChIP-seq library, 3 million cells
383 were harvested and fixed as previously described [27]. Cell lysis, chromatin fragmentation, and immuno-
384 precipitation were performed as described [93], with minor modifications. For H3K27ac ChIP, sodium butyrate
385 was added to hypotonic lysis buffer, MNase buffer, RIPA buffer (post-overnight antibody conjugation), RIPA
386 wash buffer, and TE to inhibit histone deacetylase activity. Immunoprecipitation was performed using
387 antibodies against H3K27ac (12 µL, abcam, ab4729, lot# GR312658-1; Cell Signaling Technology, 8173, lot#
388 9) and H3K9me3 (6 µL, Active Motif, 39161, lot# 09919003). ChIP-seq libraries were constructed using the
389 KAPA HyperPrep Kit (Roche Sequencing and Life Science). Quality and concentration of libraries were
390 assessed using the High Sensitivity D5000 ScreenTape (Agilent Technologies) and KAPA Library
391 Quantification Kit (Roche Sequencing and Life Sciences). Libraries were sequenced to a read depth of about
392 40 million reads per sample (75 bp) on an Illumina NextSeq or NovaSeq X Plus.

393 **HiChIP sample preparation and sequencing**

394 *Restriction enzyme based HiChIP*

395 HiChIP experiments were performed as described [39], using two samples from independently derived cell
396 lines per strain. We used 10 million cells, DpnII (NEB, R05443), and the following 3C modifications [94] to
397 improve digestion, incorporation, ligation, and removal of dangling ends. Specifically, 1) biotin incorporation
398 was performed at 23°C for 4 hours with interval shaking (10 sec at 900 rpm, 5 min off). 2) Ligation was
399 performed at 16°C for 4 hours with interval shaking (1 min at 500 rpm, 4 min off). 3) Samples were sheared
400 using a Diagenode Biorupter with the following parameters: tube size = 1.5 mL, time on = 30 sec, time off = 30
401 sec, cycle # = 8. 4) A total of 3 µL of H3K27ac antibody (Abcam #ab4729) was added to the sample and
402 incubated at 4°C overnight with rotation. 5) After ChIP washes, samples were resuspended in 20 µL removal of
403 biotin-dATP master mix: 5 µL 10X NEB Buffer 2.1, 0.125 µL of dATP and dGTP at 10 mM each, 5 µL T4 DNA
404 polymerase (NEB, M0203), and 39.8 µL distilled water. Samples were incubated at 20°C for 4 hours with
405 interval shaking (1 min at 400 rpm, 4 min off). Lastly, size selection of libraries was performed using a two-
406 sided selection with AMPureXP beads. Briefly, 22.5 µL of beads were added to the sample to capture

407 fragments less than 700 bp. Supernatant was transferred to a new tube and 13.5 µL of beads were added to
408 the supernatant sample to capture fragments greater than 300 bp. Libraries were eluted in 50 µL 10 mM Tris
409 pH 8.0. Libraries were quantified using qPCR against Illumina primers and then paired end sequenced with
410 read length of 75 bp to a sequencing depth of 200 million reads per sample over two Illumina NextSeq runs.
411 *Dovetail HiChIP MNase Kit*
412 The Dovetail HiChIP MNase Kit was used for HiChIP for validation experiments repeating B6, D2, and
413 including B6-13D2 and D2-13B6. HiChIP was performed using two samples from independent cultures from
414 the same cell line per strain. The User Guide [95] (version 2.0) was used to process samples of 10 million
415 cells. H3K27ac (Cell Signaling Technology, 8173, lot #9) was used for immunoprecipitation. Libraries were
416 paired end sequenced with a read length of 150 bp to a sequencing depth of 300 million reads per sample
417 (over two flow cell lanes) on an Illumina NovaSeq X Plus.
418 **ATAC-seq and ChIP-seq data processing**
419 Illumina adaptors were trimmed from ChIP reads using Trimmomatic [96] (version 0.39) and then aligned to
420 mouse reference genome (mm10), modified to incorporate known D2 variants reported in Mouse Genomes
421 Project Database, REL-1505 [26,97] using hisat2 [98] (version 2.2.1). Duplicate reads were removed using
422 Picard Tools [99] (version 2.0.1), and peaks were called for each sample using MACS [100] (version 1.4.2).
423 While independent peak calling was performed on H3K27ac and H3K9me3 to determine quality and
424 reproducibility, to compare ATAC- and ChIP-seq signal at QTL targets the comprehensive set of open
425 chromatin locations (i.e., peakome) previously developed [28] was used to generate read counts across all
426 samples using bedtools [101] (version 2.9.2) similar to ATAC-seq. Read matrices were TMM-normalized and
427 $\log_2$-transformed downstream differential analysis of histone modification using edgeR [102] (version 3.34.1).
428 For visualization on genome browser tracks with karyoploteR [103] (version 1.20.3) we used bwa [104]
429 (version 0.7.5; filter mapping quality < 30, unmapped, secondary, QC fail, duplicates, and supplementary
430 reads) alignment to generate bigwig files.
431 **HiChIP data processing and analysis**
432 *Restriction enzyme-based HiChIP*
433 HiChIP reads were processed using HiC-Pro [105] (v2.11.4) with the following parameters:
434 N_CPU = 4, SORT_RAM = 768M

435 REFERENCE_GENOME = mm10

436 LIGATION_SITE = GATCGATC

437 MIN_FRAG_SIZE = 30

438 MAX_FRAG_SIZE =100000

439 MIN_INSERT_SIZE = 30

440 MAX_INSERT_SIZE = 600.

441 Long-range interactions were called with hichipper [106] (v.0.7.8b0) using H3K27ac ChIP-seq peaks generated
442 in this paper. PET counts were imported into RStudio (version 4.1) using diffloop [107] (v1.22.0). Before
443 filtering data, read matrices were TMM-normalized. Anchors of interactions were not merged and interactions
444 less than 5,000 bp were filtered out. P-values were adjusted using an FDR cutoff of 0.01. Interactions were
445 filtered for at least four read counts in at least two samples before identifying differential interactions using
446 DESeq2 [108] within the diffloop [107] package. Significant differential interactions were cutoff at FDR $\leq$ 0.2.
447 Differential interactions (DIs) showed higher average PET counts than non-differential interactions (Wilcoxon
448 rank sum test $p < 2.2e-16$; Additional file 1: Fig. S10), consistent with increased statistical power to detect
449 differences at well-covered loci.

450 *Dovetail HiChIP MNase Kit*

451 HiChIP reads were processed using Dovetail's HiChIP Loop Calling [109] (readthedocs Revision 8f2e0b16).
452 Reads were aligned to mm10 using bwa [104] (v0.7.18). pairtools [110] (version 1.1.2) was used to record and
453 sort valid ligation events, remove PCR duplicates, and generate .pairs and bam files. Samtools [111] (version
454 1.19.2) was used to create and index the final bam file. Libraries were verified to have no-dup read pairs $\geq$
455 50%, about 20% or greater of the total mapped non-dup pairs to be in *cis* and > 1 kb in length, and total reads
456 in 1 kb around center of ChIP peaks > 2%. Filtered pairs were converted to Hi-C Pro valid pairs format. Loops
457 were called using FitHiChIP [112] (version 11.0) with the following parameters in the bias correction/coverage
458 bias configuration file:

459 Interaction type = peak to peak

460 Bin size = 2500

461 Lower distance threshold = 10000

462 Upper distance threshold = 2000000

463 Bias types = coverages bias regression

464 Merge filtering was employed (option 1)

465 q-value = 0.01

466 Chromosome sizes correspond to mm10.

467 Peak-to-peak interactions were imported into RStudio (version 4.1). diffloop [107] (version 1.22.0) was used to
468 adjust p-values using an FDR cutoff of 0.01. Interactions were filtered for at least four read counts in at least
469 two samples. PET counts were normalized, and edgeR [102] (version 3.34.1) was used to identify differential
470 interactions (FDR < 0.2).

471 **Spatial Enrichment Analysis**

472 Region Associated Differentially expressed gene (RAD) framework [68] was adapted to test whether chromatin
473 accessibility changes and gene expression changes regulated by *trans*-QTL spatially co-localize. The genomic
474 background was defined as genes expressed in ESCs consistent with cutoff used for QTL mapping [28], rather
475 than all annotated genes genome-wide. Strand-specific signed distances between each gene's TSS to nearby
476 chromatin accessibility peaks were calculated, with negative values indicating peak positions upstream of the
477 TSS in the direction of transcription and positive values indicating downstream positions. Distances were
478 binned into 100 kb intervals spanning -500 kb to +500 kb relative to the TSS. Binomial tests were used to
479 determine significance treating each gene-peak connection within 500 kb as an independent trial (binom.test in
480 R, alternative = 'greater'), with the expected probability defined by the genome-wide proportion of focal genes
481 (p = M_focal / N_total), and Benjamini-Hochberg correction was applied for multiple testing within each
482 analysis.

483 **3C-qPCR**

484 3C-qPCR experiments were performed as previously described [65] with the following modifications. Briefly, 1)
485 10 million cells were fixed with fresh formaldehyde at a final concentration of 1% for 10 minutes with rotation
486 before quenching the reaction with 10X glycine for 5 minutes with rotation. Cells were pelleted, washed with
487 PBS, and flash frozen; 2) 800 U restriction enzyme (NEB, #R0101) and 1 mM ATP were used for digestion; 3)
488 4,000 U ligase (NEB, #M0202) was used for an overnight incubation at 16°C; 4) Qiagen MaXtract HD tubes
489 were used for 3C sample purification. BACs overlapping loci of interest (Additional file 4, Table S2) were used
490 for PCR control templates. Primers used for 3C-qPCR experiments are listed in Additional file 4.

491 **Declarations**

492 **Ethics approval and consent to participate**

493 All animal procedures were approved by the Animal Care and Use Committee of The Jackson Laboratory
494 under protocol number 04008.

495 **Availability of data and materials**

496 Original data discussed in this study in NCBI's Gene Expression Omnibus and are accessible through GEO
497 Series accession number GSE312918 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE312918).
498 Along with raw sequencing data, processed data tables in the accession include raw read counts for ATAC-
499 seq, H3K27ac ChIP-seq, H3K9me3 ChIP-seq, and HiChIP. Additionally, there are bigwig files for each ATAC,
500 ChIP, and HiChIP sample. H3K4me3 ChIP-seq [27], ATAC-seq, and RNA-seq [28] for B6 and D2 were
501 collected previously and are available through GEO accessions (H3K4me3: GSE113192,
502 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE113192; ATAC-seq and RNA-seq: GSE164935,
503 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE164935).

504 **Funding**

505 This work was supported by the National Institute of General Medical Sciences (NIGMS) grant R35GM133724.

506 **Acknowledgements**

507 We would like to thank all the members of the Baker laboratory for comments and discussion. Additionally, we
508 would like to thank Steve Munger for comments on the manuscript and Natalie Powers and Petko Petkov for
509 their early breeding of the congenic lines. We recognize contributions from The Jackson Laboratory Genome
510 Technologies services. The Jackson Laboratory scientific services are supported in part through the National
511 Cancer Institute's Cancer Core Grant P30CA034196.

References

1. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. Science. American Association for the Advancement of Science; 2012;337:1190–5. https://doi.org/10.1126/science.1222794

2. Tak YG, Farnham PJ. Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. Epigenetics & Chromatin. 2015;8:57. https://doi.org/10.1186/s13072-015-0050-4

3. Ward LD, Kellis M. Evidence of abundant purifying selection in humans for recently-acquired regulatory functions. Science. 2012;337:1675–8. https://doi.org/10.1126/science.1225057

4. Guan Y, Liu H, Ma Z, Li S-Y, Park J, Sheng X, et al. Dnmt3a and Dnmt3b-Decommissioned Fetal Enhancers are Linked to Kidney Disease. J Am Soc Nephrol. 2020;31:765–82. https://doi.org/10.1681/ASN.2019080797

5. Ma Y, Buttitta L. Chromatin organization changes during the establishment and maintenance of the postmitotic state | Epigenetics & Chromatin | Full Text. Epigenetics & Chromatin [Internet]. 2017 [cited 2021 June 22];10. https://doi.org/10.1186/s13072-017-0159-8

6. Fagnocchi L, Cherubini A, Hatsuda H, Fasciani A, Mazzoleni S, Poli V, et al. A Myc-driven self-reinforcing regulatory network maintains mouse embryonic stem cell identity. Nat Commun. 2016;7:11903. https://doi.org/10.1038/ncomms11903

7. Atlasi Y, Stunnenberg HG. The interplay of epigenetic marks during stem cell differentiation and development | Nature Reviews Genetics. Nature Reviews Genetics. 2017;18:643–58. https://doi.org/10.1038/nrg.2017.57

8. Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, et al. Chromatin architecture reorganization during stem cell differentiation. Nature. Nature Publishing Group; 2015;518:331–6. https://doi.org/10.1038/nature14222

9. Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, et al. Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. Cell. Elsevier; 2016;167:1369-1384.e19. https://doi.org/10.1016/j.cell.2016.09.037

10. Zhang F, Lupski JR. Non-coding genetic variants in human disease. Hum Mol Genet. 2015;24:R102–10. https://doi.org/10.1093/hmg/ddv259

11. Rockman MV, Kruglyak L. Genetics of global gene expression. Nat Rev Genet. Nature Publishing Group; 2006;7:862–72. https://doi.org/10.1038/nrg1964

12. Liu X, Mefford JA, Dahl A, He Y, Subramaniam M, Battle A, et al. GBAT: a gene-based association test for robust detection of trans-gene regulation. Genome Biology. 2020;21:211. https://doi.org/10.1186/s13059-020-02120-1

13. Mostafavi H, Spence JP, Naqvi S, Pritchard JK. Systematic differences in discovery of genetic effects on gene expression and complex traits. Nat Genet. Nature Publishing Group; 2023;55:1866–75. https://doi.org/10.1038/s41588-023-01529-1

