Load Data

Reproducibility.read_data

For every valid sample

read_peak_file

Create BedGraph obj

Create GenomeLoopData obj

Add GenomeLoopData obj to dict

GenomeLoopData

Initializes a ChromLoopData obj for every valid chromosome

Uses bedgraph to find peak values

Reads loop file adding to appropriate chromosome for each line

Cleans up all chromosomes

ChromLoopData.finish_init

Converts lists to numpy arrays

ChromLoopData.find_loop_anchor_points

Finds specific start/end anchor inside given start/end intervals
Weights each loop according to bedgraph
Removes loops that have overlapping start/end

Preprocess

Reproducibility.preprocess

Calls GenomeLoopData.preprocess for every given GenomeLoopData obj

GenomeLoopData.preprocess

Finds number of peaks for each chromosome after chr1

Calls ChromLoopData.preprocess for every chromosome

Removes problematic chroms

ChromLoopData.filter_with_peaks

Get coverage of peaks across chromosome

Check each loop to see if it has peak support

Comparison

Reproducibility.compare

Loops through all possible comparisons and calls GenomeLoopData.compare for each

GenomeLoopData.compare

Calls ChromLoopData.compare for every chromosome and window

Combines values for a genome-wide value

ChromLoopData.compare

Compares two chromosomes in a given window

get_removed_area - get area that will not be compared

get_loops - get loops in window that are not in removed area

create_graph - use loops to create a graph

get_stats - get the emd/j value for the window