Interpretation of neurological disorders cohort through in-house developed tool, festiVAR, showed an improved diagnostic yield: data from over 1000 Indian patients

A. Mannan, B. Forona, S. Sur, K. Karmalkar, K. Nisha, J. S. Vasudevan, R. Dave, S. Nayanala, R. H. J., S. Sridharan, S. Phalke, C. Judith, S. Katragadda, A. Janakiraman, V. Veeramachaneni, R. Hariharan

Strand Life Sciences, Bangalore, India

Abstract

Background: Neurological disorders are one of the major causes of disability and death worldwide. These disorders are clinically heterogeneous group and a significant proportion of these are 'neurogenetic disorders', which include disorders, such as epilepsy, ataxia, leukodystrophy, neuropathy, dystonia, myopathy, autism, intellectual disability and others, that are mostly caused due to defects in one or more genes. A precise and confirmatory diagnosis in the patients in a timely manner is essential for appropriate therapeutic and management strategies. Due to the complexity of the clinical presentations across various neurological disorders, arriving at an accurate diagnosis remains a challenge.

Methods: We sequenced 1082 unrelated patients from India with suspected neurological disorders, using whole exome panel. Genetic variations were identified using the Strand NGS software and interpretation was done by using in-house tools: festiVAR and StrandOmics platform.

Results: The diagnostic yield in our cohort was 40% (pathogenic and likely pathogenic variants) and in 26% of cases, we detected VUS (variant of uncertain significance). The highest diagnostic rate was observed among patients with muscular dystrophy (65%) followed by leukodystrophy (51%). Interestingly, the detection rate in autism cases was lowest (22%). In terms of mutations types, we detected all types of variants with 50% of pathogenic/likely pathogenic variants were missense and 12.1% were structural variants.

Conclusion: In our study, we observed an improved performance of whole exome testing, with an overall diagnostic yield of 40%. Furthermore, we show that our recently developed in-house interpretation tool, festiVAR, is fast and efficient, which can minimize the time required to perform interpretation and it can be very effective in identifying causative genes/variants in complex neurological cases.

Material and methods

Cases and sample collection: The study involved 1082 unrelated patients affected with different neurological conditions, referred for whole exome sequencing (WES) at >100x coverage. Genomic deoxyribonucleic acid (gDNA) was isolated from the patient's blood, saliva or any other tissue specimen for preparation of the 'DNA sequencing ready' library.

NGS - data analysis: Next-generation sequencing (NGS) was performed on the patient's gDNA on Illumina NovaSeq instrument. We used the xGen™ DNA Library Prep EZ Kits (IDT) protocol to produce next-generation sequencing (NGS) libraries. The reads from the FASTQ files generated via bcl2fastq were processed using the on-instrument DRAGEN Enrichment pipeline on NovaSeqXPlus. The whole genome build hg19-hs37d5 was used for the analysis involving mapping/ aligning, sorting, duplicate marking, and variant calling. The variants called by the Dragen pipeline and the STRAND®NGS (http://www.strand-ngs.com/) pipeline were combined for variant annotation and prioritization. STRAND® NGS is an NGS analysis platform from Strand Life Sciences. It comprises algorithms for alignment, variant calling, exon deletion/duplication analysis, and structural variant calling.

Variant interpretation: Post variant-calling, single nucleotide variants (SNVs) and small indels were prioritized based on festiVAR (fast estimation of variants for automated reporting) (v0.97.6). festiVAR prioritizes the variants taking into account of the HPO terms and various gene and variant-level factors including the inheritance model, the number of variants, predicted effects of these variants, allele frequencies and presence of the variants in disease-database. Furthermore, the variants prioritized in festiVAR are selected for reporting in StrandOmics (v6.28.0), which is a clinical genomics interpretation and reporting platform.

CNV Analysis: In addition to SNVs and small indels, copy number analysis was performed to identify large deletions or insertions ranging from exon-level to chromosomal-level based on the consensus between the STRAND®NGS (v3.3.5) and Illumina Dragen enrichment pipelines.

Results

- The overall diagnostic yield (pathogenic and likely pathogenic variants) in our cohort was 40%.
- The highest diagnostic rate was observed among patients with muscular dystrophy (65%) followed by leukodystrophy (51%).
- Interestingly, the detection rate in autism and intellectual disability cases was lowest (22%).
- Variants of uncertain significance (VUS) were detected in 26% of the cases. Among the VUS variants, the highest detection rate was in autism and intellectual disability, at 38%.
- The most significant contributor of mutations for neurological conditions was the *DMD* gene (12 occurrences), followed closely by the *CAPN3* gene (10 occurrences).
- The lowest contributors were the *SLC2A1, CEP290, PLA1G6* genes (1 occurrence).
- In our cohort, we detected all types of 'pathogenic'/'likely pathogenic' variants. The mutation spectrum showed 50.4% missense, 15.3% nonsense, 12.8% frameshift, 8.6% splice-site, 0.7% in-frame and 12.1% structural variants (CNV).
- Among the copy number variants (CNV), we observed that band level events constituted 54.02% of the total variants observed, while intragenic CNVs contributed to 45.98% of the total.
- The intragenic CNVs could be further categorized as 80% multi-exon events and 20% single exon events.

