Copy number analysis using next-generation sequencing: Comprehensive genetic testing and its application to neuromuscular and epilepsy panels

Ali Entezam, Thomas L. Winder, Rebecca Truty, Nila Patel, Jody Westbrook, Christopher A. Tan, Paige Taylor, Marina Rabideau, Hannah White, Swaroop Aradhya

Abstract

Objective: The diagnosis of Mendelian disorders is routinely performed through gene sequence or chromosomal copy number variant (CNV) analysis. Although exon-level CNVs contribute significantly to disease burden, their testing has traditionally been limited to a few loci. We evaluated a next-generation sequencing (NGS) method that simultaneously evaluates sequence changes and gene-level CNVs in hundreds of genes. Our data show that a high frequency of gene and exonic CNVs are associated with neurological disorders.

Methods: We investigated subsets of 300 genes curated for neuromuscular disorders in 4,358 unrelated individuals and up to 186 genes curated for epilepsy in 2,008 unrelated individuals. Validated coverage-based CNV detection algorithms and custom algorithms designed to flag split-read signals were applied to all genes in all samples.

Results: Among individuals with neuromuscular disorders, a total of 424 CNVs were detected, 386 of these CNVs were pathogenic and diagnostic of the patient's conditions. These variants represented 40% of all neuromuscular diagnostic findings. The majority of pathogenic CNVs were in *PMP22* or *DMD*, but CNVs were also identified in several other genes not traditionally tested for CNVs. These genes include *SPAST*, *SPG7*, *SPG11*, *LAMA2*, *TRIM32*, *GCH1*, *FGD4*, *SPG7*, and *REEP1*. The highest observed diagnostic yields were obtained in SMA (68%) and DMD (63%) testing. The single most frequently observed pathogenic variant was gene duplication of *PMP22*. Among individuals with epilepsy, 43 had CNV diagnostic pathogenic variants reported out of 55 observed whole-gene or exonic CNVs in 23 genes, representing 13% of the pathogenic variants in epilepsy genes.

Conclusion: Exonic CNVs explain a substantial (~15%) proportion of pathogenic variants across several disorders and occur in a broader variety of genes than previously appreciated. This NGS-based sequence and CNV detection method can be used routinely in germline genetic testing in neurological disorders to detect both common and rare events that together contribute to a high diagnostic yield.

Introduction

The implementation of NGS panels has significantly improved the ability to identify molecular causes of neuromuscular conditions and epilepsy. Previously, CNV analysis for genes associated with these conditions was either unavailable or offered as a separate test with additional costs and time. Furthermore, the frequency of these events has been relatively unknown.

The purpose of this study was to evaluate the frequency of CNVs detected in patients with neuromuscular and epilepsy disease using a single NGS assay to detect both sequence changes and gene-level CNVs simultaneously.

Methods

A total of 4,358 and 2,008 unrelated patient samples were tested at Invitae for sequence and deletion/duplication analysis of genes associated with neuromuscular conditions and epilepsy, respectively. Lists of genes included in each panel can be found at www.invitae.com.

Germline DNA from blood was tested with neuromuscular and epilepsy panels. Sequence and CNV analyses were performed with validated NGS methods, allowing for concurrent analysis of sequence variants and exonic CNVs, thereby providing a method for evaluating the contribution of CNVs in inherited neuromuscular conditions and epilepsy. The reported CNVs were confirmed with an alternative method.

Invitae uses a proprietary, validated algorithm to detect deletions and duplications with NGS. The algorithm calls exonic deletions and duplications by calculating the statistical likelihood of each copy number state through comparison of the depth of sequence coverage at the targeted exons with the depth measured from a set of baseline samples (Figure 1).[1]

SMN1 and *SMN2* analyses are accomplished using a validated bioinformatic approach and high-depth NGS reads derived from both *SMN1* and *SMN2*. Combined reads are aligned to *SMN1*, and combined *SMN1/2* copy number is determined using Invitae's read count-based copy number variant detection algorithm, CNVitae. *SMN1*- and *SMN2*-specific exon 7 copy number is resolved by counting reads with the gene determining variant in exon 7.[2]

Figure 1. Read-depth approach to deletion/duplication analysis by NGS. Example of a duplication.

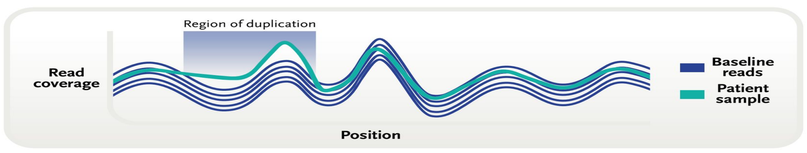
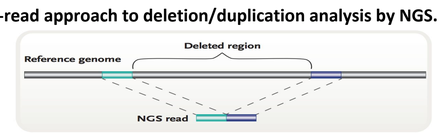


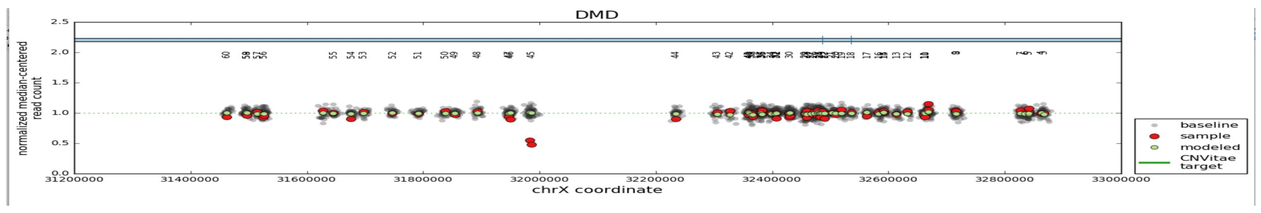
Figure 2: Split-read approach to deletion/duplication analysis by NGS.



