Differential Analysis Pipeline README

Table of Contents

Analysis Outputs ... **1**

Primary report outputs ... 1

Capture-Specific Differential Analysis Output ... 2

HiChIP-Specific Differential Analysis Output ... 3

Gene Ontology Analysis Output ... 3

Feature-level comparison results ... 4

Detailed description of relevant files ... **4**

Description of differential analysis output for whole genome hic ... 4

Description of differential analysis output for capture ... 5

Description of differential analysis output for hichip ... 6

Description of top differential results integrated with epigenetic features ... 7

Description of optionally provided regions of interest integrated with epigenetic features ... 7

Description of collapsed GO analysis table ... 8

Detailed methodology for feature-level comparisons ... **9**

A/B Compartment Comparative Analysis ... 9

TAD Comparative Analysis ... 10

Loop Comparative Analysis ... 11

List of relevant tools used for differential analysis ... **12**

Analysis Outputs

Primary report outputs

The following section provides an overview of the primary outputs generated by the differential analysis pipeline. These files represent the key results of the analysis and are intended to guide interpretation, downstream biological investigation, and reporting. Users should begin with the main summary report, which provides a high-level overview of data quality, normalization, statistical testing, and significant findings.

The table below highlights the most important result files produced by the pipeline. Core statistical outputs from whole-genome Hi-C (*HiCcompare*) or capture Hi-C (*chicdiff*) are provided in results.txt, depending on the analysis type. Loops_EdgeR_Default.bed is the differential analysis result for HiChIP. These files contain the complete interaction-level statistical results, including fold changes, significance values, and adjusted p-values. For convenience and biological interpretation, we additionally provide summary tables that integrate differential interactions with structural and epigenetic features, as well as a consolidated Gene Ontology enrichment summary table. Where applicable, region-of-interest analyses are also summarized in a dedicated file.

Selected file names in the table are hyperlinked to detailed descriptions later in this document. Those sections provide column-by-column explanations, methodological context, and guidance on how each file can be used for interpretation, figure generation, and manuscript preparation. Together, these outputs provide a comprehensive and reproducible record of the differential chromatin interaction analysis and its functional annotation.

File Name	Description
summary_report.html	Main summary report
results.txt (wgs) results.txt (capture) Loops_EdgeR_Default.bed (hichip)	Main differential interaction results table containing statistics (log fold change, p-value, adjusted p-value/FDR, counts per group). Note: *HiCcompare* was used for whole genome differential analysis. *Chicdiff* is used for capture differential analysis. Fithichip is used for differential hichip analysis.
diff_annotated_with_features.tsv	Top 5,000 differentially interacting regions from results.txt integrated with epigenetic feature calls from the epigenetics pipeline
top_differential_regions_contact_maps.tar.gz	Package of contact maps for top identified differential regions
roi_annotated_with_features.tsv	If provided, a summary of fold change and p value (if applicable) between groups, and associated epigenetic features for the given region of interest
regions_of_interest_contact_maps.tar.gz	Package of contact maps for regions of interest (if applicable)
GO_BP_UPDOWN_enrichGO_readable_collapsed_summary.tsv	Combined summary file (one table) containing the collapsed, readable depth-filtered GO terms from both UP and DOWN, annotated with direction (UP/DOWN) and group (labels from --groupA_label / --groupB_label) and a top_metric_value column (computed using --top_metric). Always written. If no terms in either direction, an empty TSV with columns is written.
GO_BP_outputs.tar.gz	Package of expanded GO Analysis outputs
ab_summary.csv (whole genome hic only)	Summary of A/B compartment bins unique and shared between groups, for each resolution
tad_summary.csv (whole genome hic only))	Summary of TADs unique and shared between groups, for each resolution
loop_summary.csv	Summary of loops unique and shared between groups, for each resolution
pairtoolsStats.csv	Summary of pairtools statistics for each library used in the analysis
download_all.sh	Script enabling download of all files

Capture-Specific Differential Analysis Output

The following files are generated by *chicdiff*, the statistical framework used for differential interaction analysis in capture Hi-C experiments. These files are supplemental outputs specific to capture-based analysis and are provided in addition to the primary report.

