# Loop Catalog: a comprehensive HiChIP database of human and mouse samples

**Joaquin Reyna[1,2*], Kyra Fetter[1,3*], Romeo Ignacio[1], Cemil Can Ali Marandi[1,2], Nikhil Rao[1,4], Zichen Jiang[1,5], Daniela Salgado Figueroa[1,2], Sourya Bhattacharyya[1], Ferhat Ay[1,2,6]**

*** These authors contributed equally**

**Corresponding author: ferhatay@lji.org**

**[1] Centers for Cancer Immunotherapy and Autoimmunity, La Jolla Institute for Immunology, La Jolla, CA 92037 USA.**

**[2] Bioinformatics and Systems Biology Graduate Program University of California, San Diego, La Jolla, CA 92093 USA.**

**[3] Department of Bioengineering, University of California San Diego, La Jolla, CA 92093 USA.**

**[4] Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA 92093 USA.**

**[5] Department of Mathematics, University of California San Diego, La Jolla, CA 92093 USA.**

**[6] Department of Pediatrics, University of California San Diego, La Jolla, CA 92093 USA.**

## ABSTRACT

HiChIP enables cost-effective and high-resolution profiling of regulatory and structural loops. To leverage the increasing number of publicly available HiChIP datasets from diverse cell lines and primary cells, we developed the Loop Catalog (https://loopcatalog.lji.org), a web-based database featuring HiChIP loop calls for 1319 samples across 133 studies and 44 high-resolution Hi-C loop calls. We demonstrate its utility in interpreting fine-mapped GWAS variants (SNP-to-gene linking), in identifying enriched sequence motifs and motif pairs at loop anchors, and in network-level analysis of loops connecting regulatory elements (community detection). Our comprehensive catalog, spanning over 4M unique 5kb loops, along with the accompanying analysis modalities constitutes an important resource for studies in gene regulation and genome organization.

**Keywords:** HiChIP; chromatin loops; database; GWAS; SNP-to-gene linking

## BACKGROUND

Changes in chromatin folding caused by mutations and variants can impact cell-type-specific function and disease risk via altered 3D interactions between genetic loci (Chakraborty & Ay, 2019; S. Wang et al., 2023). An important feature of chromatin folding is the existence of chromatin loops that connect regions separated by large genomic distances, often up to a megabase but with notable cases spanning even larger distances (V. Chandra et al., 2021; Friman et al., 2023; Levo et al., 2022; Long et al., 2020). These loops can be broadly categorized into: *i)* structural loops demarcating domains of interactions such as topological domains (TADs) and marked by the binding of CTCF and cohesin at their anchors; *ii)* regulatory loops which join distal gene regulatory elements such as enhancers and promoters to modulate gene expression (Boltsis et al., 2021; Brandão et al., 2021; L.-F. Chen & Long, 2023; Hu et al., 2023; Ibrahim & Mundlos, 2020; Ito et al., 2022; Popay & Dixon, 2022; Portillo-Ledesma et al., 2023; Qiu et al., 2023; Schaeffer & Nollmann, 2023; Xu et al., 2020). Our understanding of the exact role these loops/interactions play in cell-type specific gene regulation and ultimately disease susceptibility is far from complete.

To capture such interactions, the Hi-C procedure was developed as a high-throughput genome-wide assay that carries proximity ligation either in dilution (Lieberman-Aiden et al., 2009) or in-situ within intact nuclei (Rao et al., 2014). As a subset of interactions, regulatory loops are often associated with regions enriched for histone modifications such as H3K27ac and H3K4me3, transcription factors or for accessible chromatin (Davies et al., 2016; Fang et al., 2016; Fullwood et al., 2009; Liu et al., 2022; Mumbach et al., 2016; Wei et al., 2022). The development of the HiChIP assay represents an extension of the in-situ Hi-C methodology that can facilitate capture of these protein-specific regulatory loops by using immunoprecipitation to enrich for interactions involving anchors associated with the binding of transcription factors (TFs) or histone modifications of interest. Given the ability to enrich signal for a subset of the genome (e.g., active regulatory elements), HiChIP enables high-resolution profiling of chromatin interactions of interest with lower sequencing depth compared to Hi-C (Fang et al., 2016; Mumbach et al., 2016; Yu et al., 2021). Another advantage of HiChIP, at the time of its initial development, was that it allowed working with lower numbers of cells as input (1-5M), hence enabling the characterization of 3D organization in primary cells. Since the introduction of HiChIP in 2016, the number of publicly available human and mouse HiChIP studies published has consistently increased each year (**Figure 1A**). This indicates the growing popularity of the HiChIP assay, particularly as a method of investigating distal interactions between genetic variants, often in non-coding regions of the

genome, and their potential target genes in a cell-type and context-specific manner (V. Chandra et al., 2021; Duan et al., 2021; Giambartolomei et al., 2021; Mumbach et al., 2017; O'Mara et al., 2019; Shi et al., 2021; W. Wang et al., 2021).

In the context of human disease, HiChIP and similar assays provide a 3D view for the annotation of disease associations of non-coding genetic variants identified from GWAS (Fu et al., 2018; Gazal et al., 2022; Giambartolomei et al., 2021; Orozco et al., 2022; Zhong et al., 2022). Combined with efforts from multiple large consortia for cataloging putative regulatory elements spanning distinct cell types (e.g., ENCODE, BLUEPRINT and Roadmap Epigenomics), mapping such 3D maps of chromatin organization has become a critical piece of the puzzle, which led to formation of the 4D Nucleome Consortium (Dekker et al., 2017; Reiff et al., 2022). In parallel, other large-scale efforts identified SNPs associated with gene expression (i.e., expression quantitative trait loci or eQTLs) for different tissues and primary cell types (Bernstein et al., 2010; ENCODE Project Consortium et al., 2022; GTEx Consortium, 2020; Martens & Stunnenberg, 2013; Ota et al., 2021; Schmiedel et al., 2021). Although multiple studies incorporated GWAS, eQTL and high-resolution chromatin looping across many different matched/related cell types, it still remains a challenge to find a catalog of uniformly processed loops identified from datasets generated outside the ENCODE and 4DN consortia. Recent publications and databases, namely HiChIPdb (Zeng et al., 2023), ChromLoops (Zhou et al., 2023), and Cohesin-DB (J. Wang & Nakato, 2023), were first attempts to catalog loops by curating them from the broader literature. We provide a detailed comparison of Loop Catalog with these previous databases on the basis of the number of samples processed, the choice of tools for data processing including loop calling, integrated visualization and additional data types compiled (**Table 1**). In addition to having the largest number of HiChIP samples, unique features of Loop Catalog include utilization of matched ChIP-seq data when available, enabling a broader set of data download and visualization capabilities and additional modalities including SGL mapping, traditional and pairwise motif enrichment analysis and regulatory network analysis (**Table 1**).

Overall, the Loop Catalog developed here is a database containing the largest set of uniformly processed HiChIP data to date, curated from over 130 publications and a total of 1319 samples (from 2016 to January 2024) leading to over 5 million unique looping interactions at 5 kb resolution (10M total across all resolutions with 7.7M for human, 3.1M for mouse) with an accompanying web server that enables visualization, querying and bulk download of looping data (**Figure 1B, Supp Figure 1**). In addition to allowing easy access to this comprehensive catalog of chromatin looping data, we also demonstrated three different applications of how this data can be utilized.

In the first application, we intersected fine-mapped GWAS SNPs from CAUSALdb for four autoimmune diseases (J. Wang et al., 2020) with HiChIP loops from various immune cell types to identify potential target genes for these disease associated SNPs in each cell type (SNP to gene linking with loops or SGLs). Across all diseases we located 3048 unique SNPs, 1486 genes, 3411 loops and 13672 SGLs that span a median genomic distance of ~140kb. One example was a fine-mapped Type 1 Diabetes (T1D) GWAS SNPs which was linked to multiple genes through HiChIP loops in multiple lymphoid cells but not in monocytes. As the second application, we investigated TF motifs in loop anchors using motif enrichment analysis (Bailey et al., 2015) and pairs of TF motifs at loop anchors through pairwise motif enrichment using a bootstrapping approach. Our analysis of regulatory loop anchors conserved across a large majority of samples identified hundreds of enriched motifs including known and novel zinc finger transcription factors some of which were also enriched for significantly overrepresented combinations of motif pairs across samples (e.g., ZNF460, ZNF135). For the last application, we constructed enhancer-promoter networks using H3K27ac HiChIP loops as the underlying connections between these regulatory elements. Analysis of this network via community detection and community ranking algorithms resulted in the identification of tightly connected network communities, which may act as modular units of gene regulation. The large set of loops cataloged in this work will enable not only 3D-informed prioritization of genetic variants and enhancers involved in gene regulation but also will stimulate development of machine learning, deep learning, and network construction approaches that require large scale data.

## CONSTRUCTION AND CONTENT

### *Curating HiChIP and ChIP-seq Samples*

To identify a comprehensive list of publicly-released HiChIP datasets, we developed a pipeline that scans NCBI's Gene Expression Omnibus (GEO) database (Barrett et al., 2013) for studies performing HiChIP experiments. To extract information on these studies the BioPython.Entrez (Buchmann & Holmes, 2019) and metapub.convert (https://pypi.org/project/metapub/) packages were used. Raw sequencing data associated to these studies was then identified from the SRA database using the *pysradb* Python package (https://github.com/saketkc/pysradb) and the results were manually examined to extract HiChIP samples. ChIP-seq samples corresponding to these studies were also extracted if there was a record of them within the same GEO ID as the HiChIP sample.

### *Populating Sample Metadata*

To automatically add metadata, such as organ and cell type, to the curated HiChIP data, the BioPython.Entrez package was used to perform a search query within the GEO database. GSM IDs from the previous query were used with an esummary query to the biosample database (Buchmann & Holmes, 2019). From these results organism, biomaterial, celltype, GSM ID, SRA ID, disease, organ, treatment, tissue, and strain (only for mouse) were extracted. To improve classes for organ and biomaterial, a dictionary of classes as keys and synonyms as value (e.g. heart is a key and its synonyms are cardiovascular, atrium, aorta, etc) as used to search across GEO and Biosample reports. “N/A " was used to indicate fields that are not applicable for a given sample and for the remaining fields, “Undetermined” was assigned when the information could not be retrieved.

### *Downloading Raw Sequencing Datasets*

HiChIP FASTQ files were systematically downloaded using one of three methods and prioritized in the following order: SRA-Toolkit’s prefetch and fasterq-dump (2.11.2) (https://hpc.nih.gov/apps/sratoolkit.html), grabseqs (0.7.0) (Taylor et al., 2020), or EBI URLs generated from SRA Explorer (1.0) (Ewels, 2024). In addition, FASTQ files for phs001703v3p1 and phs001703v4p1 were downloaded that pertain to two previously published HiChIP studies from the Database of Genotypes and Phenotypes (dbGaP) (Schmiedel et al., 2021; V. Chandra et al., 2021; Sayers et al., 2022) (**Supp Table 1**). Additionally, ChIP-seq FASTQ files were downloaded from GEO using grabseqs (0.7.0).

### *Processing Reads from HiChIP Data*

Human samples were aligned to hg38 while mouse samples were aligned to mm10 using the HiC-Pro (3.1.0) pipeline (Servant et al., 2015) (**Supp Table 2**). Reference genomes were downloaded for hg38 (https://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/), and for mm10 (https://hgdownload.soe.ucsc.edu/goldenPath/mm10/bigZips/) and indexed with bowtie2 (2.4.5) using default parameters. All technical replicates were grouped into their respective biological replicate before read alignment. Restriction enzymes and ligation sequences were determined during the HiChIP literature search (**Supp Table 3**) and restriction fragments were generated with the HiC-Pro *digestion.py* tool using default parameters. HiC-Pro was then run in parallel mode by splitting FASTQ files into chunks of 50,000,000 reads using HiC-Pro’s *split_reads.py* utility with a minimum MAPQ threshold of 30. HiC-Pro generated contact maps at

the 1kb, 2kb, 5kb, 10 kb, 20kb, 40kb, 100kb, 500kb, and 1Mb resolutions. For all other HiC-Pro configuration parameters, the defaults were used.

### *Processing Reads and Calling Peaks for ChIP-seq Data Using ChIPLine*

Our previously developed ChIPLine pipeline (https://github.com/ay-lab/ChIPLine) was used to align ChIP-seq reads with Bowtie2 (Langmead et al., 2019) and call peaks using MACS2 (2.2.7.1) (Zhang et al., 2008) (**Supp Table 4**). If available, the corresponding input ChIP-seq files were used by MACS2 in ChIPLine for peak calling using input vs treatment mode. MACS2 peak calls were made without an input file if no such file was available for the sample. For ChIP-seq datasets with multiple biological replicates, the replicate with the largest number of peak calls was selected as the peak set to be used in downstream analyses for loop calling.

### *Calling Peaks from Reads of HiChIP Data Using FitHiChIP*

In the absence of matching ChIP-seq data, we turned to inferring peaks directly from HiChIP data. We called HiChIP peaks the PeakInferHiChIP.sh utility function of FitHiChIP 10.0 version (Bhattacharyya et al., 2019), which takes processed interaction pairs generated by HiC-Pro as input (i.e., one valid read pair corresponds to two individual reads) and utilizes MACS2 for peak calling using the *PeakInferHiChIP.sh* script by specifying the correct HiC-Pro output directory, reference genome, and read length as reported by SRA Run Selector entry for a given sample (**Supp Table 4**). If read lengths differed across technical replicates for a single HiChIP biological replicate, the mode read length was used. For cases in which there was no single mode read length, the longest read length was selected. For merged biological replicates, the mode read length of all individual biological replicates for a given sample was chosen, and similarly, the longest read length was selected in the case of multiple modes. We additionally used HiChIP-Peaks (0.1.2) (Shi et al., 2021) with its default parameters to infer peaks from HiChIP data; however, these peak calls spanned untypically large-sized genomic regions (median 2.15 kb and up to 122 kb) and we decided to not utilize these peak calls.

### *Performing Recall Analysis of 1D Peaks Inferred from HiChIP Data*

Peaks called from ChIP-seq datasets were considered the ground truth set and peaks inferred directly from HiChIP datasets were assessed for their overlap with ChIP-seq peaks (**Supp Figure 2**). To measure the validity of HiChIP-inferred peaks, we compared them to the peaks from ChIP-seq data and computed the percentage of ChIP-seq peaks recovered using HiChIP-inferred

peaks. To obtain the intersection, each pair of corresponding HiChIP and ChIP-seq peak sets were intersected using bedtools allowing for 1kb slack on both sides.

### *Integration of Biological Replicates of HiChIP Data*

In order to generate more deeply sequenced contact maps from the initial set of samples, we grouped together samples originating from the same study, pulldown and biological replicate set for both human and mouse datasets (**Supp Table 1**). Before merging the biological replicates, the reproducibility was assessed using hicreppy (0.1.0) which generates a stratum-adjusted correlation coefficient (SCC) as a measure of similarity for a pair of HiChIP contact maps (https://github.com/cmdoret/hicreppy) (**Supp Table 5**). Briefly, contact maps in hic format were converted to cool format for hicreppy input at 1kb, 5kb, 10kb, 25kb, 50kb, 100kb, 250kb, 500kb, and 1mb resolutions using *hic2cool* (0.8.3) (https://github.com/4dn-dcic/hic2cool). *hicreppy* was run on all pairwise combinations of biological replicates from a given HiChIP experiment as follows: first, hicreppy *htrain* was used to determine the optimal smoothing parameter value, or *h*, for a pair of input HiChIP contact matrices at 5kb resolution. *htrain* was run on a subset of chromosomes (chr1, chr10, chr17, and chr19) and with a maximum possible *h-value* of 25. Default settings were used otherwise. Next, hicreppy *scc* was ran to generate a SCC for the matrix pair using the optimal *h-value* reported by htrain and at 5kb resolution considering chr1, chr10, chr17, and chr19 only. A group of biological replicates were merged if all pairwise combinations of replicates in that group resulted in a SCC greater than 0.8 (**Supp Figure 3**). Merging of biological replicates was performed by concatenating HiC-Pro validpairs files from which we performed all downstream peak and loop calling steps of our pipeline.

### *Identifying Significant Chromatin Loops from HiChIP Data*

Loop calling was performed for both unmerged and merged HiChIP biological replicates by (a) HiCCUPS (JuicerTools 1.22.01) (Rao et al., 2014), (b) FitHiChIP with HiChIP-inferred peaks (FH loops), and (c) FitHiChIP with ChIP-seq peaks (FC loops), when available (Durand et al., 2016) (**Supp Table 6, Supp Figure 4**). Briefly, HiC-Pro validpairs files for each sample were converted to .hic format using HiC-Pro's *hicpro2juicebox* utility with default parameters. HiCCUPS loop calling was initially performed for chr1 only with the following parameters: --cpu, --ignore-sparsity, -c chr1, -r 5000,10000,25000, and -k VC. Samples which passed the thresholds of at least 200 loops from chromosome 1 for human samples and at least 100 loops from chromosome 1 for mouse samples, both at 10kb resolution, were processed further with HiCCUPS for genome-wide loop calling using the same parameters. FitHiChIP loop calling was run with HiC-Pro validpairs

as input at the 5kb, 10kb, and 25kb resolutions for the Stringent (S) background model with coverage bias regression, merge filtering, peak-to-all interactions and an FDR threshold of 0.01. Default values were used for all other FitHiChIP parameters.

### *Assigning QC Flags to Peak and Loops Calls*

To assign a set of comprehensive QC flags to all HiChIP samples (unmerged and merged biological replicates), we derived criteria from quantifying numbers of peak and loop calls (**Supp Table 7, Supp Figure 5**). For both FH and ChIP-seq peaks, we assigned a "Poor" flag to samples with <=1000 peaks, "Warning" to samples with >1000 and <5000 peaks, and "Good" to samples with >=5000 peaks. To establish intermediate flags for loops, each one of the six FitHiChIP loop call configurations (Stringent 5kb, Loose 5kb, Stringent 10kb, Loose 10kb, Stringent 25kb, and Loose 25kb) was assigned a flag as follows: "Poor" if the number of Stringent loops was greater than the number of Loose loops at a given resolution or the sample possessed <=100 loops, "Warning" for >100 and <1000 loops, or "Good" for >= 1000 loops. Lastly, we set a final loop flag on each loop call configuration which considers both the peak and intermediate loop flags. This final loop flag similarly takes on the values of "Poor", "Warning", or "Good". "Good" was only assigned in cases where both the peak and intermediate flags were "Good". All other possible cases are further described in **Supp Table 7**.

### *Establishing a Set of High-Confidence Regulatory HiChIP Loops*

We established a high-confidence set of 53 unique human H3K27ac HiChIP samples from the pool of 120 human H3K27ac HiChIP datasets with either merged or unmerged biological replicates and over 10,000 stringent 5kb FC loops, henceforth called the HCRegLoops-All sample set. We additionally established two subsets of the HCRegLoops-All sample set as follows: HCRegLoops-Immune contains 27 samples from immune-associated cell types and HCRegLoops-Non-Immune contains 26 samples from non-immune cell types (**Supp Table 8**).

### *Identifying Significant Chromatin Loops from High Resolution Hi-C Data*

Loop calling was performed for 44 high resolution Hi-C samples gathered from the 4DN data Portal's (https://data.4dnucleome.org) "High-Resolution Hi-C Datasets" Collection using Mustache (Roayaei Ardakany et al., 2020). Processed .hic files were downloaded and Mustache was run with default parameters across chr1 through chr22 using raw, KR, and VC normalized contact matrices (-norm) to determine loops at 1kb, and 5kb resolutions (-r) (**Supp Table 9**).

