Submitter Name: Tristan de Jong
Submitted email: Tristan_dejong@hotmail.com
PI Name (if different): Hao Chen
PI email (if different): hchen@uthsc.edu

Towards an infinite genetic map for the rat hybrid diversity panel

Tristan V. de Jong[1], David G. Ashbrook[2], Clifton L. Dalgard[3], Camille Alba[3], Abraham A. Palmer[4], Mary Shimoyama[5], Aron Geurts[6], Melinda Dwinell[6], Michal Pravenec[7], Victor Guryev[8], Yanchao Pan[10], Huda Akil[9], Jun Li[10], Robert W. Williams[2], Hao Chen[1]

[1]Department of Pharmacology, Addiction Science, and Toxicology, [2]Department of Genetics, Genomics and Informatics, University of Tennessee Health Science Center; [3]The American Genome Center, Uniform Service University; [4]Department of Psychiatry, University of California San Diego; [5]Department of Biomedical Engineering, [6]Department of Physiology, Medical College of Wisconsin; [7]Institute of Physiology, Czech Academy of Sciences; [8]Faculty of Medical Sciences, University of Groningen; [9]Molecular and Behavioral Neuroscience Institute; [10]Department of Human Genetics, University of Michigan

Whole genome sequencing and *de novo* assembly of highly diverse genomes greatly facilitates efforts to link genomic variation to disease risk and innovate therapeutic strategies for beneficial outcomes. Despite rapid initial progress, characterization of the rat genome has languished for the past decade. We have begun to systematically sequence the Hybrid Rat Diversity Panel (HRDP)—an extend family of 90 isogenic rat strains that includes 30 classical inbred strains, 30 HXB/BXH recombinant inbred strains, and 30 diallel cross (DX) F1 progeny. The high level of genetic diversity within the HRDP requires state-of-the-science profiling methodologies. Using 10X Chromium linked-reads libraries, we have sequenced 43 strains, including the entire HXB/BXH family. Analysis of 14 genomes has defined more than 1.7 million SNPs, of which ~57% are shared by two or more strains, and 42% by more than seven strains. To ensure the quality of calls, all SNPs with a phred score below 30 or coverage below 15x were ignored. We also discounted SNPs overlapping deletions in BN (the reference genome) that are likely to be errors in the rn6 assembly. Analysis of structural variants is in progress. The availability of sequencing data for the HRDP will greatly facilitate both forward and reverse genetic studies, including many related to substance use disorders and related behaviors. Our currently available infinite marker maps for the HXB/BXH family significantly improves mapping precision for 100s of phenotypes and many large expression data sets.