Tri-4C: efficient identification of cis-regulatory loops at hundred base pair resolution

Yizhou Zhu[1,*], Yousin Suh[1,2,3,*]

Affiliations:
1. Department of Genetics, Albert Einstein College of Medicine, Bronx, New York 10461, USA
2. Department of Medicine, Albert Einstein College of Medicine, Bronx, New York 10461, USA
3. Department of Ophthalmology & Visual Sciences, Albert Einstein College of Medicine, Bronx, New York 10461, USA

*Corresponding Authors:

Yizhou Zhu
Department of Genetics,
Albert Einstein College of Medicine
Michael F. Price Center
1301 Morris Park Avenue, Room 469
Bronx, NY 10461, USA.
Phone: 718-678-1112
E-mail: yizhou.zhu@alumni.einsteinmed.org

Yousin Suh
Department of Genetics,
Albert Einstein College of Medicine
Michael F. Price Center
1301 Morris Park Avenue, Room 475
Bronx, NY 10461, USA.
Phone: 718-678-1200
E-mail: yousin.suh@einstein.yu.edu

Abstract

Here we present Tri-4C, a targeted chromatin conformation capture method for ultrafine mapping of chromatin interactions. Tri-4C quantitatively reveals cis-regulatory loops with unprecedented resolution, identifying functional enhancer loops devoid of typical epigenomic marks and uncovering allele-specific loop alterations in enhancer interaction networks underlying dynamic gene control. The Tri-4C approach is applicable to general 3C-derived methods for the study of single-allele enhancer loop networks.

Full Text

The human genome harbors a large number of cis-regulatory elements (CREs), such as enhancers, that play fundamental roles in gene regulation[1]. Since many of these elements exert their regulatory function through long-range interactions, methodologies that map distal chromatin contacts, such as chromatin conformation capture (3C) and its derivatives, have been developed[2, 3]. However, the ability of current 3C methods to identify distal interactions is limited by the restriction site availability on the genome[4]. Due to significant variation in restriction site distribution (**Supplementary Fig 1a**), a typical 4 bp-cutter is not able to tag all CREs (e.g. 200-300 bp core size for enhancers) for proximity ligation[5, 6]. In addition, to compensate for the digestion heterogeneity, most 3C-derived methods produce a uniformly spaced contact matrix by binning multiple restriction sites into larger windows of ~1-5 kb size at minimum[7-9]. This suboptimal resolution for analysis poses a significant challenge to distinguish the regulatory loop interaction signals from the relatively high background interaction often found within topologically-associating domains (TADs). Therefore, improved 3C methods are needed to comprehensively tag CREs, and to provide sufficient resolution for loop signal-background separation.

To overcome the limitations of current methods to comprehensively detect CRE loops, we developed Tri-4C, a novel targeted chromatin conformation capture (4C) method. In Tri-4C, distal chromatin interactions are probed by *in situ* digestion of genomic DNA using three 4 bp-cutter restriction enzymes (REs), DpnII (MboI), Csp6I (CviQI), and NlaIII (**Fig 1a**). The sticky ends are then blunted, allowing free re-ligation of cutting sites generated by three different REs, which dramatically increases ligation complexity. After sonication, the enrichment of contacts at the target viewpoint is achieved by two rounds of nested PCR using two sequential primers in the vicinity of the cutting site. The chromatin contacts are then identified through paired-end sequencing. Similar to UMI-4C[7], the sonication ends are utilized as unique molecular identifiers (UMI), to generate a PCR bias-free quantitative interaction map. The Tri-4C protocol is multiplexible, and the procedure can be completed in 2-3 days. *In silico* analysis of the human genome showed that the fragment size of Tri-4C is 1.9- to 5.2-fold shorter than single 4 bp-cutters (**Supplementary Fig 1a**).

To evaluate the performance of Tri-4C, we conducted a comprehensive comparative analysis of CRE interaction profiles between Tri-4C and UMI-4C in the 9p21 interferon B1 (IFNB1) TAD in IMR90 cells[10]. This locus, which shows a complex multi-sub-TAD architecture, harbors 85 putative CREs marked by DNase I hypersensitive sites (DHSs), including 46 showing H3K27Ac active enhancer marks (**Supplementary Fig 2**). We performed multiplexed Tri-4C on three viewpoints, two promoters (MLLT3, IFNB1) and a sub-TAD boundary (Boundary) showing strong CTCF/cohesin binding (**Supplementary Table 1**, **Supplementary Fig 2**). To map the ligated contacts, we developed a modified UMI-4C analysis pipeline to adapt to the multiple ligation ends generated by tripledigestion (Methods). Reads with low mapping quality and PCR duplicates were removed to generate quantitative interaction profiles. In parallel, we performed three UMI-4C assays (with minor modifications - see Methods), digested individually by DpnII, Csp6I, or NlaIII[7]. We found that the read count yield of Tri-4C was on average 5.3-fold higher than UMI-4C

(Supplementary Table 2, **Supplementary Fig 1b**), suggesting that the detectability of distal interaction proportionally increased with digestion frequency. Consistently, the reproducibility of Tri-4C was significantly higher, especially at sub-kilobase resolution **(Supplementary Fig 1c**).

In order to differentiate interaction loops from the local background interactions that occur with high frequency within TADs, we developed an algorithm resembling MACS[11] to identify loop sites with over-represented interaction read counts. Since 97% of fragments generated by the triple digestion are smaller than 500 bp, we binned reads into 500 bp windows in 100 bp sliding steps, a resolution comparable to the size CREs, and quantified their enrichment against a local background within a 5-50 kb dynamic range (**Fig 1b, Supplementary Fig 1a**). We applied the algorithm to Tri-4C, yielding 233, 138, and 21 reproducible intra-TAD loops, respectively, for the MLLT3, Boundary, and IFNB1 viewpoints. These loops significantly overlapped with a total of 70 CREs marked by DHS, 37 of which were also marked by H3K27Ac (**Supplementary Fig 3a**). *In-situ* Hi-C of IMR90, by comparison, revealed 4, 4, and 0 loops for these three viewpoints (**Supplementary Fig 2**)[9]. The cis-regulatory loop (CRL) profiles showed significant overlap among the viewpoints, similar to observations from HiChIP (**Supplementary Fig 3b**)[8]. We examined the mappability, GC content, and restriction site density around the identified loops, and found loop calling was not siginificantly affected by these factors (**Supplementary Fig 3c-e)**.

The 500 bp resolution (bin size) we chose to perform loop calling for Tri-4C was significantly higher than that with UMI-4C (3-5 kb) or Hi-C/HiChIP (5 kb). To test the impact of higher resolution on CRL detection, we re-analyzed the Tri-4C data with a larger bin size (3000 bp), comparable to previous methods[7-9]. At 3 kb resolution, Tri-4C identified on average 35% of the CRLs found at 500 bp resolution (**Supplementary Fig 4**), with lower signal-to-noise ratios at the overlapping loops, and produced merged loop signals between closely located CREs. Consistently, the 500 bp resolution analysis revealed that CRLs were less than 1 kb long, with the pinnacle precisely

aligning with DHS peaks (**Fig 1c**). Hence, sub-kilobase resolution mapping was essential to prevent excess convolution with background and robustly identify CRLs.

We compared the Tri-4C loop caller with the UMI-4C and the 1D adaptation of *in-situ* Hi-C algorithms, both of which estimate background interactions by using distance modeling based on global interaction profiling (Methods). Receiver operating characteristic (ROC) analysis at 100 bp resolution showed that Tri-4C loops were a strong predictor of DHS-marked CREs regardless of the algorithm used, while loop scores determined by the Tri-4C caller showed the highest accuracy (**Supplementary Fig 5a**). Furthermore, the CRL strengths (fold enrichment against background) determined by the Tri-4C algorithm were distance-independent and strongly correlated between viewpoints (r=0.82 between Boundary and MLLT3). The correlations obtained by the Hi-C and UMI-4C algorithms were less significant (r=0.48 and 0.29, respectively), most likely due to their tendency to over-correct for the distance (**Supplementary Fig 5b-d**).

To compare the performance of Tri-4C with UMI-4C in identifing CRLs, we analyzed the UMI-4C results using the Tri-4C loop caller at 500 bp resolution. For all viewpoints, Tri-4C identified on average 4.6-fold more CRLs compared to UMI-4C (**Fig 1d**, **Supplementary Fig 6a**). The loop score of Tri-4C also more accurately predicted the positions of DHS-marked CREs and H3K27Ac-marked enhancers than UMI-4C, suggesting that its higher loop detection sensitivity was not compromised by specificity (**Fig 1e**, **Supplementary Fig 6b,c**). We found that the UMI-4C profiles generated by different 4 bp-cutters each revealed a unique subset of the CRLs identified by Tri-4C, and these subsets overlapped poorly with each other (intersection over union < 0.2). In all UMI-4C profiles, CREs that did not show looping were significantly more distal to the closest restriction site, indicating that UMI-4C could detect CRLs only when the restriction site was sufficiently close to the CRE (**Fig 1f,g**). Such distance bias was not detected in Tri-4C, suggesting

that the ultrafine digestion of the genome of Tri-4C was necessary and sufficient to accurately identify looped CREs without bias.

We investigated Tri-4C-identified loop sites that did not overlap with enhancer marks, including histone modifications and DHS. We found they partially overlapped with ENCODE ChIP-seq signals, suggesting the presence of transcription factor binding sites and regulatory potential (**Supplementary Fig 7a**). To determine the regulatory function of these loops, we used CRISPR/Cas9 to delete ~1 kb regions of 4 sites that looped with the MLLT3 promoter but were devoid of enhancer marks (**Fig 2a, Supplementary Fig 7b, Supplementary Table 3**). Deletion of two of the sites significantly down-regulated MLLT3 expression, indicating that these were *bona fide* enhancers (**Fig 2b**). These functional enhancer loops were not revealed by DpnII UMI-4C.

To quantitatively analyze the CRLs called by Tri-4C, we compared the loop strength (i.e. log fold enrichment against local background) with the DHS fold enrichment for all CREs in the locus, and found they were significantly correlated (**Supplementary Fig 8a**). Motif analysis indicated that CREs harboring the CTCF motif formed significantly stronger loops with all three viewpoints (**Supplementary Fig 8b**), consistent with the role of CTCF in mediating chromatin interactions[9, 12]. In contrast, this correlation was not revealed by UMI-4C, due to the significant distortion of loop strength caused by the distance between CREs and the nearest restriction sites (**Supplementary Fig 9**).

To test if Tri-4C can reveal the CRL networks underlying dynamic gene control, we induced robust expression of IFNB1 through activation of well-defined antiviral signaling and performed Tri-4C on all three viewpoints[13]. The induction of IFNB1 caused its promoter to interact more frequently with the majority of CREs in the locus (**Supplementary Fig 10a-c**). However, many of these gains were not significant against the similarly increased local background, and after nomalization only

13 CREs showed induced-looping with IFNB1 (**Fig 2c, Supplementary Fig 10d**). The alterations in loop strengths with IFNB1 promoter significantly correlated with those from the MLLT3 and Boundary viewpoints, as well as the CRE activities indicated by the ATAC-seq peak strengths (**Supplementary Fig 10e,f**). The CREs gaining loop strength upon induction were enriched with the motifs of IRF family members, which are key regulators for IFNB1 activation (**Fig 2d, Supplementary Fig 10f**)[14]. These results indicated that Tri-4C was capable of revealing quantitative loop alterations underlying the activities of CREs in the CRL networks.

To test whether Tri-4C could differentiate the allelic impact of regulatory variants on CRL networks, we applied Tri-4C to examine the 9p21.3 locus. This locus harbors multiple coronary artery disease (CAD) risk variants, including two functional variants reported to abrogate the function of an enhancer (ECAD9) by disrupting TEAD3 and STAT1 binding, thereby misregulating the expression of the target genes, CDKN2A/B[15-17]. Using vascular smooth muscle cells (VSMC) derived from a human embryonic stem cell line (H7) that is heterozygous for the risk variants, we performed allele-specific (AS) Tri-4C on ECAD9 (**Supplementary Fig 11a**)[16]. The AS-Tri-4C profile showed highly cis-specific interaction (>99%), revealing looping of ECAD9 with 25 ATAC-seq-marked CREs in the locus, including both CDKN2A and CDKN2B promoters (**Supplementary Fig 11b,c)**. Among the looped CREs, 10 showed differential loop strength between alleles, and in all cases loops on the non-risk alleles were significantly stronger than the risk alleles. The stronger loop activity of ECAD9 on the non-risk allele was consistent with its higher accessibility indicated by ATAC-qPCR (**Supplementary Fig 11d)**[18]. Lastly, we found that stronger loops were formed between ECAD9 and CREs harboring TEAD3, STAT1, or SMAD family motifs (**Supplementary Fig 11e**), consistent with the roles of these factors in regulating CDKN2A/B, which were diminished by the CAD risk variants[15, 16, 19].

Finally, to demonstrate that Tri-4C can be applied to 3C-derived methods requiring end-filling of biotinylated nucleotides for pull-downs, such as Hi-C and Hi-ChIP[8, 20-22], we replaced NlaIII with its isoschizomer, CviAII, which generates a 5' AT overhang, and performed Tri-4C on the Boundary and MLLT3 viewpoints. The interaction profiles of the alternatively digested Tri-4C were highly consistent with the Tri-4C profiles generated using NlaIII digestion (r = 0.94) (**Supplementary Fig 12**).

We have described Tri-4C, an effective method for comprehensively and quantitatively identifying cis-regulatory loops. Unlinke 3C methods which are designed to quantify interaction frequencies among known regulatory elements, such as NG-Capture C and 3e-HiC, Tri-4C is capable of discovering cis-regulatory loops without prior knowledge of 1D CRE landscape[20, 23]. The ability of Tri-4C to identify regulatory elements and to quantitate their looping activities in an allele-specifc manner will be of great value to understanding how their looping activities are affected by the regulatory disease risk variants. Tri-4C could be applied to general 3C-derived methods, including the recently reported multi-contact 4C[4], for the detection of regulatory network landscapes with higher resolution and accuracy at broader scales.