14. Tyler AL, Mahoney JM, Keller MP, Baker CN, Gaca M, Srivastava A, et al. Transcripts with high distal heritability mediate genetic effects on complex metabolic traits. Nat Commun. Nature Publishing Group; 2025;16:5507. https://doi.org/10.1038/s41467-025-61228-9

15. Westra H-J, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet. Nature Publishing Group; 2013;45:1238–43. https://doi.org/10.1038/ng.2756

16. Wheeler HE, Ploch S, Barbeira AN, Bonazzola R, Andaleon A, Fotuhi Siahpirani A, et al. Imputed gene associations identify replicable trans-acting genes enriched in transcription pathways and complex traits. Genet Epidemiol. 2019;43:596–608. https://doi.org/10.1002/gepi.22205

17. Yao DW, O'Connor LJ, Price AL, Gusev A. Quantifying genetic effects on disease mediated by assayed gene expression levels. Nat Genet. Nature Publishing Group; 2020;52:626–33. https://doi.org/10.1038/s41588-020-0625-2

18. Albert FW, Bloom JS, Siegel J, Day L, Kruglyak L. Genetics of trans-regulatory variation in gene expression. Wittkopp PJ, editor. eLife. eLife Sciences Publications, Ltd; 2018;7:e35471. https://doi.org/10.7554/eLife.35471

19. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet. 2017;101:5–22. https://doi.org/10.1016/j.ajhg.2017.06.005

20. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet. Nature Publishing Group; 2010;11:415–25. https://doi.org/10.1038/nrg2779

21. Zuk O, Schaffner SF, Samocha K, Do R, Hechter E, Kathiresan S, et al. Searching for missing heritability: Designing rare variant association studies. Proceedings of the National Academy of Sciences. Proceedings of the National Academy of Sciences; 2014;111:E455–64. https://doi.org/10.1073/pnas.1322563111

22. Churchill GA, Gatti DM, Munger SC, Svenson KL. The Diversity Outbred Mouse Population. Mamm Genome. 2012;23:713–8. https://doi.org/10.1007/s00335-012-9414-2

23. Svenson KL, Gatti DM, Valdar W, Welsh CE, Cheng R, Chesler EJ, et al. High-Resolution Genetic Mapping Using the Mouse Diversity Outbred Population. Genetics. 2012;190:437–47. https://doi.org/10.1534/genetics.111.132597

24. Williams RW, Gu J, Qi S, Lu L. The genetic structure of recombinant inbred mice: High-resolution consensus maps for complex trait analysis. Genome Biol. 2001;2:preprint0007.1. https://doi.org/10.1186/gb-2001-2-8-preprint0007

25. Taylor BA, Wnek C, Kotlus BS, Roemer N, MacTaggart T, Phillips SJ. Genotyping new BXD recombinant inbred mouse strains and comparison of BXD and consensus maps. Mammalian Genome. 1999;10:335–48. https://doi.org/10.1007/s003359900998

26. Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, et al. Mouse genomic variation and its effect on phenotypes and gene regulation. Nature. 2011;477:289–94. https://doi.org/10.1038/nature10413

27. Baker CL, Walker M, Arat S, Ananda G, Petkova P, Powers NR, et al. Tissue-Specific Trans Regulation of the Mouse Epigenome. Genetics. Genetics; 2019;211:831–45. https://doi.org/10.1534/genetics.118.301697

28. Byers C, Spruce C, Fortin HJ, Hartig EI, Czechanski A, Munger SC, et al. Genetic control of the pluripotency epigenome determines differentiation bias in mouse embryonic stem cells. The EMBO Journal. John Wiley & Sons, Ltd; 2021;40:e109445. https://doi.org/10.15252/embj.2021109445

29. Johnson KR, Lane PW, Ward-Bailey P, Davisson MT. Mapping the Mouse Dactylaplasia Mutation,Dac,and a Gene That Controls Its Expression,mdac. Genomics. 1995;29:457–64. https://doi.org/10.1006/geno.1995.9981

592 30. Kano H, Kurahashi H, Toda T. Genetically regulated epigenetic transcriptional activation of retrotransposon
593 insertion confers mouse dactylaplasia phenotype. PNAS. National Academy of Sciences; 2007;104:19034–9.
594 https://doi.org/10.1073/pnas.0705483104

595 31. Plamondon JA, Harris MJ, Mager DL, Gagnier L, Juriloff DM. The clf2 gene has an epigenetic role in the
596 multifactorial etiology of cleft lip and palate in the A/WySn mouse strain. Birth Defects Research Part A: Clinical
597 and Molecular Teratology. 2011;91:716–27. https://doi.org/10.1002/bdra.20788

598 32. Swanzey E, McNamara TF, Apostolou E, Tahiliani M, Stadtfeld M. A Susceptibility Locus on Chromosome
599 13 Profoundly Impacts the Stability of Genomic Imprinting in Mouse Pluripotent Stem Cells. Cell Reports.
600 Elsevier; 2020;30:3597-3604.e3. https://doi.org/10.1016/j.celrep.2020.02.073

601 33. Treger RS, Pope SD, Kong Y, Tokuyama M, Taura M, Iwasaki A. The Lupus Susceptibility Locus Sgp3
602 Encodes the Suppressor of Endogenous Retrovirus Expression SNERV. Immunity. 2019;50:334-347.e9.
603 https://doi.org/10.1016/j.immuni.2018.12.022

604 34. Skelly DA, Czechanski A, Byers C, Aydin S, Spruce C, Olivier C, et al. Mapping the Effects of Genetic
605 Variation on Chromatin State and Gene Expression Reveals Loci That Control Ground State Pluripotency. Cell
606 Stem Cell. Elsevier; 2020;27:459-469.e8. https://doi.org/10.1016/j.stem.2020.07.005

607 35. Cheung VG, Spielman RS. Genetics of human gene expression: mapping DNA variants that influence
608 gene expression. Nat Rev Genet. 2009;10:595–604. https://doi.org/10.1038/nrg2630

609 36. Mattioli K, Oliveros W, Gerhardinger C, Andergassen D, Maass PG, Rinn JL, et al. Cis and trans effects
610 differentially contribute to the evolution of promoters and enhancers. Genome Biology. 2020;21:210.
611 https://doi.org/10.1186/s13059-020-02110-3

612 37. Fullwood MJ, Wei C-L, Liu ET, Ruan Y. Next-generation DNA sequencing of paired-end tags (PET) for
613 transcriptome and genome analyses. Genome Res. 2009;19:521–32. https://doi.org/10.1101/gr.074906.107

614 38. Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, et al. Mapping long-range
615 promoter contacts in human cells with high-resolution capture Hi-C. Nat Genet. Nature Publishing Group;
616 2015;47:598–606. https://doi.org/10.1038/ng.3286

617 39. Mumbach MR, Rubin AJ, Flynn RA, Dai C, Khavari PA, Greenleaf WJ, et al. HiChIP: efficient and sensitive
618 analysis of protein-directed genome architecture. Nat Methods. 2016;13:919–22.
619 https://doi.org/10.1038/nmeth.3999

620 40. Fang R, Yu M, Li G, Chee S, Liu T, Schmitt AD, et al. Mapping of long-range chromatin interactions by
621 proximity ligation-assisted ChIP-seq. Cell Res. Nature Publishing Group; 2016;26:1345–8.
622 https://doi.org/10.1038/cr.2016.137

623 41. Krijger PHL, Geeven G, Bianchi V, Hilvering CRE, de Laat W. 4C-seq from beginning to end: A detailed
624 protocol for sample preparation and data analysis. Methods. 2020;170:17–32.
625 https://doi.org/10.1016/j.ymeth.2019.07.014

626 42. van Berkum NL, Lieberman-Aiden E, Williams L, Imakaev M, Gnirke A, Mirny LA, et al. Hi-C: A Method to
627 Study the Three-dimensional Architecture of Genomes. J Vis Exp. 2010;1869. https://doi.org/10.3791/1869

628 43. Beagrie RA, Scialdone A, Schueler M, Kraemer DCA, Chotalia M, Xie SQ, et al. Complex multi-enhancer
629 contacts captured by genome architecture mapping. Nature. Nature Publishing Group; 2017;543:519–24.
630 https://doi.org/10.1038/nature21411

631 44. Dekker J, Mirny L. The 3D Genome as Moderator of Chromosomal Communication. Cell. 2016;164:1110–
632 21. https://doi.org/10.1016/j.cell.2016.02.007

633 45. Bortle KV, Corces VG. Nuclear Organization and Genome Function. Annu Rev Cell Dev Biol. 2012;28:163–
634 87. https://doi.org/10.1146/annurev-cellbio-101011-155824

635 46. Bonev B, Cavalli G. Organization and function of the 3D genome. Nat Rev Genet. 2016;17:661–78.
636 https://doi.org/10.1038/nrg.2016.112

637 47. Gorkin DU, Leung D, Ren B. The 3D Genome in Transcriptional Regulation and Pluripotency. Cell Stem
638 Cell. 2014;14:762–75. https://doi.org/10.1016/j.stem.2014.05.017

639 48. Hug CB, Vaquerizas JM. The Birth of the 3D Genome during Early Embryonic Development. Trends in
640 Genetics. Elsevier; 2018;34:903–14. https://doi.org/10.1016/j.tig.2018.09.002

641 49. Carter D, Chakalova L, Osborne CS, Dai Y-F, Fraser P. Long-range chromatin regulatory interactions in
642 vivo. Nature Genetics. 2002;32:623–6. https://doi.org/10.1038/ng1051

643 50. Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, et al. Extensive Promoter-Centered
644 Chromatin Interactions Provide a Topological Basis for Transcription Regulation. Cell. 2012;148:84–98.
645 https://doi.org/10.1016/j.cell.2011.12.014

646 51. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. Nature.
647 Nature Publishing Group; 2012;489:109–13. https://doi.org/10.1038/nature11279

648 52. Tolhuis B, Palstra R-J, Splinter E, Grosveld F, de Laat W. Looping and Interaction between Hypersensitive
649 Sites in the Active β-globin Locus. Molecular Cell. 2002;10:1453–65. https://doi.org/10.1016/S1097-
650 2765(02)00781-5

651 53. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes
652 identified by analysis of chromatin interactions. Nature. Nature Publishing Group; 2012;485:376–80.
653 https://doi.org/10.1038/nature11082

654 54. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, et al. Spatial partitioning of the
655 regulatory landscape of the X-inactivation centre. Nature. 2012;485:381–5.
656 https://doi.org/10.1038/nature11049

657 55. Smith EM, Lajoie BR, Jain G, Dekker J. Invariant TAD Boundaries Constrain Cell-Type-Specific Looping
658 Interactions between Promoters and Distal Elements around the *CFTR* Locus. The American Journal of
659 Human Genetics. 2016;98:185–201. https://doi.org/10.1016/j.ajhg.2015.12.002

660 56. Martin P, McGovern A, Orozco G, Duffus K, Yarwood A, Schoenfelder S, et al. Capture Hi-C reveals novel
661 candidate genes and complex long-range interactions with related autoimmune risk loci. Nat Commun. Nature
662 Publishing Group; 2015;6:10069. https://doi.org/10.1038/ncomms10069

663 57. Franke M, Ibrahim DM, Andrey G, Schwarzer W, Heinrich V, Schöpflin R, et al. Formation of new
664 chromatin domains determines pathogenicity of genomic duplications. Nature. 2016;538:265–9.
665 https://doi.org/10.1038/nature19800

666 58. Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, et al. Disruptions of Topological
667 Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. Cell. 2015;161:1012–25.
668 https://doi.org/10.1016/j.cell.2015.04.004

669 59. Cova G, Glaser J, Schöpflin R, Prada-Medina CA, Ali S, Franke M, et al. Combinatorial effects on gene
670 expression at the Lbx1/Fgf8 locus resolve split-hand/foot malformation type 3. Nat Commun. Nature Publishing
671 Group; 2023;14:1475. https://doi.org/10.1038/s41467-023-37057-z

672 60. Gorkin DU, Qiu Y, Hu M, Fletez-Brant K, Liu T, Schmitt AD, et al. Common DNA sequence variation
673 influences 3-dimensional conformation of the human genome. Genome Biology. 2019;20:255.
674 https://doi.org/10.1186/s13059-019-1855-4

675 61. Peng Y, Xiong D, Zhao L, Ouyang W, Wang S, Sun J, et al. Chromatin interaction maps reveal genetic
676 regulation for quantitative traits in maize. Nat Commun. 2019;10:2632. https://doi.org/10.1038/s41467-019-
677 10602-5

678 62. Zhao L, Wang S, Cao Z, Ouyang W, Zhang Q, Xie L, et al. Chromatin loops associated with active genes
679 and heterochromatin shape rice genome architecture for transcriptional regulation. Nat Commun.
680 2019;10:3640. https://doi.org/10.1038/s41467-019-11535-9

681 63. Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, et al. Lineage-Specific Genome
682 Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. Cell.
683 2016;167:1369-1384.e19. https://doi.org/10.1016/j.cell.2016.09.037

684 64. Schoenfelder S, Javierre B-M, Furlan-Magaril M, Wingett SW, Fraser P. Promoter Capture Hi-C: High-
685 resolution, Genome-wide Profiling of Promoter Interactions. JoVE (Journal of Visualized Experiments).
686 2018;e57320. https://doi.org/10.3791/57320

687 65. Hagège H, Klous P, Braem C, Splinter E, Dekker J, Cathala G, et al. Quantitative analysis of chromosome
688 conformation capture assays (3C-qPCR). Nat Protoc. 2007;2:1722–33. https://doi.org/10.1038/nprot.2007.243

