# CLASH (Chromatin Loop Across-sample Score Harmonizer) Quantifies a Two-Factor Model Linking Genetic and Epigenetic Variation, CTCF Occupancy, and Chromatin Loop Strength

Valmik Ranparia[1]
The Human Genome Structural Variation Consortium,
Geoffrey Fudenberg[1]
Mark JP Chaisson[1,2]

1. Department of Quantitative and Computational Biology, University of Southern California, CA, USA.
2. Corresponding author: mchaisso@usc.edu .

**Abstract:**

Chromatin loops are central regulators of gene expression, yet how naturally occurring genetic and epigenetic variation influences loop formation across individuals remains unclear. Although CTCF motifs are frequently altered by SNPs, structural variants, and 5-methylcytosine (m5C) CpG methylation, existing approaches lack the methodology needed to elucidate the relationship between these molecular changes, CTCF occupancy, and loop strength. Here, we combined high resolution Hi-C, Fiber-seq, near telomere-to-telomere phased assemblies, and $m^5C$ methylation maps across five lymphoblastoid cell lines to quantify how genetic and epigenetic variation shapes chromatin loop formation. We identified 382 differential contacts and showed that sequence variation, chromatin accessibility, and $m^5C$ CpG methylation each are significantly associated with differential chromatin contacts and strongly correlate with CTCF occupancy (combined $r = 0.48$, $p = 1.1 \times 10^{-16}$). To study differential loop formation, we developed CLASH (Chromatin Loop Across-sample Score Harmonizer), a method that enables robust, quantitative comparisons of loop strengths across individuals to address inconsistent calls across samples from existing methods. CLASH substantially improves biological concordance with Hi-C signal (global $r = 0.67$ vs 0.35 for initial calls; $p < 10^{-300}$) and reveals a significant relationship between CTCF occupancy and loop strength (global $r = 0.26$, $p = 4.6 \times 10^{-267}$). We found that 58% of PWM-associated and 71% of $m^5C$-associated effects on loop variation act through CTCF occupancy, establishing a two-factor model in which genetic and epigenetic variation modulate CTCF occupancy, and occupancy affects loop formation strength.

**Introduction:**

In eukaryotic cells, the three-dimensional spatial organization of the genome plays a critical role in intricately regulating vital cellular processes such as transcription and replication. There are three primary structures within chromatin architecture: A/B compartments, topologically associating domains (TADs), and chromatin loops (Lieberman-Aiden *et al*., 2009; Xu *et al.*, 2024; Holwerda *et al.*, 2012). Chromatin loops are local topological features that influence gene expression by bringing linearly distant DNA segments into proximity and positioning enhancers/silencers near promoters. Chromatin loops at the same locus have been shown to differ across cell types, developmental stages, and between individuals at the same cell state. This difference between individuals has been implicated in a variety of diseases, including some cancers (Chen *et al.,* 2024; Yoon *et al.*, 2024; Panarotto *et al*., 2022). Although the fundamental role of CCCTC-binding factor (CTCF) in creating chromatin loops is well established, the specific mechanisms underlying differential chromatin loop formation remain incompletely understood (Rao *et al.*, 2014; Pękowska *et al.*, 2018; Bond *et al.*, 2023).

Previous work has indicated that CTCF binding is a dynamic process, and that even modest shifts in CTCF occupancy may have the potential to alter loop stability (Hansen *et al.*, 2017; Wutz *et al.*, 2020). Additionally, multiple studies have found some correlation between genetic and epigenetic variation – primarily focused at CTCF-binding sites – and differential chromatin interactions (Gorkin *et al.*, 2019; Li *et al*., 2024; Monteagudo-Sánchez *et al.*, 2024). However, these studies suffer from four limitations: (i) these studies used Hi-C/HiChIP resolutions between 5-40 kb, which are unable to capture the specific effect of most structural variations that are under 2 kb (ii) these effects have largely been demonstrated in engineered or developmental contexts, and to our knowledge, no study has explored how naturally occurring genetic and epigenetic variation between individuals influences loop formation at the same loci genome-wide (iii) previous studies have used HiChIP (Mumbach *et al,*. 2016), which enriches chromatin contacts associated with CTCF, resulting in a readout of contact frequency dependent on CTCF binding occupancy, and (iv) existing loop-calling methods are not optimized for inter-individual comparisons: different algorithms often identify inconsistent loop sets across samples and treat loop formation as a binary event rather than a quantitative property (Chowdhury *et al.*, 2024; Greenwald *et al.*, 2019). These constraints have limited our understanding of the relationship between genetic and epigenetic variation at CTCF sites and differential chromatin loop formation, and further isolated the effect of CTCF binding from other influences on genome structure.

To address this gap, we propose that CTCF occupancy represents the mechanistic link between genetic and epigenetic variation and differential chromatin loop formation between individuals. While genetic and epigenetic changes can alter the potential for loop formation, we hypothesize that these effects are indirect and mediated through changes in the frequency with which CTCF is bound at loop anchors. To test this model, we developed CLASH (Chromatin Loop Across-sample Score Harmonizer), a method that assigns harmonized, quantitative loop-strength scores to candidate loops and enables robust cross-sample comparison of loop

strength. Using CLASH-quantified chromatin loops together with high-resolution Hi-C contact maps, phased sequence variation, 5mC methylation profiles, and single-molecule CTCF-occupancy measurements from Fiber-seq (Stergachis et al., 2020), we were able to overcome the key methodological constraints of prior low-resolution Hi-C and HiChIP studies. Our findings support a two-factor model in which sequence and methylation variation modulate CTCF occupancy, which in turn governs inter-individual differences in chromatin loop strength.

**Results.**

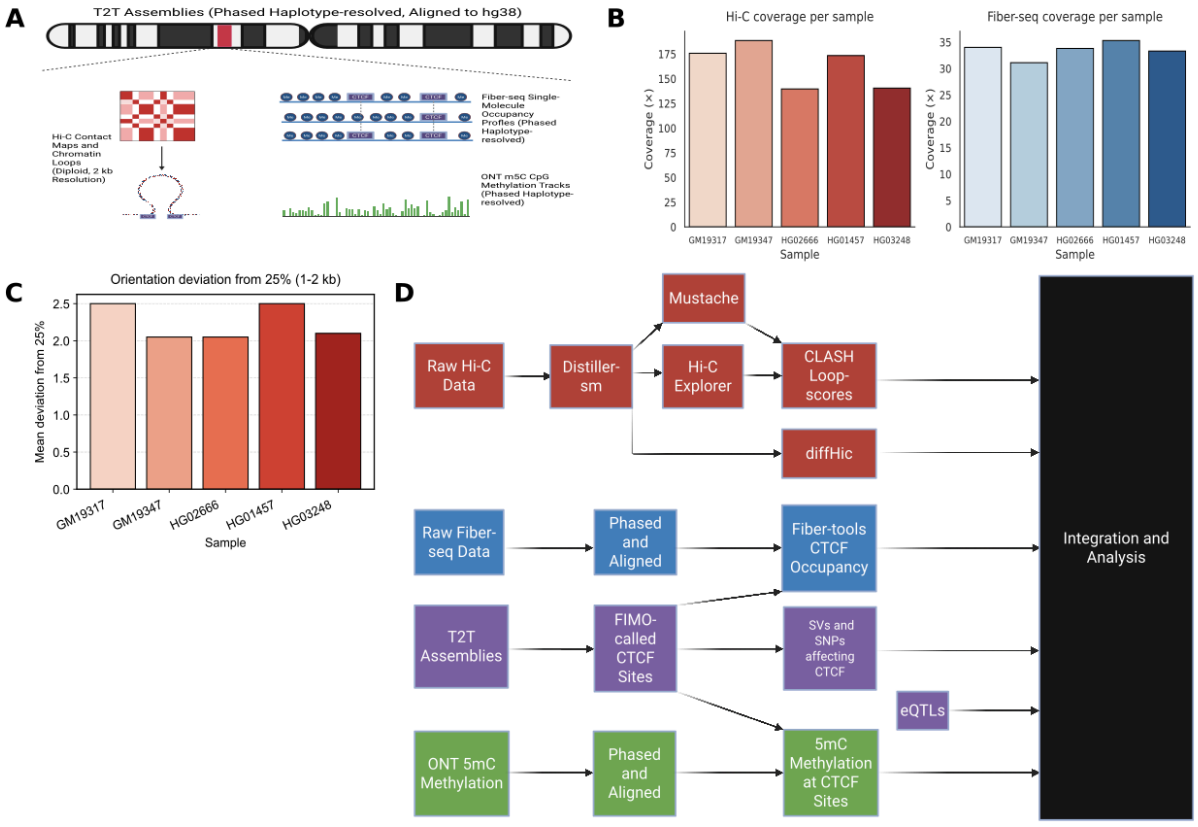



**Figure 1. Multi-omic data generation, processing, and quality assessment across five lymphoblastoid samples. a,** Overview of all datasets generated and integrated in this study, including Hi-C contact maps, Fiber-seq $m^6A$–based chromatin accessibility and CTCF occupancy, ONT-derived $m^5C$ CpG methylation, and haplotype-resolved variant calls for GM19317, GM19347, HG01457, HG02666, and HG03248. All data modalities except Hi-C are phased and haplotype-aware. Hi-C contact matrices were aligned to the GRCh38 reference genome; Fiber-seq, $m^5C$ methylation, and variant data were aligned or lifted over to GRCh38 to

enable joint analysis. Created in https://BioRender.com. **b,** Sequencing depth for Hi-C (left) and Fiber-seq (right) across all five samples. Hi-C datasets yielded an average of ~163.5× genomic coverage per individual, while Fiber-seq achieved an average of ~33.5× coverage, ensuring sufficient depth for high-resolution loop detection and CTCF occupancy quantification. **c,** Distribution of Hi-C read-pair orientation deviance for all samples from expected 25% at 1–2 kb resolution, showing near-equal representation of strand classes consistent with balanced library composition. **d,** Schematic overview of the computational workflows used in this study, including Hi-C processing and loop calling (distiller-sm, DiffHiC, Mustache, HiCExplorer, and CLASH), Fiber-seq alignment and CTCF occupancy calling (whatshap and fibertools), $m^5C$ CpG methylation phasing, structural variant integration, and CTCF motif identification using FIMO. Created in https://BioRender.com.

To enable joint analysis of chromatin architecture, DNA methylation, and protein occupancy across individuals, we generated and integrated Hi-C, haplotype-resolved Fiber-seq for $N^6$-Methyladenosine (m6a) annotation of chromatin accessibility, haplotype-resolved Oxford Nanopore (ONT) m5C methylation tracks, and near-telomere-to-telomere (T2T) assembly datasets for five lymphoblastoid cell lines, GM19317, GM19347, HG01457, HG02666, and HG03248 (Logsdon *et al.*, 2025; Figure 1a). Hi-C samples yielded an average of ~163.5x genomic coverage and the distribution of read-pair orientations at 1–2 kb binning per sample deviated by only 2.24% on average from the ideal 25% per class, which support performing downstream analyses at 2 kilobase-scale resolution (Figs. 1b,c; Table S1). Fiber-seq data were generated using PacBio HiFi sequencing with a mean coverage of 33.5× per genome (Figure 1b, Table S2). The full computational pipeline used is summarized in Fig. 1d.

All samples were confirmed to be in the same cell state by concordance of A/B compartment structures (average MSE = 0.024 ± 0.006, concordance = 94.7% ± 0.9%; Supplemental Figure 1), and consistent large-scale structure from E1 values (Supplemental Figures 2,3). Together, these indicate that compartment-level differences do not confound subsequent analyses of contacts and loop formation.

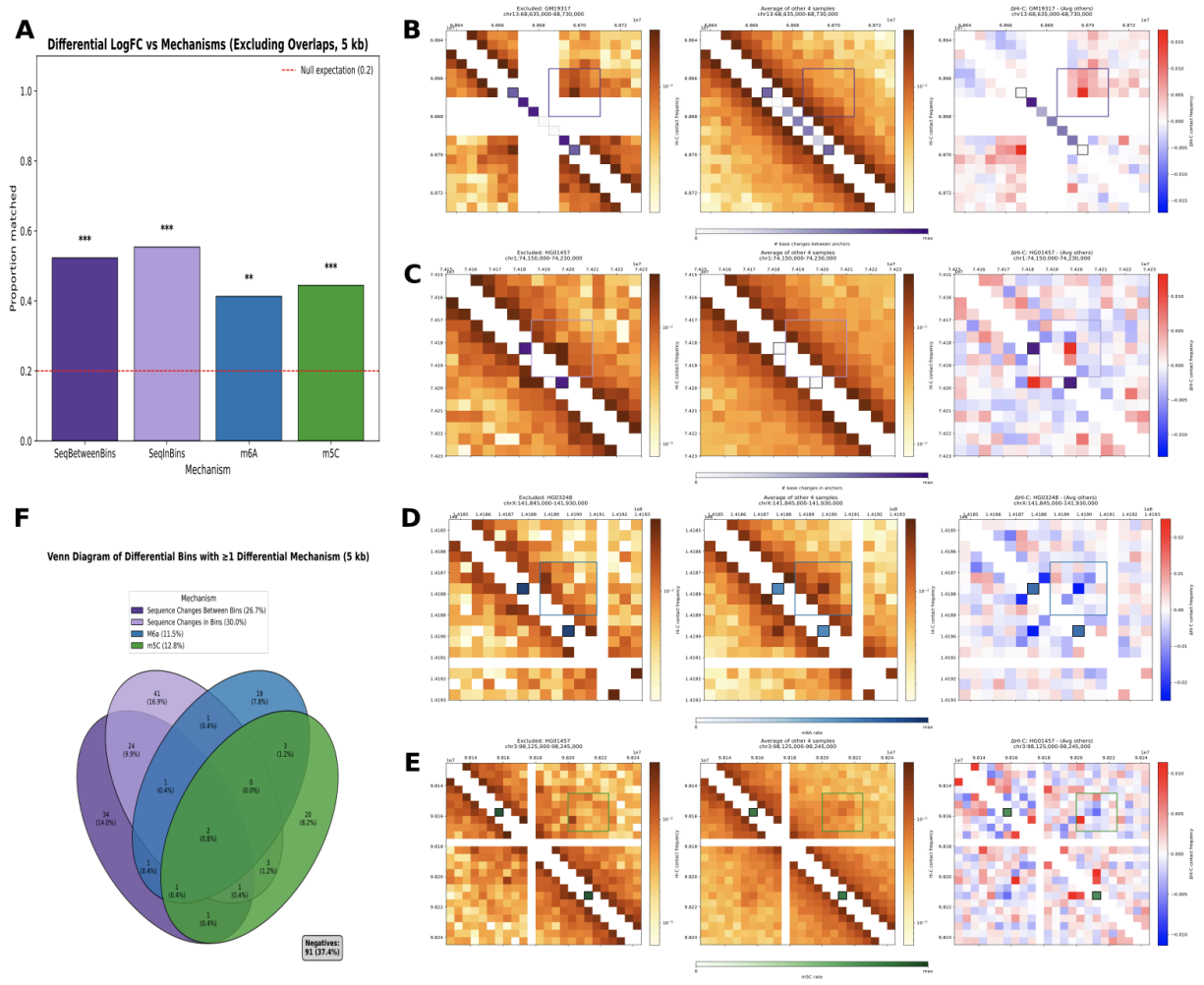



**Figure 2: Mechanisms affecting DiffHiC identified differential chromatin interactions. a,** Match rates for four mechanisms – sequence changes between anchors (dark purple), sequence changes within anchors (light purple), m6A methylation (blue), and m5C methylation (green) – after accounting for the influence of the other three mechanisms. All mechanisms remain significantly associated with differential chromatin contacts (between anchors: n = 65, p = $7.39 \times 10^{-9}$, 52.3%; in anchors: n = 74, p = $1.97 \times 10^{-11}$ 55.4%; m6A, n = 46, p = 0.0012 41.3%; m5C: n = 45, p = $2.20 \times 10^{-4}$, 44.4%). Significance levels are denoted as $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***), and were calculating using the binomial test compared against the null of 0.2. **b,** Example of a differential chromatin interaction driven by sequence changes between anchor bins. Left: Hi-C contact map for the excluded sample (GM19317) at chr13:68,635,000–68,730,000. Middle: Average Hi-C contact map of the remaining four samples. Right: ΔHi-C map (GM19317 − mean of others), highlighting the interaction that differs most strongly in GM19317. The purple diagonal bins mark the total number of base changes between the two anchor bins of the differential interaction, and the purple rectangle centers around focal contact whose strength

deviates in GM19317. Compared to the other samples, GM19317 carries a greater number of sequence changes between anchors, corresponding to an increase in contact strength (red shift). **c,** Example of a differential chromatin interaction at chr1:74,150,000-74,320,000 driven by sequence changes in anchor bins. Shown in the same layout as panel 2B. Anchor bins are annotated in purple for the number of bases changed within the bins. The highlighted interaction shows increased contact strength in HG01457 (red shift). **d,** Example of a differential chromatin interaction at chrX:141,845,000-141,930,000 driven by m6A methylation in anchor bins. Shown in the same layout as panel 2B. Anchor bins are annotated in blue for the m6A methylation rate in each bin. The highlighted interaction shows decreased contact strength in HG03248 (blue shift). **e,** Example of a differential chromatin interaction at chr3:98,125,000-98,245,000 driven by m5C methylation rate in anchor bins. Shown in the same layout as panel 2B. Anchor bins are annotated in green for the m5C methylation rate in each bin. The highlighted interaction shows decreased contact strength in HG01457 (blue shift). **f,** Venn diagram summarizing the 243 differential bins exhibiting at least one differential mechanism. In total, 62.6% of these bins can be explained by one or more of the four mechanisms shown in panel A.