Figure Legends

Figure 1 Tri-4C robustly and unbiasedly identifies CRLs. (**a**) Schematics of Tri-4C library construction. (b) Tri-4C loop calling algorithm. Each 100 bp sliding bin collects read count (***M***) from neighboring 500 bp intervals. Significant loops and their loop strengths are determined by Poisson statistics of ***M*** against expected read count ***N*** calculated from the total read counts in the 5-50 kb local backrgound. (**c**) Centerplot of Tri-4C signals at all DHS-marked CREs in the 9p21 IFNB1 TAD. (**d**) Venn diagram of reproducible CRLs (N=2) called for the Boundary viewpoint using Tri-4C and UMI-4C digested by three different restriction enzymes (**e**) ROC analysis using loop signals (-log(p)) for each 100 bp bin steps from Boundary as predictors of intra-TAD DHS peaks. (**f**) Boxplot for distance between intra-TAD DHS peak (N=85) center and the closest restriction site, separated by whether the peak is called to loop with any of the three viewpoints. (**g**) Correlation between raw signals of Tri-4C and UMI-4C and neighboring restriction site patterns at three CRE regions looped with Boundary. Statistic p values were calculated by U test. The y axis of 4C methods denote read count per 10,000 uniquely mapped reads.

Figure 2 Tri-4C reveals quantitative and functional CRLs. (**a**) Tri-4C, but not DpnII-UMI-4C, indicates looping of MLLT3 with 4 neighboring regions (S1-S4) lacking enhancer marks and CTCF/cohesin, and (**b**) Expression of MLLT3 after deletion of these regions using Cas9 and two pairs of guide RNAs (sg1, sg2) quantified by real-time PCR (N=3). (**c**) Alteration of IFNB1 interaction before (Ctrl) and after (Induced) induced expression in IMR90. Top track aligns interaction read count, while second denotes loop strength alteration (ΔlogFE) and shows the loop gain is specific to S1 despite increased count on both S1 and S2. Third track indicates ATAC-seq peak signal of the two enhancers corresponding to the two conditions. S1 is a known enhancer of IFNB1[24]. (**d**) Association between loop strength alterations after IFNB1 induction and IRF(1/2/3/7) motif presence at intra-TAD CREs. Statisitical p values were calculated by U test.

Supplementary Figure 1 Tri-4C improves both yield and reproducibility by finer digestion of the genome. (**a**) Distribution of DNA fragment size of human genome digested by indicated restriction enzymes. Numbers on top indicate median and in the parentheses indicate percentage of fragments smaller than 500 bp window size. (**b**) Yield of unique intrachromasomal reads for UMI-4C and Tri-4C on the three viewpoints. (**c**) Reproducibility of interaction profiles binned in 50 kb-50 bp.

Supplementary Figure 2 Overview of Tri-4C experimental design at the chromosome 9p21 IFNB1 TAD. Tri-4C profiles of three viewpoints (Boundary, MLLT3, and IFNB1) are displayed under IMR90 *in situ* Hi-C matrix (5kb resolution Rao 2014) obtained from 3D genome browser (Hi-C loops are highlighted with squares). Y axis of Tri-4C tracks denotes interaction frequency multiplied by 10,000. The interaction profiles are aligned with regulatory marks (DNase, H3K4me1) and boundary markers (CTCF, RAD21) for IMR90 cells obtained by the Roadmap Project. Bottom panel shows significant loop interactions between the viewpoints and CREs.

Supplementary Figure 3 Tri-4C loop annotation and quality control. (**a**) Overlap between intra-TAD Tri-4C loops and intra-TAD DHS and H3K27ac peaks. (**b**) Overlap of DHS-marked CRLs among the three viewpoints. (**c**) GC content, (**d**) mappability, and (e) restriction site density around regions looped with any of the three viewpoints. Gray background indicates confidence intervals estimated by using 1,000 randomly selected intra-TAD regions not looped with any viewpoints with mappability > 0.5.

Supplementary Figure 4 Comparison of Tri-4C profiles analyzed in 500 bp (High res) and 3000 bp (Low res) resolution. (**a**) Overlap of loops falling in CREs. (**b**) Interaction of MLLT3 with

neighboring CREs shown by Tri-4C in two resolutions. DHS peaks showing looping at 500 bp resolution are highlighted.

Supplementary Figure 5 Comparison of loop calling algorithms. (**a**) ROC analysis using loop scores for each 100 bp bin steps calculated by Tri-4C (Dynamic), 1D Hi-C, and UMi-4C algorithms as predictors of intra-TAD DHS peaks. (**b**) 2D plots comparing loop strength (logFE) on all intra-TAD CREs determined by Tri-4C normalization, or read count using (**b**) UMI-4C or (**c**) Hi-C normalization between viewpoints. Color indicates log distance ratio between the x and y viewpoint (blue = closer to x and red = closer to y). Pearson correlation coefficient r and p value from linear regression are indicated.

Supplementary Figure 6 (**a**) Venn diagram of reproducible CRLs (N=2) called for MLLT3 and IFNB1 using Tri-4C and UMI-4C digested by three different restriction enzymes (**b**) ROC analysis using loop scores for each 100 bp bin as predictors of intra-TAD DHS peaks. (**c**) ROC analysis using loop scores for each 100 bp bin as predictors of intra-TAD H3K27Ac peaks.

Supplementary Figure 7 Analysis of Tri-4C loops not overlapped with CREs (**a**) Overlap between off-CRE loops with intra-TAD regions showing binding with 5 or more ChIP-seq peaks in the ENCODE combined transcription factor binding track. (**b**) Validation of Cas9 deletion of regions indicated in **Figure 2a**.

Supplementary Figure 8 Analysis of Tri-4C loop strength. (**a**) Comparison between loop strength (logFE) from three viewpoints and DHS peak log fold enrichment on all intra-TAD CREs. Pearson correlation coefficient r and p value from linear regression model are indicated. (**b**) Association between loop strength and CTCF motif presence for Tri-4C and UMI-4C based on three enzyme digestion.

Supplementary Figure 9 2D plots between loop strength difference between UMI-4C and Tri-4C (Y axis, ΔlogFE) and the distance between the nearest restriction site and the DHS peak center (X axis, log scale) of all intra-TAD CREs. The p values were calculated by fitting with linear regression.

Supplementary Figure 10 CRL alterations after IFNB1 induction. (**a**) Expression profiling of IFNB1 and MLLT3 expression before and after IFNB1 induction by using real-time PCR (N=3, t test). (**b**) Comparison of Tri-4C yield for three viewpoints (N=2) (**c**) 2D plot showing read count of IFNB1 Tri-4C (normalized against Boundary) at all intra-TAD CREs before (X axis) and after (Y axis) induction. (**d**) Venn diagram showing the overlap of CRLs called from IFNB1 before and after induction. (**e**) Comparisons of loop strength alterations (ΔlogFE) between three viewpoints and (**d**) with ATAC-seq peak log fold enrichment changes on all intra-TAD CREs. Pearson correlation coefficient r and p value from linear regression model are indicated.

Supplementary Figure 11 Allele-specific Tri-4C for ECAD9. (**a**) Schematics for the allele-specific study design. The viewpoint primer is designed to include a heterozygote flag variant in the padding sequence. Reads are sorted and mapped separately according to the variant genotype. Allele-specific interaction loops are identified by differential loop analysis. (**b**) Allele-specific (AS) Tri-4C profile for ECAD9 in H7-derived VSMCs. Intervals below indicate loop regions called by each allele. Loops on ATAC-seq-marked CREs showing significant allelic bias are highlighted with * marks, with its color indicating the stronger allele. The VSMC ATAC-seq and ENCODE aortic smooth muscle cell (AoSMC) DNase tracks are shown below. (**c**) Fraction of *cis* interaction mapped to each AS profile (N=2) (**d**) ATAC-dPCR (digital PCR) on ECAD9. Significant p value was calculated by t test (N=2).

Supplementary Figure 12 Reproduction of Tri-4C using alternative digestion by CviAII. (**a**) 2D plot showing read count in 500 bp bins obtained from original Tri-4C and alternative digestion protocol. Pearson correlation coefficient r is indicated. (**b**) Loop score (log(-log(p)) comparison for all bins scored above 0 ($p<0.1$).

Supplementary Table 1 Sequences for viewpoint-specific primers. See Online Methods for adaptor and universal primer sequences.

Supplementary Table 2 Statistics for Tri-4C and UMI-4C libraries. Total Read indicates actual sequencing depth. On Target Ratio indicates reads with matched padding sequence. Unique Read indicates yield after deduplication. Intra-TAD Ratio indicates reads falling into the same TAD as the viewpoint.

Supplementary Table 3 Sequences for gRNA and validation primers used for MLLT3 putative enhancer deletion.

Methods

Cell Culture

IMR90 cells (ATCC CCL-186) were maintained in EMEM (Corning 10-009-CV) with 10% FBS (GEMINI 100-500). To induce *IFNB1* expression, cells were treated with 20 μM 2'3'-cyclic GMP-AMP (Invivogen tlrl-nacga23), 100 ng/mL IFNγ, and 10 ng/mL TNFα for 24 hours[13, 25]. Cells were collected at full confluence for all downstream analyses.

Human embryonic stem cells H7 (WiCell WA07) were maintained in feeder-free E8 system (ThermoFisher A1517001). Differentiation of the ES cells to vascular smooth muscle cells was conducted as previously described[26]. Briefly, cells were plated on vitronectin (ThermoFisher A14700) coated surface at 5-10% density. On day 2, the medium was switched to N2B27 (50% DMEM-F12 + 50% Neurobasal medium + 1x N2 supplement (ThermoFisher 17502048) + 1x B27 supplement (ThermoFisher 17504044)) supplied with 10 μM CHIR-99021 and 25 ng/ml BMP4 to induce mesoderm differentiation. From day 5, cells were incubated in N2B27 medium supplied with 10 ng/ml PDGF-BB (Peprotech 100-14B) and 2 ng/ml Activin A (Peprotech 120-14P) to induce VSMC differentiation. Five days after, the Activin A was retrieved from the medium, and the cells were expanded for two population doublings and collected for analysis.

RNA isolation and qRT-PCR

RNA was isolated from cells using the PureLink RNA mini kit (Thermo 12183020) according to manufacturer's instructions. DNase treatment was performed using DNA-free DNase Treatment & Removal Kit (Thermo AM1906). To synthesize cDNA, reverse transcription was performed using SuperScript IV Reverse Transcriptase (Thermo 18091050) with oligo-dT primers following manufacturer's instructions. Quantitative RT-PCR was performed on the Applied Biosystems StepOne Plus platform. For *IFNB1*, a pre-designed Taqman probe (Hs01077958_s1) was used with TaqMan Universal PCR Master Mix (Thermo 4304437). For *MLLT3*, a custom primer pair

(Fw: TTTGTGGAGAAAGTCGTCTTCC; Rev: GAGGTGATTCACTGGTGGATG) was used with PowerUp SYBR Green Master Mix (Thermo A25741). Expression was quantified using the delta CT method normalized to *HPRT1* (Thermo 4326321E).

Cas9-mediated Gene Editing

Guide RNAs (gRNAs) for the Cas9 endonuclease were selected using the CRISPR design tools from Zhang lab (http://crispr.mit.edu). To generate enhancer deletions in IMR90 cells, the IDT ribonucleoprotein (RNP) system was applied following the protocol provided by the manufacturer. Briefly, synethsized crRNAs were annealed with tracrRNA and incubated with Cas9 V3 (IDT 1081058) at equimolar concentrations. To perform NHEJ-mediated deletion, we transfected $1x10^5$ IMR90 cells with 22 pmol of Cas9 RNP using the Neon electroporation system with resuspension buffer R (ThermoFisher) at 1100V, 30ms, 1 pulse. After 72 hours, cells were collected for genomic DNA and RNA extraction. To measure deletion efficiency, target sites were amplified using the validation primers flanking the deletion region as indicated in **Supplementary Table 3** and examined using electrophoresis.

ATAC-seq

ATAC-seq was performed as previously described, with minor modifications[21]. Briefly, 50,000 IMR90 cells were collected and resuspended in 50 µl cold lysis buffer (10 mM Tris-HCl pH 8.0, 10 mM NaCl, 0.2% Igepal CA630 (Sigma)). After incubation on ice for 5 min, cells were centrifuged at 800 rcf for 5 min at 4 ^{o}C. The cell pellet was resuspended with 50 µl transposition mix containing 25 µl 2X TD buffer (Illumina FC-121-1030), 3.5 µl Tn5 transposase (Illumina FC-121-1030), and 21.5 µl water. The mixture was incubated at 37 ^{o}C for 30 min with gentle rotation, and transposed genomic DNA was recovered using DNA Clean & Concentrator-5 (Zymo Research D4013). The library was amplified using NEBNext high fidelity PCR master mix (M0541) containing 1.25 µM customized Nextera universal (Ad1_noMX) and indexed primer

with the cycling condition of 72 °C for 5 min, 10 cycles of 98 °C for 30 s, 63 °C for 10 s, 72 °C for 1 min, and final extension at 72 °C for 5 min. The amplified library was purified using SPRI beads (Beckman B23318). A double size selection was performed with 0.5x/1.8x bead volume to remove amplicons > 1000 bp or < 100 bp. Libraries were subjected to 150 bp pair end sequencing on the Illumina HiSeq platform with an expected read depth of 70 million. The FASTQ data was aligned and analyzed using the pipeline from the Kundaje Lab (Github https://github.com/kundajelab/atac_dnase_pipelines).

