

Supplemental Information

The effect of cohesin mutations on HLA-class II expression in the myeloid leukemia of Down syndrome.

Austin C. Boucher, Wojciech Rosikiewicz, Anitria Cotton, Amber L. Broadhurst, Te Ling, Rishika Vadlamudi, Melanie Bremm, Anna-Lena Schmell, Lahiri Konada, Madeline Niederkorn, Josiah Murray, Yichao Li, Shilpa Narina, Sarah Inglott, Anindita Roy, Philip Ancliff, Ricky Tirtakusuma, Jing Ma, Michael P. Walsh, Jeffery M. Klco, Paul Thomas, Yiwen Wang, Aristotelis Tsirigos, Yong Cheng, Beisi Xu, Xiaotu Ma, Soheil Meshinchi, Elvin Wagenblast, Shondra M. Pruett-Miller, Sridhar Rao, Gang Wu, Yunusa Olufadi, Guolian Kang, Zuzana Tothova, Jeffrey W. Taub, Yubin Ge, Paresh Vyas, Jan-Henning Klusmann, John D. Crispino

Supplemental Methods

Cell culture

CMY cells were a gift from Dr. Charles Mullighan at St. Jude Children's Research Hospital. CMY and isogenic clones derived from CMY were cultured in RPMI (Gibco), 10% heat inactivated FBS, and 1x pen/strep. Cells were passaged every 2-3 days. For cell counts, cells were plated at a starting density of 100,000 cells/ml and counted via Countess II automated cell counting. Kasumi-1 cells were obtained from ATCC (Cat#CRL-2724) and maintained in RPMI (Gibco), 20% heat inactivated FBS, and 1x pen/strep. Cells were passaged every 4-5 days. Cell identity was confirmed by STR-profiling and confirmed to be free of mycoplasma.

Patient samples (expanded via xenotransplantation) were cultured in StemSpan (StemCell Technologies) supplemented with streptomycin/penicillin (Thermo Fisher Scientific) and a cytokine cocktail (FLT3-L, SCF, IL-6, IL-3, TPO; all purchased from PeproTech). Samples were taken from pediatric AML patients treated as part of the AML Berlin-Frankfurt-Münster study group. Written informed consent was given by all custodians of the participants.

Xenografts

0.5e6 CMY and corrected clones were transplanted via tail vein injection into sub lethally irradiated (200 rads) immunodeficient NSG (NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ) (The Jackson Laboratory, #05557) female mice 9 weeks of age. After transplantation, mice were treated with amoxicillin and baytril via drinking water for the duration of the study to reduce the risk of bacterial infection. No sample size estimation, randomization, or blinding were performed. Animal studies were approved by the St. Jude Children's Research Hospital Institutional Animal Care and Use Committee.

Western blotting

For whole cell extracts, 1e6 cells were lysed in 100ul of cell lysis buffer. Equal volumes of lysate were loaded onto 4-15% gradient gels (BioRad; Cat# 456-8084 and 456-8085). Gels were transferred onto nitrocellulose membranes and blocked and incubated with primary overnight in 5% milk in TBST. Imaging was performed by LiCor and chemiluminescence. Primary antibodies included the following: RAD21 (GeneTex; GTX106012), GATA1 (Abcam; Cat#Ab11852), CIITA (Santa Cruz Cat#sc-13556), and HSC70-AF790 (Santa Cruz; Cat#sc-7298)(LiCor). Secondary antibodies included the ECL anti-rabbit IgG HRP linked (Cytiva; Cat#NA934V).

CRISPR-Cas9 screen

Cas9 stable CMY and 3D11 cells were generated via lenti-Cas9 spinoculation at an MOI=2 and subsequent selection for 2 week in the presence of 5ug/mL blasticidin. Cas9+ cells were then spinoculated onto Retronectin coated plates containing the Brunello sgRNA library at an MOI=0.2-0.3 to 400x coverage and left overnight. The following day, cells were washed from virus 3x in PBS and replated in media. The following day, cells were selected for 72h in 1.25ug/mL puromycin. Following selection, cells were washed from puromycin 3x in PBS. Dead cells were removed by Ficoll gradient centrifugation. 30.5e6 cells were taken for sgRNA representation. 30.5e6 cells were placed into culture for an additional 7 days. On day 4, cells were counted and diluted 1:4 with fresh media. On day 7, 100e6 cells were washed from media 3x in PBS. Cells were subsequently stained with HLA-DR,DP,DQ antibody in PBS+0.5%FBS for 25 minutes on ice. Cells were washed, resuspended in PBS+0.5%FBS, and sorted for the top and bottom 15% of HLA-class II expression. DNA was collected from these sorted fractions for guide enrichment which was determined by MAGeCK analysis. The HLA-DR,DP,DQ screen was performed independently twice.