To test the two-factor model, we first investigated whether genetic and epigenetic mechanisms influence chromatin interactions genome-wide. Using DiffHiC, we identified 382 significantly different contact bins across the five samples, comparing each sample to a set formed of the remaining four samples ($P < NNN$; Supplementary Figures 4-8). We then assessed the various data types to measure how sequence variation within and between contact bins, as well as m5c/m6A methylation rates correlate with contact strength.

In loci where one sample exhibited a differential number of bases changed between contact bins, 60.7% of samples exhibiting differential interactions matched the sample containing the sequence change ($n = 107$; $p = 3.95 \times 10^{-20}$, binomial test; Supplementary Figure 9). Similarly, in loci where one sample exhibited a differential number of bases changed in contact bins 61.9% of samples exhibiting differential interactions matched the sample containing the sequence change ($n = 118$; $p = 4.25 \times 10^{-23}$, binomial test; Supplementary Figure 10). These enrichments demonstrate that both large scale structural rearrangements and smaller sequence changes in interaction-pair bins are strongly associated with altered chromatin contact strength across individuals, consistent with previous findings (Gorkin *et al.*, 2019; Li *et al.*, 2024). Similarly, for epigenetic mechanisms, differentially decreased contacts were also enriched in the samples with higher methylation levels with a match rate of 37.3% for $m^6A$ accessibility ($n = 75$; $p = 8.62 \times 10^{-4}$, binomial test) and 41.9% for $m^5C$ CpG methylation ($n = 74$; $p = 1.58 \times 10^{-5}$, binomial test; Supplementary Figures 11,12). These enrichments demonstrate that epigenetic differences at loop-anchor loci, manifested as enrichments in methylation, are strongly associated

with decreased chromatin contact strength across individuals. This is expected for $m^5C$ based on prior observations (Monteagudo-Sánchez *et al.*, 2024), whereas no comparable evidence exists for $m^6A$, making the $m^6A$ association unique to this dataset.

After accounting for each of the four mechanisms as covariates, all mechanisms remained enriched among differential interactions (Fig. 2a). The adjusted match rates were 52.3% for bases changed between anchors (n = 65, $p = 7.39 \times 10^{-9}$, binomial test), 55.4% for bases changed within anchors (n = 74, $p = 1.97 \times 10^{-11}$, binomial test), 41.3% for $m^6A$ accessibility (n = 46, p = 0.0012, binomial test), and 44.4% for $m^5C$ methylation (n = 45, $p = 2.20 \times 10^{-4}$, binomial test). Representative examples of differential interactions associated with each mechanism are shown in annotated Hi-C heatmaps (Figs. 2b-e). Across the 243 differential interactions where at least one differential mechanism was present, 62.6% could be attributed to one or more of these four sources of variation (Fig. 2f). The frequency distribution of contact bins among differential interactions at 5 kb resolution shows that the majority of identified bins only appear once, while a minority recur across interactions suggesting that they are tied to larger-scale architectural features (Supplementary Figure 13). Approximately half of identified differential chromatin interactions at 1 kb (45%) and 2 kb (48%) resolution and 15% at 5 kb resolution lie within 10 kb of chromatin loops, confirming that chromatin loops account for a substantial fraction of genome-wide differential interaction patterns (Supplementary Figure 14). Together, these results show that genetic variation, differences in chromatin accessibility and protein occupancy, and CpG methylation collectively contribute to inter-individual variation in chromatin contact strength.

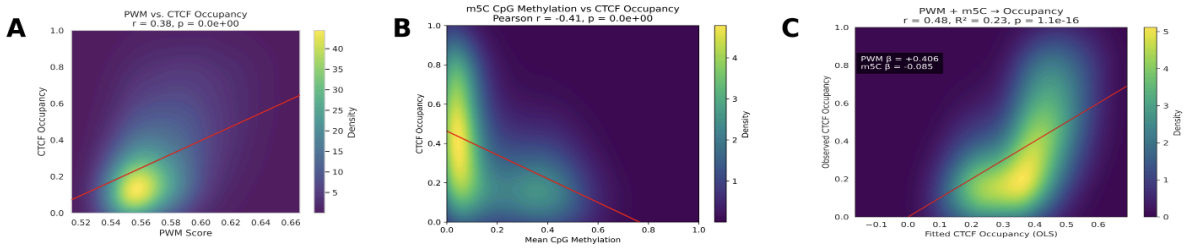



**Figure 3: Quantification of genetic and epigenetic mechanisms influencing CTCF occupancy. a,** Correlation between CTCF motif strength and occupancy. Position weight matrix (PWM) scores were computed for 509,826 CTCF motifs across 10 haplotypes and correlated with Fiber-seq–derived CTCF occupancy (n = 262,463 sites; Pearson r = 0.38; $p < 10^{-300}$). **b,** Relationship between $m^5C$ CpG methylation and CTCF occupancy. Average $m^5C$ methylation percentage across CpG sites within each motif was correlated with occupancy (n = 90,672 sites;

Pearson r = −0.41; $p < 10^{-300}$). **c,** Linear model estimating the combined effects of PWM score and $m^5C$ methylation on CTCF occupancy using ordinary least squares regression (n = 90,672 sites). Both predictors showed significant associations with occupancy (PWM β = +0.406; $m^5C$ β = −0.085), and the model explained a moderate fraction of variance (Pearson r = 0.48; $R^2 = 0.23$; $p = 1.1 \times 10^{-16}$).

We next quantified how genetic and epigenetic variation correlates with CTCF occupancy, the intermediate component of our two-factor model. To assess the contribution of genetic variation, we correlated global motif strength with occupancy across 262,463 CTCF sites (from 10 haplotypes), observing a positive association between PWM score and binding level (r = 0.38; $p < 10^{-300}$; Pearson r; Fig. 3a). This was confirmed by a complementary metric – the number of bases altered within each motif – which showed a consistent negative association with occupancy (r = –0.21; $p < 10^{-300}$; Pearson r; Supplementary Figure 15). We next examined epigenetic variation. Averaged $m^5C$ methylation across CpGs within each motif was negatively correlated with occupancy (n = 90,672; r = –0.41; $p < 10^{-300}$; Pearson r; Fig. 3B), and this trend was independently supported by comparing occupancy across hypo-, hyper-, and mixed-methylation states (hypomethylated n = 58,094 sites; hypermethylated n = 140 sites; mixed n = 14 sites; r = –0.64; $p = 4.4 \times 10^{-39}$; Pearson r; Supplementary Figure 16). Finally, an ordinary least squares association model combining PWM score and $m^5C$ methylation explained more variance in CTCF occupancy than either factor alone (n = 90,672; PWM β = +0.406; $m^5C$ β = –0.085; Pearson r = 0.48; $R^2 = 0.23$; $p = 1.1 \times 10^{-16}$; Fig. 3C), indicating complementary effects. This outcome aligns with prior results showing that SNPs and variation in $m^5C$ at CTCF sites can affect CTCF binding (Zeng *et al.*, 2023).

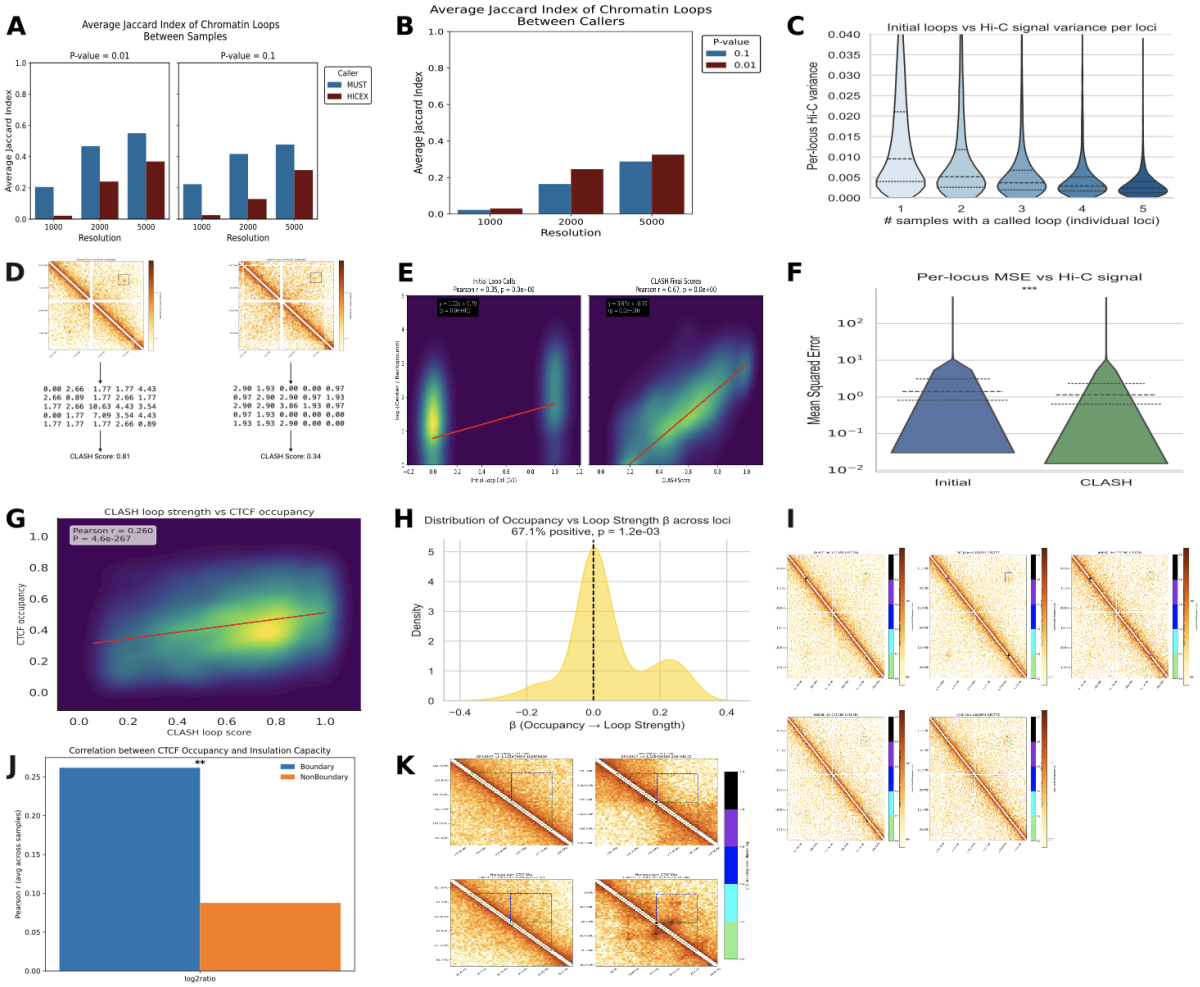



**Figure 4: CLASH improves loop-calls and helps reveal the relationship between CTCF occupancy, loop strength, and TAD organization. a,** Average pairwise Jaccard index of loop calls between samples with a leniency of 10kb, shown separately for Mustache (MUST; blue) and Hi-C-Explorer (HICEX; red), across all resolutions. Values are displayed for loops filtered at significance thresholds of P = 0.1 (left) and P = 0.01 (right). **b,** Average Jaccard index of loop calls within each sample between Mustache and Hi-C-Explorer at matched resolutions, shown for P = 0.1 (blue) and P = 0.01 (red). **c,** Hi-C contact variance, measured by center contact count / local background count, at each locus across samples versus the number of samples with initial loop calls. Calls are pooled across Mustache and Hi-C-Explorer. **d,** Schematic illustration of how CLASH scores loops at the same locus (representative region of chr2:13741565-134936468 for samples GM19347 and HG01457) by extracting the local matrix around the local maximum count. Created in [https://BioRender.com](https://BioRender.com). **e,** Global biological concordance between Hi-C signal vs initial binary loop-scores and Hi-C signal vs CLASH scores shows that CLASH nearly doubles the initial Pearson correlation from r = 0.35 to r = 0.67 ($\Delta r = +0.32$; $p < 10^{-300}$; Fisher

r-to-z). **f,** The per-locus improvement by CLASH compared to initial binary callers quantified using mean squared error (Average ΔMSE per locus = -0.39; $p < 10^{-300}$; Wilcoxon signed-rank test). **g,** Global correlation between CTCF occupancy and loop strength across all samples (n = 17,439, Pearson r = 0.26, $p = 4.6 \times 10^{-267}$). **h,** Per-locus regression of CTCF occupancy vs loop strength for loci with CLASH score standard deviation ≥ 0.27, showing that 67.1% of loci exhibit positive β values (n = 76, p = 0.0012; binomial sign test). **i,** Example locus (chr1:115216363-115527735) across all 5 samples illustrating a positive relationship between CTCF occupancy and loop strength. Both CTCF occupancy and CLASH loopscore are annotated for each sample with light green = 0-0.2, light blue = 0.2-0.4, blue = 0.4-0.6, purple = 0.6-0.8, and black = 0.8-1.0. Loop score is annotated by the rectangle surrounding the loop, and CTCF occupancy of each CTCF loop-contact site is annotated as rectangles along the diagonal. GM19347 has higher CTCF occupancy than other samples, corresponding to an increase in loop-strength. **j,** Average within-sample correlations between CTCF occupancy log2ratio insulation score correlations (representing TAD strength) at boundary vs non-boundary sites in chromosome 1 (boundary n = 740, boundary Pearson r = 0.2617, non-boundary n = 3,063, non-boundary Pearson r = 0.0876; Δr = 0.1741; p = 0.00326, Welch's T test). **k,** Demonstrative example showing results for **j** as four loci within GM19317. CTCF sites classified as boundary sites are presented as the top two panels, while CTCF sites classified as non-boundary are presented at the bottom two panels. CTCF sites with low occupancy are presented on the left, while CTCF sites with high occupancy are presented on the right. Green triangles are overlaid on each Hi-C map representing cis-CTCF site contacts considered, while the blue square represents trans-CTCF site contacts considered. Significance levels are denoted as p < 0.05 (*), p < 0.01 (**), and p < 0.001 (***).

We next sought to quantify the relationship between CTCF occupancy and chromatin loop formation. Loops were identified using Mustache and HiCExplorer at 1, 2, and 5 kb resolution. At 2 kb resolution, the pooled call set contained 18,879 loop loci, but only 5,314 (28%) were anchored by two CTCF sites in at least one haplotype (Supplementary Figure 17). Additionally, reproducibility across samples and callers was low (Figs. 4a,b, Supplementary Figures 18,19), and many loci showed stable Hi-C contact frequencies but inconsistent loop detection (Fig. 4c). These observations demonstrated that conventional loop-calling approaches alone could not robustly capture biological variation or reliably represent cross-sample loop presence.

To overcome these limitations, we developed CLASH, a method that aggregates loop calls across tools and quantifies loop strength directly from the underlying Hi-C matrices as a

continuous score (Fig. 4d). This approach assigns a standardized loop-strength score to every locus in every individual, enabling harmonized, quantitative comparison of loop strength across samples. CLASH substantially improved loop calls across global and per-locus metrics relative to binary callers, including correlation to biological signal ($\Delta r = +0.32$; $p < 10^{-300}$, Fisher r-to-z) and MSE (Average $\Delta$MSE per locus = -0.39; $p < 10^{-300}$; Wilcoxon signed-rank test; Figs. 4e,f Supplementary Figures 20,21), providing a more reliable basis for downstream analyses of how genetic and epigenetic mechanisms modulate chromatin loop formation.

With the 26,750 CLASH-scored loops derived from the 5,314 CTCF-anchored loci, we quantified the association between CTCF occupancy and loop strength. Across all samples, occupancy showed a positive genome-wide correlation with loop strength (n = 17,439; r = 0.26; $p = 4.6 \times 10^{-267}$, Pearson correlation; Fig. 4g). To assess locus-specific effects, we examined loci with high inter-sample variability in CLASH scores (standard deviation $\geq 0.27$). Among these, 67.1% exhibited positive $\beta$ coefficients (n = 76; P = 0.0012, binomial sign test; Figs. 4h,i), indicating a directionally consistent association between higher occupancy and increased loop strength at variable loci. This association is supported by prior work showing that binary changes in CTCF binding coincide with corresponding gains or losses of chromatin loops (Pugacheva *et al.*, 2020). However, to our knowledge, no previous studies have focused on examining continuous quantitative variation in either occupancy or loop strength.

We also tested whether CTCF occupancy affects higher-order chromatin organization at TAD boundaries. Using insulation profiles from Cooltools (Open2C *et al.*, 2024), CTCF sites on chromosome 1 were classified as boundary or non-boundary sites. Chromosome 1 was selected because insulation score calculations were completed for this chromosome whereas several other chromosomes exhibited technical issues preventing score calculations. As chromosome 1 is the largest human chromosome, it provides sufficient representative loci for assessing boundary-associated patterns. Although no significant global or per-locus correlation was observed across samples, within-sample analyses showed a positive association between occupancy and insulation strength, with stronger associations at boundary sites than non-boundary sites (boundary n = 740, boundary r = 0.2617, non-boundary n = 3,063, non-boundary r = 0.0876, Pearson r ; $\Delta r = 0.1741$; p = 0.00326, Welch's T test; Figs. 4j,k). A similar relationship between CTCF occupancy and TAD strength has been reported previously although, like earlier, these studies only examined whether CTCF is present or absent and did not quantify occupancy as a continuous variable (Davidson *et al.*, 2023; Nora *et al.*, 2017).

