

RESEARCH ARTICLE

# Exome sequencing covers >98% of mutations identified on targeted next generation sequencing panels

Holly LaDuca<sup>1\*</sup>, Kelly D. Farwell<sup>1\*</sup>, Huy Vuong<sup>1</sup>, Hsiao-Mei Lu<sup>1</sup>, Wenbo Mu<sup>1</sup>, Layla Shahmirzadi<sup>1</sup>, Sha Tang<sup>1</sup>, Jefferey Chen<sup>1</sup>, Shruti Bhide<sup>1</sup>, Elizabeth C. Chao<sup>2</sup>

**1** Department of Clinical Diagnostics, Ambray Genetics, Aliso Viejo, California, United States of America,

**2** Department of Pediatrics, Division of Genetics and Genomic Medicine, University of California Irvine, Irvine, California, United States of America

\* [hladuca@ambrygen.com](mailto:hladuca@ambrygen.com) (HL); [kfarwell@ambrygen.com](mailto:kfarwell@ambrygen.com) (KDF)



**OPEN ACCESS**

**Citation:** LaDuca H, Farwell KD, Vuong H, Lu H-M, Mu W, Shahmirzadi L, et al. (2017) Exome sequencing covers >98% of mutations identified on targeted next generation sequencing panels. PLoS ONE 12(2): e0170843. doi:10.1371/journal.pone.0170843

**Editor:** Noam Shomron, Tel Aviv University, ISRAEL

**Received:** September 1, 2016

**Accepted:** January 11, 2017

**Published:** February 2, 2017

**Copyright:** © 2017 LaDuca et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** Ambray Genetics provided support in the form of salaries for all authors, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** All authors were full time paid employees of Ambray Genetics at the time the study was conducted. This does not alter our

## Abstract

### Background

With the expanded availability of next generation sequencing (NGS)-based clinical genetic tests, clinicians seeking to test patients with Mendelian diseases must weigh the superior coverage of targeted gene panels with the greater number of genes included in whole exome sequencing (WES) when considering their first-tier testing approach. Here, we use an *in silico* analysis to predict the analytic sensitivity of WES using pathogenic variants identified on targeted NGS panels as a reference.

### Methods

Corresponding nucleotide positions for 1533 different alterations classified as pathogenic or likely pathogenic identified on targeted NGS multi-gene panel tests in our laboratory were interrogated in data from 100 randomly-selected clinical WES samples to quantify the sequence coverage at each position. Pathogenic variants represented 91 genes implicated in hereditary cancer, X-linked intellectual disability, primary ciliary dyskinesia, Marfan syndrome/aortic aneurysms, cardiomyopathies and arrhythmias.

### Results

When assessing coverage among 100 individual WES samples for each pathogenic variant (153,300 individual assessments), 99.7% (n = 152,798) would likely have been detected on WES. All pathogenic variants had at least some coverage on exome sequencing, with a total of 97.3% (n = 1491) detectable across all 100 individuals. For the remaining 42 pathogenic variants, the number of WES samples with adequate coverage ranged from 35 to 99. Factors such as location in GC-rich, repetitive, or homologous regions likely explain why some of these alterations were not detected across all samples. To validate study findings, a similar analysis was performed against coverage data from 60,706 exomes available through the Exome Aggregation Consortium (ExAC). Results from this validation confirmed

adherence to PLOS ONE policies on sharing data and materials.

that 98.6% (91,743,296/93,062,298) of pathogenic variants demonstrated adequate depth for detection.

## Conclusions

Results from this *in silico* analysis suggest that exome sequencing may achieve a diagnostic yield similar to panel-based testing for Mendelian diseases.

## Introduction

With the expanded availability of next generation sequencing (NGS)-based clinical genetic tests, clinicians are faced with the decision to pursue targeted gene panels versus whole exome sequencing (WES) as their first-tier testing approach [1]. This decision may be particularly challenging for diseases with significant genetic and phenotypic heterogeneity. It is important to consider the benefits and limitations of each approach when deciding on the best testing strategy.

Coverage is superior on NGS panels compared to WES when the amount of sequenced nucleotides is the same. In most cases, sequence-specific enrichment is performed prior to NGS to help achieve high coverage and reduce or eliminate low-coverage regions (typically ~10-50X). Commonly, Sanger sequencing is applied to regions which are recalcitrant to NGS enquiry, due to technical or biological limitations, including GC-rich regions and regions of high homology. In contrast, current estimates of coverage achieved from whole exome capture and sequencing are 90–95% at >20X, with factors such as target enrichment design, off-target capture, repetitive and GC- or AT-rich regions, copy-number variations, and structural variations posing challenges to complete capture [2–5]. Unlike NGS panels, the addition of Sanger sequencing for regions with inadequate coverage would not be time or cost-effective for WES, where more than 20,000 genes are being analyzed.

In contrast, one key drawback of targeted NGS panels is that they may become outdated rather quickly. During the time a panel is developed and validated for clinical utilization, new studies are already published identifying newly characterized disease genes. In fact, among positive WES findings, 23% are within genes characterized within the last two years and 7% are novel gene discoveries [6]. As such, an advantage of WES is the ability to sequence the entire exome at once, allowing for the analysis and interpretation of all alterations in both well characterized and novel genes, and also allowing for re-interpretation as new genetic associations are established. Additional advantages of an exome sequencing approach include the ability to analyze a significantly larger number of genes at a reasonable cost, the potential to identify novel genes, and the ability to sequence the exomes of multiple family members simultaneously in labs that offer sequencing of trios.

