Blueprint Genetics

Analytic Validation of Whole Exome Sequencing for Clinical Diagnostics of Inherited Disorders

Mikko Muona, Massimiliano Gentile, Juha W Koskenvuo, Samuel Myllykangas
Blueprint Genetics, Biomedicum 2U, Tukholmankatu 8, Helsinki, 00290, Finland

Analytic validation of SNV and indel detection in Whole Exome Sequencing assay shows high sensitivity and specificity

Utility of Whole Exome Sequencing (WES) in clinical diagnostics has been limited by the non-uniform sequencing coverage across exonic regions, leaving typically 5-10% of the regions with limited sequencing coverage. We have aimed to develop a WES assay of which breadth of coverage would approach that of high-coverage gene-panel based assays, leading to high sensitivity in variant detection.

We have performed analytic validation of the WES assay, which utilizes Agilent SureSelect Human All Exon V6 hybrid capture selection to capture 60 Mb of exonic targets. Sequencing was performed using an Illumina HiSeq sequencing system. Analytic validity of the WES assay was demonstrated for SNVs and indels using reference samples from The Genome In a Bottle Consortium with high-quality variant calls (**Table 1**). The validation showed clinical-grade performance in all assessed metrics.

Table 1. Analytic validation of SNV and indel detection in WES.

Performance metric	Value	Measurements
Accuracy (SNVs)	0.999	TN: 326,571,803
Sensitivity (SNVs)	0.995	TP: 152,827
Specificity (SNVs)	0.999	FP: 896
Positive predictive value (SNVs)	0.994	FN: 791
Sensitivity (1-10 bp indels)	0.973	TP/FN: 6033 / 169
Sensitivity (11-20 bp indels)	0.957	TP/FN: 356 / 16
Sensitivity (21-30 bp indels)	0.970	TP/FN: 162 / 5
Sensitivity (>= 31 bp indels)	1.000	TP/FN: 10 / 0
Nucleotides with >=15x sequencing depth	98.9%	
Mean sequencing depth at nucleotide level	148x	
Reportable range (SNVs)	Hom, Het	
Reportable range (indels)	0-35 bp	
Repeatability	0.994	

High and uniform sequencing coverage combined with thorough bioinformatics analysis of the complete WES data leads to high discovery rate

Clinical-grade WES assay translates into a high diagnostic yield: the genetic diagnosis was established in >40% of cases (**Figure 1**).

Key factors maximizing the diagnostic yield in a WES assay:

1) Deep and uniform sequencing coverage across all protein coding genes (Figure 2)
   - ➢ In production samples, average sample specific mean coverage 158x with 98.9% of coding regions targeted >=15x

2) Taking advantage of the complete WES data using a 'genotype first' approach
   - ➢ Analysis not limited to strict phenotype-driven gene lists – WES data analysed primarily for all >3,500 clinically associated genes
   - ➢ Genotype first approach allows identification of atypical gene-phenotype presentations
   - ➢ Constantly updated gene-phenotype databases allow diagnosis of recently established genetic disorders

3) Cases who remain without diagnosis after analysis of established disease genes are analysed for candidate variants in genes that have not been associated with human disease
   - ➢ E.g. truncating in variants in genes predicted to be tolerant for loss-of-function variation are reported

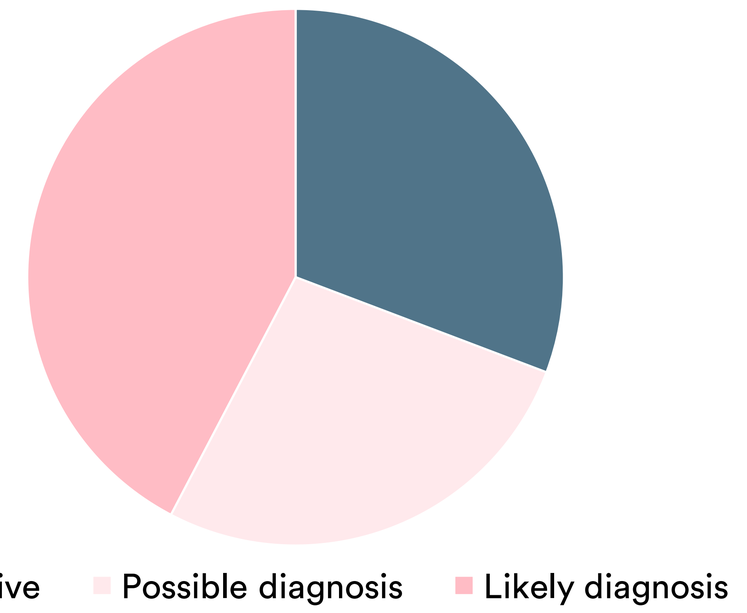
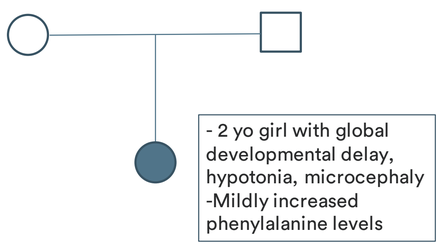