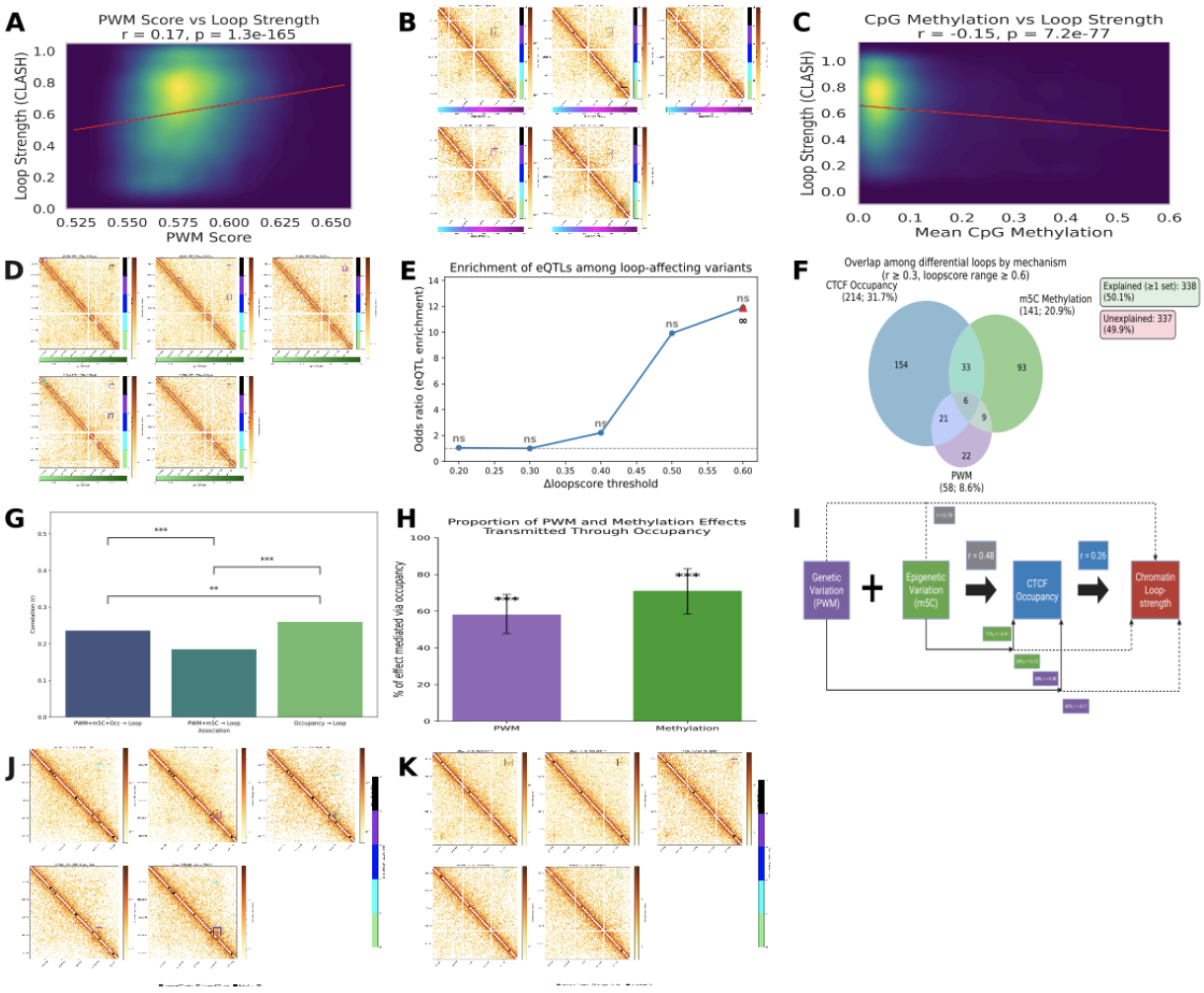



**Figure 5: Direct mechanism by loopscore analysis, eQTL analysis, mediation analysis, and structural variation analysis. a,** Correlation between PWM score (averaged across both haplotypes and both CTCF anchor sites) and loop strength (n = 26,512 loci, r = 0.17, p = 1.3 × $10^{-165}$). **b,** Representative locus (chr2:134,714,609–134,963,448) illustrating the expected relationship between PWM score, CTCF occupancy, and loop strength across samples. In HG01457, the upstream CTCF anchor contains a heterozygous SNP at the 14th base position (G → A), resulting in a reduced PWM score (lighter purple). This nucleotide change corresponds to both the lowest CTCF occupancy at that site and the weakest loop formation across samples. **c,** Correlation between m5C methylation level at CTCF anchor sites and loop strength (n = 15,879 loci, r = –0.15, p = 7.2 × $10^{-77}$). **d,** Representative locus (chr4:85373292–85784871) illustrating the expected relationship between m5C methylation, CTCF occupancy, and loop strength across samples. In HG02666, the upstream CTCF anchor is significantly more methylated (darker green) than the other samples. This corresponds to both the lowest CTCF occupancy at that site and the weakest loop formation across samples. **e,** Enrichment of eQTLs in loop-altering SNPs

(putative iQTLs; n = 668 total SNP-containing CTCF loci) at Δloop-strength thresholds of 0.2 - 0.6. Although the analysis lacks statistical power ($p > 0.05$ for all thresholds) due to small sample size (n = 11 iQTLs at CLASH ≥ 0.4; n = 2 overlapping eQTLs) the observed trend is consistent with expectations. **f,** Proportion of differential loops (n = 675 loci with CLASH score range ≥ 0.6) explained by each mechanism using $r \geq 0.3$ as the explanatory threshold (CTCF occupancy = 31.7%; m5C methylation = 20.9%; PWM score = 8.6%; together explaining 50.1% of differential loops). **g,** Comparison of correlation of loop strength with combined genetic, epigenetic, and occupancy features (left) versus correlation of loop strength with combined genetic and epigenetic features (center) versus correlating loop strength with occupancy alone (right). Including all three mechanisms together reduces the correlation (n = 4,472, $\Delta r = -0.024$, $p = 0.0058$; Steiger test). Excluding occupancy substantially reduces the correlation (n = 4472; $\Delta r = -0.0748$; $p = 2.7 \times 10^{-6}$; Steiger test). **H:** Product of coefficients mediation analysis quantifying the proportion of PWM (58%, $p < 10^{-300}$) and m5C (71%, $p < 10^{-300}$) effects on loop strength that act through CTCF occupancy. Significance was calculated using the Bootstrap-estimated Sobel test. **I:** Schematic summary of the two-factor model: genetic and epigenetic variation modulates CTCF occupancy, which in turn regulates loop strength. Created in https://BioRender.com. **J:** Example locus (chr2: 119456242-119717829) showing a structural-variant insertion that inserts a homozygous CTCF site (light green) in GM19347 (as opposed to the other samples with heterozygous insertions – dark green), leading to increased loop strength in GM19347. **K:** Example locus (chr18: 53883045-54145059) showing a structural-variant deletion (heterozygous – dark red – in GM19347 and HG02666 and homozygous – light red – in HG03248) that removes a CTCF site. Samples without any deletions exhibit strong loop formation, samples with heterozygous deletions vary between strong and minimal loop formation (as expected with not haplotype-resolved Hi-C maps), and homozygous deletions exhibit no loop formation. Significance levels are denoted as $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***).

Having demonstrated each component of the proposed two-factor mechanism – first, that genetic and epigenetic variation alter CTCF occupancy, and second, that occupancy regulates loop strength – we next evaluated whether each mechanism was individually correlated with loop-strength. Using loops anchored by two CTCF sites, we quantified the association between sequence or methylation variation and loop strength, averaging PWM score and $m^5C$ methylation across both anchors. PWM score showed a positive correlation with loop strength (n = 26,512; Pearson $r = 0.17$; $p = 1.3 \times 10^{-165}$; Figure 5A), and $m^5C$ methylation showed a negative correlation (n = 15,879; Pearson $r = -0.15$; $p = 7.2 \times 10^{-77}$; Figure 5C), as expected (Gorkin *et al.*, 2019; Monteagudo-Sánchez *et al.*, 2024). Examples of loci exhibiting these associations are

shown in Figures 5B and 5D. In both cases, the PWM- or methylation-affected CTCF site that lost loop formation also showed the lowest CTCF occupancy across samples.

After showing that SNP's in CTCF sites showed measurable effects on loop-formation, we also sought to determine whether these putative iQTLs were enriched for eQTLs. We intersected these putative iQTLs with GTEx Whole Blood v8 eQTL calls and, although statistical power was limited due to small sample size (iQTL n = 11; eQTL n = 2 at CLASH threshold ≥ 0.4), we observed a consistent increase in eQTL enrichment as loop-strength thresholds were made more stringent (Figure 5E). This trend supports recent results indicating that iQTLs are enriched for eQTLs (Bhattacharyya et al., 2024).

We next quantified how many differential loops could be associated with each mechanism. Among the 675 loci where loop strength varied by more than 0.6 CLASH units, 31.7% showed a correlation between occupancy and loop strength above $r \geq 0.3$, while 20.9% were associated with $m^5C$ methylation and 8.6% with PWM score. In total, these mechanisms collectively accounted for 50.1% of differential loops (Figure 5F).

To determine if occupancy accounts for the observed mechanistic effect on loop-strength, first calculated the joint association between PWM score, $m^5C$ methylation, and occupancy with loop strength, as well as between PWM score and $m^5C$ methylation with loop strength. Including PWM score and $m^5C$ methylation together with occupancy yielded a lower correlation with loop strength than occupancy alone ($n = 4{,}472$; $\Delta r = -0.024$; $p = 0.0058$, Steiger test), while excluding occupancy resulted in a significantly lower correlation with loop strength than occupancy alone ($n = 4472$; $\Delta r = -0.075$; $p = 2.7 \times 10^{-6}$, Steiger test; Figure 5G). We also directly quantified the extent to which variation in loop strength associated with PWM score or $m^5C$ methylation was statistically mediated by occupancy by applying a product-of-coefficients mediation analysis. This analysis indicated that 58% of the association between PWM score and loop strength and 71% of the association between $m^5C$ methylation and loop strength were mediated by occupancy (both $p < 10^{-300}$, Bootstrap-estimated Sobel test; Figure 5H). A summary of these relationships is shown in Figure 5I. To our knowledge, prior studies have not evaluated whether genetic or epigenetic effects on loop strength are statistically mediated through CTCF occupancy.

Because the occupancy-based model does not incorporate structural variants that insert or delete CTCF sites, we separately evaluated whether such variants were associated with differences in loop strength. For each locus where a structural variant created or removed a loop-anchoring CTCF site, we compared average loop scores between samples carrying the variant and those without it. Among 71 loops containing insertions (present in at least one but not all haplotypes), the mean loop-score difference was $-0.008 \pm 0.19$ SD, and among 64 loops containing deletions (present in at least one but not all haplotypes), the mean difference was $-0.015 \pm 0.20$ SD (Supplementary Figure 22). Contrary to the expectation that creating or removing loop-anchoring CTCF sites would consistently alter loop strength, these results

indicate that variants showed minimal average effects at the genome-wide level (Li *et al*., 2024). We further characterized the structural variants affecting loop anchors, with 75.7% being heterozygous and 44% located within 20 kb of another CTCF site (Supplementary Figure 23,24). Despite the lack of a consistent genome-wide effect, several individual loci exhibited clear loop-strength changes associated with variant CTCF sites (Figures 5J–5K).

**Discussion:**

Diffhic:

Our diffHic analysis establishes that inter-individual variation in chromatin interactions is pervasive and reflects contributions from multiple molecular mechanisms, including sequence variation, DNA methylation, and protein occupancy. These findings align with prior results showing that models in which genetic and epigenetic perturbations modulate the local 3D contact landscape and that variation in these features can lead to differential contact interactions between individuals. Although $m^6A$ accessibility has not previously been examined in this context, the observed reduction in contact strength is consistent with increased m6a's indication of reduced protein or histone protection, which would decrease the number of interactions as proteins are required to form chromatin interactions. Ultimately, our diffHic analysis confirmed that inter-individual variation in chromatin structure exists and is significantly influenced by multiple molecular mechanisms, providing the rationale to expand our analysis specifically to chromatin loops.

Occupancy Model:

Although prior studies have independently shown that motif strength, CpG methylation, and CTCF binding each influence genome architecture, none have evaluated these mechanisms jointly or quantified binding effects through occupancy. Our results directly address this gap by demonstrating a two-factor model in which CTCF occupancy acts as the central quantitative mediator linking genetic and epigenetic variation at a CTCF site to differences in chromatin loop strength. First, multivariate association models showed that of 3 models explaining variation in loop-strength, the model using just occupancy as a feature performed significantly better than the model incorporating occupancy, PWM scores, and m5C methylation and the model incorporating PWM scores and m5C methylation, despite PWM scores and m5C methylation individually exhibiting significant correlations with loop-strength. This pattern suggests that the effects of sequence and methylation variation on looping are largely transmitted through their influence on occupancy, which serves as the main contributor to loop-strength. To quantitatively test this, we performed mediation analysis which supports this interpretation by revealing that the majority of PWM scores and m5C methylation's effect on loop-strength is captured by occupancy. This framework explains why loci with substantial sequence or methylation differences between

individuals may show no change in chromatin loop strength, and conversely, why differential loops can arise at the same locus between individuals even in the absence of sequence or methylation differences.

The importance of CTCF occupancy in loop formation is supported by prior studies demonstrating that loop stability depends on the residence time of CTCF–cohesin interactions: single-molecule imaging revealed that CTCF binding and cohesin loading are highly dynamic, with loops forming and dissolving on minute timescales (Hansen *et al.*, 2017), while complementary work showed that cohesin can remain chromatin-bound for hours when stabilized by CTCF and ESCO1, producing long-lived loops (Wutz *et al.*, 2020). The importance of CTCF occupancy in 3D genomics was reinforced by our result that occupancy correlates with TAD insulation strength,

Predictive models such as Akita (Fudenberg *et al.*, 2020) may benefit from incorporating CTCF occupancy data and as a feature in predicting chromatin loop-strength. Further work could also aim to elucidate and quantify the effect to which other regulatory factors that affect CTCF occupancy, as well as the effect of other mechanisms such as cohesion loading and loop extrusion, to enable full prediction of chromatin loop-strength. Future studies of chromatin loop regulation should aim to incorporate direct measurements of protein occupancy, rather than relying solely on assays such as Hi-ChIP which conflate binding with looping and obscure the intermediate mechanistic step. Additionally, future work could extend the framework presented here to examine genetic and epigenetic variation beyond the CTCF motif itself to the flanking regions that regulate CTCF binding, as recent studies indicate that local chromatin context – including adjacent motifs, nucleosome positioning signals, and cofactor-binding elements – can modulate loop formation.

eQTL:

Although our analysis is statistically underpowered, we observed a clear enrichment of putative iQTLs among eQTLs, indicating that SNP's associated with loop strength tend to overlap with SNP's known to modulate gene expression. This pattern is consistent with recent work (Bhattacharyya et al., 2024), and suggests that at least a subset of expression differences across individuals may be mediated through variation in chromatin looping.

Structural Variations:

Although structural variants can create or delete CTCF motifs, we surprisingly found that structural variants showed little global impact on loop strength across individuals. We hypothesized that this could be due to three reasons: (i) the majority of structural variants that affect loops in at least one – but not all – samples are heterozygous, making their effects difficult to resolve without haplotype-specific Hi-C maps (ii) inserted or deleted CTCF sites were located within 20 kb of an existing CTCF site, suggesting that redundant motifs may buffer the structural

impact of structural variants. Prior work has shown that cohesin extrusion is predominantly halted by CTCF sites spaced 10–30 kb apart, meaning that nearby motifs in this range can effectively function as alternative loop anchors (Fudenberg et al., 2016) (iii) inserted CTCF sites may have low CTCF occupancy, which we could not test without haplotype-resolved Fiber-seq assembly data.

Despite these proposed limitations and global buffering, several loci showed clear structural variant-associated loop alterations, demonstrating that structural variants can modulate chromatin structure when they occur in specific regions. To accurately capture structural variant-driven structural changes that are masked in aggregate analyses, future studies could perform a similar analysis using haplotype-resolved Hi-C maps and haplotype-resolved Fiber-seq data.

CLASH:

The loop-strength variation highlighted throughout this study could not be quantified by traditional binary loop callers, underscoring the necessity of CLASH. Unlike existing loop callers that provide binary classifications (“loop” or “no loop”), CLASH converts the underlying Hi-C signal into a single continuous value between 0-1. This shift from detection to measurement enables direct comparisons of subtle loop strength differences across individuals at matched genomic coordinates, which is essential for population-scale analyses. This makes it uniquely suited for applications in identifying loci where subtle changes in loop-strength across samples/individuals are still significant, as in this project. CLASH could be used in future work to explore whether variation in loop strength contributes to disease-associated regulatory differences by testing whether loop strength at specific functional loci may serve as a predictive biomarker in precision-medicine contexts.

A key advantage of CLASH is that it generates reproducible scores directly from the Hi-C matrix, rather than relying on any particular loop-calling algorithm. This design makes CLASH compatible with any set of genomic loci provided by the user, resolving the inconsistencies among current loop-calling methods and ensuring that it remains broadly useful even as loop-calling methods improve. Furthermore, CLASH can serve as an orthogonal benchmark for assessing the robustness of new loop-calling algorithms by providing a continuous, caller-independent measure of signal strength. While CLASH is currently calibrated to operate optimally at 2 kb resolution – reflecting the depth of Hi-C data used here – future work expanding CLASH to support multi-resolution scoring will broaden its applicability across datasets with different sequencing depths.

A limitation of CLASH is the lack of an external ground truth for validating quantitative loop strength. Because no experimental assay provides a fully independent measurement of loop intensity, our validation necessarily relied on Hi-C–derived features that are not fully

independent of the scoring function itself. Future integration with orthogonal datasets may enable more robust and independent validation of CLASH.

Summary:

Our two-factor model provides a mechanistic basis for interpreting how genetic and epigenetic diversity contributes to chromatin architecture across individuals, while CLASH offers a quantitative framework for detecting these effects. Together, these analyses reveal a unified view of inter-individual variation in 3D genome structure: molecular variation primarily affects the potential for CTCF occupancy, which in turn affects chromatin loop-strength and potentially downstream gene regulation. The methods and findings described here lay the groundwork for future work to predict chromatin loop strength from additional biological factors as well as determine whether changes in chromatin loop strength may have functional and health-related consequences.