Chicdiff is designed to identify statistically significant differences in chromatin interaction frequencies between experimental conditions. It operates on interaction-level count data derived from capture Hi-C experiments and applies normalization procedures to account for library size differences and other technical biases. After filtering low-count interactions, the method performs statistical testing to compare interaction frequencies between defined groups. Resulting p-values are adjusted to control the false discovery rate (FDR), and both fold changes and adjusted significance values are reported. The files listed below include intermediate analysis objects to ensure reproducibility, as well as processed result tables suitable for downstream interpretation, visualization, and inclusion in publications.

File Name	Description
names.txt	List of sample names or condition labels used in the analysis.
peaks.Rds	R object containing significant differential interactions.
peaks.txt	Raw counts per sample for each bait
peaks_peaklist.txt	Simplified list of significant peaks for reporting or visualization.
peaks_tree.pdf	Hierarchical clustering/tree visualization of significant interactions (if enabled).
results_countput.Rds	.Rds object file containing raw counts for each library
results_filteredBaits.txt	List of baits retained after filtering (low-count filtering, QC steps).
results_results.Rds	Full differential testing results object (model output, statistics).
results_settings.Rds	Stores pipeline parameters and configuration used for the run.

HiChIP-Specific Differential Analysis Output

The following files are generated by *fithichip-differential*, the statistical framework used for differential interaction analysis in hichip experiments. These files are supplemental outputs specific to fithichip.

FitHiChIP differential analysis identifies loops that change between two conditions by comparing contact counts for a union set of loops across replicates using an edgeR-based negative binomial model. It treats loop interaction counts analogously to RNA-seq counts, applying normalization, dispersion estimation, and statistical testing to compute log2 fold change and FDR for each loop. The output is a table of differential loops prioritized by effect size and significance, and—optionally—annotated based on whether changes are driven by differences in 3D chromatin contacts alone or accompanied by changes in underlying 1D ChIP-seq signal at loop anchors.

File Name	Description
bedToBigBed	Utility script/binary for converting BED → BigBed
count_matrix.bed	Raw interaction counts for all loops across samples
DiffLoops_ALL.bed	All tested loops with differential statistics
DiffLoops_ALL_IGV.bed	Same as above formatted for IGV
DiffLoops_ALL_WashU.bed	Same as above formatted for WashU browser
Input_Parameters_*.log	Logs of parameters used for each run
interact.as	AutoSQL schema for UCSC interact format
Interacting_segments.bed	All interacting anchors/segments
Loops_EdgeR_Default.bed	Loops with edgeR statistics (logFC, p-value)
Loops_EdgeR_Default_SIG.bed	Significant differential loops (FDR-filtered)
MasterSheet_group1_group2_Loops.bed	Union set of loops used for comparison

Gene Ontology Analysis Output

The following files are generated as part of the Gene Ontology (GO) enrichment analysis and are provided as supplemental outputs to support biological interpretation of the differential interaction results. These files summarize functional enrichment for genes associated with significantly changing interactions and provide both tabular results and visualization-ready outputs.

GO enrichment analysis is performed separately for genes associated with increased (UP) and decreased (DOWN) interaction signals. Gene sets derived from the differential analysis are tested for overrepresentation of Gene Ontology Biological Process (GO:BP) terms relative to an appropriate background. Enrichment is evaluated using standard overrepresentation testing, and resulting p-values are adjusted to control the false discovery rate (FDR). Depth filtering allows restriction of terms to specified ontology levels, and redundancy reduction may be applied using gene-overlap–based similarity thresholds to collapse highly overlapping GO terms. Readable versions of result tables map gene identifiers to gene symbols for ease of interpretation. Visualization outputs, including enrichment maps and dot plots, are generated from filtered and optionally collapsed term sets to facilitate reporting and downstream presentation. Empty placeholder files are written in cases where no significant terms pass filtering criteria, ensuring pipeline consistency and reproducibility.

File Name	Description
GO_BP_UP_enrichGO_all.tsv	Full enrichGO result (raw ego data frame) for UP genes. If too few genes mapped (< --min_genes) or enrichGO returns 0 terms, an empty TSV with the canonical columns is written.