689 66. Friedman JR, Fredericks WJ, Jensen DE, Speicher DW, Huang XP, Neilson EG, et al. KAP-1, a novel
690 corepressor for the highly conserved KRAB repression domain. Genes Dev. 1996;10:2067–78.
691 https://doi.org/10.1101/gad.10.16.2067

692 67. Moosmann P, Georgiev O, Thiesen H-J, Hagmann M, Schaffner W. Silencing of RNA Polymerases II and
693 III-Dependent Transcription by the KRAB Protein Domain of KOX1, a Krüppel-Type Zinc Finger Factor.
694 Biological Chemistry. De Gruyter; 1997;378:669–78. https://doi.org/10.1515/bchm.1997.378.7.669

695 68. Guo Y, Xue Z, Yuan R, Li JJ, Pastor WA, Liu W. RAD: a web application to identify region associated
696 differentially expressed genes. Bioinformatics. 2021;37:2741–3. https://doi.org/10.1093/bioinformatics/btab075

697 69. Ray-Jones H, Sung CK, Chan LT, Haglund A, Artemov P, Della Rosa M, et al. Genetic coupling of
698 enhancer activity and connectivity in gene expression control. Nat Commun. Nature Publishing Group;
699 2025;16:970. https://doi.org/10.1038/s41467-025-55900-3

700 70. Shi C, Zhao D, Butler J, Frantzeskos A, Rossi S, Ding J, et al. Multi-omics analysis in primary T cells
701 elucidates mechanisms behind disease-associated genetic loci. Genome Biology. 2025;26:26.
702 https://doi.org/10.1186/s13059-025-03492-y

703 71. Greenwald WW, Li H, Benaglio P, Jakubosky D, Matsui H, Schmitt A, et al. Subtle changes in chromatin
704 loop contact propensity are associated with differential gene regulation and expression. Nat Commun.
705 2019;10:1054. https://doi.org/10.1038/s41467-019-08940-5

706 72. Arthur TD, Nguyen JP, Henson BA, D'Antonio-Chronowska A, Jaureguy J, Silva N, et al. Multiomic QTL
707 mapping reveals phenotypic complexity of GWAS loci and prioritizes putative causal variants. Cell Genom.
708 2025;5:100775. https://doi.org/10.1016/j.xgen.2025.100775

709 73. Thiesen HJ. Multiple genes encoding zinc finger domains are expressed in human T cells. New Biol.
710 1990;2:363–74.

711 74. Bellefroid EJ, Poncelet DA, Lecocq PJ, Revelant O, Martial JA. The evolutionarily conserved Krüppel-
712 associated box domain defines a subfamily of eukaryotic multifingered proteins. Proceedings of the National

713 Academy of Sciences. Proceedings of the National Academy of Sciences; 1991;88:3608–12.
714 https://doi.org/10.1073/pnas.88.9.3608

715 75. Schultz DC, Ayyanathan K, Negorev D, Maul GG, Rauscher FJ. SETDB1: a novel KAP-1-associated
716 histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic
717 genes by KRAB zinc-finger proteins. Genes Dev. 2002;16:919–32. https://doi.org/10.1101/gad.973302

718 76. Lupo A, Cesaro E, Montano G, Zurlo D, Izzo P, Costanzo P. KRAB-Zinc Finger Proteins: A Repressor
719 Family Displaying Multiple Biological Functions. Curr Genomics. 2013;14:268–78.
720 https://doi.org/10.2174/1389202913149990002

721 77. Sripathy SP, Stevens J, Schultz DC. The KAP1 Corepressor Functions To Coordinate the Assembly of De
722 Novo HP1-Demarcated Microenvironments of Heterochromatin Required for KRAB Zinc Finger Protein-
723 Mediated Transcriptional Repression. Mol Cell Biol. 2006;26:8623–38. https://doi.org/10.1128/MCB.00487-06

724 78. Francis BA, Gozashti L, Costello K, Kasahara T, Harringmeyer OS, Lilue J, et al. Complete genome
725 assemblies of two mouse subspecies reveal structural diversity of telomeres and centromeres. Nat Genet.
726 Nature Publishing Group; 2025;57:2852–62. https://doi.org/10.1038/s41588-025-02367-z

727 79. Bruno M, Maisha S, Mitra A, Costello K, Watkins-Chow D, Logsdon GA, et al. Young KRAB-zinc finger
728 gene clusters are highly dynamic incubators of ERV-driven genetic heterogeneity in mice [Internet]. bioRxiv;
729 2025 [cited 2025 July 3]. p. 2025.02.26.640358. https://doi.org/10.1101/2025.02.26.640358

730 80. Groner AC, Meylan S, Ciuffi A, Zangger N, Ambrosini G, Dénervaud N, et al. KRAB–Zinc Finger Proteins
731 and KAP1 Can Mediate Long-Range Transcriptional Repression through Heterochromatin Spreading. PLOS
732 Genetics. Public Library of Science; 2010;6:e1000869. https://doi.org/10.1371/journal.pgen.1000869

733 81. Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS. CRISPR interference (CRISPRi) for
734 sequence-specific control of gene expression. Nat Protoc. 2013;8:2180–96.
735 https://doi.org/10.1038/nprot.2013.132

736 82. Bruno M, Maisha S, Mitra A, Costello K, Watkins-Chow D, Logsdon GA, et al. Young KRAB-zinc finger
737 gene clusters are highly dynamic incubators of ERV-driven genetic heterogeneity in mice. bioRxiv.
738 2025;2025.02.26.640358. https://doi.org/10.1101/2025.02.26.640358

739 83. Imbeault M, Helleboid P-Y, Trono D. KRAB zinc-finger proteins contribute to the evolution of gene
740 regulatory networks. Nature. Nature Publishing Group; 2017;543:550–4. https://doi.org/10.1038/nature21683

741 84. Hansen TJ, Fong SL, Day JK, Capra JA, Hodges E. Human gene regulatory evolution is driven by the
742 divergence of regulatory element function in both cis and trans. Cell Genom. 2024;4:100536.
743 https://doi.org/10.1016/j.xgen.2024.100536

744 85. Becker JS, Nicetto D, Zaret KS. H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes.
745 Trends Genet. 2016;32:29–41. https://doi.org/10.1016/j.tig.2015.11.001

746 86. Soufi A, Donahue G, Zaret KS. Facilitators and Impediments of the Pluripotency Reprogramming Factors'
747 Initial Engagement with the Genome. Cell. 2012;151:994–1004. https://doi.org/10.1016/j.cell.2012.09.045

748 87. Kaemena DF, Yoshihara M, Beniazza M, Ashmore J, Zhao S, Bertenstam M, et al. B1 SINE-binding
749 ZFP266 impedes mouse iPSC generation through suppression of chromatin opening mediated by
750 reprogramming factors. Nat Commun. 2023;14:488. https://doi.org/10.1038/s41467-023-36097-9

751 88. Kim S-K, Seo S, Stein-O'Brien G, Jaishankar A, Ogawa K, Micali N, et al. Individual variation in the
752 emergence of anterior-to-posterior neural fates from human pluripotent stem cells. Stem Cell Reports.
753 2024;19:1336–50. https://doi.org/10.1016/j.stemcr.2024.07.004

754 89. Perdomo-Sabogal Á, Nowick K. Genetic Variation in Human Gene Regulatory Factors Uncovers
755 Regulatory Roles in Local Adaptation and Disease. Genome Biol Evol. 2019;11:2178–93.
756 https://doi.org/10.1093/gbe/evz131

757 90. Sigmon JS, Blanchard MW, Baric RS, Bell TA, Brennan J, Brockmann GA, et al. Content and Performance
758 of the MiniMUGA Genotyping Array: A New Tool To Improve Rigor and Reproducibility in Mouse Research.
759 Genetics. 2020;216:905–30. https://doi.org/10.1534/genetics.120.303596

760 91. Czechanski A, Byers C, Greenstein I, Schrode N, Donahue LR, Hadjantonakis A-K, et al. Derivation and
761 characterization of mouse embryonic stem cells (mESCs) from permissive and non-permissive strains. Nat
762 Protoc. 2014;9:559–74. https://doi.org/10.1038/nprot.2014.030

763 92. Corces MR, Buenrostro JD, Wu B, Greenside PG, Chan SM, Koenig JL, et al. Lineage-specific and single-
764 cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat Genet. Nature Publishing
765 Group; 2016;48:1193–203. https://doi.org/10.1038/ng.3646

766 93. Billings T, Parvanov ED, Baker CL, Walker M, Paigen K, Petkov PM. DNA binding specificities of the long
767 zinc-finger recombination protein PRDM9. Genome Biology. 2013;14:R35. https://doi.org/10.1186/gb-2013-14-
768 4-r35

769 94. Belaghzal H, Dekker J, Gibcus JH. Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide
770 mapping of chromosome conformation. Methods. 2017;123:56–65. https://doi.org/10.1016/j.ymeth.2017.04.004

771 95. Dovetail-HiChip-MNase-Kit-Version-2.0.pdf [Internet]. [cited 2025 July 7]. https://cantatabio.com/wp-
772 content/uploads/2023/12/Dovetail-HiChip-MNase-Kit-Version-2.0.pdf. Accessed 7 July 2025

773 96. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics.
774 2014;30:2114–20. https://doi.org/10.1093/bioinformatics/btu170

775 97. Yalcin B, Wong K, Agam A, Goodson M, Keane TM, Gan X, et al. Sequence based characterization of
776 structural variation in the mouse genome. Nature. 2011;477:326–9. https://doi.org/10.1038/nature10432

777 98. Pertea M, Kim D, Pertea G, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq
778 experiments with HISAT, StringTie, and Ballgown. Nat Protoc. 2016;11:1650–67.
779 https://doi.org/10.1038/nprot.2016.095

780 99. Picard Tools - By Broad Institute [Internet]. [cited 2025 July 9]. https://broadinstitute.github.io/picard/.
781 Accessed 9 July 2025

782 100. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based Analysis of ChIP-
783 Seq (MACS). Genome Biol. 2008;9:R137. https://doi.org/10.1186/gb-2008-9-9-r137

784 101. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics.
785 2010;26:841–2. https://doi.org/10.1093/bioinformatics/btq033

786 102. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression
787 analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.
788 https://doi.org/10.1093/bioinformatics/btp616

789 103. Gel B, Serra E. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying
790 arbitrary data. Bioinformatics. 2017;33:3088–90. https://doi.org/10.1093/bioinformatics/btx346

791 104. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics.
792 2009;25:1754–60. https://doi.org/10.1093/bioinformatics/btp324

793 105. Servant N, Varoquaux N, Lajoie BR, Viara E, Chen C-J, Vert J-P, et al. HiC-Pro: an optimized and flexible
794 pipeline for Hi-C data processing. Genome Biology. 2015;16:259. https://doi.org/10.1186/s13059-015-0831-x

795 106. Lareau CA, Aryee MJ. hichipper: a preprocessing pipeline for calling DNA loops from HiChIP data. Nat
796 Methods. 2018;15:155–6. https://doi.org/10.1038/nmeth.4583

797 107. Lareau CA, Aryee MJ. diffloop: a computational framework for identifying and analyzing differential DNA
798 loops from sequencing data. Bioinformatics. 2018;34:672–4. https://doi.org/10.1093/bioinformatics/btx623

799 108. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with
800 DESeq2. Genome Biol [Internet]. Springer Science and Business Media LLC; 2014 [cited 2025 July 7];15.
801 https://doi.org/10.1186/s13059-014-0550-8

802 109. HiChIP 0.1 documentation [Internet]. [cited 2025 July 7].
803 https://hichip.readthedocs.io/en/latest/before_you_begin.html. Accessed 7 July 2025

804 110. Open2C, Abdennur N, Fudenberg G, Flyamer IM, Galitsyna AA, Goloborodko A, et al. Pairtools: From
805 sequencing data to chromosome contacts. Ay F, editor. PLoS Comput Biol. Public Library of Science (PLoS);
806 2024;20:e1012164. https://doi.org/10.1371/journal.pcbi.1012164

807 111. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format
808 and SAMtools. Bioinformatics. 2009;25:2078–9. https://doi.org/10.1093/bioinformatics/btp352

809 112. Bhattacharyya S, Chandra V, Vijayanand P, Ay F. Identification of significant chromatin contacts from
810 HiChIP data by FitHiChIP. Nat Commun. Nature Publishing Group; 2019;10:4221.
811 https://doi.org/10.1038/s41467-019-11950-y

