Telomere-to-Telomere Assembly of the SHRSP/BbbUtx (SHR-A3) Rat

Kai Li[1], Sergey Koren[2], Adam Phillippy[2], Elizabeth Hudson[3], Melissa Laird-Smith[3], Ted, Kalbfleisch[1], Peter Doris[4]

[1] Department of Veterinary Science, University of Kentucky, Lexington, KY 40546

[2] Genome Informatics Section, Center for Genomics and Data Science Research, National Human Genome Research Institute, National Institutes of Health (Bethesda, MD, USA)

[3] Department of Biochemistry and Molecular Genetics, University of Louisville, Louisville, KY, USA

[4] Center for Human Genetics, Brown Foundation Institute of Molecular Medicine, University of Texas McGovern School of Medicine, Houston, TX

It has been recently demonstrated that genomes built using Telomere-to-Telomere (T2T) best practices are more contiguous, more complete, and more accurate that genomes built with less comprehensive methods and source data. Of interest is the fact that T2T assemblies use Oxford Nanopore Technologies ultra-long reads, specifically ~30 fold coverage of reads in excess of 100kb in length that can span, and consequently order genome regions flanking large repeat regions including across centromeres. In earlier work, a reference genome published January 2022 was derived from a SHRSP/BbbUtx male using PacBio HiFi reads, HiC proximity ligation data, and BioNano optical mapping data (HiFi assembly). This genome was comparable in contiguity to the then existing rat reference, but more accurate at the base level. In our preliminary T2T work using the same SHRSP/BbbUtx strain we demonstrate increased contiguity having reduced the number of gaps from 1,610 in the HiFi assembly to only 13 in the T2T assembly capturing the telomeres at the ends of all 20 autosomal and the sex chromosomes. **Also of note, this assembly is derived from tissue and not blood cells and will be unaffected by recombination of germline immune receptor genes.** We are continuing to work to close gaps and improve base level accuracy. Interesting challenges remain such as resolving a clear path across ribosomal DNA arrays on chromosomes 3, 11, and 12. We anticipate resolving these, and other complex, difficult to assemble regions that would have been impossible with other approaches. We will report the progress made along with a target publication date for the updated SHRSP/BbbUtx assembly.