Stratification of RNA seq gene expression

RNA-seq normalized gene expression matrices of the different patient sample AML datasets were imported into Python (v3.12.3) using pandas (v2.2.2). The analyses were focused on the set of cohesin complex genes (STAG1, STAG2, PDS5A, PDS5B, RAD21, NIPBL, SMC1A, SMC3, WAPAL) and HLA-class II genes (HLA-DPA1, HLA-DPB1, HLA-DRA, HLA-DRB1, HLA-DRB3, HLA-DRB4, HLA-DRB5, HLA-DQA1, HLA-DQB1, HLA-DQA2, HLA-DQB2, HLA-DMA, HLA-DMB, HLA-DOA, HLA-DOB, CIITA and CD74). For each cohesin gene, the patient samples were stratified into quartiles based on normalized expression. Z-score normalization was applied across all genes for each sample using numpy (v1.26.4) and scipy (v1.13.1) to compare between genes with different expression levels. Median Z-scored expression for the defined set of HLA class II genes was calculated within each quartile stratified group, and statistical significance between low (Q1) and high (Q4) cohesin gene expression groups was assessed using a two-sided Mann–Whitney U test. Data visualization through boxplots of HLA-class II expression across quartiles was generated using matplotlib (v3.8.4) and seaborn (v0.13.2).

CUT&RUN

CUT&RUN was performed per the manufacturer's protocol using the Epicypher CUTANA kit (Cat#14-1048) with the following antibodies: H3K27Ac (Active motif; Cat#39133), H3K4me1 (Active motif; Cat#39297), RAD21 (GeneTex; Cat#GTX106012), CTCF (Millipore Sigma; CAT#07-729), GATA1; (Cat#Ab11852). The MNase reaction time was modified from 2 hours to 1.5 hours. Libraries were prepared using the CUTANA CUT&RUN Library Prep Kit (Cat#14-1001) according to manufacturer protocol. Samples were sequenced using an Illumina NovaSeq.

ChIP-seq

ChIP-seq was performed as previously described(1). Briefly, 10 million cells were crosslinked with 1% formaldehyde for 10 minutes at 37 °C, followed by two washes with ice-cold PBS. Cells were collected and resuspended in 300 µL SDS lysis buffer. Chromatin was sheared to 200–500 bp fragments. After sonication, samples were centrifuged at 13,000 RPM for 10 minutes at 4 °C to pellet debris. Supernatants were diluted to a final volume of 2 mL in ChIP dilution buffer. 1% input was set aside for normalization, and the remaining samples were incubated overnight at 4 °C with anti-CIITA antibody (Rockland; Cat#100-401-249,1:2000) or anti-IgG antibody (CST, 2729S, 1:500). The following day, samples were incubated for 3 hours at 4 °C with protein A/G magnetic beads (Thermo; Cat#88802). Antibody-conjugated magnetic beads were sequentially washed for 3 minutes per wash at room temperature with low salt buffer, high salt buffer, and LiCl buffer. Finally, beads were washed twice with 1x TE buffer. Chromatin was eluted from beads in two rounds using ChIP elution buffer incubated for 15 minutes at 65 °C with shaking at 1000 rpm. Eluates were combined. DNA–protein crosslinks (from both ChIP and input samples) were reversed by adding 20 mM NaCl and incubating for 4 hours at 37 °C. Residual proteins were digested with Proteinase K. DNA was purified via phenol/chloroform extraction. DNA concentration was determined using Qubit dsDNA High Sensitivity assay. For sequencing, libraries were prepared using a KAPA Biosciences paired-end library prep kit compatible with Illumina platforms. Library fragment size (~350–450 bp) was verified on a Tapestation (Agilent). Replicates were normalized to equal DNA concentration, pooled, and sequenced on an Illumina NovaSeq.

RNA sequencing

RNA was extracted from 1e6 cells per group with the Qiagen RNeasy Kit (Cat#74104). Quality control, library preparation, sequencing, and read alignment were performed as previously described(2).

ATAC sequencing

50,000 cells were washed, nuclei extracted, and tagmented following the Omni-ATAC protocol(3). Quality control, library prep, and sequencing were performed by the Hartwell Center at St. Jude. Samples were sequenced on an Illumina NovaSeq.

Hi-C

Hi-C was performed according to Rao et al. 2014(4). Briefly, 2e7 cells were crosslinked at 1% formaldehyde and quenched at 0.2M glycine. Samples were then lysed and subsequently digested overnight with 100U Mbol (NEB,R0147) at 37 degrees in a thermomixer set at 900rpm. Restriction fragments were then filled and marked with biotinylated dATP (Life Technologies, Cat#19524-016). Samples were then ligated, and crosslinks were reversed. DNA was then sheared and size selected to 300-500bp. Biotin pulldown was then performed, and samples were prepared for Illumina sequencing. Samples were sequenced using a NovaSeq.

ChIP-seq data processing

Raw reads were trimmed with Trim Galore (v0.4.4; cutadapt with Q20)(5). Reads were aligned to hg38/GRCh38.p12 using BWA (v0.7.17-r1198)(6), converted to BAM and filtered with samtools (v1.2)(7). Duplicate reads were marked using biobambam2 bamsormadup (v2.0.87). Uniquely mapped, non-duplicate reads were retained with samtools (MAPQ ≥ 1). For QC and fragment size estimation, deduplicated reads were analyzed by strand cross-correlation using SPP (v1.11)(8). Next, fragments were extracted to BED format and extended with bedtools (v2.17.0)(9), followed by calculating genome-wide coverage using bedtools genomecov in bedGraph mode. Samples were scaled to 15 million sequencing depth and converted to bigWig format using UCSC bedGraphToBigWig (v4)(10). Peaks were called with MACS2 (v2.1.1.20160309)(11) using the inferred fragment size with settings “--nomodel --extsize <fragment_size>” and matched IgG samples as background control. High-confidence peaks were defined at $q \leq 0.05$, and a

companion low-confidence set with $q \leq 0.5$. Peaks overlapping ENCODE blacklist regions were removed by subtracting blacklist intervals from each peak set using bedtools(12).