For ATAC-dPCR for ECAD9, 20 ng of final library was loaded to Quantstudio 3D digital PCR system (version 2, ThermoFisher) and amplified using Taqman array (rs4977575, ThermoFisher C__27869497_10). Allele-specific signal quantification was performed using the online cloud application provided by the manufactuer.

Tri-4C and single RE UMI-4C library construction

To generate the preamplificiation library, Tri-4C adapted the *in situ* Hi-C and UMI-4C protocols[7,10]. 10^7 cells were fixed with 1% (v/v) formaldehyde (Thermo Fisher 28906) in PBS for 10 min at room temperature. Fixation was quenched by adding 2.5M glycine, dropwise, to a 0.2M final concentration and incubating for 5 min at RT. Cells were washed with cold PBS twice and pelleted (300 rcf, 4 min, 4 °C) in 2ml Eppendorf LoBind tubes. Pellets could be immediately used for downstream procedures, or snap-frozen with liquid nitrogen and stored at -80 °C.

To prepare crude nuclei, the cell pellet was resuspended in a premixture of 250 µl cold lysis buffer (10 mM Tris-HCl pH 8.0, 10 mM NaCl, 0.2% Igepal CA630 (Sigma)) and 50 µl protease inhibitors (Sigma P8340). After mixing thoroughly, the suspension was incubated on ice for 15 min and centrifuged (1000 rcf, 5 min, 4 °C). The pellet was washed once with 500 µl of cold lysis

buffer and carefully resuspended in 50 µl of 0.5% SDS. The suspension was then incubated in a 62 °C heating block for 7 min, followed by mixing with 145 µl water and 25 µl 10% Triton X-100 (Sigma), and incubated at 37 °C for 15 min for quenching. To carry out triple digestion, the suspension was mixed with 50 µl of buffer G (ThermoFisher), 120 U MboI (DpnII) (Thermo Fisher ER0811, 10 U/µl), 120 U Csp6I (CviQI) (Thermo Fisher ER0211, 10 U/µl), 100 U Hin1II (NlaIII) (Thermo Fisher ER1831, 5 U/µl), and x µl of water, where x is determined to bring the total volume to 500 µl, empirically within a range of 100-150 µl, depending on the pellet size.

The genomic triple digestion can be alternatively performed by using a combination of MboI (Thermo Fisher), Csp6I (Thermo Fisher) and CviAII (NEB) to generate consistent 5' TA overhangs for other 3C-derived protocols requiring biotin-dA filling. In this case, after Triton quenching, the nuclei suspension was mixed with 50 µl of 10x Custmart buffer (NEB) and 100 U CviAII (NEB), and diluted to 500 µl. The mixture was incubated at 25 °C with rotation for 2 hours, and then 37 °C for two hours or overnight after adding 120 U MboI and 120 U Csp6I.

For Single RE UMI-4C[7] experiments, the suspension was mixed with 50 µl of buffer R, buffer B, and buffer G, respectively, for digestion using 100 U MboI, 100 U Csp6I, or 100 U Hin1II in 500 µl final volume. All digestions were conducted at 37 °C overnight with rotation.

On the second day, the restriction enzymes were inactivated by incubating at 65 °C for 20 min. After cooling to room temperature, end blunting was performed by adding 3 µl of 10 mM dNTP, 8 µl of DNA polymerase I Klenow (NEB M0210, 5U/µl), and 4 µl of T4 DNA polymerase (NEB M0203, 3U/µl), and incubating at 37 °C for 1 hour with rotation. For blunt end ligation, the suspension was then mixed with 460 µl water, 120 µl T4 ligase buffer (NEB B0202), 100 µl 10% Triton X-100, 6 µl 20 mg/ml BSA (NEB B9000S), and 5 µl of 400 U/µl T4 ligase (NEB M0202S/L), and incubated for 4 hours at room temperature with rotation. The processed nuclei

were pelleted (1000 rcf, 5 min, 4 °C) and resuspended in 500 µl 1x T4 ligation buffer supplemented with 50 µl 20 mg/ml proteinase K (Thermo AM2546) and 50 µl 10% SDS, and incubated at 55 °C for 30 min. For de-crosslinking, 60 µl 5M sodium chloride was added and the mixture was incubated at 68 °C for 2 hours. Note that this step can also be prolonged to overnight. Phenol-cholorform extraction was performed to recover DNA as follows: The suspension was washed once with an equal volume of Phenol:Chloroform:Isoamyl alcohol (25:24:1), and once with an equal volume of Chloroform:Isoamyl alcohol (24:1). After phase separation, the aqueous phase was transferred to a new 2 ml LoBind Eppendorf tube, mixed with 60 µl 3M sodium acetate, 1 µl GlycoBlue coprecipitant (Thermo AM9515) and 1.5 ml pure ethanol, and incubated at -80 °C for 15 min. The mixture was centrifuged at max speed at 4 °C for 15 min. The supernatant was carefully removed and precipitated DNA was washed twice with 1 ml cold 70% ethanol. The DNA pellet was air dried for 15 min, and resuspended in 130 µl 10 mM Tris-HCl (pH 8.0) for 1 hour at room temperature, or overnight at 4 °C.

The re-arranged genomic DNA was sonicated to 300-400 bp fragments using a Covaris S2 ultrasonicator. The parameter guidelines from the manufacturer were used, with settings of Intensity (4), Duty cycle (10%), cycles per burst (200), and time (80 sec) as starting points. In general, multiple rounds (typically 2) with the above parameters were run to obtain the desired fragment size peak of 300-400 bp, which was confirmed by Bioanalyzer (Applied Biosystems). The fragmented DNA was double size-selected using 0.40x/1.0x of SPRI beads (Beckman) to remove fragments below 100 bp and above 1000 bp, and eluted in a final volume of 70 µl of 10 mM Tris-HCl. To repair the sonicated fragment ends, the eluent was mixed with 10 µl 10x T4 ligation buffer (NEB), 10 µl 100 mM ATP, 5 µl 10 mM dNTP mix, 4 µl T4 DNA polymerase (NEB), 1 µl DNA polymerase Klenow (NEB), and 5 µl T4 PNK (NEB M0201, 5 U/µl), and incubated for 30 min at room temperature. The repaired DNA was purified by using 1.0x SPRI beads, and eluted in a master mix of 94.5 µl 1x NEB buffer 2 and 0.5 µl 100 mM dATP. After

removing the beads, 5 µl of Klenow exo- (NEB M0212S/L, 5U/µl) was added, and the mixture was incubated at 37 °C for 30 min for dA tailing. The processed DNA was purified by using 1.0x SPRI beads, and eluted in 20 µl 10 mM Tris-HCl.

We designed a custom Y-shape adaptor to generate Illumina next-generation sequencing libraries:

Foward: GATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T

Reverse: /5Phos/GATCGGAAGAGCCATACAGC

The oligos were synthesized using the IDT Ultramer service (Integrated DNA Technologies). The forward and reverse single strand oligos were annealed (95 °C for 5 min, down to 25 °C at 0.1 °C/sec temperature gradient) and prepared at 30 µM stock concentrations. Five µl of adaptor was added to the A-tailed libraries, and mixed with 25 µl blunt/TA ligase master mix (NEB M0367). The mixture was incubated at room temperature for 15 min, and purified with 1.0x SPRI beads. After eluting in 50 µl Tris-HCl, the libraries were purified a second time with 1.0x SPRI beads to completely remove the residual adaptors. The final preamplification libraries were eluted in 100 µl Tris-HCl, and examined by Bioanalyzer (Applied Biosystems) to ensure correct size distributions and absence of unligated adaptors. The size distributions of mature libraries after incorporating adaptors were centered around 500 bp.

To generate the final Tri-4C and single RE UMI-4C libraries, we designed a pair of outer and inner primers, based on the restriction enzyme, for each viewpoint to increase amplification specificity (**Supplementary Table 1**). For amplification with outer primers, eight 100 µl reactions, each containing 400 µg preamplification library, 2 µM universal primer (AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC), 0.5 µM viewpoint-specific outer primer for each multiplexed viewpoint, and 1x SuperFi PCR master mix (Thermo 12358010) with 20% GC enhancer, were amplified with the following conditions: 98 °C

for 30s, 14 cycles of 98 °C for 10s, 62 °C for 10s, and 72 °C for 60s, and final extention at 72 °C for 5 min. All primers were synthesized using the IDT Ultramer service (Integrated DNA Technologies). The products were pooled and purified with 1.0x SPRI beads, and amplified with the inner primer pair (Illumina P5 + bait-specific P7 index-attached reverse primer) for 14 cycles using the same conditions. After purification with 1.0x SPRI beads, the products were quantified using the Qubit DNA assay kit (Thermo Q32851), examined by Bioanalyzer (Applied Biosystems), and diluted to 10 nM to be sequenced on Illumina platforms.

We aimed for a read depth of 5 million reads for each viewpoint. For a typical library containing 100,000 unique fragments, this results in 50x coverage. The high coverage is desired as Tri-4C generates more reads than single RE UMI-4C with the same DNA input. In practice, the actual yield varies in multiplexed libraries, possibly due to primer efficiency and off-target amplification. A minimum depth of 1 million reads was required for all our experiments. Sequencing was performed on Illumina platforms (MiSeq/HiSeq) in paired read mode with read lengths of 75-150 bp.

Data Analysis

Due to the utilization of the multiple restriction enzymes, most currently existing C pipelines are not applicable to Tri-4C. However, alignment and processing is straightforward, as described below. After demultiplexing, the reverse end (R2, viewpoint end) of the FASTQ file was used to filter reads that were correctly ligated with the viewpoint by matching the sequence head with the inner primer sequence and the padding sequence. We used FASTX Barcode Splitter (http://hannonlab.cshl.edu/fastx_toolkit/index.html) for this step, with an allowance of 1 mismatch. The tool also trimmed the viewpoint sequence during the process. For allele-specific analysis, the reads were splitted by matching the allele with the tag SNP on the padding sequence using an awk command. The undigested/unligated ratio was calculated at this step by

measuring the fraction of trimmed reads, starting with the immediate downstream sequence from the viewpoint. After trimming, the residual reverse end was mapped together with the forward end (R1, sonication end) using BWA mem with default paired end alignment settings against the hg19 genome. The aligned reads were deduplicated according to the mapped position of the sonication end (5' for the reads on the + strand and 3' for the - strand) by using a simple AWK script. We considered reads with sonication ends separated by 1 bp as duplicates based on the observation that the Illumina platform occasionally bypassed the first nucleotide of the read. Reads with low quality (MAPQ = 0) were removed from analysis. The complexity of the library (number of unique reads) and intrachromosome ratio were meansured at this stage. Reads from 1kb upstream to 2kb downstream of the viewpoint were removed as these regions were consistently highly interactive and subjected to over deduplication due to saturation of unique sonication ends. Interchromosomal interactions were also excluded from downstream analysis since no loop-like interaction hotspots outside the same chromosome were observed. For standard analysis, the processed reads were binned in 500 bp, with a sliding step of 100 bp for both visualization and other downstream analyses, with the exception of reproducibility tests (**Fig 1C, Supplementary Fig 3C, D**), where reads were binned with the indicated bin size and equal step size. For comparison of the analysis at the conventional lower resolution, reads were binned in 3000 bp, sliding at 100 bp. Of note, we used the entire aligned sequence for this step, instead of assigning each read to its corresponding restriction site, for two reasons. First, we observed a clear directional bias on the target restriction sites, which indicated that the real contact point was not at the RE site but in its close vicinity. Thus, a short overhang during alignment created a benign bias pointing toward the real interaction spot. Second, the Tri-4C method was designed for promoter/enhancer viewpoints which often contained high/low GC contents. The difficulty of designing primers for these regions sometimes results in a long padding sequence and an unmappable short residual sequence on the reverse end after trimming, yielding reads that cannot be matched to the corresponding RE sites. The piled raw

read count bedGraphs were used to perform peak calling. To ensure fair comparison, single RE UMI-4C libraries were analyzed in parallel using the same pipeline.

Loop Peak Calling

We used a local fold enrichment-based strategy to identify significant interaction loop peaks for the Tri-4C and singe RE UMI-4C data. Thus, the expected number of reads (background) for a given bin with read count *M* was estimated by taking the average of neighboring bins. We used the smallest mean value of 5 kb, 10 kb, 20 kb, and 50 kb intervals centered at the bin location to represent the background *N*. Then, significance *p* values were calculated by $p = Pr\{X \geq M\}$ given $X \sim Poisson(N)$. Of note, this step can be achieved by feeding the MACS2 bdgcmp function with the background and signal tracks using –m ppois mode[18].

To identify significant and reproducible peak regions, bins were scored with the –log10(-log10(p)) value, and those with a score > 0 ($p < 0.1$) in all replicates were collected and analyzed by IDR package[19] (Github https://github.com/nboley/idr) using the following settings

--initial-mu 1.5 --initial sigma 0.3 --initial-rho 0.8

Bins with IDR < 0.05 (score ≥ 540) were considered significant and merged. We defined a minimum length of 300 bp for calling significant distal loop peaks.

The UMI-4C[7] and 1D adaptation of *in situ* Hi-C[9] loop calling algorithms were used for comparison. For both algorithms, distance-dependent decay of interaction frequency was calculated at genome-wide level using IMR90 *in situ* Hi-C data at 5 kb resolution. The decay function was further smoothed to 500 bp bins (*W*) in 100 bp step size resolution by using linear interpolation to obtain the *F(d)* (UMI-4C) or *E** (Hi-C) suitable for high resolution analysis in Tri-

4C. For the UMI-4C algorithm, the background of each Tri-4C profile was obtained by re-allocating the total intra-TAD Tri-4C read counts (N) to each bin according to $F(d)$. The enrichment p value of a bin with expected read count of E and actual count of E_1 was calculated by fitting to binomial distribution: $Pr(B(N, E/N) > E_1)$. For the 1D *in situ* Hi-C algorithm, the adjusted expected read count for each bin $E^{d^*}_i$ can be calculated by the filter fomula

$$E_i^{d^*} = \frac{\sum_{a=i-w}^{i+w} M_a^* - \sum_{a=i-p}^{i+p} M_a^*}{\sum_{a=i-w}^{i+w} E_a^* - \sum_{a=i-p}^{i+p} E_a^*} \times E_i^*$$

Where p and w were set at 2 and 100 to be corresponding with 500 bp peak and 20 kb background size. This expected count was compared with actual read count M_i using the Poisson statistics. The p values obtained from UMI-4C and Hi-C algorithms were directly compared with Tri-4C raw p value by the CRE ROC analysis.