**Methods:**

Data processing:

All data used corresponded to five lymphoblastoid cell lines, GM19317, GM19347, HG01457, HG02666, and HG03248 (Logsdon et al., 2025). Hi-C data for all samples was generated, aligned to HG38 and processed using the distiller-sm pipeline to resolutions as low as 1 kb. Phased SNP and structural-variant calls for all samples were obtained from the HGSVC Phase 3 dataset (*Ebert et al.*, 2021), generated from telomere-to-telomere haplotype assemblies aligned to hg38, and processed with dipcall (Li *et al.*, 2018) and whatshap (Martin *et al.*, 2023). to annotate CTCF-overlapping insertions, deletions, and SNPs. Single-molecule Fiber-seq reads generated with PacBio HiFi for all samples were aligned, phased, and processed with pbmm2 (Li, 2018), whatshap, and fibertools (Jha *et al.*, 2024) to determine CTCF occupancy, with occupancy defined as the fraction of fibers showing a footprint at each motif. Haplotype-specific CpG methylation tracks were downloaded from HGSVC via globus, aligned to HG38 with minimap2, phased with whatshap, and summarized per genomic bin and per CTCF site.

AB compartment analysation:

A/B compartments were computed from ICE-balanced Hi-C matrices at 100 kb resolution using Cooltools eigs_cis, with GC content used to orient the first eigenvector such that positive values correspond to active (A) chromatin. Compartment similarity across individuals was assessed using pairwise eigenvector sign concordance and mean-squared error.

DiffHiC analysis:

Differential chromatin interactions were identified using DiffHiC (Lun *et al.*, 2015), applying quasi-likelihood tests across all five samples at 1 kb, 2 kb, and 5 kb resolutions and retaining contacts with sufficient coverage (logCPM > –4) and that passed Bonferroni multiple test correction at genomic distances between 10 kb and 1 Mb. Each differential interaction was then annotated for four potential mechanisms – large structural variants, small sequence changes, $m^6A$ adenine methylation, and $m^5C$ CpG methylation – using outlier frameworks to identify the sample or haplotype exhibiting the most extreme deviation for each feature. For each mechanism, we quantified whether the sample with the strongest molecular deviation corresponded to the sample with the largest Hi-C log fold-change and assessed significance using binomial tests. Loci explained by multiple mechanisms were removed to confirm that each mechanism remained significant.

CTCF site determination and PWM score calculation:

Because 85% of chromatin loops form at CCCTC-binding factor (CTCF; Rao *et al.*, 2014), CTCF sites were identified in each haplotype by running FIMO (Grant *et al.*, 2011) on haplotype-resolved assemblies and mapping the resulting coordinates back to hg38 using the Long Read Aligner software (Ren & Chaisson, 2021), yielding ~50,000 sites per haplotype. This matches expected CTCF site counts per haplotype (ENCODE, 2012).

For each motif, the underlying 19-bp sequence was scored using a position weight matrix (PWM) derived from the JASPAR MA0139.1 log-odds matrix (Khan et al., 2018), producing a continuous measure of motif strength that reflects the impact of SNPs and small indels. PWM scores were computed by summing position-specific weights and normalizing by motif length to facilitate comparison across haplotypes and individuals. Sequence mismatches relative to the canonical 15-bp CTCF consensus motif were enumerated by aligning each haplotype sequence to the consensus and counting non-matching bases.

CTCF site m5C CpG methylation:

CpG methylation levels were quantified at single-base resolution from phased ONT methylation calls, with per-position methylation fractions computed as the proportion of reads supporting a methylated cytosine. For each haplotype-resolved CTCF motif, all CpG positions within the motif boundaries were extracted and averaged to obtain a site-level $m^5C$ methylation value. Additionally, a genome-wide hidden Markov model was used to classify each CpG as hypo- or hypermethylated, and these state assignments were integrated with CTCF annotations.

Quantifying correlations between genetic and epigenetic factors and occupancy:

Each CTCF site across all ten haplotypes was annotated with its PWM score, number of sequence mismatches, CpG methylation level and state, and CTCF occupancy value. Global pairwise correlations were computed to assess how each genetic or epigenetic feature individually relates to CTCF occupancy. To evaluate their combined contributions, we fit an ordinary least squares model using PWM score and CpG methylation as joint predictors of occupancy after appropriate normalization. Because the goal was to quantify association rather than perform prediction, the model was fit to the full dataset and significance was evaluated using standard linear regression statistics.

Calling loops with Mustache and Hi-C Explorer:

Chromatin loops were initially identified using both Mustache (Roayaei *et al.*, 2020) and HiCExplorer (Wolff *et al.*, 2020) across 1 kb, 2kb, and 5kb resolutions filtering for p values of 0.1 and 0.01 to generate a comprehensive union of candidate loop calls for each sample. To assess agreement between callers and across individuals, we compared loop sets using Jaccard indices after standardizing loop coordinates within a small genomic window. These values were interpreted in the context of prior work showing that ~75% of chromatin loops are conserved between human samples (Pękowska *et al.*, 2018). Additionally, we computed the variance in Hi-C signal vs the number of samples that the initial callers called a loop in.

CLASH:

We developed CLASH, a quantitative loop-scoring framework that re-evaluates pooled loop calls across samples by extracting an adaptively sized Hi-C submatrix around each candidate loop and normalizing interaction strength to a distance-matched background. For each sample, CLASH identifies a refined loop center, adjusts the extraction window based on local matrix structure, and applies depth normalization to ensure comparability across genomes. The method integrates center enrichment, hierarchical decay patterns, and matrix coherence into a unified 0–1 score that reflects the Hi-C signal of each loop. These continuous scores replace inconsistent binary loop calls and enable robust cross-sample comparison of loop strength. Full algorithmic details, scoring functions, and parameter choices are provided in the Supplementary Methods.

CLASH validation:

CLASH scores were compared to initial pooled Mustache/HiCExplorer binary loop calls using multiple quantitative metrics. We compared their global correlation with local Hi-C signal, evaluated differences in explained variance using nested linear models, assessed improvements in mutual information via bootstrap resampling and computed per-locus mean-squared error between predicted and observed Hi-C signal across samples. The proportion of loci showing

improved fit under CLASH was quantified, and paired improvements were tested using Wilcoxon signed-rank tests.

Quantifying the correlation between CTCF occupancy and CLASH loop-score:

To assess how CTCF protein binding relates to chromatin loop strength, we restricted analysis to loops whose two anchor bins each contained a CTCF site within 10 kb and computed a per-loop occupancy value by averaging the Fiber-seq–derived occupancy of its two anchors. These occupancy values were paired with CLASH loop scores across all samples to quantify the global relationship between CTCF binding and loop intensity. To evaluate locus-specific effects, we performed a per-locus regression analysis in which the association between occupancy values across the five samples and the CLASH loop scores was quantified, retaining only loci with sufficient occupancy and loop-score measurements and adequate loop-score variance. For each locus, we obtained a shrinkage-stabilized effect-size coefficient (β), which is more reliable than Pearson correlation in the small-n (n=5) setting. We then tested whether β coefficients were significantly positive using both a one-sample t-test and a sign test and summarized their distribution across all loci.

Quantifying the correlation between CTCF occupancy and CTCF insulation capability:

To evaluate how CTCF occupancy affects a CTCF site's ability to function as an insulator, insulating CTCF sites on chromosome 1 for each sample were identified using cooltools insulation. Identified sites were classified as either boundary sites, and all the other CTCF sites were classified as non-boundary sites. For each CTCF site, we computed a quantitative insulation score by extracting a 200-kb Hi-C window around the site and calculating an insulation score, based on an existing function (Crane *et al.*, 2015), comparing upstream/downstream interaction versus cross-boundary interactions. Within each sample, Pearson correlations between occupancy and insulation score were computed across all CTCF sites, and the difference between boundary and non-boundary correlations was assessed using a two-sided Welch's t-test.

Quantifying the correlation between PWM scores, m5C methylation, CTCF occupancy and CLASH loop-scores:

Per-loop PWM and CpG methylation values were obtained by averaging anchor-level measurements and pairing them with CLASH loop scores and occupancy. We quantified individual and joint associations using Pearson correlations and Steiger's test for dependent correlations, enabling direct evaluation of whether PWM and m5C methylation contributed explanatory power beyond occupancy alone, including in a model excluding occupancy.

Mediation analysis:

We quantified the extent to which CTCF occupancy mediates the effects of PWM motif strength and CpG methylation on loop strength using a correlation-based mediation analysis, in which indirect, direct, and proportion-mediated effects were derived from pairwise correlations and their significance assessed through bootstrap resampling.

eQTL analysis:

To assess whether SNPs within loop-anchor CTCF sites that modulate loop strength act as interaction QTLs (iQTLs), we identified anchor SNP–loop pairs and tested whether allele differences were associated with changes in CLASH loop scores across individuals. We then compared these iQTL candidates to the GTEx v8 Whole Blood significant eQTL catalog (GTEx Consortium, 2020) to determine whether loop-modulating variants were enriched for known expression QTLs. Enrichment across increasing loop-strength difference thresholds was quantified by computing the log odds ratio of eQTL overlap.

Structural variation analysis:

To assess whether structural variants (SVs) that add or remove CTCF motifs influence chromatin loop formation, we identified insertions and deletions overlapping CTCF sites that serve as loop-anchor regions and quantified the resulting gain or loss of motif instances. For each loop and individual, we compared CLASH loop scores between samples carrying a CTCF-altering SV and those without, and evaluated the direction and magnitude of loop-strength changes across the genome. We additionally examined SV zygosity and the presence of nearby "backup" CTCF motifs to partially explain the observed lack of genome-wide trend.

**References:**


1) Bhattacharyya, S., Ay, F. Identifying genetic variants associated with chromatin looping and genome function. *Nat Commun* 15, 8174 (2024). https://doi.org/10.1038/s41467-024-52296-4

2) Bond, M. L., Davis, E. S., Quiroga, I. Y., Dey, A., Kiran, M., Love, M. I., Won, H., & Phanstiel, D. H. (2023). Chromatin loop dynamics during cellular differentiation are associated with changes to both anchor and internal regulatory features. Genome research, 33(8), 1258–1268. https://doi.org/10.1101/gr.277397.122

3) Chowdhury, H. M. A. M., Boult, T., & Oluwadare, O. (2024). Comparative study on chromatin loop-callers using Hi-C data reveals their effectiveness. BMC bioinformatics, 25(1), 123. https://doi.org/10.1186/s12859-024-05713-w

4) Crane, E., Bian, Q., McCord, R. *et al.* Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* 523, 240–244 (2015). https://doi.org/10.1038/nature14450

5) Davidson, I. F., Barth, R., Zaczek, M., van der Torre, J., Tang, W., Nagasaka, K., Janissen, R., Kerssemakers, J., Wutz, G., Dekker, C., & Peters, J. M. (2023). CTCF is a DNA-tension-dependent barrier to cohesin-mediated loop extrusion. *Nature*, *616*(7958), 822–827. https://doi.org/10.1038/s41586-023-05961-5

6) Ebert, P., Audano, P. A., Zhu, Q., Rodriguez-Martin, B., Porubsky, D., Bonder, M. J., Sulovari, A., Ebler, J., Zhou, W., Serra Mari, R., Yilmaz, F., Zhao, X., Hsieh, P., Lee, J., Kumar, S., Lin, J., Rausch, T., Chen, Y., Ren, J., Santamarina, M., … Eichler, E. E. (2021). Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science (New York, N.Y.)*, *372*(6537), eabf7117. https://doi.org/10.1126/science.abf7117

7) ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature, 489(7414), 57–74. https://doi.org/10.1038/nature11247

8) Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., & Mirny, L. A. (2016). Formation of Chromosomal Domains by Loop Extrusion. *Cell reports*, *15*(9), 2038–2049. https://doi.org/10.1016/j.celrep.2016.04.085

9) Fudenberg, G., Kelley, D. R., & Pollard, K. S. (2020). Predicting 3D genome folding from DNA sequence with Akita. *Nature Methods, 17*(11), 1111–1117. https://doi.org/10.1038/s41592-020-0958-x

10) Chen, W., Zeng, Y. C., Achinger-Kawecka, J., Campbell, E., Jones, A. K., Stewart, A. G., Khoury, A., & Clark, S. J. (2024). Machine learning enables pan-cancer identification of mutational hotspots at persistent CTCF binding sites. *Nucleic acids research*, *52*(14), 8086–8099. https://doi.org/10.1093/nar/gkae530

11) Gong, Y., Lazaris, C., Sakellaropoulos, T. et al. Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries. Nat Commun 9, 542 (2018). https://doi.org/10.1038/s41467-018-03017-1

12) Grant, C. E., Bailey, T. L., & Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics (Oxford, England)*, *27*(7), 1017–1018. https://doi.org/10.1093/bioinformatics/btr064

13) Greenwald, W. W., Li, H., Benaglio, P., Jakubosky, D., Matsui, H., Schmitt, A., Selvaraj, S., D'Antonio, M., D'Antonio-Chronowska, A., Smith, E. N., & Frazer, K. A. (2019). Subtle changes in chromatin loop contact propensity are associated with differential gene regulation and expression. *Nature communications*, *10*(1), 1054. https://doi.org/10.1038/s41467-019-08940-5

14) GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science (New York, N.Y.)*, *369*(6509), 1318–1330. https://doi.org/10.1126/science.aaz1776

15) Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R., & Darzacq, X. (2017). CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *eLife*, *6*, e25776. https://doi.org/10.7554/eLife.25776

16) Hashimoto, H., Wang, D., Horton, J. R., Zhang, X., Corces, V. G., & Cheng, X. (2017). Structural Basis for the Versatile and Methylation-Dependent Binding of CTCF to DNA. Molecular cell, 66(5), 711–720.e3. https://doi.org/10.1016/j.molcel.2017.05.004

17) Holwerda, S., & de Laat, W. (2012). Chromatin loops, gene positioning, and gene expression. Frontiers in genetics, 3, 217. https://doi.org/10.3389/fgene.2012.00217

18) Jha, A., Bohaczuk, S. C., Mao, Y., Ranchalis, J., Mallory, B. J., Min, A. T., Hamm, M. O., Swanson, E., Dubocanin, D., Finkbeiner, C., Li, T., Whittington, D., Noble, W. S., Stergachis, A. B., & Vollger, M. R. (2024). DNA-m6A calling and integrated long-read epigenetic and genetic analysis with fibertools. Genome Research. https://doi.org/10.1101/gr.279095.124

19) Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J. A., van der Lee, R., Bessy, A., Chèneby, J., Kulkarni, S. R., Tan, G., Baranasic, D., Arenillas, D. J.,

Sandelin, A., Vandepoele, K., Lenhard, B., Ballester, B., Wasserman, W. W., Parcy, F., & Mathelier, A. (2018). JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic acids research*, *46*(D1), D260–D266. https://doi.org/10.1093/nar/gkx1126

20) Li, C., Bonder, M. J., Syed, S., Jensen, M., Human Genome Structural Variation Consortium (HGSVC), HGSVC Functional Analysis Working Group, Gerstein, M. B., Zody, M. C., Chaisson, M. J. P., Talkowski, M. E., Marschall, T., Korbel, J. O., Eichler, E. E., Lee, C., & Shi, X. (2024). An integrative TAD catalog in lymphoblastoid cell lines discloses the functional impact of deletions and insertions in human genomes. *Genome research*, *34*(12), 2304–2318. https://doi.org/10.1101/gr.279419.124

21) Li H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics (Oxford, England)*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

22) Li H, Bloom JM, Farjoun Y, Fleharty M, Gauthier L, Neale B, MacArthur D (2018) A synthetic-diploid benchmark for accurate variant-calling evaluation. Nat Methods, 15:595-597. [PMID:30013044]

23) Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., & Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (New York, N.Y.)*, *326*(5950), 289–293. https://doi.org/10.1126/science.1181369

24) Logsdon, G. A., Ebert, P., Audano, P. A., Loftus, M., Porubsky, D., Ebler, J., Yilmaz, F., Hallast, P., Prodanov, T., Yoo, D., Paisie, C. A., Harvey, W. T., Zhao, X., Martino, G. V., Henglin, M., Munson, K. M., Rabbani, K., Chin, C-S., Gu, B., Ashraf, H., Scholz, S., Austine-Orimoloye, O., Balachandran, P., Bonder, M. J., Cheng, H., Chong, Z., Crabtree, J., Gerstein, M., Guethlein, L. A., Hasenfeld, P., Hickey, G., Hoekzema, K., Hunt, S. E., Jensen, M., Jiang, Y., Koren, S., Kwon, Y., Li, C., Li, H., Li, J., Norman, P. J., Oshima, K. K., Paten, B., Phillippy, A. M., Pollock, N. R., Rausch, T., Rautiainen, M., Song, Y., Söylev, A., Sulovari, A., Surapaneni, L., Tsapalou, V., Zhou, W., Zhou, Y., Zhu, Q., Zody, M. C., Mills, R. E., Devine, S. E., Shi, X., Talkowski, M. E., Chaisson, M. J. P., Dilthey, A. T., Konkel, M. K., Korbel, J. O., Lee, C., Beck, C. R., Eichler, E. E., Marschall, T. (2025). Complex genetic variation in nearly complete human genomes. *Nature*, 644(8076), 430-441. https://doi.org/10.1038/s41586-025-09140-6

25) Lun, A.T., Smyth, G.K. diffHic: a Bioconductor package to detect differential genomic interactions in Hi-C data. *BMC Bioinformatics* 16, 258 (2015). https://doi.org/10.1186/s12859-015-0683-0

26) Martin, M., Ebert, P., & Marschall, T. (2023). Read-Based Phasing and Analysis of Phased Variants with WhatsHap. *Methods in molecular biology (Clifton, N.J.)*, *2590*, 127–138. https://doi.org/10.1007/978-1-0716-2819-5_8

27) Misteli T. (2020). The Self-Organizing Genome: Principles of Genome Architecture and Function. *Cell*, *183*(1), 28–45. https://doi.org/10.1016/j.cell.2020.09.014

28) Monteagudo-Sánchez, A., Richard Albert, J., Scarpa, M., Noordermeer, D., & Greenberg, M. V. C. (2024). The impact of the embryonic DNA methylation program on CTCF-mediated genome regulation. *Nucleic acids research*, *52*(18), 10934–10950. https://doi.org/10.1093/nar/gkae724

29) Mumbach, M. R., Rubin, A. J., Flynn, R. A., Dai, C., Khavari, P. A., Greenleaf, W. J., & Chang, H. Y. (2016). HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nature methods*, *13*(11), 919–922. https://doi.org/10.1038/nmeth.3999

30) Nora, E. P., Goloborodko, A., Valton, A. L., Gibcus, J. H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L. A., & Bruneau, B. G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell*, *169*(5), 930–944.e22. https://doi.org/10.1016/j.cell.2017.05.004

31) Open2C. *distiller-sm: A modular Hi-C mapping pipeline for reproducible data analysis.* GitHub repository. Available at: https://github.com/open2c/distiller-sm (accessed November 15, 2025).