CUT&RUN data processing

Raw reads were trimmed with Trim Galore (v0.4.4) (cutadapt Q20)(5). Reads were aligned to a human-Drosophila hybrid reference (GRCh38.p12 + dm6) using BWA (v0.7.12-r1039)(6), converted to BAM with samtools (v1.2)(7), and duplicate reads were marked with biobambam2 (v2.0.87). Spike-in normalization factors were derived from the fraction of reads mapping to dm6 relative to all reads mapped to the hybrid reference. For downstream processing only the portion of the reads mapping to human reference was used, alignments were filtered to uniquely mapped reads ($\text{MAPQ} \geq 1$) and low-quality/secondary/supplementary alignments were excluded. BAM files were name-sorted with samtools and converted to paired-end fragment representations (BEDPE) using bedtools (v2.17.0)(9), retaining properly paired fragments. For visualization, a single genome-wide coverage track was generated from all fragments shorter than 2 kb by extracting the central 80 bp of each fragment (midpoint ± 40 bp), computing genome coverage using bedtools genomecov, and converting bedGraph to bigWig using UCSC bedGraphToBigWig(10). Coverage was first normalized to a fixed depth of 10 million fragments per sample (based on the number of <2 kb fragments), and then additionally scaled by the spike-in factor using $1/(\text{spike-in percentage})$. The resulting bigWig files were used for visualization and are provided as supplementary data. Whenever applicable, narrow peaks were called using MACS2, using matched IgG samples as controls. High-confidence peaks were defined at $q \leq 0.05$, and a companion, low-confidence peak set was called with $q \leq 0.5$, while broad peaks were called with SICER (v1.1; p-value $1e-5$)(13). Peaks overlapping ENCODE blacklist regions were removed using bedtools(12).

ATAC-seq data processing

Reads were processed with Trim Galore (v0.4.4) and cutadapt (Q20)(5); to mitigate GC bias, the first 15 bp of each read were clipped. Mapping and duplicate marking followed the GATK workflow implemented by NVIDIA Parabricks fq2bam (v4.3.0-1), which uses BWA-mem (v0.7.17)(6) to align reads to hg38/GRCh38.p12, and GATK MarkDuplicates(14) to mark PCR duplicates. Properly paired, uniquely mapped reads (MAPQ ≥ 1) were retained using samtools (v1.2)(7), and mitochondrial reads were excluded from downstream analyses. Paired-end reads were converted to BEDPE format using bedtools (v2.24.0)(9), enabling fragment-length-based classification. Nucleosome-free fragments (<100 bp insert size, Tn5-adjusted) were isolated for downstream signal generation and peak calling, while mono- and multi-nucleosome fragments were retained for QC checks. Fragment length distributions and mapping statistics were computed to assess library quality. Genome-wide coverage tracks were generated from nucleosome-free fragments by computing the midpoint of each fragment (± 40 bp window), followed by genome coverage calculation with bedtools genomecov. Coverage was normalized using a library-size scaling factor corresponding with 20 million nucleosome-free fragments per sample, and converted to bigWig format using UCSC bedGraphToBigWig(10). Peak calling was performed on nucleosome-free fragments using MACS2 (v2.1.1.20160309)(11) with the parameters “--nomodel --keep-dup all” and genome-specific effective genome sizes. High-confidence peaks were identified at $q \leq 0.05$, with an additional low-confidence peak set generated at $q \leq 0.5$. Peaks overlapping ENCODE blacklist regions were removed using bedtools(12).

RNA-seq data processing

Reads were trimmed with Trim Galore (v0.4.4) (cutadapt Q20)(5), aligned to hg38/GRCh38.p12 with STAR (v2.7.9a)(15), PCR duplicates were marked using GATK MarkDuplicates (v4.1.2.0)(14), followed by per gene read count quantification with RSEM and using GENCODE Release 31 as the reference gene annotation(16, 17). Only protein-coding genes with GENCODE annotation confidence levels 1 or 2 were retained for analysis. Lowly expressed genes were

filtered using an adaptive, threshold dependent on library-size: the median library size across samples was used to define a minimum expression cutoff equivalent to approximately one read per 10 million mapped reads (with a minimum of 1 read for libraries <10 million reads and a maximum of 10 reads for libraries >100 million reads). Differential expression analysis was performed using edgeR(18) and limma-voom approach(19, 20). Differentially expressed genes were defined as those with absolute $\log_2$ fold-change value > 1, and false discovery rate (FDR) < 0.05.