GO_BP_UP_enrichGO_readable_all.tsv	Same results as (1) but with gene IDs converted to readable gene symbols (via setReadable). Empty placeholder if no terms or too few genes.
GO_BP_UP_enrichGO_depth<min>-<max>.tsv	Subset of (1) filtered by GO BP depth (--go_min_depth–--go_max_depth). Empty placeholder if zero terms pass depth filter or upstream failures. Default depth 4-7
GO_BP_UP_enrichGO_readable_depth<min>-<max>.tsv	Same as (3) but readable gene symbols. Empty placeholder if none.
GO_BP_UP_enrichGO_readable_depth<min>-<max>_collapsed.tsv	Depth-filtered and redundancy-collapsed (by gene-overlap / Jaccard threshold when --collapse_go=overlap) readable table. If collapse requested but no geneID column present, collapse is skipped (original depth table saved). If there are zero terms, an empty placeholder is written.
GO_BP_UP_emap.png / GO_BP_UP_emap.pdf	Enrichment network (emapplot) built from the depth-filtered + collapsed IDs (uses pairwise_termsim + emapplot). If there are no terms after depth filter or after optional qvalue filtering for plotting, the emap is skipped and no plot files are written for the network. (Note: the code only writes emap files when there are terms to plot.)
GO_BP_DOWN_enrichGO_all.tsv	Same as (1) but for DOWN genes. Empty placeholder behavior identical.
GO_BP_DOWN_enrichGO_readable_all.tsv	Same as (2) but for DOWN genes.
GO_BP_DOWN_enrichGO_depth<min>-<max>.tsv	Same as (3) but for DOWN genes.
GO_BP_DOWN_enrichGO_readable_depth<min>-<max>.tsv	Same as (4) but for DOWN genes.
GO_BP_DOWN_enrichGO_readable_depth<min>-<max>_collapsed.tsv	Same as (5) but for DOWN genes.
GO_BP_DOWN_emap.png / GO_BP_DOWN_emap.pdf	Same as (6) but for DOWN genes. Writes only if there are terms to plot after depth/collapse/qvalue filtering.
GO_BP_TOP_updown_dot.png / GO_BP_TOP_updown_dot.pdf	Faceted dotplot showing Top N GO:BP terms (both UP and DOWN) ranked by --top_metric. Always written: if there are no terms available the script writes a placeholder image that contains the message "No enriched GO:BP terms (faceted top plot)".
GO_BP_TOP_updown_single_dot.png / GO_BP_TOP_updown_single_dot.pdf	"Single column" customer-facing dotplot of top terms (UP vs DOWN at two X positions). Always written: if empty, a placeholder is written with "No enriched GO:BP terms (top plot)".

Feature-level comparison results

This section summarizes the outputs generated from feature-level comparisons between subgroups, highlighting differences in chromatin organization across multiple structural features. The files below contain results for loop calls, topologically associating domains (TADs), and A/B compartments, including both shared and subgroup-specific features.

File Name	Description
loop_analysis_outputs.tar.gz	A package containing the files associated with feature-level loop calls between subgroups. BEDPE files represent loop calls unique and shared between subgroups. Loops are identified using mustache for whole genome hic, Chicago for capture-c, and fithichip for hichip experiments.
tad_analysis_outputs.tar.gz (whole genome hic only)	A package containing the files associated with feature-level TAD calls between subgroups. BED files represent TAD calls unique and shared between subgroups. TADs are identified using arrowHead for whole genome hic.
ab_analysis_outputs.tar.gz	A package containing the files associated with AB compartment comparisons between subgroups. BEDPE files represent loop calls unique and shared between subgroups. Loops are identified via mustache for whole genome hic, Chicago for capture-c, and fithichip for hichip experiments. AB compartments are identified using Fan-C compartments.

Detailed description of relevant files

Description of differential analysis output for whole genome hic

The following table describes the primary output file generated by the *HiCcompare*-based whole-genome differential analysis. This file (results.txt) contains interaction-level statistics comparing two Hi-C datasets and forms the basis for identifying significantly differential chromatin contacts.