Early reports comparing targeted NGS panels to WES focused on exon or gene level coverage, as related to a specific gene or set of diagnostic genes [7]. Here, we aim to calculate the analytic sensitivity of WES for pathogenic variants identified on targeted NGS panels.

## Materials and methods

The internal database at Ambry Genetics (Aliso Viejo, CA) was queried for all alterations classified as pathogenic and likely pathogenic (herein referred to collectively as ‘pathogenic variants’) detected on germline targeted NGS multi-gene panel testing of over 50,000 patients in the clinical diagnostic laboratory from April 2010 until July 2014. Variants underwent

thorough assessment and review of available evidence and were classified using a five-tiered classification algorithm [8] based on guidelines from the American College of Genetics and Genomics and the International Agency for Research on Cancer [9, 10]. Pathogenic variants in this study included single nucleotide substitutions and small insertion and deletion events up to 40 basepairs. Gross rearrangements, large insertions, and large deletions detected on array or multiplex ligation-dependent probe amplification-based assays were excluded from analysis. Multi-gene panels targeted a range of hereditary (Mendelian) disorders including cancer susceptibility, X-linked intellectual disability (XLID), primary ciliary dyskinesia (PCD), Marfan syndrome, thoracic aortic aneurysms and dissections, and related disorders (Marfan/TAAD), and other cardiovascular diseases such as cardiomyopathies and arrhythmias, and were selected by the ordering clinician based on the patients' clinical history. Corresponding nucleotide positions for these pathogenic variants were interrogated against data from 100 randomly-selected patients whose samples were submitted to Ambry Genetics for clinical WES, and the coverage at each position was assessed. Pathogenic variants were interpreted as having adequate depth for detection if coverage at the respective nucleotide position on WES was  $\geq 10X$ . Coverage at the first and last nucleotides was averaged for insertions, and for deletions and indels coverage was assessed for the first and last nucleotides, with the lower of the two values being used for analysis. Since all data was accessed anonymously, this research was determined to be exempt from review based on 45.CFR46.101 and is in compliance with the Helsinki Declaration (Solutions Institutional Review Board, Reference Number 1OCT14-93).

### NGS panel sequencing

Library preparation, sequencing, bioinformatics, and data analysis were performed as previously described [8, 11]. Briefly, samples were enriched for sequence targets using Raindance Thunderstorm technology (RainDance Technologies, Billerica, MA), and sequenced using paired-end, 100-cycle chemistry on the Illumina HiSeq 2000 (Illumina, San Diego, CA). Importantly, PCR duplicated sequences were removed from the dataset prior to alignment, and quantification of basepair level coverage. The sequence data were aligned to the reference human genome (GRCh37) and variant calls were generated using CASAVA and Pindel [12]. No/low-coverage regions (i.e.  $< 50X$ ) and variant calls other than known non-pathogenic alterations were analyzed using automated fluorescence dideoxy sequencing.

### Whole exome sequencing

Exome library preparation, sequencing, bioinformatics, and data analysis were performed as previously described [6, 13, 14]. Briefly, samples were prepared using SeqCap EZ VCRome 2.0 (Roche NimbleGen, Madison, WI) and sequenced using paired-end, 100-cycle chemistry on the Illumina HiSeq 2000 (Illumina, San Diego, CA). The sequence data were aligned to the reference human genome (GRCh37) and variant calls were generated using GATK and CASAVA. Variant filtering on WES was performed using a custom bioinformatics pipeline as previously described in detail [6]. Briefly, stepwise filtering included the removal of variants with quality scores  $< 20$  and allele counts  $< 10X$ , common SNPs, intergenic and 3'/5' UTR variants, non-splice-related intronic variants, and synonymous variants. Variants were filtered further based on family history and possible inheritance models. Data were annotated with the Ambry Variant Analyzer tool (AVA) [8]. All samples were required to meet minimum quality standards, with at least 90% of bases covered at  $\geq 10X$  and having base call quality scores  $\geq Q20$ , which translates to a base-calling error rate of 1:100. Identified candidate alterations were confirmed using automated fluorescence dideoxy sequencing. All gene/bases covered in gene panel design were covered at  $> 20X$  in the exome design.

## Results and discussion

A total of 1533 different pathogenic variants identified on targeted NGS multi-gene panel testing were included in this analysis, representing 91 genes implicated in 5 disease categories identified through analysis of greater than ~100,000 alleles (S1 Table). Pathogenic variants in cancer susceptibility genes accounted for 88.1% of pathogenic variants analyzed (n = 1350), with each of the other disease categories accounting for <4% of pathogenic variants studied. Exonic single nucleotide substitutions were the most common type of variant included in this analysis (n = 665, 43.4%), followed by small deletions (n = 485, 31.6%), intronic variants (n = 184, 12.0%), small duplications and insertions (n = 169, 11.0%), and indels (n = 30, 2.0%).