### *Identifying SGLs in Immune-based Diseases*

SGLs were identified utilizing fine-mapped GWAS SNPs from CAUSALdb for Type 1 Diabetes (T1D), Rheumatoid Arthritis (RA), Psoriasis (PS), and Atopic Dermatitis (AD) which included 7, 7, 3, and 1 individual studies, respectively (**Supp Table 10**). The fine-mapped data was downloaded and lifted over from hg19 to hg38 using the MyVariantInfo Python package (Xin et al., 2016). We downloaded the GENCODE v30 transcriptome reference, filtered transcripts for type equal to "gene" and located coordinates of the transcription start site (TSS) (Frankish et al., 2021). For genes on the plus strand, the TSS would be assigned as a 1bp region at the start site and, for the minus strand, the 1bp region at the end site. Lastly, we extracted all HiChIP samples whose organ was classified as "Immune-associated". To integrate these datasets, loop anchors were intersected with fine-mapped GWAS SNPs and TSSs independently using bedtools *pairtobed*. Subsequently, loops were extracted as an SGL if at least one anchor contained a GWAS SNP and the opposing anchor contained a TSS.

### *Identifying SGLs with Immune-associated eQTL Studies*

SGLs were identified utilizing eQTLs from the eQTL Catalog which included CD4 T-cells, CD8 T-cells, B cells, Natural Killer cells and monocytes (**Supp Table 11**). Similarly to GWAS-SGLs, we used GENCODE v30 to locate the TSS and extracted a subset of the GWAS-SGL HiChIP samples whose cell type matched the eQTL studies. To integrate these datasets, loops were intersected with pairs of SNP-gene pairs using bedtools *pairtopair*. Subsequently, loops were extracted as an SGL if at least one anchor contained an eQTL SNP and the opposing anchor contained a promoter.

### *Building a Consensus Gene List for T1D*

From the MalaCards database (https://www.malacards.org/) a query was made using the term "*type_1_diabetes_mellitus"* and the list of associated genes was downloaded. For eDGAR (https://edgar.biocomp.unibo.it/gene_disease_db/), "*diabetes mellitus, 1*" was queried under the Main Tables tab and all gene symbols were extracted. OpenTargets hosts a disease based search with gene association scores and a query was made to the MONDO ID for T1D (MONDO_0005147). The corresponding genes were downloaded and filtered for an association score > 0.5. For the GWAS Catalog, a query was made using the previous MONDO ID and all associated genes were extracted. Table 1 of Klak et al (Klak et al., 2020) summarizes genes that have been associated with T1D and gene symbols were extracted from this resource.

### *Performing Motif Enrichment Analysis across Conserved Anchors*

Motif enrichment analysis on the HCRegLoops-All, HCRegLoops-Immune, and HCRegLoops-Non-Immune sample sets was performed. Briefly, for each sample set, we identified highly conserved loop anchors by compiling loops across all samples in the set, extracting anchors, and filtering the anchors for those involved in at least one loop call in at least 90% of samples from the given sample set. We downloaded 727 known human motifs from the 2022 JASPAR CORE database (Castro-Mondragon et al., 2022). Motif enrichment analysis was directly applied to the conserved anchor sites using MEME Suite SEA (Simple Enrichment Analysis) version 5.5.0 (Bailey et al., 2015) using a match p-value threshold of 1e-6 and default values for all other parameters (**Supp Table 12**).

### *Identifying Pairs of Motifs Overlapping Loop Anchors*

Briefly, for each sample in HCRegLoops-All, unique loop anchors were extracted and these were intersected with sample-specific corresponding ChIP-seq peaks using *bedtools intersect* with no slack. We selected the peak with the highest signal value to represent that anchor for cases in which multiple H3K27ac peaks overlapped one loop anchor. The resulting peak sets were deduplicated since a single peak may overlap multiple loop anchors. To determine the genomic coordinates of motifs from the 2022 JASPAR CORE database (n = 727 motifs) in these representative peak sets, MEME Suite FIMO (Find Individual Motif Occurrences) version 5.5.0 (Grant et al., 2011) was applied to the HCRegLoops-All sample set on a sample-by-sample basis using a match p-value threshold of 1e-6 and default values for all other parameters. For each sample, we intersected motif coordinates with loop coordinates using *bedtools pairtobed* and annotated loop anchors with the motifs falling within the anchor.

### *Performing Enrichment Analysis for Motif Pairs*

Statistical analyses were performed via a bootstrapping method for HCRegLoops-All HiChIP samples (**Supp Table 13, Supp Figure 6**). Bootstrapping was performed using a block bootstrap with anchors as the unit analysis and anchors were randomly shuffled within their corresponding chromosomes. To get the null distribution, we shuffled the dataset through the location of unique instances of anchors. From there, we assigned a uniform probability of drawing a given anchor within each chromosome with replacement and used a total of 100,000 simulations. P-values were then determined for each sample by counting up the total number of simulated pairs that were greater than or equal to observed pairs in the original sample. Due to the large number of pairs ranging from tens to hundreds of thousands for each sample, we only focused on motif pairs

with both motifs within the top 50 most frequently enriched motifs for that sample. A multiple testing correction using Benjamini-Hochberg was then utilized with these filtered pairs to obtain adjusted p-values.

### *Constructing Chromatin Interaction Networks using Loops*

To construct a network from chromatin loops, each anchor was considered a node and each significant loop an edge. In addition, anchors were labeled as promoters by intersecting with TSS coordinates (slack of +/- 2.5kb) and allowing the promoter label to take priority over any other possible label. For H3K27ac HiChIP data, non-promoter nodes/anchors are labeled as enhancers when they overlap with ChIP-seq peaks (no slack). All other nodes that are not a promoter or an enhancer were designated as “other”. After obtaining the annotated anchors, we created a weighted undirected graph using the loops as edges and loop strength as edge weights calculated as the -log10(q-value) of FitHiChIP loop significance. To trim outliers, we set values larger than 20 to 20 and further scaled these values to between 0 and 1 for ease of visualization.

### *Detecting and Prioritizing Communities*

Community detection was applied to the networks created using FC loops at 5kb. Two levels of community detection were applied, the first detected communities within the overall network created for each chromosome (high-level) followed by a second round that detected subcommunities of the communities reported in the first round (low-level). High-level communities were located by running the Louvain algorithm using default parameters as implemented by the CDlib Python package. Starting with each high-level community, subcommunities were called using the same Louvain parameters. CRank (Zitnik et al., 2018) was then applied at both levels to obtain a score that aggregates several properties related to the connectivity of a community into a single score for ranking.

### *Designing the Internal Database and Filesystem*

The database is composed of two main parts, the first contains high sample-level and low-level loop information while the second contains SGL specific tables with auxiliary tables used to add additional annotations. For the first part, we atomized the data into the following tables: *hic_sample, celltype, hicpro, chipseq_merged, fithichip_cp, fithichip_cp_loop, hiccup, fithichip_hp* and *reference* which allowed us to capture important metadata and the uniqueness of loops using different loop callers. The second part includes the *gwas_study, gwas_snp, snp, gene* and

*fithichip_cp_fm_sgls* tables, which allowed us to capture important relationships for SGLs and to facilitate their query. The full database schema can be found in **Supp Figure 7.**

### *Designing and Implementing the Web Interface*

Loop Catalog was built using Django v3.2 (https://www.djangoproject.com/) as the backend framework with all data stored using the Postgresql v9 (https://www.postgresql.org/) database management system. To style the frontend interface, we used Bootstrap 5.2 (https://getbootstrap.com/docs/5.2/). For implementing advanced tables and charts DataTables v1.12.1 (https://datatables.net), Charts.js v4.0 (https://www.chartjs.org/), and D3 v4.13.0 (https://d3js.org/) were used. To visualize genetic and epigenetic data the WashU Epigenome Browser v53.6.3 (https://epigenomegateway.wustl.edu/) was used which maintains an easily accessible web embedding. Lastly, CytoscapeJS v3.26.0 (https://js.cytoscape.org/) was used to visualize enhancer-promoter networks.

## UTILITY AND DISCUSSION

### *Curating a set of publicly available HiChIP samples for human and mouse cells*

In total, we collected 750 human and 281 mouse samples from 133 studies that cover a diverse set of cell types and cell lines. For primary human samples there is a concentration of immune cell types that include monocytes, natural killer (NK) cells, T cell and B cell subsets, among others (175 designated samples; 140 and 35 for human and mouse, respectively). As expected, cancer cell lines are well represented (e.g., HCC15, NCI-H1105, MCF7) together with other cell lines from normal tissue including heart-derived samples (e.g., aortic valve interstitial, coronary artery smooth muscle, and aortic smooth muscle cells) as listed in **Supp Table 1**.

Regarding the target protein in the HiChIP experiment, active regulatory element-associated histone mark H3K27ac was the most highly represented, accounting for 62% of human samples and 53% of mouse samples. Other frequently represented ChIP targets include CTCF for human datasets and H3K4me3 for mouse datasets. Cohesin subunits such as SMC1A was also a frequent protein of choice in both human and mouse samples (**Figure 1C**). It is evident that, while the majority of studies target functional interactions via H3K27ac or H3K4me3 pulldown, structural interactions via CTCF and cohesin pull-down are also well-represented among human and mouse HiChIP studies. Overall, this dataset provides a comprehensive coverage of HiChIP samples that investigate structural as well as regulatory loops profiled across hundreds of samples.

***Uniform processing of HiChIP samples and quality controls***

The raw HiChIP data aligned to human (hg38) or mouse (mm10) genomes were processed to generate high-confidence loops utilizing multiple loop calling approaches: **i)** FitHiChIP with ChIP-seq peaks (FC), **ii)** FitHiChIP with HiChIP-inferred peaks (FH), and **iii)** HiCCUPS (**Supp Figure 1**). To visualize these loops, we developed a webpage with a highly interactive table that allows selection of multiple samples, loop calling settings, resolutions and corresponding peaks (**Figure 2**). These selections can then be plugged into an embedding of the WashU Epigenome Browser, downloaded as track files or WashU hub files for additional use cases (**Figure 2B**). The Loop Catalog was initialized with loops called for 1031 unmerged HiChIP samples and 282 merged samples that were created after combining biological replicates from the same study that pass a correlation threshold for pairwise similarity of individual replicates (**Supp Figure 3**). An additional 6 immune cell-based *mega-merged* samples were created by merging across all donors and all biological replicates using data from two previous publications (V. Chandra et al., 2021; Schmiedel et al., 2021).

Overall, the Loop Catalog provides access to loops for 1319 HiChIP samples with FH loop calls for all samples and, of these, 386 samples had matching ChIP-seq data that enable loop calling using FC. In addition, we called HiCCUPS loops genome-wide for 426 samples, for which we had at least a certain number of loop calls from chr1 (200 for human, 100 for mouse). In particular, when comparing these figures to other databases, the Loop Catalog stands out with the highest number of HiChIP samples, totaling 1031 distinct and 1319 overall with merged samples. ChromLoops provides loop calls for 816 samples (772 HiChIP), however this total is spread across 13 different species, whereas HiChIPdb and CohesinDB offer significantly fewer HiChIP samples, with only 200 and 42, respectively (**Table 1**). It is noteworthy that various loop-calling methods were employed across all databases. With this consideration, HiChIPdb is most similar to the Loop Catalog, using mainly FitHiChIP for loop calling (**Table 1**). However, unlike the Loop Catalog, HiChIPdb derives peaks solely from HiChIP data, which tends to have a low recall rate with respect to ChIP-seq based peaks (**Supp Figure 2**). Conversely, Chromloops opted to use ChIA-PET Tool (V3), a tool specifically designed to address concerns for ChIA-PET experiments and may fail to correct for HiChIP biases. With regards to Cohesin-DB, this database used HiCCUPS, a software that is originally designed for high-depth Hi-C data (27467249_Durand_2016) (**Table 1**). Despite using HiCCUPS in addition to FitHiChIP, we found that, for most of our curated HiChIP samples, the number of HiCCUPS loops were quite low (e.g.,

less than 200 for human chr1) likely limiting the utility of those loop calls for downstream analysis. Another important feature unique to Loop Catalog is the extent of quality control that is performed on both peak (ChIP-seq or HiChIP-derived) and loop level to provide QC flags for each set of loop calls to inform users of the inferred quality of the data they are about to visualize or download for further analysis (**Supp Figure 5, Supp Table 7**).

***Visualization and exploration of loop calls through a web interface***

The Loop Catalog is underpinned by a comprehensive database that incorporates processed HiChIP data from GEO and dbGaP, high-resolution Hi-C data from the 4DN data portal, along with fine-mapped GWAS data for a number of immune-associated diseases and eQTLs for major immune cell types (**Figure 1**). When first accessing the platform, users are presented with an entry page featuring a selection for reference genome and analysis type including loops, SNP-Gene linking (SGLs) and statistics (**Figure 2A**). Subsequently, once an analysis has been selected, a secondary page will render the corresponding analysis with a navigation menu to switch between other analyses and website-related information. The data page offers immediate and extensive visualization of HiChIP and Hi-C samples with their associated loop calls, spanning various methods and resolutions, as illustrated in **Figure 2B**. For specific sample information, each sample is linked to a dedicated page displaying metadata, loop data, regulatory network analyses and a motif scanning analysis (**Figure 2C**). Furthermore, the SGL page grants access to the integration of immune-based HiChIP samples and fine-mapped GWAS SNPs and eQTLs. Lastly, the statistics page provides a higher-level overview of all HiChIP data stored for a given reference genome.

***Expanding the use of HiChIP data for annotating GWAS SNPs and eQTLs using SNP-to-Gene Loops***

Leveraging the Loop Catalog we integrated loops for immune-related HiChIP samples, we identified 79 samples with nonzero FC loops at 5kb (63 unmerged, 10 bio-rep merged, and 6 all donors merged). We then used these loop calls together with fine-mapped GWAS SNPs from CAUSALdb to find target genes for each SNP that we term a SNP-gene pair with a loop (SGL) (**Figure 3A**). Briefly, across T1D, RA, PS and AD there are 7729, 1121, 590, and 674 unique fine-mapped GWAS SNPs, respectively. After overlapping these SNPs with our loop anchors and genes connected through such loops, we found 74 samples with at least one SGL and a total of 182,306 SGL instances across 18 studies covering the four diseases. When removing duplicate SGLs we found 28,162 distinct SGLs which included 4241 SNPs, 2354 genes and 7269 loops (**Figure 3B, Supp Table 10**). To store these results into the database and to allow querying , we

constructed SNP, GWAS, gene and loop-level tables with genomic coordinates. To then browse these results, an SGL entry page allows users to first select their target disease, locus and samples from which the loops will be derived (**Figure 2D**). Subsequently, the Loop Catalog returns an SGL analysis page that includes an embedded WashU Epigenome Browser element loaded with a track for fine-mapped SNPs and loops tracks for each sample. Below the browser users will find an interactive table that lists all mapped SGLs for their selection. This table also allows navigating between loops and within loops including the left or right anchor and SNP positions (**Figure 2E**).

Similar to GWAS variants, we expanded our annotation of genomic elements using eQTL SNP-gene pairs. We started by downloading uniformly processed eQTL studies from the eQTL Catalogue and focused on cell types from Schmiedel et al 2018 dataset that include eQTLs for naive CD4 T-cells (n=64386), naive CD8 T-cells (n=67793), naive B cells (n=60629), NK cells (n=48221), monocytes (n=66024) and nonclassical monocytes (58069). For these cell types, we also have the HiChIP data derived from a subset of the same donors (V. Chandra et al., 2021; Schmiedel et al., 2018, 2021). For these cell types, we then located 11604 unique eQTL-SGLs that cover 11053 SNPs and 1128 genes (**Supp Figure 8**, **Supp Table 11**). These results are made available through a similar web interface as SGLs derived from GWAS (**Supp Figure 9**). It is possible to extend our eQTL-based SGLs analysis to the remainder of the eQTL Catalogue, however, matching cell types from eQTL studies to those from HiChIP studies is not a trivial task for most of the cases.

***Utilizing T1D SGLs for SNP and Gene Prioritization***

To better understand the utility of using SGLs for linking GWAS SNPs and genes, we focused our attention on analyzing SGLs in T1D. When compared to all other diseases, T1D has the highest number of unique SGLs (n=16,534) mainly due to the high number of fine-mapped GWAS SNPs as our starting point (**Figure 3B**). In post-GWAS analyses, it is important to distinguish putative causal SNPs among those that are in linkage disequilibrium (high LD). In cases where the phenotypic effect of the SNP is mainly through regulation of a distal gene, SGLs can corroborate important information to accurately annotate SNP function and to prioritize GWAS genes and SNPs while utilizing information on their 3D proximity. Investigating this for T1D, we observed that at least half of the CAUSALdb SNPs participate in an SGL, these SNPs are often in contact with multiple genes (**Figure 3B**). We further explored the multiplicity of SNP and gene links within T1D and found that 26% of SGL genes are linked to a single SNP, and the median number of SNP

links per gene is 3. On the other hand, 27% of SGL SNPs are linked to a single gene and the median number of gene links per gene is 3 (**Figure 3C**).

To understand if SGL genes overlap genes with known T1D associations, we built a consensus gene list using MalaCards (Rappaport et al., 2013), eDGAR (Babbi et al., 2017), OpenTargets (Ochoa et al., 2023), and GWAS Catalog (Sollis et al., 2023) and a T1D review paper (Klak et al., 2020). The union of the gene lists across these five resources had 497 genes in total, of which 106 overlapped with our 1532 SGL genes identified for T1D. The 1426 genes uniquely found by our SGL approach, although likely to involve false positives, are potential targets for future investigations (**Figure 3D**). One of these genes was *IL15RA*, a cytokine receptor that binds the pro-inflammatory cytokine IL-15 with high affinity and through *cis* and *trans* presentation of IL-15 impacts cellular functions of CD8 T cells as well as Natural Killer (NK) cells (Waldmann, 2015). Despite *IL15RA* not being within the consensus T1D gene list, the IL15RA/IL-15 axis has been associated with T1D but whether this axis has a pathogenic or protective role has not been clear (Bobbala et al., 2017; J. Chen et al., 2013; Lu et al., 2020). Through our SGL analysis of H3K27ac HiChIP loop calls from major immune cells, we identified looping between the *IL15RA* promoter and rs61839660, a SNP 75kb away that is highly associated with T1D and has been further prioritized by fine-mapping in three out of four T1D-GWAS studies with a posterior probability greater than 0.70. The corresponding loops are found for T cells, Naive B cells and NK cells but not for monocytes suggesting an important role for this SGL within the adaptive immune system. The more likely scenario is that rs61839660's T1D association is mediated through *IL2RA* given that this SNP falls within a constituent intron. However, specific loops connecting rs61839660 to *IL15RA* (P-value < $10^{-9}$) as well as to *RBM17* (P-value < $10^{-11}$) suggest the possibility of a pleiotropic effect for this SNP (**Figure 3E**, **Supp Table 10**). In addition, *RBM17*, an RNA-binding protein, has been previously shown to affect other autoimmune diseases such as rheumatoid arthritis (Ibáñez-Costa et al., 2022). As exemplified here, SGL analysis with the Loop Catalog may provide further evidence and/or mechanisms of action for a genetic variant and its target gene. In addition, it may help to find targets for GWAS variants whose target gene remains elusive.