Differentially accessible and differentially binding regions identification

Following the previously described approach(21), reproducible peaks across biological replicates were identified for each target's condition. Peaks were considered reproducible if (i) overlapping high-confidence peaks were present in all biological replicates, or (ii) a high-confidence peak in one replicate was supported by either a high- or low-confidence peak in all remaining replicates. Differential accessibility or binding analysis was performed on an union of reproducible peaks from all conditions of the target (e.g. ATAC-seq, GATA1s etc.). Fragment counts per sample were calculated using bedtools coverage (v2.24.0)(9) based on fragment intervals corresponding to the analyzed signal (i.e. ATAC-seq / ChIP-seq / Cut-and-Run). Raw count matrices were imported into edgeR to construct a DGEList object(18). For samples without spike-in controls, library size normalization was performed using the trimmed mean of M-values (TMM) method via calcNormFactors function. For samples containing spike-in material, normalization factors were instead derived from spike-in read counts using an alternative, geometric-mean-centered strategy, and these externally computed scaling factors were supplied as edgeR normalization factors for downstream modelling(18). Low-information regions were filtered using custom CPM/FPKM-based procedure with data-driven cutoff, after which differential analysis was performed using limma-voom approach (voom transformation followed by lmFit and eBayes)(19, 20). In parallel, reproducible peaks were annotated to genes and genomic features. Peaks were

assigned to genes if they overlapped a promoter region, defined as transcription start site (TSS) ± 2 kb, using GENCODE reference annotations (v31)(17). In addition, each peak was annotated with a single genomic context using a prioritized, mutually exclusive hierarchy: (1) promoter upstream (≤ 2 kb upstream of TSS), (2) promoter downstream (≤ 2 kb downstream of TSS), (3) exon, (4) intron, (5) transcription end site (TES ± 2 kb), (6) distal upstream (up to 50 kb upstream of TSS, excluding promoter regions), (7) distal downstream (up to 50 kb downstream of TES), and (8) intergenic regions. Peaks overlapping multiple annotation categories were assigned to the highest-priority category. For downstream gene-level analyses, ranked gene lists were generated from the differential peak results by collapsing reproducible peaks to genes based on promoter overlap. For genes associated with multiple peaks, the statistical values from the peak with the strongest absolute ranking statistic, were assigned as metric to the gene. The π -value, which corresponds with $\log_2(\text{fold-change}) \times -\log_{10}(\text{p-value})$, was used as the rank metric(22).

Gene set enrichment analysis (GSEA)

Gene set enrichment analysis was performed using GSEAPy (v1.1.1; <https://gseapy.readthedocs.io/en/latest/>) in preranked mode(23). Genes were ranked using π -value rank(22), based on $\log_2(\text{fold-change}) \times -\log_{10}(\text{p-value})$ values, derived from differential genes or regions analyses, with genes ordered by the absolute value of the ranking metric and duplicate gene symbols collapsed by retaining the highest-ranking entry. As reference gene signatures, the MSigDB (24-26) H, C2, C3, C4, C5, C6 and C8 gene set collections (v2023.2.Hs), supplemented by the custom gene set collections, were used. Each collection was analyzed separately as per MSigDB authors recommendation. Prior to analysis, gene identifiers in gene rank files, and in-house gene set collections, were harmonized using MSigDB-provided chip files to ensure consistent gene symbol namespaces. GSEA was conducted using a classic enrichment score, with gene set sizes restricted to a minimum of 15 and a maximum of 500 genes and using

1000 permutations for significance calculations. For visualization and interpretation of enrichment results, gene set enrichment networks were constructed with nodes in the network representing enriched gene sets, and edges representing pairwise overlap between gene sets based on shared leading-edge genes. Network layouts were generated using a force-directed algorithm to group highly overlapping gene sets. Node size was scaled according to normalized enrichment scores.

Hi-C data analysis

For Hi-C, paired-end reads of 100bp reads were processed with Juicer (v1.5, default parameters)(27), based on hg38 reference (without ALT chromosomes) and Mbol ligation sites GATCGATC. Four replicates were performed for each cell type and processed independently, replicates are merged after visual confirmation of reproducibility and achieved 680M, 730M, 830M read pairs respectively. Juicebox have been used for visualization (28, 29).

Supplemental Methods References

1. Delint-Ramirez I, Konada L, Heady L, Rueda R, Jacome ASV, Marlin E, et al. Calcineurin dephosphorylates topoisomerase II β and regulates the formation of neuronal-activity-induced DNA breaks. *Mol Cell*. 2022;82(20):3794-809 e8.
2. Melo-Cardenas J, Bezavada L, Crawford JC, Gurbuxani S, Cotton A, Kang G, et al. IL-13/IL-4 signaling contributes to fibrotic progression of the myeloproliferative neoplasms. *Blood*. 2022;140(26):2805-17.
3. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods*. 2017;14(10):959-62.
4. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014;159(7):1665-80.
5. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011. 2011;17(1):3.
6. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-60.
7. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-9.
8. Kharchenko PV, Tolstorukov MY, Park PJ. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol*. 2008;26(12):1351-9.
9. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841-2.
10. Kuhn RM, Haussler D, Kent WJ. The UCSC genome browser and associated tools. *Brief Bioinform*. 2013;14(2):144-61.
11. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9(9):R137.
12. Dale RK, Pedersen BS, Quinlan AR. Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. *Bioinformatics*. 2011;27(24):3423-4.
13. Zang C, Schones DE, Zeng C, Cui K, Zhao K, Peng W. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics*. 2009;25(15):1952-8.
14. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297-303.
15. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21.
16. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011;12:323.
17. Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res*. 2019;47(D1):D766-D73.
18. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139-40.
19. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*. 2014;15(2):R29.
20. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.

