

1 **Common DNA sequence variation influences 3-dimensional conformation of the human**
2 **genome**

3
4 David U. Gorkin^{1,2*}, Yunjiang Qiu^{1,3*}, Ming Hu^{4*†}, Kipper Fletez-Brant^{5,6}, Tristin Liu¹, Anthony D.
5 Schmitt^{1,10}, Amina Noor², Joshua Chiou^{7,11}, Kyle J Gaulton⁷, Jonathan Sebat², Yun Li⁸, Kasper
6 D. Hansen^{5,6}, Bing Ren^{1,2,9†}

7
8 * equal contribution

9 † correspondence to biren@ucsd.edu, hum@ccf.org

10

11 ¹ Ludwig Institute for Cancer Research. La Jolla, California USA.

12 ² Department of Cellular and Molecular Medicine, University of California San Diego. La Jolla,
13 California USA.

14 ³ Bioinformatics and Systems Biology Graduate Program, University of California San Diego. La
15 Jolla, California USA.

16 ⁴ Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic
17 Foundation. Cleveland, Ohio USA.

18 ⁵ McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine.
19 Baltimore, Maryland USA.

20 ⁶ Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health. Baltimore,
21 Maryland USA.

22 ⁷ Department of Pediatrics, University of California San Diego. La Jolla, California, USA.

23 ⁸ Department of Genetics, Biostatistics, and Computer Science, University of North Carolina,
24 Chapel Hill, Chapel Hill, North Carolina USA.

25 ⁹ Institute of Genomic Medicine and Moores Cancer Center, University of California San Diego.
26 La Jolla, California USA.

27 ¹⁰ Current address: Arima Genomics, San Diego, CA, 92121, USA

28 ¹¹ Biomedical Sciences Graduate Program, University of California San Diego. La Jolla,
29 California USA.

30

31 **ABSTRACT**

32

33 The 3-dimensional (3D) conformation of chromatin inside the nucleus is integral to a
34 variety of nuclear processes including transcriptional regulation, DNA replication, and DNA

35 damage repair. Aberrations in 3D chromatin conformation have been implicated in
36 developmental abnormalities and cancer. Despite the importance of 3D chromatin conformation
37 to cellular function and human health, little is known about how 3D chromatin conformation
38 varies in the human population, or whether DNA sequence variation between individuals
39 influences 3D chromatin conformation. To address these questions, we performed Hi-C on
40 Lymphoblastoid Cell Lines (LCLs) from 20 individuals. We identified thousands of regions
41 across the genome where 3D chromatin conformation varies between individuals and found that
42 this conformational variation is often accompanied by variation in gene expression, histone
43 modifications, and transcription factor (TF) binding. Moreover, we found that DNA sequence
44 variation influences several features of 3D chromatin conformation including loop strength,
45 contact insulation, contact directionality and density of local *cis* contacts. We mapped hundreds
46 of Quantitative Trait Loci (QTLs) associated with 3D chromatin features and found evidence that
47 some of these same variants are associated at modest levels with other molecular phenotypes
48 as well as complex disease risk. Our results demonstrate that common DNA sequence variants
49 can influence 3D chromatin conformation, pointing to a more pervasive role for 3D chromatin
50 conformation in human phenotypic variation than previously recognized.

51

52 INTRODUCTION

53

54 3-dimensional (3D) organization of chromatin is essential for proper regulation of gene
55 expression¹⁻³, and plays an important role in other nuclear processes including DNA
56 replication^{4,5}, X chromosome inactivation⁶⁻⁹, and DNA repair^{10,11}. Many recent insights about 3D
57 chromatin conformation have been enabled by a suite of technologies based on Chromatin
58 Conformation Capture (3C)¹². A high-throughput version of 3C called “Hi-C” enables the
59 mapping of 3D chromatin conformation at genome-wide scale¹³, and has revealed several key
60 features of 3D chromatin conformation including: 1) compartments (often referred to as “A/B
61 compartments”), which refer to the tendency of loci with similar transcriptional activity to
62 physically segregate in 3D space¹³⁻¹⁵, 2) chromatin domains (often referred to as Topologically
63 Associating Domains, or TADs) demarcated by sharp boundaries across which contacts are
64 relatively infrequent¹⁶⁻¹⁸, 3) chromatin loops, which describe point-to-point interactions that occur
65 more frequently than would be expected based on the linear distance between interacting loci,
66 and often anchored by convergent CTCF motif pairs¹⁴, and 4) Frequently Interacting Regions
67 (FIREs), which are regions of increased local interaction frequency enriched for tissue-specific
68 genes and enhancers^{19,20}.

69 Previous studies have used Hi-C to profile 3D chromatin conformation across different
70 cell types^{14,16,21}, different primary tissues¹⁹, different cell states²², and in response to different
71 genetic and molecular perturbations²³⁻²⁷, producing a wealth of knowledge about key features of
72 3D chromatin conformation. However, to our knowledge no study to date has measured
73 variation in 3D chromatin conformation across more than a handful of unrelated individuals.
74 Several observations demonstrate that at least in some cases DNA sequence variation between
75 individuals can alter 3D chromatin organization with pathological consequences²⁸. Pioneering
76 work by Mundlos and colleagues described several cases in which rearrangements of TAD
77 structure lead to gene dysregulation and consequent developmental malformations^{29,30}. In
78 cancer, somatic mutations and aberrant DNA methylation can disrupt TAD boundaries leading
79 to dysregulation of proto-oncogenes^{31,32}. Moreover, many genetic variants associated with
80 human traits by GWAS occur in distal regulatory elements that loop to putative target gene
81 promoters in 3D, and in some cases, the strength of these looping interactions has been shown
82 to vary between alleles of the associated SNP^{33,34}. Although these studies demonstrate that that
83 both large effects as well as more subtle aberrations of 3D chromatin conformation are potential
84 mechanisms of disease, population-level variation in 3D chromatin conformation more broadly
85 has remained unexplored.

86 In the present study, we set out to characterize inter-individual variation in 3D chromatin
87 conformation by performing Hi-C on Lymphoblastoid Cell Lines (LCLs) derived from individuals
88 whose genetic variation has been cataloged by the HapMap or 1000 Genomes Consortia³⁵.
89 LCLs have been used as a model system to study variation in several other molecular
90 phenotypes including gene expression, histone modifications, transcription factor (TF) binding,
91 and chromatin accessibility³⁶⁻⁴². These previous efforts provide a rich context to explore variation
92 in 3D chromatin conformation identified in this model system. Through integrative analyses, we
93 found that inter-individual variation in 3D chromatin conformation occurs on many levels
94 including compartments, TAD boundary strengths, FIREs, and looping interaction strengths.
95 Moreover, we found that variation in 3D chromatin conformation coincides with variation in
96 activity of the underlying genome sequence as evidenced by transcription, histone
97 modifications, and TF binding. Although our sample size is small, we observe reproducible
98 effects of DNA sequence variation on 3D chromatin conformation and identify hundreds of
99 Quantitative Trait Loci (QTLs) associated with multiple features of 3D chromatin conformation.
100 Our results demonstrate that variation in 3D chromatin conformation is readily detectable from
101 Hi-C data, often overlaps with regions of transcriptomic and epigenomic variability, and is
102 influenced in part by genetic variation that may contribute to disease risk.

103

104 RESULTS

105

106 Mapping 3D chromatin conformation across individuals

107

108 To generate maps of 3D chromatin conformation suitable for comparison across
109 individuals, we performed “dilution” Hi-C on LCLs derived from 13 Yoruban individuals (including
110 one trio), one Puerto Rican trio, and one Han Chinese trio (19 individuals total; Supplemental
111 Table 1). We also include published Hi-C data from one European LCL (GM12878) generated
112 previously by our group using the same protocol⁴³, for a total of 20 individuals from four different
113 populations. Many of these same LCLs have been used in previous genomic studies^{38,40,42},
114 allowing us to leverage multiple transcriptomic and epigenomic datasets in our analysis below
115 (Supplemental Table 2). Importantly, 18 of these individuals have had their genetic variation
116 cataloged by the 1000 Genomes Consortium^{35,44} (Supplemental Table 1), which allowed us to
117 examine the influence of genetic variation on 3D chromatin conformation. Two replicates of Hi-C
118 were performed on each LCL, with each replicate performed on cells grown independently in
119 culture for at least two passages (Supplemental Table 3).

120 All Hi-C data were processed using a uniform pipeline that incorporates the WASP
121 approach^{40,45} to eliminate allelic mapping biases (see **methods** section 2a). For each sample,
122 we mapped a series of well-established Hi-C-derived features including 40Kb resolution contact
123 matrices, Directionality Index (DI)¹⁶, Insulation score (INS)⁷, and compartmentalization¹³ (Figure
124 1a; Supplemental Figure 1a-c). Compartmentalization is measured by the first Principal
125 Component (PC1) of Hi-C contact matrices, and thus we use the acronym “PC1” below to refer
126 to this measure of compartmentalization. We also identified regions known as Frequently
127 Interacting Regions (FIREs)¹⁹ and their corresponding “FIRE scores”, which measure how
128 frequently a given region interacts with its neighboring regions (15~200kb). The concept of FIRE
129 is based on the observation that the frequency of contacts at this distance is not evenly
130 distributed across the genome, but rather, tends to peak in regions showing epigenomic
131 signatures of transcriptional and regulatory activity (Supplemental Figure 2). As we have shown
132 previously^{19,20}, FIRE regions often overlap putative enhancer elements (Supplemental Figure
133 1d-e). We did not call “chromatin loops” in this study because our data was not of sufficient
134 resolution, but we use a set of loops called previously in the LCL GM12878¹⁴ to examine
135 variation in loop strength among the LCLs in our study. Aggregate analysis shows that these
136 published LCL loops are generally reproduced in our data (Supplemental Figure 3).

137

138 **3D chromatin conformation variations between individuals**

139

140 After uniformly processing all Hi-C data (see **methods** section 2), we compared
141 chromatin conformation across LCLs at the level of contact matrices and multiple derived
142 features (PC1, DI, INS, and FIRE). From a genome-wide perspective, each of these 3D
143 chromatin features shows a signature consistent with reproducible inter-individual variation
144 whereby replicates from the same individual (i.e. same LCL) are more highly correlated than
145 datasets from different individuals (PC1 $p=2.4e-7$, INS $p=1.6e-7$, DI $p=3.3e-7$, FIRE $p=0.0157$
146 by Wilcoxon rank sum test; Figure 1b-d, Supplemental Figure 4a-f). The Hi-C data also cluster
147 by population (Supplemental Figure 4f-g) consistent with an influence from genetic background,
148 but we note that this population-level clustering can be caused by other factors such as batch of
149 sample acquisition⁴⁶.

150 Despite generally high correlations of Hi-C data across individuals, we frequently
151 observed regions where 3D chromatin conformation varies reproducibly between individuals
152 (example shown in Figure 2a, Supplemental Figure 5a). To more systematically identify regions
153 of variable 3D chromatin conformation, we used the “*limma*” package⁴⁷ to identify regions where
154 variation between individuals was more significant than variation between two replicates from
155 the same individual. We applied this approach to DI, INS, FIRE, and PC1. For each metric, we
156 first defined a set of testable 40kb bins across the genome by filtering out bins with low levels of
157 signal across all individuals or near structural variants that can appear as aberrations in Hi-C
158 maps⁴⁸ (see **methods** section 4a). We then applied a False Discovery Rate (FDR) threshold of
159 0.1 and merged neighboring variable bins, resulting in the identification of 2,318 variable DI
160 regions, 2,485 variable INS regions, 1,996 variable FIRE regions, and 7,732 variable PC1
161 regions (Figure 2b, Supplemental Table 4, Supplemental Figure 5b). We note that there is
162 strong overlap between the variable DI, INS, FIRE, and PC1 regions detected across all 20
163 LCLs and those detected using only the 11 unrelated YRI LCLs, which suggests that potential
164 confounding effects of variation between different populations are not driving the identification of
165 these variable regions (Supplemental Figure 5c). Although each metric has a unique set of
166 testable bins, we found significant enrichment for bins that are variable in more than one metric
167 (Figure 2c, Supplemental Figure 5d-e), indicating that the same regions often vary across
168 multiple features of 3D chromatin conformation.

169 We next used Fluorescent In Situ Hybridization (FISH) to examine whether variable
170 regions detected by Hi-C are consistent with distance measurements from imaging data (Figure

171 2d-e). Focusing on a variable DI region on chromosome 15 (chr15:96720000-96920000; hg19),
172 we performed FISH in LCLs from four individuals with different levels of DI at the variable region
173 being evaluated (YRI-3, YRI-4, YRI-5, YRI-8). We used three BAC probes that hybridize
174 respectively to the variable DI region (“center”, probe covers chr15:96715965-96898793), a
175 region approximately 668Kb upstream (“upstream”, probe covers chr15:95897555-96047720),
176 or a region approximately 590Kb downstream (“downstream”, probe covers chr15:97488414-
177 97648104). We found that distances between the center probe and these flanking probes vary
178 significantly between individuals with strong upstream contact bias as measured by DI (YRI-4,
179 YRI-8) and individuals without this upstream contact bias (YRI-3, YRI-5)(Figure 2d-e, center-
180 upstream distance $p=0.017$, center-downstream distance $p=1.7e-5$ by Wilcoxon rank sum test).
181 Moreover, we found that the center probe is closer to the upstream than the downstream probe
182 in the two individuals with strong upstream DI signal at the central variable DI region ($p=3.2e-3$
183 for YRI-3, $p=1.5e-4$ for YRI-5 by Wilcoxon rank sum test). However, this trend is reversed in
184 individuals without upstream DI signal where the center probe is now closer to the downstream
185 probe ($p=0.021$ for YRI-4, $p=0.1$ for YRI-8 by Wilcoxon rank sum test) (Supplemental Figure
186 6a).

187 We also sought to identify variable entries in the Hi-C contact matrix itself (“matrix cells”).
188 To facilitate this search, we used a method called Bandwise Normalization and Batch Correction
189 (BNBC) that we recently developed to normalize Hi-C data across individuals (Fletez-Brant et al.
190 Pre-print: <https://doi.org/10.1101/214361>). BNBC takes contact distance into account as a co-
191 variate because batch effects in Hi-C data can be distance-dependent. To identify variable
192 matrix cells, we performed a variance decomposition on Hi-C contact matrix cells which
193 exhibited statistically-significant variability between individuals, resulting in a measure of
194 biological variability for each bin in the contact matrix (see example in Figure 2a and
195 Supplemental Figure 5a). To identify matrix cells with significant levels of biological variability,
196 we estimated FDR using the IHW framework⁴⁹ to include the distance between anchor bins as
197 an informative covariate. At an FDR threshold of 0.1, we identified 115,817 matrix cells showing
198 significant variability between samples (Supplemental Table 5). These variable bins are heavily
199 skewed toward shorter contact distances (Figure 2f, Supplemental Figure 6b), likely due in part
200 to higher read counts and thus increased power at these distances. We observed that the
201 anchor regions of variable matrix cells overlap with variable regions of DI, INS, FIRE, and PC1
202 more often than would be expected by chance (Figure 2g; Supplemental Figure 6c). We also
203 observed that variable matrix cells tend to occur in groups (Figures 2a, 3a), suggesting that
204 variation in 3D chromatin conformation often affects more than one adjacent genomic window.

205

206 **Coordinated variation of the 3D genome, epigenome, and transcriptome**

207

208 To investigate the relationship between variation in 3D chromatin conformation and gene
209 regulation, we analyzed multiple published datasets including RNA-seq, ChIP-seq, and DNase-
210 seq data generated from some of the same LCLs in our study (Supplemental Table 2).
211 Strikingly, for all external datasets examined here, we see an enrichment for regions at which
212 3D chromatin conformation across individuals is correlated with measures of genome activity in
213 the same 40Kb bin (see example in Figure 3a and Supplemental Figure 7a). To assign a level of
214 statistical significance to these observations, we approximated the null distribution by randomly
215 permuting the sample labels of external datasets, thus disrupting the link between Hi-C and
216 ChIP/RNA/DNase-seq data from the same individual, but not changing the underlying data
217 structure (see schematic in Supplemental Figure 7b). We used these permutations to calculate
218 the bootstrap p -values in Figure 3b. Among variable PC1 regions, we observed a significant
219 enrichment for regions at which PC1 values across individuals are positively correlated with
220 histone modifications indicative of transcriptional activity including H3K27ac (bootstrap
221 $P < 0.001$), H3K4me1, and H3K4me3 (but notably less so with H3K27me3, which is marker of
222 transcriptional repression) (bootstrap $P < 0.001$ for all histone modifications, Figure 3b). The
223 correlations between PC1 and marks of transcriptional activity occur in the expected direction --
224 i.e. higher PC1 values are associated with higher gene expression and more active histone
225 modifications. Similar correlations were apparent in two distinct sets of ChIP-seq data generated
226 by different groups^{40,42}, and observed whether we use variable regions identified across all 20
227 LCLs or only across the 11 unrelated YRI LCLs (Supplemental Figure 7c).