32) Open2C, Abdennur, N., Abraham, S., Fudenberg, G., Flyamer, I. M., Galitsyna, A. A., Goloborodko, A., Imakaev, M., Oksuz, B. A., Venev, S. V., & Xiao, Y. (2024). Cooltools: Enabling high-resolution Hi-C analysis in Python. PLoS computational biology, 20(5), e1012067. https://doi.org/10.1371/journal.pcbi.1012067

33) Panarotto, M., Davidson, I. F., Litos, G., Schleiffer, A., & Peters, J. M. (2022). Cornelia de Lange syndrome mutations in NIPBL can impair cohesin-mediated DNA loop extrusion. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(18), e2201029119. https://doi.org/10.1073/pnas.2201029119

34) Pękowska, A., Klaus, B., Xiang, W., Severino, J., Daigle, N., Klein, F. A., Oleś, M., Casellas, R., Ellenberg, J., Steinmetz, L. M., Bertone, P., & Huber, W. (2018). Gain of CTCF-Anchored Chromatin Loops Marks the Exit from Naive Pluripotency. Cell systems, 7(5), 482–495.e10. https://doi.org/10.1016/j.cels.2018.09.003

35) Pugacheva, E. M., Kubo, N., Loukinov, D., Tajmul, M., Kang, S., Kovalchuk, A. L., Strunnikov, A. V., Zentner, G. E., Ren, B., & Lobanenkov, V. V. (2020). CTCF mediates chromatin looping via N-terminal domain-dependent cohesin retention. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(4), 2020–2031. https://doi.org/10.1073/pnas.1911708117

36) Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S., & Aiden, E. L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, *159*(7), 1665–1680. https://doi.org/10.1016/j.cell.2014.11.021

37) Roayaei Ardakany, A., Gezer, H. T., Lonardi, S., & Ay, F. (2020). Mustache: Multi-scale detection of chromatin loops from Hi-C and Micro-C maps using scale-space representation. *Genome Biology, 21*, 256. https://doi.org/10.1186/s13059-020-02167-0

38) Sabaté, T., Lelandais, B., Bertrand, E., & Zimmer, C. (2023). Polymer simulations guide the detection and quantification of chromatin loop extrusion by imaging. *Nucleic acids research*, *51*(6), 2614–2632. https://doi.org/10.1093/nar/gkad034

39) Salameh, T.J., Wang, X., Song, F. et al. A supervised learning framework for chromatin loop detection in genome-wide contact maps. Nat Commun 11, 3428 (2020). https://doi.org/10.1038/s41467-020-17239-9

40) Stergachis, A. B., Debo, B. M., Haugen, E., Churchman, L. S., & Stamatoyannopoulos, J. A. (2020). Single-molecule regulatory architectures captured by chromatin fiber sequencing. *Science (New York, N.Y.)*, *368*(6498), 1449–1454. https://doi.org/10.1126/science.aaz1646

41) Wang, J., Wu, L., Wei, J. et al. CGLoop: a neural network framework for chromatin loop prediction. BMC Genomics 26, 342 (2025). https://doi.org/10.1186/s12864-025-11531-y

42) Wolff, J., Rabbani, L., Gilsbach, R., Richard, G., Manke, T., Backofen, R., & Grüning, B. A. (2020). Galaxy HiCExplorer 3: A web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization. *Nucleic Acids Research, 48*(W1), W177–W184. https://doi.org/10.1093/nar/gkaa220

43) Wutz, G., Ladurner, R., St Hilaire, B. G., Stocsits, R. R., Nagasaka, K., Pignard, B., Sanborn, A., Tang, W., Várnai, C., Ivanov, M. P., Schoenfelder, S., van der Lelij, P., Huang, X., Dürnberger, G., Roitinger, E., Mechtler, K., Davidson, I. F., Fraser, P., Lieberman-Aiden, E., & Peters, J. M. (2020). ESCO1 and CTCF enable formation of long chromatin loops by protecting cohesinSTAG1 from WAPL. *eLife*, *9*, e52091. https://doi.org/10.7554/eLife.52091

44) Xu, J., Xu, X., Huang, D. et al. A comprehensive benchmarking with interpretation and operational guidance for the hierarchy of topologically associating domains. Nat Commun 15, 4376 (2024). https://doi.org/10.1038/s41467-024-48593-7

45) Yoon, I., Kim, U., Jung, K. O., Song, Y., Park, T., & Lee, D. S. (2024). 3C methods in cancer research: recent advances and future prospects. Experimental & molecular medicine, 56(4), 788–798. https://doi.org/10.1038/s12276-024-01236-9

46) Zeng, Y., Jain, R., Lam, M., Ahmed, M., Guo, H., Xu, W., Zhong, Y., Wei, G. H., Xu, W., & He, H. H. (2023). DNA methylation modulated genetic variant effect on gene transcriptional regulation. *Genome biology*, *24*(1), 285. https://doi.org/10.1186/s13059-023-03130-5

**Supplemental Methods:**

Data processing:

Hi-C data were generated for five lymphoblastoid cell lines (GM19317, GM19347, HG01457, HG02666, and HG03248), each sequenced in four independent runs. Raw FASTQ files were aligned to the GRCh38 (hg38) reference genome and processed using the distiller-sm pipeline (Open2C). PCR duplicates, self-circles, and dangling ends were removed following standard Hi-C quality-control steps. Balanced contact matrices were generated at 1 kb, 2 kb, and 5 kb resolutions using the cooler (v0.10.3) framework and iterative correction (ICE) was applied for normalization.

Phased variant call files (VCFs) containing both single-nucleotide polymorphisms and structural variants were obtained from the HGSVC Phase 3 dataset. These VCFs were produced by aligning telomere-to-telomere haplotype assemblies generated from PacBio HiFi and Oxford Nanopore reads to hg38, calling variants with dipcall (v0.3) and phasing with whatshap (v2.8). The resulting phased VCFs were used to annotate deletions, insertions, and SNPs overlapping CTCF motifs and chromatin loops.

Single-molecule chromatin Fiber-seq data were produced using PacBio Sequel II HiFi sequencing (30 h movie time per SMRT Cell) at the University of Washington. Each sample achieved a mean genome-wide coverage of 33.5x, an average read length of 20.7 kb, HiFi yields of ~103 Gb, and read-quality scores ranging from Q30–Q33. See Table S1B for full Fiber-seq statistics. Reads containing pre-annotated $N^6$-methyladenine ($m^6A$) modifications were aligned to hg38 using pbmm2 (v1.10.0). Resulting BAM files were phased using whatshap phase guided by corresponding variant calls. Processed reads were analyzed using the standard fibertools command suite (v0.6.4) pipeline to identify nucleosomes. Footprinting was performed against known CTCF motif coordinates (JASPAR MA0139.1). For each motif, both the total number of fibers spanning the site and the number containing a CTCF-sized footprint were counted. The ratio of footprinted to total fibers defined the CTCF occupancy frequency for that site.

Per-base 5mC CpG methylation calls for all five samples were obtained from the Human Genome Structural Variation Consortium (HGSVC) via Globus. Base-called reads were aligned to each sample's telomere-to-telomere assembly using minimap2 (v2.30-r1287). Methylation profiles were phased using whatshap phase (v2.8) guided by dipcall-derived VCFs, yielding haplotype-specific methylation tracks. For downstream integration with Hi-C and Fiber-seq, the number of methylated CpG sites per fiber per genomic bin and per CTCF site was used as a proxy for total methylation signal.

AB compartment analysation:

A/B compartments were computed using the Cooltools (v0.7.0) eigs_cis function applied to balanced Hi-C matrices at 100 kb resolution. GC content was used as the phasing track to orient the first eigenvector (E1), such that positive values correspond to GC-rich, transcriptionally active (A) compartments. The resulting E1 eigenvalues for each genomic bin were compared between samples using pairwise mean squared error (MSE) and sign concordance to quantify compartment similarity. For visualization, E1 profiles across a representative 35 Mb region were plotted for all samples. Hi-C maps showing compartment state and inter-sample compartment changes were generated using custom *matplotlib* utilities from sample cooler files. For all AB compartment, TAD domain, chromatin loop, and chromatin interaction-related analysis in this manuscript, we highlight representative loci and regions that illustrate the general trends observed genome-wide.

Identification of differential interactions using diffHiC:

Genome-wide differential chromatin interactions between the 5 lymphoblastoid samples were identified using the DiffHiC package in R (v1.38.0). For each chromosome, raw Hi-C contact matrices for the 5 samples were generated at 1 kb, 2 kb, and 5 kb resolution, and merged into one spreadsheet. The glmQLFit quasi-likelihood model was run on this spreadsheet five times, each comparing one sample against the remaining four. Resulting interaction lists were filtered to retain bins with logCPM > –4 and interaction distances between 10 kb and 1 Mb. The logCPM threshold was selected based on the distributions of logCPM and logFC values, optimizing retention of informative contacts while preserving the expected unimodal logFC distribution centered around zero at 1 kb resolution. Across filtered loci, logFC standard deviations followed an approximately negative-binomial distribution with a peak near $\sigma \approx 0.2$, reflecting that most interactions are conserved and a minority are differential.
Significant interactions were identified using Bonferroni correction at each resolution, and genome-wide logFC profiles were visualized for each sample. The set of 382 significant interactions at 5 kb was used for mechanistic analyses.

Mechanistic analysis of diffhic differential interactions:

Four mechanisms were evaluated for their association with differential Hi-C signal:

1. Large-scale sequence differences between anchor bins (structural variation)
2. Small-scale sequence differences within anchor bins (SNPs and indels)
3. $m^6A$ adenine methylation (Fiber-seq)
4. $m^5C$ CpG methylation (ONT)

For genetic variation, a leave-one-out (LOO) approach was used to identify the sample with the most extreme value at each differential bin. For each bin containing two anchors that represent an interaction locus, the sample with a differential number of bases changed (separately including total number of bases changed, insertions, and deletions) both between anchors and in anchors was determined, as well as the sample with a differential logFC value. Samples were considered differential if they exhibited a MAD-based Z-score $\geq 2$ and their LOO t-test p-value fell below a Bonferroni-adjusted 0.02 threshold. This was chosen as the threshold because it corresponds to strong outlier deviation in a MAD-based framework while maintaining sensitivity given the sample size ($n = 5$). For bins where both (i) a differential genetic-alteration sample and (ii) a differential Hi-C logFC sample were identified, we quantified the match rate between the two. Statistical significance was assessed using a two-sided binomial test under a null probability of 0.2 (reflecting the 1-in-5 chance of a match by random expectation).

M6A methylation rates were computed by determining, for each fiber and each bin, the number of methylated adenines divided by total adenines. Rates were averaged across fibers and across both interacting bins for each differential Hi-C contact. To normalize baseline differences between individuals, genome-wide expected methylation ratios were computed and compared to observed ratios. Bins with observed log-ratios deviating from expectation by > 0.2 (log space)

were flagged as differential, and the direction of the deviation was recorded. Because increased m6A methylation is expected to correlate with reduced chromatin contacts, loci were classified into four categories (+m6A, +logFC; +m6A, -logFC; -m6A, +logFC; -m6A, -logFC) based on whether the differential-methylation sample aligned with the minimum or maximum logFC value: The match rate of each set was calculated and statistical significance was assessed using a two-sided binomial test under a null probability of 0.2.

CpG methylation levels were obtained from phased ONT reads. For each resolution (1 kb, 2 kb, 5 kb), per-bin methylation fractions were computed as the number of methylated cytosines divided by the total number of cytosines within that bin, separately for each haplotype (H1, H2). The methylation fractions of the two anchor bins were averaged to obtain a single value per haplotype per differential interaction. To identify whether a sample exhibited aberrant m5C levels at a locus, we applied a similar robust outlier framework. Across the 10 haplotypes, methylation values were converted into median–absolute-deviation (MAD) Z-scores, and the haplotype with the largest absolute Z-score was considered a candidate differential sample. A leave-one-out (LOO) Z-score was then computed by recalculating the median and MAD excluding the candidate haplotype-sample. A haplotype was marked as differential if both the MAD Z-score and the LOO Z-score were $\geq 2$, and the direction of the deviation ("+" for higher methylation, "–" for lower) was recorded. For each differential m5C event, we compared the direction of methylation deviation with the haplotype's sample's Hi-C log fold-change (logFC). Because increased CpG methylation is expected to reduce chromatin contacts, loci were classified into four categories (+m5C, +logFC; +m5C, -logFC; -m5C, +logFC; -m5C, -logFC) based on whether the differential-methylation sample aligned with the minimum or maximum logFC value. For each category, we computed the proportion of loci in which the m5C deviation correctly predicted the logFC extremum and assessed significance using a two-sided binomial test (null probability 0.2).

To assess joint explanatory power, differential bins were annotated with any mechanism for which a differential sample was detected. To evaluate independence, match-rate analyses (total bases changed between anchors and in anchors; +m6a, -logFC; +m5C, - logFC) were recomputed after removing any differential interactions explained by 2 or more mechanisms, demonstrating that each mechanism retains significant predictive power when controlling for the others. A Venn diagram generated using the venn Python package (v0.1.3) summarized the overlap between mechanisms match rate in bins that exhibited at least one differential mechanism. We also computed the frequency with which each genomic bin appeared across all differential interactions and later quantified the proportion that fell within 10 kb of a chromatin loop anchor.

CTCF site determination and PWM score calculation:

Because 85% of chromatin loops form at CCCTC-binding factor (CTCF), CTCF motif locations in each sample was derived by running the Fimo software (v5.3.0) on haplotype-resolved assemblies with default parameters for each sample and mapping the results back to the reference genome HG38 using the Long Read Aligner software (v1.3.7.2). This process successfully mapped ~90% of CTCF sites for each sample back to the reference genome (averaging ~ 49,983 sites per haplotype). This matches expected CTCF site counts per haplotype.

The Fimo output also included the base sequence for each CTCF site. The position weight matrix (PWM) scores were calculated for each sequence for each haplotype. A PWM matrix was constructed from JASPAR's 2018 CTCF frequency matrix by using the provided log-odds weights for each base and position in the MA0139.1 motif. For each haplotype-resolved CTCF site, the corresponding 19-bp sequence was scored by summing the position-specific weights associated with each base and dividing the total by motif length to obtain an average per-position motif score, which enables direct comparison across the ten haplotypes in our dataset and integrates naturally with downstream quantitative models.

For each CTCF sequence, the number of bases mutated compared to the canonical 15-bp CTCF consensus sequence 5′-NCANNAGRNGGCRSY-3′ (IUPAC) was calculated. The consensus was expanded into explicit base-allowance rules, and the 15-bp window was aligned across each haplotype sequence to identify the highest-scoring local match. Within the best alignment, bases that violated the consensus rule were counted as mismatches, and the number of mismatches per sequence was recorded.

CTCF site m5C methylation:

CpG methylation was quantified at single-base resolution using ONT-derived per-nucleotide methylation calls for each haplotype. This data was processed to determine for each genomic position, the number of reads overlapping that cytosine and the number of reads supporting a methylated call for each sample and haplotype. For each position, a methylation fraction was computed as methylated/total. For every haplotype-resolved CTCF motif, all CpG positions falling between the start and end coordinates were retrieved. Multiple CpG methylations on the same site were averaged, producing a site-level methylation profile for every motif and haplotype.

In parallel, a hidden Markov model (HMM) was applied genome-wide to classify each CpG position as hypomethylated or hypermethylated. These HMM state calls were merged into the CTCF annotations using the same coordinate-based procedure, yielding for each haplotype-specific CTCF site a categorical label of hypomethylated, hypermethylated, or mixed (if CpG positions within the motif belonged to different HMM states).

Quantifying correlations between genetic and epigenetic factors and occupancy:

Each CTCF site across all ten haplotypes was annotated with the computed PWM score, the number of sequence mutations relative to the consensus motif, the mean CpG methylation level, the CpG methylation state (hypo, mixed, hyper), and the Fiber-seq–derived CTCF occupancy value. Global pairwise correlations between each feature and occupancy were computed to assess individual relationships. To quantify the joint association between sequence strength and CpG methylation and their relationship to CTCF occupancy, we fit an ordinary least squares (OLS) regression model using PWM score and methylation as predictors of occupancy. PWM scores were min–max normalized to the range [0,1], whereas methylation values were z-scored across all motif instances. Rows containing missing values for either predictor or occupancy were excluded. Because the goal of this analysis was association rather than prediction, the model was fit to the full dataset without a train–test split. Global significance of the association model was assessed using the F-test for linear regression with two predictors. All analyses were performed in Python using pandas, numpy, scikit-learn, scipy, matplotlib, and seaborn.