21. Wang X, Rosikiewicz W, Sedkov Y, Martinez T, Hansen BS, Schreiner P, et al. PROSER1 mediates TET2 O-GlcNAcylation to regulate DNA demethylation on UTX-dependent enhancers and CpG islands. *Life Sci Alliance*. 2022;5(1).
22. Xiao Y, Hsiao TH, Suresh U, Chen HI, Wu X, Wolf SE, et al. A novel significance score for gene selection and ranking. *Bioinformatics*. 2014;30(6):801-7.
23. Fang Z, Liu X, Peltz G. GSEAPy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics*. 2023;39(1).
24. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27(12):1739-40.
25. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545-50.
26. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1(6):417-25.
27. Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, et al. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst*. 2016;3(1):95-8.
28. Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, et al. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst*. 2016;3(1):99-101.
29. Zhou X, Wang J, Patel J, Valentine M, Shao Y, Newman S, et al. Exploration of Coding and Non-coding Variants in Cancer Using GenomePaint. *Cancer Cell*. 2021;39(1):83-95 e4.

Supplemental Figure Legends

Supplemental Figure 1: Differential chromatin accessibility and occupancy in *RAD21*-corrected cells. A) Volcano plots of differential CUT&RUN and ATAC regions, data presented as *RAD21* corrected clones vs. CMY. B) Annotated genomic locations of the differential regions in A. Promoter.up indicates upstream and promoter.down indicates downstream.

Supplemental Figure 2: HLA class II genes are regulated by cohesin in ML-DS. A) Volcano plot of differentially expressed genes presented as *RAD21*-corrected clones vs. CMY. B,C,D) Word clouds of leading-edge gene frequency called in respective GSEA clusters. E) Next generation sequencing detecting R305* *STAG2* mutation in CMK and CMK *STAG2* R305* clones. F,G) Representative plots and quantification of HLA-DR,DP,DQ flow cytometry of CMK and CMK R305* *STAG2* mutant isogenic clones. Statistics: G: Student's t-test.

Supplemental Figure 3: Genomic architecture and *CIITA* expression are not impacted by *RAD21* correction in CMY cells. A) Differential TAD analysis from Hi-C data. Red dots represent TADs situated in the HLA locus. B) Insulation scores across the HLA class II locus generated from Hi-C. C) AAVS1 safe harbor editing within the PPP1R12C gene. Editing percent of cell pool determined by next generation sequencing. D) Log2FC of *CIITA* from RNA seq from this study and Wagenblast *et al.* E) Representative western blot of *CIITA* protein in CMY and *RAD21*-corrected clones developed via chemiluminescence. F) Densitometry of biological replicates of E. Statistics: C: Paired t-test. F: Student's t-test.

Supplemental Figure 4: Analysis of second ML-DS progression cohort. Second cohort of HLA-DR expression in TAM patients (n=28) with known ML-DS progression (3/28).

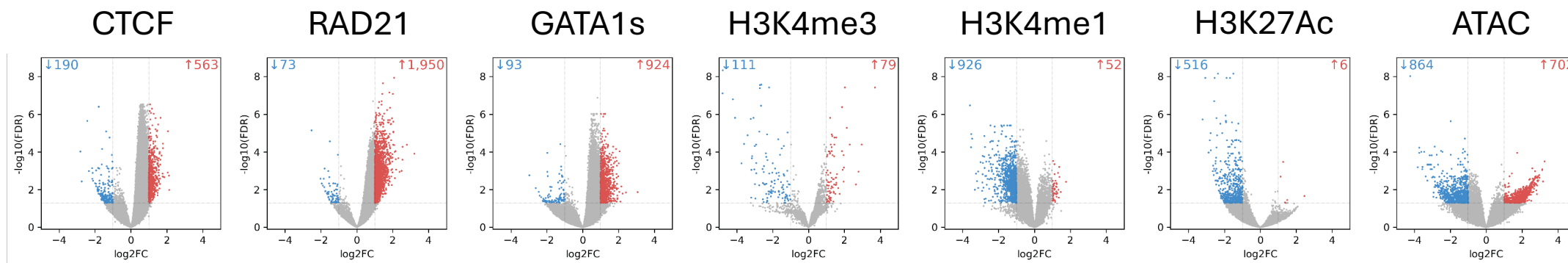
Supplemental Figure 5: CRISPR-Cas9 HLA-DR,DP,DQ screen validation. A) Table showing number of candidate HLA-DR,DP,DQ regulators filtered by guide representation 3 or more, log fold change +/-0.5 or beyond, and p-value cutoff of either 0.01 or 0.05. B) Representative normalized guide count for CIITA. C) Representative normalized guide count for FBXO11. D) Representative normalized guide count for MEAF6. E) Editing efficiency of cell pools for two independent guides for MEAF6, KAT7, KAT6a, or KAT6b in CMY (black/white), 2B10 (blue), and 3D11 (orange). F) HLA-DR,DP,DQ mean fluorescent intensity of cell pools for two independent guides for MEAF6, KAT7, KAT6a, or KAT6b from E in CMY (black/white), 2B10 (blue), and 3D11 (orange) 5 days post RNP introduction. Control, non-targeting guide. Statistics: E,F: Exact paired Wilcoxon signed-rank test or paired Student's t-test.