228 The relationship between variation in 3D chromatin conformation and underlying
229 genome activity extends beyond A/B compartmentalization. At variable FIRE regions, we found
230 an abundance of regions where FIRE score is positively correlated with marks of cis-regulatory
231 activity including H3K27ac and H3K4me1 (Bootstrap $P < 0.001$; Figure 3b, Supplemental Figure
232 9a), consistent with the previously reported relationship between FIREs and cis-regulatory
233 activity^{19,20}. DI and INS values at variable regions tend to be correlated histone modification
234 levels as well as CTCF and Cohesin subunit SA1 binding (Bootstrap $P < 0.001$; Figure 3b,
235 Supplemental Figure 8a-b), which are known to influence these 3D chromatin features^{16,50,51}.
236 For INS, the relationship is directional such that higher CTCF/Cohesin binding corresponds to
237 more contact insulation (i.e. lower INS score). However, at variable DI regions the correlations
238 are not as clearly directional, reflecting current understanding that the direction of DI (i.e.

239 upstream vs downstream contact bias) is arbitrary relative to strength of CTCF/Cohesin binding.
240 We performed similar analysis on variable cells in the contact matrix, and found that the
241 interaction frequency in these matrix cells across individuals tends to be correlated with
242 epigenetic or transcriptional properties of one or both corresponding “anchor” bins (Bootstrap
243 $P < 0.001$; Figure 3b, Supplemental Figure 9b). Importantly, for all types of variable regions
244 examined here we found correlation with RNA-seq signal, indicating that at least at some
245 regions, variation in 3D chromatin features accompanies variation in gene expression.

246 We examined further whether 3D chromatin conformation at a given variable region
247 tends to be correlated with only one epigenomic property, or with several properties
248 simultaneously. We found that PC1, FIRE, INS, and DI values across individuals are often
249 correlated with multiple features of active regions (e.g. H3K27ac, H3K4me1, RNA), and anti-
250 correlated with the repressive H3K27me3 histone modification (Figure 3c,d). For DI, where
251 direction is not as clearly linked to magnitude of gene regulatory activity, we note a larger set of
252 regions with anti-correlation to features of active regions (e.g. H3K27ac, H3K4me1, RNA) and
253 positive correlation with H3K27me3 (Figure 3e,f). These results demonstrate that variation in 3D
254 chromatin conformation is often accompanied by variation in transcriptional and regulatory
255 activity of the same region. Moreover, the correlations between multiple molecular phenotypes
256 at the same region suggest that shared mechanism(s) underlie variation in these phenotypes
257 across individuals.

258

259 **Genetic loci influencing 3D chromatin conformation**

260

261 To examine genetic influence on 3D chromatin conformation we first considered genetic
262 variants overlapping CTCF motifs at chromatin loop anchors¹⁴, because disruption of these
263 CTCF motifs by genome engineering has been shown to alter chromatin looping²³. Focusing on
264 SNPs at variation-intolerant positions in anchor CTCF motifs (“anchor disrupting SNPs”, at
265 sequence weight matrix positions where a single base has a probability of > 0.75 , Figure 4a), we
266 observed a significant linear relationship between SNP genotype and the strength of
267 corresponding loops ($p = 7.6e-5$ by linear regression; Figure 4b,c). We also examined whether
268 individuals heterozygous for anchor disrupting SNPs showed allelic imbalance in loop strength.
269 To facilitate this analysis, we used the HaploSeq⁴³ method to generate chromosome-span
270 haplotype blocks for each LCL (Supplemental Table 6). Although few Hi-C read pairs overlap a
271 SNP allowing haplotype assignment (mean 7.89% of usable reads per LCL), we do observe that
272 the haplotype bearing the stronger motif allele tends to show more reads connecting the

273 corresponding loop anchors ($p=5.9e-4$ by one-sided t-test of mean > 0.5; Figure 4d). Our
274 observation that CTCF motif SNPs can modulate 3D chromatin conformation is consistent with
275 similar findings reported from ChIA-PET data⁵², and a recent report of haplotype-associate
276 chromatin loop published while this manuscript was in preparation²⁷.

277 Motivated by these preliminary observations of genetic effects on 3D chromatin
278 conformation, we next searched directly for QTLs associated with Hi-C derived features of 3D
279 chromatin conformation. Power calculations indicated that, despite limited sample size, we were
280 moderately powered to find QTLs with strong effect sizes using a linear mixed effect model
281 (LMM) approach that takes advantage of the Hi-C replicates for each LCL (Supplemental Table
282 7). Thus, we conducted a targeted search for QTLs associated with variation in FIRE, DI, INS,
283 and contact frequency. We did not include PC1 in the QTL search because we reasoned that
284 individual genetic variants would be more likely to have detectable effects on local chromatin
285 conformation rather than large-scale features like compartmentalization. For this same reason,
286 we used modified versions of DI and INS scores for the QTL search calculated with a window
287 size of 200Kb upstream and downstream of the target bin, rather than the standard 2Mb window
288 size for DI¹⁶ or 480Kb for INS⁴⁷. We also limited our QTL searches to the 11 unrelated YRI
289 individuals in our study (referred to below as the “discovery set”) to mitigate potential
290 confounding differences between populations.

291 For each 3D genome phenotype under study we identified a list of testable bins that
292 showed appreciable levels of signal in at least one individual in our discovery set (see **methods**
293 section 7 for full description of test bin and SNP selection). We also identified a set of test SNPs
294 that includes at most one tag SNP among those in perfect LD in each 40Kb bin. Response
295 variables (i.e. 3D chromatin phenotype values) were quantile normalized across the discovery
296 set. For each testable bin, we measured the association of the given 3D chromatin phenotype
297 with all test SNPs in that bin. In cases where multiple SNPs in the same bin were significantly
298 associated with the phenotype we selected only the most significantly associated SNP per bin
299 for our final QTL list. Ultimately, at an FDR of 0.2, we identified 387 FIRE-QTLs (i.e. testable
300 bins in which FIRE score is associated with at least one SNPs in that bin; comprising 6.6% of
301 tested bins), 545 DI-QTLs (4.2% of tested bins), and 911 INS-QTLs (12.0% of tested
302 bins)(Figure 4e, Supplemental Figure 10a, Supplemental Table 8). For analysis of DI-QTLs, we
303 separated the testable bins into those with upstream bias and those with downstream bias (see
304 **methods** section 7d), because we observed a Simpson's paradox when we analyzed the
305 genotype trend at all DI-QTL regions together (Supplemental figure 10b).

306 We also searched for QTLs associated directly with interaction frequency in individual
307 contact matrix cells using an LMM approach like that described above for FIRE, DI, and INS.
308 The large number of cells in a Hi-C contact matrix, together with limited sample size, made a
309 true genome-wide QTL search unfeasible. However, power calculations indicated that if we
310 limited our QTL search to a subset of cells in the matrix we could have moderate power to
311 detect strong genetic signals (Supplemental Table 7). Thus, we limited our QTL search for
312 contact matrix QTLs (“C-QTLs”) to matrix cells that showed significant biological variability in our
313 samples, as described above. We tested for association in our discovery set between the
314 BNBC-normalized interaction frequency in these variable matrix cells and the genotype of test
315 SNPs in either of the two anchor bins. We selected at most one QTL SNP per matrix cell, using
316 association p -value to prioritize, finally yielding 345 C-QTL SNPs associated with 463 matrix
317 cells at an IHW-FDR threshold of 0.2 (Figure 4f, Supplemental Table 8).

318 To evaluate the reproducibility of each of these QTLs sets (FIRE-QTLs, DI-QTLs, INS-
319 QTLs, and C-QTLs), we examined Hi-C data from 6 individuals who were not included in our
320 discovery set (we refer to these 6 individuals our “validation set”; Supplemental Table 1). These
321 individuals represent four different populations (CEU, PUR, CHS, YRI), and they include a child
322 of two individuals in the discovery set (YRI-13/NA19240 is child of YRI-11/NA19238 and YRI-
323 12/NA19239). In each case, we find a significant linear relationship in the validation set between
324 QTL genotype and the corresponding 3D chromatin phenotype ($p=1.8e-14$ for FIRE-QTLs,
325 $p=2.5e-7$ for DI-QTLs at positive DI bins, $p=0.008$ for DI-QTLs at negative DI bins, $p=3e-4$ for
326 INS-QTLs, $p=4.1e-9$ for C-QTLs; Figure 4g). To provide an additional and more stringent
327 estimate of the significance of these observations, we performed permutations by randomly
328 selecting sets of test SNPs and measuring the linear relationship between genotype and
329 phenotype in the validation set. In all cases, the observed relationship was also significant by
330 this more conservative bootstrap approach ($p<0.001$ for FIRE-QTLs, $p<0.001$ for DI-QTLs at
331 positive DI bins, $p=0.041$ for DI-QTLs at negative DI bins, $p=0.005$ for INS-QTLs, $p=0.006$ for C-
332 QTLs; Figure 4h).

333 There is little direct overlap between our different QTL sets (Supplemental figure 10c),
334 likely due to limited power and the fact that the testable bins were different for each metric.
335 However, we observed genotype-dependent INS score at FIRE-QTLs and C-QTLs, and
336 genotype-dependent FIRE score at INS-QTLs and DI-QTLs (Supplemental Figure 10d), which
337 suggested that overlapping signal between different types of 3D chromatin QTLs in present
338 below the level of test-wide significance. To more rigorously assess overlapping signal between
339 our QTL sets we examined shared association below the threshold of multiple test correction,

340 inspired by similar approaches reported elsewhere⁵³. Our underlying hypothesis is that genetic
341 association studies of two different phenotypes “X” and “Y” with overlapping (or partially
342 overlapping) genetic architecture may have few direct overlaps between significant hits due to
343 limited power or differing study designs, but the shared signal should become apparent when
344 the full range of association results are considered. To quantify this, we calculated the fraction
345 of QTLs for a given phenotype X that exceed a nominal level of significance ($p < 0.05$) when
346 tested for association with a different phenotype Y . We refer to this value as the “nominal
347 fraction” below and in figure 4i. To test whether the nominal fraction of X -QTLs was significantly
348 higher than would be expected by chance, we approximated the null distribution by calculating
349 nominal fractions for 10,000 sets of SNPs selected randomly from among all X test SNPs. In
350 almost all pairwise comparisons between 3D chromatin QTL types examined here, we find that
351 the observed nominal fractions are significantly higher than would be expected in the absence
352 of shared genetic architecture (Figure 4i,j).

353

354 **Contribution of 3D chromatin QTLs to molecular phenotypes and disease risk**

355

356 Given the correlation observed between 3D chromatin variation and epigenome
357 variation, we next investigated whether 3D chromatin QTLs could modulate both the epigenome
358 and 3D genome. Here, we made use of published ChIP-seq data for histone modifications
359 (H3K4me1, H3K4me3, H3K27ac) in a large set of 65 YRI LCLs³⁹, DNase-seq data from 59 YRI
360 LCLs³⁸, and CTCF ChIP-seq data from 15 CEU LCLs⁵⁴. Notably, most individuals in these
361 datasets were not included in our QTL discovery or validation sets (54/65 for histone
362 modification ChIP-seq, 48/59 for DNase-seq, 15/15 for CTCF ChIP-seq). In many cases, we
363 found a significant linear relationship between 3D chromatin QTL genotypes and these different
364 epigenetic phenotypes (Figure 5a, Supplemental figure 11a). For example, at FIRE-QTLs, the
365 high-FIRE allele is also associated with higher levels of active histone modifications and
366 chromatin accessibility (Figure 5a). We note that although these associations are all significant
367 by linear regression, only H3K27ac and H3K4me1 passed more conservative permutation
368 testing in which the null distribution is approximated by selecting random SNPs from the full set
369 of tested SNPs (Figure 5b). At C-QTLs, the high-contact alleles show higher levels of the
370 enhancer-associated mark H3K4me1 in the two anchor bins that connect the corresponding
371 matrix cell. Moreover, the nominal fraction of C-QTLs (i.e. fraction of c-QTLs with $p < 0.05$) in a
372 published set of H3K4me1-QTLs is significantly higher than expected in the absence of shared
373 genetic association ($p=6.9e-6$ by chi square test, bootstrap $p=0.028$; Supplemental Figure

374 11b,d). At INS-QTLs, the slope of these genotype-phenotype relationships is inverted such that
375 higher levels of histone modifications and chromatin accessibility are associated with the low
376 INS score allele (i.e. more contact insulation), although only the association with chromatin
377 accessibility is significant by both linear regression and permutation test ($p=1.6e-40$ by linear
378 regression, bootstrap $p=0.023$; Supplemental Figure 11b,d). The genotype-phenotype
379 relationships observed at DI-QTLs are not as clear as for other metrics (Figure 5b,
380 Supplemental figure 11a), but this is expected because increased histone modifications or
381 chromatin accessibility can influence DI in either direction, potentially confounding this type of
382 aggregate analysis. Anecdotally, we do observe examples of individual DI-QTLs where
383 genotype appears to correlate with epigenomic phenotype (Figure 5c).

384 Finally, we sought to examine whether 3D chromatin QTLs might contribute risk for
385 complex diseases. There are 44 direct overlaps between our 3D chromatin QTLs (or SNPs in
386 perfect LD in the same 40Kb bin) and NHGRI-EBI GWAS catalog⁵⁵ (Supplemental Table 9).
387 However, the significance of these direct overlaps is hard to assess given the differences
388 between the populations and study designs in question. Thus, here again we examined
389 overlaps below the level of genome-wide significance by looking at nominal fractions to assess
390 shared signal between association studies. We compiled full summary statistics for large GWAS
391 (>50,000 individuals) of the related immune-relevant phenotypes Crohn's Disease (CD),
392 Ulcerative Colitis (UC), and Inflammatory Bowel Disease (IBD)⁵⁶, as well as studies of the non-
393 immune phenotypes height⁵⁷ and Body Mass Index (BMI)⁵⁸. We observed striking enrichments
394 for INS-QTLs among variants with nominal associations to UC and IBD risk (1.67- and 1.65-fold,
395 respectively), and these enrichments are significant by both chi square and permutation tests
396 (INS-QTL with UC chi square $p=2.5e-16$ and bootstrap $p=0.024$; INS-QTL with IBD chi square
397 $p=5.5e-17$ and bootstrap $p=0.018$; Figure 5d,e). We also note a trend in which FIRE-QTLs show
398 nominal association with UC and IBD (1.36- and 1.58-fold enrichment, respectively), although
399 these observations fall just below the threshold of significance by the more stringent
400 permutation test (FIRE-QTL with UC chi square $p=7.6e-6$ and bootstrap $p=0.090$; FIRE-QTL
401 with IBD chi square $p=4.2e-8$ and bootstrap $p=0.056$; Figure 5d,f).

402

403 **DISCUSSION**

404

405 Our results provide the first systematic characterization of how chromatin conformation
406 varies between unrelated individuals at the population level, and as a consequence of genetic
407 variation. The most important finding of our study is that genetic variation influences multiple

408 features of 3D chromatin conformation, and does so to an extent that is detectable even with
409 limited sample size and Hi-C resolution. To the best of our knowledge, this represents the first
410 report of QTLs directly associated with 3D chromatin conformation. However, there are
411 limitations to our QTL search that are important to note here. First, the small sample size means
412 that our power to detect QTLs is limited, and in order to identify QTL sets that could be analyzed
413 in aggregate we tolerated elevated type I error by using an FDR threshold of 0.2 (as done
414 previously for molecular QTL studies with limited power⁴⁰). Second, the limited resolution of our
415 Hi-C data (40Kb) and extensive LD in our study population prevented us from identifying
416 specific causal variant(s) for validation through genetic perturbation experiments. Nonetheless,
417 we were able to validate the 3D chromatin QTL sets through aggregate analysis of Hi-C data
418 from a small set of individuals who were not included in the QTL search, and with independently
419 generated ChIP-seq and DNase-seq data from a larger set of individuals. Taken together, our
420 results show that genetic variation influences several features of 3D chromatin conformation,
421 which is an important step forward to evaluate the role of 3D chromatin conformation in
422 mediating disease risk.

423 Another key finding of our study is that regions which vary in 3D chromatin conformation
424 across individuals also tend to vary in measures of transcriptional and regulatory activity. This
425 supports the existence of shared mechanisms that underlie variation in 3D chromatin
426 conformation, transcription, and epigenomic properties. We suspect that no single mechanism
427 or causal hierarchy applies to all regions of the genome with variation in one or more of these
428 properties. However, in at least some cases, this shared mechanism is likely genetic. This
429 raises the question of whether 3D chromatin QTLs are fundamentally the same as QTLs
430 previously described for other molecular phenotypes (e.g. eQTLs, dsQTLs, histoneQTLs;
431 collectively referred to below as “molQTLs”), or represent a separate set of QTLs not detectable
432 with other methods. This question is difficult to answer in the present study for two main
433 reasons: 1) Our power is limited and thus we cannot say with confidence that a given SNP is *not*
434 a 3D chromatin QTL. Many molQTL studies also have limited power and are thus prone to type
435 II error. 2) Our QTL searches, like most molQTL studies, are not truly genome-wide because
436 subsets of testable regions and testable SNPs are preselected to focus the search space.
437 These selection criteria can differ widely between studies, making direct QTL-to-QTL
438 comparisons challenging. The observation of genotype dependent epigenetic signal at 3D
439 chromatin QTLs suggest that at least some 3D chromatin QTLs could also be detected as other
440 types of molQTLs if those studies had sufficient statistical power. However, the limited overlap
441 between 3D chromatin QTLs and published molQTLs (even when considering SNPs with only a

442 nominal level of significance) points to a lack of power in current studies, and suggests further
443 that the QTLs with largest effects on 3D chromatin conformation are not necessarily the same
444 as those with large effects on other molecular phenotypes, and vice versa. Therefore, it is likely
445 that QTL studies directed toward different types of molecular phenotypes (including 3D
446 chromatin features) are likely to be complimentary rather than redundant.