To investigate for potential artifacts during Tri-4C loop calling due to mapping bias, the GC content and restriction site density under triple enzyme digestion for the locus analyzed by Tri-4C were obtained by directly analyzing the hg19 genome sequence of the region. Mappability of the region was obtained from the ENCODE mappability track available on the USCS genome browser. The average GC content, restriction site density, and mappability for the 10 kb neighboring regions for all loop sites called by Tri-4C were calculated at 100 bp resolution. To generate the background for comparison, a set of 10,000 equal size genomic intervals were randomly selected in the locus. The mean for each set was calculated after removing intervals whose center fell within Tri-4C loops or repeat regions.

Data Reproducibility

Each Tri-4C and SRE-4C was performed in two technical replicates. Reproducibility of intrachromosomal interaction was measured by Pearson's correlation r using the Python numpy.corr function after binning the contacts in the size described in **Fig1C**.

Hi-C and Topologically-Associated Domain (TAD) Definition

The IFNB1 TAD (chr9:19480000-2120000) was defined by IMR90 Hi-C data from the Aiden[7] lab. Hi-C Browser (http://promoter.bx.psu.edu/hi-c/index.html) was used to visualize the Hi-C interactions in the TAD.

Analysis of Interaction Frequency and Loop Strength

For frequency-based analysis, the read count for each bin was converted to interaction frequency by normalizing against (1) total intrachromosomal interactions, which we referred to as normalized interaction or (2) total intrachromosomal interactions of a reference viewpoint in a multiplexed run, which we referred to as relative interaction frequency. The purpose of the latter method was to control for the significant change in total read count generated between different experiental conditions, as in the IFNB1 viewpoint at the baseline control compared to the IFNB1 induced condition (**Fig 2C, Supp Fig 9B**). Thus, the reference point was selected based on the standard of exhibiting the least variation in unit read count yield among different conditions, and in our experiment the Boundary viewpoint was chosen for that reason. Subtraction analysis was performed by directly calculating the interaction frequency difference between two tracks in comparison after normalization. A multiplication factor of 10,000 was applied to the normalized interaction frequency to simplify the display when presenting the interaction map on UCSC/WashU track.

For loop-based analysis, loop strength, i.e. log fold enrichment (logFE) was calculated by *logFE = log10(M/N)*, where M and N are the actual and expected read count for each sliding window

as indicated in the Peak Calling section. This step can be perfomed with MACS2 using –m logFE. A 1.0 pseudo count was given to calculate logFE to resolve zero division as well as attenuating noise level at regions with sparse mapped counts. Differential loop strength was calculated by measuring the logFE difference between two tracks. For presentation in **Fig 2C**, the logFE was weighed by frequency at basal condition.

Allele-specific Loop Calling

We used a likelihood ratio test, based on the loop strength, to determine whether a interation loci displayed allelic bias. Specifically, for each 100 bp sliding window we obtained four vectors: M_{ref}, M_{alt}, N_{ref}, and N_{alt}, where *ref* and *alt* denoted the allele genotype, and *M* and *N* denoted the actual and expected read count, as indicated above. Each element in the vector represented one replicate. Firstly, the elements in N_{ref} and N_{alt} were normalized to their respective mean, and the scalars were used to normalize their corresponding *M*. Then, we had H_1: $\hat{M}_{ref} \sim Pois(\hat{N}_{ref})$; $\hat{M}_{alt} \sim Pois(\hat{N}_{alt})$, and H_0: $\hat{M}_{ref} \sim Pois(N_0)$; $\hat{M}_{alt} \sim Pois(N_0)$, where N_0 denoted the mean of $\hat{N}_{ref}$ and $\hat{N}_{alt}$. The likelihood ratio *L* was converted to p value by applying the Wilks' theorem, namely $-2ln(L) \sim \chi^2(1)$, subjected to subsequent Bonferroni correction where n equaled to total number of assayed intervals in the locus.

Analysis of Cis-regulatory Element Interaction Network

To annotate CREs and active enhancers, respectively, DNase and H3K27Ac peak position and intensity for IMR90 cells were obtained from the Roadmap Project web portal (https://egg2.wustl.edu/roadmap/web_portal/). For the comparison of loop strength and interaction frequency between viewpoints with DNase peak intensity, linear regression models were built in Python using the scipy.stats.linregress function.

Receiver Operating Characteristic (ROC) Analysis

CRE positions, defined by Roadmap DNase peaks, were converted to 100 bp 1/0 tracks, with 1 indicating the presence of peaks inside the bin. The ROC curves were built by using loop scores (-log(p)) obtained from Tri-4C and UMI-4C as predictors for the peak positions in the converted DNase track, using Python sklearn.metrics: roc_curve and auc function with default settings.

Motif Analysis

The DNA sequences for all Tri-4C peak regions in both baseline control and IFNB1-induced conditions were extracted. Motif prediction was performed using TFBSTools (R platform) with accession to the JASPAR2018 database[27, 28]. A minimum score of 90% was set to discover matched motifs. The Lasso CV model (CV=10, iter=10,000) was applied to all motifs to identify factors correlated with ΔlogFE of Tri-4C peaks during induction. Significance of correlation was determined by F statistic and subject to Bonferroni correction.

Code Availability

The Shell pipeline used to align Tri-4C data and interaction loop analysis is available on Github (https://github.com/kimagure/Tri-4C).

Accession codes

Raw and processed data available at NCBI Gene Expression Omnibus, accession number GSE119189.

Acknowledgements

We thank X. Dong and D. Zheng for discussion on analysis, T. Wang for discussion on statisitcs, and M.G. Rosenfeld and J. Vijg for their critical reading of the manuscript. Research in the Suh lab was supported by the National Institute of Health (AG017242, GM104459, AG056278, AG057341, AG057433, AG057706, CA180126, HG008153) and by the Glenn Center for the Biology of Human Aging (Paul Glenn Foundation for Medical Research).

Author Contributions

Y.Z. developed the method, performed experiments and analyzed the data. Y.Z. and Y.S. interpreted the results and wrote the manuscript. Y.S. supervised research.

Competing interests

The authors declare no competing or financial interests.

Reference

1. Shlyueva, D., Stampfel, G. & Stark, A. *Nat Rev Genet* **15**, 272-286 (2014).
2. Calo, E. & Wysocka, J. *Mol Cell* **49**, 825-837 (2013).
3. Catarino, R.R. & Stark, A. *Genes Dev* **32**, 202-223 (2018).
4. Oudelaar, A.M. et al. *Nat Genet* **50**, 1744-1751 (2018).
5. Hsieh, T.H. et al. *Cell* **162**, 108-119 (2015).
6. Andersson, R. et al. *Nature* **507**, 455-461 (2014).
7. Schwartzman, O. et al. *Nat Methods* **13**, 685-691 (2016).
8. Mumbach, M.R. et al. *Nat Methods* **13**, 919-922 (2016).
9. Rao, S.S. et al. *Cell* **159**, 1665-1680 (2014).
10. Bernstein, B.E. et al. *Nat Biotechnol* **28**, 1045-1048 (2010).
11. Zhang, Y. et al. *Genome Biol* **9**, R137 (2008).
12. Splinter, E. et al. *Genes Dev* **20**, 2349-2354 (2006).
13. Liu, Y. et al. *N Engl J Med* **371**, 507-518 (2014).
14. Sato, M. et al. *Immunity* **13**, 539-548 (2000).
15. Almontashiri, N.A. et al. *Circulation* **132**, 1969-1978 (2015).
16. Harismendy, O. et al. *Nature* **470**, 264-268 (2011).
17. Holdt, L.M. & Teupser, D. *Arterioscler Thromb Vasc Biol* **32**, 196-206 (2012).
18. Yost, K.E., Carter, A.C., Xu, J., Litzenburger, U. & Chang, H.Y. *Nat Methods* **15**, 304-305 (2018).
19. Jarinova, O. et al. *Arterioscler Thromb Vasc Biol* **29**, 1671-1677 (2009).
20. Davies, J.O. et al. *Nat Methods* **13**, 74-80 (2016).
21. Bonev, B. et al. *Cell* **171**, 557-572 e524 (2017).
22. Allahyar, A. et al. *Nat Genet* **50**, 1151-1160 (2018).
23. Ren, G. et al. *Mol Cell* **67**, 1049-1058 e1046 (2017).
24. Banerjee, A.R., Kim, Y.J. & Kim, T.H. *Nucleic Acids Res* **42**, 12537-12554 (2014).
25. Yarilina, A. & Ivashkiv, L.B. *Curr Dir Autoimmun* **11**, 94-104 (2010).
26. Patsch, C. et al. *Nat Cell Biol* **17**, 994-1003 (2015).
27. Tan, G. & Lenhard, B. *Bioinformatics* **32**, 1555-1556 (2016).
28. Khan, A. et al. *Nucleic Acids Res* **46**, D260-D266 (2018).