Supplemental Table 1: Oligonucleotide sequence information for CRISPR-Cas9 editing.

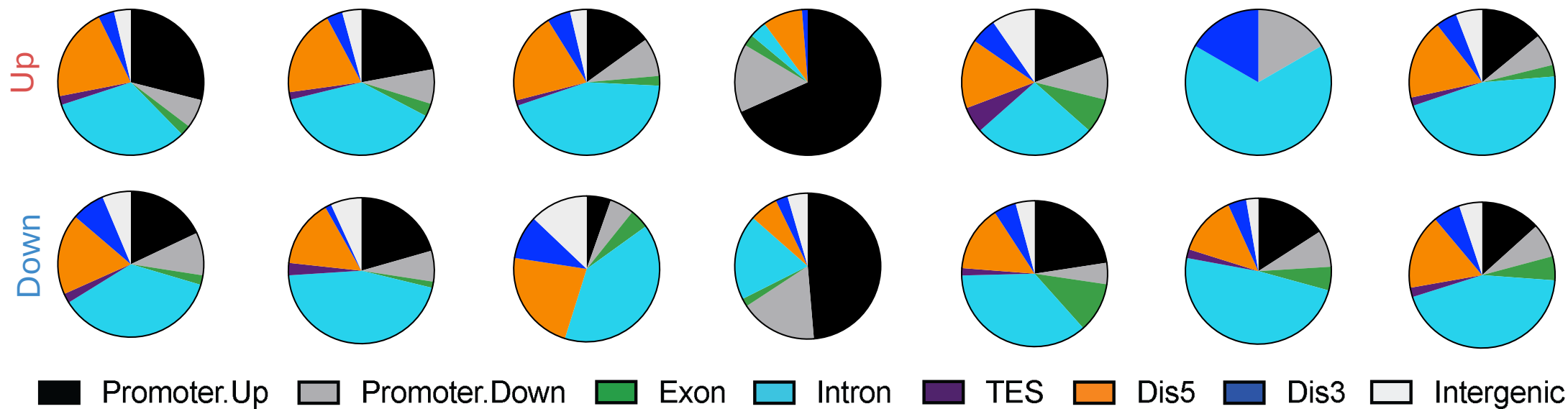
Supplemental Table 2: List of overlapping gene sets. GSEA was performed on RNA seq data contrasted as 2B10 vs. CMY, 3D11 vs. CMY, and T21+GATA1s vs. T21+GATA1s+STAG2 knockout (WT vs cohesin mutant for each contrast). These three datasets were then analyzed for which gene sets were common between them. These common gene sets were subsequently clustered based on the leading-edge genes in each gene set.

Supplemental Table 3: HLA-DR,DP,DQ expression CRISPR-Cas9 Brunello screen replicates.