447 Future studies with higher resolution Hi-C data and larger sample sizes will be important
448 to identify functional variants modulating 3D chromatin conformation, and to further dissect the
449 mechanistic relationships between genetics, 3D chromatin conformation, and other molecular
450 phenotypes. We anticipate that these studies will continue to reveal cases in which perturbation
451 of 3D chromatin conformation is a molecular mechanism through which disease-associated
452 genetic variants confer disease risk. The present study provides initial discoveries of genetic
453 influence on 3D chromatin conformation and an analytical framework and that we believe will
454 facilitate future efforts to unravel the molecular basis of genetic disease risk.

455

456 **AUTHOR CONTRIBUTIONS**

457

458 Study was conceived and overseen by B.R., M.H., K.H., J.S., K.G., and D.U.G. Hi-C
459 experiments performed by D.U.G. and A.S. Data analysis performed by Y.Q., M.H., K.F-B,
460 (variable matrix cells, C-QTLs, related contact matrix analyses, power calculations), A.N., Y.L.,
461 J.C., and D.U.G. FISH experiments performed by T.L. Manuscript written by D.U.G., Y.Q., M.H.,
462 K.F-B., with input from all authors.

463

464 **FIGURE LEGENDS**

465

466 **Figure 1. Biological variability in multiple aspects of 3D chromatin.** (a) Browser view to
467 illustrate the Hi-C-derived molecular phenotypes examined here: contact matrices, FIRE, DI,
468 INS, and PC1 (chr8:125,040,000-132,560,000; hg19). Only 4 individuals shown here for
469 illustrative purposes. The full set of individuals is shown in Supplemental Figure 1. (b) Boxplots
470 show correlation between biological replicates from the same cell line (Individuals = “within”, N =
471 20), and between replicates from different cell lines (Individuals = “between”, N = 760).
472 Statistical significance calculated by two-sided Wilcoxon rank sum test. See Supplemental
473 Figure 4a for schematic of shuffling strategy. These and all boxplots in this manuscript show
474 median as a horizontal line, interquartile range (IQR) as a box, and the most extreme value
475 within 1.5*IQR or -1.5*IQR as whiskers extending above or below the box, respectively. (c) The

476 Pearson correlation coefficient between quantile normalized Hi-C matrix replicates from the
477 same cell line or different cell lines is plotted as a function of genomic distance between anchor
478 bins. Distances between 0.1-1Mb are highlighted in the magnified sub-panel to the right. (d)
479 Significance of the difference between the “within” and “between” values in (c) was calculated at
480 multiple points along the distance-correlation curve by two-sided Wilcoxon rank sum test. Note
481 that the scale of contact distance here is linear. Yellow bars indicate significance exceeding a
482 nominal p-value of 0.05 (dotted line).

483

484 **Figure 2. Variable regions of 3D chromatin conformation.** (a) Example of a variable region
485 (chr15:93,040,000-100,560,000; hg19). Triangular heatmaps from top to bottom: Four Hi-C
486 contact heatmaps in red from individuals showing variable 3D chromatin architecture, a
487 heatmap in blue showing the degree of variation measured across LCLs, and a heatmap in blue
488 showing variable cells in the matrix at IHW-FDR < 0.1 (var=variable, ns=not significant).
489 Standard tracks from top to bottom, and zoomed in more closely on the region of interest
490 (chr15:95,482,152-98,025,591; hg19): BAC probes used for FISH experiment in panels 2d-e,
491 variable DI regions called using all 20 LCLs, variable DI regions called using just 11 YRI LCLs,
492 twelve DI tracks from four different individuals. For each individual, DI tracks are shown from
493 two biological replicates and from Hi-C data merged across both replicates. Note that two
494 individuals have strong upstream contact skew in the boxed regions (YRI-4, YRI-8), while the
495 other two individuals have weak or no upstream contact skew in that region (YRI-3, YRI-5). (b)
496 The number of testable bins and significantly variable regions for each 3D chromatin phenotype
497 examined here. (c) Significance of pairwise overlap between different sets of variable regions. P
498 values calculated by chi square test. Additional details in methods and Supplemental Figures 5
499 and 6. (d) Boxplots showing the distance between indicated probe sets in four different LCLs.
500 Probe labels same as in panel (a). P values calculated by two-sided Wilcoxon rank sum test.
501 Number of nuclei measured for each LCL and probe pair, from left to right, are: 140, 91, 111,
502 70, 128, 124, 219, 70. (e) Representative images of nuclei corresponding to panel (d). (f) Blue
503 line shows the fraction of variable matrix cells distributed across a range of interaction
504 distances. Black shows the fraction of all matrix cells distributed across the same range of
505 interaction distances. (g) Top panel shows the percentage of variable matrix cell anchor bins
506 that overlap variable DI, FIRE, INS, or PC1 regions, respectively. The shade of blue is scaled
507 with overlap percentage. Bottom panel shows the statistical significance of these overlaps as
508 calculated by chi square test, and plotted with same color scale as (c).

509

510 **Figure 3. Coordinated variation of the 3D genome, epigenome, and transcriptome.** (a)
511 Example of a variable region where 3D chromatin phenotypes are correlated with epigenomic
512 and transcriptomic phenotypes (chr6:126,280,000-131,280,000; hg19). Six triangular heatmaps
513 from top to bottom: Hi-C contact heatmaps from four individuals, variability matrix, and variable
514 cells in the matrix (var=variable, ns=not significant). Standard tracks below show 3D chromatin,
515 epigenomic, and transcriptomic properties from four individuals in zoomed in region
516 (chr6:127,680,918-129,416,097; hg19). All ChIP-seq and RNA-seq data in the figure from
517 Kasowski et al, 2013⁴². (b) Density plots show the distribution of Spearman correlation
518 coefficients at variable regions (see Figure 2b for numbers) between the epigenomic or
519 transcriptomic phenotype indicated in the top margin of panel and the 3D chromatin phenotype
520 indicated in the right margin of panel. Grey lines show the distributions from 100 random
521 permutations (selected randomly from the 10,000 permutations performed) in which the sample
522 labels were shuffled (see Supplemental Figure 7b). *** $p < 0.0001$ by permutation test as
523 described in methods section 4c, which applies to all observations in this panel except RNA-seq
524 at INS regions ($p = 0.0018$) and RNA-seq at FIRE regions ($p = 0.0096$). (c) Heatmap showing
525 Spearman correlation coefficients between PC1 and multiple epigenomic/transcriptomic
526 phenotypes, arranged by k-means clustering ($k = 4$). Tick marks to the right show boundaries
527 between clusters. Each row ($N = 518$) is one variable PC1 region, limited to the subset of variable
528 PC1 regions that contain RNA-seq signal and at least one peak in at least one individual for
529 each ChIP-seq target included here (H3K27ac, H3K4me1, H3K27me3, CTCF, Cohesin). (d)
530 Similar to (c), showing correlations with FIRE at $N = 132$ variable FIRE regions. (e) Similar to (c),
531 showing correlations with DI $N = 265$ variable DI regions. (f) Similar to (c), showing correlations
532 with INS at $N = 154$ variable INS regions.

533
534 **Figure 4. A genetic contribution to variations in 3D chromatin conformation.** (a) A graphic
535 representation of the CTCF Position Weight Matrix (PWM) is shown. Eight positions boxed by
536 dashed lines have probability > 0.75 for a single base. We refer to SNPs at these positions as
537 “motif disrupting SNPs”. Alleles matching the consensus base in the motif are labeled “strong
538 motif alleles (S)”, and alleles matching any other base are labeled “weak motif alleles (W)”. (b)
539 Boxplot shows the distribution of interaction frequencies at loops with exactly one anchor
540 containing a CTCF motif disrupting SNP ($N = 138$), separated according to genotype. For each
541 SNP, loop strengths are normalized to the mean value of the heterozygous genotype (WS).
542 There is significant linear relationship between normalized loop strength and genotype by linear
543 regression ($p = 7.6e-5$). (c) Aggregate contact map shows the average difference in interaction

544 frequency per loop between SS and SW genotypes (top; N=117 SNPs), and between SW and
545 WW genotypes (bottom; N=31 SNPs). The cross point of dotted lines indicates the 40Kb bin
546 containing the loop being evaluated. (d) Histogram shows the allelic imbalance in reads
547 connecting loop anchors on the S vs W haplotypes in WS heterozygotes (N=135 loops). The
548 mean percentage of reads on the S haplotypes is significantly larger than 0.5 ($p= 5.9e-4$ by one-
549 sided t-test). (e) Line plots show the genotype-dependent signal of FIRE-QTL, INS-QTL and DI-
550 QTL using 11 independent YRI individuals. Each plot show the indicated phenotype as lines
551 with light color, medium color and dark color representing average signal across LCLs with the
552 low signal genotype, medium signal genotype, and high signal genotype, respectively. For DI-
553 QTL, we split all 40Kb QTL bin into two groups, based on either upstream DI bias (upper panel)
554 or downstream DI bias (bottom panel). (f) For C-QTLs, an aggregate contact plot analogous to
555 panel c is used to show the average difference in BNBC corrected interaction frequency (“ Δ
556 $\log(\text{norm contacts})$ ”) between the high and medium contact genotypes (top; N=138
557 interactions), and between the genotypes medium and low genotypes (bottom; N=94
558 interactions). The cross point of dotted lines indicates the 40Kb test bin in question. (g) Boxplots
559 show the genotype-dependent signal at QTLs using additional 6 individuals as a validation set.
560 In each boxplot, three boxes with light color, median color and dark color represent the average
561 signal in the 40Kb QTL bin from individuals with the low signal genotype, medium signal
562 genotype, and high signal genotype, respectively. (h) Results of permutation test to evaluate the
563 statistical significance of results in (g). The solid vertical lines show the estimated linear
564 regression slope values obtained from the validation set (N=6 individuals). The grey curves
565 show the distributions of slope values obtained from 1,000 random permutations.
566 Corresponding bootstrap p values indicated in the upper left corner of each subpanel. (i) Line
567 plot shows the fraction of foreground SNPs with nominal significance in the background
568 association study (“nominal fraction”). Red dashed lines show the values for all SNPs in a given
569 background set, yellow diamonds show values for SNPs in a given foreground set, and open
570 triangles show values for SNPs tested in the foreground QTL search. Black circles and lines
571 indicate the median and middle 95% range, respectively, of 10,000 permutations in which SNPs
572 were selected from the “foreground, tested” set. The number to the right of each line indicates
573 the fraction of permutations with a value higher than observed for “foreground, QTL” set. (j) QQ
574 Plot shows FIRE-QTL search results, including all SNPs tested for FIRE association (black
575 points, N=128,137), and several subsets as follows: DI-QTL tested (light green, N=46,784), INS-
576 QTLs tested (light red; N=6,238), C-QTL tested (light blue; N=69,847), DI-QTLs (dark green,
577 N=152), INS-QTLs (dark red, N=60), C-QTLs (dark blue, N=53).

578

579 **Figure 5. Contribution of 3D chromatin QTLs to other molecular and organismal**

580 **phenotypes.** (a) Boxplots show signal for epigenetic phenotypes separated by genotype at
581 FIRE-QTLs (top row), C-QTLs (middle row), and INS-QTLs (bottom row). Epigenetic signals
582 averaged across all peaks in 40Kb bin. Linear regression p and beta values shown above each
583 plot. P-value <0.05 in bold, others in italics. Yellow boxes highlight relationships that are
584 significant by linear regression and in permutation testing as shown in (b). (b) Line plots shows
585 beta values of linear relationships between QTL genotypes as indicated to the left and
586 epigenetic phenotype indicated above each subpanel. Yellow diamonds show values for the
587 true QTLs sets as shown in (a) and Supplemental figure 11a. Black circles and lines indicate the
588 median and middle 95% range, respectively, of 1,000 permutations in which SNPs were
589 selected from the “foreground, tested” set. The number to the right of each line indicates the
590 fraction of permutations with abs(beta) higher than observed for the true QTL set. Yellow boxes
591 highlight values < 0.05. (c) Genome browser view (chr2:201,222,342-201,386,844; hg19)
592 showing examples of a DI-QTL (chr2:201333312) and FIRE-QTL (chr2:201254049). All signals
593 plotted as a function of DI-QTL genotype (L=Low DI, M=medium DI, H=High DI). Grey boxes
594 highlight region where epigenetic signals stratify by DI-QTL genotype. (d) Left subpanel shows
595 the enrichment values for 3D QTL SNPs with nominal significance in the indicated GWAS study
596 calculated as follows: (fraction of indicated 3D QTL SNPs with nominal significance in the
597 indicated GWAS) / (fraction of SNPs tested in the indicated 3D QTL search with nominal
598 significance in the indicated GWAS). Asterisks mark values with $P < 0.05$ by chi-square test
599 (middle panel), and permutation test (right panel). Right panel shows the proportion of 1000
600 random subsets selected from the tested SNPs with enrichment values higher than the
601 indicated true QTL set. Dotted lines mark $p = 0.05$. (e) QQ plot shows the results of UC GWAS
602 with all tested SNPs shows as black points, and two subsets as follows: SNPs also tested in our
603 INS-QTL search (light red), and SNP called as INS-QTLs or in perfect LD with INS-QTLs in the
604 same 40Kb bin (dark red). (f) QQ plot shows the results of IBD GWAS with all tested SNPs
605 shows as black points, and two subsets as follows: SNPs also tested in our FIRE-QTL search
606 (light green), and SNP called as FIRE-QTLs or in perfect LD with FIRE-QTLs in the same 40Kb
607 bin (dark green).

608

609 **SUPPLEMENTAL FIGURE LEGENDS**

610

611 **Supplemental Figure 1. Hi-C derived molecular phenotypes measured across 20 LCLs.** (a)
612 Hi-C contact matrices show for all 20 LCLs. For comparison, we also show data from H1 human
613 embryonic stem cells (H1-ES), and 4 lineages derived from H1 by in vitro differentiation²¹ (H1-
614 ME = Mesendodermal cells, H1-NPC = Neuronal precursor cells, H1-TR = Trophoblast-like cell,
615 MSC = Mesenchymal stem cells). These H1-derived cell line represent different cell types from
616 the same individual (i.e. same genetic background). The region shown here is the same as
617 Figure 1a (chr8:125,040,000-132,560,000; hg19). (b) Same region as above, but showing PC1
618 and FIRE values. ChIP-seq data for several histone modifications, CTCF, and Cohesin subunit
619 SA1 are shown for one LCL (YRI-13, GM19240) as a reference for the epigenomic landscape⁴².
620 (c) Same region as above, but showing DI and INS values. (d) Bar plots show the percentage of
621 super-enhancers (left) or typical enhancers (right) in GM12878⁵⁹ that overlap with 6,980 LCL
622 FIRE bins (called as FIRE in at least one individual in our dataset) and 6,980 random 40kb bins.
623 (e) Biological Process Gene Ontology terms associated with genes proximate to FIRE regions
624 as defined by GREAT⁶⁰.

625
626 **Supplemental Figure 2. FIRE measures density of local interactions.** Illustrative example
627 showing that overall density of Hi-C reads (all reads irrespective of location of interacting
628 partner, all *cis* interactions, or all *trans* interactions) is highly consistent across the genome.
629 However, interactions between partners separated by 15-200kb (“FIRE” distance) show
630 enrichment in regions of the genome with marks of regulatory and transcription activity
631 (H3K4me1, H3K4me3, H3K36me3 from ENCODE for CEU-1 / GM12878 shown for
632 reference)⁶¹. ChromHMM functional annotations for CEU-1 / GM12878 are also shown⁶². pr-a =
633 active promoter, pr-w = weak promoter, pr-i/p = inactive/poised promoter, en-s = strong
634 enhancer, en-w/p = weak/poised, ins = insulator, tr-el/tr = transcriptional elongation or
635 transcriptional transition, tr-w = weak transcribed, rep-p = polycomb-repressed, het/rep/cnv =
636 heterochromatin, low signal, repetitive, or copy number variation. Top panel (a) shows the long
637 arm of chr14 (chr14:24,406,737-104,693,368; hg19). Bottom panel (b) is a zoomed-in view of
638 region boxed by dotted lines above (chr14:58,000,000-63,500,000; hg19).

639
640 **Supplemental Figure 3. Aggregate looping interactions in each sample.** Aggregate plots
641 show the interaction frequencies at GM12878 HiCCUPS loops from Rao et al 2014 in each
642 sample examined here. All autosomal *cis* loops with anchor bins separated by more than 40kb
643 are included here (N=8,893). The middle bin represents the interaction frequency between two
644 40kb bins containing loop anchor bins. The full submatrix extends 10 bins (400kb) upstream and

645 downstream of the interaction bin (x and y axis). The color scale indicates the average
646 interaction frequency per loop from Hi-C contact matrices.

647

648 **Supplemental Figure 4. 3D chromatin variation among 20 LCLs and H1-derived lineages.**

649 (a) Graphical representation of the shuffling scheme used to assess biological variability in
650 Figure 1b-d, and here in panels b-e. (b)-(e) Boxplots show Pearson correlation coefficient
651 between biological replicates from the same cell line (Replicates = “True”), and between
652 replicates from difference cell lines (Replicates = “Shuff”; short for “shuffled”). The set of cell
653 lines considered is indicated below each box. LCL = 20 LCL cell lines (40 replicates); H1 = H1-
654 ES and the four derived lineages H1-ME, H1-NPC, H1-TB, H1-MSK (5 cell lines, 10 replicates);
655 H1+LCL = 20 LCLs and 5 H1-derived lines considered together (50 replicates). All phenotypes
656 examined here (DI, PC1, INS, or FIRE) show a signature of cell-type specificity whereby they
657 are more similar across individuals when looking at the same (or highly similar) cell types (i.e.
658 LCLs), relative to comparing across cell types within an individual (i.e. same genetic
659 background, H1s). Statistical significance calculated by two-sided Wilcoxon rank sum test. (f)
660 Dendrograms from hierarchical clustering of 40 Hi-C replicates based on one of four Hi-C-
661 derived phenotypes, as indicated above each dendrogram (DI, PC1, INS, or FIRE). In the most
662 of cases, replicates from the same cell lines cluster together. (g) Principal Component Analysis
663 of 20 LCLs using one of four Hi-C-derived phenotypes, as indicated above each plot. Each
664 population in our study is represented by a different color as indicated in the color key to the
665 right. Note that the LCLs tend to separate by population in each plot.

