Pangenome mapping enhances genotype-phenotype associations in BXD mouse family

Flavia Villani[1], Andrea Guarracino[1], Farnaz Salehi[1], Hao Chen[2], Robert W. Williams[1], Pjotr Prins[1], David G. Ashbrook[1], Erik Garrison[1], Vincenza Colonna[1]

[1]Department of Genetics, Genomics and Informatics, University of Tennessee Health Science Center, Memphis, TN, USA. [2] Department of Pharmacology, Addiction Science, and Toxicology, University of Tennessee Health Science Center, Memphis, TN, USA.

The recombinant inbred BXD family of mice is a valuable resource for mapping complex traits, and they offer a unique insight into genetic traits through epoch-specific segregating variants and rare de novo mutations that accumulate over generations. Traditional genomic analyses of BXDs rely on a single linear reference genome, which limits the detection of sequences diverging from it, thus reducing the range of genetic variation to be tested when mapping quantitative traits. To address this limitation, we implemented a pipeline that maps short-read against the pangenome. This approach aims to mitigate reference bias—the propensity for reads to align preferentially to the reference allele rather than to alternate alleles—and facilitates the identification of novel genetic variations.

As a case study, we focus on mapping a known quantitative trait locus (QTL) for *Cacna1e* gene expression in the BXD mouse population using markers derived from mapping short-read against the pangenome. The *Cacna1e* QTL was previously identified and catalogued in GeneNetwork database alongside millions of other phenotypic traits. We built a pangenome graph of the parental strains C57BL6J (B) and DBA2J (D) using long-read de novo assemblies, and aligned linked-read sequences from all extant BXD family members, projecting the alignments into the pangenome graph. This process produced a node coverage matrix, extending the idea of variation from a linear genome to nodes, or continuous sequences, within the pangenome. We then regress pangenome nodes coverage information over phenotypic variation to rediscover *Cacna1e* QTL on chromosome 1 and identify novel peaks on other genomic regions.

Our findings demonstrate the power of pangenome analysis in uncovering new genotype-phenotype associations that may be missed by traditional linear genome approaches, potentially enhancing our understanding of complex genetic traits in model organisms.