Fig 1

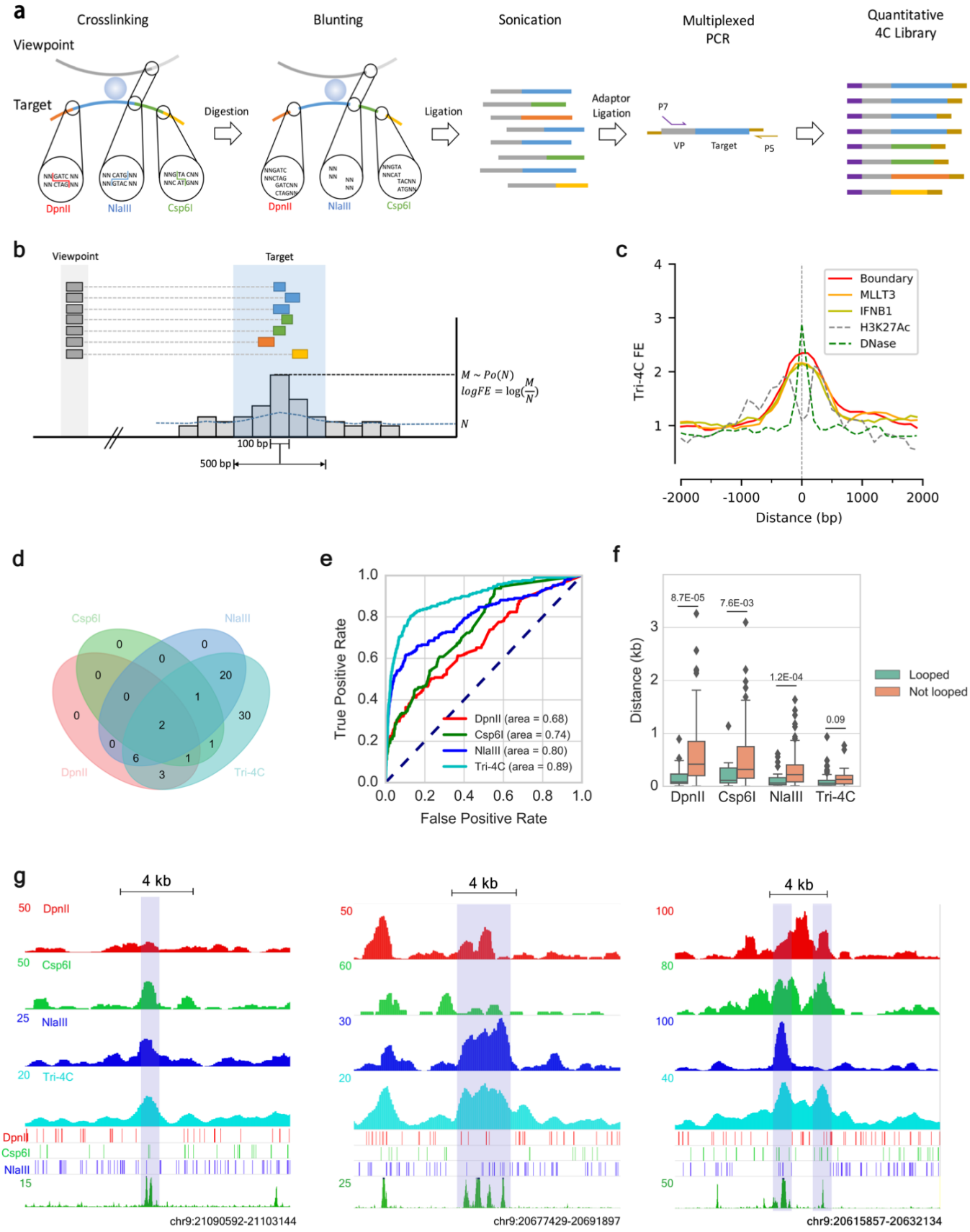
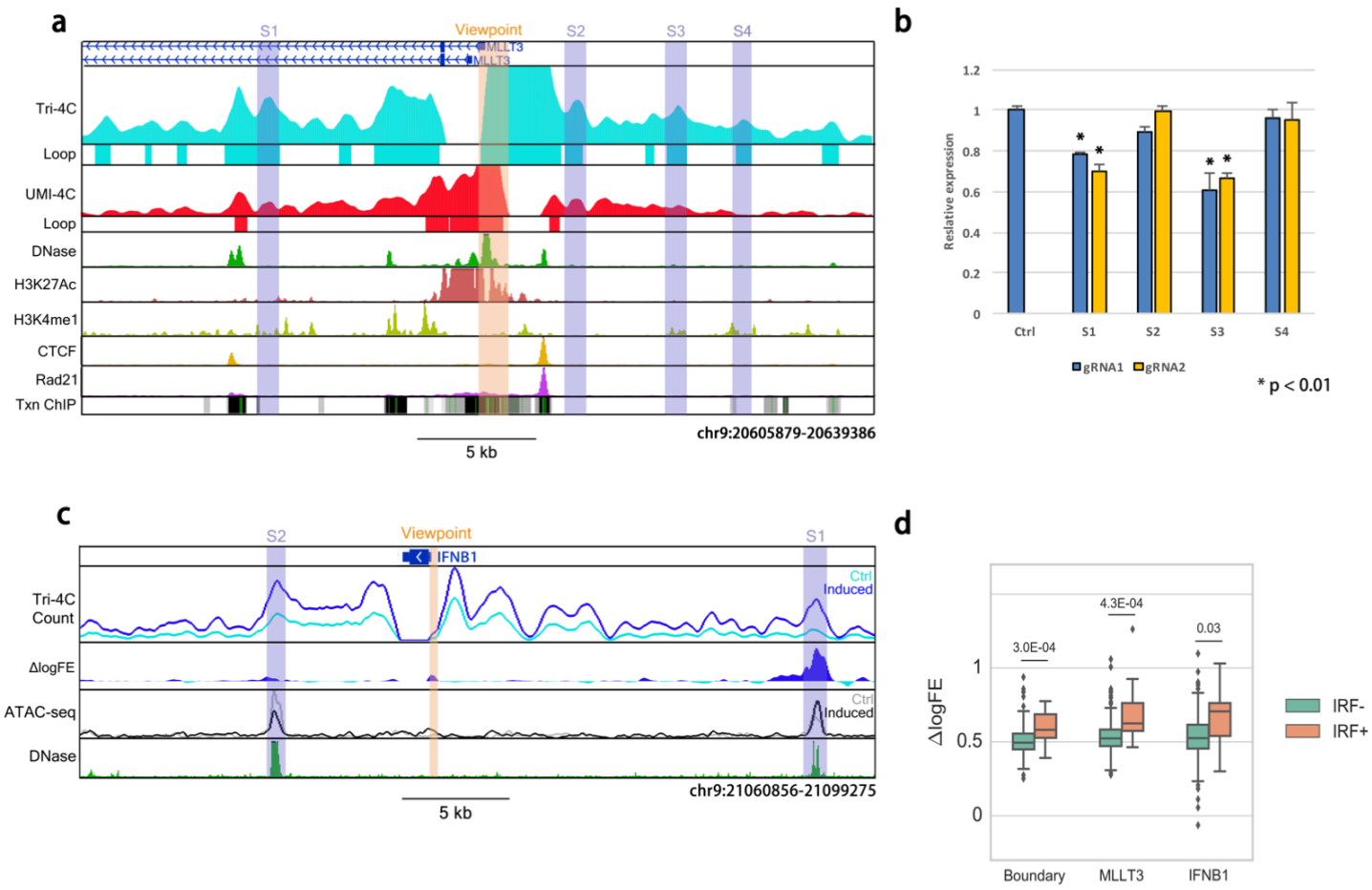
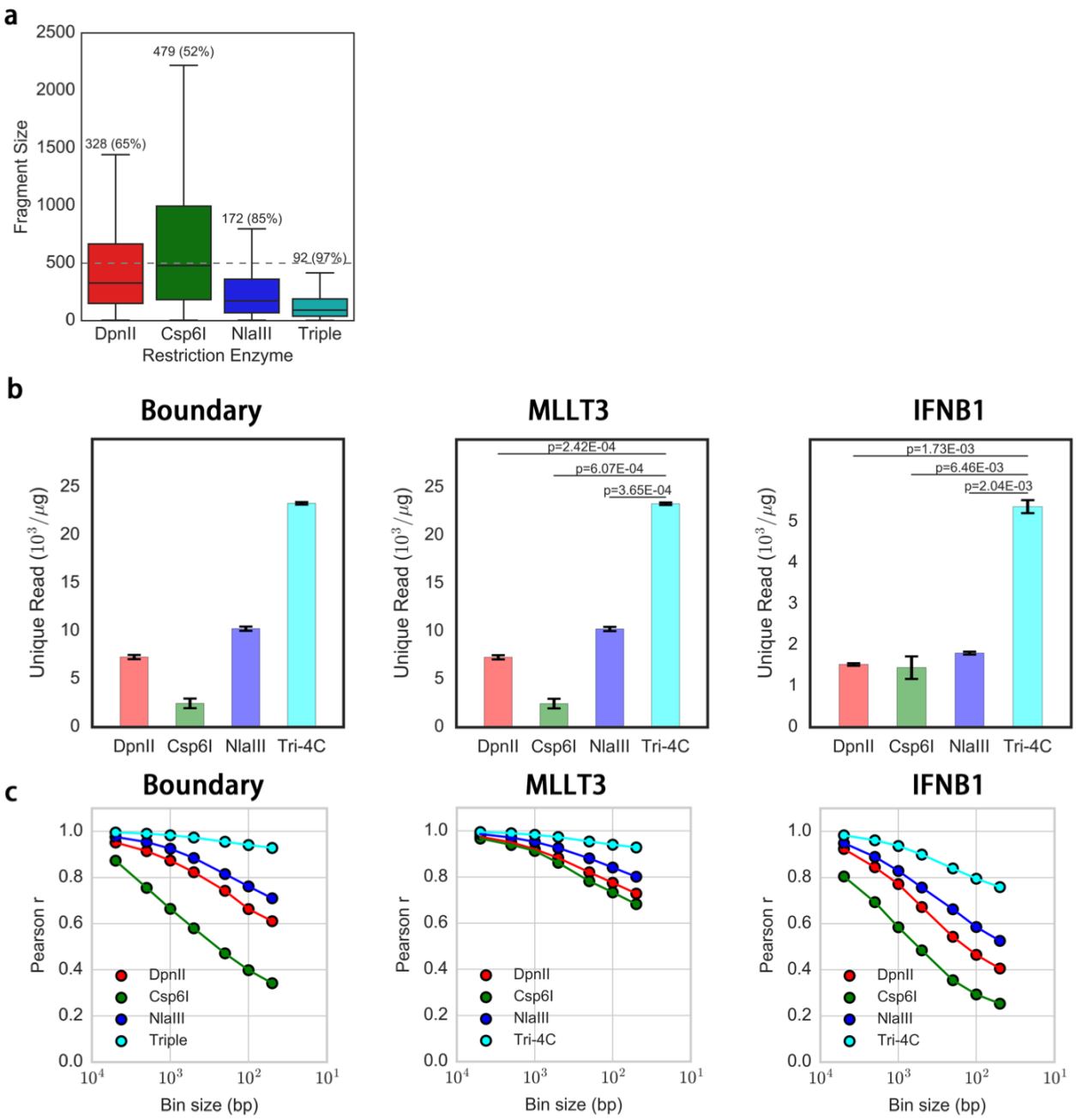