Figure 1. Diagnostic yield in initial 26 BpG WES cases. 42% of cases obtained a likely diagnosis, 27% a possible diagnosis, and 31% remained negative. Of the negative cases, we reported candidate variants in 38% individuals

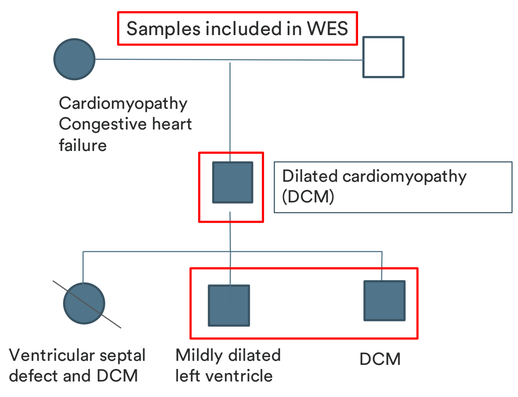
Case 1: Dual diagnosis



Genetic diagnosis #1
De novo ZBTB18 c.1390C>T, p.(Arg464Cys)
- De novo *ZBTB18* mutations recently linked to intellectual disability with variable features
- The same variant described earlier in two patients

Genetic diagnosis #2
Homozygous *PAH* c.1139C>T, p.(Thr380Met)
- Known variant for non-phenylketonuria hyperphenylalaninemia

Case 2: Atypical gene-phenotype association



Heterozygous GATA4 c.838A>T, p.(Thr280Ser) variant in affected individuals that were subjected to WES
- GATA4 is a transcription factor that is essential for heart function
- GATA4 heterozygous variants cause various CHDs, and is only recently also associated with DCM

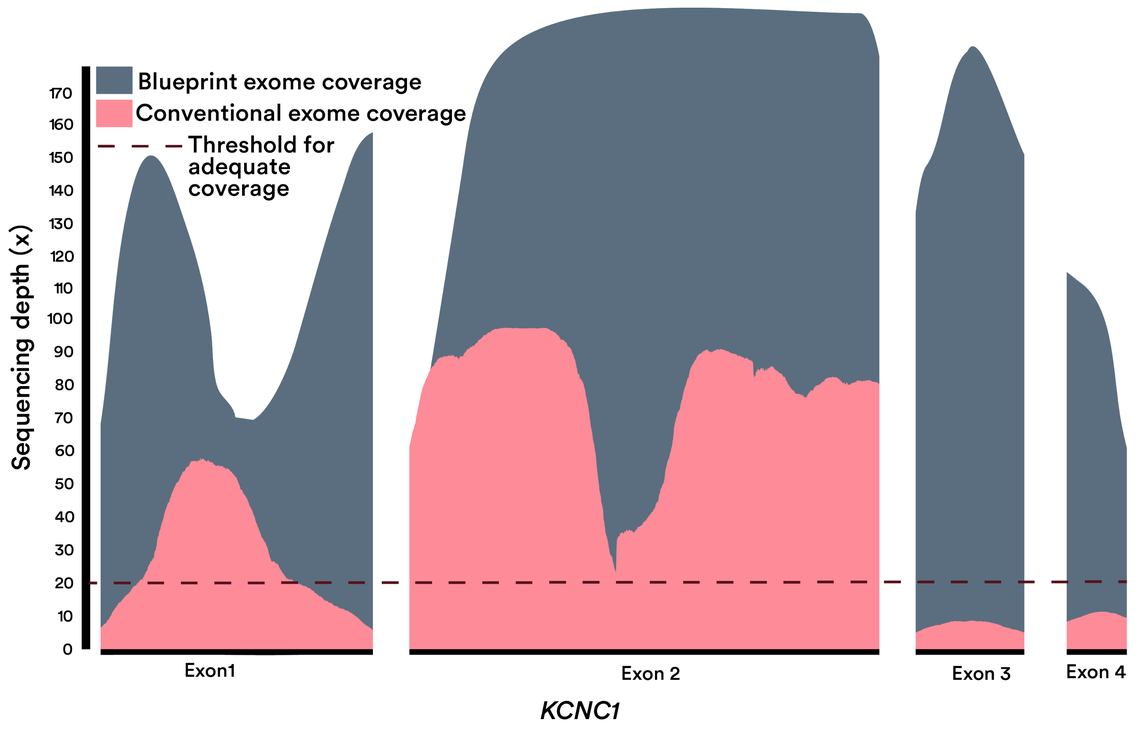


Figure 2. Deep and uniform sequencing depth is required for reliable detection of disease-causing variants across all clinically relevant genes. The above figure illustrates how deep, clinical-grade sequencing coverage minimizes gaps in the clinically relevant genes such as *KCNC1*. On the contrary, conventional exomes suffered from low coverage, risking detection of disease-causing variants. Coverage data for 'Blueprint exome' represent data from our assay validation and data for 'Conventional exome' represent mean coverages of the corresponding genes in the ExAC database.

Conflict of interest statement: All authors are employed by Blueprint Genetics.