Figures

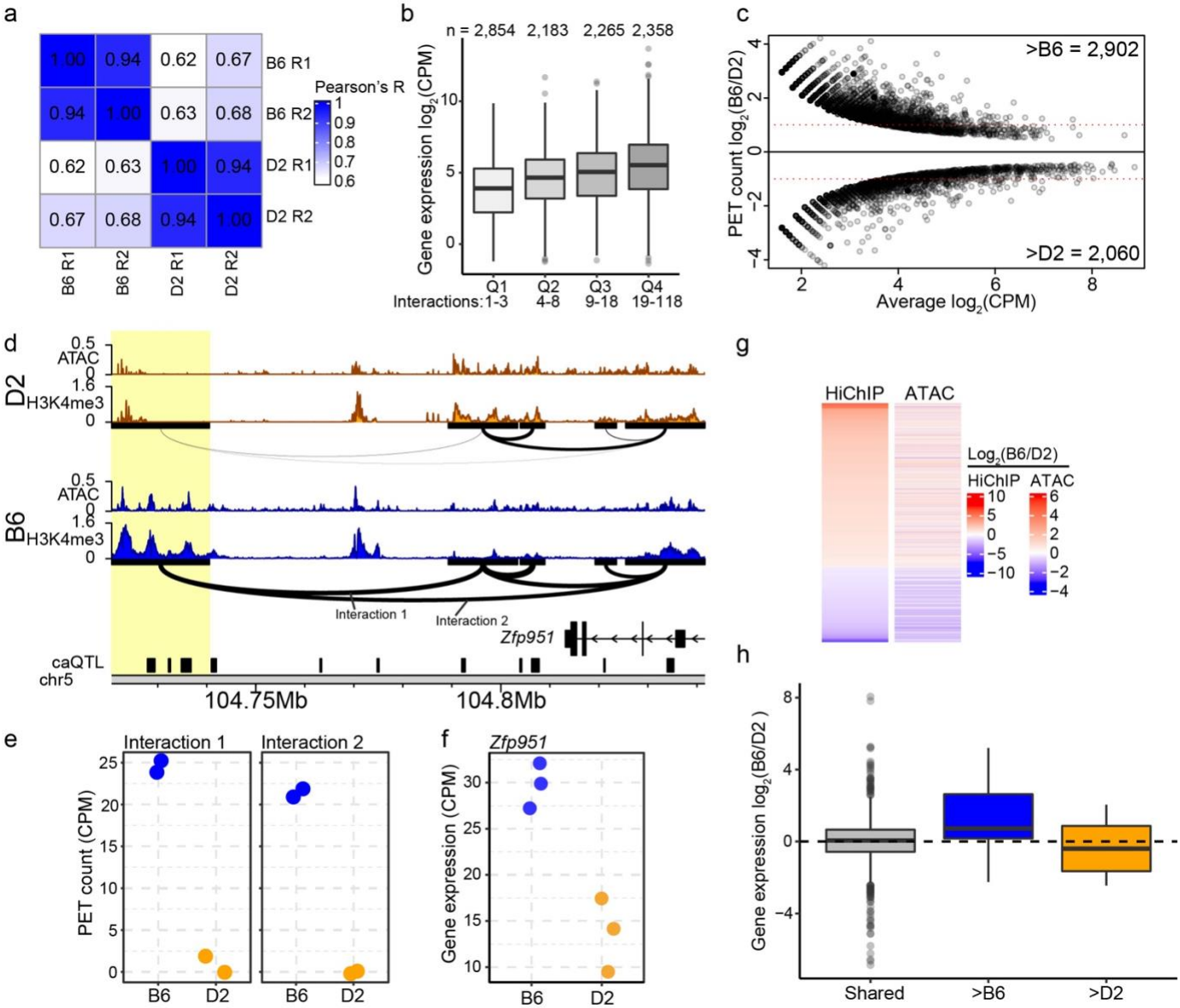


Fig. 1. Genetic variation coordinates chromatin interactions, accessibility, and gene expression. a Pearson correlation of H3K27ac HiChIP from biological replicates (R1 and R2) of mouse embryonic stem cells (ESCs) derived from B6 and D2 strains. **b** Box and whisker plots indicating mean gene expression values from RNA-seq for B6 and D2 mESCs (for all figure boxes: first quartile, median, and third quartile, whiskers: minimum and maximum, outliers are indicated as dots). Genes were stratified into quartiles (Q1-4) based on the number of contacts per promoter (interactions below graph, gene number above). One-way ANOVA: $F_{(3, 9656)} = 232$, $p < 2e\text{-}16$; Tukey HSD post-hoc test: all pairwise comparisons between quartiles were significant. **c** MA-plot for differential interactions B6 and D2 ESCs (FDR < 0.2). **d** Genome browser profiles of H3K4me3 ChIP-seq and ATAC-seq at the *Zfp951* locus for B6 (blue) and D2 (orange). DIs are drawn below H3K4me3 tracks. The thickness of the arc connecting two anchors represent contact frequency. Black boxes represent chromatin accessibility regions mapped as quantitative trait loci (caQTL) [28]. Yellow highlight – contact anchor overlapping multiple QTL targets. **e**. HiChIP PET counts for B6 and D2 from interactions 1 (p = 7.44e-9) and 2 (p = 4.48e-10) at *Zfp951* locus. **f** Expression of *Zfp951* from B6 and D2 ESCs (p = 2.81e-5). **g** Heatmap indicating fold-change (B6/D2) in HiChIP PET counts for DIs (left) and corresponding fold-change for ATAC-seq signal at peaks that overlap contact anchors (right). **h** Box and whisker plots of fold-change (B6/D2) in RNA-seq signal at genes which overlap contact anchors for interactions without significant differences (shared, n = 943), and either greater in B6 (n = 17) or D2 (n = 6).

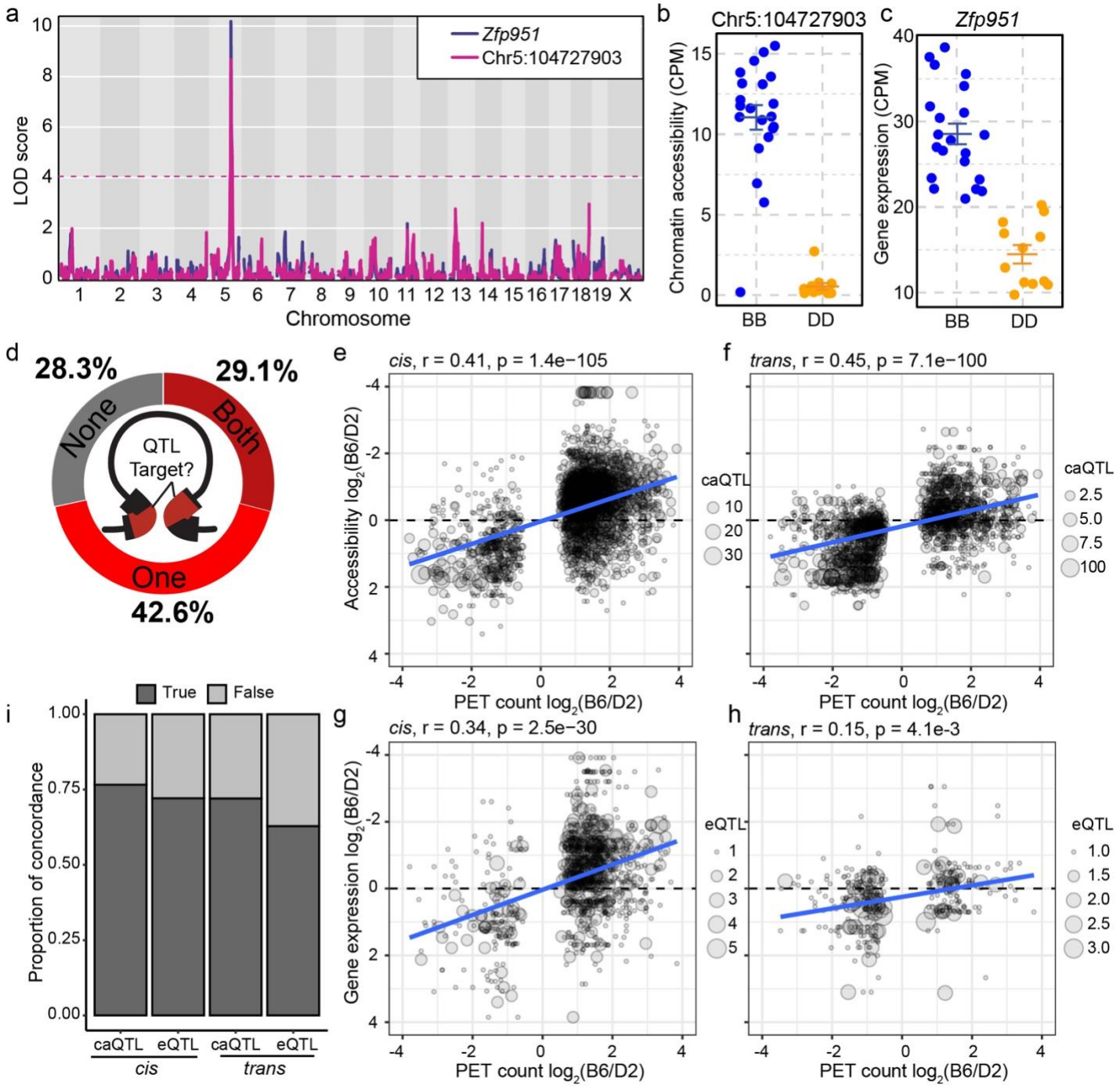


Fig. 2. Differential interactions overlap QTL targets and show allelic concordance across regulatory layers. **a** LOD score plot for *Zfp951* and a regulatory element QTL (Fig 1d, leftmost caQTL target in highlighted anchor). Both chromatin accessibility and expression of *Zfp951* are regulated locally on Chr 5. **b** Chromatin accessibility at *Zfp951* regulatory element (Fig 2a pink line) in 33 BXD ESC lines based on genotype at the QTL locus (line represents mean ± standard error). **c** Gene expression of *Zfp951* in 33 BXD ESC lines. **d** Donut plot indicating proportions of DIs for which contact anchors overlap caQTL targets at one, both, or no anchors. Correlation between PET count and chromatin accessibility (**e**, **f**) or gene expression (**g**, **h**), separated into *cis*- (left) and *trans*-associations (right). **i** Proportion of allele concordance between interaction strength and chromatin accessibility (caQTL) or gene expression (eQTL). *cis* caQTL: χ^2 = 529.5, df = 2, p < 2.2e-16; *trans* caQTL: χ^2 = 440.2, df = 2 p < 2.2e-16; *cis* eQTL: χ^2 = 148.2, df = 1 p < 2.2e-16; *trans* eQTL χ^2 = 20.3, df = 1 = 6.7e-6.

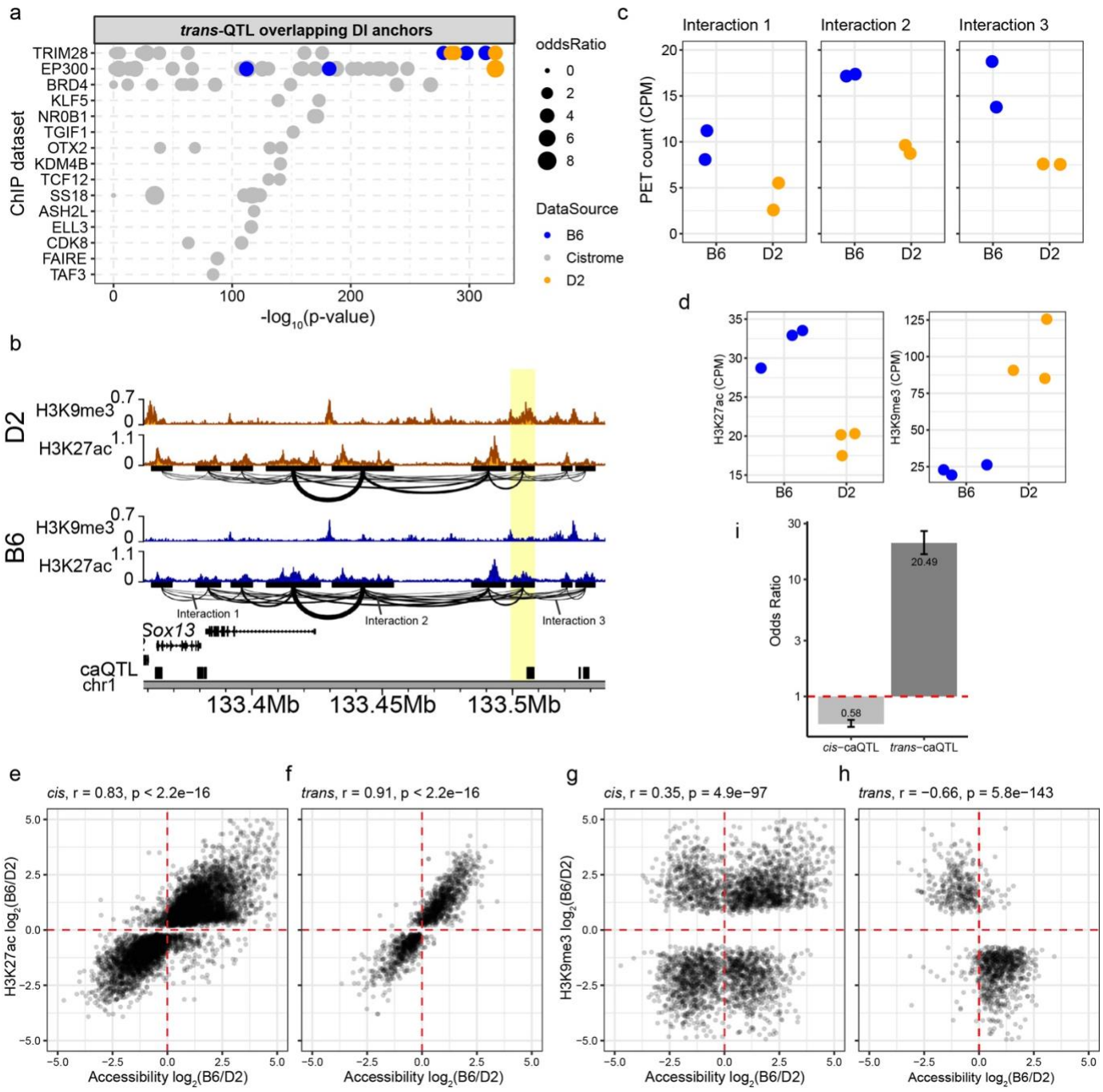


Fig. 3. Heterochromatin formation distinguishes *trans* from *cis* regulatory mechanisms. a Enrichment of the top 15 DNA-binding proteins at *trans*-QTL targets overlapping DI anchors. Each dot represents a ChIP-seq experiment either from publicly available database (grey - cistrome), or our previously published B6 (blue) and D2 (orange) datasets. Binding sites were evaluated for overlap between *trans*-targets and DI contact anchors (Fisher's exact test, top 15 factors with q-value cutoff < 0.01). **b** Genome profiles for H3K27ac and H3K9me3 ChIP enrichment from D2 (orange) and B6 (blue) at *Sox13* locus. HiChIP interactions indicated below H3K27ac tracks. Highlight indicates a contact anchor that overlaps a *trans*-target (black box). **c** HiChIP PET counts for three significant DIs indicated in panel *b* (Interaction 1 p = 9.53e-3, Interaction 2 p = 1.05e-2, Interaction 3 p = 3.04e-3). **d** Difference in H3K27ac (left) and H3K9me3 (right) at caQTL target highlighted in panel *b* between B6 and D2 (H3K27ac p = 2.24e-3, H3K9me3 p = 3.58e-3). Correlation between fold change (B6/D2) in H3K27ac enrichment and chromatin accessibility at *cis*- and *trans*-QTL targets (**e, f**). Correlation between fold change in H3K9me3 enrichment and chromatin accessibility at *cis*- and *trans*-QTL targets (**g, h**). **i** Odds ratio showing that *trans*-targets are 20.49 times more likely to show discordant chromatin states resulting in high H3K9me3 and low ATAC.