HiCcompare performs joint normalization of two Hi-C contact matrices to remove systematic biases, including distance-dependent interaction decay. The method models interaction differences using an MA/MD framework, where interaction intensities are compared as a function of genomic distance. A

loess-based normalization is applied to correct for systematic trends, and the resulting adjusted interaction frequencies are used to compute log-scale differences (M values). Statistical significance is assessed using Z-scores derived from normalized M values, and p-values are calculated under a standard normal assumption. Multiple testing correction is then applied (e.g., FDR), producing adjusted p-values used to define significant differential interactions. The columns described below capture raw interaction frequencies, normalized values, model-derived statistics, and significance measures necessary for interpretation, filtering, and downstream analysis.

Column Name	Description
chr1	Chromosome for anchor/bin 1.
start1	Start coordinate for anchor/bin 1.
end1	End coordinate for anchor/bin 1.
chr2	Chromosome for anchor/bin 2.
start2	Start coordinate for anchor/bin 2.
end2	End coordinate for anchor/bin 2.
IF1	Interaction frequency (contact value) from dataset 1 (input / pre-normalization value in the table).
IF2	Interaction frequency (contact value) from dataset 2.
D	Genomic distance between the two anchors/bins (used on the x-axis of the MD plot)
M	"Difference" between the two datasets used in the MD plot (the M in MD). In *HiCcompare* this is the log-scale difference between the two interaction frequencies (the value being modeled vs distance).
adj.IF1	Adjusted (jointly normalized) interaction frequency for dataset 1, produced by hic_loess().
adj.IF2	Adjusted (jointly normalized) interaction frequency for dataset 2, produced by hic_loess().
adj.M	Adjusted/normalized M value after joint normalization (i.e., M computed after applying the loess-based correction).
mc	The M correction factor added by joint normalization (a distance-dependent correction used to center/adjust M).
A	Average expression/intensity ("A" in an MA/MD-style representation) added by joint normalization; used for filtering via A.min.
Z	Z-score for the interaction derived from M values (Z becomes NA for rows filtered out by the A threshold).
p.value	Unadjusted p-value computed from the standard normal distribution using the Z-score.
p.adj	Multiple-testing adjusted p-value (method controlled by p.method, e.g. FDR).

Description of differential analysis output for capture

The following table describes the primary output file generated by the *chicdiff*-based capture Hi-C differential analysis. This file (results.txt) contains interaction-level statistics comparing predefined experimental groups and represents the main statistical result used to identify significant differential interactions.

Chicdiff models interaction counts anchored at capture baits using a generalized linear modeling framework applied to normalized count data across samples. Interactions are stratified by genomic distance to account for distance-dependent contact decay, and statistical testing is performed to estimate log2 fold changes between comparison groups. Wald statistics are computed for each interaction, and resulting p-values are adjusted for multiple testing. To further improve power while controlling false discovery, *chicdiff* applies independent hypothesis weighting (IHW), which assigns data-driven weights to interactions based on distance strata. The final significance metric is the weighted, FDR-adjusted p-value, which is used to define differential interactions. The columns described below include model coefficients, normalization summaries, distance metrics, weighting parameters, and both raw and weighted significance values, enabling transparent interpretation and reproducibility of the differential analysis.

Column Name	Description
group	Distance bin (stratum) assigned to the interaction based on genomic distance. Used for independent hypothesis weighting (IHW), not a biological group label.
baseMean	Mean normalized interaction count across all samples included in the comparison.

log2FoldChange	Log2 fold change of interaction counts between comparison groups. Positive values indicate enrichment in the numerator group; negative values indicate enrichment in the denominator group.
lfcSE	Standard error of the log2 fold change estimate.
stat	Wald test statistic from the differential interaction model.
pvalue	Raw p-value from the statistical test.
padj	Multiple testing–adjusted p-value (FDR), typically before weighting.
baitID	Identifier of the bait fragment anchoring the interaction.
maxOE	Maximum observed/expected (OE) interaction value across samples.
minOE	Minimum observed/expected (OE) interaction value across samples.
regionID	Identifier of the interacting fragment (other end region).
OEchr	Chromosome of the interacting fragment (other end).
OEstart	Start coordinate of the interacting fragment.
OEend	End coordinate of the interacting fragment.
baitchr	Chromosome of the bait fragment.
baitstart	Start coordinate of the bait fragment.
baitend	End coordinate of the bait fragment.
avDist	Average genomic distance between bait and interacting fragment.
uniform	P-value under uniform weighting (used as a reference in weighting procedure).
shuff	P-value derived from shuffled/background data used to estimate distance-dependent behavior.
avgLogDist	Average log-transformed genomic distance for the interaction.
avWeights	Average weight assigned within the distance stratum during IHW.
weight	Final weight applied to this interaction’s p-value.
weighted_pvalue	P-value after applying distance-based weighting.
weighted_padj	FDR-adjusted p-value after weighting (final significance metric used to call differential interactions).