***Identifying Significant Sequence Motifs at Regulatory Loop Anchors***

In order to examine binding patterns of TFs in regulatory loops (H3K27ac HiChIP), we performed 1D motif enrichment analysis on highly conserved regulatory loop anchors from three high-confidence (HC) sample sets (**Figure 4A**). The HCRegLoops-All sample set contains the 53 H3K27ac HiChIP samples with over 10000 stringent 5kb FC loops. These samples encompass

diverse cell types including immune cells, heart cells, and various cancer cell lines. The HCRegLoops-Immune sample set contains the subset of 27 samples from the HCRegLoops-All set which are from immune-associated cell types and the remaining 26 samples are contained in HCRegLoops-Non-Immune. In terms of conserved anchors that were derived for motif enrichment analysis, we annotated an anchor as conserved if was involved in at least one loop call in 90% or more of the samples from the given sample set and identified 879, 2766, and 721 anchors fitting this criterion for HCRegLoops-All, HCRegLoops-Immune, and HCRegLoops-Non-Immune, respectively. Using the Simple Enrichment Analysis tool SEA from the MEME Suite, we identified 205, 313, and 177 significantly enriched motifs (p-value < 1e-6) for the three sample sets, respectively (**Supp Table 12**). The top 3 most significantly enriched motifs were *ZNF460*, *ZNF135*, and *MEF2D* or *MEF2A* across all three sets (**Figure 4B**). *ZNF460* demonstrated strikingly high enrichment relative to other motifs, especially in the HCRegLoops-Immune set, with over 58-fold enrichment compared to control regions in this sample set, in comparison to 2.15-fold enrichment for CTCF (**Figure 4C**). Although similar to CTCF in regard to possessing 11 zinc finger domains (Dai et al., 2003), *ZNF460* also harbors a KRAB domain (like one third of all ZNFs (Lupo et al., 2013) that is associated with transcriptional repression. It is therefore surprising to have ZNF460 motif enrichment in anchors of loops detected through enrichment of H3K27ac, an active histone mark. It is possible that ZNF460 may be playing an important role in looping and one that is more pronounced for immune cells. However, recent studies highlight the importance of caution and the requirement for functional validation by showing that, for ZNF143, another ZNF with a presumed role in looping, such an association was a result of antibody cross-reactivity (Magnitov et al., 2024; Narducci & Hansen, 2024). Our motif enrichment analysis also identified two other ZNFs, MAZ and PATZ1, which are recently shown to have novel chromatin insulating activities, similar to CTCF (Ortabozkoyun et al., 2022, 2023). Regardless of the specific transcription factors, our provided analysis of TF enrichment at loop anchors for a diverse set of samples is useful for developing hypotheses about generic and cell-type-specific regulators or correlates of looping.

### *Identifying Significant Motif-Pairs Across Loop Anchors*

Transcription factors often work together, through co-binding, dimerization, or multimerization, to promote gene expression. With this motivation, we expanded our motif analysis at conserved anchors to a search for significantly overrepresented motif pairs connected by loops. Since the brute-force approach of scanning all possible motifs across the genome for enrichment would have been infeasible, we performed bootstrap analysis to calculate an empirical p-value. We used

the frequency of a given motif pair across the entire set of loops as our statistic and tested whether the frequency of this motif pair appeared greater than expected by chance for a given loop set. Bootstrap was used to build a simulated null distribution by shuffling loop anchors and thus motif pairs (see methods "Performing Enrichment Analysis for Motif Pairs"). A p-value was then calculated as the fraction of simulations greater than our observed frequency (**Figure 4A** (bottom); **Supp Figure 6**). We performed this analysis on the HCRegLoops-ALL sample sets (**Supp Table 13**) and, to ensure our bootstrapping approach does not lead to inflated p-values, we investigated the Q-Q plots of our samples. We observed that most p-values lie along the diagonal for non-significant p-values suggesting no inflation (**Figure 4D**). Summarizing this analysis across all of our samples, we found signals for ZNF460 paired with motifs such as ZNF135 and THRA across a majority of samples (**Figure 4E**). Self pairs of ZNF135 and ZNF460 were also overrepresented across multiple samples. Overall, this is an interesting finding which suggests an association for zinc finger proteins other than CTCF with chromatin loops and their anchor regions.

***Investigating topological properties of enhancer-promoter interactions networks derived from HiChIP loops***

In addition, for a subset of human H3K27ac HiChIP samples, we provide a network analysis with nodes being genomic regions (including regulatory elements such as enhancers and promoters) and edges representing connections by significant loops we detected. The network analysis was run for 240 human samples with 5kb FC loops of which 151 had at least one community (strongly connected set of nodes also referred to as network modules) detected. For these, we performed a two-level community detection for each chromosome where we first defined communities and then defined sub-communities for those with a large number of nodes. For both communities and sub-communities, we used CRank, an unsupervised technique for prioritization of network communities depending on their connectivity and topology parameters such as conductance, modularity and randomness (Zitnik et al., 2018). Sub-communities are then visualized using CytoscapeJS to provide a dynamic interface and users can switch between sub-communities using an interactive table that allows filtering on various CRank variables (**Figure 2C**). The provided functionalities for interrogating and ranking network communities at multiple levels will allow downstream analysis such as hyperconnected cliques (Petrovic et al., 2019) and multi-enhancer hubs (A. Chandra et al., 2022) that were previously studied in the context of oncogenes and immune cell development, respectively.

## CONCLUSION

The HiChIP assay has empowered the field of chromatin structure by providing a relatively inexpensive, high-resolution and targeted approach to mapping chromatin interactions. It is no surprise that the number of HiChIP studies published annually keeps growing each year and will continue to do so until superior methods are developed. The Loop Catalog is a public hub that centralizes 1319 HiChIP samples from these studies and, through the use of user-facing tools, lowers the bar for exploring 3D genome organization datasets, which would require substantial bioinformatics skills otherwise. As previewed by our applications, we foresee the Loop Catalog becoming a valuable resource for a broad range of chromatin studies including but not limited to variant-gene prioritization, machine learning, and deep learning approaches for loop prediction or utilizing loops to predict other functional measurements including gene expression as well as benchmarking analyses of such methods. At its current state, the Loop Catalog allows the community to access and bulk download uniformly processed chromatin looping data from HiChIP experiments and intermediate files, as desired. In addition, the SGL analysis, initially made available for GWAS studies in four immune-related diseases, makes it possible to query variants or genes of interest with GWAS SNPs or a prioritized subset of them such as fine-mapped SNPs. Lastly, by adding SGLs derived from eQTL, we provide epigenetic context to gene regulation via genetic variation.

The Loop Catalog portal is a powerful way to access HiChIP data while providing additional analyses that are of strong interest to the scientific community. We plan to continue expanding the Loop Catalog as newly published datasets become available and our overarching, ongoing goal is to build a chromatin-centric database that seamlessly expands in terms of data as well as computational utilities offered that will eventually become the go-to platform for access and analysis of all published HiChIP data.

### Abbreviations

- 4DN : 4D Nucleome
- AD : Atopic Dermatitis
- ChIP-seq : Chromatin Immunoprecipitation Followed by Sequencing
- ChIA-PET : Chromatin Interaction Analysis with Paired-End Tag
- CTCF : CCCTC-Binding Factor
- dbGaP : Database of Genotypes and Phenotypes
- EBI : European Bioinformatics Institute

- eQTL : Expression Quantitative Trait Loci
- ENCODE : ENCyclopedia Of DNA Elements
- FDR : False Discovery Rate
- FC Loops : FitHiChIP with ChIP-seq Peaks
- FH Loops : FitHiChIP with HiChIP Peaks
- FIMO : (MEME Suite) Find Individual Motif Occurrences
- GEO : Gene Expression Omnibus
- GWAS : Genome-wide Association Study
- HCRegLoops : High Confidence Regulatory Loops
- Hi-C : High-Throughput Chromosome Conformation Capture
- HiCCUPS : Hi-C Computational Unbiased Peak Search
- HiChIP : High-Throughput Chromosome Conformation Capture with ChIP
- KR : Knight and Ruiz (normalization)
- LD : Linkage Disequilibrium
- MAPQ : Mapping Quality (score)
- NK : Natural Killer Cells
- QC : Quality Control
- RA : Rheumatoid Arthritis
- SCC : Stratum-Adjusted Correlation Coefficient
- SEA : (MEME Suite) Simple Enrichment Analysis
- SGL : SNP-to-Gene Linking
- SNP : Single Nucleotide Polymorphism
- SRA : Sequence Read Archive
- PS : Psoriasis
- T1D : Type 1 Diabetes
- TF : Transcription Factor
- TSS : Transcription Start Site
- VC : Vanilla Coverage (normalization)
- ZNF : Zinc Finger Protein

**Experimental design and statistics**

No new experiments were performed. All data were downloaded from published data sources.

**Resources**

All software and resources utilized in our work are reported in **Supp Table 14**.

**Declarations**

**Ethics approval and consent to participate:** Not applicable

**Consent for publication:** Not applicable

**Availability of data and materials:**

The Loop Catalog is freely available at https://loopcatalog.lji.org to all users without any log-in or registration requirements. The main processing pipeline has been released on Github as Loop-Catalog-Pipelines (https://github.com/ay-lab-team/Loop-Catalog-Pipelines). Similarly, we developed GEO-Resources (https://github.com/ay-lab-team/GEO-Resources) to locate HiChIP datasets in NCBI GEO, motif_pair_enrichment (https://github.com/ay-lab-team/motif_pair_enrichment) to perform enrichment analyses for motif pairs, Community-Detection-Using-Chromatin-Loops (https://github.com/ay-lab-team/Community-Detection-Using-Chromatin-Loops) to locate communities formed by chromatin loops, and T1D-Loop-Catalog (https://github.com/ay-lab-team/T1D-Loop-Catalog) to detect immune-associated SGLs. Versions of all software included in our pipeline are recorded (**Supp Table 14**). Raw sequencing reads for HiChIP and ChIP-seq were downloaded from NCBI GEO (https://www.ncbi.nlm.nih.gov/geo/) and NCBI dbGaP (https://www.ncbi.nlm.nih.gov/gap). Hi-C contact matrices were retrieved from the 4DN Data Portal (https://data.4dnucleome.org). Fine-mapped GWAS SNPs were retrieved from CAUSALdb (http://www.mulinlab.org/causaldb).

**Competing interests:** None

**Funding:** This work was supported by National Institutes of Health (NIH) grants R03-OD034494 and R35-GM128938 to F.A.. K.F. was supported by the UCSD Regents Scholarship and the Philip and Elizabeth Hiestand Scholarship for Research in SIO and Engineering.

**Authors' contributions:**

Joaquin Reyna: Code development, Web Development, Formal analysis, Writing—original draft.

Kyra Fetter: Data collection, Formal analysis, Writing—original draft.

Romeo Ignacio: Formal analysis, Writing—original draft.

Cemil Can Ali Marandi: Formal analysis, Writing—original draft.

Nikhil Rao: Data collection, Web development.
Zichen Jiang: Data collection, Web development.
Daniela Salgado Figueroa: Data collection.
Sourya Bhattacharya: Conceptualization, Writing—review & editing.
Ferhat Ay: Conceptualization, Writing—review & editing.

**Acknowledgements:**
We would like to thank Michael Talbott from La Jolla Institute for Immunology (LJI) whose web development expertise helped us set up the website and troubleshoot along the way.

## REFERENCES

Babbi, G., Martelli, P. L., Profiti, G., Bovo, S., Savojardo, C., & Casadio, R. (2017). eDGAR: A database of Disease-Gene Associations with annotated Relationships among genes. *BMC Genomics*, *18*(Suppl 5), 554. https://doi.org/10.1186/s12864-017-3911-3

Bailey, T. L., Johnson, J., Grant, C. E., & Noble, W. S. (2015). The MEME Suite. *Nucleic Acids Research*, *43*(W1), W39-49. https://doi.org/10.1093/nar/gkv416

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C. L., Serova, N., Davis, S., & Soboleva, A. (2013). NCBI GEO: Archive for functional genomics data sets--update. *Nucleic Acids Research*, *41*(Database issue), D991-995. https://doi.org/10.1093/nar/gks1193

Bernstein, B. E., Stamatoyannopoulos, J. A., Costello, J. F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M. A., Beaudet, A. L., Ecker, J. R., Farnham, P. J., Hirst, M., Lander, E. S., Mikkelsen, T. S., & Thomson, J. A. (2010). The NIH Roadmap Epigenomics Mapping Consortium. *Nature Biotechnology*, *28*(10), 1045–1048. https://doi.org/10.1038/nbt1010-1045

Bhattacharyya, S., Chandra, V., Vijayanand, P., & Ay, F. (2019). Identification of significant chromatin contacts from HiChIP data by FitHiChIP. *Nature Communications*, *10*(1), 4221. https://doi.org/10.1038/s41467-019-11950-y

Bobbala, D., Mayhue, M., Menendez, A., Ilangumaran, S., & Ramanathan, S. (2017). Trans-presentation of interleukin-15 by interleukin-15 receptor alpha is dispensable for the pathogenesis of autoimmune type 1 diabetes. *Cellular & Molecular Immunology*, *14*(7), 590–596. https://doi.org/10.1038/cmi.2015.102

Boltsis, I., Grosveld, F., Giraud, G., & Kolovos, P. (2021). Chromatin Conformation in Development and Disease. *Frontiers in Cell and Developmental Biology*, *9*, 723859. https://doi.org/10.3389/fcell.2021.723859

Brandão, H. B., Gabriele, M., & Hansen, A. S. (2021). Tracking and interpreting long-range chromatin interactions with super-resolution live-cell imaging. *Current Opinion in Cell Biology*, *70*, 18–26. https://doi.org/10.1016/j.ceb.2020.11.002

Buchmann, J. P., & Holmes, E. C. (2019). Entrezpy: A Python library to dynamically interact with the NCBI Entrez databases. *Bioinformatics (Oxford, England)*, *35*(21), 4511–4514. https://doi.org/10.1093/bioinformatics/btz385

Castro-Mondragon, J. A., Riudavets-Puig, R., Rauluseviciute, I., Lemma, R. B., Turchi, L., Blanc-Mathieu, R., Lucas, J., Boddie, P., Khan, A., Manosalva Pérez, N., Fornes, O., Leung, T. Y., Aguirre, A., Hammal, F., Schmelter, D., Baranasic, D., Ballester, B., Sandelin, A., Lenhard, B., … Mathelier, A. (2022). JASPAR 2022: The 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Research*, *50*(D1), D165–D173. https://doi.org/10.1093/nar/gkab1113

Chakraborty, A., & Ay, F. (2019). The role of 3D genome organization in disease: From compartments to single nucleotides. *Seminars in Cell & Developmental Biology*, *90*, 104–113. https://doi.org/10.1016/j.semcdb.2018.07.005

Chandra, A., Yoon, S., Michieletto, M. F., Goldman, N., Ferrari, E. K., Fasolino, M., Joannas, L.,

Kee, B. L., Henao-Mejia, J., & Vahedi, G. (2022). *A multi-enhancer hub at the Ets1 locus controls T cell differentiation and allergic inflammation through 3D genome topology* (p. 2022.10.28.514213). bioRxiv. https://doi.org/10.1101/2022.10.28.514213

Chandra, V., Bhattacharyya, S., Schmiedel, B. J., Madrigal, A., Gonzalez-Colin, C., Fotsing, S., Crinklaw, A., Seumois, G., Mohammadi, P., Kronenberg, M., Peters, B., Ay, F., & Vijayanand, P. (2021). Promoter-interacting expression quantitative trait loci are enriched for functional genetic variants. *Nature Genetics*, *53*(1), 110–119. https://doi.org/10.1038/s41588-020-00745-3

Chen, J., Feigenbaum, L., Awasthi, P., Butcher, D. O., Anver, M. R., Golubeva, Y. G., Bamford, R., Zhang, X., St Claire, M. B., Thomas, C. J., Discepolo, V., Jabri, B., & Waldmann, T. A. (2013). Insulin-dependent diabetes induced by pancreatic beta cell expression of IL-15 and IL-15Rα. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(33), 13534–13539. https://doi.org/10.1073/pnas.1312911110

Chen, L.-F., & Long, H. K. (2023). Topology regulatory elements: From shaping genome architecture to gene regulation. *Current Opinion in Structural Biology*, *83*, 102723. https://doi.org/10.1016/j.sbi.2023.102723

Dai, J., Li, Y., Ji, C., Jin, F., Zheng, Z., Wang, X., Sun, X., Xu, X., Gu, S., Xie, Y., & Mao, Y. (2003). Characterization of two novel KRAB-domain-containing zinc finger genes, ZNF460 and ZNF461, on human chromosome 19q13.1—>q13.4. *Cytogenetic and Genome Research*, *103*(1–2), 74–78. https://doi.org/10.1159/000076292

Davies, J. O. J., Telenius, J. M., McGowan, S. J., Roberts, N. A., Taylor, S., Higgs, D. R., & Hughes, J. R. (2016). Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nature Methods*, *13*(1), 74–80. https://doi.org/10.1038/nmeth.3664

Dekker, J., Belmont, A. S., Guttman, M., Leshyk, V. O., Lis, J. T., Lomvardas, S., Mirny, L. A., O'Shea, C. C., Park, P. J., Ren, B., Politz, J. C. R., Shendure, J., & Zhong, S. (2017). The 4D nucleome project. *Nature*, *549*(7671), 219–226.

https://doi.org/10.1038/nature23884

Duan, A., Wang, H., Zhu, Y., Wang, Q., Zhang, J., Hou, Q., Xing, Y., Shi, J., Hou, J., Qin, Z., Chen, Z., Liu, Z., & Yang, J. (2021). Chromatin architecture reveals cell type-specific target genes for kidney disease risk variants. *BMC Biology*, *19*(1), 38. https://doi.org/10.1186/s12915-021-00977-7

Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S. P., Huntley, M. H., Lander, E. S., & Aiden, E. L. (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Systems*, *3*(1), 95–98. https://doi.org/10.1016/j.cels.2016.07.002

ENCODE Project Consortium, Moore, J. E., Purcaro, M. J., Pratt, H. E., Epstein, C. B., Shoresh, N., Adrian, J., Kawli, T., Davis, C. A., Dobin, A., Kaul, R., Halow, J., Van Nostrand, E. L., Freese, P., Gorkin, D. U., Shen, Y., He, Y., Mackiewicz, M., Pauli-Behn, F., … Weng, Z. (2022). Author Correction: Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*, *605*(7909), E3. https://doi.org/10.1038/s41586-021-04226-3

Ewels, P. (2024). *SRA Explorer* [Computer software]. https://sra-explorer.info/

Fang, R., Yu, M., Li, G., Chee, S., Liu, T., Schmitt, A. D., & Ren, B. (2016). Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. *Cell Research*, *26*(12), 1345–1348. https://doi.org/10.1038/cr.2016.137

Frankish, A., Diekhans, M., Jungreis, I., Lagarde, J., Loveland, J. E., Mudge, J. M., Sisu, C., Wright, J. C., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Boix, C., Carbonell Sala, S., Cunningham, F., Di Domenico, T., Donaldson, S., Fiddes, I. T., García Girón, C., … Flicek, P. (2021). GENCODE 2021. *Nucleic Acids Research*, *49*(D1), D916–D923. https://doi.org/10.1093/nar/gkaa1087

Friman, E. T., Flyamer, I. M., Marenduzzo, D., Boyle, S., & Bickmore, W. A. (2023). Ultra-long-range interactions between active regulatory elements. *Genome Research*, *33*(8), 1269–1283. https://doi.org/10.1101/gr.277567.122

Fu, Y., Tessneer, K. L., Li, C., & Gaffney, P. M. (2018). From association to mechanism in complex disease genetics: The role of the 3D genome. *Arthritis Research & Therapy*, *20*(1), 216. https://doi.org/10.1186/s13075-018-1721-x

Fullwood, M. J., Liu, M. H., Pan, Y. F., Liu, J., Xu, H., Mohamed, Y. B., Orlov, Y. L., Velkov, S., Ho, A., Mei, P. H., Chew, E. G. Y., Huang, P. Y. H., Welboren, W.-J., Han, Y., Ooi, H. S., Ariyaratne, P. N., Vega, V. B., Luo, Y., Tan, P. Y., … Ruan, Y. (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature*, *462*(7269), 58–64. https://doi.org/10.1038/nature08497

Gazal, S., Weissbrod, O., Hormozdiari, F., Dey, K. K., Nasser, J., Jagadeesh, K. A., Weiner, D. J., Shi, H., Fulco, C. P., O'Connor, L. J., Pasaniuc, B., Engreitz, J. M., & Price, A. L. (2022). Combining SNP-to-gene linking strategies to identify disease genes and assess disease omnigenicity. *Nature Genetics*, *54*(6), 827–836. https://doi.org/10.1038/s41588-022-01087-y

Giambartolomei, C., Seo, J.-H., Schwarz, T., Freund, M. K., Johnson, R. D., Spisak, S., Baca, S. C., Gusev, A., Mancuso, N., Pasaniuc, B., & Freedman, M. L. (2021). H3K27ac HiChIP in prostate cell lines identifies risk genes for prostate cancer susceptibility. *American Journal of Human Genetics*, *108*(12), 2284–2300. https://doi.org/10.1016/j.ajhg.2021.11.007

