Enhanced detection of large indels in diagnostic exome sequencing

Deepali N. Shinde, Layla Shahmirzadi, Dima El-Khechen, Zoe Powis, Cameron Mroske, David J. Salvador, Hsiao-Mei Lu, Kelly D. Farwell Gonzalez, Sha Tang

Ambry Genetics, 15 Argonaut, Aliso Viejo, CA 92656

Diagnostic exome sequencing (DES) has been remarkably successful as both a diagnostic and novel gene discovery tool. It represents one of the most comprehensive tests available and requires a robust bioinformatics pipeline for base calling, sequence alignment, variant calling, annotation, filtering, and prioritization. While single nucleotide substitutions are generally efficiently identified in well-covered exonic regions in DES, accurate mapping of indels, especially those larger than 20 nt, is challenging due to complicated gapped alignment and paired-end sequence inference. There are very limited data on the detection limit of indels of different pipelines based on the current exome sequencing platforms.

Of 153 positive/likely positive cases in the first 500 unselected DES cases referred to Ambry Genetics, 4 (2.6%) were associated with large indels (> 40 nt) up to 115 nucleotides in size. These large indels were identified by Ambry Variant Analysis (AVA) pipeline and confirmed by Sanger sequencing. In the first family, a maternally inherited heterozygous *UBE3A* deletion of 89 nt was detected in a pediatric patient with a differential diagnosis of Angelman syndrome (AS). It is presumed that the asymptomatic mother's deletion occurred on her paternally inherited allele, which was inactivated due to imprinting. In a second pediatric patient with suspected arthrogryposis, AVA detected a *de novo* heterozygous 115 nt deletion in the *MYH3* gene. In a pediatric patient with recurrent infections, immunodeficiency and significantly decreased expression of CD127, compound heterozygosity of a splice site mutation and a 58 nt deletion in the *RFXANK* gene was observed. Lastly, a homozygous 41 nt indel in the *PDE6B* gene caused by a 47 nt deletion coupled with 6 nt insertion was detected in a 14 year old patient with retinitis pigmentosa consistent with the gene finding.

In conclusion, although detection of large indels is inherently difficult for DES, indels larger than 40nt account for 2.6% of the positive cases in our cohort and indels up to 100 nt can be identified by AVA. Our data highlight the importance of an optimized bioinformatics pipeline for the detection of large indels to improve clinical sensitivity and diagnostic yield.