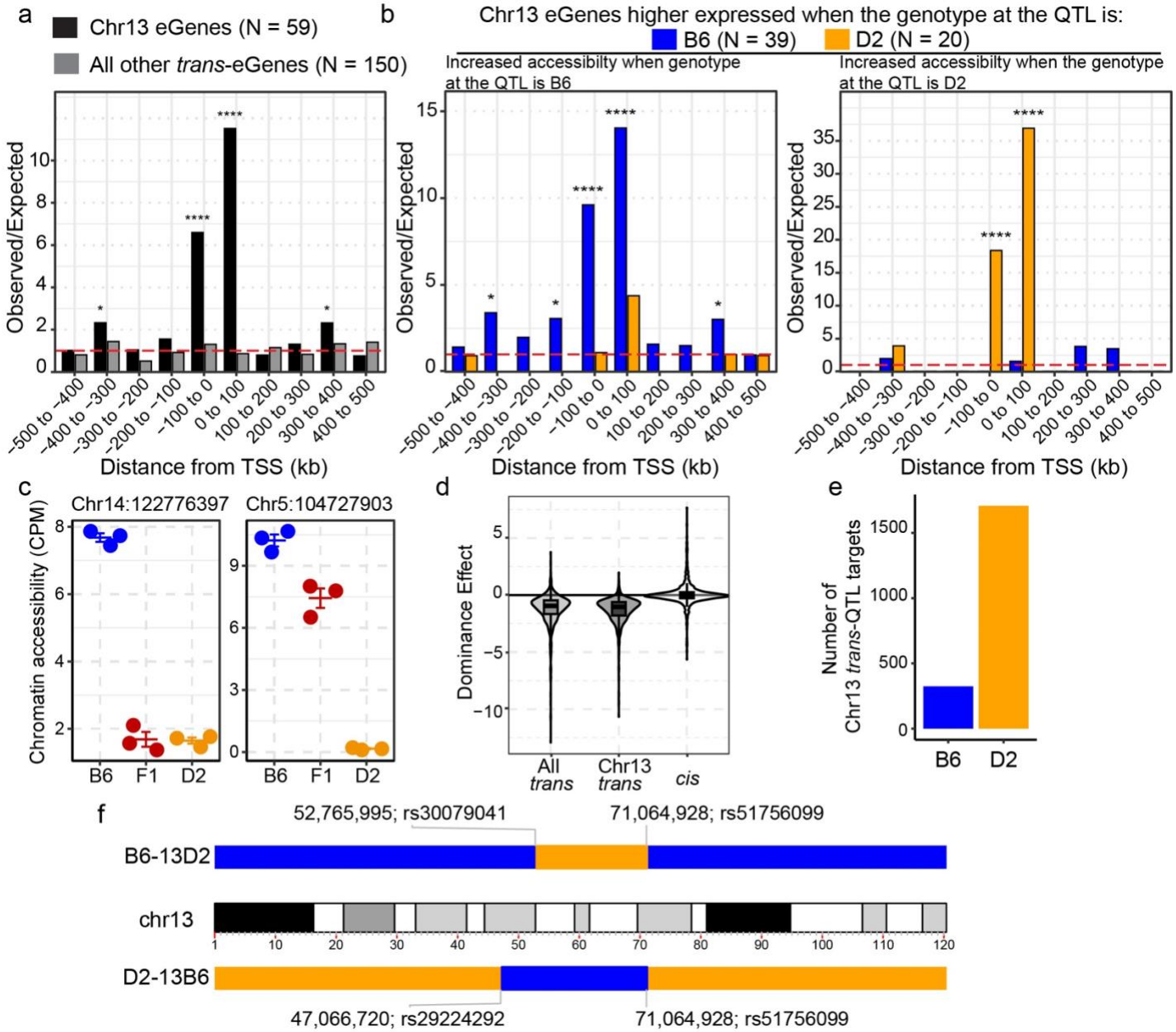


Fig. 4. Chr13 *trans*-QTL coordinates spatial regulation of chromatin accessibility and gene expression through dominant repression. **a** Observed-over-expected ratio for the spatial proximity of Chr13 *trans*-targets to transcription start sites (TSS) of either Chr13 eGenes (black) or all other *trans*-eGenes (grey). Red dashed line is null expectation (binomial test, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$). **b** Similar to *a* except stratifying genes that either higher expressed when Chr13 QTL haplotype is B6 (blue) or D2 (orange). Spatial proximity was tested for peaks with higher chromatin accessibility when Chr13 is B6 (left) or D2 (right). **c** Scatterplot of chromatin accessibility based on B6, D2, or F1-hybrid genotype of ESCs at two caQTL. Left – representative Chr13 *trans*-target indicating dominant repressive effect on chromatin accessibility. Right - *cis*-QTL target representative of additive effect (line – mean, bars – standard error). **d** Distribution of dominance effect values in all *trans*-targets (N = 836), Chr13 *trans*-targets (N = 575), and all *cis*-caQTL (N = 7049). Negative values indicate repressive effect; values near zero indicate additive effect. **e** Classification of Chr13 *trans*-targets by repressive allele. To establish directional predictions targets were stratified by which QTL genotype reduces chromatin accessibility, consistent with dominant repression. **f** Schematic of chromosome 13 showing the base-pair position and identifying SNP for congenic regions in B6-13D2 (D2 region on B6 background) and D2-13B6 (B6 region on D2 background) strains.

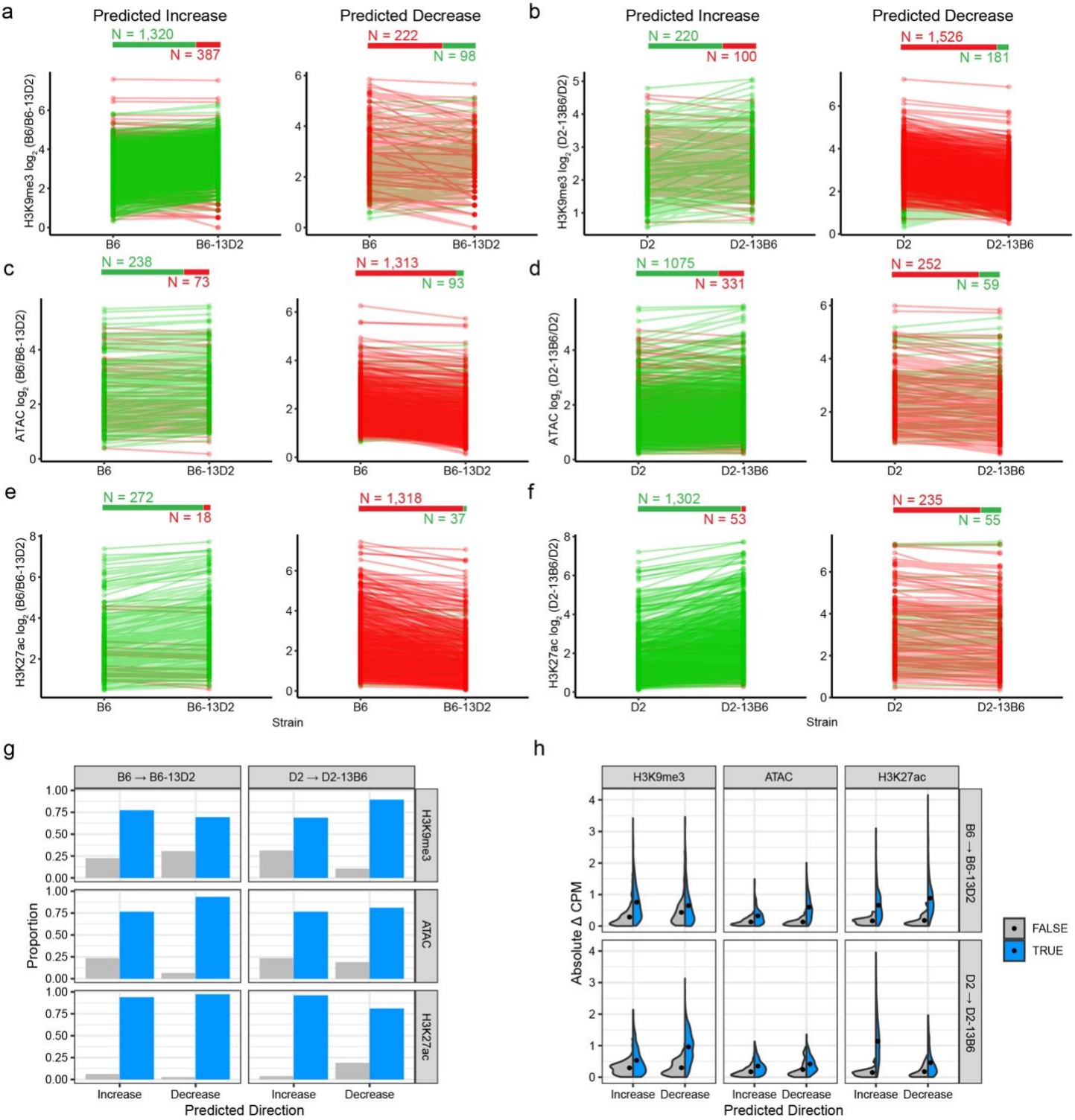


Fig. 5. Chr13 *trans*-QTL regulates heterochromatin formation at hundreds of distal loci. **a-f** Histone modification or chromatin accessibility at Chr13 *trans*-targets in parental (B6 and D2) and reciprocal congenic (B6-13D2 and D2-13B6) strains for H3K9me3 ChIP-seq (**a, b**), ATAC-seq (**c, d**), and H3K27ac ChIP-seq (**e, f**). Each line represents a single *trans*-target, connecting measurements between the parental strain (left) and its corresponding congenic (right). B6→B6-13D2 panels (**a, c, e**) test the effect of introducing the D2 Chr13 region into the B6 background, while D2→D2-13B6 panels (**b, d, f**) test the reciprocal introduction of the B6 Chr13 region into the D2 background. Within each panel, loci are separated by predicted direction of change based on QTL effect direction from BXD mapping: "Predicted Increase" (left) indicates targets expected to gain signal in the congenic strain, while "Predicted Decrease" (right) indicates targets expected to lose signal. Green lines indicate increasing values, while red lines indicate decreasing values. Numbers above each plot show total

880 *trans*-targets tested that increase (green) or decrease (red). Statistical significance assessed by exact binomial
881 test. H3K9me3 enrichment: **a** increase p < 2.2e-16, decrease p = 3.3e-12; **b** increase p = 1.7e-11, decrease p
882 < 2.2e-16 ATAC-seq: **c** increase p < 2.2e-16, decrease p < 2.2e-16; **d** increase p < 2.2e-16, decrease p <
883 2.2e-16. H3K27ac: **e** increase p < 2.2e-16, decrease p < 2.2e-16; **f** increase p < 2.2e-16, decrease p < 2.2e-
884 16. **g** Observed proportion of Chr13 *trans*-targets for which chromatin state changes in the direction predicted
885 (i.e., True) based on QTL effect and dominant repression. **h** Violin plot showing distribution of absolute
886 changes in counts per million (CPM) at Chr13 *trans*-QTL targets. *Trans*-targets that change as predicted (blue)
887 or do not change as predicted (grey). Dots indicated median value.