Considering that coverage was assessed among 100 individual WES samples for each pathogenic variant (153,300 individual assessments), adequate depth for variant detection was observed for a total of 99.7% (n = 152,798) of pathogenic variants (Table 1). The percentage of pathogenic variants with adequate depth for detection on WES was highest among Marfan/TAAD (99.8%) and lowest among XLID (98.5%) related gene panels. A total of 97.3% (n = 1491) of the pathogenic variants demonstrated adequate depth for detection across all 100 WES samples. The percentage of pathogenic variants with adequate depth for detection across all 100 WES samples was highest among PCD (98.2%) and lowest among XLID (73.9%) related genes. Of the diseases included in this analysis, genes involved in XLID represented the smallest number of pathogenic variants (n = 23). The lower proportion of pathogenic variants with adequate depth for detection for XLID may be a result of this small sample size as well as fewer analyzed alleles in that male subjects only have one copy of the X chromosome. The average percentage of bases covered  $\geq 10X$  for the 100 WES samples was 94.8% (range 92.9–96.0), and the average depth per sample was 94X (range 80X–114X) (S2 Table). Furthermore, among these 100 WES samples, 98% bases were covered  $> 20X$ , 48% bases were covered  $> 100X$  and 0% of bases had no coverage.

All pathogenic variants had at least some coverage on exome sequencing; however, there were 42 pathogenic variants (2.7%) that were covered  $< 10X$  in at least one of the 100 WES samples. Subsequent review of these 42 pathogenic variants revealed that 11 alterations (26.2%) were in GC-rich regions (defined as GC-content  $> 60\%$ ), 8 were in repetitive regions (19.0%) (defined as polymer stretching  $\geq 9$  basepairs), and 3 (7.1%) were located in regions with known pseudogene interference. For the remaining 20 pathogenic variants (47.6%), there was no obvious explanation for no/low exome coverage at the respective nucleotide position (Table 2). The pathogenic variant with the lowest level of coverage- c.325DELG (p.E109Kfs\*3) in *PMS2*, a gene with high pseudogene homology- was detectable in 35 of the 100 WES samples. Low inter-sample variability was observed regarding the number of pathogenic variants not covered at  $\geq 10X$ . Across all 100 exomes, the median number of variants lacking adequate coverage ( $< 10X$ ) in each sample was 5 (range 0–12). Coverage for all pathogenic variants per sample is provided in S3 Table.

The lengthiest alterations assessed in this dataset were a 40-nucleotide deletion in *BRCA1*, and a 20-nucleotide duplication in *BARD1*. Since this study was based on coverage analysis at respective nucleotide positions and did not directly assess the performance of the exome hybridization or the alignment and variant calling algorithms to detect these deletions and duplications, it is difficult to know whether these would truly have been detected. However, based on a retrospective analysis of 500 WES cases performed at Ambry Genetics indels larger than 40 nucleotides accounted for 2.6% of positive results [6], demonstrating that such alterations are detectable by WES at our laboratory.

To validate study findings, a similar analysis was performed against coverage data from 60,706 exomes available through the Exome Aggregation Consortium (ExAC) [15]. The ExAC

**Table 1. Exome sequencing coverage by disease type.**

Disease	Pathogenic variant positions with $\geq 10X$ coverage across all 100 WES samples			Total pathogenic variant positions with $\geq 10X$ coverage in each WES sample		
	Positions covered, n	Positions assessed, n	% Positions covered	Positions covered, n	Positions assessed, n	% Positions covered
Cancer Susceptibility	1319	1350	97.7%	134,584	135,000	99.7%
Cardiovascular	51	53	96.2%	5280	5300	99.6%
Marfan/TAAD	50	52	96.2%	5187	5200	99.8%
PCD	54	55	98.2%	5481	5500	99.7%
XLID	17	23	73.9%	2266	2300	98.5%
All	1491	1533	97.3%	152,798	153,300	99.7%

TAAD, thoracic aortic aneurysms and dissections; PCD, primary ciliary dyskinesia; XLID, X-linked intellectual disability; WES, whole exome sequencing.

doi:10.1371/journal.pone.0170843.t001

database was queried for the percentage of samples with coverage  $\geq 10X$  at the first nucleotide position for each pathogenic variant. Considering that coverage was assessed among 60,706 individual ExAC WES samples for each alteration (93,062,298 individual assessments), a total of 98.6% (n = 91,743,296) of pathogenic variants demonstrated adequate depth for detection (S4 Table). A total of 86.2% (n = 1321) of the pathogenic variants demonstrated adequate depth for detection in  $\geq 99\%$  (n = 60,099) samples. Twenty-five percent (388/1533) of pathogenic variants in this analysis were reported in ExAC, demonstrating the ability of WES to detect at least a portion of these alterations in actuality.

To further compare the analytic sensitivity of targeted panels to WES, our internal database was queried for patients who underwent previous targeted panel testing, either through Ambry Genetics or an outside laboratory, prior to WES at Ambry. Sixteen patients had a total of 21 alterations detected on targeted panel testing, all of which were detected on WES at Ambry (S5 Table). Though limited in size, this dataset also demonstrates the ability of WES to equally detect alterations reported on targeted panel testing.