Split-read: NGS reads that span a breakpoint of a deletion or duplication will show distinct patterns when mapped to the reference genome, and these patterns are detected by a process called split-read analysis (Figure 2). Some events detected by split-read analysis are exactly those that are harder for read-depth approaches, such as events that start or end in the middle of an exon. Not only do the two approaches complement each other, but used together they can resolve novel complex events that can be difficult for probe-based methods like qPCR, MLPA, or microarrays.

Methods

Figure 3: Use of read depth to visualize the loss of one exon in the *DMD* gene in a female



Results

A total of 4,632 and 2,008 evaluations were performed at Invitae for sequence and deletion/duplication analysis of genes associated with neuromuscular conditions and epilepsy, respectively. The overall diagnostic yield for neuromuscular conditions was 22.5% with pathogenic CNVs accounting for 40% of all positive results (15% when *DMD* and *PMP22* are excluded). The overall diagnostic yield for epilepsy was 17%, with pathogenic CNVs accounting for 13% of all positive results.

Analysis (Neuromuscular panels) / number of physician orders	Number of positive results (positive rate)	Positive CNV (% of positive results)	Positive SNV (% of positive results)	Number of uncertain results (VUS rate)	Most common gene (% of positive results)
Neuromuscular / 602	148 (25%)	20 (14%)	128 (86%)	352 (58%)	*DMD* (13%)
Neuropathy / 1215	144 (12%)	65 (45%)	79 (55%)	597 (49%)	*PMP22* (46%)
Myopathy / 227	27 (12%)	0	27 (12%)	121 (53%)	*RYR1* (31%)
Muscular dystrophy / 282	117 (51%)	22 (19%)	95 (81%)	104 (46%)	*DMD* (63%)
Limb-girdle muscular dystrophy / 117	34 (33%)	5 (15%)	29 (85%)	38 (28%)	*DMD* (32%)
Charcot-Marie-Tooth disease / 897	252 (28%)	118 (47%)	134 (53%)	251 (28%)	*PMP22* (50%)
Dystrophinopathies (*DMD* gene) / 141	84 (60%)	52 (62%)	32 (38%)	7 (8%)	n/a
Dystonia / 366	40 (12%)	5 (12%)	35 (88%)	43 (13%)	*SGCE* (23%)
Hereditary spastic paraplegia / 662	108 (16%)	16 (15%)	92 (85%)	232 (35%)	*SPAST* (39%)
SMA (*SMN1/2* genes) / 123	84 (68%)	83 (99%)	1 (1%)	0	n/a

Analysis (Epilepsy panels)	Number of positive results (positive rate)	Positive CNV (% of positive results)	Positive SNV (% of positive results)	Number of uncertain results (VUS rate)	Most common gene (% of positive results)
Comprehensive Epilepsy / 1832	307 (17%)	38 (12.5%)	269 (87.5%)	1069 (58%)	SCN1A (16%)
EIEE / 77	11 (14%)	4 (36%)	7 (63%)	26 (34%)	KCNQ2 (45%)
Rett and Angelman and related disorders / 99	20 (20%)	1 (5%)	19 (95%)	14 (14%)	MECP2 (45%)

Table 1. Percentage of orders for each gene panel in which likely pathogenic or pathogenic variants (positive rate), variant(s) of uncertain significance in the absence of positive findings (VUS rate) were detected. Among positive results, the percentage that are copy number variants (CNVs) and sequence variants (single nucleotide variants [SNVs]) are called out. For each panel, the gene contributing to the most positive report is also given.

Conclusion

The results suggest that CNV analysis is a key component of genetic testing for inherited neuromuscular conditions and epilepsy. For CMT disease panels, CNVs accounted for 47% of positive results, the majority of which were whole-gene copy gains of *PMP22* (which is the most common causative CMT variant). For *DMD*, 62% of the positive findings were CNVs, while the remaining 38% were SNVs. CNVs accounted for 12% and 15% of positive findings in dystonia and HSP panels, respectively. Among individuals with epilepsy, a total of 55 whole-gene or exonic CNVs were identified among 23 genes. Of these CNVs, 43 were pathogenic and diagnostic for epilepsy, thereby resulting in a diagnostic yield of 13%.

The development and validation of this method for clinical diagnostic testing requires a considerable investment in bioinformatics. An NGS-based CNV detection method allows for integrated simultaneous sequence and copy number calling, which lowers costs, reduces turnaround time, decreases sample input requirements, and increases resolution.

References

1. Invitae white paper. Detecting deletions and duplications using next-generation sequencing (NGS) (WP100-2). https://marketing.invitae.com/acton/attachment/7098/f-01ab/1/-/-/-/-/WP100_Invitae_WhitePaper_DeletionDuplication.pdf

2. Invitae white paper. Invitae's unique approach to testing *SMN1* and *SMN2* for spinal muscular atrophy. https://marketing.invitae.com/acton/attachment/7098/f-05d2/1/-/-/-/-/WP108_Invitae_WhitePaper_SMN1-SMN2.pdf