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

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

Hu, Y., Salgado Figueroa, D., Zhang, Z., Veselits, M., Bhattacharyya, S., Kashiwagi, M., Clark, M. R., Morgan, B. A., Ay, F., & Georgopoulos, K. (2023). Lineage-specific 3D genome

organization is assembled at multiple scales by IKAROS. *Cell*, *186*(24), 5269-5289.e22. https://doi.org/10.1016/j.cell.2023.10.023

Ibáñez-Costa, A., Perez-Sanchez, C., Patiño-Trives, A. M., Luque-Tevar, M., Font, P., Arias de la Rosa, I., Roman-Rodriguez, C., Abalos-Aguilera, M. C., Conde, C., Gonzalez, A., Pedraza-Arevalo, S., Del Rio-Moreno, M., Blazquez-Encinas, R., Segui, P., Calvo, J., Ortega Castro, R., Escudero-Contreras, A., Barbarroja, N., Aguirre, M. A., … Lopez-Pedrera, C. (2022). Splicing machinery is impaired in rheumatoid arthritis, associated with disease activity and modulated by anti-TNF therapy. *Annals of the Rheumatic Diseases*, *81*(1), 56–67. https://doi.org/10.1136/annrheumdis-2021-220308

Ibrahim, D. M., & Mundlos, S. (2020). The role of 3D chromatin domains in gene regulation: A multi-facetted view on genome organization. *Current Opinion in Genetics & Development*, *61*, 1–8. https://doi.org/10.1016/j.gde.2020.02.015

Ito, S., Das, N. D., Umehara, T., & Koseki, H. (2022). Factors and Mechanisms That Influence Chromatin-Mediated Enhancer-Promoter Interactions and Transcriptional Regulation. *Cancers*, *14*(21), 5404. https://doi.org/10.3390/cancers14215404

Klak, M., Gomółka, M., Kowalska, P., Cichoń, J., Ambrożkiewicz, F., Serwańska-Świętek, M., Berman, A., & Wszoła, M. (2020). Type 1 diabetes: Genes associated with disease development. *Central-European Journal of Immunology*, *45*(4), 439–453. https://doi.org/10.5114/ceji.2020.103386

Langmead, B., Wilks, C., Antonescu, V., & Charles, R. (2019). Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics (Oxford, England)*, *35*(3), 421–432. https://doi.org/10.1093/bioinformatics/bty648

Levo, M., Raimundo, J., Bing, X. Y., Sisco, Z., Batut, P. J., Ryabichko, S., Gregor, T., & Levine, M. S. (2022). Transcriptional coupling of distant regulatory genes in living embryos. *Nature*, *605*(7911), 754–760. https://doi.org/10.1038/s41586-022-04680-7

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

Liu, S., Cao, Y., Cui, K., Tang, Q., & Zhao, K. (2022). Hi-TrAC reveals division of labor of transcription factors in organizing chromatin loops. *Nature Communications*, *13*(1), 6679. https://doi.org/10.1038/s41467-022-34276-8

Long, H. K., Osterwalder, M., Welsh, I. C., Hansen, K., Davies, J. O. J., Liu, Y. E., Koska, M., Adams, A. T., Aho, R., Arora, N., Ikeda, K., Williams, R. M., Sauka-Spengler, T., Porteus, M. H., Mohun, T., Dickel, D. E., Swigut, T., Hughes, J. R., Higgs, D. R., … Wysocka, J. (2020). Loss of Extreme Long-Range Enhancers in Human Neural Crest Drives a Craniofacial Disorder. *Cell Stem Cell*, *27*(5), 765-783.e14. https://doi.org/10.1016/j.stem.2020.09.001

Lu, J., Liu, J., Li, L., Lan, Y., & Liang, Y. (2020). Cytokines in type 1 diabetes: Mechanisms of action and immunotherapeutic targets. *Clinical & Translational Immunology*, *9*(3), e1122. https://doi.org/10.1002/cti2.1122

Lupo, A., Cesaro, E., Montano, G., Zurlo, D., Izzo, P., & Costanzo, P. (2013). KRAB-Zinc Finger Proteins: A Repressor Family Displaying Multiple Biological Functions. *Current Genomics*, *14*(4), 268–278. https://doi.org/10.2174/13892029113149990002

Magnitov, M. D., Maresca, M., Saiz, N. A., Teunissen, H., Braccioli, L., & Wit, E. de. (2024). *ZNF143 is a transcriptional regulator of nuclear-encoded mitochondrial genes that acts independently of looping and CTCF* (p. 2024.03.08.583864). bioRxiv. https://doi.org/10.1101/2024.03.08.583864

Martens, J. H. A., & Stunnenberg, H. G. (2013). BLUEPRINT: Mapping human blood cell epigenomes. *Haematologica*, *98*(10), 1487–1489.

https://doi.org/10.3324/haematol.2013.094243

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

Mumbach, M. R., Satpathy, A. T., Boyle, E. A., Dai, C., Gowen, B. G., Cho, S. W., Nguyen, M. L., Rubin, A. J., Granja, J. M., Kazane, K. R., Wei, Y., Nguyen, T., Greenside, P. G., Corces, M. R., Tycko, J., Simeonov, D. R., Suliman, N., Li, R., Xu, J., … Chang, H. Y. (2017). Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nature Genetics*, *49*(11), 1602–1612. https://doi.org/10.1038/ng.3963

Narducci, D. N., & Hansen, A. S. (2024). *Putative Looping Factor ZNF143/ZFP143 is an Essential Transcriptional Regulator with No Looping Function* (p. 2024.03.08.583987). bioRxiv. https://doi.org/10.1101/2024.03.08.583987

Ochoa, D., Hercules, A., Carmona, M., Suveges, D., Baker, J., Malangone, C., Lopez, I., Miranda, A., Cruz-Castillo, C., Fumis, L., Bernal-Llinares, M., Tsukanov, K., Cornu, H., Tsirigos, K., Razuvayevskaya, O., Buniello, A., Schwartzentruber, J., Karim, M., Ariano, B., … McDonagh, E. M. (2023). The next-generation Open Targets Platform: Reimagined, redesigned, rebuilt. *Nucleic Acids Research*, *51*(D1), D1353–D1359. https://doi.org/10.1093/nar/gkac1046

O'Mara, T. A., Spurdle, A. B., Glubb, D. M., & Endometrial Cancer Association Consortium. (2019). Analysis of Promoter-Associated Chromatin Interactions Reveals Biologically Relevant Candidate Target Genes at Endometrial Cancer Risk Loci. *Cancers*, *11*(10), 1440. https://doi.org/10.3390/cancers11101440

Orozco, G., Schoenfelder, S., Walker, N., Eyre, S., & Fraser, P. (2022). 3D genome organization links non-coding disease-associated variants to genes. *Frontiers in Cell and Developmental Biology*, *10*, 995388. https://doi.org/10.3389/fcell.2022.995388

Ortabozkoyun, H., Huang, P.-Y., Cho, H., Narendra, V., LeRoy, G., Gonzalez-Buendia, E., Skok, J. A., Tsirigos, A., Mazzoni, E. O., & Reinberg, D. (2022). CRISPR and biochemical screens identify MAZ as a cofactor in CTCF-mediated insulation at Hox clusters. *Nature Genetics*, *54*(2), 202–212. https://doi.org/10.1038/s41588-021-01008-5

Ortabozkoyun, H., Huang, P.-Y., Cho, H., Tsirigos, A., Mazzoni, E., & Reinberg, D. (2023). *Novel Chromatin Insulating Activities Uncovered upon Eliminating Known Insulators in vivo* (p. 2023.04.25.538167). bioRxiv. https://doi.org/10.1101/2023.04.25.538167

Ota, M., Nagafuchi, Y., Hatano, H., Ishigaki, K., Terao, C., Takeshima, Y., Yanaoka, H., Kobayashi, S., Okubo, M., Shirai, H., Sugimori, Y., Maeda, J., Nakano, M., Yamada, S., Yoshida, R., Tsuchiya, H., Tsuchida, Y., Akizuki, S., Yoshifuji, H., … Fujio, K. (2021). Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. *Cell*, *184*(11), 3006-3021.e17. https://doi.org/10.1016/j.cell.2021.03.056

Petrovic, J., Zhou, Y., Fasolino, M., Goldman, N., Schwartz, G. W., Mumbach, M. R., Nguyen, S. C., Rome, K. S., Sela, Y., Zapataro, Z., Blacklow, S. C., Kruhlak, M. J., Shi, J., Aster, J. C., Joyce, E. F., Little, S. C., Vahedi, G., Pear, W. S., & Faryabi, R. B. (2019). Oncogenic Notch Promotes Long-Range Regulatory Interactions within Hyperconnected 3D Cliques. *Molecular Cell*, *73*(6), 1174-1190.e12. https://doi.org/10.1016/j.molcel.2019.01.006

Popay, T. M., & Dixon, J. R. (2022). Coming full circle: On the origin and evolution of the looping model for enhancer-promoter communication. *The Journal of Biological Chemistry*, *298*(8), 102117. https://doi.org/10.1016/j.jbc.2022.102117

Portillo-Ledesma, S., Li, Z., & Schlick, T. (2023). Genome modeling: From chromatin fibers to genes. *Current Opinion in Structural Biology*, *78*, 102506. https://doi.org/10.1016/j.sbi.2022.102506

Qiu, Y., Feng, D., Jiang, W., Zhang, T., Lu, Q., & Zhao, M. (2023). 3D genome organization and epigenetic regulation in autoimmune diseases. *Frontiers in Immunology*, *14*, 1196123.

https://doi.org/10.3389/fimmu.2023.1196123

Rao, S. S. P., 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

Rappaport, N., Nativ, N., Stelzer, G., Twik, M., Guan-Golan, Y., Stein, T. I., Bahir, I., Belinky, F., Morrey, C. P., Safran, M., & Lancet, D. (2013). MalaCards: An integrated compendium for diseases and their annotation. *Database: The Journal of Biological Databases and Curation*, *2013*, bat018. https://doi.org/10.1093/database/bat018

Reiff, S. B., Schroeder, A. J., Kırlı, K., Cosolo, A., Bakker, C., Mercado, L., Lee, S., Veit, A. D., Balashov, A. K., Vitzthum, C., Ronchetti, W., Pitman, K. M., Johnson, J., Ehmsen, S. R., Kerpedjiev, P., Abdennur, N., Imakaev, M., Öztürk, S. U., Çamoğlu, U., … Park, P. J. (2022). The 4D Nucleome Data Portal as a resource for searching and visualizing curated nucleomics data. *Nature Communications*, *13*(1), 2365. https://doi.org/10.1038/s41467-022-29697-4

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*(1), 256. https://doi.org/10.1186/s13059-020-02167-0

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., … Sherry, S. T. (2022). Database resources of the national center for biotechnology information. *Nucleic Acids Research*, *50*(D1), D20–D26. https://doi.org/10.1093/nar/gkab1112

Schaeffer, M., & Nollmann, M. (2023). Contributions of 3D chromatin structure to cell-type-specific gene regulation. *Current Opinion in Genetics & Development*, *79*, 102032.

https://doi.org/10.1016/j.gde.2023.102032

Schmiedel, B. J., Rocha, J., Gonzalez-Colin, C., Bhattacharyya, S., Madrigal, A., Ottensmeier, C. H., Ay, F., Chandra, V., & Vijayanand, P. (2021). COVID-19 genetic risk variants are associated with expression of multiple genes in diverse immune cell types. *Nature Communications*, *12*(1), 6760. https://doi.org/10.1038/s41467-021-26888-3

Schmiedel, B. J., Singh, D., Madrigal, A., Valdovino-Gonzalez, A. G., White, B. M., Zapardiel-Gonzalo, J., Ha, B., Altay, G., Greenbaum, J. A., McVicker, G., Seumois, G., Rao, A., Kronenberg, M., Peters, B., & Vijayanand, P. (2018). Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell*, *175*(6), 1701-1715.e16. https://doi.org/10.1016/j.cell.2018.10.022

Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C.-J., Vert, J.-P., Heard, E., Dekker, J., & Barillot, E. (2015). HiC-Pro: An optimized and flexible pipeline for Hi-C data processing. *Genome Biology*, *16*, 259. https://doi.org/10.1186/s13059-015-0831-x

Shi, C., Ray-Jones, H., Ding, J., Duffus, K., Fu, Y., Gaddi, V. P., Gough, O., Hankinson, J., Martin, P., McGovern, A., Yarwood, A., Gaffney, P., Eyre, S., Rattray, M., Warren, R. B., & Orozco, G. (2021). Chromatin Looping Links Target Genes with Genetic Risk Loci for Dermatological Traits. *The Journal of Investigative Dermatology*, *141*(8), 1975–1984. https://doi.org/10.1016/j.jid.2021.01.015

Sollis, E., Mosaku, A., Abid, A., Buniello, A., Cerezo, M., Gil, L., Groza, T., Güneş, O., Hall, P., Hayhurst, J., Ibrahim, A., Ji, Y., John, S., Lewis, E., MacArthur, J. A. L., McMahon, A., Osumi-Sutherland, D., Panoutsopoulou, K., Pendlington, Z., … Harris, L. W. (2023). The NHGRI-EBI GWAS Catalog: Knowledgebase and deposition resource. *Nucleic Acids Research*, *51*(D1), D977–D985. https://doi.org/10.1093/nar/gkac1010

Taylor, L. J., Abbas, A., & Bushman, F. D. (2020). grabseqs: Simple downloading of reads and metadata from multiple next-generation sequencing data repositories. *Bioinformatics (Oxford, England)*, *36*(11), 3607–3609. https://doi.org/10.1093/bioinformatics/btaa167

Waldmann, T. A. (2015). The shared and contrasting roles of IL2 and IL15 in the life and death of normal and neoplastic lymphocytes: Implications for cancer therapy. *Cancer Immunology Research*, *3*(3), 219–227. https://doi.org/10.1158/2326-6066.CIR-15-0009

Wang, J., Huang, D., Zhou, Y., Yao, H., Liu, H., Zhai, S., Wu, C., Zheng, Z., Zhao, K., Wang, Z., Yi, X., Zhang, S., Liu, X., Liu, Z., Chen, K., Yu, Y., Sham, P. C., & Li, M. J. (2020). CAUSALdb: A database for disease/trait causal variants identified using summary statistics of genome-wide association studies. *Nucleic Acids Research*, *48*(D1), D807–D816. https://doi.org/10.1093/nar/gkz1026

Wang, J., & Nakato, R. (2023). CohesinDB: A comprehensive database for decoding cohesin-related epigenomes, 3D genomes and transcriptomes in human cells. *Nucleic Acids Research*, *51*(D1), D70–D79. https://doi.org/10.1093/nar/gkac795

Wang, S., Luo, Z., Liu, W., Hu, T., Zhao, Z., Rosenfeld, M. G., & Song, X. (2023). The 3D genome and its impacts on human health and disease. *Life Medicine*, *2*(2), lnad012. https://doi.org/10.1093/lifemedi/lnad012

Wang, W., Song, F., Feng, X., Chu, X., Dai, H., Tian, J., Fang, X., Song, F., Liu, B., Li, L., Li, X., Zhao, Y., Zheng, H., & Chen, K. (2021). Functional Interrogation of Enhancer Connectome Prioritizes Candidate Target Genes at Ovarian Cancer Susceptibility Loci. *Frontiers in Genetics*, *12*, 646179. https://doi.org/10.3389/fgene.2021.646179

Wei, X., Xiang, Y., Peters, D. T., Marius, C., Sun, T., Shan, R., Ou, J., Lin, X., Yue, F., Li, W., Southerland, K. W., & Diao, Y. (2022). HiCAR is a robust and sensitive method to analyze open-chromatin-associated genome organization. *Molecular Cell*, *82*(6), 1225-1238.e6. https://doi.org/10.1016/j.molcel.2022.01.023

Xin, J., Mark, A., Afrasiabi, C., Tsueng, G., Juchler, M., Gopal, N., Stupp, G. S., Putman, T. E., Ainscough, B. J., Griffith, O. L., Torkamani, A., Whetzel, P. L., Mungall, C. J., Mooney, S. D., Su, A. I., & Wu, C. (2016). High-performance web services for querying gene and variant annotation. *Genome Biology*, *17*(1), 91. https://doi.org/10.1186/s13059-016-

0953-9

Xu, H., Zhang, S., Yi, X., Plewczynski, D., & Li, M. J. (2020). Exploring 3D chromatin contacts in gene regulation: The evolution of approaches for the identification of functional enhancer-promoter interaction. *Computational and Structural Biotechnology Journal*, *18*, 558–570. https://doi.org/10.1016/j.csbj.2020.02.013

Yu, M., Juric, I., Abnousi, A., Hu, M., & Ren, B. (2021). Proximity Ligation-Assisted ChIP-Seq (PLAC-Seq). *Methods in Molecular Biology (Clifton, N.J.)*, *2351*, 181–199. https://doi.org/10.1007/978-1-0716-1597-3_10

Zeng, W., Liu, Q., Yin, Q., Jiang, R., & Wong, W. H. (2023). HiChIPdb: A comprehensive database of HiChIP regulatory interactions. *Nucleic Acids Research*, *51*(D1), D159–D166. https://doi.org/10.1093/nar/gkac859

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., & Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9), R137. https://doi.org/10.1186/gb-2008-9-9-r137

Zhong, W., Liu, W., Chen, J., Sun, Q., Hu, M., & Li, Y. (2022). Understanding the function of regulatory DNA interactions in the interpretation of non-coding GWAS variants. *Frontiers in Cell and Developmental Biology*, *10*, 957292. https://doi.org/10.3389/fcell.2022.957292

Zhou, Q., Cheng, S., Zheng, S., Wang, Z., Guan, P., Zhu, Z., Huang, X., Zhou, C., & Li, G. (2023). ChromLoops: A comprehensive database for specific protein-mediated chromatin loops in diverse organisms. *Nucleic Acids Research*, *51*(D1), D57–D69. https://doi.org/10.1093/nar/gkac893

Zitnik, M., Sosič, R., & Leskovec, J. (2018). Prioritizing network communities. *Nature Communications*, *9*(1), 2544. https://doi.org/10.1038/s41467-018-04948-5

## FIGURE CAPTIONS

**Figure 1.** Data overview and high-level summary of the Loop Catalog. **A)** Breakdown of the number of HiChIP samples generated from 2016 to 2024. The top panel shows the number of studies broken down by human (blue), mouse (teal), or both (orange). Bottom panel shows the cumulative trend. **B)** Breakdown of samples by target protein or histone modification and by organism. **C)** Schema for the development of the Loop Catalog starting from raw sequencing files to processing (top left), database storage (bottom) and web accessible analyses (top right). **D-E)** Number of peak calls (left) and FitHiChIP loop calls at three different resolutions (right) for HiChIP samples with ChIP-seq data for **D)** human and **E)** mouse samples.

**Figure 2.** Layout of the Loop Catalog portal. **A)** Screenshot of the main entry page to the Loop Catalog. **B)** Main data page which includes an embedding of the WashU Epigenome Browser followed by a table of HiChIP samples with various metadata fields. **C)** Screenshot of a HiChIP sample page with statistics on the loop calls and the enhancer-promoter network visualization with an associated table on ranking of identified network communities. Subcommunities are color coded with nodes belonging to either enhancers (circles), promoters (squares) or other (triangles). **D)** Screenshot of the SGL entry page where disease, locus/gene and HiChIP samples are selected. **E)** Screenshot of the SGL analysis page which includes an embedded browser and a table with SGL metadata including buttons for navigation.