666

667 **Supplemental Figure 5. Characterization of variable regions of 3D chromatin**

668 **conformation.** (a) Same region as in Figure 2a (chr15:94,280,000-99,280,000), but showing
669 reproducible variation in PC1, and full square matrices for contact matrix variability as opposed
670 to the half-matrices shown in 2a. FISH probes used in Figure 2 are represented as grey boxes
671 to the top and right of square heatmaps. (b) Similar to Figure 2b, but with additional data
672 columns. The “merged regions” column shows the number of regions after merging variable
673 bins that are immediately adjacent to each other. The empirical false positives and FDR
674 columns show the number and percentage of false positive variable regions detected in 1000
675 permutations with shuffled labels. (c) Venn diagrams showing the overlap of variable regions
676 identified using either all 20 LCLs (“LCL20”) or only the 11 unrelated YRI LCLs (“YRI11”). Only
677 bins tested for both LCL20 and YRI11 are included here. (d) Venn diagrams showing the
678 number of variable bins for each phenotype or combination of phenotypes. The diagram on the

679 left includes bins that were testable for at least one of the four phenotypes. The diagram on the
680 right only includes bins that were testable for all four phenotypes. (e) Mosaic plots show the
681 significance of overlaps between variable regions in a pairwise fashion. P-values calculated by
682 two-sided chi square test.

683

684 **Supplemental Figure 6. Additional characterization of variable regions of 3D chromatin**
685 **conformation.** (a) Same underlying FISH data as in Figure 2e, but here comparing the distance
686 between U and C probes to the distance between C and D probes within the same LCL. P-
687 values calculated by two-sided Wilcoxon rank sum test. (b) As in Figure 2d, blue line shows the
688 fraction of variable matrix cells distributed across a range of interaction distances. Black line
689 shows the fraction of all matrix cells distributed across the same range of interaction distances.
690 (c) Mosaic plots show the significance of overlap between variable regions and anchor bins of
691 variable matrix cells. P-values calculated by two-sided chi square test.

692

693 **Supplemental Figure 7. Coordinated variation between 3D chromatin conformation and**
694 **multiple molecular phenotypes.** (a) Same region as in Figure 3a (chr6:126,280,000-
695 131,280,000; hg19), but showing additional individuals and additional data types as indicated.
696 (b) Representation of permutation scheme used to calculate P-values in panels c as well as in
697 Figure 3b and Supplemental Figures 8 and 9. (c) Density plots in the top left quadrant show
698 Spearman correlation coefficients (SCC) between PC1 and molecular phenotypes as indicated
699 in the top margin of panel. Density plots in the top right quadrant show the same underlying data
700 but plotted as absolute SCC to highlight the shift of real correlations (red lines) towards one,
701 relative to the permuted values (grey). We show both SCC and abs(SCC) because SCC is
702 useful to observe bias toward positive or negative correlations, but the bootstrap p -values are
703 calculated based on the abs(SCC) to reflect correlations that are more extreme (positive or
704 negative) than expected based on approximation of the null hypothesis. For the top left and top
705 right quadrants, the ChIP-seq and RNA-seq data in the figure from Kasowski *et al*, 2013⁴²
706 (same as in Figure 3b), with the addition of histone modifications H3K4me1, H3K4me3, and
707 H3K36me3, as well as DNase-seq data from Degner *et al*, 2012³⁸. ChIP-seq, RNA-seq, and
708 DNase-seq plots include eight, seven, and eleven individuals, respectively (i.e. all individuals for
709 which both Hi-C and the data type in question are available). Bottom left quadrant shows SCC
710 like above, but using variable regions called in only 11 individuals. Bottom right quadrant shows
711 SCC using ChIP-seq data from McVicker *et al*, 2012⁴⁰. These plots include ten individuals for

712 which ChIP-seq and Hi-C data are available. In all cases, p -values are calculated as the number
713 of permutations with a mean absolute SCC greater than the observed values.

714

715 **Supplemental Figure 8. Correlations between DI, INS and multiple molecular phenotypes.**

716 Similar schema to Figure 7c, but focusing on DI in (a), and INS in (b).

717

718 **Supplemental Figure 9. Correlations between FIRE, interaction frequency and multiple**
719 **molecular phenotypes.** Similar schema to Figure 7c, but focusing on FIRE in (a), and contact

720 frequency (examining the anchor bins of variable matrix cells) in (b).

721

722 **Supplemental Figure 10. 3D chromatin QTLs.** (a) QQ plots for each QTL search. In each QQ

723 plot, the X-axis is the $-\log_{10}$ theoretical quantiles calculated from the uniform distribution. The Y-

724 axis is the $-\log_{10}$ p -value calculated from linear mixed effects model for each type of QTL

725 search. The grey area represents the 5% - 95% confidence bands based on Beta distribution

726 $\text{Beta}(i, M-i+1)$, where i is the i th order statistics and M is the total number of tested SNPs. (b)

727 Genotype trend for bins with positive DI (left), negative DI (right), and all QTLs (right). A

728 Simpson's paradox is observed when all bins are considered together. (c) Number of direct

729 overlaps between QTL sets. (d) Similar schema to Figure 4e, but showing FIRE, INS, DI score

730 (indicated on the Y axis) as a function of genotype for each QTL set as indicated above each

731 column. Grey boxes highlight the cases plotted in Figure 4e where the signal type and QTL set

732 are the same.

733

734 **Supplemental Figure 11. Influence of 3D chromatin QTLs on epigenomic and disease**

735 **phenotypes.** (a) Similar schema to Figure 5a, but showing DI-QTLs in positive bins (top) and

736 negative DI bins (bottom). (b) Left subpanel shows the enrichment for 3D QTL SNPs with

737 nominal significance in the indicated epigenetic or eQTL study calculated as follows: (fraction of

738 indicated 3D QTL SNPs with nominal significance in the indicated molQTL study) / (fraction of

739 SNPs tested in the indicated 3D QTL search with nominal significance in the indicated molQTL

740 study). Asterisks mark values with $p < 0.05$ by chi-square test (middle panel), and permutation

741 test (right panel). Right panel shows the proportion of 1,000 random subsets selected from the

742 tested SNPs with enrichment values higher than the indicated true QTL set. Dotted lines mark

743 $p = 0.05$. (c) QQ plot shows the results of H3K4me1 QTL search from Grubert et al., with all

744 tested SNPs shown as black points, and two subsets as follows: SNPs also tested in our C-QTL

745 search (light blue), and SNP called as C-QTLs or in perfect LD with C-QTLs in the same 40Kb
746 bin (dark blue).

747

748 **TABLES**

749

750 Supplemental Table 1. LCLs included in the study.

751 Supplemental Table 2. Public datasets re-analyzed in this study.

752 Supplemental Table 3. Hi-C mapping statistics.

753 Supplemental Table 4. Regions showing evidence of biological variability in 3D chromatin
754 conformation.

755 Supplemental Table 5. Matrix cells showing evidence of biological variability.

756 Supplemental Table 6. Summary of phasing results.

757 Supplemental Table 7. Power calculations.

758 Supplemental Table 8. 3D chromatin conformation QTLs.

759 Supplemental Table 9. Overlaps between 3D chromatin QTL and GWAS catalog.

760

761 **METHODS**

762

763 **1. Hi-C data generation.** Hi-C was performed as previously described¹³. We note that all Hi-C
764 experiments were performed using a “dilution” HindIII protocol, rather than the newer “in situ”
765 version of the protocol, for consistency because data generation began before the invention of
766 in situ Hi-C. In addition, the resolution of 40kb used here for most analysis was determined
767 primarily by sequencing depth rather than choice of a restriction enzyme. Thus, even if a 4-
768 cutter like MboI had been used, the prohibitive cost of sequencing would have prevented us
769 taking advantage of the additional possible resolution.

770

771 **2. Hi-C data processing.**

772 **2.a. Alignment with WASP.** Read ends were aligned to the hg19 reference genome using
773 BWA-MEM⁶³ v0.7.8 as single-end reads with the following parameters: -L 13,13. We used the
774 WASP pipeline^{40,45} to control for potential allelic mapping biases, which some modifications to
775 account for unique aspects of Hi-C data. BWA-MEM can produce split alignments where
776 different parts of a read are aligned to different parts of the genome. This is critical for Hi-C data,
777 because a read can span a Hi-C ligation junction between two interacting fragments. In the case
778 of a split alignment, BWA-MEM will mark the higher-scoring alignment as the primary alignment.

779 For Hi-C data this is not ideal – we want the five-prime-most alignment (before the ligation
780 junction) to be the primary alignment. To account for this, we further processed the alignments
781 from BWA-MEM to select the five-prime-most alignment in cases where one read was split.
782 Reads without an alignment to the five-prime end of the read were filtered out, as were
783 alignments with low mapping quality (<10). The WASP pipeline was then used to generate
784 alternative reads by flipping the allele in reads overlapping SNPs, and these reads were then
785 realigned using the same pipeline. As input to WASP, we included all SNPs and indels present
786 in the PUR individuals in our set (HG00731, 732, 733), CHS individuals in 1000 genomes (we
787 included all CHS to account for the fact that no 1000 genomes genotype calls were available for
788 HG00514), YRI individuals in 1000 genomes (we included all YRI individuals to account for the
789 fact that no 1000 genomes genotype calls were available for GM19193), and the H1 cell line²¹
790 (to facilitate uniform processing and comparisons between LCLs and H1-derived datasets).
791 After alignment of the alternative reads, alignment of the original reads and alternative reads
792 were compared by WASP, and only the original reads for which all alternative reads aligned at
793 the same location with same CIGAR string were kept. Reads overlapping indels were removed.
794 Reads were then re-paired, and only pairs in which both reads survived this filtering were kept.
795 PCR duplicates were removed using Picard tool (<http://broadinstitute.github.io/picard/>) with
796 default parameters. To ensure that our adapted WASP pipeline removed allelic mapping biases
797 effectively, we simulated all possible 100bp single end reads spanning SNPs in our LCLs and
798 aligned them back to the genome using our adapted WASP pipeline. We found no SNPs which
799 depart from 50/50 mapping ration between reference and alternative allele in these simulations.

800 We also took steps to remove any potential artifacts due to HindIII polymorphisms. Hi-C
801 data was obtained by cutting the genome with HindIII, so we reasoned that SNPs or indels that
802 disrupt existing HindIII sites or create novel HindIII sites could lead to differential cutting of two
803 alleles and thus the appearance of differential contact frequency. To mitigate these potential
804 artifacts, we identified all HindIII sites that would be disrupted or created by genetic variants
805 present in our samples, and removed all reads within 1Kb of these polymorphisms in all
806 individuals.

807 **2.b. Contact Matrix Calculations.** Matrices were generated and normalized as previously
808 described^{21,64}. Briefly, intra-chromosomal read pairs were divided into 40Kb bin pairs based on
809 five prime positions. The number of read pairs connecting each pair of 40Kb bins were tallied to
810 produce contact matrices for each chromosome. Raw counts in the contact matrices were then
811 normalized using HiCNorm⁶⁴ to correct for known sources of bias in Hi-C contact matrices (GC
812 content, mappability, fragment length). Bins that are unmappable (effective fragment length, GC

813 content or mappability is 0) were assigned NA values. These normalized matrices were further
814 quantile normalized across samples to account for differing read depths and mitigate potential
815 batch effects. One such quantile normalized matrix was generated for each chromosome in
816 each replicate, as well as in each sample (replicates pooled together). We eliminated
817 chromosomes X and Y from all downstream analyses due to the gender differences between
818 our samples.

819 **2.c. PC1 Score.** PC1 scores were computed using methods defined previously¹³. Briefly,
820 quantile normalized matrices for each chromosome were transformed to Observed/Expected
821 (O/E) matrices by dividing each entry in the matrix by the expected contact frequency between
822 regions in that matrix at a given genomic distance. For a given matrix, the expected contact
823 frequencies were computed by averaging contact frequencies at the same distance in that each
824 matrix. The O/E matrices were further transformed to Pearson correlation matrices by the “cor”
825 function in R and eigen vectors (principal components) were computed using the “cov” function
826 in R. Generally, the first eigenvector (“PC1”) reflects A/B compartmentalization. However, for
827 some chromosomes we have seen that the second eigenvector sometimes reflects
828 compartmentalization, while the first eigenvector reflects other features like the two
829 chromosome arms. To systematically account for this effect, we examined the first three
830 eigenvectors for each chromosome in each replicate by correlating them with the gene density
831 (compartmentalization is correlated with gene density, while other properties like chromosome
832 arms generally are not). We required that PC1 show the highest correlation with gene density
833 among the first three eigenvectors in every replicate. If this was not the case for a given
834 chromosome, we eliminated that chromosome from all downstream analyses in all individuals to
835 be conservative. Six chromosomes were eliminated in this way: chr1, chr9, chr14, chr19, chr21
836 and chr22. For the chromosomes that passed this filter, the sign of the first eigenvector (which is
837 arbitrary) was adjusted such that the correlation between PC1 and gene density is positive, and
838 this positive PC1 values correspond to compartment A. Finally, PC1 tracks were manually
839 inspected to ensure that they are consistent with expected checkerboard patterns of
840 compartmentalization.

841 **2.d. Directionality Index.** Directionality Index was computed as previously described¹⁶. Briefly,
842 upstream and downstream contacts within 2Mb window for each 40Kb bin were counted, and
843 chi-square statistics were calculated under equal assumption. The sign of the chi-square
844 statistics was adjusted such that positive values represent upstream biases. For some bins,
845 there are more than five NA bins within 2Mb window and DI for those bins are not calculated. As

846 noted in the main text, we made a slight variation of these DI scores for the QTL searches in
847 which DI was recalculated using a window size of 200Kb to capture more local features.

848 **2.e. Insulation Score.** Insulation scores were computed as previously described¹⁶ with some
849 adjustments. Briefly, contacts linking upstream and downstream 400Kb windows for each 40Kb
850 bin were calculated in the O/E matrices instead of raw matrices. We further divided the contact
851 frequency by the average of upstream and downstream 400Kb windows, to account for
852 differences in contact density across the chromosome. The Insulation Scores were then ranged
853 from 0 to 1, representing absolute insulation and no insulation respectively. Insulation scores for
854 bins, for which more than 50% cells in the 400Kb window as NA values, were not computed. For
855 the QTL search, we also calculated insulation scores using 200Kb window.

856 **2.f. TADs Calling.** TADs were called using the same approach as described previously¹⁶. DI
857 values for each 40Kb bins were used to build a Hidden Markov Model and predict the probability
858 being upstream bias, no bias, and downstream bias. Regions switching from upstream bias to
859 downstream bias were called as boundaries.

860 **2.g. FIRE.** We first calculated FIRE score for each of 20 individuals, as described in our
861 previous study¹⁹. Specifically, we mapped the raw reads to the reference genome hg19 as
862 described above. Next, we removed all intra-chromosomal reads within 15Kb, and created 40Kb
863 raw Hi-C contact matrix for each individual for each autosome. For each 40Kb bin, we
864 calculated the total number of intra-chromosomal reads in the distance range of 15-200Kb. We
865 then filtered bins as follows, starting from 72,036 autosomal 40Kb bins: First, we removed 40Kb
866 bins with zero effective fragment size, zero GC content, or zero mappability score⁶⁴. Next, we
867 filtered out 40Kb bins within 200Kb of the bins removed in the previous step. We further filtered
868 out 40Kb bins overlapping with the chr6 MHC region (chr6:28,477,797-33,448,354; hg19), which
869 has extremely high SNP density that can make it difficult to correct for allelic mapping artifacts.
870 This left 64,222 40Kb bins for downstream analysis. Next, we applied HiCNormCis¹⁹ to remove
871 systematic biases from local genomic features, including effective fragment size, GC content
872 and mappability. The normalized total number of cis intra-chromosomal reads is defined as
873 FIRE score. We further performed quantile normalization across multiple individuals using R
874 package “preprocessCore”. The final FIRE score is log transformation $\log_2(\text{FIRE score} + 1)$ and
875 converted into a Z-score to create a mean of 0 and standard deviation of one. To identify
876 significant FIRE bins in each individual, we used one-sided P-value < 0.05 . Ultimately, merging
877 across all individuals, we identified 6,980 40Kb bins which are FIRE bin in at least one of 12 YRI
878 individuals. Consistent with our previous findings¹⁹, we observed significant enrichment of
879 GM12878 typical enhancers and super enhancers among these 6,980 40Kb FIRE bins

880 (Supplemental Figure 1d). GREAT analysis⁶⁰ further showed immune-related biological
881 pathways and disease ontologies are enriched in these 6,980 40Kb FIRE bins (Supplemental
882 Figure 1e).