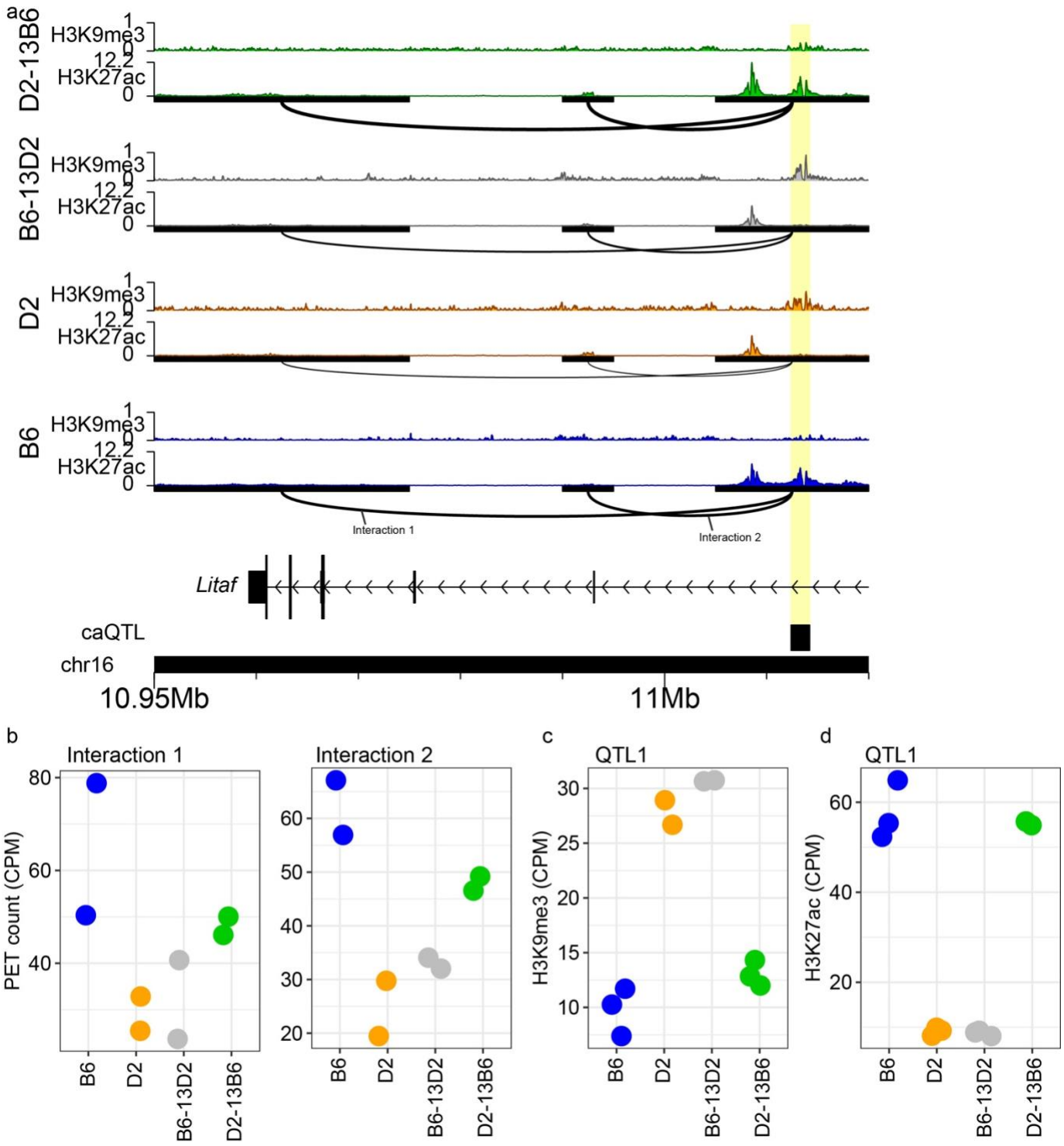


Fig. 6. Chr13 coordinates heterochromatin and 3D architecture at *Litaf* locus. **a** Genome profiles for H3K27ac and H3K9me3 ChIP enrichment from B6 (blue), D2 (orange), B6-13D2 (grey) and D2-13B6 (green) at the *Litaf* locus. HiChIP interactions indicated below H3K27ac tracks with the thickness of the arc connecting two anchors representing contact frequency. Highlight indicates a *trans*-target (black box) overlapping a contact anchor and variable histone modifications. **b** HiChIP PET counts for all four strains for two interactions indicated in panel a (Interaction 1: BD p = 4.67e-4, BB13D p = 3.14e-4, DD13B p = 2.68e-2; Interaction 2: BD p = 9.94e-5, BB13D p = 1.51e-3, DD13B p = 3.15e-3). **c** H3K9me3 enrichment at Chr13 *trans*-target highlighted panel a (BD p = 8.25e-12, BB13D p = 2.98e-15, DD13B p = 4.81e-8). **d** H3K27ac enrichment at Chr13 *trans* caQTL target highlighted panel a (BD p = 6.62e-108, BB13D p = 4.54e-119, DD13B p = 9.24e-103). BD is the comparison between B6 and D2; BB13D is the comparison between B6 and B6-13D2; DD13B is the comparison between D2 and D2-13B6.

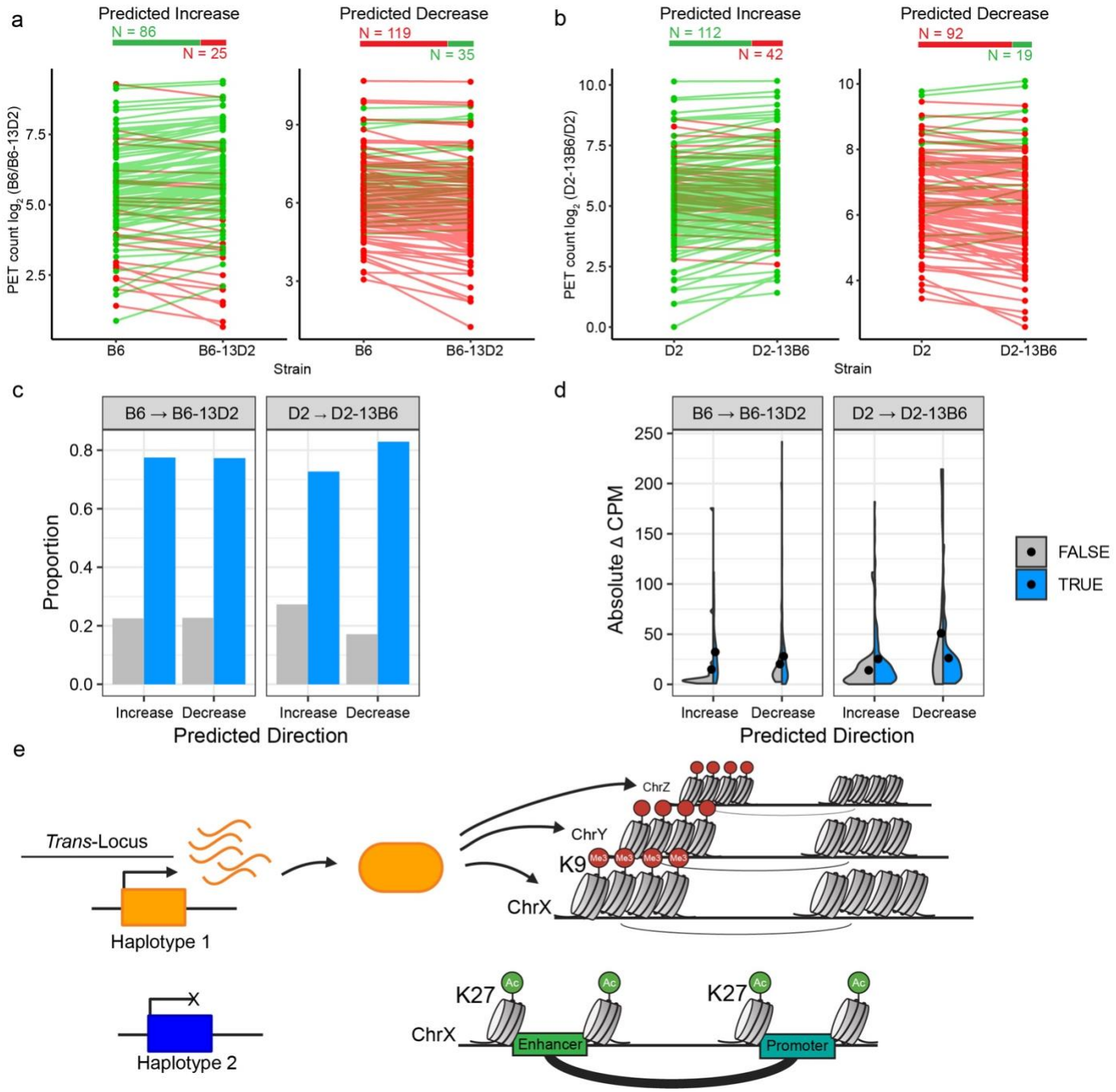
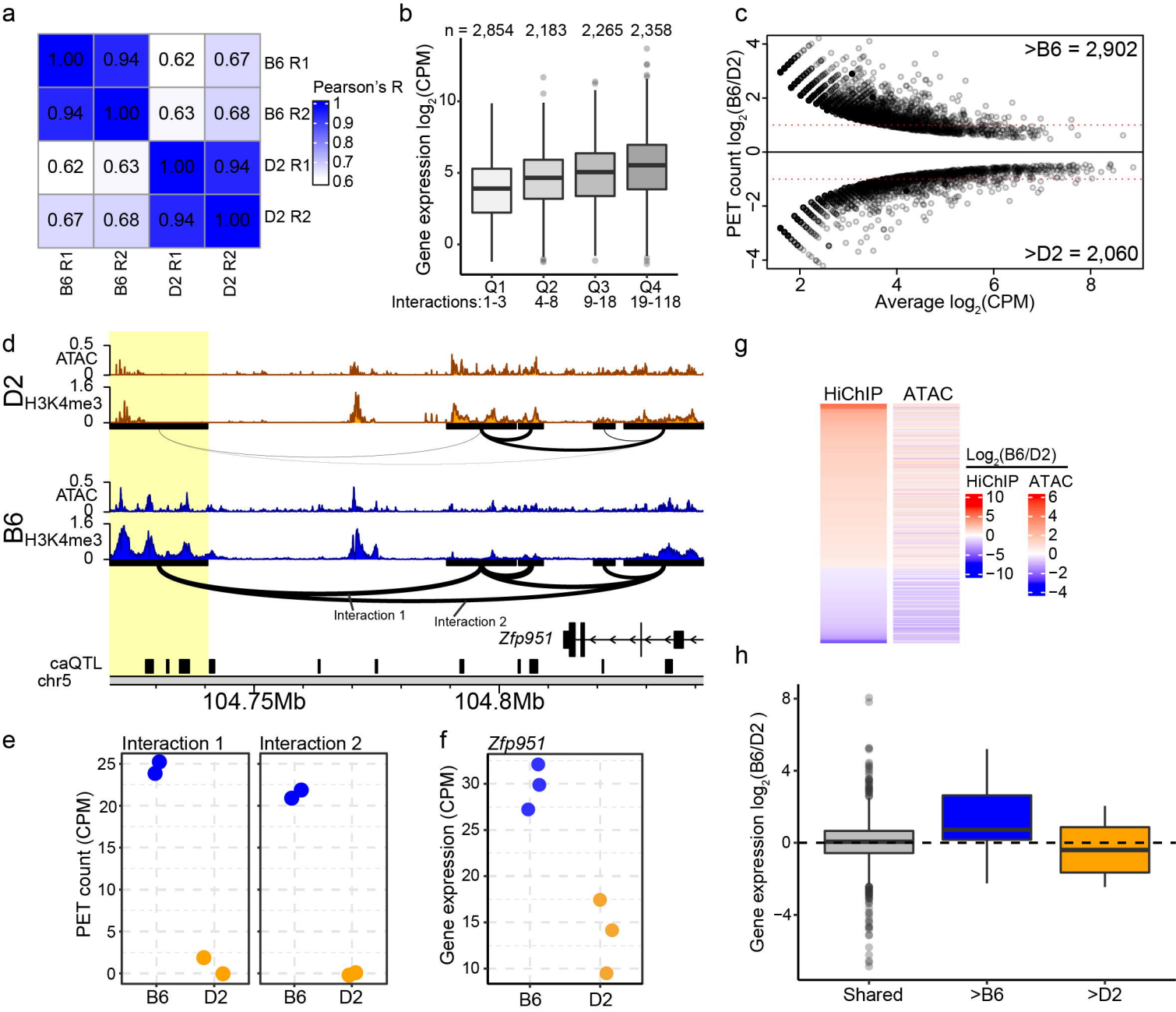
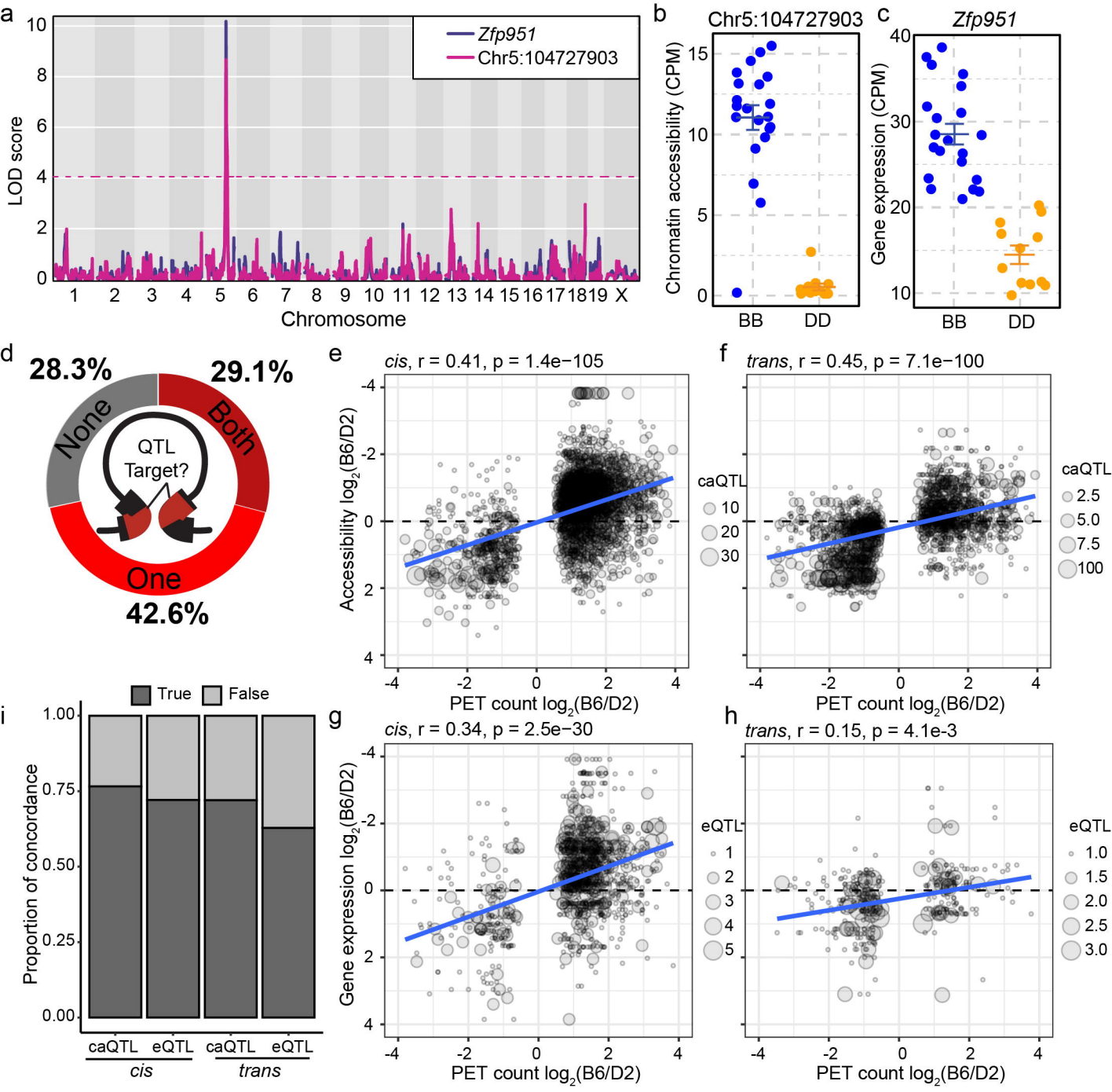
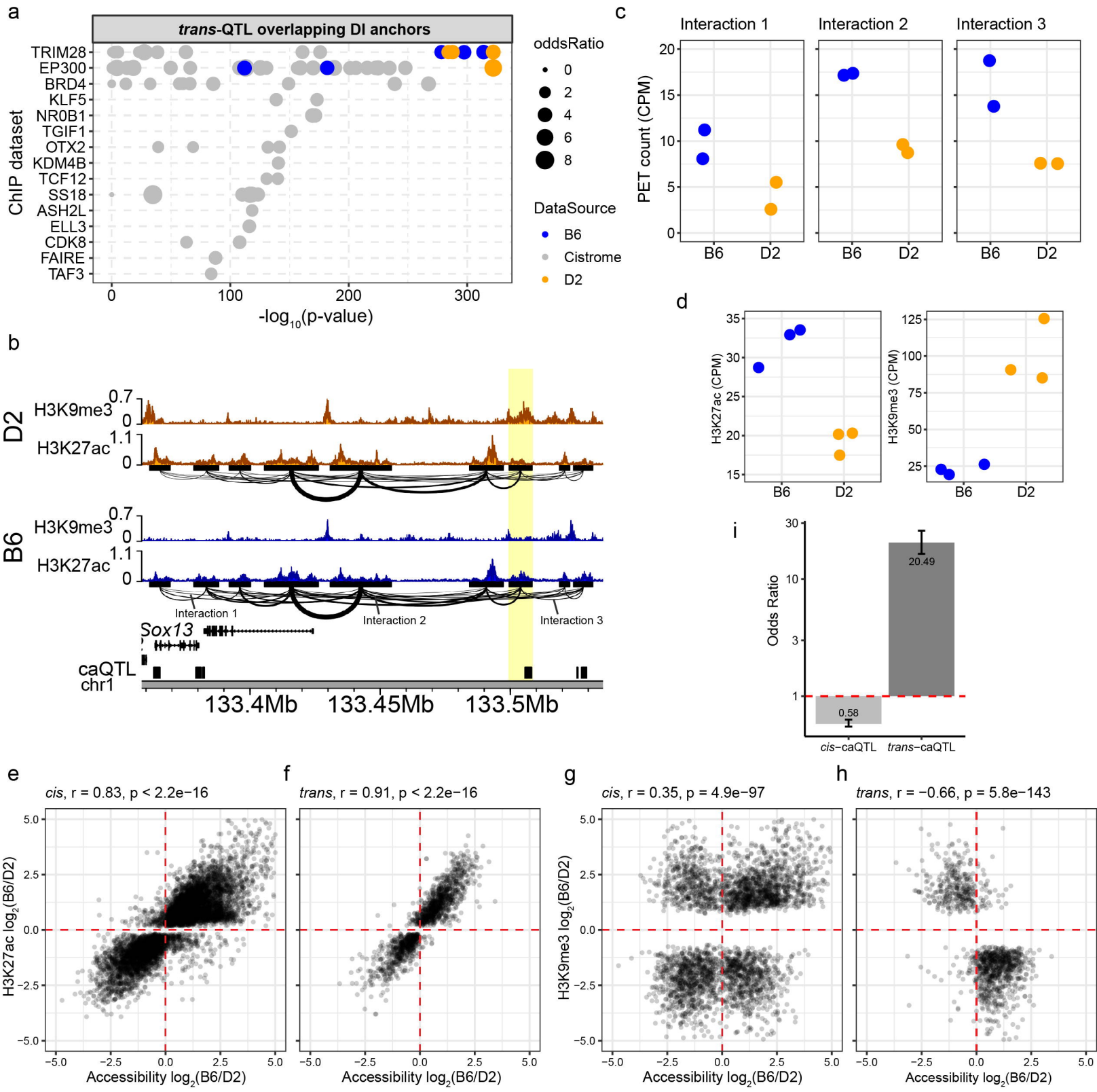
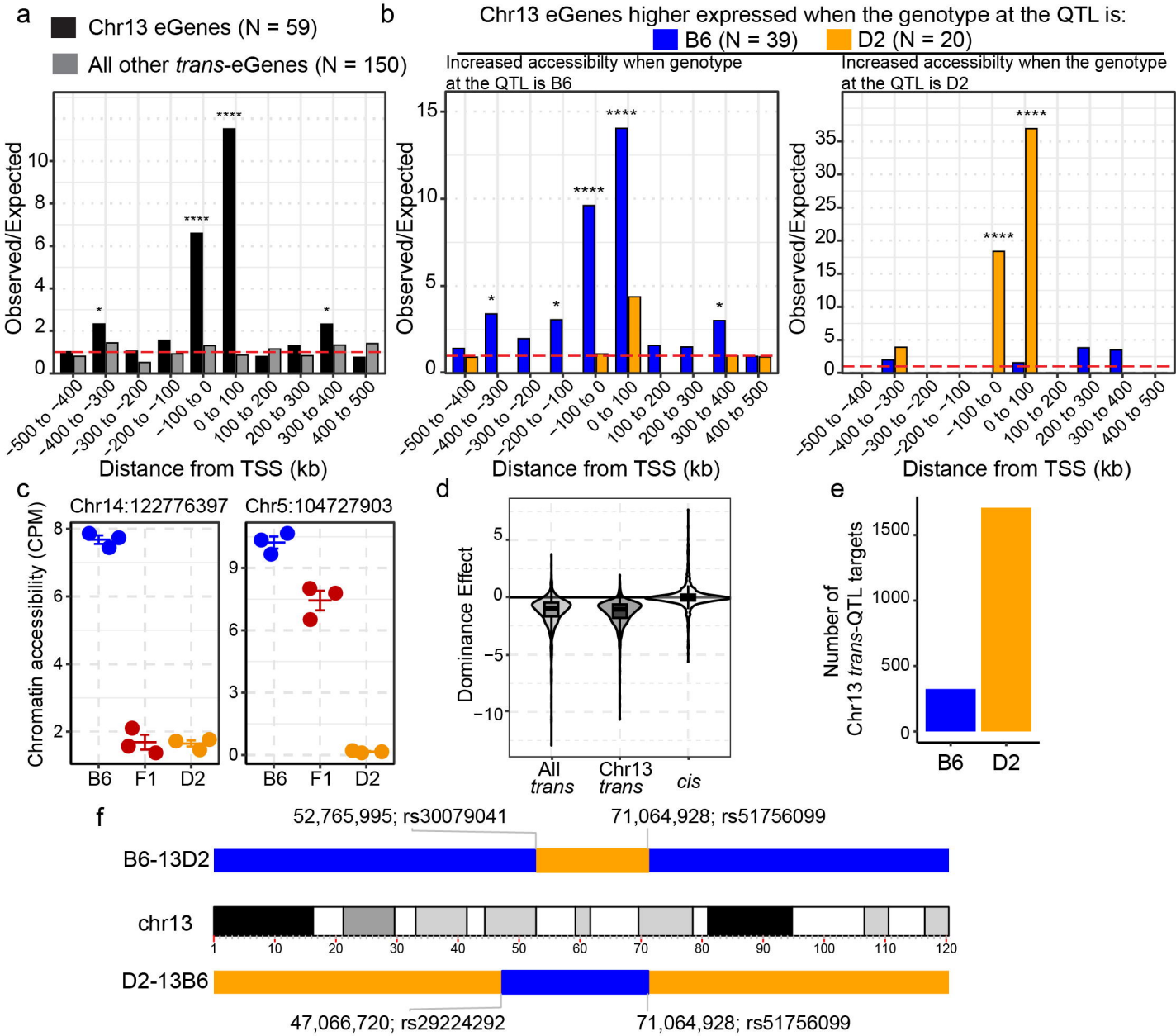