Recently, Park *et al.* investigated the performance of exome sequencing for the 56 genes in the American College of Genetics and Genomics' incidental finding recommendation by determining coverage at the nucleotide positions for all 18,336 nucleotide variants annotated in HGMD for these genes [16]. Authors identified inadequate coverage for the majority of variants in 7 genes (*SDHC*, *SDHD*, *GLA*, *TGFBR2*, *COL3A1*, *PMS2*, and *PCSK9*) and also identified six GC-rich exons with a high failure rate in their coverage analysis of 12 clinical exomes. In this study, adequate depth for detection was observed for 100% pathogenic variants in five of these genes: *SDHC* (n = 3), *SDHD* (n = 3), *GLA* (n = 1), *TGFBR2* (n = 4), and *COL3A1* (n = 3). *PMS2* is known to harbor several homologous exons; thus, WES is not expected to yield accurate results for this gene. While *SDHC* and *SDHD* also harbor homologous exons, alterations in these genes that were included in our dataset were not located in exons with significant homology. There were not any *PCSK9* alterations included in our dataset, as this gene is not analyzed as part of any targeted NGS panels at our laboratory. A limitation to the Park *et al.* data, as pointed out by the authors, is that only 7.5% of 'disease causing' variants in HGMD are actually pathogenic/likely pathogenic [17]. While our study reported on a lesser number of alterations, they are classified based on our clinical laboratory standards to be pathogenic or likely pathogenic [8]. The vast majority of benign variants are single basepair substitutions which are well represented in this data set, along with insertions and deletions. The set of pathogenic variants is likely to provide the most complete and heterogeneous representation of variant subtypes including increased representation of variants in critical domains which

**Table 2. Pathogenic variant positions not covered across all 100 WES samples.**

Gene	Nucleotide variant	Protein variant	Coding exon/ intron	Exon %GC content	Intron %GC content <sup>a</sup>	Proposed reason for incomplete coverage	Samples covered ≥10X, n
<i>APC</i>	c.423-3T>A	NA	Intron 3	NA	28.0%	polymer stretch of 13 As	99
<i>ATM</i>	c.1236-2A>T	NA	Intron 8	NA	28.0%	polymer stretch of 15 Ts	77
<i>ATM</i>	c.1236-3_1236-2DEL	NA	Intron 8	NA	28.0%	polymer stretch of 15 Ts	77
<i>ATM</i>	c.2250G>A	p.K750K	CDS 13	42.1%	NA	NA	98
<i>ATM</i>	c.3994-2A>G	NA	Intron 25	NA	29.0%	NA	97
<i>ATM</i>	c.4776+2T>C	NA	Intron 30	NA	30.0%	NA	99
<i>ATM</i>	c.7875_7876DELGTGINSGC	p.D2625_A2626delinsEP	CDS 52	39.6%	NA	NA	99
<i>ATM</i>	c.7913G>A	p.W2638*	CDS 52	39.6%	NA	NA	93
<i>ATRX</i>	c.109C>T	p.R37*	CDS 2	40.7%	NA	NA	98
<i>BRCA2</i>	c.6944_6947DELTA	p.I2315Kfs*12	CDS 12	38.6%	NA	NA	84
<i>BRCA2</i>	c.7007G>A	p.R2336H	CDS 12	38.6%	NA	NA	91
<i>BRIP1</i>	c.2392C>T	p.R798*	CDS 16	44.2%	NA	NA	44
<i>BRIP1</i>	c.2400C>G	p.Y800*	CDS 16	44.2%	NA	NA	57
<i>CBS</i>	c.1105C>T	p.R369C	CDS 10	67.9%	NA	GC-rich	88
<i>CDKN2A</i>	c.335_337DUPGTC	p.R112dup	CDS 2	72.3%	NA	GC-rich	51
<i>CUL4B</i>	c.2493G>A	p.T831T	CDS 19	38.2%	NA	NA	97
<i>EMD</i>	c.153dupC	p.S52Qfs*9	CDS 2	59.0%	NA	NA	97
<i>FBN1</i>	c.7C>T	p.R3*	CDS 1	60.4%	NA	GC-rich	99
<i>MECP2</i>	c.1152_1155DEL	p.P385Cfs*23	CDS 3	61.4%	NA	GC-rich	91
<i>MECP2</i>	c.1208DEL	p.P403Lfs*6	CDS 3	61.4%	NA	GC-rich	89
<i>MECP2</i>	c.1213C>T	p.P405S	CDS 3	61.4%	NA	GC-rich	95
<i>MRE11A</i>	c.21-6_26DEL12	NA	Intron 1-CDS 2	32.3%	NA	NA	85
<i>MSH2</i>	c.942+2T>C	NA	Intron 5	NA	27.0%	polymer stretch of 27 As	96
<i>MSH2</i>	c.942+3A>T	NA	Intron 5	NA	26.0%	polymer stretch of 27 As	94
<i>MSH6</i>	c.3172+1G>A	NA	Intron 4	NA	41.0%	NA	99
<i>MSH6</i>	c.3978_3979INSA	p.N1327Kfs*14	CDS 9	40.0%	NA	NA	98
<i>MSH6</i>	c.3980_3983DUPATCA	p.L1330Vfs*12	CDS 9	40.0%	NA	NA	99
<i>MSH6</i>	c.3984_3987DUPGTCA	p.L1330Vfs*12	CDS 9	40.0%	NA	NA	99
<i>PALB2</i>	c.226DELA	p.I76Yfs*101	CDS 4	40.1%	NA	NA	98
<i>PMS2</i>	c.2404C>T	p.R802*	CDS 14	55.3%	NA	pseudogene	50
<i>PMS2</i>	c.2500_2501DELATINS	p.M834Gfs*17	CDS 15	52.1%	NA	pseudogene	97
<i>PMS2</i>	c.325DELG	p.E109Kfs*3	CDS 4	46.6%	NA	pseudogene	35
<i>PTEN</i>	c.802-2A>T	NA	Intron 7	NA	24.0%	polymer stretch of 15 As	99
<i>RAD50</i>	c.2156DUPT	p.E723Gfs*5	CDS 13	45.8%		variant located at edge of polymer stretch of 9 As	99
<i>RAD50</i>	c.2165DUPA	p.E723Gfs*5	CDS 13	45.8%		polymer stretch of 9 As	98
<i>RAD50</i>	c.2202DELC	p.M735*	CDS 13	45.8%		NA	98
<i>RET</i>	c.2410G>A	p.V804M	CDS 14	64.7%		GC-rich	98
<i>RPGR</i>	c.28+1G>A	NA	Intron 1	NA	75.0%	GC-rich	81
<i>SLC16A2</i>	c.453DELC	p.E152Sfs*6	CDS 1	68.4%		GC-rich	97
<i>STK11</i>	c.608DUPC	p.F204Vfs*62	CDS 5	68.6%		GC-rich	86
<i>STK11</i>	c.913C>T	p.Q305*	CDS 7	60.3%		GC-rich	99
<i>TTN</i>	c.59933-1G>C	NA	Intron 267	NA	39.0%	NA	83