Description of differential analysis output for hichip

The following table describes the primary output file generated by the FitHiChIP-based HiChIP differential analysis. This file (Loops_EdgeR_Default.bed or Loops_EdgeR_Default_SIG.bed) contains loop-level statistics comparing predefined experimental groups and represents the main statistical result used to identify significant differential chromatin interactions.

FitHiChIP performs differential analysis by constructing a union set of loops across all samples and quantifying interaction counts for each loop in each replicate. These counts are modeled using a negative binomial generalized linear framework implemented in edgeR, analogous to RNA-seq differential expression analysis. Library size normalization and dispersion estimation are applied to account for differences in sequencing depth and variability across replicates. Statistical testing is performed to estimate log2 fold changes between comparison groups, with likelihood-based methods used to compute p-values for each loop. Multiple testing correction is applied using the Benjamini–Hochberg procedure to control the false discovery rate (FDR).

The resulting output reports, for each loop, genomic coordinates, raw interaction counts, normalized abundance (logCPM), estimated log2 fold change between groups, and both raw and adjusted significance values. Additional columns summarize replicate-level support for each interaction, enabling assessment of reproducibility across samples. Together, these metrics provide a statistically rigorous and interpretable framework for identifying and prioritizing differential chromatin loops between experimental conditions.

Column Name	Description
chr1	Chromosome of anchor 1
start1	Start of anchor 1
end1	End of anchor 1
chr2	Chromosome of anchor 2
start2	Start of anchor 2
end2	End of anchor 2
group1_R1_RawCC	Raw contact count in Group1 replicate 1

group1_R1_QVal	Peak significance (q-value) in Group1 replicate 1
group2_R1_RawCC	Raw contact count in Group2 replicate 1
group2_R1_QVal	Peak significance (q-value) in Group2 replicate 1
logFC	Log2 fold change from edgeR
logCPM	Log counts per million (normalized abundance)
PValue	Statistical p-value from edgeR
FDR	Multiple-testing corrected p-value
group1_SigRepl (Test group)	Number of replicates in Group1 where loop is significant
group2_SigRepl (Control Group)	Number of replicates in Group2 where loop is significant

Description of top differential results integrated with epigenetic features

The following table describes the summary file (diff_annotated_with_features.tsv) that integrates statistically significant differential chromatin interactions with higher-order chromatin features identified by external structural callers. This file provides a biologically contextualized view of the top differential interactions by linking them to loops and domain-level features detected in the corresponding datasets.

Significant differential interactions identified by *HiCcompare* (whole-genome Hi-C) or *chicdiff* (capture Hi-C) are annotated with structural features derived from established tools. For whole-genome analyses, loop calls from *Mustache* and topologically associating domains (TADs) from *Arrowhead* are incorporated. For capture analyses, loop calls from *CHiCAGO* are used. Each differential interaction is evaluated for overlap with group-specific and shared structural features, allowing interactions to be classified according to whether they are unique to one condition or common to both. Statistical values reported in this table correspond to the primary differential testing framework used (*HiCcompare* or *chicdiff*), including raw and adjusted p-values and the appropriate fold-change metric. The resulting summary enables direct interpretation of differential interaction signals within the context of 3D genome architecture and provides a streamlined table suitable for downstream biological interpretation and publication reporting.

Column Name	**Description**
Left_position	5' interacting co-ordinate
Right_position	3' interacting co-ordinate
pvalue	p-value ("pvalue" from *HiCcompare* for wgs; "p.value" from *chicdiff* for capture)
FDR	multiple hypothesis-corrected p-value ("p.adj" from *HiCcompare* for WGS; "weighted_padj" from *chicdiff* for capture)
FoldChange	fold change between groups ("adj.M" from *HiCcompare* for wgs; "log2FoldChange" from *chicdiff* for capture)
groupA_features (Test group)	Epigenetic features unique to group A. For WGS, features are both loops (from *Mustache*) and TADs (from *Arrowhead*) and are delimited with a semi-colon (loops ; TADs). For capture, only loops identified from *ChiCAGO* are displayed
groupB_features (Control group)	Same as above, for group B.
Shared_features	Features shared between both groups.