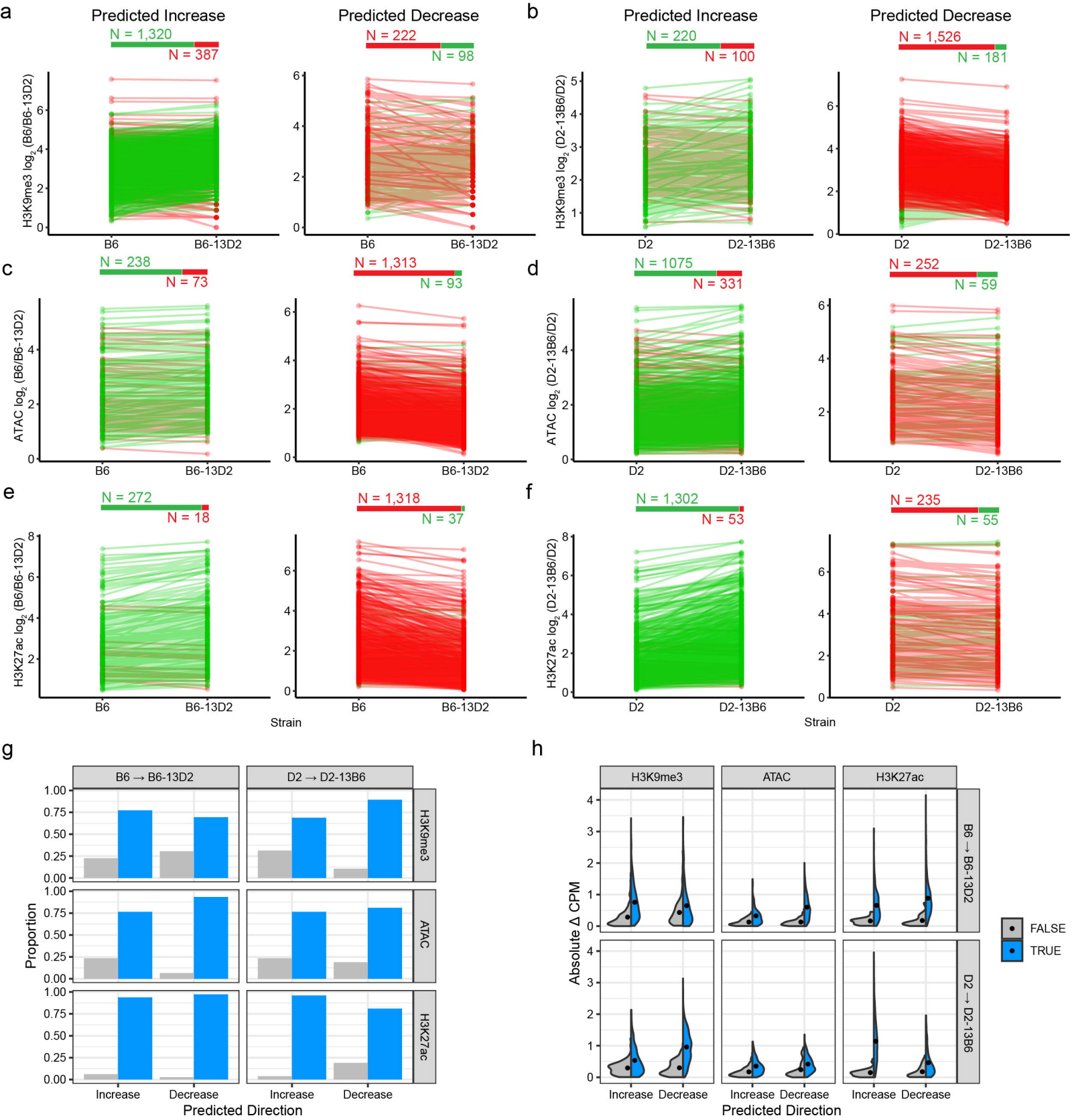
Fig. 7. Heterochromatin coordinates genome-wide changes in 3D architecture. **a**, **b** Validation of predicted changes in 3D contact frequency at Chr13 *trans*-targets using H3K27ac HiChIP. Experimental design and visualization as in Fig. 5a-f, with each line representing a single interaction overlapping a Chr13 *trans*-target. B6→B6-13D2 (**a**) and D2→D2-13B6 (**b**). Loci were split by predicted direction of change based on BXD QTL mapping. Green lines indicate increasing PET counts; redlines indicate decreasing PET counts. Numbers above show total interactions tested that increase (green) or decrease (red). Exact binomial test: **a** increase p = 5.03e-9, decrease p = 6.59e-12; **b** increase p = 1.53e-8, decrease p = 1.12e-12. **c** Observed proportion of interactions that changed as predicted (blue). **d** Distribution of absolute changes in PET counts at Chr13 *trans*-targets, showing validated predictions (blue) and non-validated targets (grey). Dots indicate median values. **e** Model illustrating *trans*-regulation through heterochromatin formation. In this example, the orange haplotype *encodes* diffusible repressor(s) that establish H3K9me3-marked heterochromatin at distal *trans*-targets (ChrX, Y, Z), reducing chromatin accessibility and weakening 3D contacts. Whereas the blue haplotype lacks active repressor(s), permitting open chromatin and stronger enhancer-promoter interactions and gene expression at the same distal locus.