NA, not available.

<sup>a</sup> Intron % GC content for a 100 basepair window surrounding the variant

doi:10.1371/journal.pone.0170843.t002

are most likely to impact function. While variant subtype (single basepair substitution, indel, etc.) is subject to different NGS detection rates, clinical classification of alterations is not expected to have any impact on their detectability. It should be noted that the exome enrichment technique as well as the bioinformatics aligning and variant calling pipeline may result in differing detection rates.

Reports from multiple groups, including ours, have shown that the diagnostic yield of exome sequencing varies by indication [6, 18]. In addition, studies comparing targeted NGS panels with exome sequencing are, for the most part, disease-specific. Studies have supported targeted NGS panel testing as a first-tier testing approach over exome sequencing for several diseases based on diagnostic yield, coverage, and cost-savings [19–22]. For example, Wang and colleagues published results from a study on the diagnostic yield of a clinically-validated targeted NGS panel testing for retinitis pigmentosa (RP) [23]. Based on the high diagnostic yield reported in their study (82%) and the distinct phenotype observed with RP, the authors propose targeted panel testing as the first-tier approach for RP and WES as a second-tier option in cases where a molecular diagnosis is not made via panel testing. However, this study did not include any cross comparisons with the diagnostic yield of exome sequencing for RP. Results from another recent study of the performance of NGS and WES for inherited eye disorders also supported the use of targeted panel testing over WES based on superior analytic sensitivity and high diagnostic yield of the panel [24].

For diseases with a lower diagnostic yield on targeted NGS panel testing and/or less distinct clinical phenotypes, the choice between panel and exome as the first-tier testing approach may not be as straightforward. The major advantages of exome sequencing over targeted NGS panel testing is the detection of alterations in newly characterized genes, potential for novel gene characterizations, and ability to sequence nearly all genes in the genome. The rate of newly discovered gene characterizations is increasing rapidly and OMIM phenotypes for which the molecular basis is known almost doubled in the recent 6 years [25]. For each disease represented in this study, there have been published reports of novel genes after the NGS panels became clinically available. For example, approximately 30 genes have been associated with PCD, and a number of these have recently been discovered by WES [26–34]. According to current estimates, characterized PCD genes account for 66% of PCD cases, leaving the remaining third unexplained. Based on results from our study, WES coverage depth was adequate for detection for close to all (99.7%) pathogenic variants identified on targeted NGS panel testing, along with newly-discovered PCD genes not yet available on targeted NGS panels and identified potentially novel genes. An additional advantage of exome sequencing is the option for data reanalysis at a later point in time. Furthermore, the diagnostic yield of exome sequencing can be further maximized with approaches such as trio sequencing [6] and augmented exome sequencing [35].

Some important limitations of this study should be noted. Even though variants were detected in actual patients undergoing multigene panel testing and compared with coverage data from actual clinical exomes, this analysis is still theoretical and limited by the fact that variants studied were not directly detected prospectively on WES. While variants are confirmed with an orthogonal method prior to reporting WES results, the well-known disadvantage of exome sequencing compared to panel testing is the possibility of false-negative results as it is not feasible to confirm all no/low coverage regions across the exome without significantly impacting cost and turn-around-time. The results herein quantify the risk for false negatives on exome versus panel testing. Due to variations between exome enrichment platforms and bioinformatics methods, another limitation to this data is that it is based on the performance of SeqCap EZ VCRome 2.0 enrichment platform in combination with Ambry Genetics' custom bioinformatics pipeline, and therefore is not generalizable to other labs using the same

platform, or to other WES platforms. In addition, several alterations in genes containing regions of high sequence homology were included in this data set such as *CHEK2*, *MYH7*, and *PMS2*. Therefore, it is possible that alterations in these regions would have escaped detection if the reported coverage at the respective nucleotide position was representative of a homologous region rather than the actual gene. One final limitation to this study is that copy number variations were not assessed. Future studies are needed to directly compare the analytic and clinical sensitivity, cost analysis, and counseling implications for panels vs. WES. Examples include a two-arm study where patients are randomized to a targeted NGS panel vs. WES approach or a study where both a targeted panel and exome sequencing are performed on each patient concurrently. An important consideration is that factors beyond differences in technology and clinical and analytic sensitivity come into play in deciding to pursue targeted testing vs WES. For example, in hereditary cancer diagnostics, it is often imperative to have genetic testing results within a short turn-around-time, as results may impact surgical decisions such as lumpectomy vs. mastectomy in the setting of breast cancer or partial vs. total colon resection in the setting of colon cancer and polyposis syndromes [36, 37]. In addition to time-sensitive medical management considerations, there are also implications on genetic counseling practices. For example, targeted panels are unlikely to result in secondary findings, whereas this is a possibility for WES. Such factors are beyond the scope of this paper, but are important in considering testing approach.