Calling loops with Mustache and Hi-C Explorer:

Chromatin loops were initially called on each resolution of each sample using both Mustache (v1.3.3; MUST) and Hi-C Explorer (v3.7.5; HICEX) software with p values of both 0.1 and 0.01. Mustache loops called at 1 kb and 2 kb resolutions included an additional st parameter which was set at 0.7 following recommended settings. All Hi-C Explorer loops called with the detect loops function included additional parameters of maxLoopDistance = 2000000, windowSize = 10, and peakWidth = 6. To compare loop sets, we computed Jaccard indices between callers and between samples after expanding each loop to a ±10 kb neighborhood to account for minor positional differences in peak localization. These values were interpreted in the context of prior work showing that ~75% of chromatin loops are conserved between human samples. The same comparisons were also performed on two independent GM19317 technical replicates.

To further assess loop-calling validity, a violin plot comparing the distributions of Hi-C variation per locus across all 5 samples against the number of samples in which Mustache or Hi-C Explorer initially called loops was then generated. To determine the Hi-C variance at each sample, the Mustache and Hi-C Explorer loop-call sets for each sample at resolution = 2kb and p = 0.1 were first pooled. Multiple loops within a 50 kb window of each other were merged, and for each sample at each loop locus, the maximum contact count in the Hi-C contact map within a dynamic search window ranging from a 5×5 matrix (±2 bins) for short-range loops (< 100 kb) up to an 11×11 matrix (±5 bins) for long-range loops (> 200 kb) was selected as the refined center of the candidate loop. Then, for each sample at each locus, this center value was divided by the background mean for that sample, which was determined by calculating the average of the off-diagonal bins along the genomic distance band appropriate for loop size (±9–14 bins for

small loops ≤ 35 kb and ±21–50 bins for larger loops) to give a simple proxy of Hi-C signal. Across each locus, the variance in Hi-C signal was then calculated.

CLASH:

CLASH's implementation first takes in pooled loop-call sets of genomic coordinates where a previous method has called a chromatin loop. For each sample, candidate loop coordinates were first grouped by chromosome, and the corresponding ICE-balanced Hi-C contact matrices were loaded once per chromosome using an in-memory cache to minimize redundant file access. Balanced counts were normalized between samples for sequencing depth. For each candidate loop, bins within 10 kb of the diagonal were excluded to avoid self-interaction noise. Then, from each called-loop center the maximum contact count within a dynamic search window ranging from a 5×5 matrix (±2 bins) for short-range loops (< 100 kb) up to an 11×11 matrix (±5 bins) for long-range loops (> 200 kb) was selected as the refined center of the candidate loop. For each candidate loop center, background interaction levels were calculated from averaging off-diagonal bins along the genomic distance band appropriate for loop size (±9–14 bins for small loops ≤ 35 kb and ±21–50 bins for larger loops).

For each refined loop center, we extracted a local Hi-C submatrix with adaptive sizing. Small loops (<35 kb) always used a 5×5 window (±2 bins). For larger loops, we iteratively expanded the window from radius 2 up to a cap of radius 10, evaluating the outer "ring" at each step; expansion stopped when the mean ring intensity fell below an adaptive multiple of the background (linearly decreasing from 1.5× at the inner rings toward 1.0× at the maximum radius). In sparse neighborhoods (any zero-valued neighbor in the immediate 3×3 around the center), we fell back to a fixed 5×5 window. Sample-specific depth normalization was applied to each value in the matrix via fixed scale factors (GM19317=0.872592, GM19347=0.885916, HG01457=0.965019, HG02666=1.0, HG03248=0.934843). To detect deletions near the loop-center, we flagged submatrices with entire zero rows/columns that continue for ≥20 bins beyond the extracted window in the source matrix.

The five extracted sample matrices at a given genomic coordinate were then scored using one of three methods based on the size of the each extracted matrix at that position: 5x5 matrices used a weighted combination of center-pixel, 3×3 inner square, and full matrix enrichment over background, while larger matrices used an adaptive weighted average between functions that calculated the weighted decay of interaction strength from the loop center across all pixels and calculated the iterative shell-based decay relative to the previous shell's average, with negative decay penalties for stronger-than-expected outer pixels. Matrices larger than the median size matrix at a given genomic coordinate were rewarded based on the increase in size. All scores were then robustly normalized using a median-absolute-deviation-based sigmoid transformation. Each loop's base score came from applying a weighted average of the 2 function scores based on the size of the matrix, with smaller matrices being influenced heavier by the hierarchical

(compared to shell before) score and larger matrices being influence heavier by the division (compared to center) score.

After computing the base scores, additional matrix terms were incorporated to refine the score. A coherence term quantified the variance among pixels within the 3×3 region surrounding the loop center, with lower variance leading to rewards in the score. A contrast term captured the relative enrichment of the loop center compared to the local background, rewarding loops with sharply defined focal interactions. The support term, representing the proportion of nonzero terms, and context enrichment terms, representing how enriched the matrix surrounding the loop center were used used to penalize sparse or noisy matrices and to upweight locally enriched regions. These features were combined using a logistic scaling function to smooth transitions between high- and low-quality matrices and added to the base score to yield a score between 0-1 for each candidate loop that serves as a proxy for loop intensity at the given coordinates for each sample, thus allowing for inter-sample differential loop comparison. CLASH's threshold values, rewards, weights, and overall implementation is optimised for loop analysis at the 2 kb resolution, and all downstream loop analysis was performed at 2 kb resolution.

CLASH validation:

To validate CLASH loop scores improve upon the initial binary loop calls, several metrics were calculated and compared across the two datasets. First the correlation between loop score and Hi-C signal (as calculated above) was compared for both datasets. Next, the explained variance ($R^2$) was obtained from univariate linear regression using either the initial call or CLASH score as the predictor, and the improvement in $R^2$ ($\Delta R^2$) was evaluated with an F-test for nested models. Third, mutual information (MI) between each predictor and Hi-C signal was estimated using mutual_info_regression, and the difference in MI ($\Delta$MI) was assessed using a 200-iteration bootstrap test. Finally, per-locus mean-squared error (MSE) between predicted signal (initial or CLASH) and observed Hi-C signal across the five samples was computed, the fraction of loci showing MSE improvement under CLASH was recorded, and paired differences were tested using a Wilcoxon signed-rank test. All analyses were performed in Python using pandas, numpy, scipy, scikit-learn, seaborn, and matplotlib.

Quantifying the correlation between CTCF occupancy and CLASH loop-score:

To quantify how CTCF protein occupancy globally relates to chromatin loop strength, CLASH-scored loops were first filtered to retain only those in which both anchor bins contained a CTCF site within 10kb. For each loop, the CTCF occupancy values of its two anchors were averaged to generate a single occupancy value. If one CTCF occupancy was missing data, the other site's CTCF occupancy value was used, and if both were missing then the datapoint was

excluded from analysis. This per-loop occupancy metric was then paired with the corresponding CLASH loop score, and correlations were computed across all loci and samples to assess the global relationship between CTCF binding and loop strength.

We also quantified how strongly CTCF binding predicts loop strength at individual genomic loci, we performed a per-locus regression analysis across the five samples. For each loop locus, we assembled a vector of occupancy values (predictor) and corresponding CLASH loop scores (response) across samples. Loci with fewer than four samples containing both valid occupancy and loop-score measurements were excluded. To ensure sufficient dynamic range for meaningful regression, loci were further filtered to retain only those with loop-score standard deviation greater than 0.27 across samples as loci with minimal across-sample variation do not contain sufficient dynamic range to support reliable regression estimates and were empirically observed to produce unstable or noise-dominated β coefficients. For each remaining locus, we z-scored the occupancy predictor and fit a univariate Ridge regression model (RidgeCV, $\alpha \in 10^{-4}$–$10^{4}$) to predict loop strength across the five samples. Because each locus contains only five measurements (one per sample), per-locus Pearson correlations are statistically unstable and have extremely wide sampling variance. This approach is more robust for small-n designs and directly estimates the influence of occupancy on loop strength at each locus. We collected all per-locus β coefficients and performed both a one-sample one-sided t-test ($H_1$: $\beta > 0$) and a sign test (binomial test, null $p = 0.5$). The proportion of loci with $\beta > 0$ was also reported. Finally, the distribution of β coefficients was visualized using kernel density estimation. All analyses were conducted in Python using pandas, numpy, scikit-learn, scipy, seaborn, and matplotlib.

Quantifying the correlation between CTCF occupancy and CTCF insulation capability:

To investigate the effect of CTCF occupancy on CTCF's role as an insulator, we first used cooltools (v0.7.0) insulation to call insulating bins in our 5kb resolution mcool files with a window size of 100,000 and min-frac-valid-pixels = 0.3 This process only successfully completed chromosome 1 for each sample, but nonetheless still provided sufficient data points (~5132 bins per sample) to continue with downstream analysis. We used this list of insulating bins to classify all of the CTCF sites within chromosome 1 for each sample as either a "boundary-CTCF site", responsible for creating TAD domains, or a "non-boundary CTCF site" which does not form TAD domains. For each of the insulating bins, a boundary score was provided by cooltools insulation. For every CTCF site, boundary and non-boundary, a custom log2ratio score was implemented to quantify each CTCF site's ability to act as an insulator. At 5 kb resolution, each CTCF coordinate was rounded to the nearest bin and a 200 kb (±100 kb) window of the Hi-C contact matrix was extracted. The upper triangular portion for each of these matrices was divided into 3 sections relative to the CTCF site's reference bin:

- x1 (−/− triangle): contacts upstream of the site.
- x2 (+/− and -/+ square): contacts between downstream–upstream bins.

- x3 (+/+ triangle): contacts downstream of the site.

The log2ratio score was defined as

log2ratio=log2((x1+x3)/(x2)) and was similar to early insulation-based boundary calling methods. Log2ratio scores were calculated for each CTCF site. Each CTCF site also contained CTCF occupancy data, as previously determined. Pearson's correlation coefficients were calculated within each sample to compare protein occupancy levels with the log2ratio insulation score across both boundary and non-boundary CTCF sites and averaged across samples. The significance of the difference between the boundary and nonboundary values was determined using a two-sided Welch's t-test on the per-sample correlation coefficients.

Quantifying the correlation between PWM scores, m5C methylation, CTCF occupancy and CLASH loop-scores:

For each chromatin loop in each sample, PWM scores and CpG methylation levels from the two CTCF anchors were averaged to generate per-loop measures of motif strength and methylation. These values were paired with CLASH loop scores and occupancy values from earlier, and global Pearson correlations were computed to quantify the individual relationships between motif strength, methylation, and loop intensity. To test whether PWM and methylation provided explanatory power beyond CTCF occupancy, we compared the correlation between occupancy and loop-strength to the multiple correlation of a combined association model incorporating occupancy, PWM scores, and CpG methylation using Steiger's test for dependent correlations, which accounts for shared outcomes and predictor non-independence. This analysis provided a direct statistical assessment of whether adding genetic (PWM) or epigenetic (m5C) features improved explanatory power relative to occupancy alone. We performed an analogous comparison using PWM and methylation as the only predictors to evaluate their joint contribution independent of occupancy.

Mediation analysis:

To quantify how much of the effect of CTCF motif strength (PWM score) and CpG methylation on loop strength is transmitted through CTCF occupancy, we performed a correlation-based mediation analysis. Pairwise Pearson correlations were computed between (i) PWM and occupancy, (ii) methylation and occupancy, (iii) occupancy and loop strength, and (iv) PWM or methylation and loop strength. For each predictor X (PWM or methylation), mediator M (occupancy), and outcome Y (loop strength), the indirect (mediated) effect was estimated as the product of correlations ab = r_XM * r_MY, the direct effect was computed as c' = r_XY – ab, and the proportion mediated was defined as (a*b) / r_XY. Confidence intervals for the indirect effect were obtained using 10,000 bootstrap replicates, generated by adding small Gaussian noise

to each observed correlation to approximate sampling uncertainty. The significance of the mediated effect was assessed using a z-test comparing the observed indirect effect to its bootstrap distribution.

eQTL analysis:

To test whether SNPs within loop-anchor CTCF sites that modulate loop strength (putative interaction quantitative trait loci, iQTLs) are enriched for known expression QTLs (eQTLs), we obtained the GTEx v8 Whole Blood significant variant–gene pair dataset. For each loop containing a SNP within either CTCF anchor, we recorded whether the variant was present in the GTEx eQTL list. Separately, we quantified whether each variant significantly altered loop strength by comparing the mean CLASH loop scores between individuals carrying the reference versus alternate allele and evaluating a series of loop-strength difference thresholds. For each threshold (0.2–0.6), we computed the proportion of SNP–loop pairs classified as iQTLs and the proportion overlapping GTEx eQTLs. Enrichment was assessed by calculating the log odds ratio of eQTL overlap across increasing loop-strength thresholds.

Structural variation analysis:

We next investigated whether large structural variants (SVs)—specifically insertions and deletions that introduce or remove CTCF binding sites between haplotypes—lead to measurable changes in chromatin loop formation. Structural variant calls for each sample haplotype, aligned to the hg38 reference genome, were filtered to retain only those altering at least 19 bp (the length of the CTCF consensus motif). FIMO motif scanning was then performed on each SV sequence to identify the number and positions of CTCF motifs gained or lost. These SV-associated motif changes were subsequently intersected with CLASH-derived loop scores to assess whether the addition or removal of CTCF sites corresponded to loop strengthening or weakening.

For each loop locus, and for each sample, we then determined whether either haplotype carried an SV that added at least one new CTCF site (insertion) or removed at least one CTCF site (deletion) within either loop anchor. Across the five individuals, this yielded two groups of loop-strength measurements at each locus: (i) samples whose haplotypes contained a CTCF-altering SV at the anchor, and (ii) samples without such an SV. For each locus and for each SV class, we computed the change in loop strength as the difference between the mean CLASH loop score of the SV group and the non-SV group ($\Delta$ loopscore = mean_SV − mean_nonSV). Statistical significance of these differences was assessed using a two-sided Mann–Whitney U test comparing loop scores between SV and non-SV samples at each locus. Genome-wide distributions of $\Delta$ loopscore values for CTCF-adding insertions and CTCF-removing deletions were then summarized and visualized to quantify the directional impact of SVs on chromatin loop formation.

To investigate whether the allele-specific impact of CTCF-altering structural variants (SVs) could explain the weak global correlation between motif-disrupting SVs and loop-strength

changes, we performed two complementary analyses. First, for every loop-anchor SV that introduced or removed ≥1 CTCF motif in either haplotype of any sample, we determined the SV zygosity by checking whether CTCF-site gain or loss occurred on one haplotype (heterozygous) or both (homozygous). Loci in which all samples were SV or all were non-SV were excluded to ensure informative comparisons. Across all variable loci, we tabulated the proportion of CTCF-altering SVs that were homozygous versus heterozygous.

Second, to test whether nearby "backup" CTCF motifs might buffer the impact of motif-disrupting SVs, we scanned for non-overlapping CTCF motifs located within ±20 kb of each SV event. All phased CTCF motif coordinates were extracted from haplotype-resolved FIMO-derived BED files for each sample. Each SV <20 kb in length was intersected with these motif sets to determine whether a compensatory motif remained intact in the SV-bearing haplotype, provided it did not overlap the SV interval itself. For insertions, the SV region was treated as a minimal 30 bp span; for deletions, the full deleted interval was used. SVs were then classified as either having or lacking a nearby backup CTCF site, and the genome-wide fraction of compensatory SVs was computed. This analysis enabled us to distinguish SV-driven loop perturbations that directly remove unique CTCF anchors from those in genomic regions containing redundant, functionally substitutable CTCF sites.