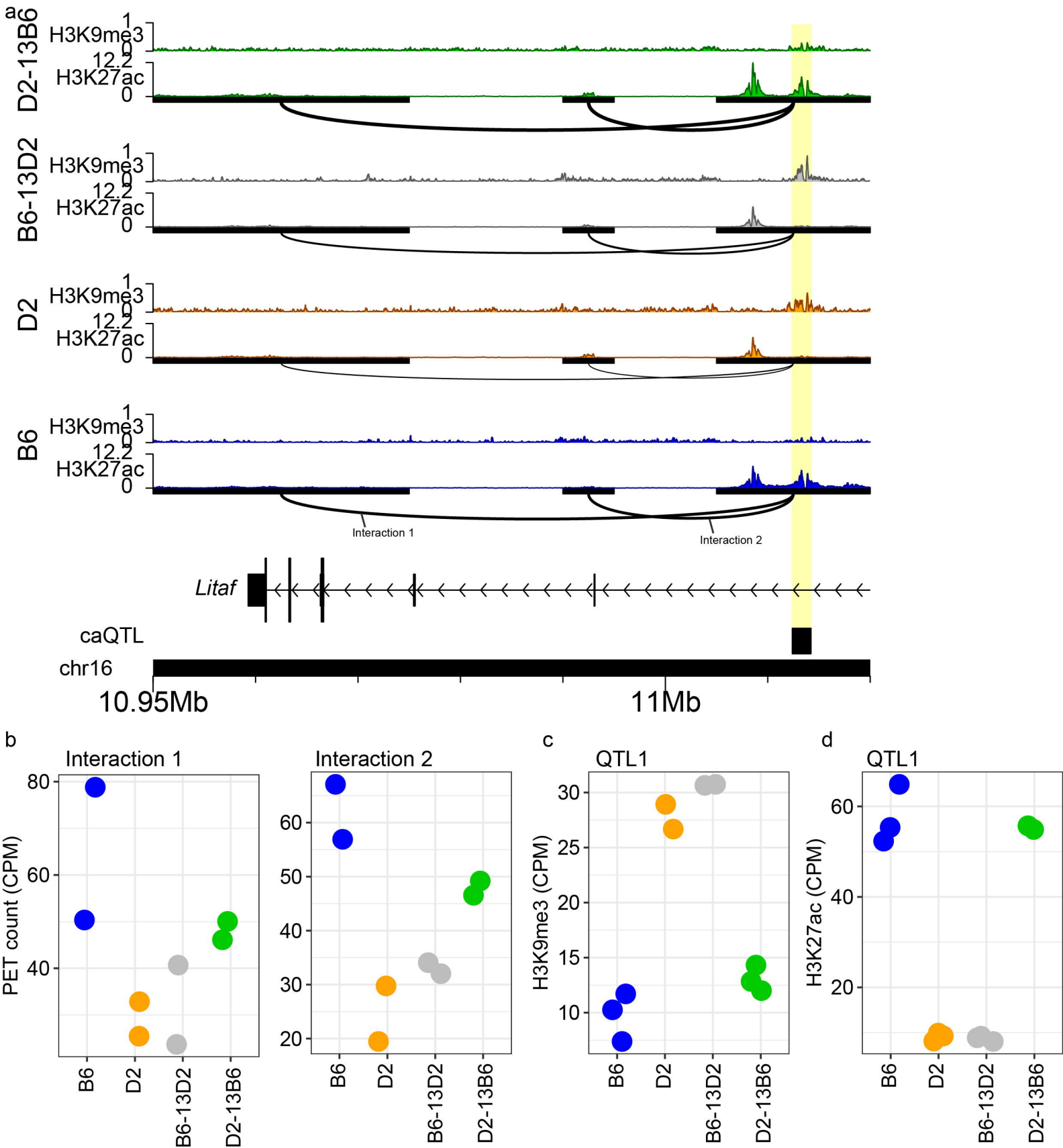

a
1.00
0.94
0.62
0.67
B6 R1
0.94
1.00
0.63
0.68
B6 R2
0.62
0.63
1.00
0.94
D2 R1
0.67
0.68
0.94
1.00
D2 R2
B6 R1
B6 R2
D2 R1
D2 R2
Pearson's R
1
0.9
0.8
0.7
0.6
b
n = 2,854 2,183 2,265 2,358
Gene expression log2(CPM)
Q1 Q2 Q3 Q4
Interactions: 1-3 4-8 9-18 19-118
c
>B6 = 2,902
>D2 = 2,060
PET count log2(B6/D2)
Average log2(CPM)
d
D2
ATAC
H3K4me3
B6
ATAC
H3K4me3
Interaction 1
Interaction 2
Zfp951
caQTL
chr5
104.75Mb
104.8Mb
e
Interaction 1
Interaction 2
PET count (CPM)
B6
D2
f
Zfp951
Gene expression (CPM)
B6
D2
g
HiChIP
ATAC
Log2(B6/D2)
HiChIP
ATAC
h
Gene expression log2(B6/D2)
Shared
>B6
>D2

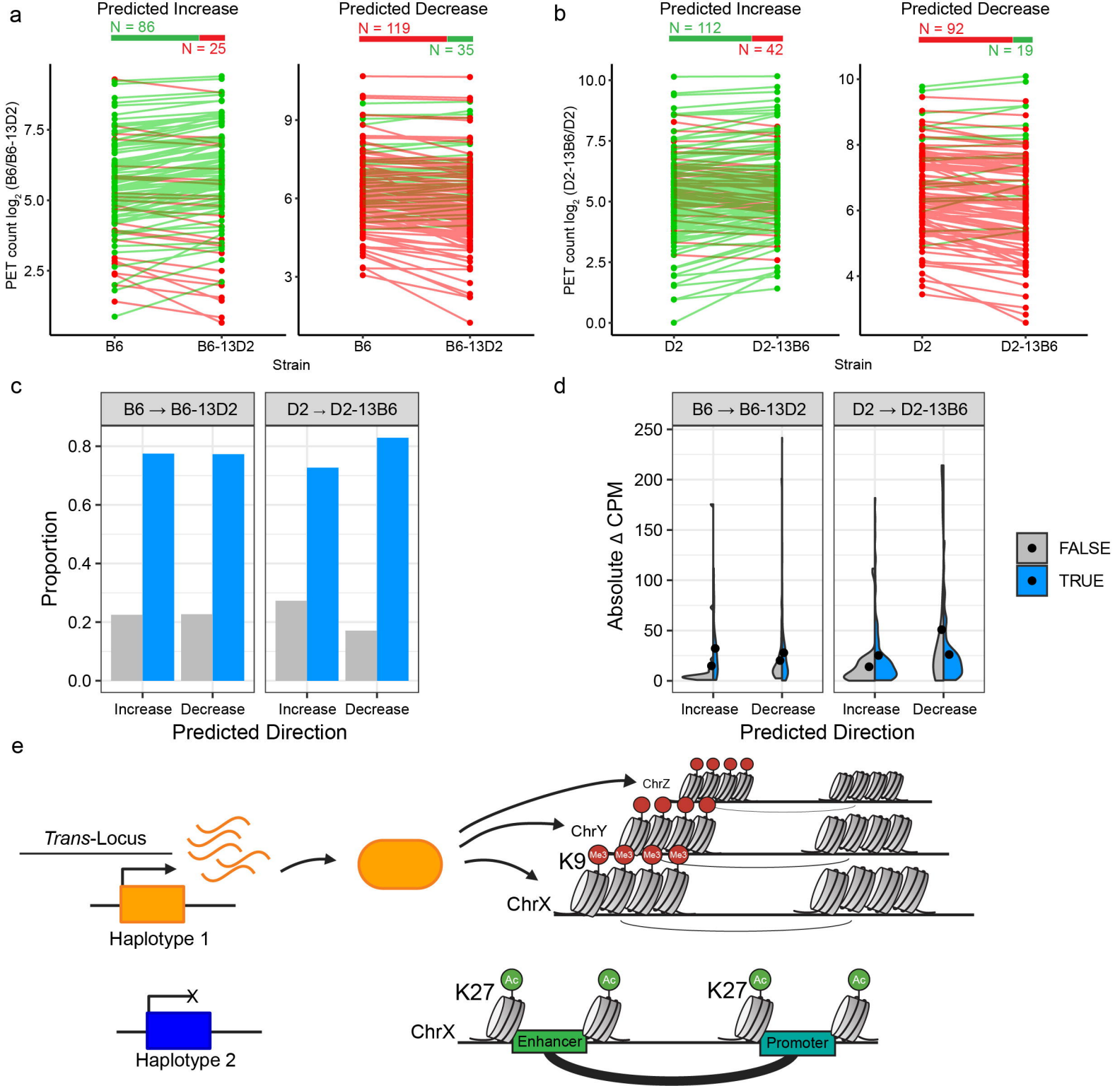

a
Zfp951
Chr5:104727903
LOD score
Chromosome
b
Chr5:104727903
Chromatin accessibility (CPM)
BB
DD
c
Zfp951
Gene expression (CPM)
BB
DD
d
28.3%
29.1%
None
Both
QTL
Target?
One
42.6%
e
cis, r = 0.41, p = 1.4e−105
Accessibility log2(B6/D2)
PET count log2(B6/D2)
caQTL
f
trans, r = 0.45, p = 7.1e−100
PET count log2(B6/D2)
caQTL
i
True
False
Proportion of concordance
caQTL
eQTL
caQTL
eQTL
cis
trans
g
cis, r = 0.34, p = 2.5e−30
Gene expression log2(B6/D2)
PET count log2(B6/D2)
eQTL
h
trans, r = 0.15, p = 4.1e-3
PET count log2(B6/D2)
eQTL


a
trans-QTL overlapping DI anchors
ChIP dataset
TRIM28
EP300
BRD4
KLF5
NR0B1
TGIF1
OTX2
KDM4B
TCF12
SS18
ASH2L
ELL3
CDK8
FAIRE
TAF3
-log10(p-value)
oddsRatio
DataSource
B6
Cistrome
D2
b
D2
H3K9me3
H3K27ac
B6
Interaction 1
Interaction 2
Interaction 3
Sox13
caQTL
chr1
133.4Mb
133.45Mb
133.5Mb
c
Interaction 1
Interaction 2
Interaction 3
PET count (CPM)
d
H3K27ac (CPM)
H3K9me3 (CPM)
i
Odds Ratio
20.49
0.58
cis−caQTL
trans−caQTL
e
cis, r = 0.83, p < 2.2e−16
f
trans, r = 0.91, p < 2.2e−16
g
cis, r = 0.35, p = 4.9e−97
h
trans, r = −0.66, p = 5.8e−143
H3K27ac log2(B6/D2)
H3K9me3 log2(B6/D2)
Accessibility log2(B6/D2)


a
Chr13 eGenes (N = 59)
All other trans-eGenes (N = 150)
Observed/Expected
Distance from TSS (kb)
b
Chr13 eGenes higher expressed when the genotype at the QTL is:
B6 (N = 39)
D2 (N = 20)
Increased accessibilty when genotype at the QTL is B6
Increased accessibilty when the genotype at the QTL is D2
c
Chr14:122776397
Chr5:104727903
Chromatin accessibility (CPM)
B6
F1
D2
d
Dominance Effect
All trans
Chr13 trans
cis
e
Number of Chr13 trans-QTL targets
B6
D2
f
52,765,995; rs30079041
71,064,928; rs51756099
B6-13D2
chr13
D2-13B6
47,066,720; rs29224292
71,064,928; rs51756099


a
Predicted Increase
N = 1,320
N = 387
Predicted Decrease
N = 222
N = 98
H3K9me3 log2 (B6/B6-13D2)
B6
B6-13D2
b
Predicted Increase
N = 220
N = 100
Predicted Decrease
N = 1,526
N = 181
H3K9me3 log2 (D2-13B6/D2)
D2
D2-13B6
c
N = 238
N = 73
N = 1,313
N = 93
ATAC log2 (B6/B6-13D2)
d
N = 1075
N = 331
N = 252
N = 59
ATAC log2 (D2-13B6/D2)
e
N = 272
N = 18
N = 1,318
N = 37
H3K27ac log2 (B6/B6-13D2)
f
N = 1,302
N = 53
N = 235
N = 55
H3K27ac log2 (D2-13B6/D2)
Strain
g
B6 → B6-13D2
D2 → D2-13B6
H3K9me3
ATAC
H3K27ac
Proportion
Increase
Decrease
Predicted Direction
h
H3K9me3
ATAC
H3K27ac
B6 → B6-13D2
D2 → D2-13B6
Absolute Δ CPM
FALSE
TRUE
Increase
Decrease
Predicted Direction


a
D2-13B6
H3K9me3
H3K27ac
B6-13D2
D2
B6
Interaction 1
Interaction 2
Litaf
caQTL
chr16
10.95Mb
11Mb
b
Interaction 1
Interaction 2
PET count (CPM)
c
QTL1
H3K9me3 (CPM)
d
QTL1
H3K27ac (CPM)
B6
D2
B6-13D2
D2-13B6


a
Predicted Increase
N = 86
N = 25
Predicted Decrease
N = 119
N = 35
PET count log2 (B6/B6-13D2)
B6
B6-13D2
Strain
b
Predicted Increase
N = 112
N = 42
Predicted Decrease
N = 92
N = 19
PET count log2 (D2-13B6/D2)
D2
D2-13B6
Strain
c
B6 → B6-13D2
D2 → D2-13B6
Proportion
Increase
Decrease
Predicted Direction
d
B6 → B6-13D2
D2 → D2-13B6
Absolute Δ CPM
FALSE
TRUE
Increase
Decrease
Predicted Direction
e
Trans-Locus
Haplotype 1
Haplotype 2
ChrZ
ChrY
K9
Me3
ChrX
K27
Ac
Enhancer
Promoter