## Conclusions

Despite current estimates that 90–95% exome-wide coverage is achieved with WES, results from this position-specific comparative coverage analysis limited to disease-causing variants identified through NGS panels demonstrate that exome sequencing is expected to perform well ( $\geq 98.5\%$ ) for a range of inherited diseases. If validated in follow-up studies, these data will help guide clinicians in deciding which type of testing to pursue for their patients. These data suggest the use of exome sequencing may achieve similar diagnostic yield when compared to panel based tests and, if cost and turn-around-time are comparable or favorable, that WES may be an appropriate first-tier option to consider when clinically indicated. The high level coverage achieved by WES reported herein, coupled with high rate of newly characterized and novel gene findings on exome (30% collectively) [6] demonstrates a major benefit of WES compared to panel testing for Mendelian diseases.

## Supporting information

### **S1 Table. Pathogenic variant coverage by gene.**

(XLSX)

### **S2 Table. Coverage summary for 100 WES samples.**

(XLSX)

### **S3 Table. Pathogenic variant position coverage for 100 WES samples.**

(XLSX)

### **S4 Table. ExAC coverage comparison.**

(XLSX)

### **S5 Table. WES samples with prior panel testing.**

(XLSX)

## Acknowledgments

We thank the laboratory, bioinformatics and clinical diagnostics teams at Ambry Genetics for their support in performing and analyzing/interpreting results for the genetic tests included in this study.

## Author Contributions

**Conceptualization:** HL.

**Formal analysis:** HL H-ML JC HV.

**Investigation:** JC WM HV.

**Methodology:** HL KDF LS ST SB ECC.

**Software:** WM JC H-ML.

**Writing – original draft:** HL KDF LS ST ECC.

**Writing – review & editing:** HL KDF HV H-ML WM LS ST JC SB ECC.

## References

1. Rehm HL. Disease-targeted sequencing: a cornerstone in the clinic. *Nature reviews Genetics*. 2013; 14(4):295–300. doi: [10.1038/nrg3463](https://doi.org/10.1038/nrg3463) PMID: [23478348](https://pubmed.ncbi.nlm.nih.gov/23478348/)
2. Asan, Xu Y, Jiang H, Tyler-Smith C, Xue Y, Jiang T, et al. Comprehensive comparison of three commercial human whole-exome capture platforms. *Genome biology*. 2011; 12(9):R95. doi: [10.1186/gb-2011-12-9-r95](https://doi.org/10.1186/gb-2011-12-9-r95) PMID: [21955857](https://pubmed.ncbi.nlm.nih.gov/21955857/)
3. Biesecker LG, Green RC. Diagnostic clinical genome and exome sequencing. *The New England journal of medicine*. 2014; 371(12):1170.
4. Chilamakuri CS, Lorenz S, Madoui MA, Vodak D, Sun J, Hovig E, et al. Performance comparison of four exome capture systems for deep sequencing. *BMC genomics*. 2014; 15:449. doi: [10.1186/1471-2164-15-449](https://doi.org/10.1186/1471-2164-15-449) PMID: [24912484](https://pubmed.ncbi.nlm.nih.gov/24912484/)
5. Clark MJ, Chen R, Lam HY, Karczewski KJ, Chen R, Euskirchen G, et al. Performance comparison of exome DNA sequencing technologies. *Nature biotechnology*. 2011; 29(10):908–14. doi: [10.1038/nbt.1975](https://doi.org/10.1038/nbt.1975) PMID: [21947028](https://pubmed.ncbi.nlm.nih.gov/21947028/)
6. Farwell KD, Shahmirzadi L, El-Khechen D, Powis Z, Chao EC, Davis BT, et al. Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. *Genetics in medicine: official journal of the American College of Medical Genetics*. 2014.
7. Wooderchak-Donahue WL, O'Fallon B, Furtado LV, Durtschi JD, Plant P, Ridge PG, et al. A direct comparison of next generation sequencing enrichment methods using an aortopathy gene panel- clinical diagnostics perspective. *BMC medical genomics*. 2012; 5:50. doi: [10.1186/1755-8794-5-50](https://doi.org/10.1186/1755-8794-5-50) PMID: [23148498](https://pubmed.ncbi.nlm.nih.gov/23148498/)
8. Laduca H, Stuenkel AJ, Dolinsky JS, Keiles S, Tandy S, Pesaran T, et al. Utilization of multigene panels in hereditary cancer predisposition testing: analysis of more than 2,000 patients. *Genetics in medicine: official journal of the American College of Medical Genetics*. 2014.
9. Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS, et al. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Human mutation*. 2008; 29(11):1282–91. doi: [10.1002/humu.20880](https://doi.org/10.1002/humu.20880) PMID: [18951446](https://pubmed.ncbi.nlm.nih.gov/18951446/)
10. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in medicine: official journal of the American College of Medical Genetics*. 2015; 17(5):405–24.
11. Chong HK, Wang T, Lu HM, Seidler S, Lu H, Keiles S, et al. The validation and clinical implementation of BRCAplus: a comprehensive high-risk breast cancer diagnostic assay. *PloS one*. 2014; 9(5):e97408. doi: [10.1371/journal.pone.0097408](https://doi.org/10.1371/journal.pone.0097408) PMID: [24830819](https://pubmed.ncbi.nlm.nih.gov/24830819/)

12. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*. 2009; 25(21):2865–71. doi: [10.1093/bioinformatics/btp394](https://doi.org/10.1093/bioinformatics/btp394) PMID: [19561018](https://pubmed.ncbi.nlm.nih.gov/19561018/)
13. Gandomi SK, Farwell Gonzalez KD, Parra M, Shahmirzadi L, Mancuso J, Pichurin P, et al. Diagnostic exome sequencing identifies two novel IQSEC2 mutations associated with X-linked intellectual disability with seizures: implications for genetic counseling and clinical diagnosis. *Journal of genetic counseling*. 2014; 23(3):289–98. doi: [10.1007/s10897-013-9671-6](https://doi.org/10.1007/s10897-013-9671-6) PMID: [24306141](https://pubmed.ncbi.nlm.nih.gov/24306141/)
14. Gonzalez KD, Li X, Lu HM, Lu H, Pellegrino JE, Miller RT, et al. Diagnostic Exome Sequencing and Tailored Bioinformatics of the Parents of a Deceased Child with Cobalamin Deficiency Suggests Digenic Inheritance of the MTR and LMBRD1 Genes. *JIMD reports*. 2014.
15. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016; 536(7616):285–91. doi: [10.1038/nature19057](https://doi.org/10.1038/nature19057) PMID: [27535533](https://pubmed.ncbi.nlm.nih.gov/27535533/)
16. Park JY, Clark P, Londin E, Sponziello M, Kricka LJ, Fortina P. Clinical Exome Performance for Reporting Secondary Genetic Findings. *Clinical chemistry*. 2014.
17. Dorschner MO, Amendola LM, Turner EH, Robertson PD, Shirts BH, Gallego CJ, et al. Actionable, pathogenic incidental findings in 1,000 participants' exomes. *American journal of human genetics*. 2013; 93(4):631–40. doi: [10.1016/j.ajhg.2013.08.006](https://doi.org/10.1016/j.ajhg.2013.08.006) PMID: [24055113](https://pubmed.ncbi.nlm.nih.gov/24055113/)
18. Lee H, Deignan JL, Dorrani N, Strom SP, Kantarci S, Quintero-Rivera F, et al. Clinical exome sequencing for genetic identification of rare Mendelian disorders. *Jama*. 2014; 312(18):1880–7. doi: [10.1001/jama.2014.14604](https://doi.org/10.1001/jama.2014.14604) PMID: [25326637](https://pubmed.ncbi.nlm.nih.gov/25326637/)
19. Ankala A, da Silva C, Gualandi F, Ferlini A, Bean LJ, Collins C, et al. A comprehensive genomic approach for neuromuscular diseases gives a high diagnostic yield. *Annals of neurology*. 2015; 77(2):206–14. doi: [10.1002/ana.24303](https://doi.org/10.1002/ana.24303) PMID: [25380242](https://pubmed.ncbi.nlm.nih.gov/25380242/)
20. Iglesias A, Anyane-Yeboah K, Wynn J, Wilson A, Truitt Cho M, Guzman E, et al. The usefulness of whole-exome sequencing in routine clinical practice. *Genetics in medicine: official journal of the American College of Medical Genetics*. 2014; 16(12):922–31.
21. Ostrer H. Changing the game with whole exome sequencing. *Clinical genetics*. 2011; 80(2):101–3. doi: [10.1111/j.1399-0004.2011.01712.x](https://doi.org/10.1111/j.1399-0004.2011.01712.x) PMID: [21749364](https://pubmed.ncbi.nlm.nih.gov/21749364/)
22. Soden SE, Saunders CJ, Willig LK, Farrow EG, Smith LD, Petrikon JE, et al. Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of neurodevelopmental disorders. *Science translational medicine*. 2014; 6(265):265ra168. doi: [10.1126/scitranslmed.3010076](https://doi.org/10.1126/scitranslmed.3010076) PMID: [25473036](https://pubmed.ncbi.nlm.nih.gov/25473036/)
23. Wang J, Zhang VW, Feng Y, Tian X, Li FY, Truong C, et al. Dependable and efficient clinical utility of target capture-based deep sequencing in molecular diagnosis of retinitis pigmentosa. *Investigative ophthalmology & visual science*. 2014; 55(10):6213–23.
24. Consugar MB, Navarro-Gomez D, Place EM, Bujakowska KM, Sousa ME, Fonseca-Kelly ZD, et al. Panel-based genetic diagnostic testing for inherited eye diseases is highly accurate and reproducible, and more sensitive for variant detection, than exome sequencing. *Genetics in medicine: official journal of the American College of Medical Genetics*. 2014.
25. Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nature reviews Genetics*. 2013; 14(10):681–91. doi: [10.1038/nrg3555](https://doi.org/10.1038/nrg3555) PMID: [23999272](https://pubmed.ncbi.nlm.nih.gov/23999272/)
26. Kurkowiak M, Zietkiewicz E, Witt M. Recent advances in primary ciliary dyskinesia genetics. *Journal of medical genetics*. 2015; 52(1):1–9. doi: [10.1136/jmedgenet-2014-102755](https://doi.org/10.1136/jmedgenet-2014-102755) PMID: [25351953](https://pubmed.ncbi.nlm.nih.gov/25351953/)
27. Hjeij R, Onoufriadis A, Watson CM, Slagle CE, Klena NT, Dougherty GW, et al. CCDC151 mutations cause primary ciliary dyskinesia by disruption of the outer dynein arm docking complex formation. *American journal of human genetics*. 2014; 95(3):257–74. doi: [10.1016/j.ajhg.2014.08.005](https://doi.org/10.1016/j.ajhg.2014.08.005) PMID: [25192045](https://pubmed.ncbi.nlm.nih.gov/25192045/)
28. Horani A, Druley TE, Zariwala MA, Patel AC, Levinson BT, Van Arendonk LG, et al. Whole-exome capture and sequencing identifies HEATR2 mutation as a cause of primary ciliary dyskinesia. *American journal of human genetics*. 2012; 91(4):685–93. doi: [10.1016/j.ajhg.2012.08.022](https://doi.org/10.1016/j.ajhg.2012.08.022) PMID: [23040496](https://pubmed.ncbi.nlm.nih.gov/23040496/)
29. Knowles MR, Leigh MW, Ostrowski LE, Huang L, Carson JL, Hazucha MJ, et al. Exome sequencing identifies mutations in CCDC114 as a cause of primary ciliary dyskinesia. *American journal of human genetics*. 2013; 92(1):99–106. doi: [10.1016/j.ajhg.2012.11.003](https://doi.org/10.1016/j.ajhg.2012.11.003) PMID: [23261302](https://pubmed.ncbi.nlm.nih.gov/23261302/)
30. Knowles MR, Ostrowski LE, Loges NT, Hurd T, Leigh MW, Huang L, et al. Mutations in SPAG1 cause primary ciliary dyskinesia associated with defective outer and inner dynein arms. *American journal of human genetics*. 2013; 93(4):711–20. doi: [10.1016/j.ajhg.2013.07.025](https://doi.org/10.1016/j.ajhg.2013.07.025) PMID: [24055112](https://pubmed.ncbi.nlm.nih.gov/24055112/)