Description of optionally provided regions of interest integrated with epigenetic features

The following table describes the summary file (roi_annotated_with_features.tsv) generated when customer-provided regions of interest (ROIs) are supplied for targeted analysis. This file integrates region-level differential interaction metrics with structural and epigenetic features, providing a focused view of condition-specific chromatin organization within user-defined loci.

For each ROI, interaction signal is aggregated within a defined genomic window centered on the submitted coordinates. Region-level structure scores are computed per sample based on the mean log-

transformed contact intensity within a specified off-diagonal distance band. Group-wise differences are then evaluated using a two-sided Welch's t-test to compare ROI structure scores between conditions. Resulting p-values may be adjusted across all tested ROIs using the Benjamini–Hochberg procedure to control the false discovery rate. In parallel, ROIs are annotated with overlapping chromatin features identified in the corresponding datasets. For whole-genome Hi-C analyses, loops (*Mustache*) and TADs (*Arrowhead*) are reported, while for capture Hi-C analyses, *CHiCAGO*-identified loops are displayed. Features are categorized as unique to each group or shared between groups. This integrated summary enables direct interpretation of user-specified loci in the context of both quantitative differential interaction testing and higher-order 3D genome architecture.

Column Name	Description
Id	Customer-provided identifier for the region of interest
Left_position	5' interacting co-ordinate
Right_position	3' interacting co-ordinate
Pvalue	Two-sided statistical test p-value (Welch's t-test) comparing per-sample ROI structure scores between group A and group B.
FDR	BH-adjusted across all ROIs, but you didn't ask for that field.
FoldChange	meanA_score − meanB_score, where each score is the mean of log1p(contact) over all bin-Difference between Group A and Group B mean ROI structure scores, calculated as the mean log1p-transformed contact intensity within the specified off-diagonal distance band (group A − group B).
groupA_features (Test group)	Epigenetic features unique to group A. For WGS, features are both loops (from *Mustache*) and TADs (from *Arrowhead*) and are delimited with a semi-colon (loops ; TADs). For capture, only loops identified from *ChiCAGO* are displayed
groupB_features (Control group)	Same as above, for group B.
Shared_features	Features shared between both groups.

Description of collapsed GO analysis table

The following table describes the final collapsed Gene Ontology (GO) enrichment summary file (GO_BP_UPDOWN_enrichGO_readable_collapsed_summary.tsv). This table represents a consolidated, redundancy-reduced view of enriched GO Biological Process terms identified from differential interaction–associated gene sets and is intended for reporting, interpretation, and publication use.

GO enrichment is performed separately for genes associated with increased (UP) and decreased (DOWN) interaction signals. Enriched GO:BP terms are identified using overrepresentation testing using the clusterProfiler R package using UCSC gene annotations (TxDb.Hsapiens.UCSC.hg38.knownGene for human; TxDb.Mmusculus.UCSC.mm10.knownGene for mouse), and p-values are adjusted using the Benjamini–Hochberg procedure to control the false discovery rate. To improve interpretability and reduce redundancy inherent to the hierarchical structure of the Gene Ontology, significantly overlapping terms are collapsed based on gene-set similarity thresholds. From each cluster of related terms, a representative term is retained, preserving enrichment statistics while minimizing repetition of highly similar biological processes. The resulting table combines UP and DOWN enrichments into a single summary, includes readable gene annotations where available, and provides ranking metrics to facilitate prioritization of biologically meaningful processes. This collapsed summary serves as the primary GO results table for downstream biological interpretation and figure generation.

Column Name	Description
ID	GO term identifier (e.g., GO:000xxxx).
Description	GO term name/description.
GeneRatio	Ratio of input genes annotated to the GO term (as "k/n").
BgRatio	Ratio of background (k) / universe (n) genes annotated to the term (as "K/N").

pvalue	Raw enrichment p-value for the GO term.
p.adjust	Multiple-testing adjusted p-value (BH) for enrichment.
qvalue	q-value reported by *enrichGO*
geneID	Slash-separated list of mapped genes contributing to the term (in “readable” form when possible).
Count	Number of mapped genes from the input set assigned to the term.
direction	"UP" or "DOWN" depending on which FoldChange side produced the mapped gene set.
group	Customer-provided label for the direction (defaults: groupA (Test) for UP, groupB (Control) for DOWN).
top_metric_value	Numeric ranking value computed per row (count * logp).

