Name: Panjun Kim
Email: pkim11@uthsc.edu
PI Name: Hao Chen
PI email: hchen@uthsc.edu

3D Chromatin Interaction Data Facilitates *de novo* Genome Assembly and Identifies Promoter-Enhancer Interaction in the PFC of Rats

Panjun Kim, Rachel Ward, Robert W. Williams, Burt Sharp, Hao Chen

University of Tennessee Health Science Center

Dynamic and cell-specific interactions among chromatin domains encode a higher order type of genomic information on transcriptional states. Hi-C is a whole genome assay that quantifies chromatin contact frequencies. These frequencies between locations on the chromatin generally decrease as the distance between loci increases on single chromosomes. This contact structure can be used at a genome-wide scale to scaffold long- and short-read sequences, and to correct mis-joined contigs during *de novo* genome assembly. Using linked-read whole genome sequencing data (~60X coverage for each strain), we have assembled the genomes of the HXB/BXH recombinant inbred family (30 strains) and their parental strains - BN-Lx/Cub and SHR/Olapcv. We optimized a pipeline that integrates Hi-C data from prefrontal cortex (PFC), and this process greatly improved the scaffold N50 length for BN-Lx from 6.8 Mb to 36Mb, and SHR/Olapcv from 2.2 to 26Mb. Analysis of Hi-C data can also identify specific features, such as protein binding-mediated "loops" that bring pairs of chromatin domains that are normally far apart into adjacency. These loops often correspond to promoter-enhancer relations. Using HiCCUP, we annotated Hi-C data for four strains and identified 1000-2000 loops per strain. The average linear distance between contacting loci was 913 Kb. We are annotating these interactions to identify associated genes. These data will be useful for the functional annotation of GWAS hits on rat models of substance abuse and understanding gene regulation in PFC.