31. Kott E, Legendre M, Copin B, Papon JF, Dastot-Le Moal F, Montantin G, et al. Loss-of-function mutations in RSPH1 cause primary ciliary dyskinesia with central-complex and radial-spoke defects. *American journal of human genetics*. 2013; 93(3):561–70. doi: [10.1016/j.ajhg.2013.07.013](https://doi.org/10.1016/j.ajhg.2013.07.013) PMID: [23993197](https://pubmed.ncbi.nlm.nih.gov/23993197/)
32. Moore DJ, Onoufriadis A, Shoemark A, Simpson MA, zur Lage PI, de Castro SC, et al. Mutations in ZMYND10, a gene essential for proper axonemal assembly of inner and outer dynein arms in humans and flies, cause primary ciliary dyskinesia. *American journal of human genetics*. 2013; 93(2):346–56. doi: [10.1016/j.ajhg.2013.07.009](https://doi.org/10.1016/j.ajhg.2013.07.009) PMID: [23891471](https://pubmed.ncbi.nlm.nih.gov/23891471/)
33. Olbrich H, Schmidts M, Werner C, Onoufriadis A, Loges NT, Raidt J, et al. Recessive HYDIN mutations cause primary ciliary dyskinesia without randomization of left-right body asymmetry. *American journal of human genetics*. 2012; 91(4):672–84. doi: [10.1016/j.ajhg.2012.08.016](https://doi.org/10.1016/j.ajhg.2012.08.016) PMID: [23022101](https://pubmed.ncbi.nlm.nih.gov/23022101/)
34. Zariwala MA, Gee HY, Kurkowiak M, Al-Mutairi DA, Leigh MW, Hurd TW, et al. ZMYND10 is mutated in primary ciliary dyskinesia and interacts with LRRC6. *American journal of human genetics*. 2013; 93(2):336–45. doi: [10.1016/j.ajhg.2013.06.007](https://doi.org/10.1016/j.ajhg.2013.06.007) PMID: [23891469](https://pubmed.ncbi.nlm.nih.gov/23891469/)
35. Patwardhan A, Harris J, Leng N, Bartha G, Church DM, Luo S, et al. Achieving high-sensitivity for clinical applications using augmented exome sequencing. *Genome medicine*. 2015; 7(1):71.
36. The NCCN Clinical Practice Guidelines in Oncology™ Genetic/Familial High-Risk Assessment: Colorectal V2.2014: National Comprehensive Cancer Network, Inc.; 2014 [Jan 2015]. <http://www.nccn.org/>.
37. The NCCN Clinical Practice Guidelines in Oncology™ Genetic/Familial High-Risk Assessment: Breast and Ovarian V2.2014: National Comprehensive Cancer Network, Inc.; 2014 [Jan 2015]. <http://www.nccn.org/>.