**Figure 3.** SGL analysis overview and results. **A)** Schema of the SGL analysis using fine-mapped SNPs from CAUSALdb, 156 Loop Catalog immune-related samples, and TSS coordinates. **B)** Summary of results across all 4 diseases including the total number of GWAS hits (blue), SNPs found in a SGL (orange), genes found in SGL (green) and total SGLs (red). **C)** Distribution of SNP counts with respect to each SGL gene (left) and the distribution of gene counts with respect to each SGL SNP (right) for T1D. **D)** Evaluating the number of SGL genes which belong to a consensus list of T1D genes (green) and unique (orange). **E)** Example of an SGL between rs61839660 (red) and the genes *IL15RA* (red arc) and *RBM17* (blue arc). Tracks contain H3K27ac HiChIP tracks for naïve CD4 T-cell, naïve CD8 T-cell, naïve B cell, Natural Killer, monocytes,

nonclassical monocytes derived from the Chanra et al 2021 and Schmiedel et al 2021 samples that were merged across all donors.

**Figure 4.** Motif and paired-motif analysis of loop anchors. **A)** Schematic of the 1D and 2D (paired) motif enrichment analysis. **B)** Bubble plot for the union of the top 15 motifs from each sample set (20 total motifs). The E-value is represented on a range from 0 (gray) to 2000 (magenta) and the log2(enrichment ratio) is represented by a circle radius from 1 to 7. **C-D)** Motif plot of ZNF460 and CTCF with the significance and enrichment ratio values as reported for each sample set. **D)** Q-Q plot testing p-values for Naive CD4 T-cell 1829-RH-1 sample. **E)** Heatmap of significant motif pairs (center) where rows and columns represent motifs on opposite anchors and each cell represents the proportion of samples where the given motif pair is significant. The distributions of a given motif across the whole genome and within the top 25 motif pairs are represented on the top and right, respectively.

## TABLES

**Table 1.** Comparison of Loop Catalog with other HiChIP databases. Comparison of HiChIP processing methods and website features of Loop Catalog, ChromLoops, HiChIPdb, and Cohesin-DB. Categories include specification and quantification of data types, implementation of loop calling (software, configurations, replicates), ease of data download and visualization, including the ability to select and visualize multiple samples simultaneously, and embedded biological analysis modules. *annotation of genes, SNPs, E/Ps, silencers, circRNA, TWAS, chromosome open access data, alternative splicing, transcription factors. ** annotation of genes, SNPs. *** annotation of genes

## SUPPLEMENTARY FIGURE CAPTIONS

**Supp Figure 1.** Schematic of the HiChIP and ChIP-seq data processing pipeline we developed for the Loop Catalog. Raw sequencing reads are downloaded from NCBI GEO and dbGaP and are aligned to the reference genome (hg38 or mm10). Loops are called for HiChIP (both unmerged and merged biological replicates, as indicated by the shaded circles) using HiCCUPS and FitHiChIP at the 5kb, 10kb, and 25kb resolutions. Peaks derived from both ChIP-seq and HiChIP are used for FitHiChIP loop calling. High-confidence sample sets of top H3K27ac HiChIP samples are curated from the final set of loop calls. Red borders indicate usage in downstream database application analyses. Yellow stars indicate that the data type is available for download from the Loop Catalog web platform.

**Supp Figure 2.** Recall analysis of ChIP-seq peaks using peaks called from HiChIP reads. **A)** Distribution of recall rates of ChIP-seq peaks by peaks inferred from HiChIP by FitHiChIP in hg38. **B)** Scatter plot of recall rate versus the number of peaks in hg38. Samples with zero percent recall or with zero peak calls derived from HiChIP are not included. **C)** Scatter plot of the recall rate versus the total peak span in hg38. Samples with zero peak calls derived from HiChIP are not included. **D-F)** Represent the same analysis as **A-C** but using mouse HiChIP samples.

**Supp Figure 3.** Reproducibility analysis for HiChIP biological replicates. Stratum adjusted correlation coefficient scores (SCCs) for all pairwise combinations of HiChIP biological replicates for 291 HiChIP samples with at least two biological replicates were generated using hicreppy. SCCs for pairwise combinations of HiChIP biological replicates are displayed for the 29 HiChIP samples with a minimum SCC below 0.90. SCCs for pairwise combinations of replicates for samples with more than 2 replicates are connected with vertical lines. Samples passed the SCC threshold if the SCC of all pairwise combinations of replicates was greater than 0.80 (n = 282 samples, a subset of 20 passing samples shown in blue). Samples which possessed at least one replicate combination with a SCC less than 0.80 did not pass the threshold (n = 9 samples, all shown in red).

**Supp Figure 4:** Number of FH loop calls by protein pulldown. Distributions of number of FH loop calls by protein pulldown are displayed for **A)** human and **B)** mouse samples with >100 loops. Proteins represented by >10 samples are individually displayed while all others are grouped into the “Other” category. These include proteins such as RNA-Pol-II, GATA1, STAG1, STAG2, RAD21, etc.

**Supp Figure 5.** Assignment of QC Flags to HIChIP Samples. 13 flags were assigned to all samples. For peak calls, a flag of “Poor” was assigned if the sample possessed <=1000 peaks, “Warning” to >1000 and <5000 peaks, and “Good” to >=5000 peaks. Each configuration of FitHiChIP loop calls were each similarly assigned “Poor” (number of Stringent loops > number Loose loops at that resolution, or <=100 loops), “Warning” (>100 and <1000 loops), or “Good” (>= 1000 loops). A final flag was assigned based on the logic in Supplemental Table 7. The distribution of each flag type is displayed for **A)** human FC loops, unmerged and merged, **B)** human FH loops, unmerged and merged, **C)** mouse FC loops, unmerged and merged, and **D)** mouse FH loops, unmerged and merged. Distributions of final flags are highlighted with black borders.

**Supp Figure 6.** Schema of Bootstrap Analysis for Motif Pairs. Loops are depicted with an arc and horizontal lines are used to denote their anchors. Each loop has a distinct color to emphasize the true loop composition and overlapping motifs are depicted with horizontal lines and a separate color for each (top right). Observed motif pairs are tabulated (bottom left). To simulate a new set of loops, anchors are shuffled thereby bringing together new combinations of motifs (top middle). Motif pairs from these simulations are tabulated across 100,000 simulations (top right) to then build a null distribution (bottom right). The observed counts for a given motif are then evaluated against their respective null distribution (bottom).

**Supp Figure 7.** Schema for the Loop Catalog database tables. Lines connect tables based on relationships with children tables represented by a forked edge and asterisk.

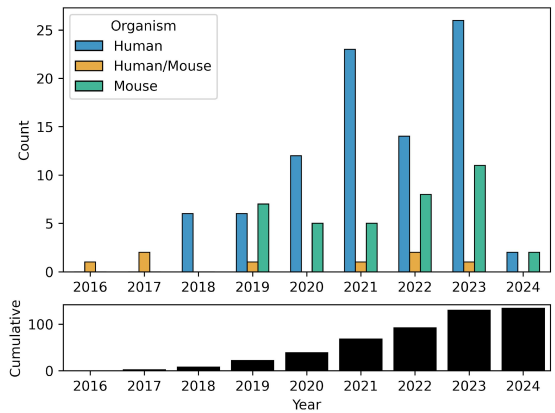
**Supp Figure 8.** Summary of SGLs derived from HiChIP and eQTL intersections. **A)** Breakdown by cell type for the total number of eQTLs (blue) and SNPs (orange) within the original eQTL dataset. **B)** Breakdown by cell type for the total number of genes from the eQTL study (green) followed by SGL summaries for unique SGLs (red), SNPs (purple) and genes (brown).

**Supp Figure 9.** Example analysis for CD4 T cells using SGLs derived from their corresponding eQTLs. Depicted is the IKZF3/ORMDL3 locus with the top gray track containing SNPs derived from eQTLs followed by H3K27ac HiChIP loops for six CD4 T cell samples.

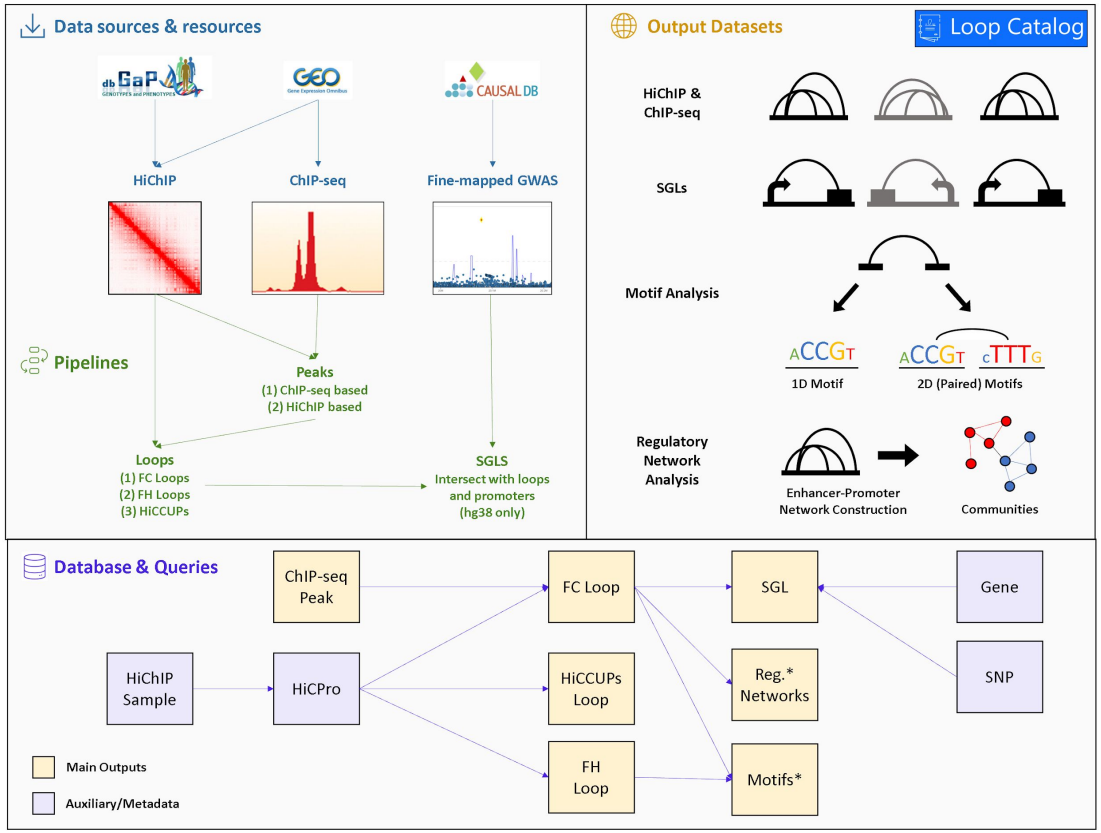
| General Feature | Specific Feature | Loop Catalog | ChromLoops (Zhou et al., 2022) | HiChIPdb (Zeng et al., 2022) | Cohesin-DB (Wang et al., 2022) |
|---|---|---|---|---|---|
| **Data Types** | **Organism** | Homo sapiens, Mus musculus | Homo sapiens, Mus musculus, 11 others | Homo sapiens | Homo sapiens |
| | **Reference Genome** | hg38, mm10 | hg38, mm10, 11 others | hg19 | hg38 |
| **HiChIP Processing** | **Total HiChIP Samples** | 1319 total (1031 distinct) | 772 | 200 | 42 |
| | **Loop Calling** | FitHiChIP, HiCCUPS | ChIA-PET Tool (V3) | FitHiChIP, hichipper | HiCCUPS |
| | **Loop Resolutions** | 5kb, 10kb, 25kb | variable | 1kb, 5kb, 10kb, 50kb, variable | 5kb, 10kb, 25kb |
| | **Peak Type Used for Loop Calling** | HiChIP-Inferred, ChIP-seq when available | HiChIP-Inferred | HiChIP-Inferred | N/A |
| | **Replicate Handling** | Technical/Biological Reps Merged,Multiple D | Biological Replicates Merged | Technical/Biological Reps Merged | Technical Reps Merged |
| | **Pipeline Code Released** | TRUE | TRUE | FALSE | TRUE |
| **Data Download** | **Loop Calls** | TRUE | TRUE | TRUE | TRUE |
| | **Peak Calls Used for Loop Calling** | TRUE | FALSE | FALSE | FALSE |
| | **Browser Track Files** | TRUE | FALSE | FALSE | FALSE |
| **Data Visualization** | **Embedded Visualization** | Epigenome Browser (WashU) | Epigenome Browser (WashU) | IGV | Epigenome Browser (WashU) |
| | **Multi-Sample Selection** | TRUE | FALSE | TRUE | TRUE |
| | **Multi-Sample Visualization** | TRUE | FALSE | FALSE | TRUE |
| **Data Analysis** | **Embedded Tools** | Community Structure Analysis,Loop Motif Pa | Functional Anchor Annotation*, Cancer High-Frequency Loops, Species-Specific High Frequency Loops | Functional Anchor Annotation** | Functional Anchor Annotation***, Loop discovery given genome regions, Prediction of regulatory sites and target genes |
| **Additional Data** | **Hi-C and Other ConformationCapture D** | High Resolution Hi-C (n=44) | PLAC-seq (n=89)<br>ChIAPET (n=789) | N/A | Hi-C (n=385)<br>ChIAPET (n=119) |

**Comparison of Loop Catalog with other HiChIP databases.** Comparison of HiChIP processing methods and website features of Loop Catalog, ChromLoops, HiChIPdb, and Cohesin-DB. Catagories include specification and quantification of data types, implementation of loop calling (software, configurations, replicates), ease of data download and visulization, including the ability to select and visualize multiple samples simultaneously, and embedded biological analysis modules.
* annotation of genes, SNPs, E/Ps, silencers, circRNA, TWAS, chromosome open access data, alternative splicing, transcription factors
** annotation of genes, SNPs
*** annotaton of genes

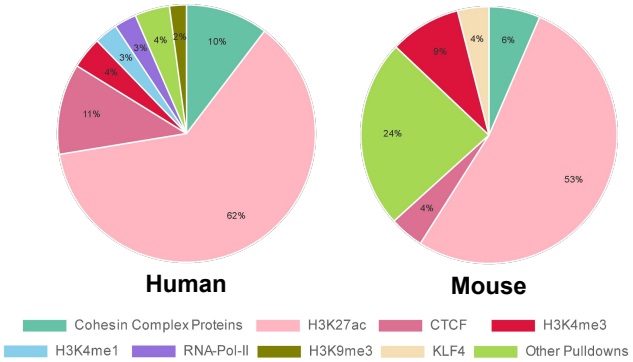
## A HiChIP Studies from 2016 – Oct. 2023




