A pangenome for the expanded BXD family of mice

Flavia Villani[1],David G. Ashbrook[1],Andrea Guarracino[2],Simon Heumos[3],Christian Fischer[1], Hao Chen[1],Pjotr Prins[1],Erik Garrison[1] Robert W.Williams[1]and Vincenza Colonna[1,4]

[1]Department of Genetics, Genomics, and Informatics, University of Tennessee Health Science Center, Memphis, TN, United States; [2]Department of Biology, University of Rome Tor Vergata, Rome Italy; [3]Quantitative Biology Center (QBiC), University of Tübingen, Tübingen, Germany; [4]Institute of Genetics and Biophysics "Adriano Buzzati-Traverso", National Research Council, Naples, Italy.

The members of BXDs family have been inbred for 20-200 generations. They are of great value for mapping complex traits and phenome-wide association analysis. Current genomic studies on BXD assume a single linear reference genome, making it difficult to observe sequences diverging from the reference, therefore limiting the accuracy and completeness of analyses. Pangenome models overcome this limitation as they contain the full genomic information of a species.

We sequenced all extant members of all BXD families at ~40X depth using linked-read libraries and constructed the pangenome graph with the PanGenome Graph Builder(PGGB). Variants were called using vg and validated in the DBA/2J mouse using as truth sets a validated VCF derived from 10X data (vcf10X), and a VCF derived from PacBio sequencing (vcfPB).

The BXD pangenome for chromosome 19 is made of 5.2M nodes and 7.9M edges (total length 233.8M bp, 82,695 paths).In DBA/2J we called from the pangenome 218,139 simple microvariants (SNPs, MNPs, and INDELs), that is 45k and 40k more variants compared to vcf10X and vcfPB respectively. From the pangenome we were also able to call 11.5k complex microvariants. In masked regions, precision and sensitivity of vg calls are 89% and 84%, respectively and for SNPs these figures are 92% and 85%.

Overall we were able to reproduce data on known genetic variants and capture novel variation that needs to be validated.We also demonstrate that pangenomes can be accurately built from linked-reads.