Enhanced Detection of Large Indels in Diagnostic Exome Sequencing

Deepali N. Shinde[1], Layla Shahmirzadi[1], Dima El-Khechen[1], Zoe Powis[1], Cameron Mroske[1], David J. Salvador[1], Hsiao-Mei Lu[1], Kelly D. Farwell Gonzalez[1], Sha Tang[1]

1. Ambry Genetics, 15 Argonaut, Aliso Viejo, CA 92656

BACKGROUND

- Exome sequencing has been remarkably successful as both a diagnostic and novel gene discovery tool.
- Diagnostic Exome Sequencing (DES) provides answers where traditional diagnostic methods have been uninformative in discovering the pathogenic etiology in patients.
- Two types of DNA sequence polymorphisms are reported following DES of patient samples – single nucleotide variants (SNVs) and insertion-deletions (indels). A robust bioinformatics pipeline is required for base calling, sequence alignment, variant calling, annotation, filtering, and prioritization.
- While SNVs are generally efficiently identified in well-covered exonic regions, accurate mapping of indels, especially those larger than 20 nucleotides, is challenging due to gapped alignment and paired-end sequence inference.
- Very limited data on the indel detection limit of different pipelines based on current exome sequencing platforms are available.
- We have developed and validated a robust bioinformatics pipeline – the Ambry Variant Analyzer (AVA) – that accurately and efficiently calls indels larger than 20 nucleotides from next generation sequencing (NGS) data and contributes to one of the highest diagnostic yields reported for clinical exome analysis.

AMBRY EXOME SEQUENCING WORKFLOW

DNA Isolation

Genomic DNA isolation from whole blood of patients and relatives.

Informed consent obtained from all family members being tested.

Exome Capture & Sequencing

Exome capture with SureSelect Target Enrichment System (Agilent Technologies, Santa Clara, CA) or SeqCap EZ VCRome 2.0 (Roche NimbleGen, Madison, WI).

Paired-end, 100-cycle exome sequencing on Illumina HiSeq 2000 (Illumina, SanDiego, CA).

Variant Filtering

Search previously described gene mutations and polymorphisms.

Removal of common SNPs, intergenic, 3'/5' UTR, non-splice related intronic & synonymous variants.

Variant filtration based on family history and possible inheritance models.

Analysis

Manual interpretive filtering by molecular geneticist.

Familial co-segregation analysis by Sanger sequencing.

AVA VARIANT DETECTION AND FILTERING

Raw Data Analysis
- Image Processing & Base Calling - NGS FASTQ file with paired-end reads

Whole Genome Mapping
- Alignment to reference genome GRCh37 using CASAVA – BAM/SAM files

Variant Calling
- Differences between sample and reference genomes – SNVs/INDELs called using CASAVA and Pindel

Annotate Variants
- Annotate using AVA - Gene, Isoform, c. and p. mutation, dbSNP, Population Frequencies (1000Genomes, ESP), Disease Information (HGMD), and predicted functional consequences (SIFT, PolyPhen)

100,000 - 200,000 annotated variants per individual in trio

Filter alterations outside the coding region (+/- 2)

Filter non-splice related synonymous alterations

***FIND**: **F**amily Hx **In**heritance-based **D**etection* – specific MAF filtering

100-200 variants

Protects alterations with HGMD or OMIM alteration ID

Protects common founder mutations and alterations classified at Ambry as mutation or VLP

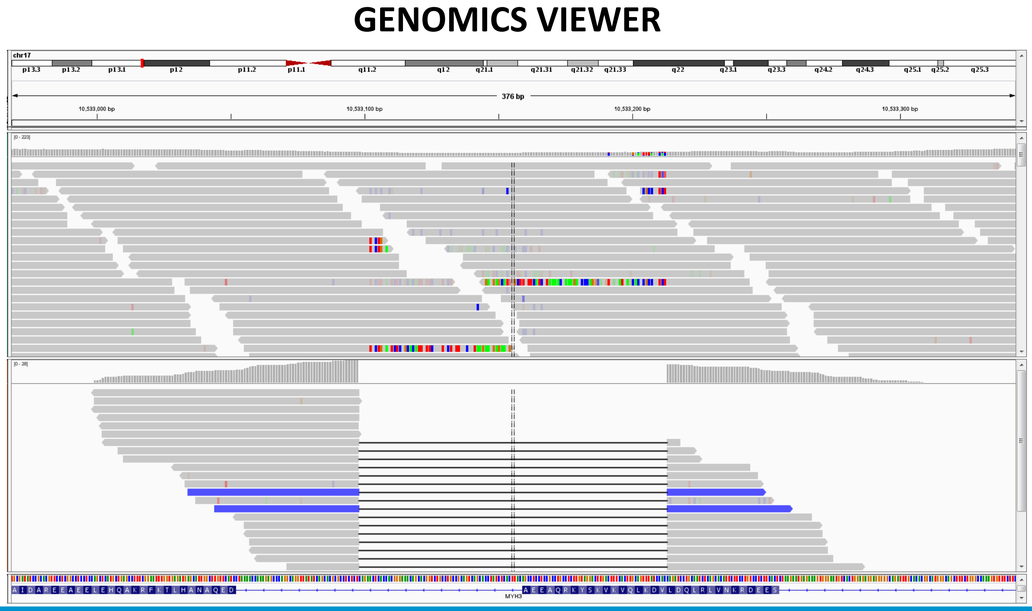
RESULTS

Gene	Mutation detected	Size of deletion	Zygosity	Reads	Q score	Confirmed by Sanger
UBE3A	c.1254_1342del89	89 nt	het	4/36	105	YES
MYH3	c.5605_5659-47del115	115 nt	het	17/69	719	YES
RFXANK	c.416_438+35del58	58 nt	het	8/32	296	YES
PDE6B	c.1923_1969ins6del47	41 nt	homo	5/5	17	YES

nt = nucleotides het = heterozygous homo = homozygous

- 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.
- 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.
- 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.

VISUALIZATION OF *MYH3* c.5605_5659-47del115 USING THE INTEGRATIVE GENOMICS VIEWER



DISCUSSION

- Although detection of large indels is inherently difficult for DES, indels larger than 40 nt account for 2.6% of the positive cases in our cohort.
- Theoretically, deletions up to 200 nucleotides and insertions up to 100 nucleotides 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.

REFERENCES

- 1000 genomes database (2010) A map of human genome variation from population-scale sequencing. *Nature* **467**: 1061–1073.
- CASAVA 1.8.2 "Consensus Assessment of Sequence And VAriation": Illumina Inc., San Diego, CA.
- dbSNP: Sherry ST, *et al.* (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* **29**,:308-311.
- ESP: NHLBI Exome Sequencing Project's Exome Variant Server. World Wide Web URL: http://evs.gs.washington.edu/EVS/
- HGMD: Stenson PD, *et al.* (2009) The Human Gene Mutation Database: 2008 update. *Genome medicine* **1** :13.
- Integrative genomics Viewer: Thorvaldsdóttir H., *et al.* (2012) *Brief in Bioinform*; Robinson J.T., *et al.* (2011) *Nat Biotech* **29**: 24–26.
- International HapMap (2003) The International HapMap Project. *Nature* **426**:789-796.
- Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD) World Wide Web URL: http://omim.org/
- Pindel: Ye K., *et al.* (2009) *Bioinformatics* **25**: 2865–2871.
- PolyPhen: Adzhubei IA, *et al.* (2010) *Nat Methods* **7**: 248–249.
- SIFT: Ng PC & Henikoff S (2006) *Hum Genet.* **7**: 61–80.