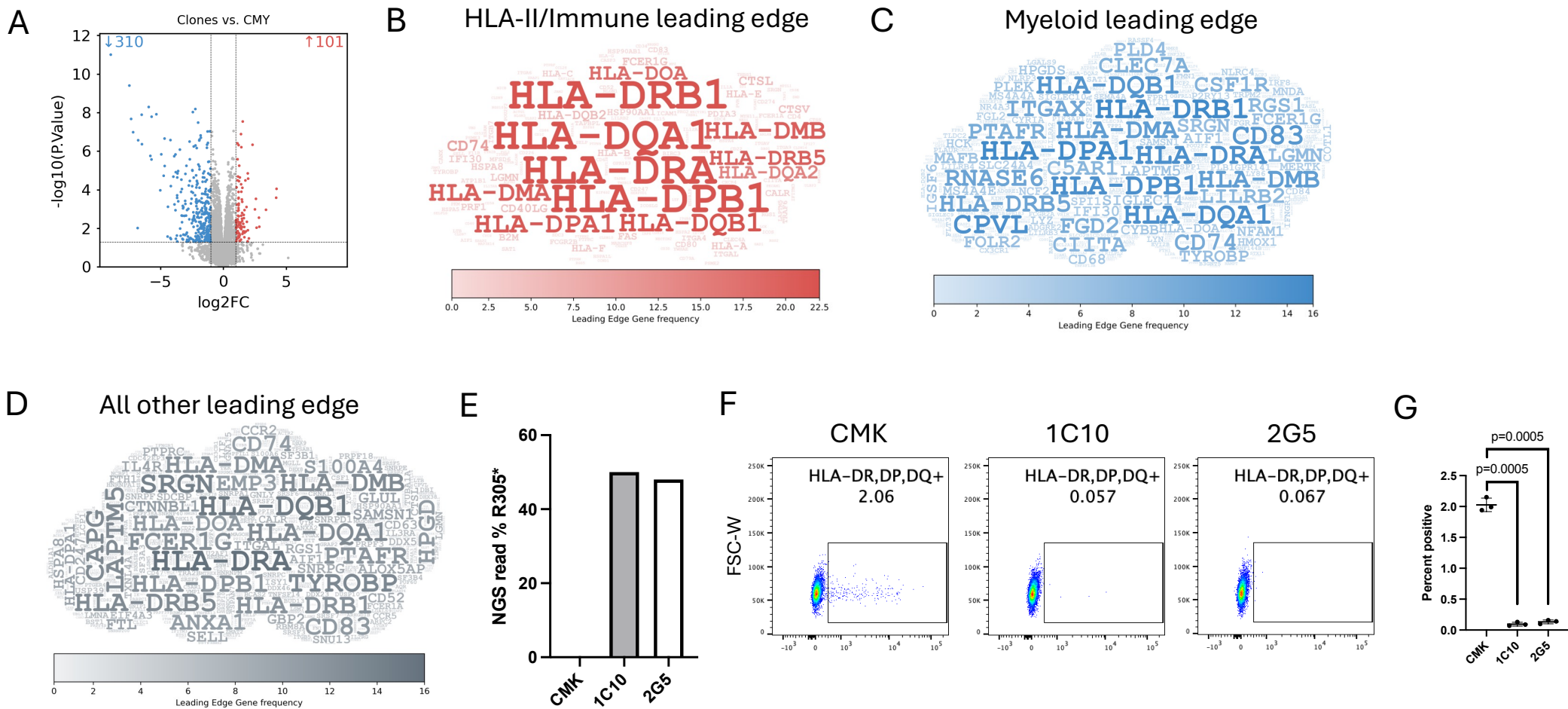
A

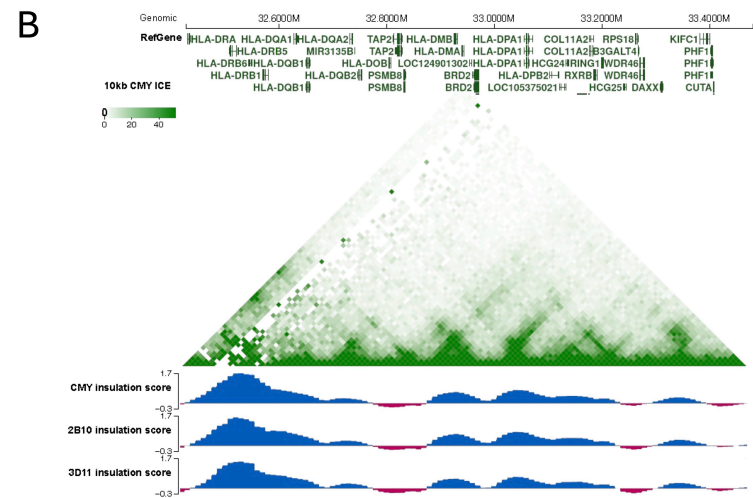
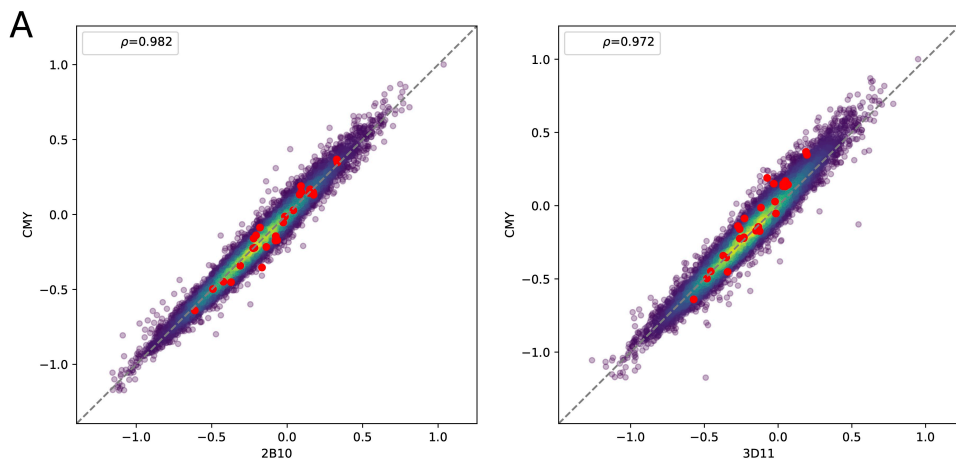


B



Supplemental Figure 1

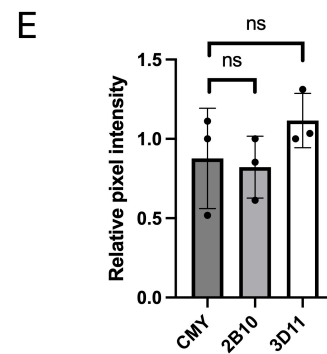
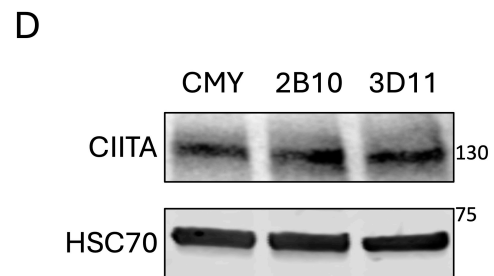




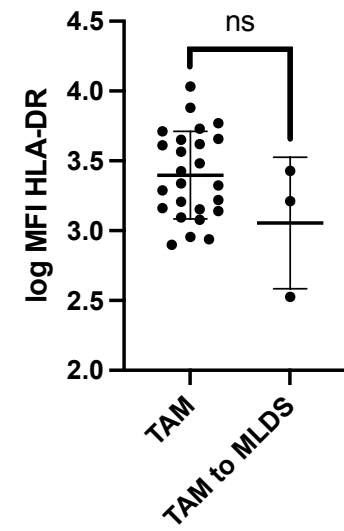
C

CIITA RNA

Sample	log2FC	FDR
2B10 v. CMY	-0.17	0.27
3D11 v. CMY	-0.35	0.35
Wagenblast et. al	0.68	0.29