## B Backend to Frontend Development of the Loop Catalog

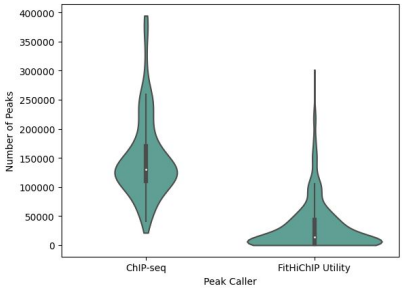



## C Characterization of HiChIP Pulldowns

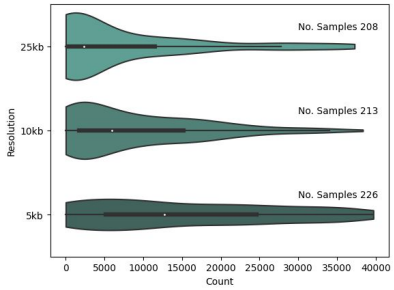



## D

### Peaks For hg38 Samples

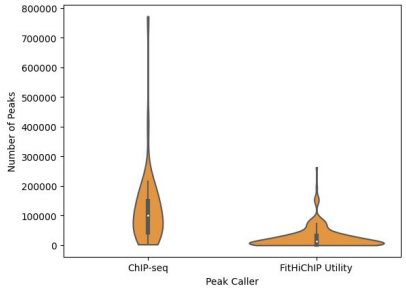



### Loops For hg38 Samples using ChIP-seq Peaks

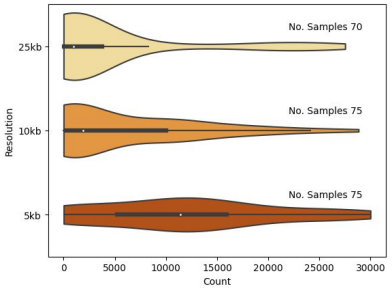



## E

### Peaks For mm10 Samples

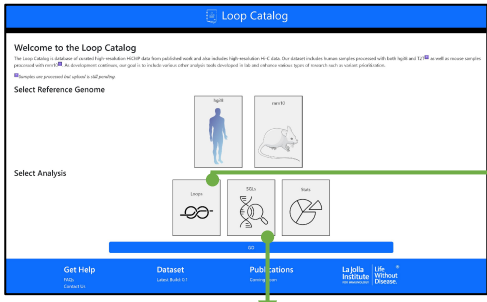



### Loops For mm10 Samples using ChIP-seq Peaks



**A**

## Landing Page

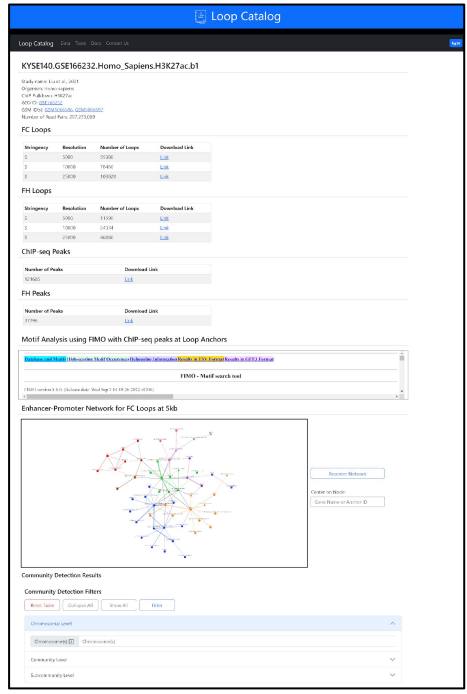



**B**

## Main HiChIP Loop Page

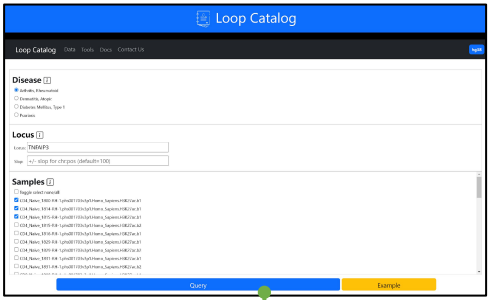
**C**

## HiChIP Sample Page



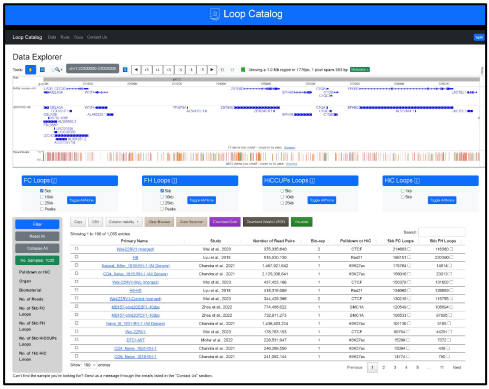

**D**

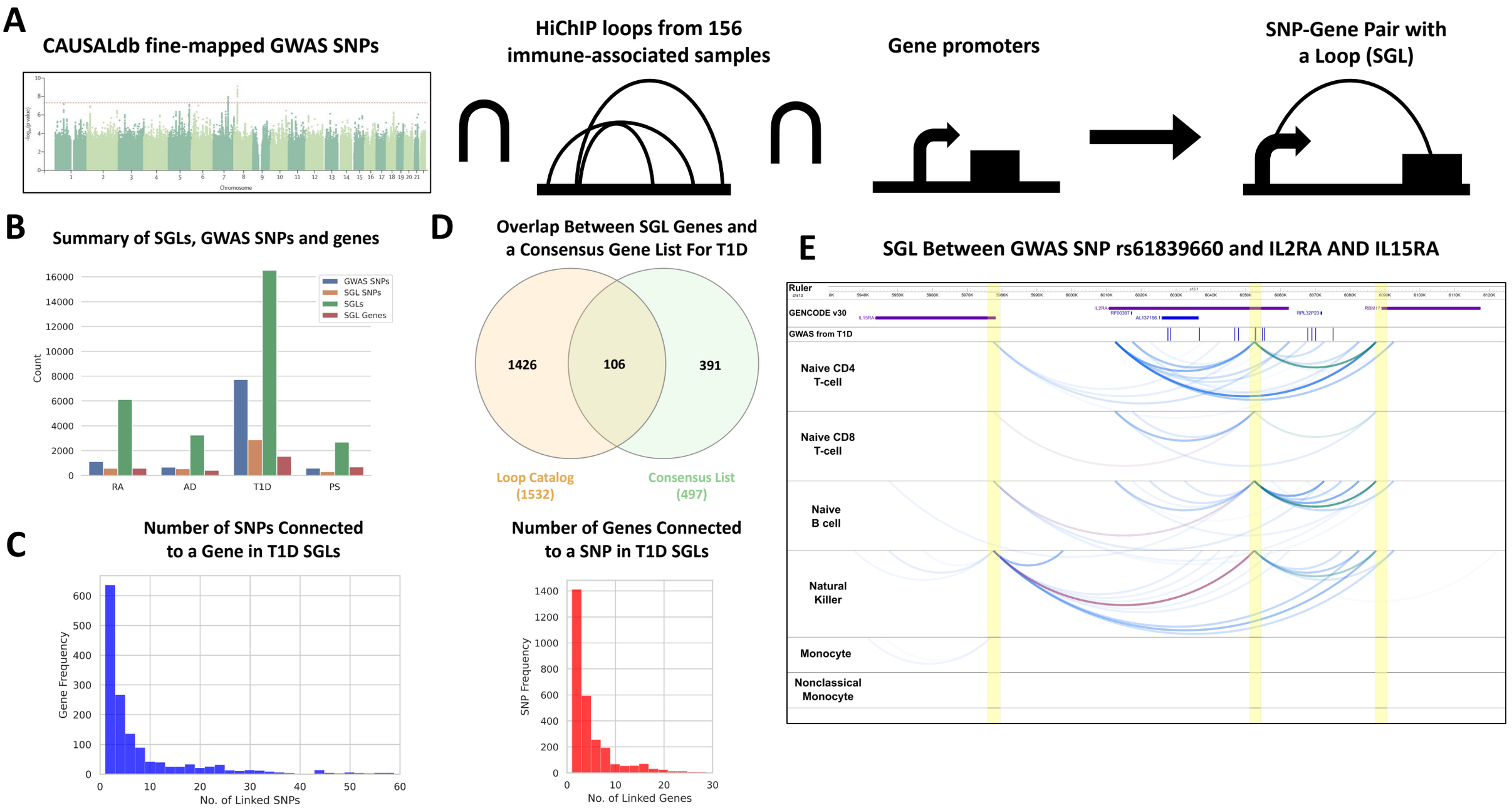
## SGL Query Page



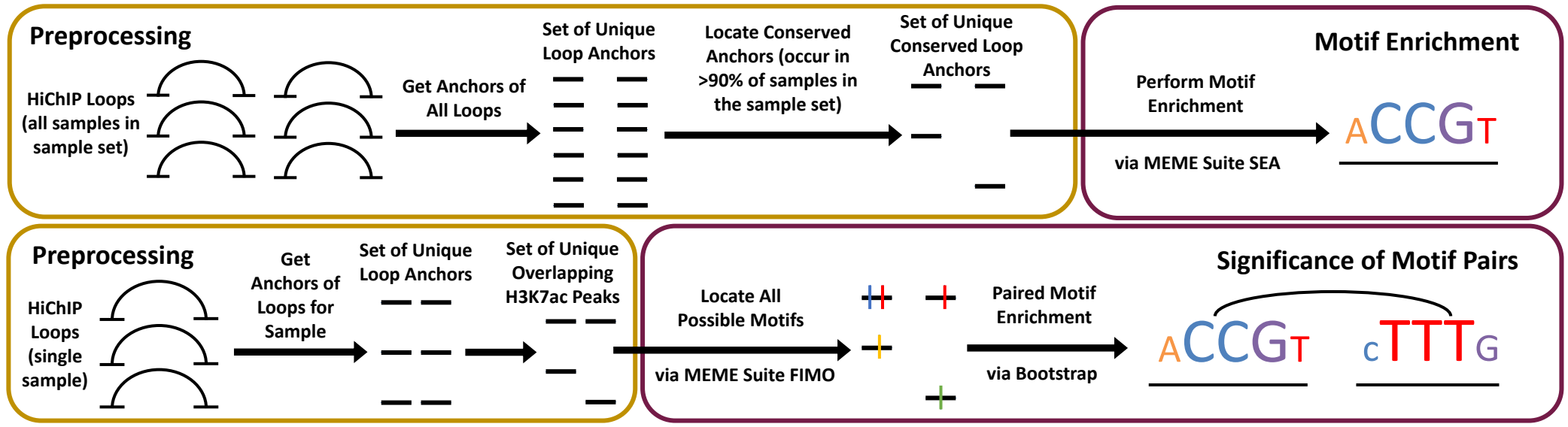

**E**

## SGL Visualizations Page

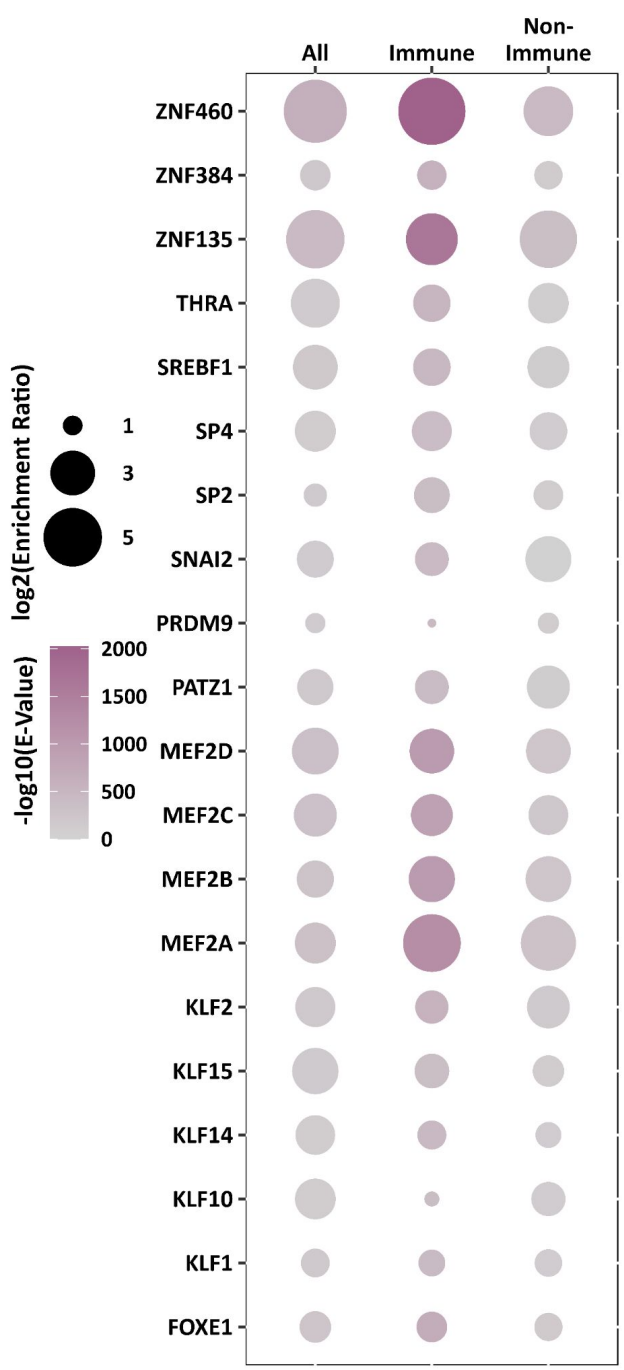



A
CAUSALdb fine-mapped GWAS SNPs
Chromosome
HiChIP loops from 156 immune-associated samples
Gene promoters
SNP-Gene Pair with a Loop (SGL)
B
Summary of SGLs, GWAS SNPs and genes
GWAS SNPs
SGL SNPs
SGLs
SGL Genes
Count
RA
AD
T1D
PS
D
Overlap Between SGL Genes and a Consensus Gene List For T1D
1426
106
391
Loop Catalog (1532)
Consensus List (497)
E
SGL Between GWAS SNP rs61839660 and IL2RA AND IL15RA
Ruler
GENCODE v30
GWAS from T1D
Naive CD4 T-cell
Naive CD8 T-cell
Naive B cell
Natural Killer
Monocyte
Nonclassical Monocyte
C
Number of SNPs Connected to a Gene in T1D SGLs
Gene Frequency
No. of Linked SNPs
Number of Genes Connected to a SNP in T1D SGLs
SNP Frequency
No. of Linked Genes

## A Pipeline for Motif Enrichment Analysis and Significance of Paired Motifs

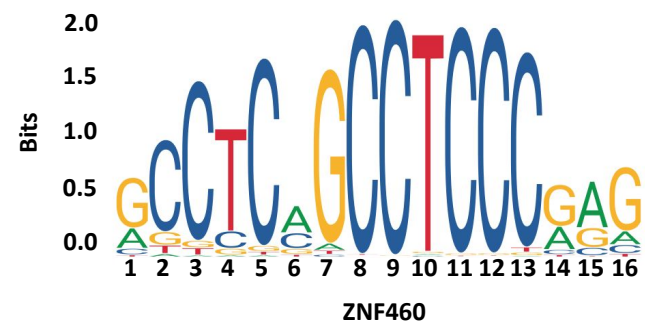



## B SEA Enrichment Ratio for the 1D Motif Analysis



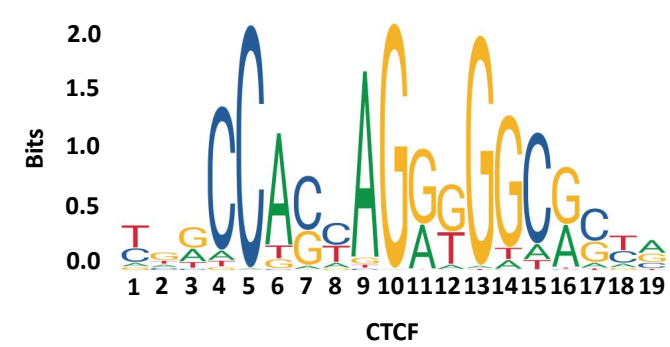

## C SEA Results for ZNF460

| Sample Set | E-Value | Enrichment Ratio |
|---|---|---|
| All | 1.28e-272 | 36.39 |
| Immune | 1.46e-878 | 58.31 |
| Non-Immune | 3.66e-189 | 9.75 |

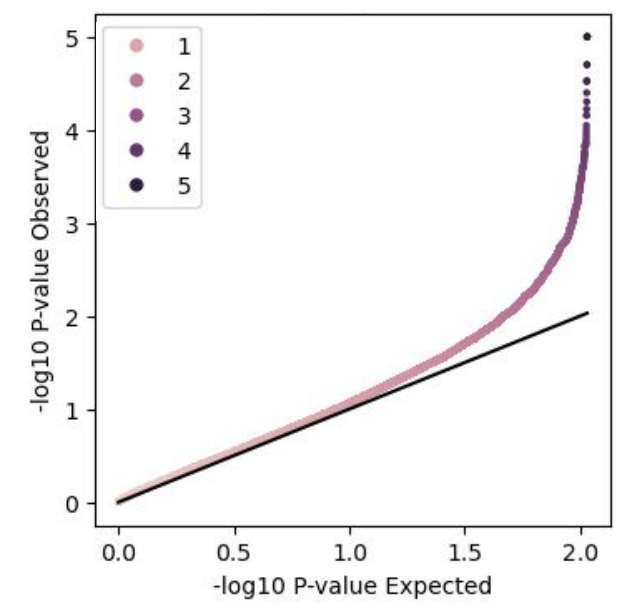



## SEA Results for CTCF

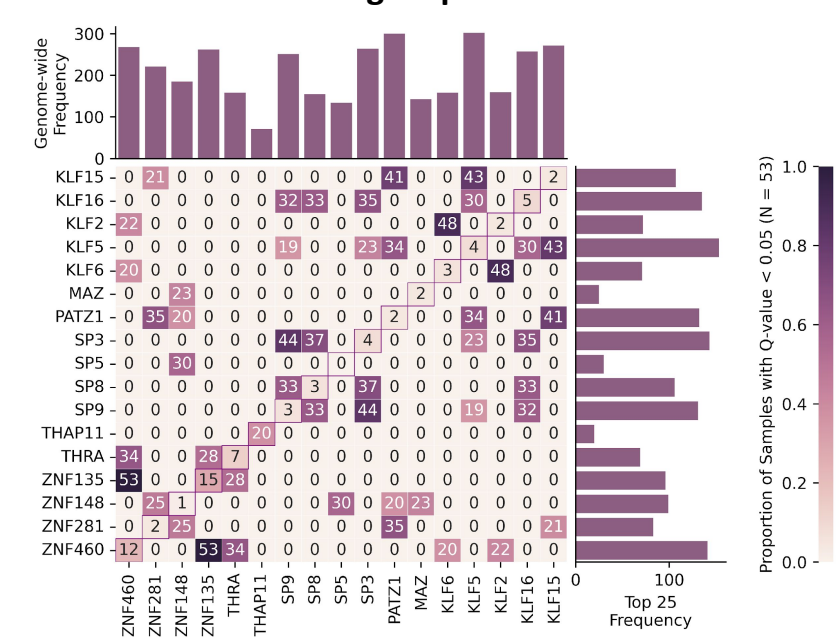
| Sample Set | E-Value | Enrichment Ratio |
|---|---|---|
| All | 2.19e-4 | 1.59 |
| Immune | 1.19e-18 | 2.15 |
| Non-Immune | 6.83e-9 | 5.38 |