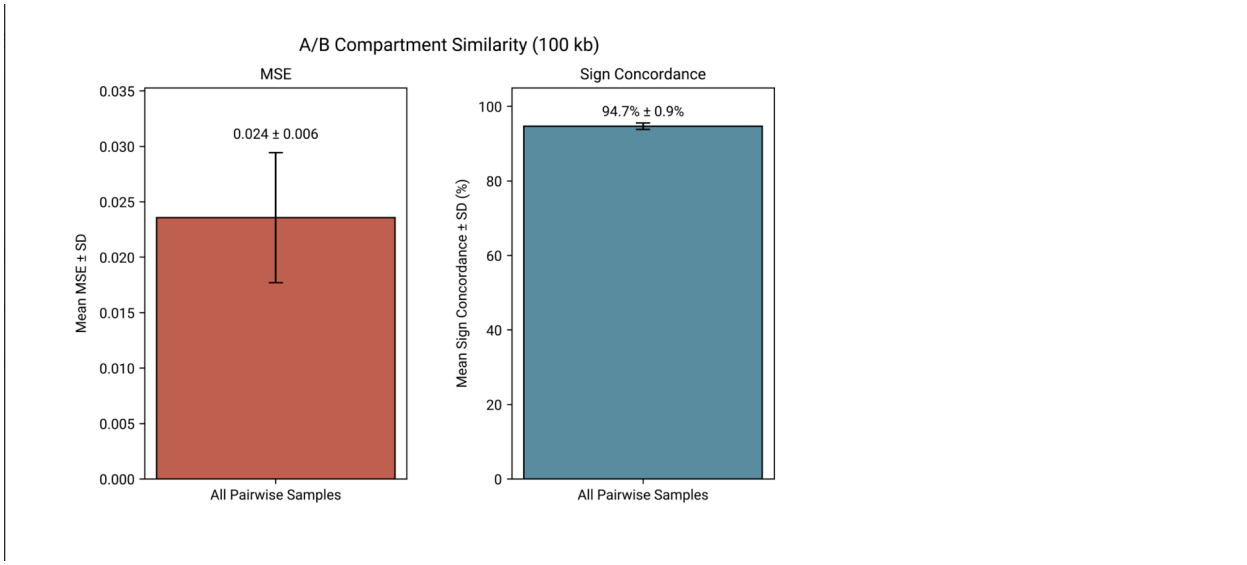
**Supplementary Results:**

**Table S1.** Hi-C sequencing summary statistics across the five samples.

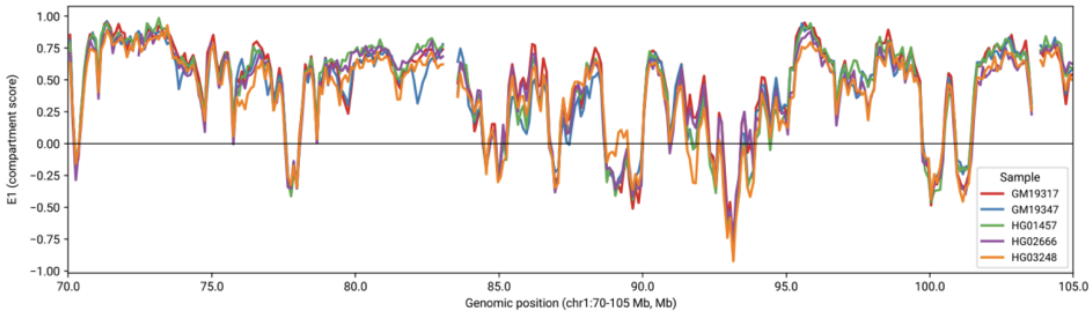
| Sample | GM19317 | GM19347 | HG02666 | HG01457 | HG03248 |
|---|---|---|---|---|---|
| Total read pairs (millions) | 1813.6 | 1946.7 | 1440.9 | 1789.8 | 1449.8 |
| % unmapped | 31.40% | 32.30% | 28.50% | 31.10% | 30.10% |
| % one-sided | 10.30% | 10.20% | 10.70% | 9.90% | 10.00% |
| % two-sided | 58.30% | 57.50% | 60.80% | 59.00% | 59.90% |
| % duplicated | 12.80% | 14.20% | 12.10% | 19.20% | 11.90% |
| Unique read pairs (millions) | 824.9 | 842.8 | 701.1 | 712 | 695.9 |
| Cis interactions (count) | 658065376 | 648963062 | 576398312 | 596436231 | 616383635 |
| Trans interactions (count) | 166854257 | 193862696 | 124683354 | 115560791 | 79549255 |
| FF (1-2kb) | 24.60% | 25.10% | 24.90% | 24.60% | 24.70% |
| RF (1-2kb) | 20.60% | 20.90% | 21.10% | 20.60% | 21.30% |
| FR (1-2kb) | 30.00% | 28.90% | 29.10% | 30.00% | 29.20% |
| RR (1-2kb) | 24.80% | 25.10% | 24.90% | 24.80% | 24.80% |
| Coverage (x) | 175.76 | 188.65 | 139.64 | 173.44 | 140.50 |

**Table S2.** Fiber-seq summary statistics across the five samples.

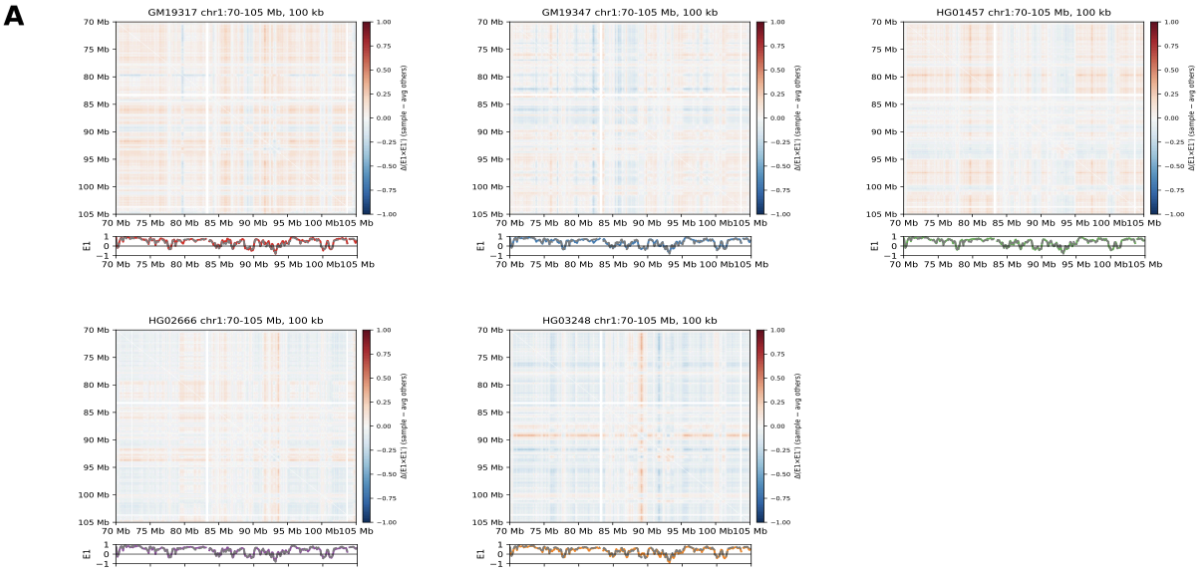
| Sample | GM19317 | GM19347 | HG02666 | HG01457 | HG03248 |
|---|---|---|---|---|---|
| Coverage (x) | 34 | 31.1 | 33.8 | 35.3 | 33.3 |
| Mean read length (bp) | 20850 | 22591 | 24462 | 21151 | 14236 |
| Hifi yield (gbp) | 105 | 96 | 105 | 109 | 103 |



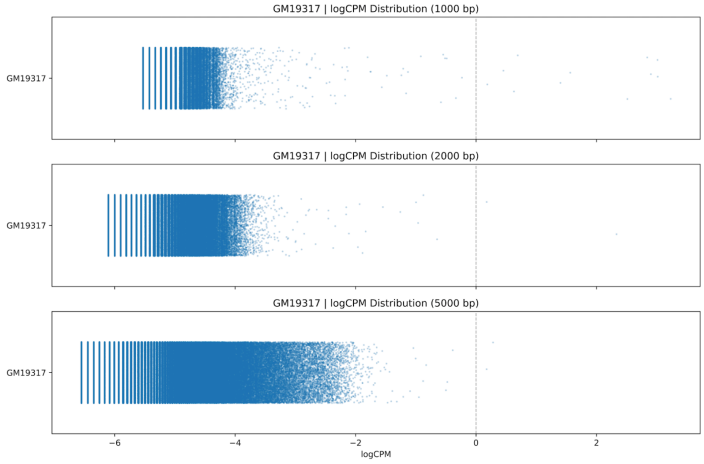

**Supplementary Figure 1. Pairwise comparisons of A/B compartment eigenvector values (E1) across all five samples.** Compartment profiles show high similarity, with an average MSE of 0.024 ± 0.006 and sign concordance of 94.7% ± 0.9% across sample pairs.



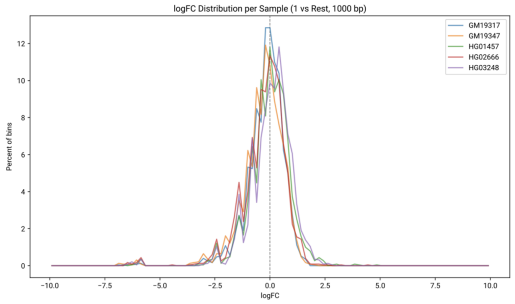

**Supplementary Figure 2. E1 compartment eigenvector profiles at 100 kb resolution for all five samples plotted across a representative region of chromosome 1 (chr1:70–105 Mb).** Colors indicate GM19317 (red), GM19347 (blue), HG01457 (green), HG02666 (purple), and HG03248 (orange).



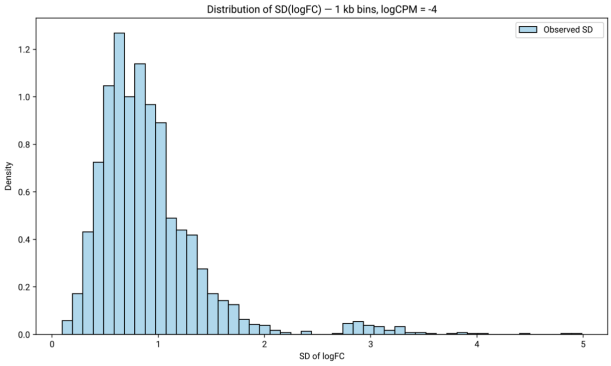

**Supplementary Figure 3. Hi-C contact maps at 100 kb resolution for the representative region chr1:70–105 Mb for all five samples.** For each individual, the mean contact map of the other four samples was subtracted to highlight sample-specific deviations. A/B compartments were computed using the first eigenvector (E1), with red indicating A compartments and blue indicating B compartments. The corresponding E1 tracks are shown below each heatmap and include both the sample-specific E1 profile and the average profile from the other four samples.



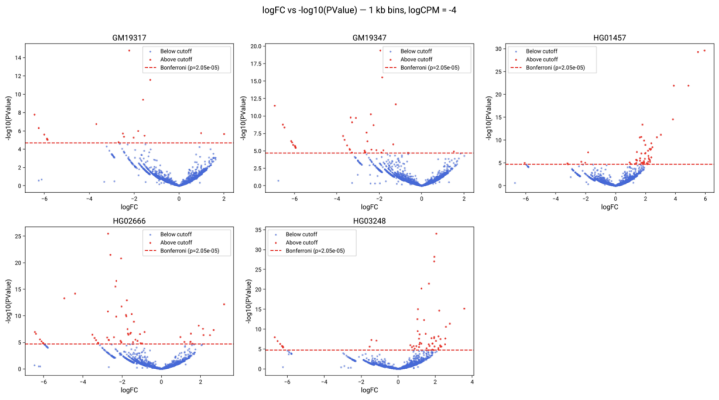

**Supplementary Figure 4. Distribution of logCPM values across all loci for GM19317 at 1 kb, 2kb, and 5kb resolution, used to determine the filtering threshold (logCPM > –4) that retains sufficiently powered interactions.**



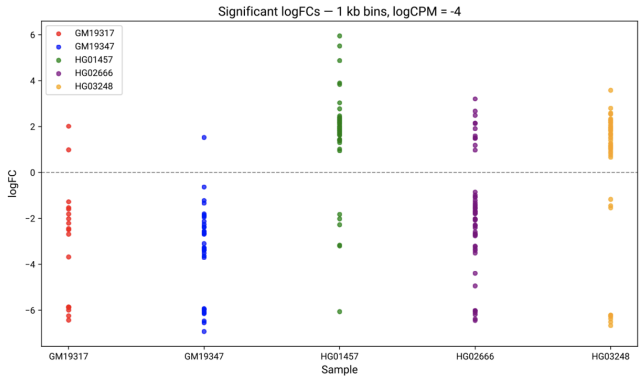

**Supplementary Figure 5. Distribution of filtered logCPM > -4 logFC values for each sample showing the expected unimodal peak centered near zero at 1 kb resolution.**



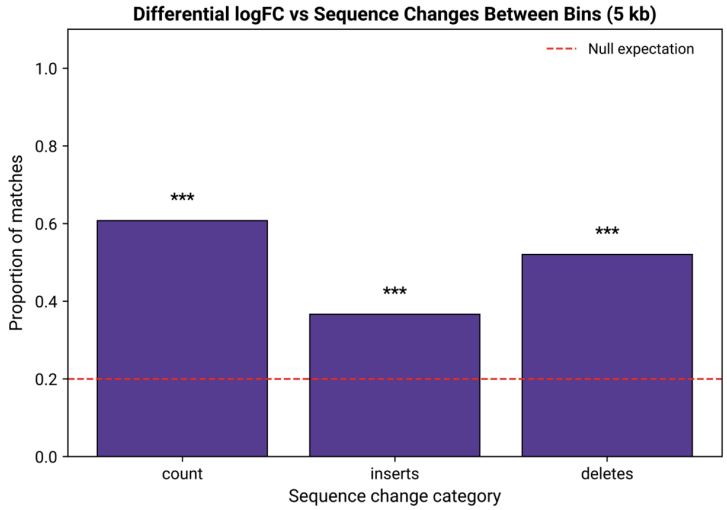

**Supplementary Figure 6. Distribution of logFC standard deviations across 1kb filtered loci, following an approximately negative-binomial shape with most interactions conserved (σ ≈ 0.5) and a minority showing strong differential signals.**



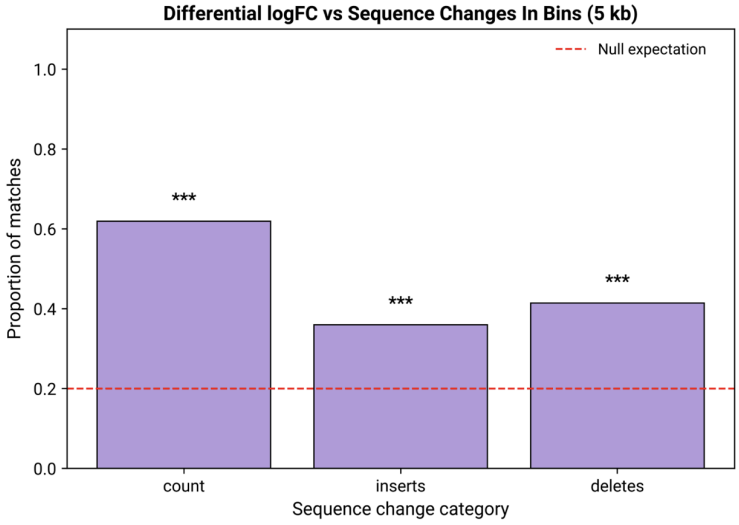

**Supplementary Figure 7. Plot of logFC vs log(p value) for 5 kb filtered bins for sample HG01457.** Bonferroni multiple-testing correction was applied with Bonferroni p = 2.05 x $10^{-5}$, and bins that passed the threshold are colored red.



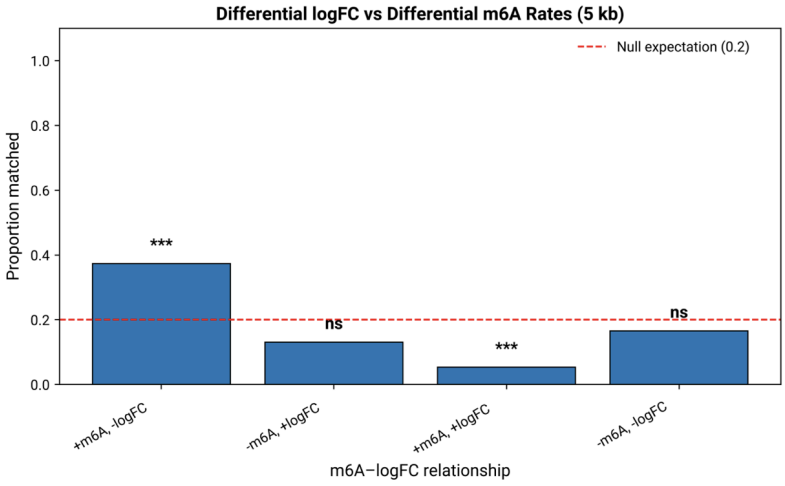

**Supplementary Figure 8. LogFC values for each differential bin for each sample at 1kb resolution that pass Bonferroni MTC.**



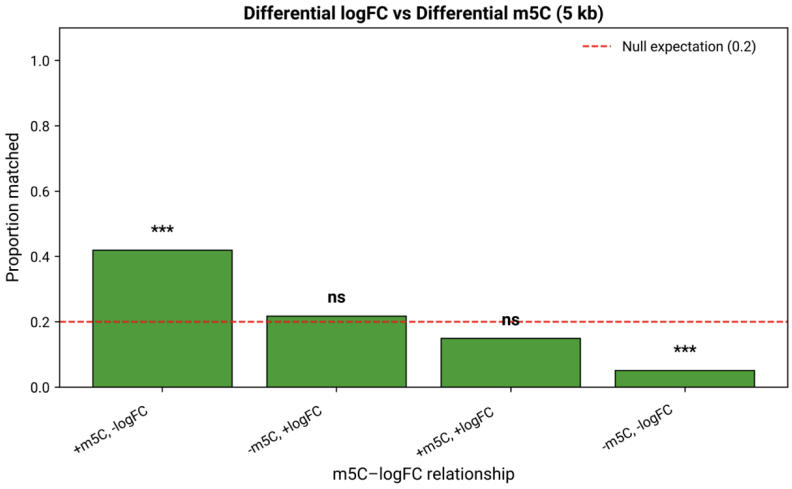

**Supplementary Figure 9. Match rate between the sample with a differential logFC value and the sample with the greatest number of base sequence changes (total changes, inserted bases, and deleted bases) compared to the null 0.2 between interaction bins.** Match rate is 60.7% (n = 107, $p = 3.95 \times 10^{-20}$, binomial test). Significance levels are denoted as $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***).



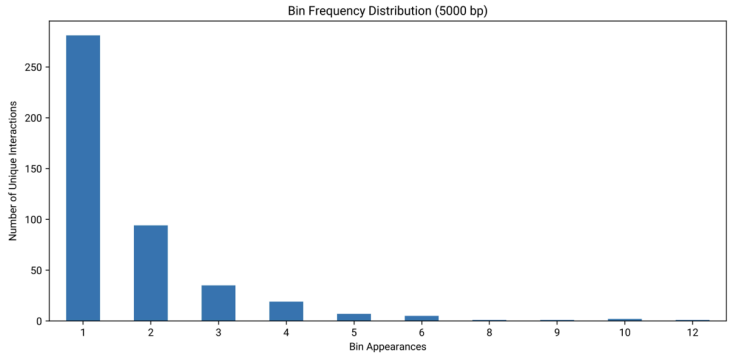

**Supplementary Figure 10. Match rate between the sample with a differential logFC value and the sample with the greatest number of base sequence changes (total changes, inserted bases, and deleted bases compared to the null 0.2 in interaction bins.** Match rate is 61.9% (n = 118, $p = 4.25 \times 10^{-23}$, binomial test). Significance levels are denoted as $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***).