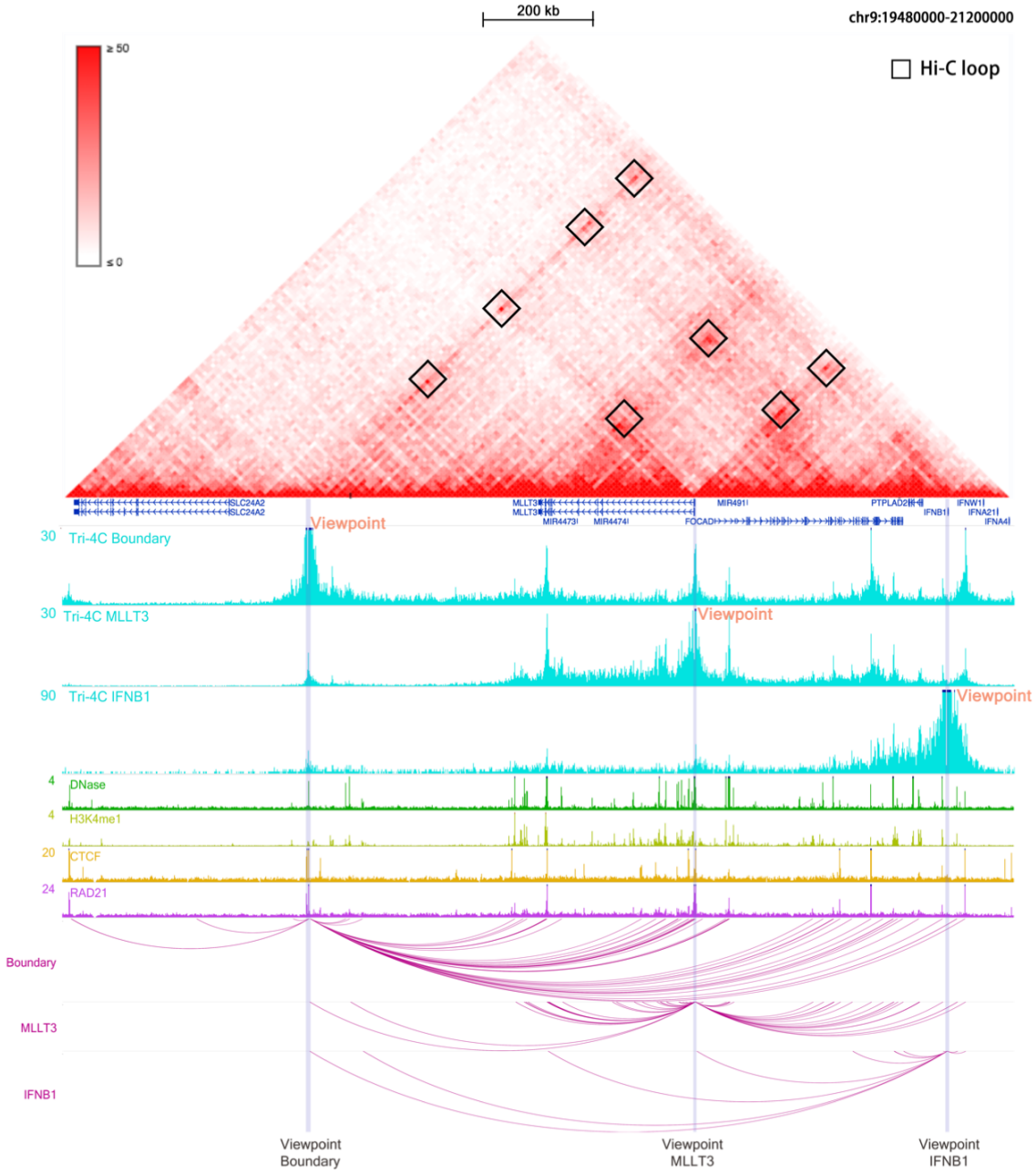
Fig 2



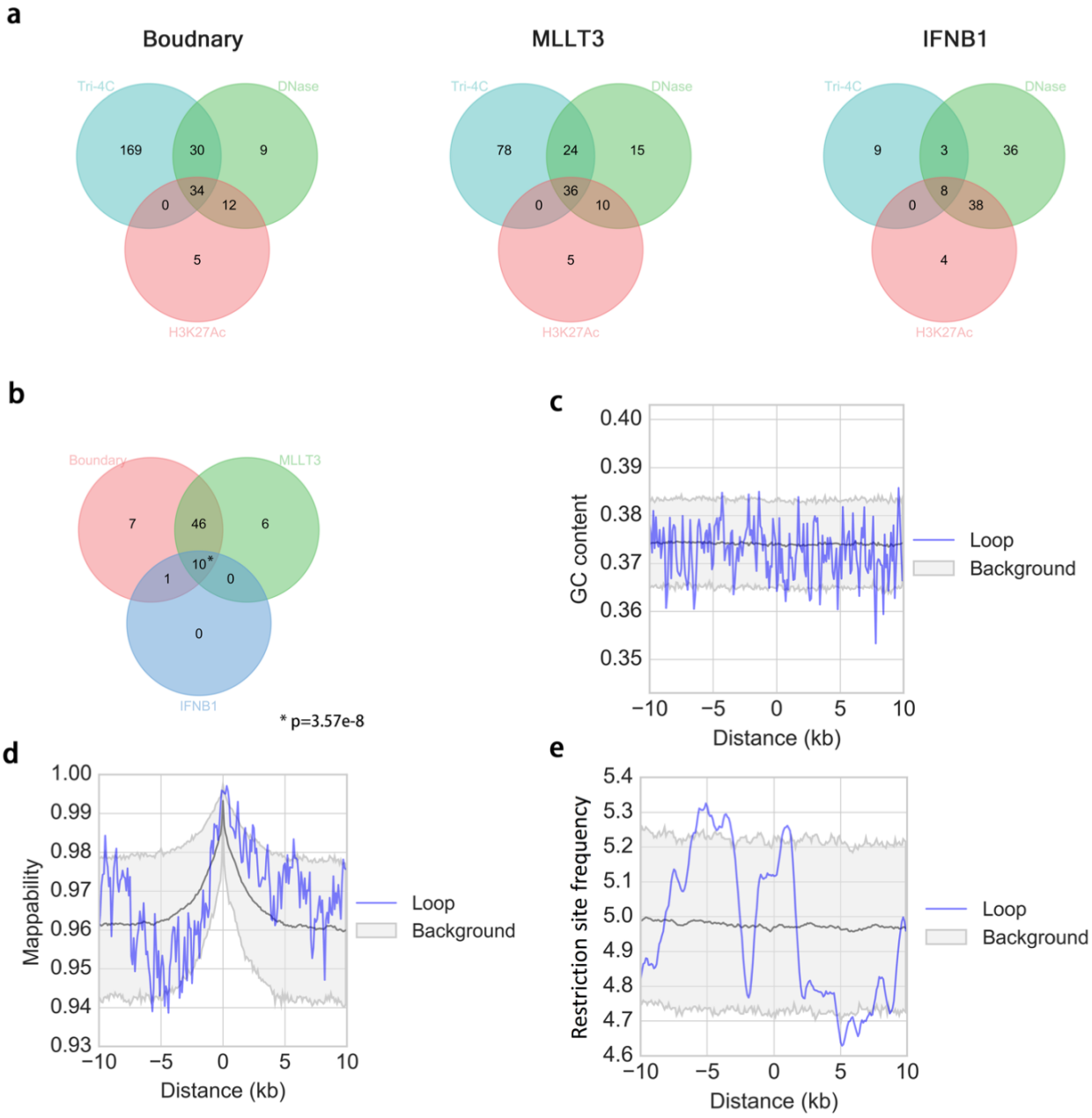
Supplementary Fig 1



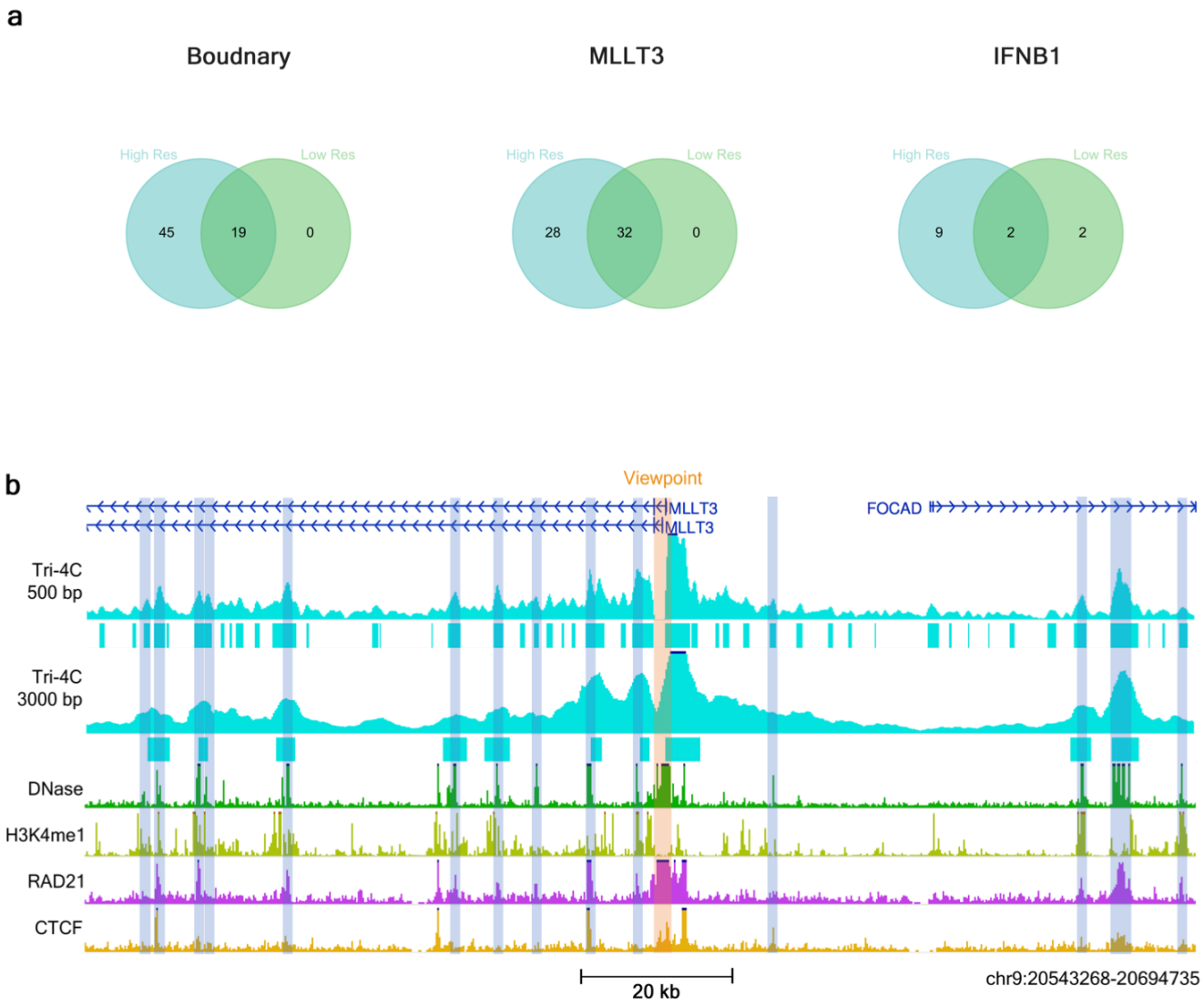
Supplementary Fig 2



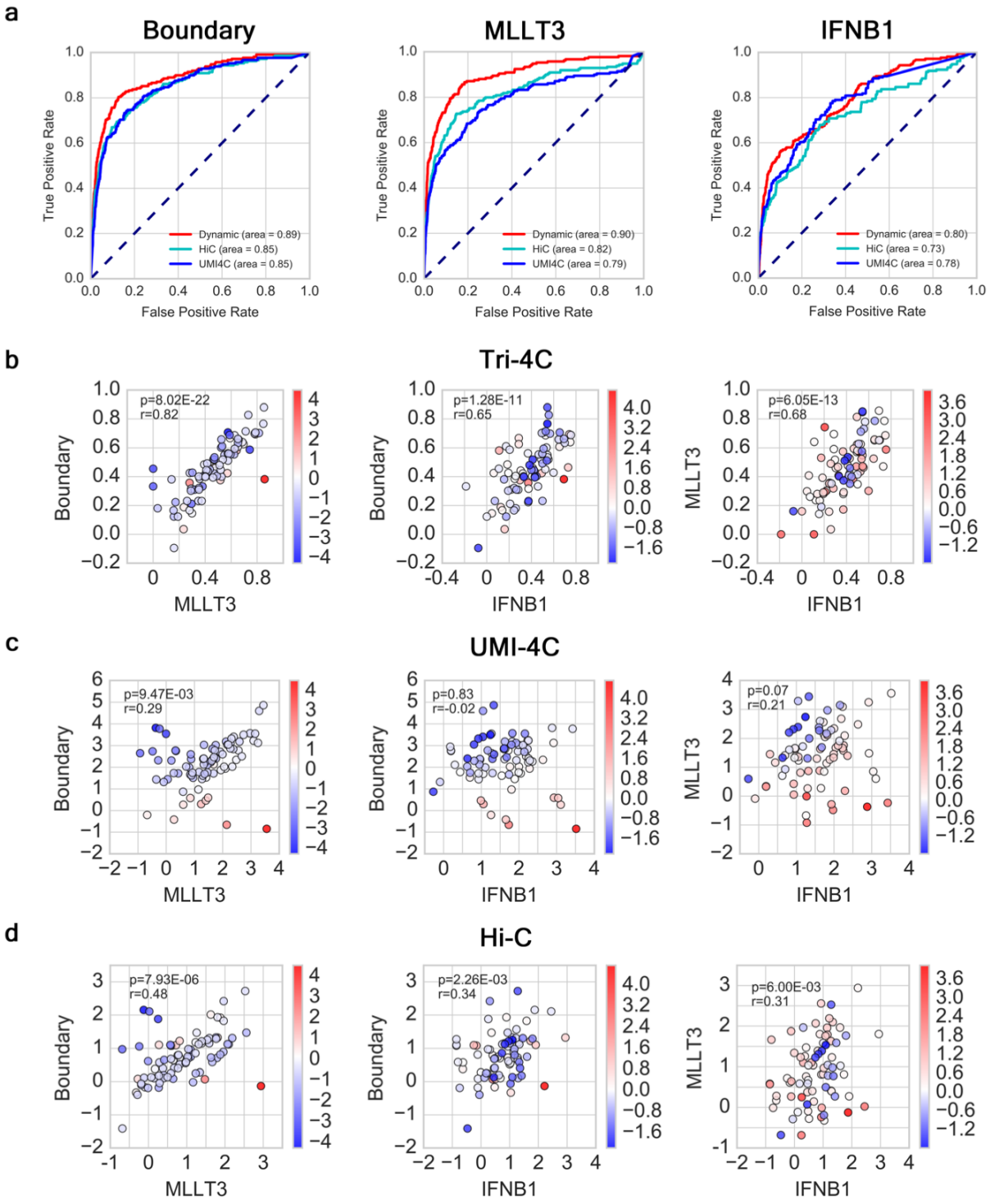
Supplementary Fig 3



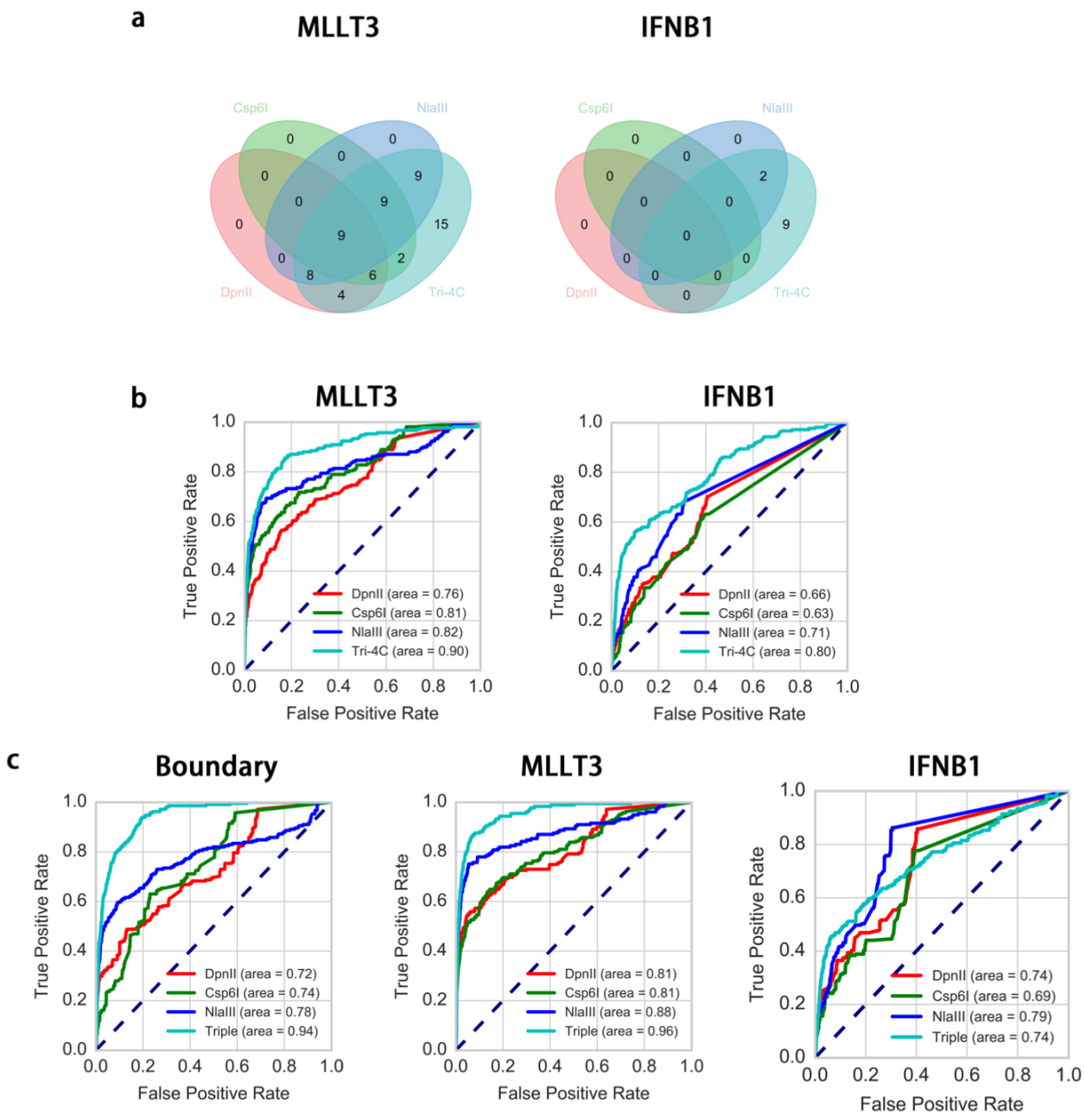
Supplementary Fig 4



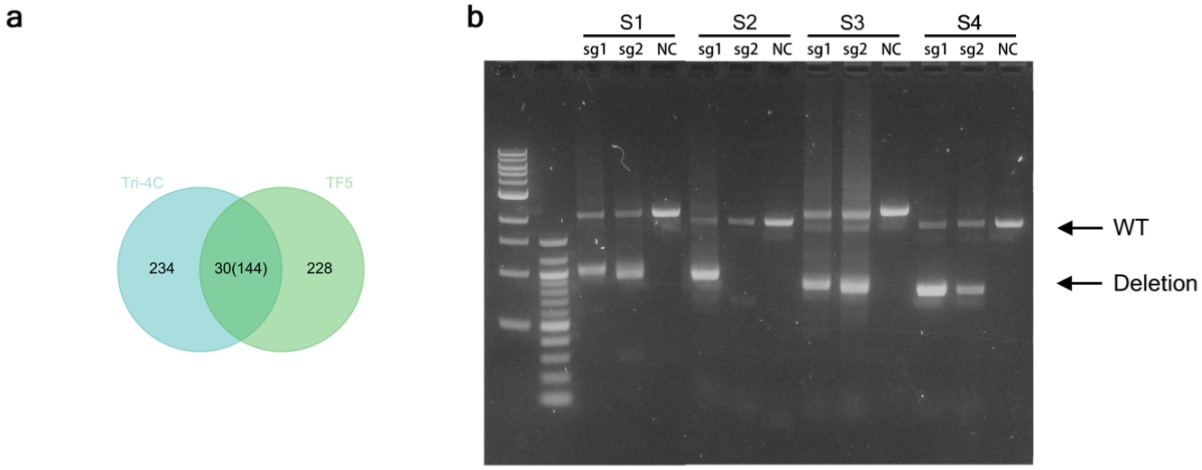
Supplementary Fig 5



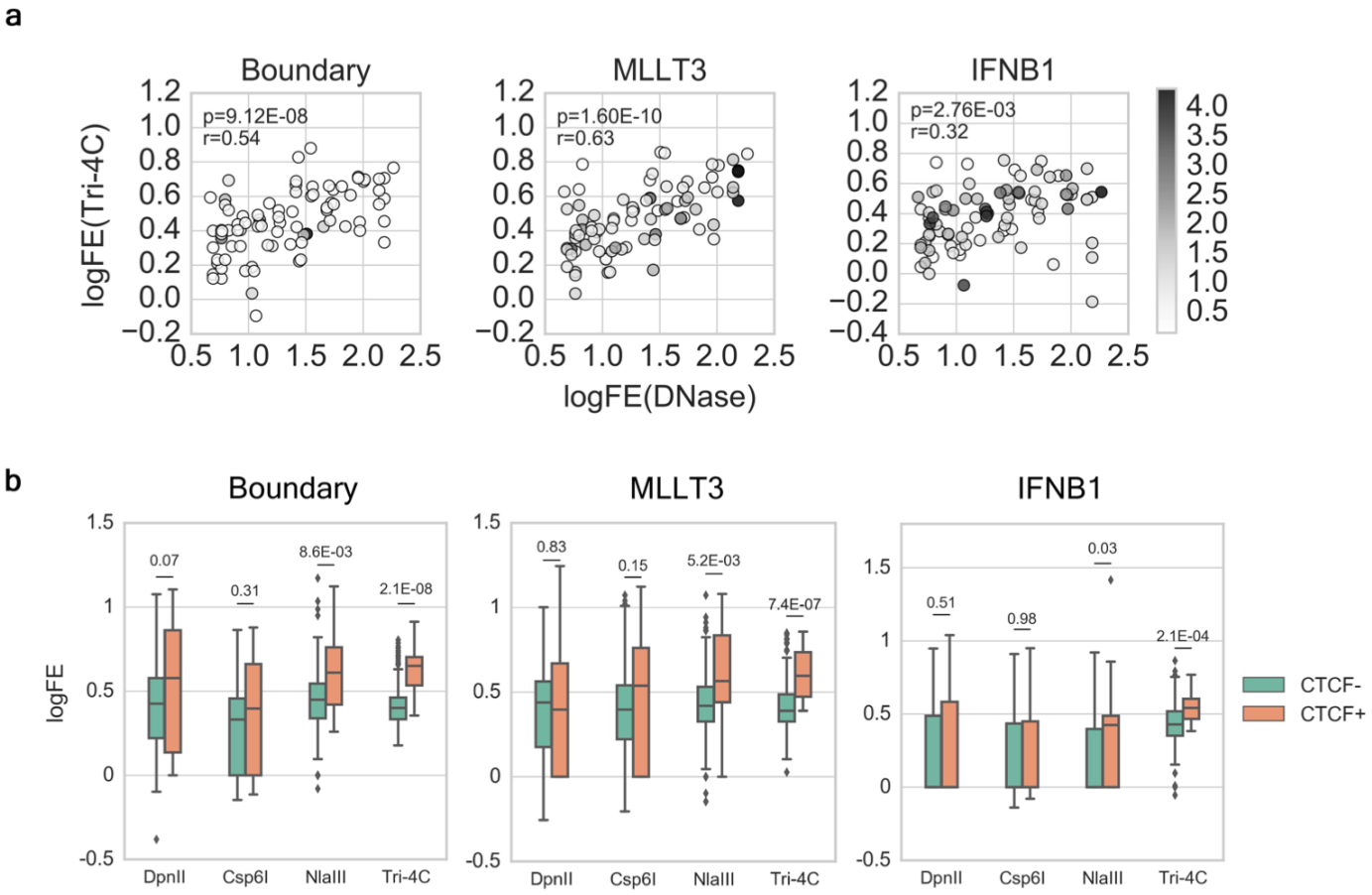
Supplementary Fig 6



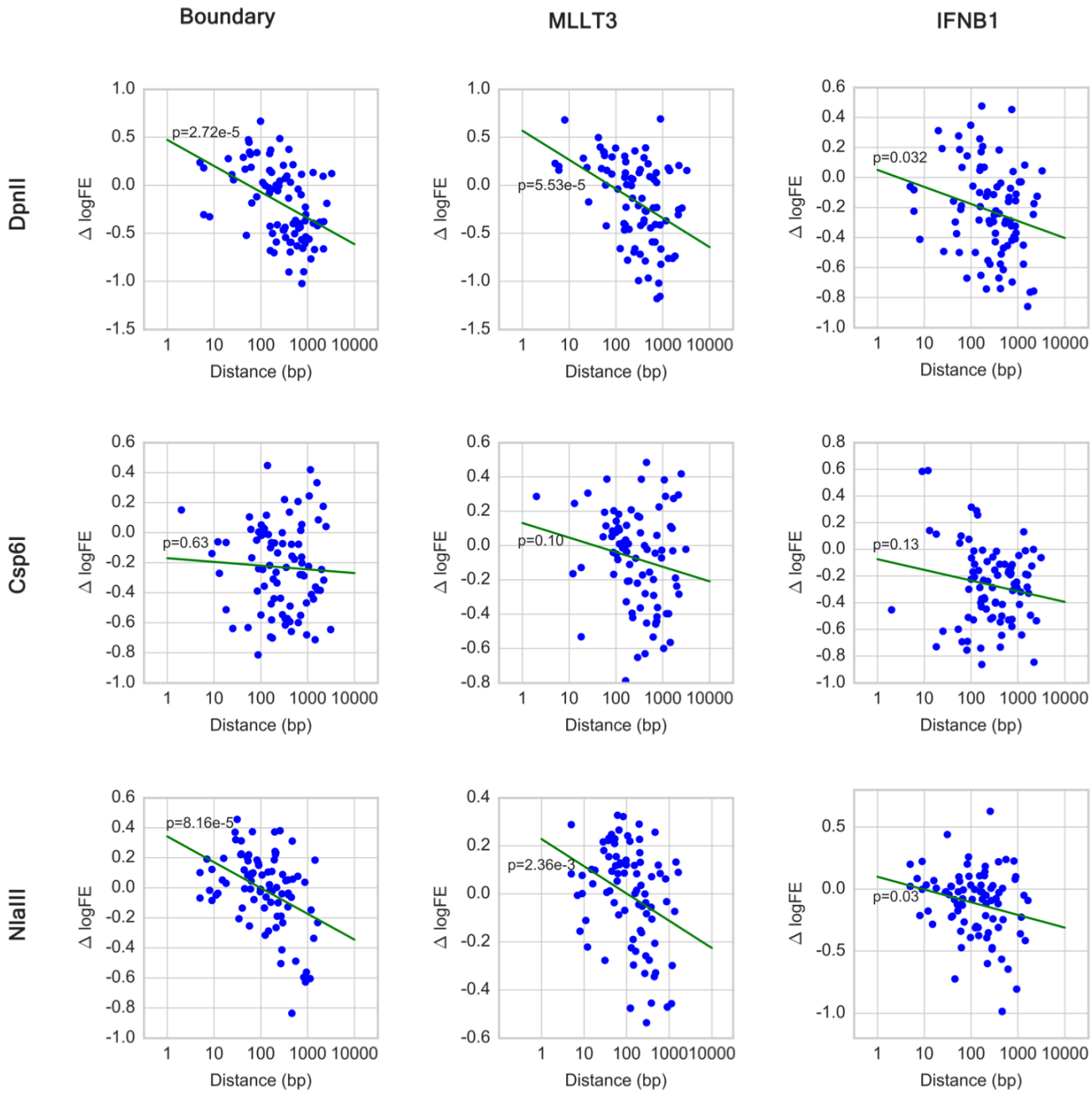
Supplementary Fig 7



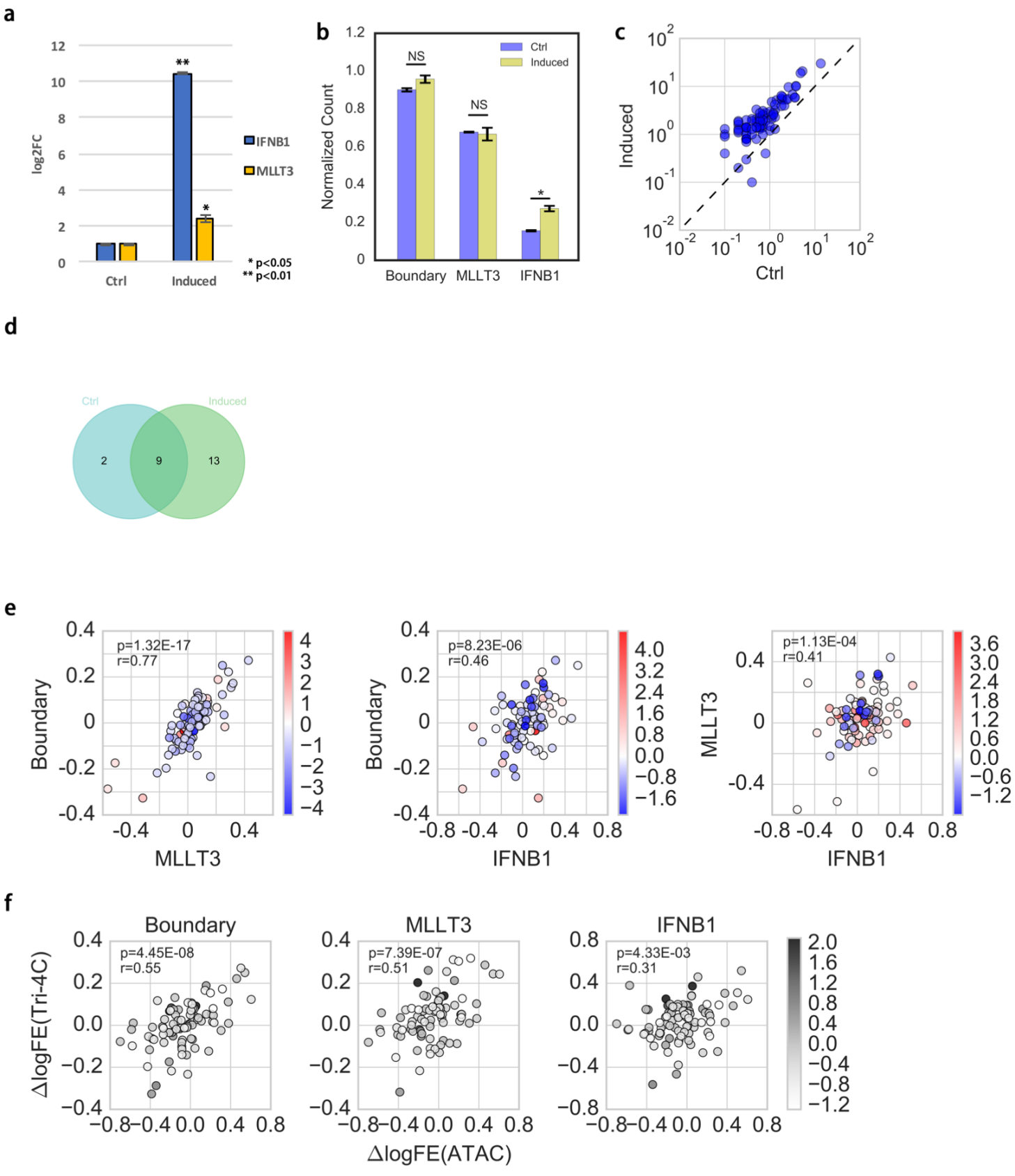
Supplementary Fig 8



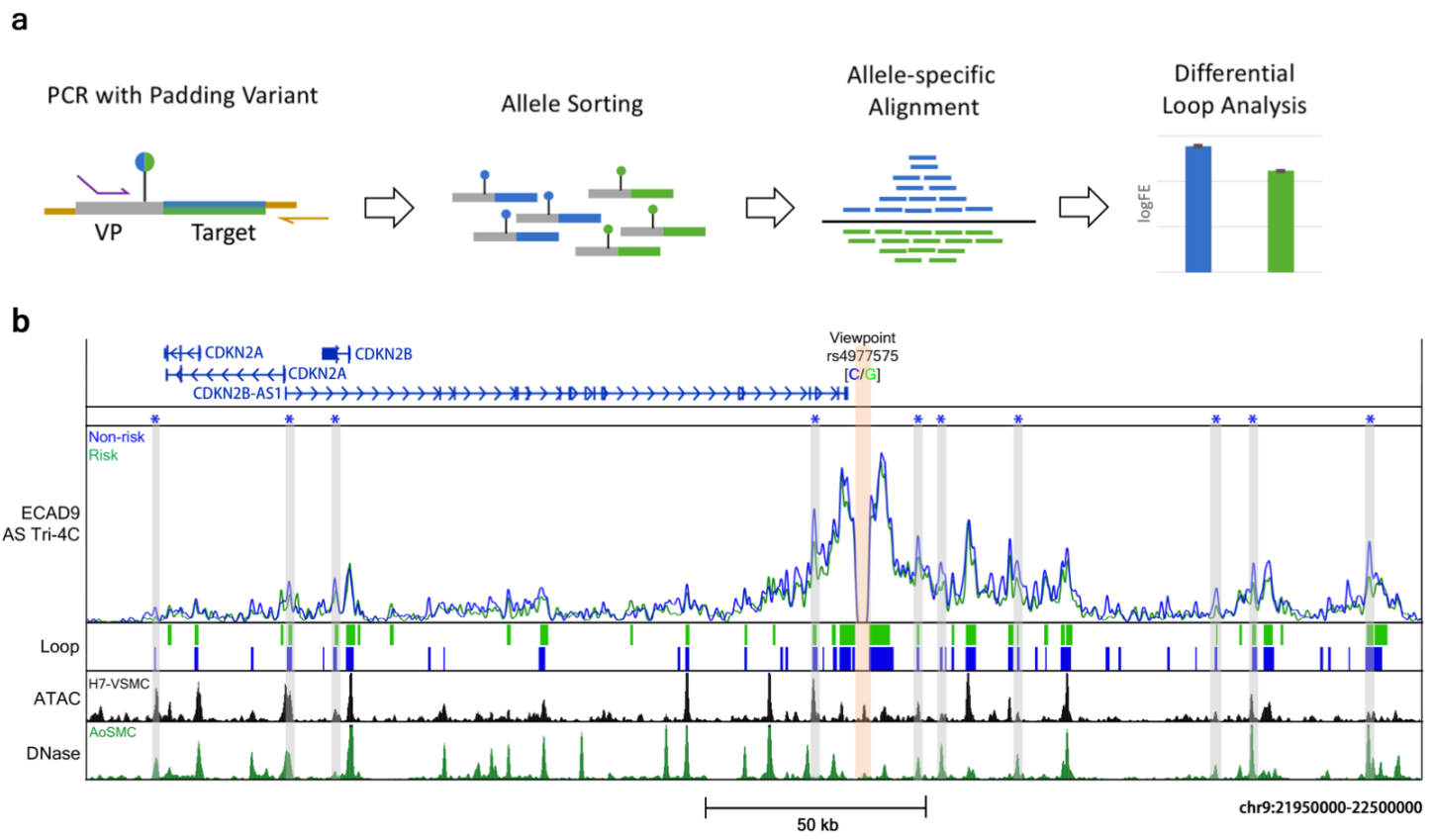
Supplementary Fig 9

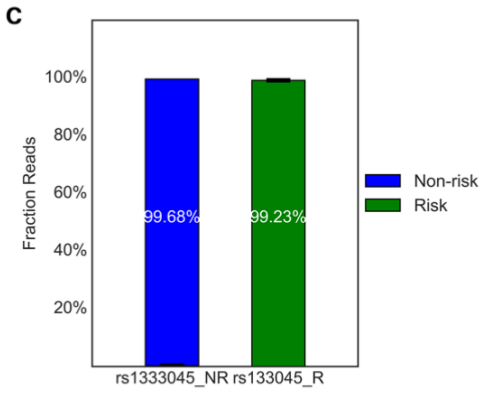


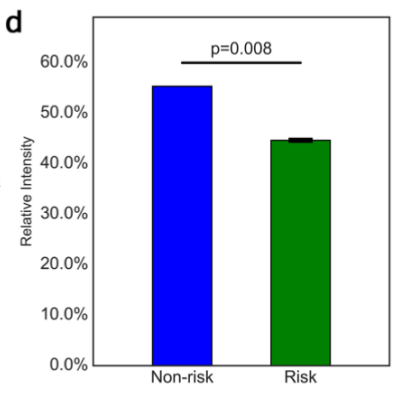
Supplementary Fig 10

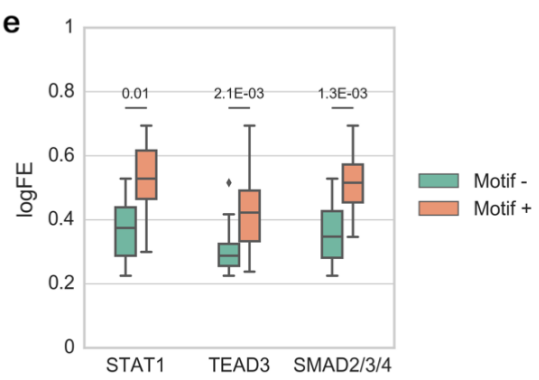


Supplementary Fig 11

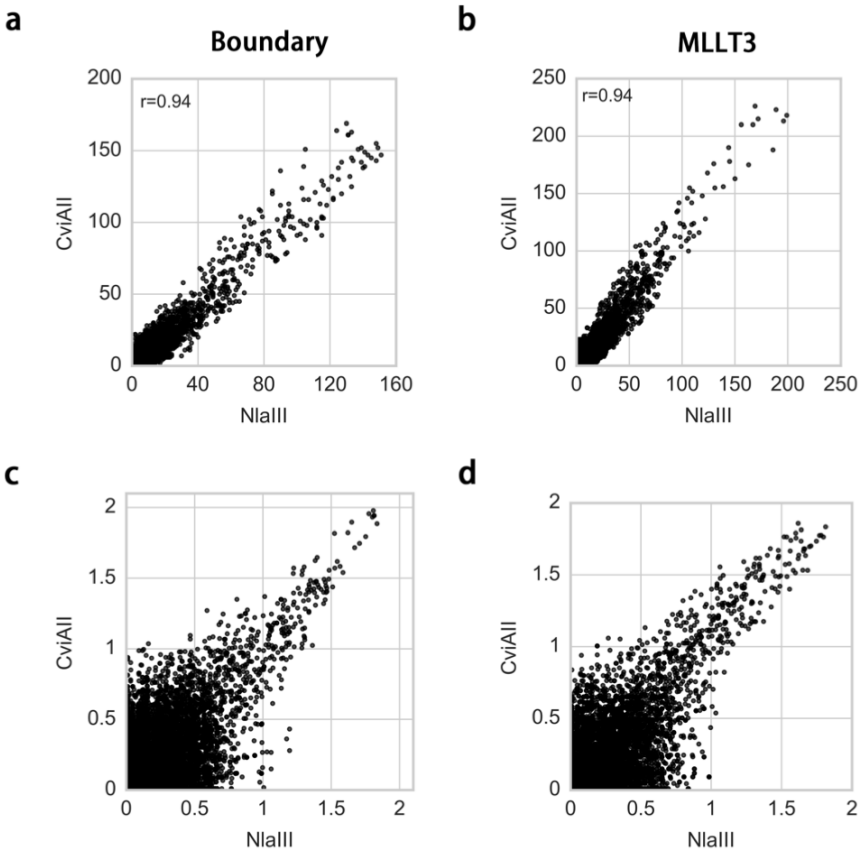








Supplementary Fig 12



Supplementary Table 1

Viewpoint	Position (hg19)	Outer Primer	Inner Primer	Padding Sequence
Boundary-DpnII	chr9:19926377	CACCTTTCATTCAGTACCTCCCTTCA	GAAAGCCCTGTCTCCTGTCCTTG	GACTGG
Boundary-Csp6I	chr9:19926333	TTTCCTTTTCGCTCCGCTGTAGT	CCGCTGTAGTTACGTGACTCACCTTT	CATTCA
Boundary-NlaIII	chr9:19926164	CGCCATAACCTCAACTCAGTCCT	CAGTCCTCTCCCGCCATCTAGT	GGT
Boundary-Triple	chr9:19926164	CGCCATAACCTCAACTCAGTCCT	CAGTCCTCTCCCGCCATCTAGT	GGT
MLLT3-DpnII	chr9:20623064	AACGAGCATGAAAATAGCAATGACTGA	AGCAATGACTGAGGCTGTTCTAACC	CGCAAC
MLLT3-Csp6I	chr9:20622883	AGCCCTCAGCAGCCGGAGA	CGGAGAGGGGGTGTTAAATCAAGT	CCTTAG
MLLT3-NlaIII	chr9:20622509	CTCCCTCCGCCCTGTGAG	CTCGGAGCCCGGGTGTCA	GGCGCCACGGCG
MLLT3-Triple	chr9:20622509	CTCCCTCCGCCCTGTGAG	CTCGGAGCCCGGGTGTCA	GGCGCCACGGCG