Detailed methodology for feature-level comparisons

A/B Compartment Comparative Analysis

A/B compartment comparison is performed in two stages: first at the level of individual genomic bins, and then at the level of larger compartment domains formed by collapsing adjacent bins of the same type. For each condition, all compartment BED files within the **test** group and all compartment BED files within the **control** group are combined separately. At each genomic bin, the eigenvector values are averaged across all samples in that group to create a single group-level compartment signal.

Because eigenvector sign can flip arbitrarily between runs, the script standardizes orientation independently within each chromosome before assigning compartments. For each chromosome, it calculates the median eigenvector value across bins, and if that median is negative, it flips the sign of all eigenvectors on that chromosome. After this orientation step, bins with positive eigenvector values are labeled as **A compartments**, and bins with zero or negative eigenvector values are labeled as **B compartments**. This makes compartment assignment deterministic within each group and prevents the comparison from depending on which condition was chosen as the reference.

At the bin level, comparison is performed by exact genomic coordinate matching. The test and control group tables are merged on chromosome, start, and end position, so a bin is only compared if it exists at the same genomic coordinates in both groups. There is no window padding or positional tolerance applied at this stage. A bin is considered **concordant** when it has the same compartment assignment in both groups. A bin is considered an **A to B switch** when it is assigned A in the control group and B in the test group, and a **B to A switch** when it is assigned B in the control group and A in the test group. In the current implementation, the merged comparison table is constructed with control bins as the first input and test bins as the second input, so these switch labels are interpreted relative to control-to-test change.

To summarize broader compartment organization, adjacent bins with the same compartment label are collapsed into continuous domains. Two bins are merged into the same domain only when they are on the same chromosome, have the same compartment assignment, and are directly contiguous, meaning the start of the next bin is equal to the end of the previous bin. No additional padding or gap tolerance is used when forming these domains.

Domain-level comparison is then performed using bedtools intersect, with the control domains supplied as the first input and the test domains as the second. Any base-pair overlap is sufficient to generate a

comparison record; there is no reciprocal overlap threshold or minimum fraction requirement. Each overlapping domain pair is then classified based on compartment identity and domain span. A **flip** is defined when overlapping control and test domains have different compartment labels. If the compartment label is the same in both domains, the script compares their lengths. A **split** is assigned when the control domain is larger than the overlapping test domain, and a **merge** is assigned when the test domain is larger than the overlapping control domain. If the overlapping domains have the same compartment and the same span, the label is **No Change**.

It is important to note that the **merge** and **split** labels in this implementation do not represent strict many-to-one or one-to-many structural events. The code evaluates each overlapping domain pair independently and classifies it based on relative size alone. As a result, a **merge** here is best understood as a same-compartment domain that is larger in the test group than in the control group, while a **split** is a same-compartment domain that is smaller in the test group than in the control group. These labels therefore describe relative domain expansion or contraction rather than literal merging or fragmentation of multiple domains.

TAD Comparative Analysis

TADs are compared at the group level by first constructing a consensus set of domains for each condition. All TAD intervals within the **test** group and the **control** group are concatenated separately, sorted by genomic position, and merged using bedtools merge to produce a non-overlapping set of domains per group. An optional padding parameter can be applied during this step, allowing nearby domains within a specified genomic distance to be merged together. This padding only affects how TADs are consolidated within each group prior to comparison.

To identify corresponding domains between the test and control groups, all pairwise overlaps are computed using bedtools intersect -wo. However, not all overlaps are considered valid matches. A pair of TADs is only considered comparable if it satisfies a **reciprocal overlap** requirement, meaning the overlapping region must cover at least a specified fraction of both the test TAD and the control TAD. By default, this threshold is 95%, ensuring that matched TADs represent nearly identical genomic structures rather than partial overlaps.