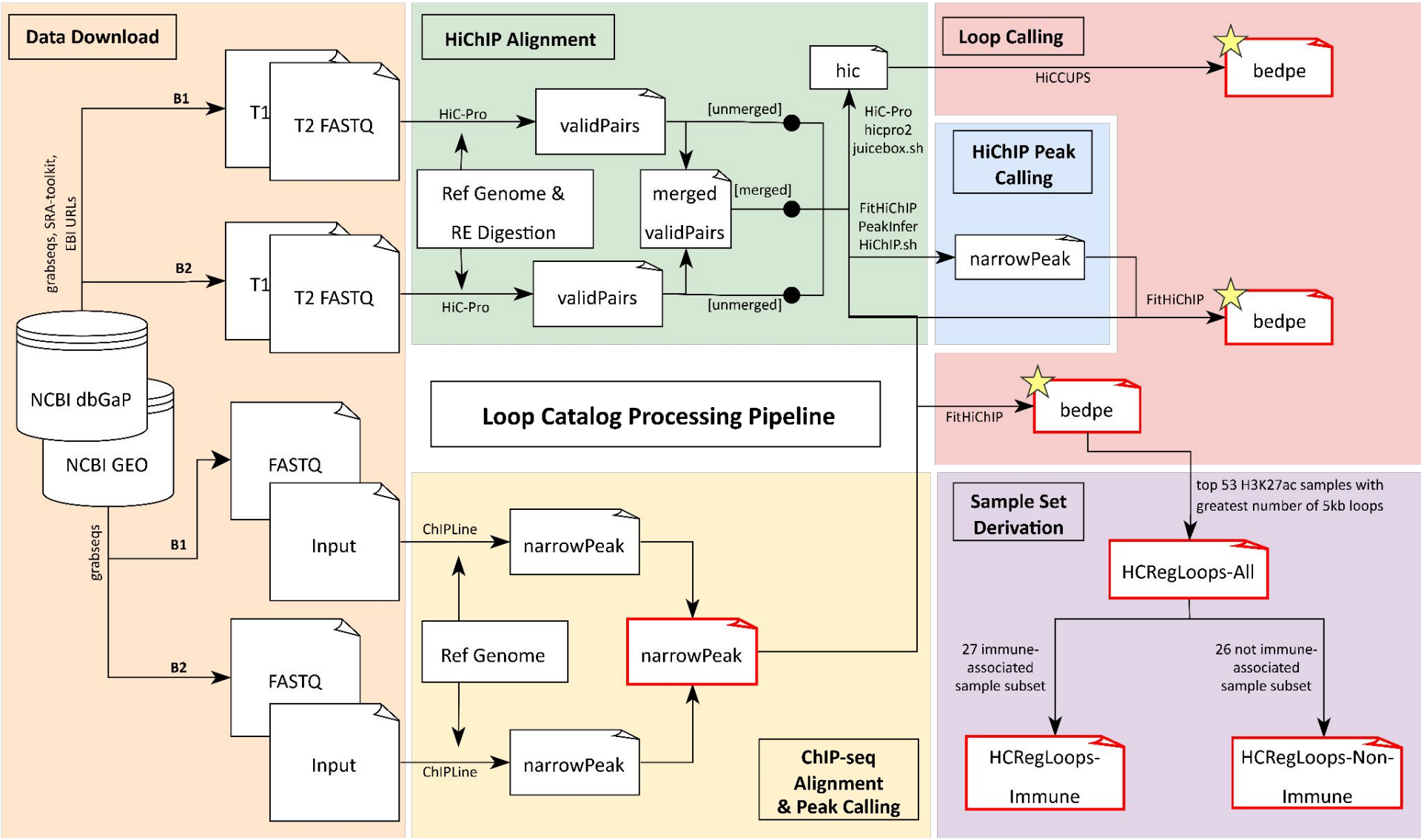


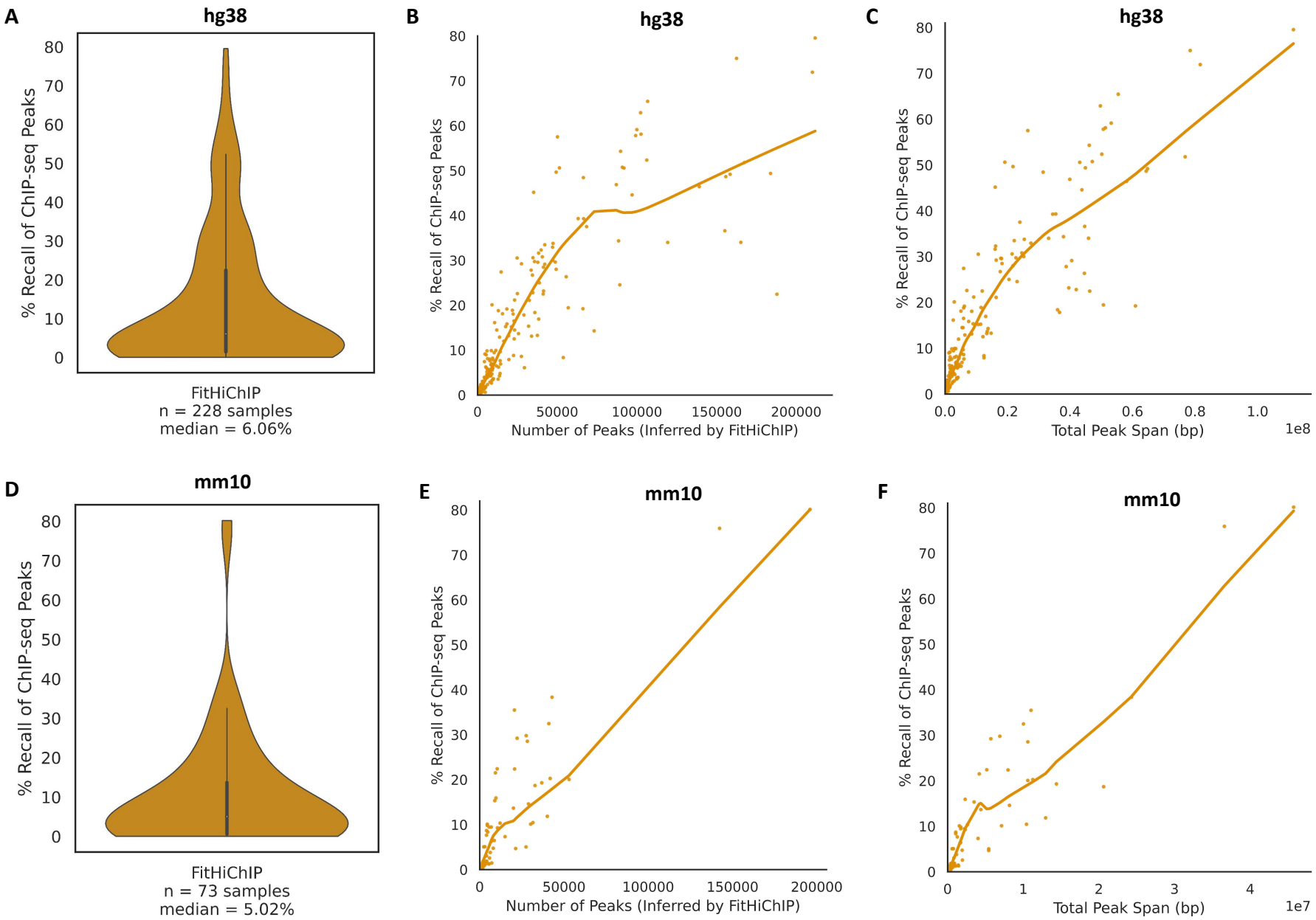

## D QQ Plot for Paired Motif Analysis of Naive CD4 T-cell 1829-RH-1

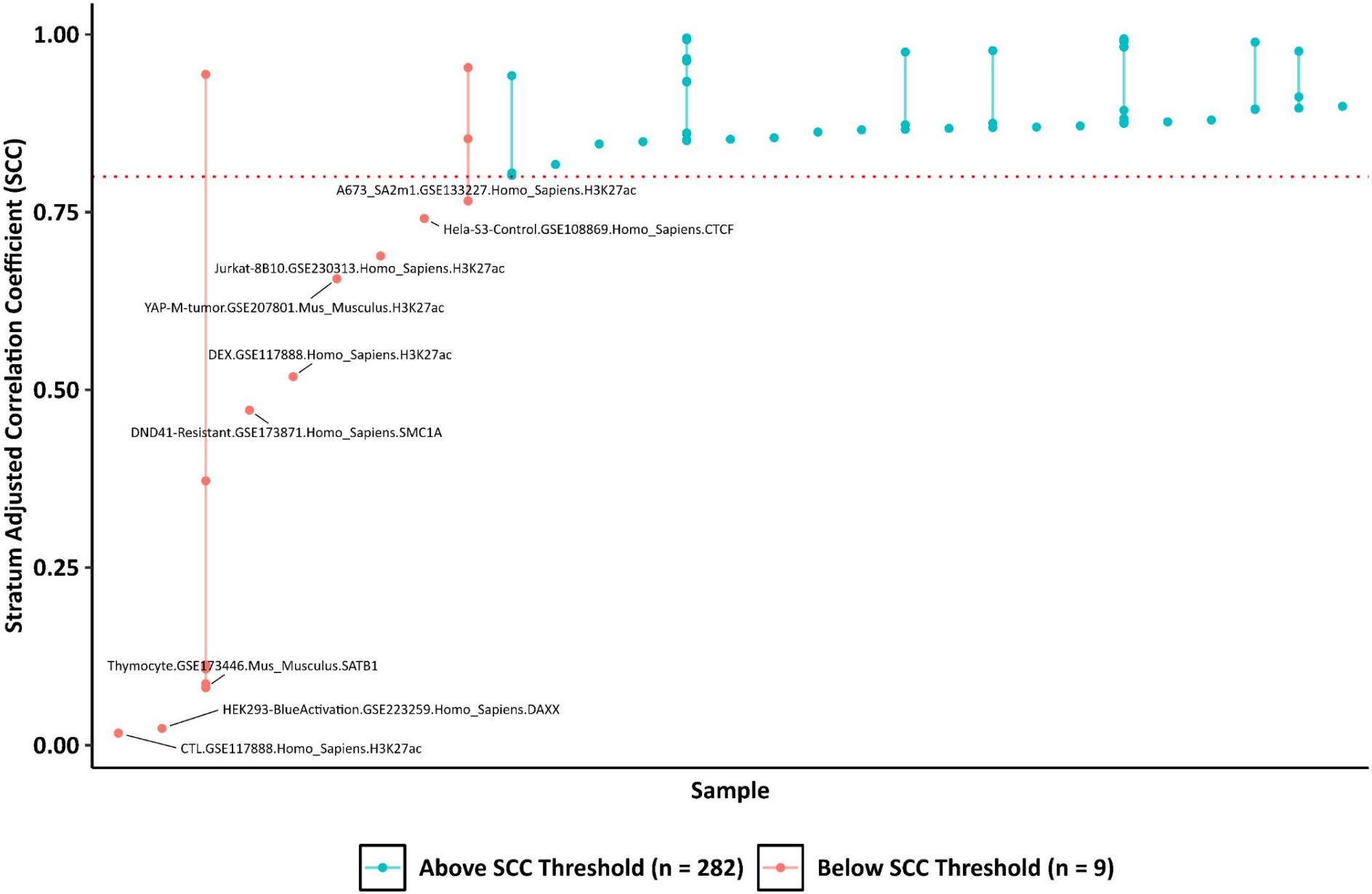


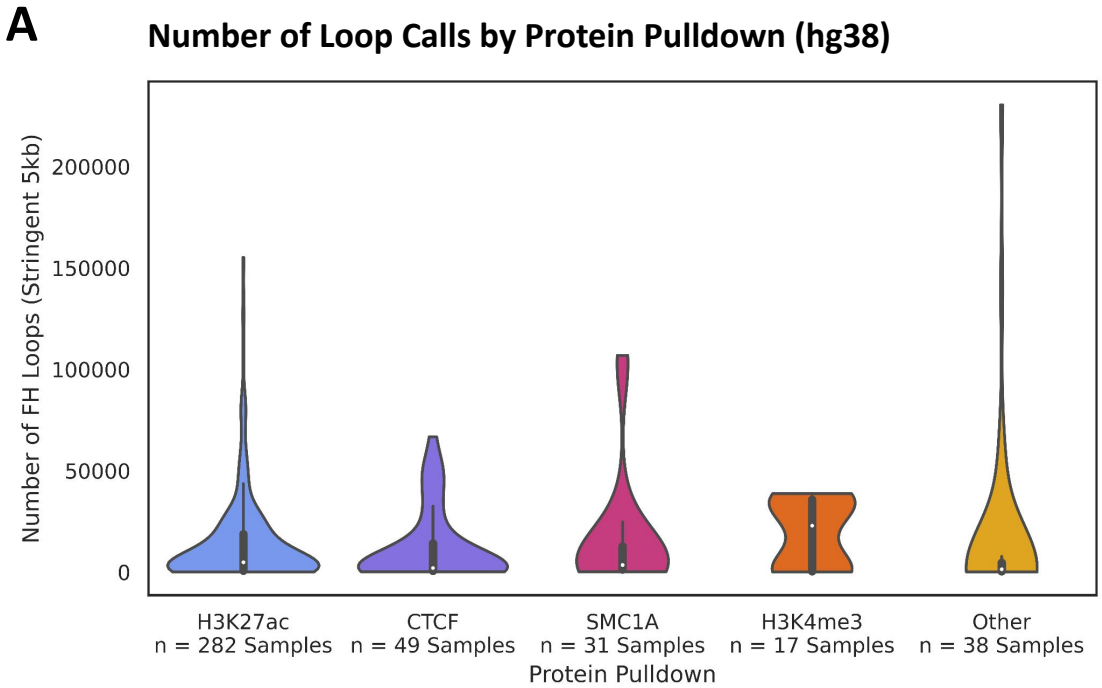

## E Breakdown of Significant Paired Motifs in HCRegLoops-All

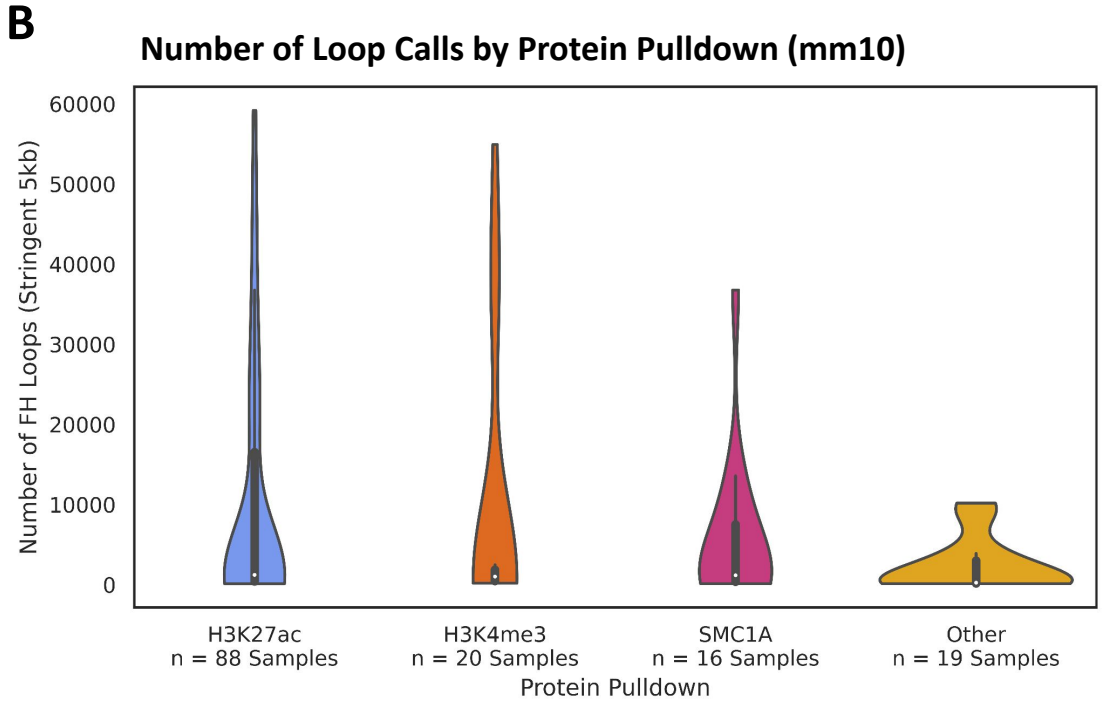


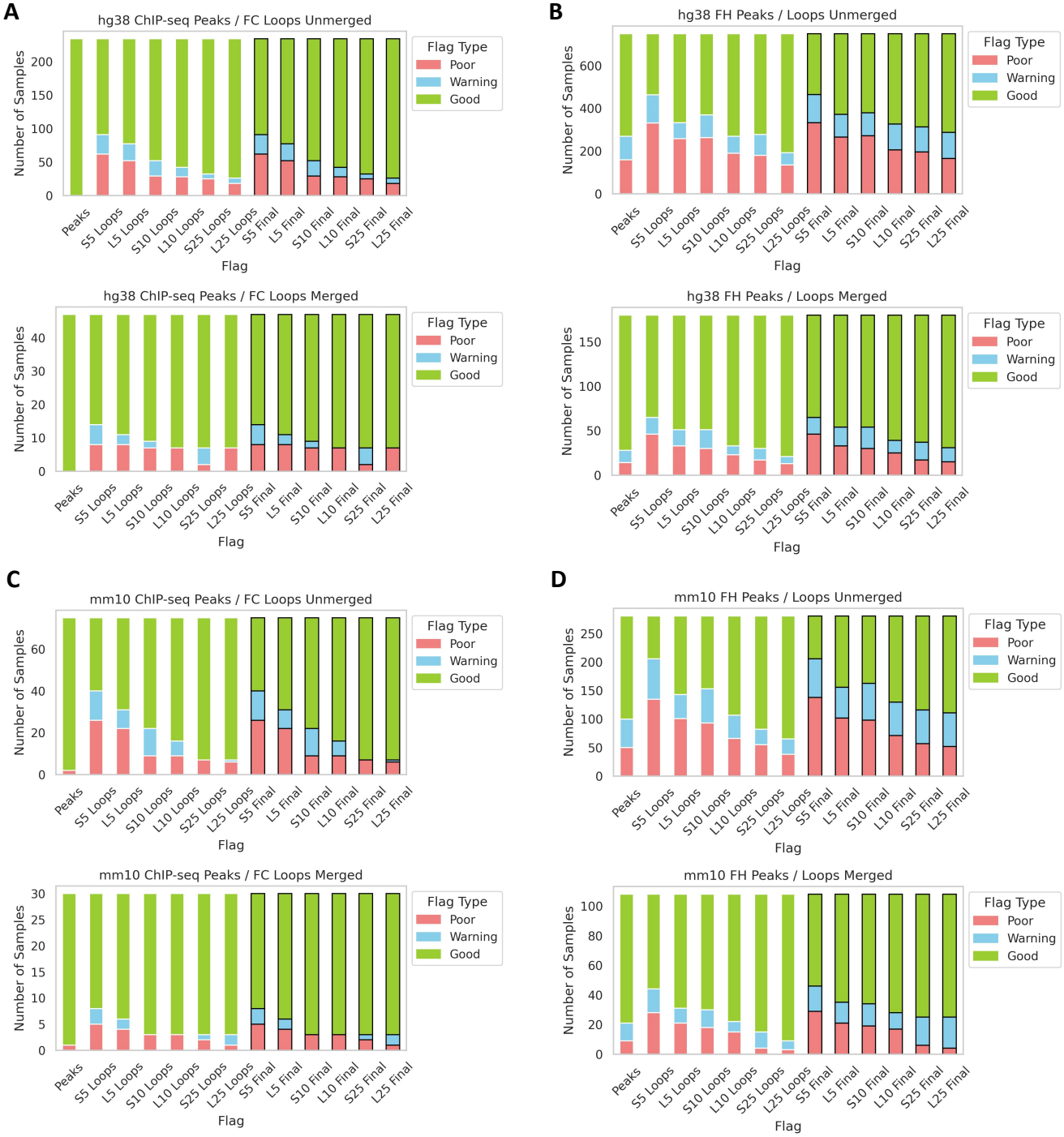

Data Download
HiChIP Alignment
Loop Calling
HiChIP Peak Calling
Loop Catalog Processing Pipeline
ChIP-seq Alignment & Peak Calling
Sample Set Derivation
grabseqs, SRA-toolkit, EBI URLs
grabseqs
B1
B2
T1
T2 FASTQ
NCBI dbGaP
NCBI GEO
FASTQ
Input
HiC-Pro
validPairs
[unmerged]
merged validPairs
[merged]
Ref Genome & RE Digestion
hic
HiC-Pro hicpro2 juicebox.sh
FitHiChIP PeakInfer HiChIP.sh
HiCCUPS
bedpe
narrowPeak
FitHiChIP
ChIPLine
Ref Genome
top 53 H3K27ac samples with greatest number of 5kb loops
HCRegLoops-All
27 immune-associated sample subset
26 not immune-associated sample subset
HCRegLoops-Immune
HCRegLoops-Non-Immune

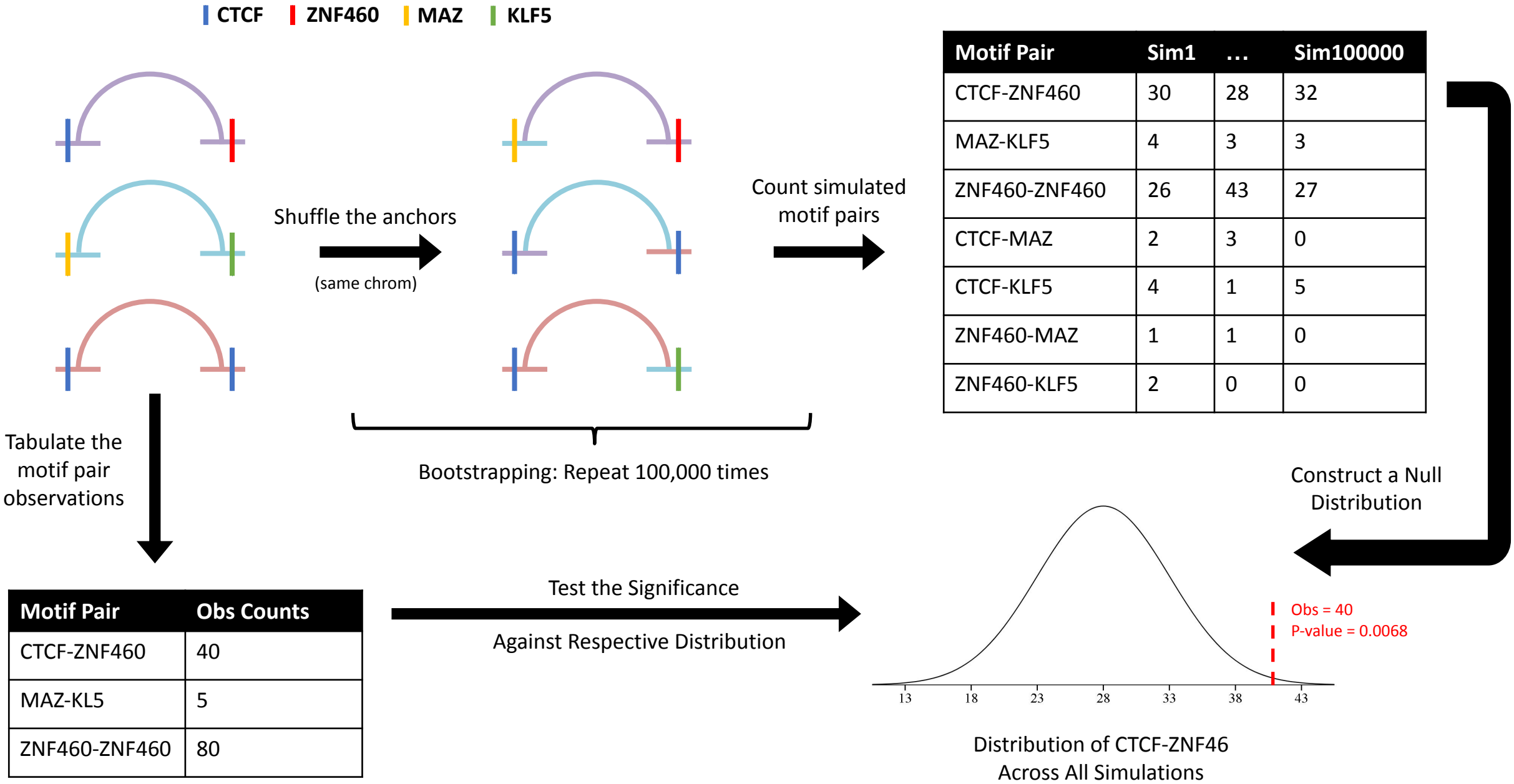

A
hg38
% Recall of ChIP-seq Peaks
FitHiChIP
n = 228 samples
median = 6.06%
B
hg38
% Recall of ChIP-seq Peaks
Number of Peaks (Inferred by FitHiChIP)
C
hg38
% Recall of ChIP-seq Peaks
Total Peak Span (bp)
1e8
D
mm10
% Recall of ChIP-seq Peaks
FitHiChIP
n = 73 samples
median = 5.02%
E
mm10
% Recall of ChIP-seq Peaks
Number of Peaks (Inferred by FitHiChIP)
F
mm10
% Recall of ChIP-seq Peaks
Total Peak Span (bp)
1e7


Stratum Adjusted Correlation Coefficient (SCC)
1.00
0.75
0.50
0.25
0.00
A673_SA2m1.GSE133227.Homo_Sapiens.H3K27ac
Hela-S3-Control.GSE108869.Homo_Sapiens.CTCF
Jurkat-8B10.GSE230313.Homo_Sapiens.H3K27ac
YAP-M-tumor.GSE207801.Mus_Musculus.H3K27ac
DEX.GSE117888.Homo_Sapiens.H3K27ac
DND41-Resistant.GSE173871.Homo_Sapiens.SMC1A
Thymocyte.GSE173446.Mus_Musculus.SATB1
HEK293-BlueActivation.GSE223259.Homo_Sapiens.DAXX
CTL.GSE117888.Homo_Sapiens.H3K27ac
Sample
Above SCC Threshold (n = 282)
Below SCC Threshold (n = 9)