IFNB1-DpnII	chr9:21077304	GAGTGGAAATCCTAAGGAACTTTTACTTCA	TTAACAGACTTACAGGTTACCTCCGAAAC	TGAA
IFNB1-Csp6I	chr9:21077396	TGAGCAGTCTGCACCTGAAAAGA	TGGGAGGATTCTGCATTACCTGA	AGGCCAAGGA
IFNB1-NlaIII	chr9:21077956	GAAGTGAAAGTGGGAAATTCCTCTGA	TTCCTCTGAATAGAGAGAGGACCATC	TCATATAAATAGGCCATACC
IFNB1-Triple	chr9:21077956	GAAGTGAAAGTGGGAAATTCCTCTGA	TTCCTCTGAATAGAGAGAGGACCATC	TCATATAAATAGGCCATACC
ECAD9[Non-risk/Risk]	chr9:22124746	CTATCTTGAAGGCAGGCCACACT	CTTGTGTAAACAATGGTATCACATTCTAACTT	AGCTGAGAC[G/C]ACTTCTGGCCCTGA

Supplementary Table 2

Viewpoint	Replication	Total Read	On Target Ratio	Unique Read	Intrachromosome Ratio	In-TAD Ratio
Boundary-DpnII	1	9,903,873	0.85	19,803	0.45	0.65
	2	12,050,430	0.87	22,113	0.45	0.65
Boundary-Csp6I	1	3,613,316	0.94	8,632	0.49	0.65
	2	2,899,486	0.92	5,704	0.50	0.69
Boundary-NlaIII	1	2,266,904	0.91	20,003	0.46	0.61
	2	2,067,982	0.91	26,365	0.49	0.60
Boundary-Triple_Ctrl	1	6,166,334	0.97	98,426	0.68	0.68
	2	8,287,678	0.95	101,384	0.69	0.68
Boundary-Triple_Induced	1	5,648,676	0.93	102,319	0.65	0.66
	2	6,113,449	0.92	108,429	0.65	0.66
Boundary-Triple_Alt_Dig	1	3,562,529	0.94	42,037	0.63	0.73
	2	3,675,999	0.94	31,493	0.60	0.76
MLLT3-DpnII	1	4,746,406	0.93	22,643	0.47	0.73
	2	6,462,079	0.94	24,034	0.46	0.72
MLLT3-Csp6I	1	5,109,338	0.95	9,467	0.49	0.74
	2	4,809,153	0.95	6,271	0.53	0.78
MLLT3-NlaIII	1	1,613,409	0.89	32,103	0.52	0.73
	2	1,522,203	0.88	33,497	0.51	0.74
MLLT3-Triple_Ctrl	1	11,108,271	0.96	75,018	0.56	0.74
	2	19,623,818	0.96	74,235	0.57	0.74
MLLT3-Triple_Induced	1	19,488,449	0.93	68,200	0.55	0.70
	2	17,435,504	0.93	78,689	0.54	0.68
MLLT3-Triple_Alt_Dig	1	4,455,496	0.93	52,089	0.54	0.77
	2	3,548,450	0.92	41,754	0.47	0.80
IFNB1-DpnII	1	8,082,998	0.55	4,983	0.39	0.75