When multiple candidate matches exist, the script selects the best match for each TAD based on the size of the overlap, followed by the strength of the reciprocal overlap, and finally by genomic coordinates to ensure deterministic behavior. This matching is performed independently from the perspective of the test and control sets. A pair of TADs is defined as **shared** only if the match is a **mutual best match**, meaning each TAD selects the other as its best partner. This results in a one-to-one, symmetric set of shared domains that is stable regardless of input order. TADs that do not participate in a mutual best match are classified as **unique to the test group** or **unique to the control group**, depending on their origin.

In addition to shared and unique domains, the script identifies higher-order structural changes based on how many overlaps occur between domains. A **merged TAD** is defined as a TAD in the test input that overlaps more than one TAD in the control input. This indicates that a single domain in the test group spans genomic regions that are partitioned into multiple domains in the control group. Conversely, a **split TAD** is defined as a TAD in the control input that overlaps more than one TAD in the test input, indicating

that a single control domain corresponds to multiple domains in the test group. In the current implementation, the merge/split detection is performed with the test and control inputs explicitly ordered, so these labels are directional: **merged** refers to consolidation in the test relative to control, and **split** refers to fragmentation in the test relative to control.

It is important to note that the merge and split classifications are based on raw overlap counts from bedtools intersect and do not use the reciprocal overlap threshold applied to shared TADs. As a result, these categories are more permissive and reflect the presence of multiple overlapping domains rather than strict one-to-one structural equivalence.

Loop Comparative Analysis

Chromatin loops are compared as paired genomic anchors, with each loop represented by two intervals. Before comparison, loops from all samples within a group are combined into a group-level set. Anchor coordinates are snapped to the requested resolution grid, anchor order is canonicalized so that the same loop is represented consistently regardless of input order, and exact duplicate loop geometries within a group are collapsed, keeping the highest-scoring instance.

Loop overlap is not determined by a simple interval intersection at one or both anchors. Instead, two loops are considered candidate matches only if they are on the same pair of chromosomes and their anchor coordinates fall within a fixed positional tolerance. Specifically, the script requires that the start and end coordinates of both anchors be within **±3 bins** of one another, where bin size is the chosen analysis resolution. In other words, the allowable matching window is **3 × resolution** for each anchor boundary. This means that both anchors must agree within the matching window; a match at only one anchor is not sufficient.

When multiple candidate matches are possible, the script does not keep all overlapping pairs. Instead, it identifies the best match for each loop based first on the smallest total coordinate difference across all four anchor boundaries, and then uses loop score as a tie-breaker. A loop pair is considered **shared** only if the match is a **mutual best match**, meaning each loop selects the other as its best partner. This makes the shared set one-to-one, deterministic, and symmetric with respect to group order. Loops that do not participate in a mutual best match are classified as **unique to test** or **unique to control**, depending on which group they belong to.

For loops in the shared set, interaction change is defined by comparing the stored loop score in the two groups. The script calculates fold change as **score_test / score_control**, where “test” is the group passed as the second argument into the intersection function. A shared loop is labeled **strengthened** when this fold change is greater than or equal to the specified threshold, **weakened** when it is less than or equal to the reciprocal of that threshold, and **unchanged** otherwise. Because this calculation is directional, the meaning of strengthened and weakened depends on group order. In the current main script, Group A is passed as the second input to the intersection step, so **strengthened** means stronger in **test relative to control**, and **weakened** means weaker in **test relative to control**.

For loops unique to one group, the script also annotates each anchor by overlap with promoter and enhancer reference tracks. Based on these annotations, unique loops are labeled by anchor type combinations such as promoter-promoter, promoter-enhancer, enhancer-enhancer, or none-containing

categories. This annotation step is used for downstream summaries of the types of loops gained or lost between groups.

List of relevant tools used for differential analysis

Tool	Source
hiccompare	https://github.com/dozmorovlab/HiCcompare
chicago	https://github.com/dovetail-genomics/chicago
chicdiff	https://github.com/RegulatoryGenomicsGroup/chicdiff
fithichip	https://github.com/ay-lab/fithichip
mustache	https://github.com/ay-lab/mustache
arrowHead	https://github.com/mdozmorov/HiC_tools
Fan-C	https://github.com/vaquerizaslab/fanc/tree/main
ClusterProfiler	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html