The Association of Structural Variation in Medically Relevant and Complex Genes, Such as GPI on Chromosome 19, With the Production of Novel Gene Transcripts and Its Influence on Gene Expression

Rodriguez, Alejandra (School: Easley High School)

Short-read RNA sequencing is limited by its inability to fully sequence long genetic transcripts, requiring computational reconstruction that often hinders accurate characterization of transcript diversity. Long-read RNA sequencing, such as Isoform sequencing (Iso-Seq), overcomes this limitation by capturing entire transcripts, enabling the identification of complex structural variants and isoforms. This study focuses on glucose-6-phosphate isomerase (GPI), a medically relevant gene on Chromosome 19 linked to disease susceptibility and drug response. Neighboring genes, GARRE1, PDCD2L, and UBA2, were also pulled due to their proximity to GPI and their canonical exons were analyzed. Using 65 human genomes (130 haplotypes) from the Human Genome Structural Variation Consortium Phase 3 (HGSVC3), repeat masking was conducted with RepeatMasker and the defam library. Canonical transcripts and exon sequences for GPI and its neighboring genes were retrieved from the Ensembl database, and a custom repeat-masked exon library was created. Intronic regions between GPI exons 9 and 10 were dissected into k-mers of varying lengths, and consensus sequences were derived. A dark region repeat consensus sequenc was discovered and further analyzed using HMMER, multiple sequence alignment, and network-based approaches. Eight network-based community consensus sequences were identified and reannotated. The analysis revealed 13 unique GPI haplotypes, including a significant structural variant involving a deletion in the haplotype of an individual that disrupted isoform production. This study highlights the utility of long-read RNA sequencing in uncovering structural variants and isoform diversity, advancing our understanding of gene function and its implications for human gene expression.

Awards Won: