The Occurrence of Noncoding Variants, Copy Number Variants, Mosaic Variants and Variants in Technically Challenging Genes in Over 10,000 Whole Exome Sequencing Tests

Kimberly Gall,[1] Julie Hathaway,[1] Alicia Scocchia,[1] Victoria Howell,[1] Allison Sluyters,[1] Inka Saarinen,[2] Tiia Kangas-Kontio,[2] Milja Kaare,[2] Kirsty Wells,[2] Maria Calvo del Castillo,[2] Mikko Muona,[2] Tuuli Pietila,[2] Matias Rantanen,[2] Massimiliano Gentile,[2] Pertteli Salmenperä,[2] Jussi Paananen,[2] Samuel Myllykangas,[2] Juha Koskenvuo[2]

[1]Blueprint Genetics Inc, Seattle, WA, USA; [2]Blueprint Genetics, Espoo, Finland

Abstract number: P465

Introduction

Whole genome sequencing (WGS) can detect variants in noncoding regions, copy number variants (CNVs), and variants in technically challenging genes (those with sequence homology or repetitive sequences). However, WGS is not optimal for detecting mosaic variants. Further, WGS is not yet routinely accepted in clinical practice. However, whole exome sequencing (WES) is currently offered routinely, and WES assays optimized for the detection of these variants have been developed, but the occurrence of such variants in an unselected referral population undergoing WES is not well understood. To better understand their impact on diagnostic testing with whole exome sequencing, we assessed the positive rate of our WES assay with particular attention to the contribution of known disease-causing noncoding variants, CNVs, mosaic variants, and variants identified in technically challenging genes.

Methods

We retrospectively examined results from patients tested consecutively with a WES assay at Blueprint Genetics. All patient samples were submitted for the investigation of one or more clinical findings. The WES assay target region included all coding exons (including +/-20 base pairs {bp} from intron/exon boundaries) and was customized to include 1,501 noncoding (regulatory or >10 base pairs at the intron exon boundary) variants classified as pathogenic (P) or likely pathogenic (LP) in ClinVar or as a disease-causing mutation (DM) in the Human Gene Mutation Database (HGMD). CNV analysis was performed bioinformatically from next-generation sequencing (NGS) data using 2 different pipelines, including a proprietary pipeline specifically developed for the detection of small, single exon deletions.

Methods (continued)

Variants that did not meet stringent internal quality metrics were confirmed with an orthogonal method.

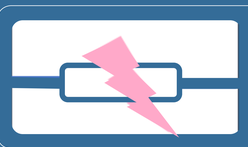
Patient sex, age at testing, and clinical history were collected from test requisition forms, while the genetic test results were extracted from the internal interpretation and reporting database. A positive result was defined as the identification of P or LP variant(s) (classified using a modified ACMG/AMP variant classification scheme) consistent with all, or part, of the patient's reported phenotype and the expected inheritance.

Results

A total of 10,078 patients underwent testing over the evaluated period; 65.4% (n=6,589) were tested as singletons and 34.6% (n=3,489) were tested with one or more family members. Of those tested, 44.9% (4,530) were female, 54.1% (5,449) were male and sex was not reported for 1.0% (99). Age at testing ranged from 0 to 88 years (median age of 2 years). Most samples were received from patients aged 0 to 18 years (72.6%, n=7,321) (Table). Validation of the WES assay demonstrated 99.4% of the target region was covered at 20X with a mean sequencing depth of 174X. Single-exon deletions were detected at >92% sensitivity, and 14.6% lowest allele fraction was detected with 90% sensitivity.

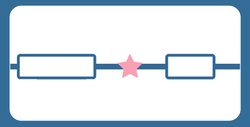
A positive result, responsible for all or part of the reported clinical history, was reported for 29.0% (n=2,923) of patients. The positive result rate was higher for patients tested with family members (32.1%, 1,121/3,489) versus index patients alone (27.3%, 1,802/6,589).