Supplemental Figure 3



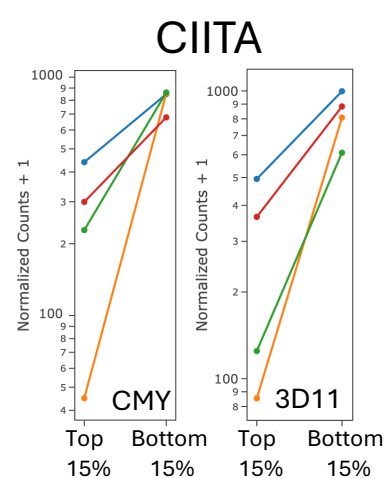
Supplemental Figure 4

A

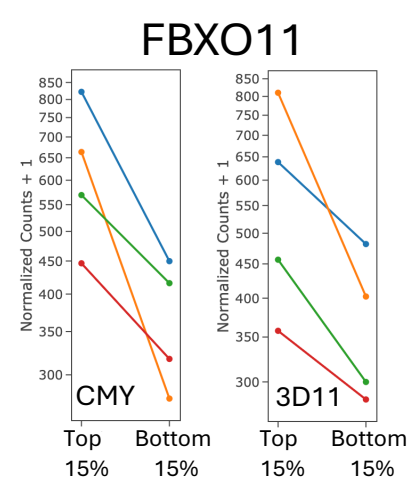
Guide rep >3, lfc>0.5

	p-val	CMY	Common	3D11
Enriched in				
	Top 15%			
	P<0.01	7	1	3
	P<0.05	27	7	15
Bottom 15%				
	P<0.01	5	5	5
	P<0.05	26	6	22

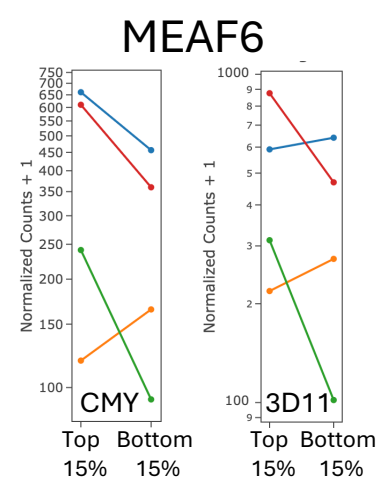
B



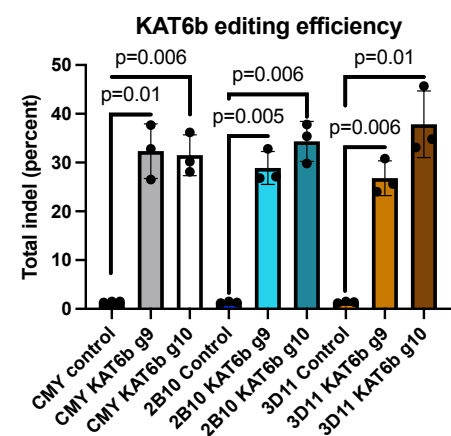
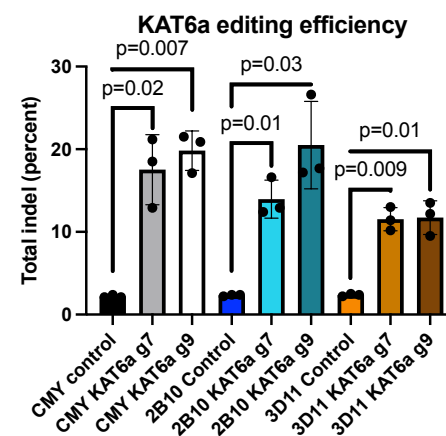
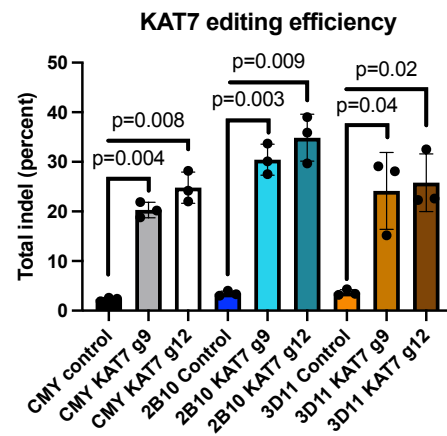
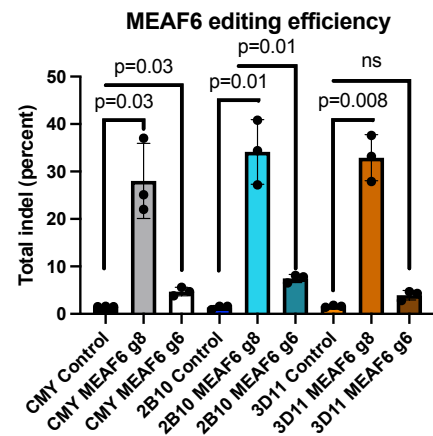
C



D



E



F