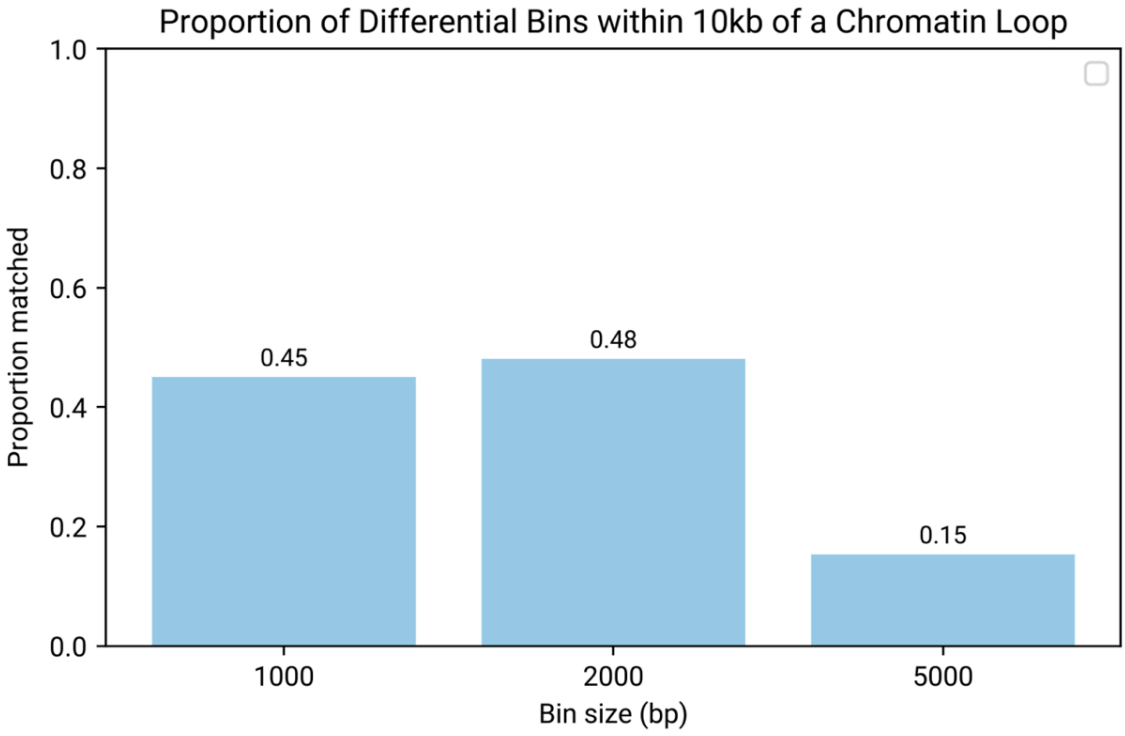

**Supplementary Figure 11. Match rate between the sample with the (highest, lowest) logFC value and the sample with the (highest, lowest) m6A methylation level at anchor bins compared to the null 0.2.** Match rate between lowest logFC and highest m6A sample is 37.3% (n = 75, $p = 8.62 \times 10^{-4}$, binomial test). Significance levels are denoted as $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***).



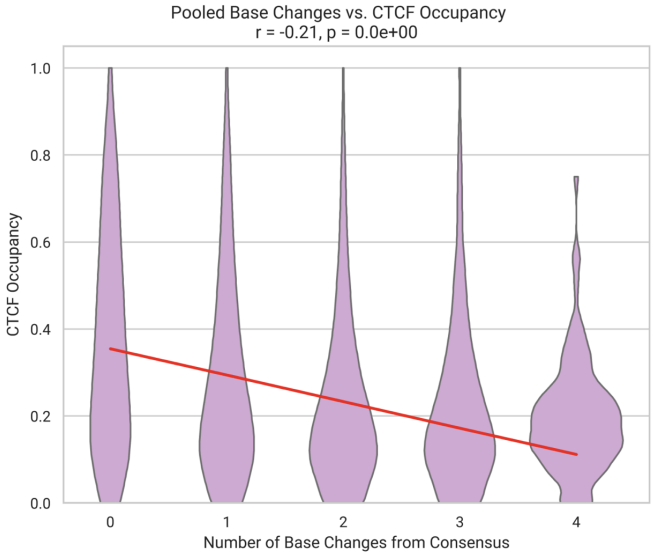

**Supplementary Figure 12. Match rate between the sample with the (highest, lowest) logFC value and the sample with the (highest, lowest) m5C methylation level at anchor bins compared to the null 0.2.** Match rate between lowest logFC and highest m5c sample is 41.9% ($n = 74$, $p = 1.58 \times 10^{-5}$, binomial test). Significance levels are denoted as $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***).



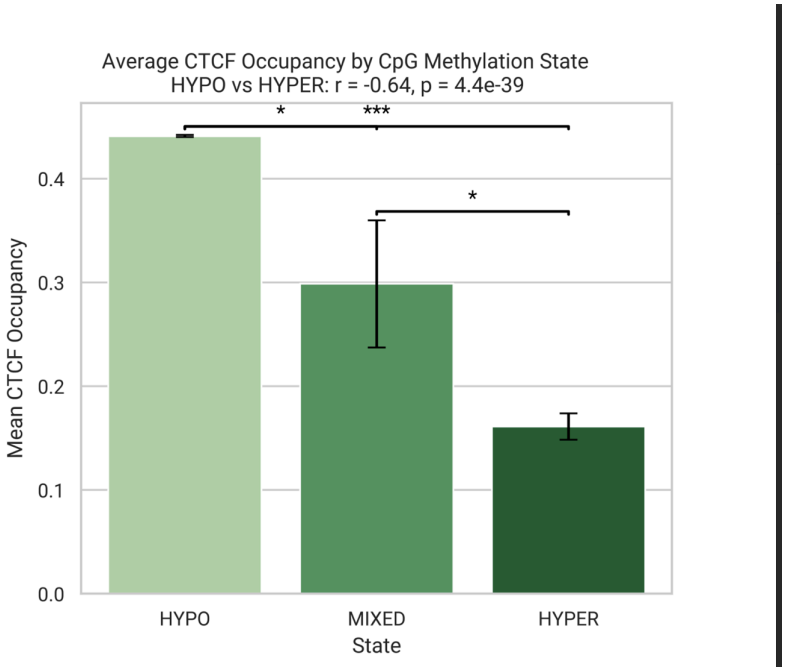

**Supplementary Figure 13. Frequency distribution of each anchor among the 382 differential interaction bins identified at 5 kb resolution across the genome.**



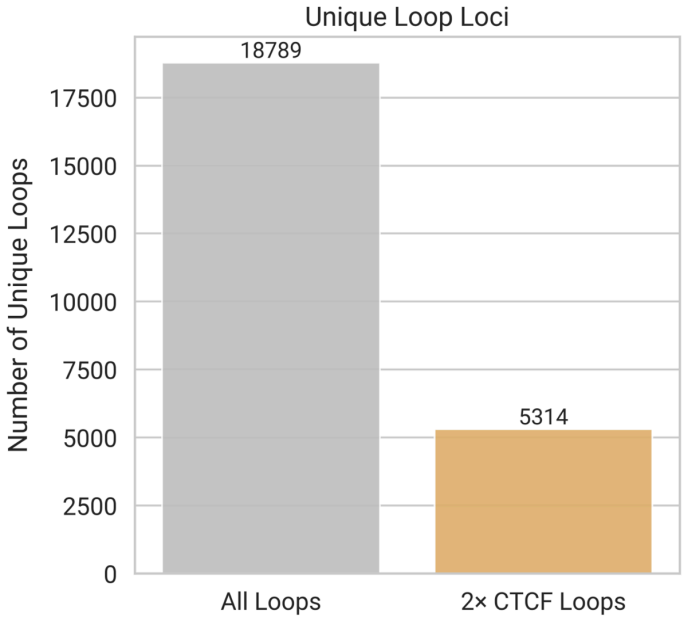

**Supplementary Figure 14.** Proportion of differential bins located within 10 kb of a loop anchor across resolutions.




**Supplementary Figure 15. Correlation between the number of base substitutions within each CTCF motif and occupancy.** Pearson p = -0.21 (n = 262,463 sites; $p < 10^{-300}$).

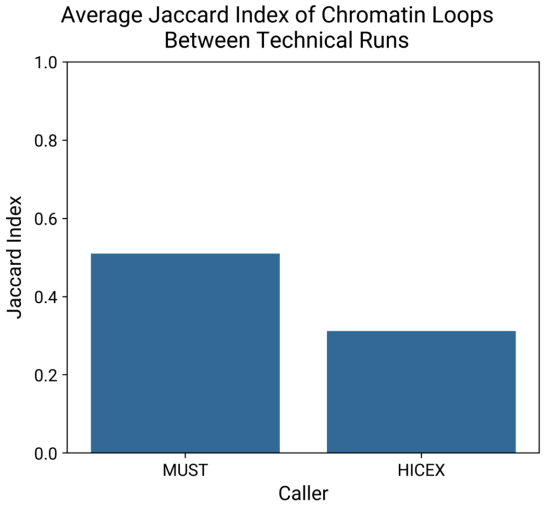



**Supplementary Figure 16. CTCF occupancy across CTCF m$^5$C CpG methylation state.** Pearson r = -0.64 between hypomethylated and hypermethylated states (hypomethylated n = 58,094 sites; hypermethylated n = 140 sites; mixed n = 14 sites; $p < 10^{-300}$).

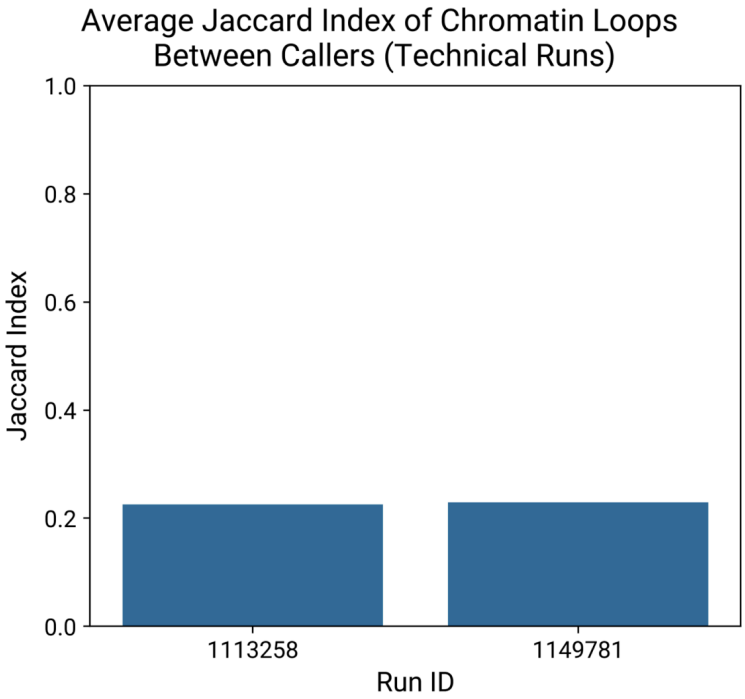



**Supplementary Figure 17. Total number of unique loci with at least one loop call identified by Mustache and HiCExplorer across samples.** At 2 kb resolution using a p-value threshold of

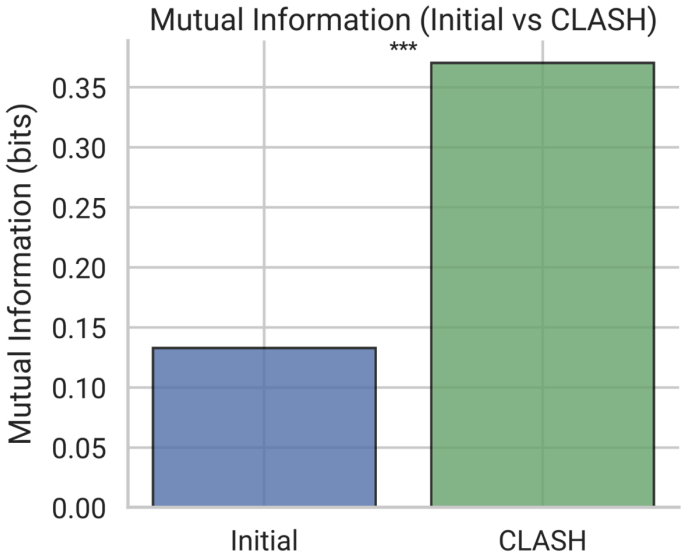
0.1, there were 18,789 loci (grey). Among these, 5,314 loci (yellow) are anchored by two CTCF sites within 10 kb in at least one haplotype.




**Supplementary Figure 18. Average Jaccard index of loop calls between technical runs for Mustache and HiCExplorer.** Technical run IDs are 1113258 and 1149781 for sample GM19317.

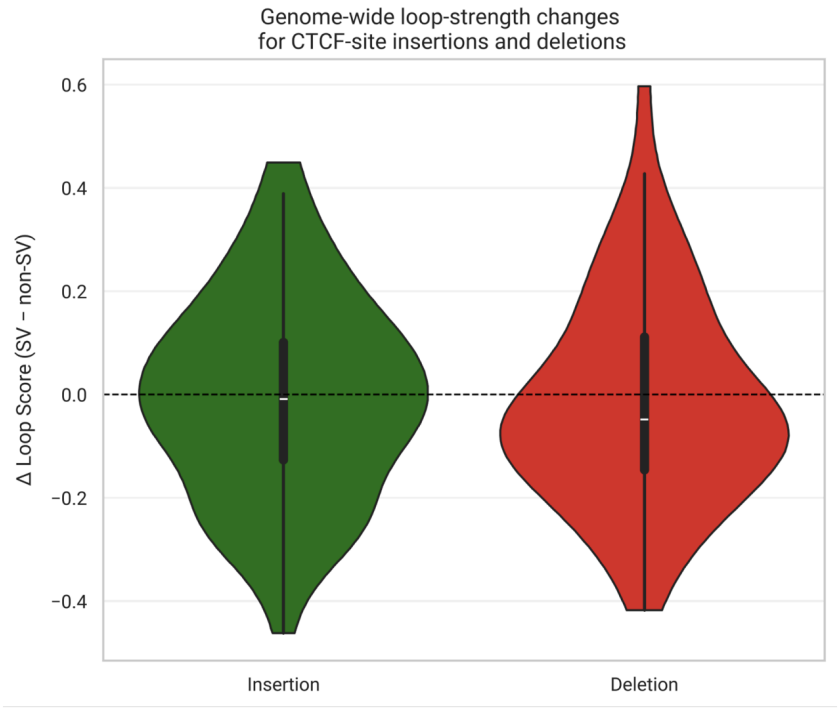


**Supplementary Figure 19. Average Jaccard index of loop calls between Mustache and HiCExplorer.** Technical run IDs are 1113258 and 1149781 for sample GM19317.

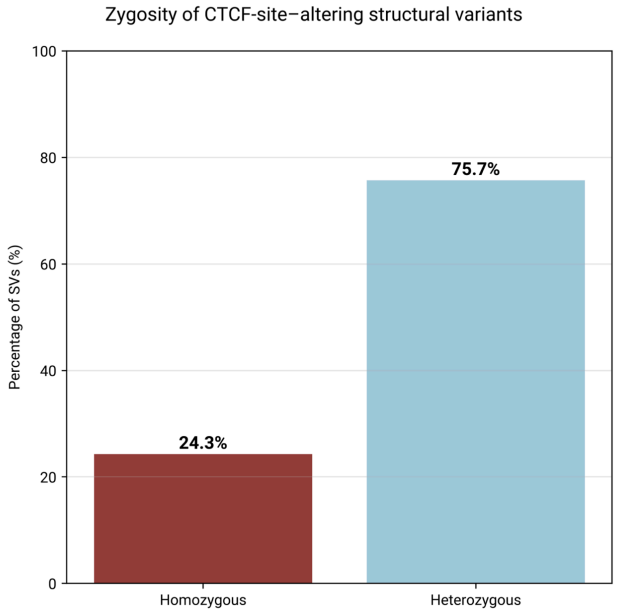



**Supplementary Figure 20. Explained variance ($R^2$; left) and mutual information (bits; right) encoded by loop calls before (Initial, blue) and after (CLASH, green) CLASH harmonization.**

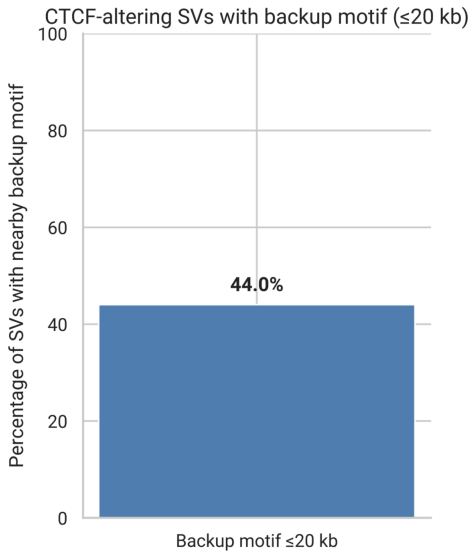
**Supplementary Figure 21. Example chromatin loop before and after CLASH harmonization.** Representative locus (chr14:76,052,000–76,352,000) illustrating loop-calling inconsistency across samples prior to CLASH (top), where Mustache and HiCExplorer fail to identify the same loop across individuals, only calling it in sample HG02666. After CLASH harmonization (bottom), the same locus is consistently scored as strong (purple) or very strong (black) across all five samples.




**Supplementary Figure 22. Distribution of loop-strength differences for loci containing structural-variant insertions (n = 71, left) or deletions (n = 64, right), compared to samples without the variant at the same locus, showing minimal global effect of structural variation on loop strength.**



**Supplementary Figure 23. The first proposed mechanism explaining the limited global impact of structural variants on loop strength.** 75.7% of loop-altering variants were heterozygous.




**Supplementary Figure 24. The second proposed mechanism explaining the limited global impact of structural variants on loop strength.** 44% of variant CTCF sites lie within 20 kb of another CTCF site.