	2	9,094,900	0.54	4,791	0.28	0.72
IFNB1-Csp6I	1	22,362,696	0.29	3,766	0.23	0.74
	2	17,115,256	0.25	5,523	0.44	0.67
IFNB1-NlaIII	1	21,198,926	0.17	5,885	0.50	0.79
	2	23,438,769	0.17	5,663	0.42	0.74
IFNB1-Triple_Ctrl	1	12,520,527	0.41	16,682	0.54	0.72
	2	10,331,398	0.39	17,689	0.59	0.73
IFNB1-Triple_Induced	1	9,631,644	0.35	27,829	0.68	0.69
	2	8,253,459	0.35	32,334	0.67	0.68
ECAD9	1	9,248,315	0.87	128,749	0.38	0.60
	2	8,074,667	0.82	89,873	0.41	0.66

Supplementary Table 3

Site	Target	Sequence	Position (hg19; chr9)
MLLT3 S1	gRNA Left 1	TATTTTTCAGAGTTGAAATT	20612755
	gRNA Right 1	AACTGGGACAATCTTTTTGG	20613816
	gRNA Left 2	ATAAAATGCTTAATCCCACC	20612715
	gRNA Right 2	ATGTTAAATTTCTAGAAGGG	20613878
	Validation Left	TCTCAAGTGTCCTTTCCAGCTCT	
	Validation Right	CCTTTCCTTCCTTCCTTCTTTCA	
MLLT3 S2	gRNA Left 1	CAAGTCTATACAGCAATGAC	20625749
	gRNA Right 1	AGGTGCTACTTATTAAATTA	20626696
	gRNA Left 2	ACACTCTAATTACACGTTAA	20625638
	gRNA Right 2	ATGCTTAGTGGAATAGTTCT	20626849
	Validation Left	TTGTGTACTTGACAATGTGGTTATAGAAA	
	Validation Right	TTGTTGAAGTTTGAGCTGTCCAA	
MLLT3 S3	gRNA Left 1	GAGGAGGATGACATAAGTGC	20629747
	gRNA Right 1	TACCCCTCACAGACATTTTA	20631067
	gRNA Left 2	TAAATTTTTTTATAGAGAGG	20629851
	gRNA Right 2	TAAATATGATGACAGTGTTC	20631240
	Validation Left	CAAAAAGACGAGGATAGGTCCAG	
	Validation Right	GGCCTTGATTGAAGGAAAGGTTA	
MLLT3 S4	gRNA Left 1	GCTCCCCAGTTTTGCCAAGT	20632843
	gRNA Right 1	AGATGTGGGTAATGTATGGT	20633966
	gRNA Left 2	CGGGTATAAGCAAGCCCACT	20632845
	gRNA Right 2	GAACAGATGTGGGTAATGTA	20633962
	Validation Left	GGGCATTTGTCATTACACAGGAT	
	Validation Right	TCCTGTACCTGTCTCAATGATGC	
GFP	Negative Ctrl	GAAGTTCGAGGGCGACACCC	