883

884 **3. Comparison of intra-individual vs inter-individual variation.** To estimate variability
885 between replicates, we computed Pearson correlation coefficient for all pairs of biological
886 replicates for each score (DI, INS, FIRE and PC1). The pairs then can be divided into two
887 groups based on whether they are from the same individuals as illustrated in Supplemental
888 Figure 4c. We then tested if the distribution of Pearson correlation coefficients were different
889 comparing two groups. Similar analysis was performed for contact matrices. For contact
890 matrices, we calculated Pearson correlation coefficient for each distance and each chromosome
891 separately as shown in Figure 1c.

892

893 **4. Variable regions.**

894 **4.a. limma test for variable bins.** To test regions that are variable across genomes, we applied
895 limma⁴⁷ with default parameters. First, values for each 40Kb bin in hg19 reference genome were
896 calculated for each metrics tested (DI, FIRE, INS, PC1) as described above. DI, PC1, and INS
897 scores were calculated based on contact matrices quantile normalized across 40 replicates.
898 FIRE scores were calculated based on raw counts using HiCNormCis¹⁹ and then quantile
899 normalized across 40 replicates. Second, we filtered out bins that are not testable. Specifically,
900 FIRE scores were only tested for bins that are FIRE regions (p -value < 0.05) in any of 40
901 replicates. DI scores were only tested for bins where strong biases are observed ($\text{abs}(\text{DI}) >$
902 10.82757 , which correspond to Chi-squared test p -value 0.001) in any of 40 replicates. INS
903 scores were only tested for bins where strong insulation is observed (z -score transformed INS
904 score < -1) in any of 40 replicates. No filterers were performed for PC1 scores. Third, we filtered
905 out any bins that overlapping large SVs (> 10,000 bp) to avoid effect caused by SVs.
906 Specifically, for FIRE, INS, and DI scores, bins that are within 200Kb, 400Kb, and 2Mb
907 respectively upstream or downstream of large SVs were removed. For PC1 scores, bins
908 overlapping large SVs were removed. Lastly, we applied limma standard model with individual
909 as a fixed factor and eBayes correction. To estimate empirical false positive rate (FDR), we
910 bootstrapped replicates to calculate the number of false positives in random background.
911 Briefly, we random selected 40 or 22 replicates with replacement for LCL20 and YRI11
912 respectively, and identified variable regions as mentioned above. We performed 1,000

913 permutations and calculated empirical FDR as the average positive hits in 1,000 permutations
914 divided by number of hits in real data.

915 **4.b. Normalizing Hi-C contact matrices using BNBC normalization.** To directly compare
916 individual Hi-C contact matrix cells across samples, we sought to remove unwanted per-cell
917 variation owing to date of processing or other unknown ‘batch’ effects. To this end we
918 developed Bandwise Normalization and Batch effect Correction (BNBC), described and
919 evaluated in a separate manuscript (preprint on bioRxiv
920 <https://www.biorxiv.org/content/10.1101/214361v1>). A brief description follows. For each
921 chromosome and for each strata of distance between loci (a matrix “band”, hence the term
922 “bandwise”), we correct for unwanted variation by taking the log counts-per-million-transformed
923 values of all samples and generating a matrix whose entries are the observations for that
924 chromosome’s matrix band across all samples (columns indexes samples and rows indexes
925 contact matrix cells with anchor bins separated by a fixed distance). We then quantile normalize
926 this matrix and regress out the impact of known batches (here, date of processing) using
927 ComBat⁶⁵ (specifically we correct both mean and variance). This procedure essentially
928 conditions on genomic distance. We correct the majority of each contact matrix for each
929 chromosome for each sample: we correct all but the 8 most distal matrix bands, for which we
930 set all values to 0. The choice of the last 8 bands is empirical and reflects the small number of
931 observations in each band matrix. The procedure is implemented in the bnb package available
932 through Bioconductor (<http://www.bioconductor.org/packages/bnb>). Correction of contact
933 matrices was performed on replicate-level data using the following LCLs: GM18486 (YRI-1),
934 GM18505 (YRI-2), GM18507 (YRI-3), GM18508 (YRI-4), GM18516 (YRI-5), GM18522 (YRI-6),
935 GM19099 (YRI-7), GM19141 (YRI-8), GM19204 (YRI-10), GM19238 (YRI-11), GM19239 (YRI-
936 12), GM19240 (YRI-13), HG00731 (PUR-1), HG00732 (PUR-2), HG00512 (CHS-1), HG00513
937 (CHS-2). We note that NA19239 (YRI-12) replicate 1 and NA19240 (YRI-13) replicate 2 were
938 excluded because the BNBC algorithm requires multiple samples from a given experimental
939 batch to estimate batch effect parameters.

940 **4.c. Identifying biological variability in Hi-C contact matrices.** To identify contacts with
941 significant levels of between-individual variability we employed the following procedure, which
942 mimics the analysis for INS, DI, FIRE and PC1, on contact matrices normalized by BNBC (see
943 section 4b). For each contact matrix cell (representing loci separated by less than 28 Mb, this is
944 a subset of the matrix cells normalized by BNBC) we used a linear model with individual
945 modeled as a fixed factor, note we have 2 growth replicates for almost every individual. We
946 used a parametric likelihood ratio test (equivalent to an F-test) to test whether there was

947 significant between-individual variation. We used the IHW framework⁴⁹ with the distance
948 between anchor bins as informative covariate, to increase power and estimate false discovery
949 rate. We used a FDR of 10% as significance threshold, resulting in 115,817 contact matrix cells
950 with significant biological variability across the autosomes. To estimate effect size (depicted in
951 Figures 2a, 3a and Supplementary Figure S5) we used a linear mixed effect model with
952 individual as random effect, to decompose the variance into between-individual variability
953 (biological) and within-individual variability (technical). As the measure of biological variability in
954 these figures, we used the estimated biological variance. For this analysis, all 16 samples we
955 normalized using BNBC were used.

956 **4.d. Correlation with other datasets.** To examine correlation between 3D genome
957 organization and other genome features, we reidentified variable regions with the same pipeline
958 mentioned above using only individuals of which data is available for other genome features, and
959 then computed Spearman correlation coefficient between 3D genome metrics (DI, INS, PC1,
960 and FIRE) and other genome features (RNA-seq, ChIP-seq, and DNase-seq) for each 40Kb bin
961 that is variable. Signals for each 40Kb bins were calculated by averaging signals for the bin.
962 Specifically, signals for ChIP-seq were the average signal of all peaks with in the bin, signals for
963 RNA-seq were the average FPKM of all genes in the bin, and DNase signals were simply
964 average signal for each base pair in the bin. In some cases, several consecutive bins were
965 identified as variable. We only kept the bin with strongest signal for other genome features
966 among consecutive bins. To generate random backgrounds, we permuted individual labels for
967 the same set of bins and recomputed Spearman correlation coefficient. 10,000 such
968 permutations were used to calculate the statistical significance of departure from the null
969 hypothesis in which the median value of true correlation values and permuted correlation
970 values are equal. Similar analysis was performed for variable matrix cells with the following
971 modifications. First, we used the variable matrix cells in the preceding section 4c. Second, to
972 correlate matrix-cell-level contacts with bin-level DNase and ChIP-seq signals, anchor bins of
973 variable matrix cells were used. Since each anchor bin may belong to more than one matrix
974 cells, we only used each bin once and selected the one with the highest Spearman correlation
975 coefficient. Exactly same approach was performed during permutation to ensure a fair
976 comparison.

977
978 **5. Phasing variants.** Phasing of variants was performed based on HaploSeq pipeline⁴³. Briefly,
979 1) Variants were filtered to keep only bi-allelic SNPs heterozygous in a given individual; 2)
980 Aligned Hi-C bam files were realigned and recalibrated using GATK 3.4.0⁶⁶ based on SNPs in

981 the individual; 3) Filtered SNPs and realigned bam files were then used as input to run
982 HAPCUT⁶⁷; 4) Results from HAPCUT were further filtered to keep only the largest haplotype
983 block and combined with homozygous alt SNPs as input for imputation using Beagle 4.0⁶⁸ using
984 1000 Genome Phase 3 data excluding individual to phase as reference panel; 5) Results from
985 Beagle were then combined with results of HAPCUT by removing conflicting phased SNPs. For
986 all auto chromosomes except 1 and 9 in 18 out of 20 individuals, we were able to obtain a single
987 haplotype block. For chromosome 1 and 9, two arms were phased separately because of large
988 heterochromatin region surrounding centromere. X chromosome was only phased for female
989 individuals. We excluded NA19193 and HG00514 from phasing because of the lack of available
990 high quality of genotypes. We evaluated accuracy of phasing in three probands in trios
991 (NA19240, NA12878, HG00733) and found phasing results are of very high accuracy
992 (~97.71%). Specifically, we calculated accuracy as percentage of correctly phased variants
993 among total phased variants. Only variants whose transmission from parents can be
994 ambiguously identified were used in calculation of accuracy where at least on parent is
995 homozygous. Detailed statistics for phasing are listed in Supplemental Table 6.

996 **6. CTCF motif variation and looping strength.** GM12878 loops and motif positions were
997 obtained from Rao *et al* 2014¹⁴
998 (GSE63525_GM12878_primary+replicate_HiCCUPS_looplist_with_motifs.txt.gz; N=9,448
999 HiCCUPS loops). We limited our analysis to autosomal *cis* loops in which a CTCF motif in one
1000 of the anchor regions overlaps a SNP (N=572). To evaluate the impact of motif disruption, we
1001 first identified eight “key” positions in the CTCF PWM (Jaspar MA0139.1)⁶⁹ in which a single
1002 base has higher than 0.75 probability. We refer to SNPs at these positions in motif occurrences
1003 with one allele matching the high-probability base as “motif disrupting SNPs”. We refer to alleles
1004 matching the consensus base in the motif as strong motif alleles (S), and alleles matching any
1005 other base as weak motif alleles (W). There are N=142 loops with a motif disrupting SNP in a
1006 convergently-oriented CTCF motif, which refer to below as testable loops. For each testable
1007 loop, we extracted the Hi-C interaction frequency in the loop bin from each LCL, and classified
1008 as either “WW”, “SW”, or “SS” depending on the individual’s genotype at the corresponding
1009 motif disrupting SNP. To enable aggregation of data across different SNPs, we set the mean
1010 “SW” interaction frequency for each SNP to 1 and normalized all values for that SNP
1011 accordingly. These values are plotted in Figure 4b. In addition, for each testable loop we
1012 extracted a submatrix including the loop bin as well as 15 bins upstream and 15 bins
1013 downstream. Submatrices with missing values were discarded. For each SNP, we calculated
1014 the mean submatrix for each genotype, and then subtracted submatrices to calculate the

1015 difference in each matrix cell per “W” allele (i.e. SS-SW and SW-WW). These differences were
1016 then averaged across all SNPs and plotted in Figure 4c. Submatrices with missing values were
1017 discarded. For the allelic analysis in S/W heterozygous individuals, we used chromosome-span
1018 phasing results (see methods section 4) to split the Hi-C reads from each chromosome in each
1019 LCL into two separate haplotypes. Specifically, we required at least one base pair overlap with
1020 phased heterozygous SNPs with high base calling score (>13) and high mapping quality (>20).
1021 Reads overlapping indels or containing SNPs from both haplotypes were not used.
1022 Approximately 7.89% of Hi-C reads covered a heterozygous variant and could thus be assigned
1023 to one of the two haplotypes. The accuracy of haplotype assignment was evaluated by fraction
1024 of homologous-trans (h-trans) read, which contain SNPs from both haplotypes. On average
1025 ~1% reads were h-trans, suggesting high quality of the assignment. For each testable loop, we
1026 defined 40Kb windows around the center of each loop anchor region and calculated the number
1027 of reads connecting these two anchor windows (“loop reads”) on each haplotype. For each
1028 heterozygous LCL, we then calculated the percentage of loop reads that occur on the haplotype
1029 containing the S allele at the motif disrupting SNPs anchor. We required that at least 10 total
1030 loop reads were present for a given loop in a given heterozygous LCL, leading to a total of 218
1031 data points from 105 different loops for inclusion in Figure 4d.

1032

1033 **7. Identification of QTLs.**

1034 **7.a. Testable bins.** To identify testable bins for FIRE-QTL, DI-QTL, INS-QTL and C-QTL
1035 searches, we began with 72,036 autosomal 40Kb bins based on reference genome hg19. We
1036 eliminated “unreliable” bins with effective length, GC content, or mappability equal to zero⁷⁰,
1037 resulting in 66,597 bins remaining. We further removed any 40Kb bins within 200Kb of an
1038 unreliable bin, resulting in 64,337 40Kb bins. We also removed bins covering the chr6 MHC
1039 locus (hg19: chr6:28,477,797-33,448,354, which is extremely polymorphic and may lead to
1040 complex mapping artifacts that are difficult to correct. To eliminate false signals in Hi-C data that
1041 could arise from large structural variations (SVs), we obtained SVs from the 1000 Genomes
1042 consortium³⁵ (ftp://ftp-
1043 trace.ncbi.nih.gov/1000genomes/ftp/phase3/integrated_sv_map/ALL.wgs.integrated_sv_map_v
1044 2.20130502.svs.genotypes.vcf.gz) and removed bins which overlap one or more structural
1045 variants previously annotated in these individuals (N=123,015 SVs), or within 200Kb of large
1046 structural variations (>10Kb, N=1,253 SVs). These filtering steps yielded a set of 51,511
1047 testable bins, which represent a common starting point for FIRE-QTL, DI-QTL, INS-QTL and C-
1048 QTL searches as described below.

1049 **7.b. Testable SNPs.** We began with a list of 15,765,667 variants among all 20 LCL individuals
1050 (Supplemental Table 3). We kept 14,177,284 variants among 11 unrelated YRI individuals,
1051 removed all indels, HindIII site polymorphisms, multi-allelic SNPs, and SNPs with minor allele
1052 frequency (MAF) < 5%. We also required that remaining SNPs were within the 51,511 testable
1053 bins described above, and that both alleles were present in at least 2 individuals in the
1054 discovery set individuals. (N=4,132,791 SNPs remaining). Finally, where multiple SNPs in the
1055 same bin were in perfect LD among 11 unrelated YRI individuals, we selected one with the
1056 smallest genomic position (to avoid the introduction of a random selection that would not be
1057 perfectly reproducible), ultimately yielding 1,304,404 potentially testable SNPs that served as a
1058 common input set to all QTL searches.

1059 **7.c. Power Calculations.** To explore the power of our approach and data, we performed a
1060 Monte Carlo-based power calculation. Specifically, we varied four variables: (1) the minor allele
1061 frequency of a variant; (2) the effect size of genotype (a fixed effect); (3) the variability between
1062 subjects (a random effect); (4) the variability of the residuals. For contact QTLs, we also varied
1063 the mean of the Hi-C contact frequency in question. For analyses reported, we fixed the number
1064 of replicates-per-subject to be 2 (consistent with our study design). We explored a variety of
1065 settings for these parameters to assess power as each variable changes (see Supplemental
1066 Table 7). Each setting tested was chosen to reflect the distribution of observed values in our
1067 real Hi-C data. For each configuration of parameters, we performed the following simulation: We
1068 simulated genotypes by randomly sampling a set of alleles (one allele per subject) from a
1069 binomial distribution parameterized by the number of subjects and the MAF; we repeated this
1070 process twice and create per-subject genotypes by adding the results of the sampling of alleles.
1071 We simulated per-subject random effects, and per-sample residuals. To obtain a given sample's
1072 simulated Hi-C contact matrix value, we added the mean Hi-C contact matrix value to that
1073 sample's simulated genotype (multiplied by the pre-specified effect size), the specific subject's
1074 random intercept and the sample's random residual. After performing this for all samples, we
1075 then fitted the same LMM model used in our QTL search. We repeated this simulation and
1076 model fitting process 1,000 times and computed power as the fraction of times the null
1077 hypothesis that the effect of genotype is equal to 0 is rejected at a nominal p-value of 0.05.

1078 **7.d. FIRE, DI, and INS QTL searches. 7.d.i. FIRE tested bins and SNPs.** We limited our FIRE
1079 QTL search to the subset of testable bins that were called as FIRE in at least one YRI LCL
1080 (N=5,822 FIRE test bins), and the subset of testable SNPs therein (N=128,137 FIRE test
1081 SNPs). **7.d.ii. INS tested bins and SNPs.** For the INS-QTL search, we examined 328,530 test
1082 SNPs with 12,976 variable INS bins (see methods section 4a). **7.d.iii. DI tested bins and**

1083 **SNPs.** For the DI-QTL search, we examined 181,950 test SNPs with 7,590 variable DI bins
1084 ((see methods section 4a). For the DI-QTL search, we further classified each DI bin based on
1085 which whether it showed stronger upstream or downstream bias, because we saw a Simpson's
1086 paradox when we considered them together (see discussion above, Supplemental Figure 10b).
1087 This was done as follows: for each bin we evaluated the DI score in each of 11 unrelated YRIs
1088 and identified the DI score among these individuals with the largest absolute value. We defined
1089 a bin as "upstream DI bias" if the DI score with the highest absolute value was positive, or
1090 "downstream DI bias", if the DI score with the highest absolute value was negative. Only
1091 37/7,590 bins (0.4%) had individuals with both positive and negative DI. **7.d.iv. LMM QTL**
1092 **searches.** For each test SNP, we identified the 40Kb bin it belongs to, and fitted a linear mixed
1093 effect model, using FIRE, DI (200Kb window; see section 2d), or INS score (200Kb window; see
1094 section 2e) in each biological replicate as the response variable and genotype of that testable
1095 SNP as the explanatory variable. Since two biological replicates from the same individuals are
1096 correlated included an individual-specific random effect to account for within-individual
1097 correlation. We used the R package "nlme" and R function "gls" to fit the linear mixed effect
1098 model. The quantile-quantile plots (QQplot) showed only minor genomic inflation (median p-
1099 value = 0.4821, lambda = 1.0864 for FIRE-QTLs; median p-value = 0.4864, lambda = 1.0649 for
1100 upstream-biased DI-QTLs; median p-value = 0.4828, lambda = 1.0826 for downstream-biased
1101 DI-QTLs; median p-value = 0.4865, lambda = 1.0646 for INS-QTLs). The linear mixed effect
1102 model identified 476, 315, 315, and 1,092 SNPs with false discovery rate (FDR) less than 0.20
1103 for FIRE, upstream-biased DI, downstream-biased DI and INS, respectively. When more than
1104 one SNP in the same bin was identified, we selected the SNP with lowest p-value among them
1105 to be included in the final QTL sets. After this filtering, we ended up with 387 candidate FIRE-
1106 QTLs, 268 candidate upstream-biased DI-QTLs, 277 downstream-biased DI-QTLs, and 911
1107 candidate INS-QTLs. As a control for each of these QTL searches, we randomly shuffled the
1108 score in question (i.e. FIRE, DI, or INS) among all 11 YRI individuals and performed QTL
1109 searches on this permuted data. In each of these tests, we found no SNPs associated with the
1110 permuted scores at FDR < 0.20.