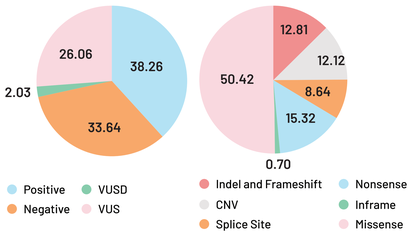


Fig. 1 Mutation detection rate and mutation spectrum in neurological disorders cohort. The pie chart shows mutation detection rate of 40% [which includes 'pathogenic' (P), "likely pathogenic' (LP) variants and 'variants of uncertain significance with probable damaging effect' (VUSD)]. VUS was identified in 33.64% of cases, while no mutation was detected in 26.06% of cases. The mutation types detected in our cohort were indel (small deletion, duplication, insertion, or insertion/deletion), missense, inframe, nonsense, splices site, CNVs (large deletion, large duplication at both gene and chromosomal levels).

Clinical Subtypes	No. Of Cases	P+LP	VUSD	VUS	Negative
Epilepsy	134	33.58	1.49	29.10	35.82
Ataxia	60	20.00	6.67	30.00	43.33
Dystonia	43	44.19	0.00	16.28	39.53
Muscular dystrophy and Myopathies	89	64.04	1.12	14.61	20.22
Spasticity and ALS	22	40.91	0.00	31.82	27.27
Autsim and Intellectual Disability	197	21.32	1.02	38.07	39.59
Leukodystrophy	31	48.39	3.23	29.03	19.35
Brain malformation and Microcephaly	162	43.83	1.85	19.75	34.57
Developmental disorders	173	43.93	1.73	27.17	27.17
Neuropathy and Others	171	39.77	3.51	20.47	36.26

Table 1: Mutation detection rate in clinical subtypes of neurological disorders in the study.

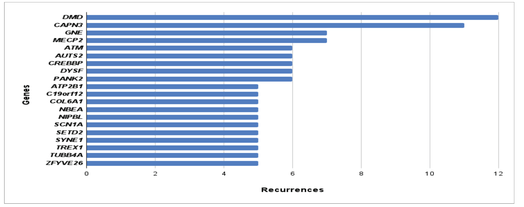
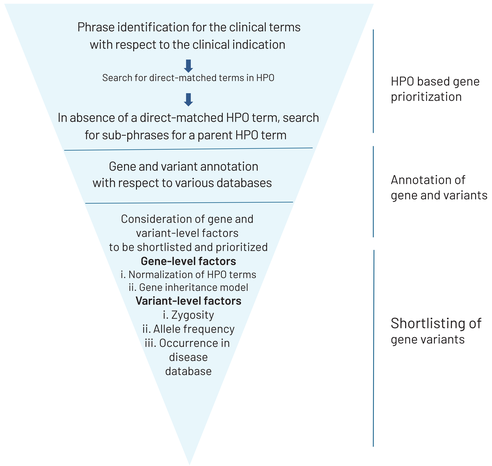


Fig. 2 Gene recurrence. The most frequently mutated gene was *DMD (12). Other recurrently mutated genes include CAPN3 (11), GNE (7), MECP2 (7), ATM, AUTS2, CREBBP, DYSF (and PANK2 (6 each).*



Conclusion

- NGS-based Whole Exome test is comprehensive for neurological disorders, as all types variants including CNV can be detected by a single test. Additionally, Whole Exome test is inherently unbiased and limitless with respect to what genes should be evaluated and is less dependent on a priori clinical information.
- In our study, we sequenced 1082 unrelated patients from India with suspected neurological disorders using Whole Exome test and observed an overall diagnostic yield of 40%.
- The highest diagnostic rate was observed among patients with muscular dystrophy (65%) and the detection rate in autism and intellectual disability cases was the lowest (22%). In epilepsy, the detection rate was ~35%, which is much higher then other large-scale studies, where detection rate of 10–20% were reported in patients affected epilepsy. Similarly, in our cohort, a higher diagnostic yield has also been observed for other conditions, such as ataxia, muscular dystrophy and myopathy as compared to other studies.
- About 12.1% of the positive cases were CNV positive cases out of which 54.02% were large CNV band events and 45.98% were intragenic events. With respect to the intragenic events, 80% are multiple exon events whereas 20% consisted of single exon events.
- Furthermore, we show that our recently developed in-house interpretation tool, festiVAR is efficient and reliable, which can considerably minimize the time required for interpretation and is effective in identifying causative genes/variants by sorting the variants in order of their importance and relevance to phenotype, which is essential for interpretation of complex neurological cases involving numerous symptoms/phenotypes.
- In conclusion, our tool, festiVAR is reliable, time saving, reduces the margin of error and helps in accelerated interpretation of neurological disorder cases.

Contact

Ashraf ul Mannan, MS, PhD, PD
Sr. Director - Medical Genetics & Genomics
(PD in Molecular Human Genetics
from UMG, Germany)

+91-7259777704 ashraf@strandls.com