For patients undergoing WES, 5.7% had a positive result due to:



Variants in technically challenging genes

•(2.0%, n=57)

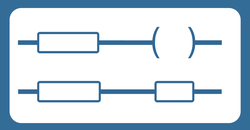


Known disease-associated noncoding variants

•(1.1%, n=32)

Mosaic Variants

•(1.0%, n=28)



CNVs <4 exons in size

•(1.6%, n=48)

Table. Demographics.

	Total Cohort		Positive Cohort	
	# (n=10,078)	% total cohort	# (n=2,923)	% positive rate
Sex				
Male	5,449	54.1%	1,513	27.8%
Female	4,530	44.9%	1,381	30.5%
Not reported	99	1.0%	29	29.3%
Age range				
Fetus	363	3.6%	84	23.1%
0–2 years	2,349	23.3%	826	35.2%
3-10 years	3,136	31.1%	901	28.7%
11-18 years	1,836	18.2%	555	30.2%
19-40 years	1,464	14.5%	397	27.1%
41-88 years	923	9.2%	157	17.0%
Age not provided	7	<0.1%	3	42.9%
Analysis				
Proband Only	6,589	65.4%	1,802	27.3%
Family	**3,489**	**34.6%**	**1,121**	**32.1%**

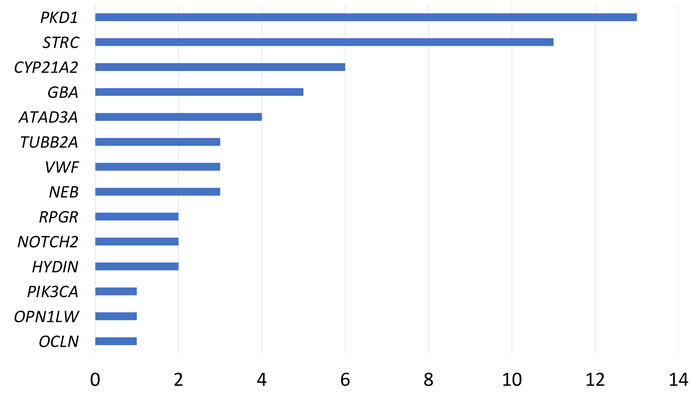


Figure. Number of positive results found in technically challenging genes.

Results (continued)

Positive results occurred in 1,051 unique genes; 348 genes were reported as a positive finding only once. Ten genes accounted for >12% of the positive results (in decreasing order from the most common: *ANKRD11, MECPS, PTPN11, ARID1B, TTN, SCN1A, KMT2A, PRRT2, PTEN, and FLG*).

Variants in noncoding regions (greater than 10 base pairs beyond the intron/exon boundary) were responsible for 1.1% (32/2923) of positive results.

CNVs contributed to 15.3% (448/2,923) of all positive results, of which 10.7% (48/448) were <1,000 bp in size. The smallest CNV reported was 128 bp in size and the largest was an entire chromosome.

Variants in genes with regions of >90% sequence homology accounted for 12.4% (362/2,923) of positive results while variants in the most technically challenging genes were responsible for 2.0% (57/2,923) of positive results (Figure 1).

Finally, a mosaic result (lowest allelic fraction 9%) was reported in 1.0% (28/2,923) of positive results.

Conclusions

- WES that included family members in the analysis resulted in a higher positive rate than testing only the proband at least in part due to easier identification of *de novo* variants and the ability to phase variants identified in autosomal recessive genes.
- Patients who were tested between 0 to 2 years had the highest positive rate (35.2%), possibly due to a higher *a prior* risk of having a genetic etiology for their symtpoms. However, even patients tested as older adults had a 17% positive rate demonstrating that WES had clinical utility across the lifespan.
- Positive results included variants in >1,000 genes with 1/3 of the positive genes only reported once highlighting the value of WES when the differential is broad.
- WES assays optimized to detect technically challenging variants may not completely overcome the limitations of short read NGS, however, greater than 1/20 patients who received a positive result had a diagnostic noncoding variant, small CNV, mosaic variant or variant in a technically challenging gene demonstrating the value of using such assays.

Blueprint Genetics

Conflict of interest statement: All authors are employees of Blueprint Genetics.