1111 **7.e. C-QTL search.** To find QTLs affecting Hi-C contact strength we first identified 115,187 Hi-C
1112 contact matrix cells exhibiting substantial biological variability as described in section 4b, and
1113 constrained our QTL search to these cells. We then intersected these contact cells with
1114 1,304,404 testable SNPs by requiring a SNP to sit in one anchor bin of one of these variable
1115 matrix cells. We also filtered out matrix cells to ensure both anchor bins of the matrix cell are
1116 among 51,511 testable bins. In total, we obtained 3,109,039 tests involving 687,655 SNPs and

1117 54,880 matrix cells on all 22 autosomes. For each test, we used the BNBC normalized data
1118 described in section 4, but used only the 11 unrelated YRI individuals with genotypes available
1119 and fit a linear mixed effect model in which genotype is a fixed effect and subject is a random
1120 intercept. We then used “lmerTest” package in R to estimate p-values for the fixed effect of
1121 genotype⁷¹. We used the IHW framework⁴⁹ to estimate FDR, with the distance between anchor
1122 bins as an informative covariate, and call any matrix cell with $FDR < 0.2$ as significant. We
1123 further filtered significant tests by selecting the most significant SNPs per matrix cell and kept
1124 the leftmost SNPs among SNPs in perfect LD in two anchor bins of the matrix cell. After filtering,
1125 we ended up with 463 tests involving 345 SNPs and 463 matrix cells. To make the aggregate
1126 contact plots in Figure 4g, we recoded the genotypes based on the direction of effect such that
1127 0, 1, 2 refer to the genotypes containing 0, 1 or 2 alleles associated with the increased
1128 phenotype, respectively. Next, to avoid aggregating the same submatrix multiple times, we
1129 filtered by 1) selecting only the most significant matrix cell associated with each QTL, 2)
1130 selecting only the most significant QTL associated with each anchor bin (in some cases the
1131 same bin anchors multiple matrix cells associated with different QTL SNPs). This filtering left
1132 165 unique matrix cell QTL interactions for plotting. For each matrix cell, we then extracted a
1133 submatrix including 25 bins upstream and 25 bins downstream. Submatrices with missing
1134 values were discarded. For each QTL, we then calculated the mean submatrix values for each
1135 genotype, and then subtracted submatrices to calculate the difference in interaction frequency
1136 between the 1 and 0 genotypes, and between the 1 and 2 genotypes. These differences were
1137 then averaged across QTLs and plotted in Figure 4g

1138 **7.f. Validation of QTLs in additional individuals.** Our validation set included six unrelated
1139 individuals not included in the discovery set: NA12878, NA19240, HG00512, HG00513,
1140 HG00731 and HG00732. For each QTL, we collected the genotype among six additional
1141 individuals, and the corresponding FIRE, DI, or INS scores. Note that a small fraction of QTLs
1142 have missing genotypes in these six individuals (coded as “-1”), and these missing data points
1143 were eliminated from validation analysis. We examined the distributions of scores for each
1144 genotype. For each QTL type (i.e. FIRE, DI, or INS), we found that the same direction of effect
1145 observed in the discovery set is observed on average in the validation set. To assess the
1146 significance of this observation, we approximated the null expectation as follows. For FIRE-
1147 QTLs, for example, we started from all 128,137 FIRE test SNPs and 5,822 FIRE test bins. Note
1148 that in our discovery set, we identified 387 FIRE-QTLs, each in a different 40Kb bin. To create a
1149 random control SNP group, we first randomly selected 387 40Kb bins from all 5,822 FIRE test
1150 bins. Next, within each select bin, we randomly selected one SNP, and combined all these 387

1151 selected SNPs into a control SNP group. We then tested their SNP effect on the six additional
1152 individuals. We repeated such sampling with replacement 1,000 times, to create a null
1153 distribution of positive and negative SNP effect, respectively. We performed the same type of
1154 permutations for DI, INS. Similar analysis was performed for C-QTLs with a few modifications.
1155 First, we only used replicates from NA19420, HG00512, HG00513, HG00731 and HG00732 as
1156 explained in methods section 4. Second, 1,000 random permutations were performed by
1157 sampling matrix cells instead of bins. Third, we used values of biological replicates separately
1158 instead of as merged data because the BNBC normalization is performed at the level of
1159 replicates.

1160 **7.g. Examining epigenetic variation at FIRE, DI, INS, and C-QTLs.** To examine epigenetic
1161 variation at 3D genome QTLs, we re-analyzed DNase-seq data from 59 LCLs³⁸, histone
1162 modification ChIP-Seq data (H3K27ac, H3K4me1 and H3K4me3) for 65 LCLs³⁹, and CTCF
1163 ChIP-seq data from 11 LCLs⁵⁴. These data were re-mapped using the WASP pipeline to control
1164 for allelic mapping artifacts and calculating the signal in 40kb bins as described above in section
1165 3.b. We examined the effect of genotype at FIRE, DI, INS or C-QTLs on DNase-seq and ChIP-
1166 seq signal by linear regression. As a control, we randomly selected the matched number of
1167 SNPs with the same approach described in section 7.f and re-did such validation analysis. We
1168 repeated such random sample 1,000 times to create the empirical null distribution of no genetic
1169 effect. For C-QTLs, we used the sum of epigenetic features in two anchor bins to calculate
1170 correlation with contact frequency.

1171

1172 **8. Nominal fraction analyses.**

1173 **8.a. Comparing between 3D chromatin QTL types.** To compare between different 3D
1174 chromatin QTLs, we took the raw test results for each QTL set and projected other 3D QTLs
1175 into the test results. For example, in Figure 4j we selected subset of SNPs that are DI-QTLs and
1176 plotted them (dark green dots) using p-values from FIRE-QTLs along with all tested in the FIRE-
1177 QTL search (black dots). We also used all tested SNPs in the DI-QTL search (light green dots)
1178 as a control set. To assign significance to the overlap, we compared the fraction of SNPs with
1179 nominal significance ($p\text{-value} < 0.05$) in each set: 1) DI-QTL tested SNPs that were not significant
1180 QTLs, and 2) DI-QTLs. We calculated p-values for this comparison by Chi-square test. To rule
1181 out the effect of sampling bias when selecting a small number of SNPs, we also performed
1182 permutation. In each permutation, we randomly selected the same number of SNPs as the real
1183 QTL set (from the full set of tested SNPs) and calculated the fraction with nominal significance.
1184 We then computed bootstrap p-values using 10,000 such permutations under the null

1185 hypothesis that the fraction of nominal significance is the same between QTLs and random
1186 selected SNPs. For C-QTLs, one SNP may be tested against multiple matrix cells, so we only
1187 keep the most significant p-value for each SNP to avoid biases towards SNPs with multiple
1188 tests.

1189 **8.b. Comparing 3D chromatin QTLs to other molQTLs.** Similar approaches were used to
1190 assess overlap between 3D chromatin QTLs and other molQTLs. We obtained full test results
1191 (all tested SNPs with the p-values) from previous molQTL studies and projected 3D chromatin
1192 QTLs into those test results. We then calculated the fraction of nominal significance and used chi-
1193 square test to evaluate significance between 3D-QTLs and non-3D-QTLs. Similarly, we
1194 performed bootstrap to estimate significance empirically. One modification is that we extended
1195 our QTL sets by incorporating all SNPs in perfect LD with the same 40Kb bin because we may
1196 not use the same tagging SNP in our study as used in other studies. To ensure a fair
1197 comparison, we performed the same extension for the control sets of all tested SNPs.

1198 **8.c. Comparing 3D chromatin QTLs to GWAS.** Comparison with the GWAS results was
1199 performed in the same manner as described above in 8.b. for other molQTLs. Instead of test
1200 results for other molQTLs, we used summary statistics from previous GWAS.

1201

1202 **9. FISH.**

1203 **9.a. Cell preparation for FISH.** Approximately 100,000 cells were adhered to center of PDL-
1204 coated coverslips (Neuvitro, GG-22-15-PDL) by placing 100 μ L of cells at 1×10^6 cells/mL. Cells
1205 on coverslips were incubated for an hour at 37°C, carefully washed with PBS, and fixed with 4%
1206 paraformaldehyde in 1X PBS for 10 mins. PFA was quenched with 0.1 M Tris-Cl, pH 7.4 for 10
1207 mins, washed with PBS, and stored in 1X PBS at 4°C for up to 1 month.

1208 **9.b. BAC probe labeling and preparation.** All BAC clones were ordered from the BACPAC
1209 Resource Center at the Children's Hospital Oakland Research Institute: "U" probe is RP11-
1210 74P5, "C" probe is RP11-337N12, and "D" probe is RP11-248M23. BAC DNAs were labeled
1211 with either Chromatide Alexa Fluor 488-5 dUTP (Invitrogen, C-11397) or Alexa Fluor 647-aha-
1212 dUTP (Invitrogen, A32764) using nick-translation kit (Roche, 10976776001), and incubated in
1213 15°C for 4 hours. The nick-translation reaction was deactivated using 1 μ L of 0.5 M EDTA, pH
1214 8.0 and heated for 10 mins at 65°C. The probes were then purified using illustra ProbeQuant G-
1215 50 Micro Columns (GE Healthcare, 28903408) and eluted to a concentration of 20 ng/ μ L.
1216 Probes were mixed with Human Cot-1 DNA (Invitrogen, 15279011) and salmon sperm
1217 (Invitrogen, 15632011), and precipitated with 1/10th volume of 3M sodium acetate, pH 5.2 and
1218 2.5 volume of absolute ethanol for at least 2 hours at -20°C. Probes were then spun down,

1219 washed with cold 70% ethanol, resuspended in formamide and 40% dextran sulfate in 8X SSC,
1220 and incubated at 55°C.

1221 **9.c. Hybridization.** Cells on coverslips were blocked with 5% BSA and 0.1% triton-X 100 in
1222 PBS for 30 mins at 37°C, and washed twice with 0.1% triton-X 100 in PBS for 10 mins each with
1223 gentle agitation at room temperature. Cells were permeabilized with 0.1% saponin and 0.1%
1224 triton-X 100 in PBS for 10 mins at room temperature. Next, they were incubated in 20% glycerol
1225 in PBS for 20 mins, freeze-thawed three times with liquid nitrogen, and incubated in 0.1M
1226 hydrogen chloride at room temperature for 30 mins. Cells were further blocked for 1 hour at
1227 37°C in 3% BSA and 100 ug/mL RNase A in PBS. Cells were permeabilized again with 0.5%
1228 saponin and 0.5% triton-X 100 in PBS for 30 mins at room temperature. Lastly, they were rinsed
1229 with 1X PBS and washed with 2X SSC for 5 mins. For hybridization of probes, the prepared
1230 probes were denatured at 73°C for 5 mins in water bath. Cells were denatured in a two-step
1231 process in a 73°C water bath: 2.5 mins in 70% formamide in 2X SSC and 1 min in 50%
1232 formamide in 2X SSC. Denatured probes were transferred onto microscope slides, and
1233 coverslips were placed on top with cell-side facing down. The coverslips were sealed with
1234 rubber cement and incubated overnight at 37°C in a dark, humid chamber. Next day, coverslips
1235 were carefully removed and transferred onto a 6-well plate. Cells were washed at 37°C with
1236 gentle agitation, twice with 50% formamide in 2X SSC for 15 mins and three times with 2X SSC
1237 for 5 mins. The cells were then stained with DAPI (Invitrogen, D1306), mounted on microscope
1238 slides with ProLong Gold Antifade Mountant (Invitrogen, P36930), sealed with nail polish, and
1239 imaged.

1240 **9.d. Microscope and analysis.** Images were acquired with DeltaVision RT Deconvolution
1241 Microscope at UC San Diego's department of neuroscience (acquired with award NS047101).
1242 Captured images were processed using the TANGO⁷² plugin in ImageJ for quantitative analysis.
1243 Each FISH experiment contained two probes labeled with different color dyes (either U-C or C-
1244 D). We limited our analysis to nuclei containing 2 labeled foci for each color (4 total foci),
1245 allowing us to more confidently distinguish foci *in cis* from those *in trans*. Distances were
1246 measured from the center of one color focus to the center of the closest focus of the other color.
1247

1248 **10. Re-analysis of public datasets**

1249 **10.a. Analysis of ChIP-seq data from Kasowski et al and McVicker et al.** Raw fastq files
1250 were downloaded from SRA database for each experiment (SRP030041 and SRP026077,
1251 respectively). Reads were aligned to hg19 reference genome using BWA MEM (Kasowski) or
1252 BWA ALN⁶³ v0.7.8 (McVicker) with WASP pipeline⁴⁵ to eliminate allelic mapping bias. Only

1253 reads with high mapping quality (>10) were kept. PCR duplicates were removed using Picard
1254 tools v1.131 (<http://broadinstitute.github.io/picard>). MACS2⁷³ v2.2.1 was then used to call peaks
1255 using corresponding input files. For CTCF and SA1, default parameters were used for MACS2.
1256 For H3K27ac, H3K4me1, and H3K4me3, peak calling was done using "--nomodel" parameter
1257 because we do not expect sharp peaks for histone modifications. For H3K27me3 and
1258 H3K36me3, peak calling was done using "--nomodel --broad" parameter. Bigwig files were
1259 generated by MACS2 using fold enrichment for viewing in genome browser. All Kasowski data
1260 were processed in pair-end mode and both replicates were merged for analysis. All McVicker
1261 data were processed in single-end mode, and the pooled input data were used for all samples
1262 because there are no individual input files. To compute signals in peaks, we used a set of
1263 merged peaks across all individuals for each mark.

1264 **10.b. Analysis of RNA-seq data from Kasowski et al.** Raw fastq files were downloaded from
1265 SRA database (SRP030041). Reads were aligned to hg19 reference genome using STAR⁷⁴
1266 v2.4.2a with the WASP pipeline in pair-end mode to eliminate allelic mapping bias. Gencode⁷⁵
1267 v24 annotation was used to construct STAR index and computing FPKM. Only uniquely mapped
1268 reads were kept. Cufflinks⁷⁶ v2.2.1 was applied to compute FPKM values. Both replicates were
1269 merged for analysis.

1270 **10.c. Analysis of DNase-seq data from Degner et al.** Raw fastq files were downloaded from
1271 SRA database for each experiment (SRP007821). Reads were aligned to hg19 reference
1272 genome using BWA ALN with the WASP pipeline in single-end mode to eliminate allelic
1273 mapping bias. Only reads with high mapping quality (>10) were kept. PCR duplicates were
1274 removed using Picard tools. Bigwig files were generated using makeUCSCfile commands in
1275 homer tools⁷⁷ v4.9.1.

1276 **10.d. Analysis of ChIP-seq data from Ding et al.** Raw fastq files were downloaded from SRA
1277 database for each experiment (SRP004714). Reads were aligned to hg19 reference genome
1278 using BWA MEM v0.7.8 with the WASP pipeline to eliminate allelic mapping bias. Only reads
1279 with high mapping quality (>10) were kept. PCR duplicates were removed using Picard tools.
1280 We performed quality control for CTCF ChIP-seq data by FRIP (Fraction of Reads In Peaks)
1281 and used datasets with FRIP > 10. Bigwig files were generated using bamCoverage commands
1282 in deepTools⁷⁸ v2.3.3. To compute signals in peaks, we used the merged CTCF peaks from
1283 Kasowski data.

1284 **10.e. Analysis of ChIP-seq data from Grubert et al.** Bigwig files and peaks for H3K27ac,
1285 H3K4me1 and H3K4me3 were downloaded from GEO database (GSE62742). Peaks for each
1286 mark were merged and then used to compute the averaged signal.

1287

1288 DATA AVAILABILITY

1289

1290 All raw sequencing data and many processed data files are available through NCBI's Gene
1291 Expression Omnibus (GEO) (GSE128678), as well as through the 4D Nucleome data portal
1292 (<https://data.4dnucleome.org>). Additional processed data not provided above in supplement, as
1293 well as code, are available on the Ren lab's website at
1294 http://renlab.sdsc.edu/renlab_website//download/iqtl/, or by request.