A
Number of Loop Calls by Protein Pulldown (hg38)
Number of FH Loops (Stringent 5kb)
200000
150000
100000
50000
0
H3K27ac
n = 282 Samples
CTCF
n = 49 Samples
SMC1A
n = 31 Samples
H3K4me3
n = 17 Samples
Other
n = 38 Samples
Protein Pulldown

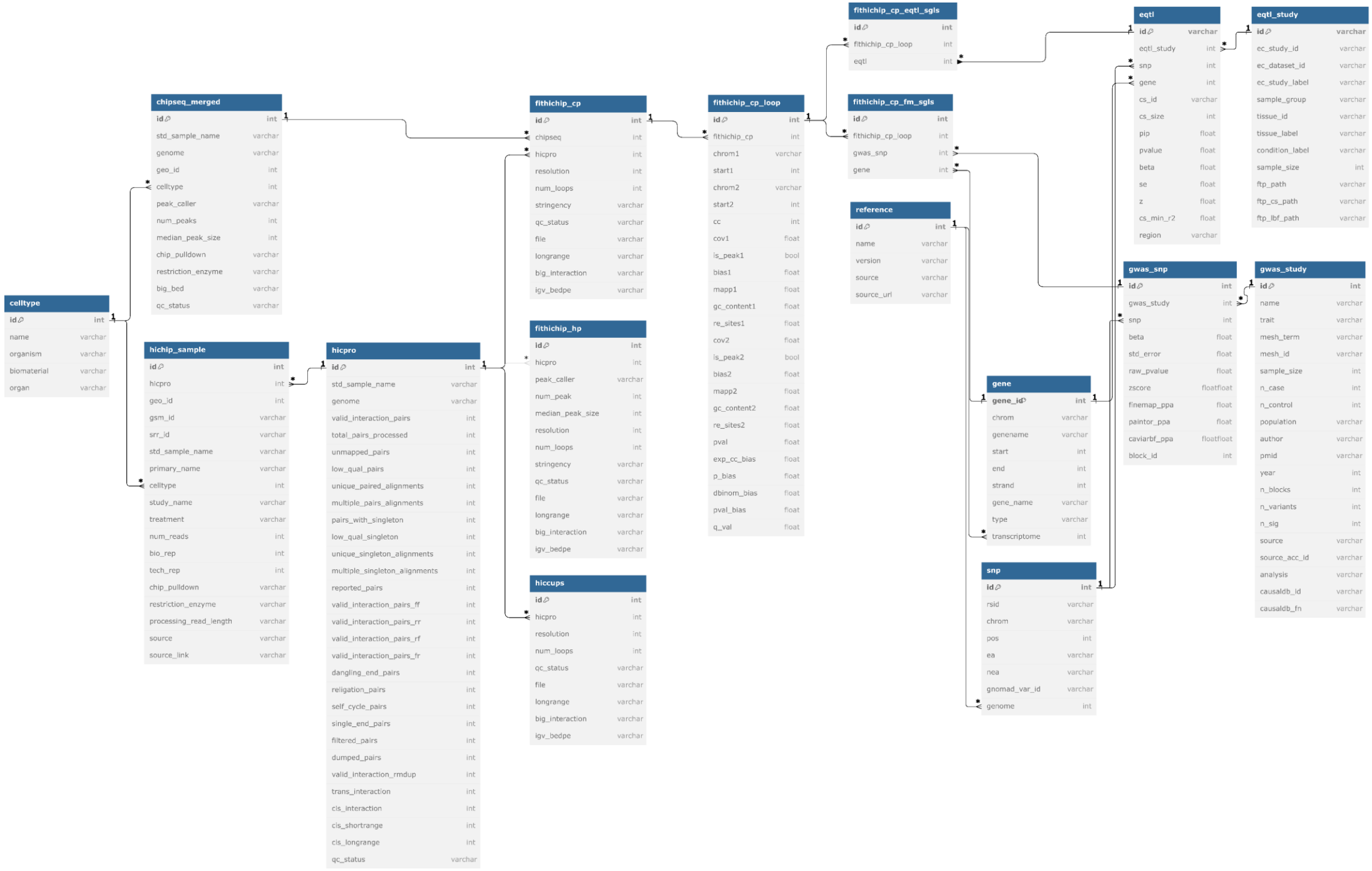

B
Number of Loop Calls by Protein Pulldown (mm10)
Number of FH Loops (Stringent 5kb)
60000
50000
40000
30000
20000
10000
0
H3K27ac
n = 88 Samples
H3K4me3
n = 20 Samples
SMC1A
n = 16 Samples
Other
n = 19 Samples
Protein Pulldown

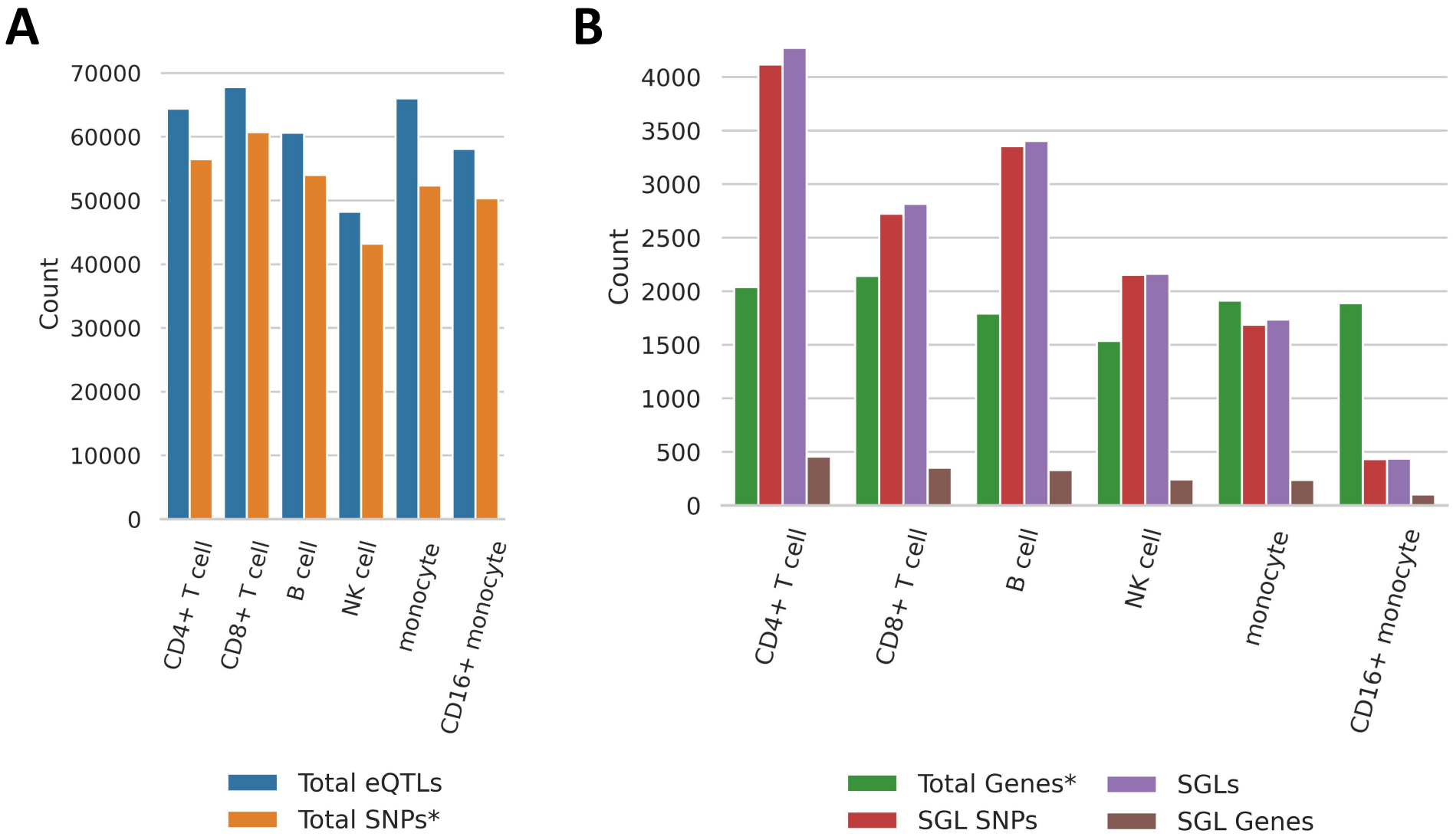

A
hg38 ChIP-seq Peaks / FC Loops Unmerged
hg38 ChIP-seq Peaks / FC Loops Merged
B
hg38 FH Peaks / Loops Unmerged
hg38 FH Peaks / Loops Merged
C
mm10 ChIP-seq Peaks / FC Loops Unmerged
mm10 ChIP-seq Peaks / FC Loops Merged
D
mm10 FH Peaks / Loops Unmerged
mm10 FH Peaks / Loops Merged
Number of Samples
Flag
Flag Type
Poor
Warning
Good
Peaks
S5 Loops
L5 Loops
S10 Loops
L10 Loops
S25 Loops
L25 Loops
S5 Final
L5 Final
S10 Final
L10 Final
S25 Final
L25 Final




| Motif Pair | Sim1 | … | Sim100000 |
|---|---|---|---|
| CTCF-ZNF460 | 30 | 28 | 32 |
| MAZ-KLF5 | 4 | 3 | 3 |
| ZNF460-ZNF460 | 26 | 43 | 27 |
| CTCF-MAZ | 2 | 3 | 0 |
| CTCF-KLF5 | 4 | 1 | 5 |
| ZNF460-MAZ | 1 | 1 | 0 |
| ZNF460-KLF5 | 2 | 0 | 0 |

| Motif Pair | Obs Counts |
|---|---|
| CTCF-ZNF460 | 40 |
| MAZ-KL5 | 5 |
| ZNF460-ZNF460 | 80 |


fithichip_cp_eqtl_sgls
id int
fithichip_cp_loop int
eqtl int
eqtl
id varchar
eqtl_study int
snp int
gene int
cs_id varchar
cs_size int
pip float
pvalue float
beta float
se float
z float
cs_min_r2 float
region varchar
eqtl_study
id varchar
ec_study_id varchar
ec_dataset_id varchar
ec_study_label varchar
sample_group varchar
tissue_id varchar
tissue_label varchar
condition_label varchar
sample_size int
ftp_path varchar
ftp_cs_path varchar
ftp_lbf_path varchar
chipseq_merged
id int
std_sample_name varchar
genome varchar
geo_id int
celltype int
peak_caller varchar
num_peaks int
median_peak_size int
chip_pulldown varchar
restriction_enzyme varchar
big_bed varchar
qc_status varchar
fithichip_cp
id int
chipseq int
hicpro int
resolution int
num_loops int
stringency varchar
qc_status varchar
file varchar
longrange varchar
big_interaction varchar
igv_bedpe varchar
fithichip_cp_loop
id int
fithichip_cp int
chrom1 varchar
start1 int
chrom2 varchar
start2 int
cc int
cov1 float
is_peak1 bool
bias1 float
mapp1 float
gc_content1 float
re_sites1 float
cov2 float
is_peak2 bool
bias2 float
mapp2 float
gc_content2 float
re_sites2 float
pval float
exp_cc_bias float
p_bias float
dbinom_bias float
pval_bias float
q_val float
fithichip_cp_fm_sgls
id int
fithichip_cp_loop int
gwas_snp int
gene int
reference
id int
name varchar
version varchar
source varchar
source_url varchar
celltype
id int
name varchar
organism varchar
biomaterial varchar
organ varchar
hichip_sample
id int
hicpro int
geo_id int
gsm_id varchar
srr_id varchar
std_sample_name varchar
primary_name varchar
celltype int
study_name varchar
treatment varchar
num_reads int
bio_rep int
tech_rep int
chip_pulldown varchar
restriction_enzyme varchar
processing_read_length varchar
source varchar
source_link varchar
hicpro
id int
std_sample_name varchar
genome varchar
valid_interaction_pairs int
total_pairs_processed int
unmapped_pairs int
low_qual_pairs int
unique_paired_alignments int
multiple_pairs_alignments int
pairs_with_singleton int
low_qual_singleton int
unique_singleton_alignments int
multiple_singleton_alignments int
reported_pairs int
valid_interaction_pairs_ff int
valid_interaction_pairs_rr int
valid_interaction_pairs_rf int
valid_interaction_pairs_fr int
dangling_end_pairs int
religation_pairs int
self_cycle_pairs int
single_end_pairs int
filtered_pairs int
dumped_pairs int
valid_interaction_rmdup int
trans_interaction int
cis_interaction int
cis_shortrange int
cis_longrange int
qc_status varchar
fithichip_hp
id int
hicpro int
peak_caller varchar
num_peak int
median_peak_size int
resolution int
num_loops int
stringency varchar
qc_status varchar
file varchar
longrange varchar
big_interaction varchar
igv_bedpe varchar
hiccups
id int
hicpro int
resolution int
num_loops int
qc_status varchar
file varchar
longrange varchar
big_interaction varchar
igv_bedpe varchar
gene
gene_id int
chrom varchar
genename varchar
start int
end int
strand int
gene_name varchar
type varchar
transcriptome int
snp
id int
rsid varchar
chrom varchar
pos int
ea varchar
nea varchar
gnomad_var_id varchar
genome int
gwas_snp
id int
gwas_study int
snp int
beta float
std_error float
raw_pvalue float
zscore floatfloat
finemap_ppa float
paintor_ppa float
caviarbf_ppa floatfloat
block_id int
gwas_study
id int
name varchar
trait varchar
mesh_term varchar
mesh_id varchar
sample_size int
n_case int
n_control int
population varchar
author varchar
pmid varchar
year int
n_blocks int
n_variants int
n_sig int
source varchar
source_acc_id varchar
analysis varchar
causaldb_id varchar
causaldb_fn varchar

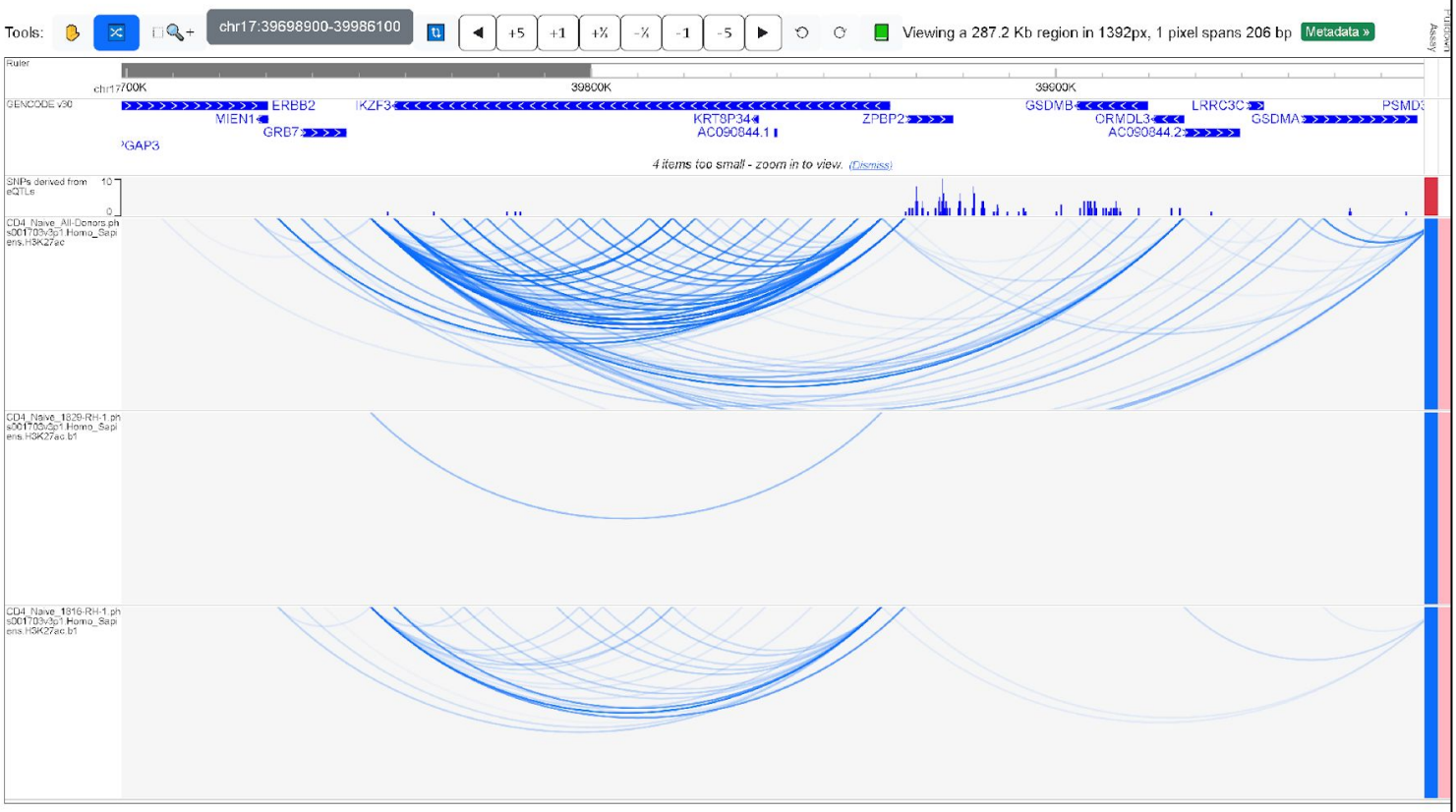

A
70000
60000
50000
40000
30000
20000
10000
0
Count
CD4+ T cell
CD8+ T cell
B cell
NK cell
monocyte
CD16+ monocyte
Total eQTLs
Total SNPs*
B
4000
3500
3000
2500
2000
1500
1000
500
0
Count
CD4+ T cell
CD8+ T cell
B cell
NK cell
monocyte
CD16+ monocyte
Total Genes*
SGL SNPs
SGLs
SGL Genes

# Loop Catalog

Loop Catalog Data Tools Docs Contact Us hg38

## eQTL-SGL Analysis for Schmiedel 2018 - CD4+ T cell

eQTL Catalogue Dataset ID: QTD000479
Condition: Naive




### Locus

Locus: ORMDL3

Slop: +/- slop for chr:pos (default=100)

Query

### SGLs Derived from FitHiChIP with ChIP-seq Peaks (FC Loops)

Copy CSV

Showing 1 to 10 of 16 entries

Search:

| rsID | Chrom | SNP BP | eQTL PIP | Gene Name | Anchor1 | Anchor2 | Distance | Loop -log(Q) | Sample | eQTL Study |
|---|---|---|---|---|---|---|---|---|---|---|
| rs11078926 | chr17 | 39906723 | 0.0212476740936706 | ORMDL3 | chr17:39905000-39910000 | chr17:39925000-39930000 | 20000 | 2.32 | CD4_Naive_All-Donors.phs001703... | Schmiedel_2018 |
| rs11078927 | chr17 | 39908152 | 0.0209544969855168 | ORMDL3 | chr17:39905000-39910000 | chr17:39925000-39930000 | 20000 | 2.32 | CD4_Naive_All-Donors.phs001703... | Schmiedel_2018 |
| rs11078928 | chr17 | 39908216 | 0.0177822560302694 | ORMDL3 | chr17:39905000-39910000 | chr17:39925000-39930000 | 20000 | 2.32 | CD4_Naive_All-Donors.phs001703... | Schmiedel_2018 |
| rs12939832 | chr17 | 39908623 | 0.0209544969855168 | ORMDL3 | chr17:39905000-39910000 | chr17:39925000-39930000 | 20000 | 2.32 | CD4_Naive_All-Donors.phs001703... | Schmiedel_2018 |
| rs12946510 | chr17 | 39756124 | 0.0067490842570102 | ORMDL3 | chr17:39755000-39760000 | chr17:39925000-39930000 | 170000 | 4.07 | CD4_Naive_1800-RH-1.phs001703v... | Schmiedel_2018 |

Show 10 entries

Previous 1 2 Next