1295

1296 ACKNOWLEDGEMENTS

1297

1298 This work was supported by NIH grant U54DK107977 to B.R. and M.H. D.U.G. was supported
1299 by fellowships from the A.P. Giannini Foundation and the NIH Institutional Research and
1300 Academic Career Development Awards (IRACDA) program. The authors would like to
1301 acknowledge members of the Ren lab, and Dr. Graham McVicker for important discussions and
1302 feedback during preparation of this manuscript. B.R. is co-founder and share holder
1303 of Arima Genomics. A.D.S. is employee and share holder of Arima Genomics.

1304

1305 REFERENCES

1306

- 1307 1 Gorkin, D. U., Leung, D. & Ren, B. The 3D genome in transcriptional regulation and
1308 pluripotency. *Cell Stem Cell* **14**, 762-775, doi:10.1016/j.stem.2014.05.017 (2014).
- 1309 2 Dekker, J. & Mirny, L. The 3D Genome as Moderator of Chromosomal Communication.
1310 *Cell* **164**, 1110-1121, doi:10.1016/j.cell.2016.02.007 (2016).
- 1311 3 Bouwman, B. A. & de Laat, W. Getting the genome in shape: the formation of loops,
1312 domains and compartments. *Genome Biol* **16**, 154, doi:10.1186/s13059-015-0730-1
1313 (2015).
- 1314 4 Pope, B. D. *et al.* Topologically associating domains are stable units of replication-timing
1315 regulation. *Nature* **515**, 402-405, doi:10.1038/nature13986 (2014).
- 1316 5 Dileep, V. *et al.* Topologically associating domains and their long-range contacts are
1317 established during early G1 coincident with the establishment of the replication-timing
1318 program. *Genome Res* **25**, 1104-1113, doi:10.1101/gr.183699.114 (2015).
- 1319 6 Engreitz, J. M. *et al.* The Xist lncRNA exploits three-dimensional genome architecture to
1320 spread across the X-chromosome*. *Science* **341**, 1237973,
1321 doi:10.1126/science.1237973 (2013).
- 1322 7 Crane, E. *et al.* Condensin-driven remodelling of X chromosome topology during dosage
1323 compensation. *Nature* **523**, 240-244, doi:10.1038/nature14450 (2015).
- 1324 8 Giorgetti, L. *et al.* Structural organization of the inactive X chromosome in the mouse.
1325 *Nature* **535**, 575-579, doi:10.1038/nature18589 (2016).
- 1326 9 Deng, X. *et al.* Bipartite structure of the inactive mouse X chromosome. *Genome Biol* **16**,
1327 152, doi:10.1186/s13059-015-0728-8 (2015).

- 1328 10 Lukas, J., Lukas, C. & Bartek, J. More than just a focus: The chromatin response to DNA
1329 damage and its role in genome integrity maintenance. *Nat Cell Biol* **13**, 1161-1169,
1330 doi:10.1038/ncb2344 (2011).
- 1331 11 Marnef, A. & Legube, G. Organizing DNA repair in the nucleus: DSBs hit the road. *Curr*
1332 *Opin Cell Biol* **46**, 1-8, doi:10.1016/j.ceb.2016.12.003 (2017).
- 1333 12 Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation.
1334 *Science* **295**, 1306-1311, doi:10.1126/science.1067799 (2002).
- 1335 13 Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals
1336 folding principles of the human genome. *Science* **326**, 289-293,
1337 doi:10.1126/science.1181369 (2009).
- 1338 14 Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals
1339 principles of chromatin looping. *Cell* **159**, 1665-1680, doi:10.1016/j.cell.2014.11.021
1340 (2014).
- 1341 15 Gibcus, J. H. & Dekker, J. The hierarchy of the 3D genome. *Mol Cell* **49**, 773-782,
1342 doi:10.1016/j.molcel.2013.02.011 (2013).
- 1343 16 Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of
1344 chromatin interactions. *Nature* **485**, 376-380, doi:10.1038/nature11082 (2012).
- 1345 17 Dixon, J. R., Gorkin, D. U. & Ren, B. Chromatin Domains: The Unit of Chromosome
1346 Organization. *Mol Cell* **62**, 668-680, doi:10.1016/j.molcel.2016.05.018 (2016).
- 1347 18 Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation
1348 centre. *Nature* **485**, 381-385, doi:10.1038/nature11049 (2012).
- 1349 19 Schmitt, A. D. *et al.* A Compendium of Chromatin Contact Maps Reveals Spatially Active
1350 Regions in the Human Genome. *Cell Rep* **17**, 2042-2059,
1351 doi:10.1016/j.celrep.2016.10.061 (2016).
- 1352 20 Yan, J. *et al.* in *Cell Res* Vol. 28 387 (2018).
- 1353 21 Dixon, J. R. *et al.* Chromatin architecture reorganization during stem cell differentiation.
1354 *Nature* **518**, 331-336, doi:10.1038/nature14222 (2015).
- 1355 22 Naumova, N. *et al.* Organization of the mitotic chromosome. *Science* **342**, 948-953,
1356 doi:10.1126/science.1236083 (2013).
- 1357 23 Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain
1358 formation in wild-type and engineered genomes. *Proc Natl Acad Sci U S A* **112**, E6456-
1359 6465, doi:10.1073/pnas.1518552112 (2015).
- 1360 24 Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320.e324,
1361 doi:10.1016/j.cell.2017.09.026 (2017).
- 1362 25 Jin, F. *et al.* A high-resolution map of the three-dimensional chromatin interactome in
1363 human cells. *Nature* **503**, 290-294, doi:10.1038/nature12644 (2013).
- 1364 26 Darrow, E. M. *et al.* Deletion of DXZ4 on the human inactive X chromosome alters
1365 higher-order genome architecture. *Proc Natl Acad Sci U S A* **113**, E4504-4512,
1366 doi:10.1073/pnas.1609643113 (2016).
- 1367 27 Greenwald, W. W. *et al.* Subtle changes in chromatin loop contact propensity are
1368 associated with differential gene regulation and expression. *Nat Commun* **10**, 1054,
1369 doi:10.1038/s41467-019-08940-5 (2019).
- 1370 28 Krijger, P. H. & de Laat, W. Regulation of disease-associated gene expression in the 3D
1371 genome. *Nat Rev Mol Cell Biol* **17**, 771-782, doi:10.1038/nrm.2016.138 (2016).
- 1372 29 Lupianez, D. G. *et al.* Disruptions of topological chromatin domains cause pathogenic
1373 rewiring of gene-enhancer interactions. *Cell* **161**, 1012-1025,
1374 doi:10.1016/j.cell.2015.04.004 (2015).
- 1375 30 Franke, M. *et al.* Formation of new chromatin domains determines pathogenicity of
1376 genomic duplications. *Nature* **538**, 265-269, doi:10.1038/nature19800 (2016).
- 1377 31 Hnisz, D. *et al.* Activation of proto-oncogenes by disruption of chromosome
1378 neighborhoods. *Science* **351**, 1454-1458, doi:10.1126/science.aad9024 (2016).

- 1379 32 Flavahan, W. A. *et al.* Insulator dysfunction and oncogene activation in IDH mutant
1380 gliomas. *Nature* **529**, 110-114, doi:10.1038/nature16490 (2016).
- 1381 33 Smemo, S. *et al.* Obesity-associated variants within FTO form long-range functional
1382 connections with IRX3. *Nature* **507**, 371-375, doi:10.1038/nature13138 (2014).
- 1383 34 Visser, M., Kayser, M. & Palstra, R. J. HERC2 rs12913832 modulates human
1384 pigmentation by attenuating chromatin-loop formation between a long-range enhancer
1385 and the OCA2 promoter. *Genome Res* **22**, 446-455, doi:10.1101/gr.128652.111 (2012).
- 1386 35 Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74,
1387 doi:10.1038/nature15393 (2015).
- 1388 36 Kilpinen, H. *et al.* Coordinated effects of sequence variation on DNA binding, chromatin
1389 structure, and transcription. *Science* **342**, 744-747, doi:10.1126/science.1242463 (2013).
- 1390 37 Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional
1391 variation in humans. *Nature* **501**, 506-511, doi:10.1038/nature12531 (2013).
- 1392 38 Degner, J. F. *et al.* DNase I sensitivity QTLs are a major determinant of human
1393 expression variation. *Nature* **482**, 390-394, doi:10.1038/nature10808 (2012).
- 1394 39 Grubert, F. *et al.* Genetic Control of Chromatin States in Humans Involves Local and
1395 Distal Chromosomal Interactions. *Cell* **162**, 1051-1065, doi:10.1016/j.cell.2015.07.048
1396 (2015).
- 1397 40 McVicker, G. *et al.* Identification of genetic variants that affect histone modifications in
1398 human cells. *Science* **342**, 747-749, doi:10.1126/science.1242429 (2013).
- 1399 41 Banovich, N. E. *et al.* Methylation QTLs are associated with coordinated changes in
1400 transcription factor binding, histone modifications, and gene expression levels. *PLoS*
1401 *Genet* **10**, e1004663, doi:10.1371/journal.pgen.1004663 (2014).
- 1402 42 Kasowski, M. *et al.* Extensive variation in chromatin states across humans. *Science* **342**,
1403 750-752, doi:10.1126/science.1242510 (2013).
- 1404 43 Selvaraj, S., J. R. D., Bansal, V. & Ren, B. Whole-genome haplotype reconstruction
1405 using proximity-ligation and shotgun sequencing. *Nat Biotechnol* **31**, 1111-1118,
1406 doi:10.1038/nbt.2728 (2013).
- 1407 44 The International HapMap Project. *Nature* **426**, 789-796, doi:10.1038/nature02168
1408 (2003).
- 1409 45 van de Geijn, B., McVicker, G., Gilad, Y. & Pritchard, J. K. WASP: allele-specific
1410 software for robust molecular quantitative trait locus discovery. *Nat Methods* **12**, 1061-
1411 1063, doi:10.1038/nmeth.3582 (2015).
- 1412 46 Stark, A. L. *et al.* Population differences in the rate of proliferation of international
1413 HapMap cell lines. *Am J Hum Genet* **87**, 829-833, doi:10.1016/j.ajhg.2010.10.018
1414 (2010).
- 1415 47 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing
1416 and microarray studies. *Nucleic Acids Res* **43**, e47, doi:10.1093/nar/gkv007 (2015).
- 1417 48 Dixon, J. R. *et al.* Integrative detection and analysis of structural variation in cancer
1418 genomes. *Nat Genet* **50**, 1388-1398, doi:10.1038/s41588-018-0195-8 (2018).
- 1419 49 Ignatiadis, N., Klaus, B., Zaugg, J. B. & Huber, W. Data-driven hypothesis weighting
1420 increases detection power in genome-scale multiple testing. *Nat Methods* **13**, 577-580,
1421 doi:10.1038/nmeth.3885 (2016).
- 1422 50 Zuin, J. *et al.* Cohesin and CTCF differentially affect chromatin architecture and gene
1423 expression in human cells. *Proc Natl Acad Sci U S A* **111**, 996-1001,
1424 doi:10.1073/pnas.1317788111 (2014).
- 1425 51 Sofueva, S. *et al.* Cohesin-mediated interactions organize chromosomal domain
1426 architecture. *EMBO J* **32**, 3119-3129, doi:10.1038/emboj.2013.237 (2013).
- 1427 52 Tang, Z. *et al.* CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin
1428 Topology for Transcription. *Cell* **163**, 1611-1627, doi:10.1016/j.cell.2015.11.024 (2015).

- 1429 53 Li, Y. I. *et al.* RNA splicing is a primary link between genetic variation and disease. *Science* **352**, 600-604, doi:10.1126/science.aad9417 (2016).
- 1430
- 1431 54 Ding, Z. *et al.* Quantitative genetics of CTCF binding reveal local sequence effects and
1432 different modes of X-chromosome association. *PLoS Genet* **10**, e1004798,
1433 doi:10.1371/journal.pgen.1004798 (2014).
- 1434 55 Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association
1435 studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* **47**, D1005-
1436 d1012, doi:10.1093/nar/gky1120 (2019).
- 1437 56 de Lange, K. M. *et al.* Genome-wide association study implicates immune activation of
1438 multiple integrin genes in inflammatory bowel disease. *Nat Genet* **49**, 256-261,
1439 doi:10.1038/ng.3760 (2017).
- 1440 57 Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological
1441 architecture of adult human height. *Nat Genet* **46**, 1173-1186, doi:10.1038/ng.3097
1442 (2014).
- 1443 58 Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity
1444 biology. *Nature* **518**, 197-206, doi:10.1038/nature14177 (2015).
- 1445 59 Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-
1446 947, doi:10.1016/j.cell.2013.09.053 (2013).
- 1447 60 McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions.
1448 *Nat Biotechnol* **28**, 495-501, doi:10.1038/nbt.1630 (2010).
- 1449 61 Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell
1450 types. *Nature* **473**, 43-49, doi:10.1038/nature09906 (2011).
- 1451 62 Ernst, J. & Kellis, M. in *Nat Methods* Vol. 9 215-216 (2012).
- 1452 63 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
1453 (2013).
- 1454 64 Hu, M. *et al.* HiCNorm: removing biases in Hi-C data via Poisson regression.
1455 *Bioinformatics* **28**, 3131-3133, doi:10.1093/bioinformatics/bts570 (2012).
- 1456 65 Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression
1457 data using empirical Bayes methods. *Biostatistics* **8**, 118-127,
1458 doi:10.1093/biostatistics/kxj037 (2007).
- 1459 66 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing
1460 next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303,
1461 doi:10.1101/gr.107524.110 (2010).
- 1462 67 Bansal, V. & Bafna, V. HapCUT: an efficient and accurate algorithm for the haplotype
1463 assembly problem. *Bioinformatics* **24**, i153-159, doi:10.1093/bioinformatics/btn298
1464 (2008).
- 1465 68 Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-
1466 data inference for whole-genome association studies by use of localized haplotype
1467 clustering. *Am J Hum Genet* **81**, 1084-1097, doi:10.1086/521987 (2007).
- 1468 69 Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W. W. & Lenhard, B. JASPAR: an
1469 open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids*
1470 *Res* **32**, D91-94, doi:10.1093/nar/gkh012 (2004).
- 1471 70 Yaffe, E. & Tanay, A. Probabilistic modeling of Hi-C contact maps eliminates systematic
1472 biases to characterize global chromosomal architecture. *Nat Genet* **43**, 1059-1065,
1473 doi:10.1038/ng.947 (2011).
- 1474 71 Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. ImerTest Package: Tests in
1475 Linear Mixed Effects Models. **82**,
1476 doi:<https://www.jstatsoft.org/index.php/jss/article/view/v082i13> (2017).
- 1477 72 Ollion, J., Cochenec, J., Loll, F., Escudé, C. & Boudier, T. in *Bioinformatics* Vol. 29
1478 1840-1841 (2013).

1479 73 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137,
1480 doi:10.1186/gb-2008-9-9-r137 (2008).
1481 74 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21,
1482 doi:10.1093/bioinformatics/bts635 (2013).
1483 75 Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE
1484 Project. *Genome Res* **22**, 1760-1774, doi:10.1101/gr.135350.111 (2012).
1485 76 Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals
1486 unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*
1487 **28**, 511-515, doi:10.1038/nbt.1621 (2010).
1488 77 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime
1489 cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-
1490 589, doi:10.1016/j.molcel.2010.05.004 (2010).
1491 78 Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible
1492 platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187-191,
1493 doi:10.1093/nar/gku365 (2014).
1494

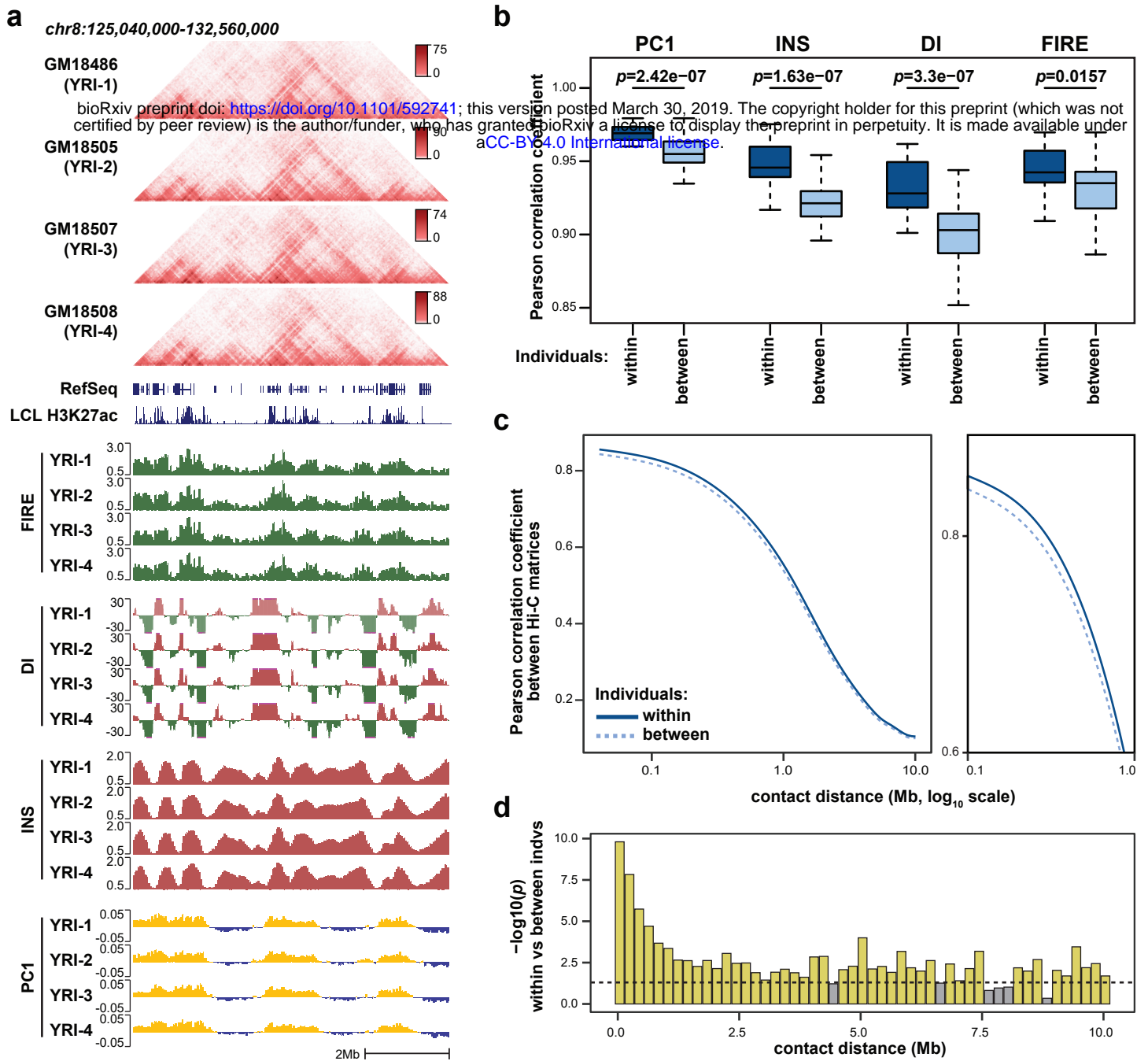


Figure 1

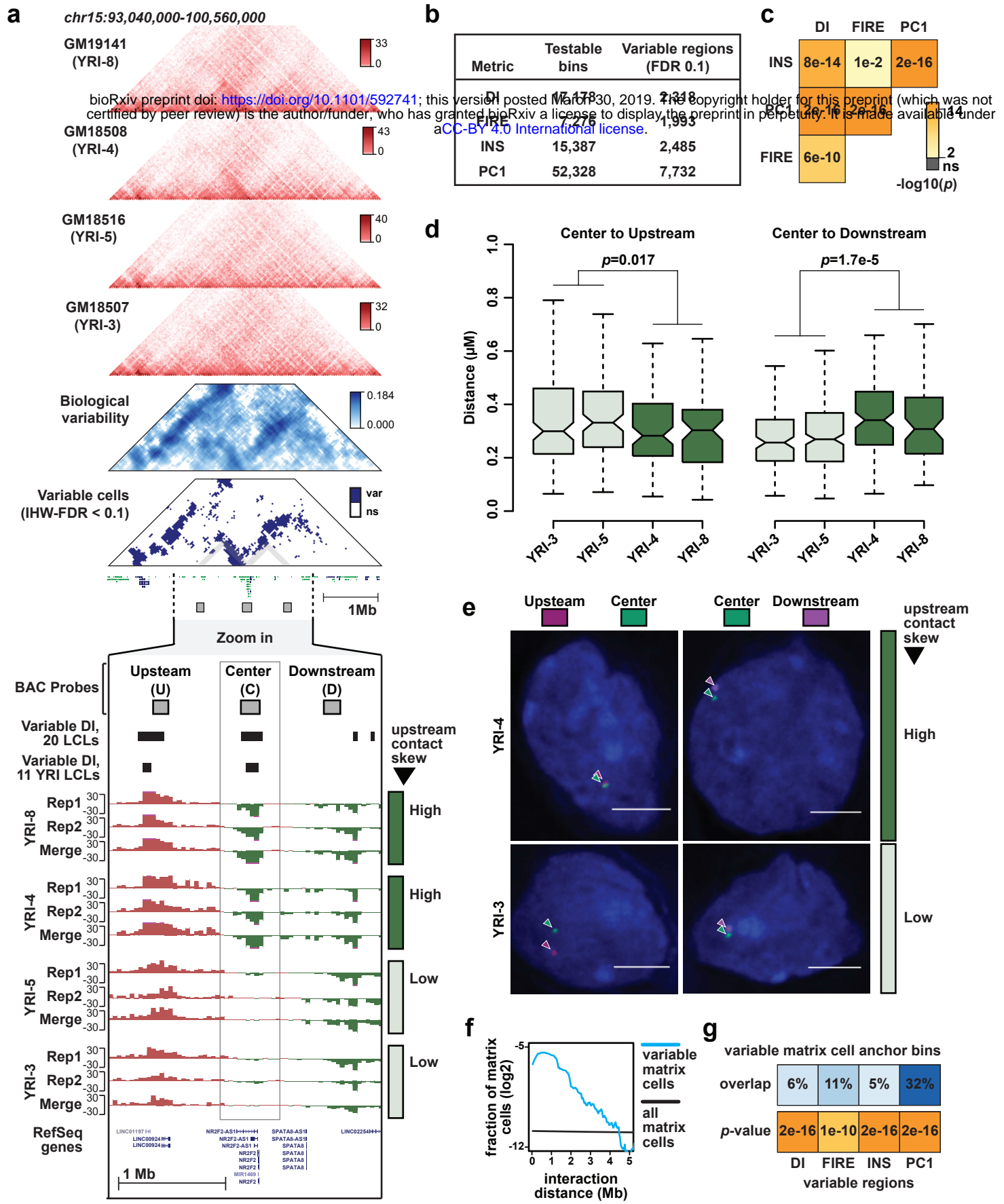


Figure 2

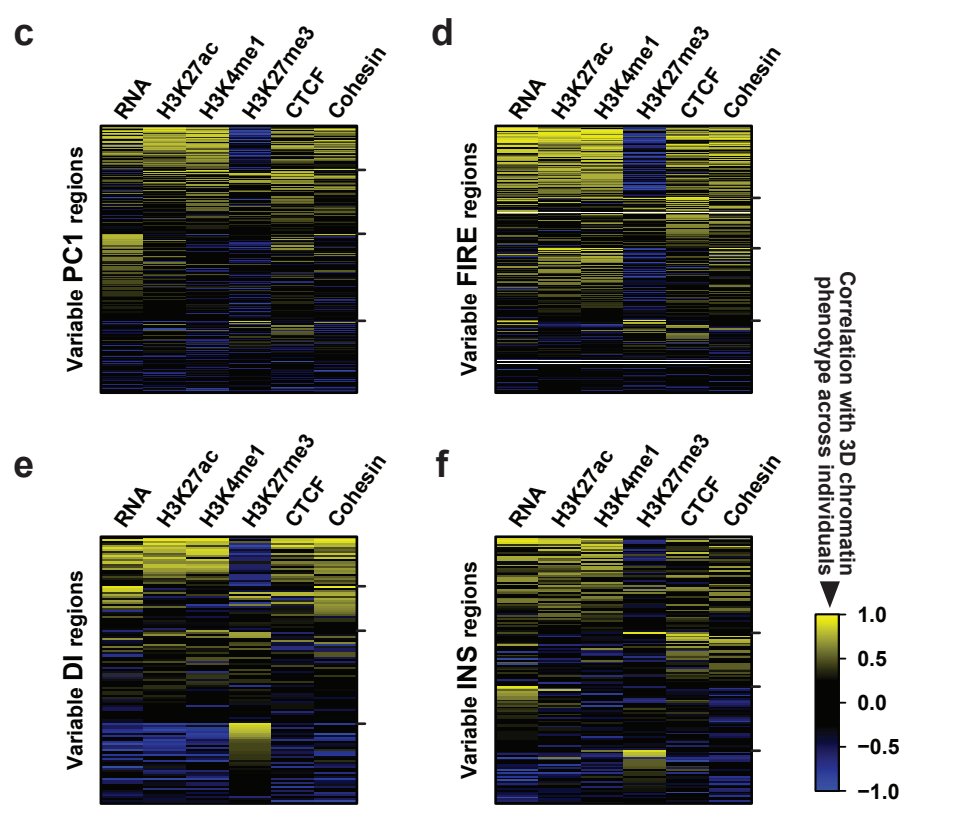
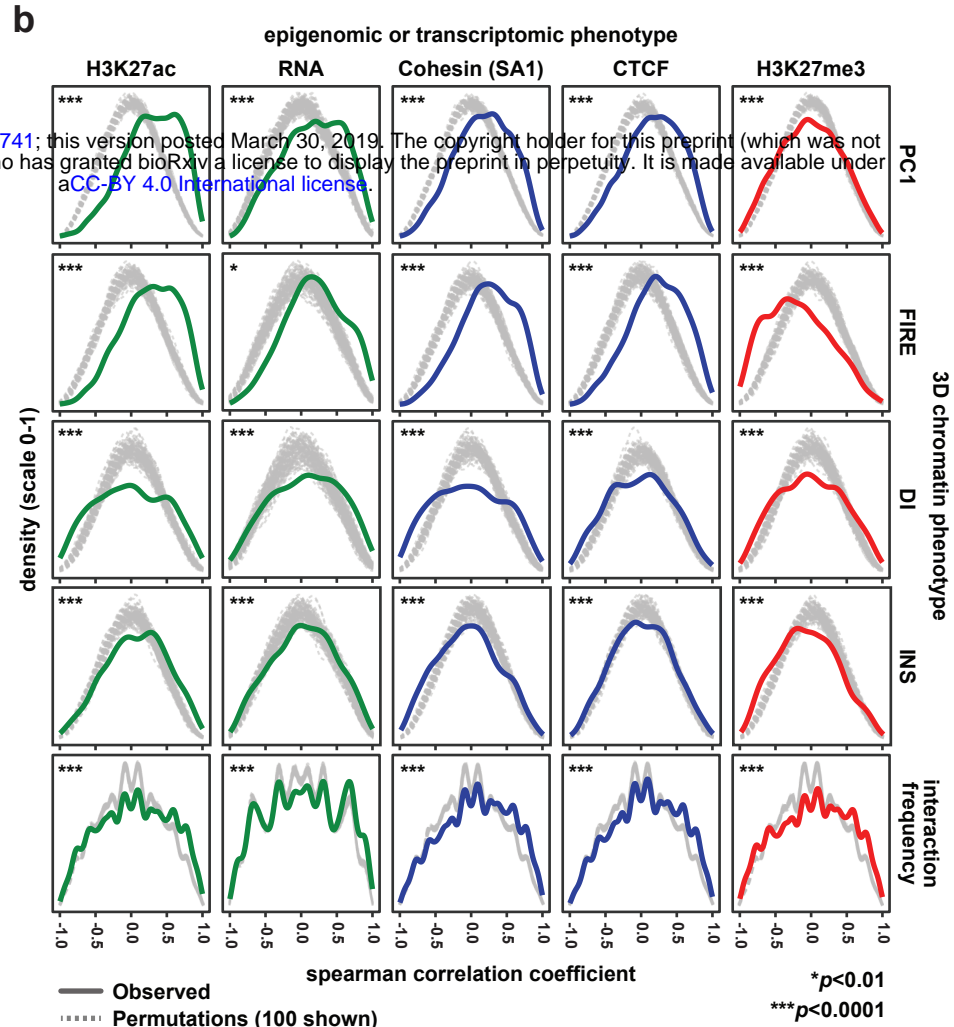
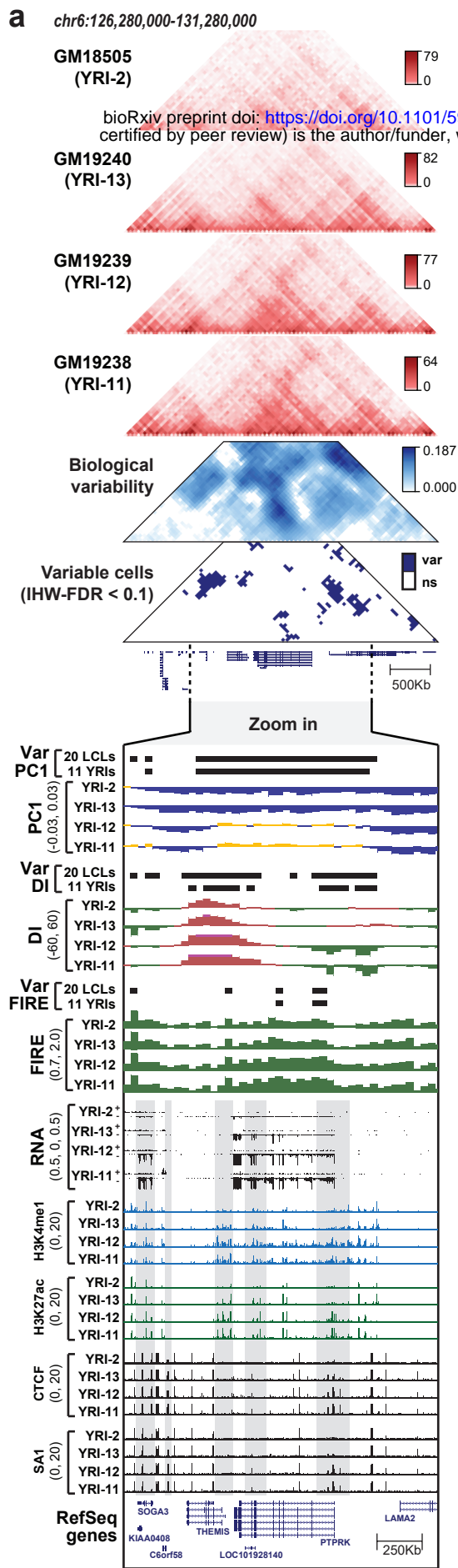


Figure 3

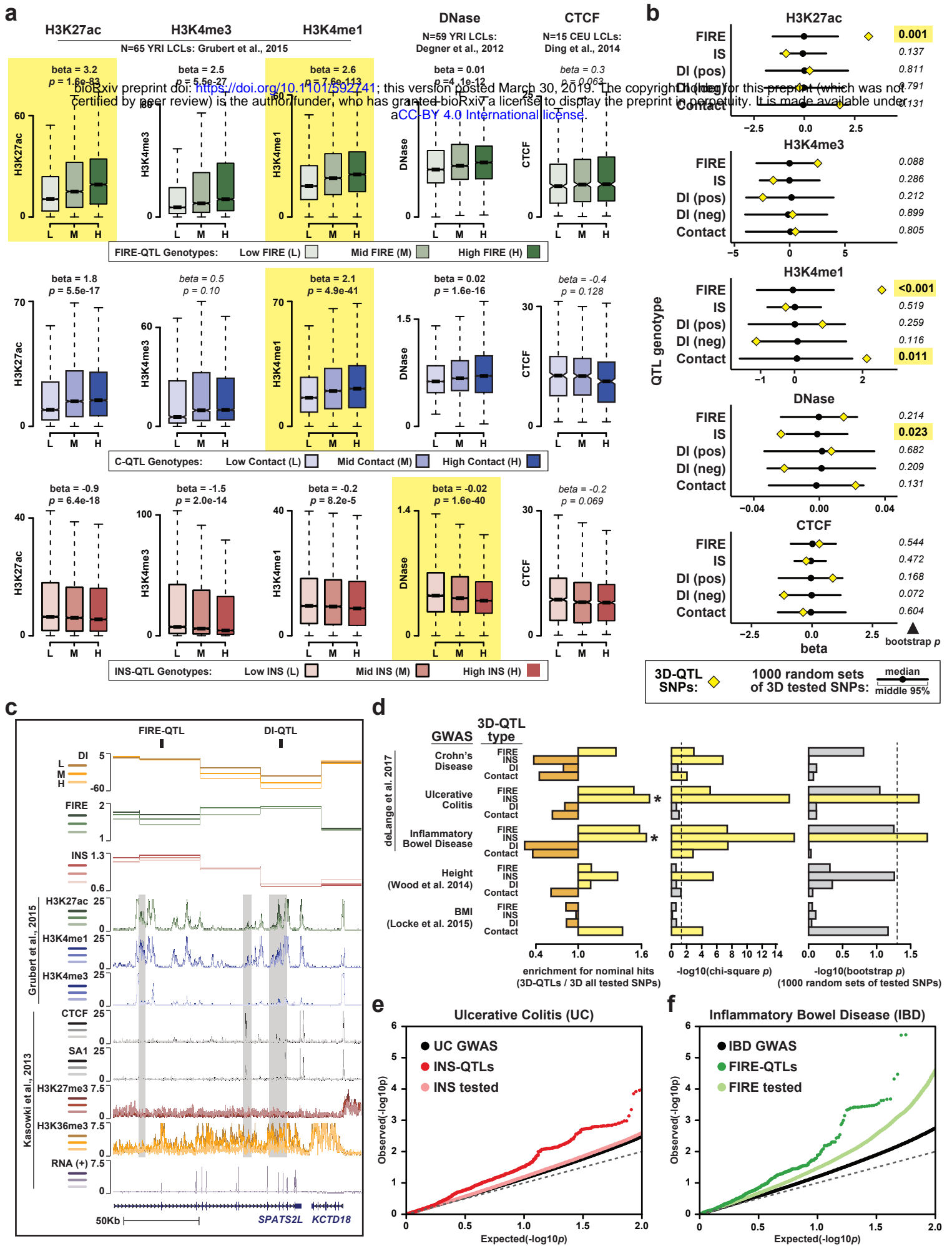


Figure 5