**Multi-gene testing in neurological disorders showed an improved diagnostic yield: data from over 1000 Indian patients**

Aparna Ganapathy1, Avshesh Mishra1,Megha Rani Soni1, Priyanka Kumar1, Mukunth Sadagopan1, Anil Vittal Kanthi1, Irene Rosetta Pia Patric1, Sobha George1, Aparajit Sridharan1, Thyagarajan T.C1, Aswathy SL1, Vidya HK1, Swathi M. Chinnappa1, Swetha Nayanala1, Manasa B. Prakash1, Vijayashree G. Raghavendrachar1, Minothi Parulekar1,Vykuntaraju K. Gowda2, Sheela Nampoothiri3, Ramshekhar N. Menon4, Divya Pachat5, Vrajesh Udani6, Neeta Naik7, Mahesh Kamate8, A. Radha Rama Devi9, P. A. Mohammed Kunju10, Mohandas Nair11, Anaita Udwadia Hegde12, M. Pradeep Kumar13, Soumya Sundaram4, Preetha Tilak14, Ratna D. Puri15, Krati Shah16, Jayesh Sheth17, Qurratulain Hasan18, Frenny Sheth17, Pooja Agrawal1, Shanmukh Katragadda1, Vamsi Veeramachaneni1, Vijay Chandru1,19,Ramesh Hariharan1,19, Ashraf U. Mannan1\*

**Authors’ affiliations**

1Strand Center for Genomics and Personalized Medicine, Strand Life Sciences, Bangalore, India

2Indira Gandhi Institute of Child Health, Bangalore, India

3Amrita Institute of Medical Sciences and Research Centre, Cochin, India

4Sree Chitra Tirunal Institute for Medical Sciences and Technology, Thiruvananthapuram, India

5Aster MIMS, Kozhikode, India

6 P. D. Hinduja Hospital and Medical Research Centre, Mumbai, India

7EN1 Neuro Services Pvt. Ltd., Mumbai, India

8 KLES Dr. Prabhakar Kore Hospital, Belgaum, India

9 Rainbow Children's Hospital, Hyderabad, India

10 SAT Hospital, Thiruvananthapuram, India

11 Goverment Medical College, Kozhikode, India

12 SRCC Children's Hospital, Mumbai, India

13 GeneOmm Medical Centre, Coimbatore. India

14 St. Johns Medical College Hospital, Bangalore, India

15Sir Ganga Ram Hospital, Delhi, India

16ONE - Centre for Rheumatology and Genetics, Vadodra, India

17FRIGE's Institute of Human Genetics, Ahmedabad, India.

18Kamineni Hospitals, Hyderabad, India

19Indian Institute of Science, Bangalore, India

\***Corresponding author:**

Ashraf U. Mannan, PhD

Strand Life Sciences, Kirloskar Business Park, Bellary Road, Hebbal

Bangalore-560024, India

Phone: +91-80-40787353, Fax: +91-80-40787299, Email: ashraf@strandls.com

**METHODS**

**Patients**

This study comprises of 1012 unrelated patients that included 566 males (~56%) and 446 were females (~44%). Of these, 31% (313/1012) of the patients were born to consanguineous parents and ~61% (616/1012) were born to reportedly non-consanguineous parents. Consanguinity data was not available for 8% (83/1012) of the patients.

**Sample preparation and multi-gene panel sequencing**

DNA was extracted from saliva samples using the PrepIT-L2P kit (DNA Genotek, Canada), as per the manufacturer’s instructions. For blood samples, either the QIAamp DNA Mini Kit (Qiagen, Germany) or the Nucleospin kit (Macherey-Nagel, Germany) was used for DNA isolation, as per the manufacturer’s instructions. NGS was performed using TruSight One Sequencing Panel (Illumina, USA) that provide comprehensive coverage of >4600 genes associated with known inherited diseases. An analytical validation of this panel in our laboratory has shown sensitivity of >96.5%, specificity of 99.95% and reproducibility of ~97%.

**Library preparation and target enrichment**

The Nextera DNA library preparation protocol (Illumina, USA) was used to convert input genomic DNA (gDNA) into adapter-tagged indexed libraries. Approximately 50 ng of input gDNA was used in the tagmentation process, which involves simultaneous fragmentation and adapter tagging of gDNA followed by adapter ligation. This was followed by limited cycles of PCR (ABI 9700, Life Technologies, USA) to allow the incorporation of sample-specific indices or multiplex identifier barcodes. The quality of the library was assessed using the BioAnalyzer (Agilent, USA). Next, 500 ng of individual libraries were pooled in batches of 4, 12 or 24 samples and hybridized to biotin-labeled probes specific to the targeted regions. The pool was enriched for the target genomic regions by adding streptavidin beads that bind to the biotinylated probes. The biotinylated gDNA fragments bound to the streptavidin beads were magnetically pulled down from the solution. The partly enriched gDNA fragments were then eluted from the beads and subjected to a second round of hybridization. The tagged and amplified sample libraries were checked for quality and quantified using the BioAnalyzer (Agilent, USA). Upto 6-10 pM of the pooled library was loaded and sequenced on the MiSeq or NextSeq platform (Illumina, USA), according to the manufacturer’s instructions.

**NGS-data analysis and interpretation**

The trimmed FASTQ files were generated using MiSeq/NextSeq Reporter (Illumina, USA). Reads were aligned against the whole genome build: hg19 using Strand NGSv2.6 (<http://www.strand-ngs.com/>). Strand NGS is an integrated platform that provides analysis, management and visualization tools for NGS data. It has a comprehensive DNA-Seq pipeline that includes alignment, read-quality assessment, filtering, small variant calling and copy number variation detection. NGS data analysis was performed using Strand NGS and variant interpretation was done by StrandOmicsv5.0 (a proprietary clinical genomics interpretation and reporting platform from Strand Life Sciences, India), as previously described [1]. In brief, StrandOmics is a clinical interpretation and reporting platform that combines knowledge from internal curated literature content, along with various publically available data sources such as Uniprot, OMIM, HGMD, ClinVar, ARUP, dbSNP, 1000 Genomes, Exome Variant Server (EVS), and Exome Aggregation Consortium (ExAC). In addition to databases, bioinformatics prediction tools, such as SIFT, PolyPhen-2, Mutation Taster, Mutation Assessor, FATHMM, LRT for missense variants, and NNSPLICE, SplicePort and MaxEntScan tools for variants in essential splice sites and exon–intron boundaries, have also been used to assess the pathogenicity of the variants. This integrated knowledge is then used to prioritize automatically a list of variants based on American College of Medical Genetics and Genomics (ACMG) guidelines for the interpretation of sequence variants [2]. Also, for the automated variant priorization, the inheritance model, disease phenotype, sequence conservation across various species, and allelic frequency in our laboratory’s internal pooled patient database (PPDB) and publically available data sources were used. A variant was labeled ‘novel’ when it had not been previously reported in the literature or in any public database.

**Variant calling and classification**

Reads with average base quality <Q20 were excluded from the variant calling process, and the Bayesian approach was used to identify the consensus genotype at the variant locus. Each variant that was called out was assigned a Phred equivalent score that represents base-calling error probabilities. The identified variants in this study were called with a read quality >Q30 and a confidence score >50. The identified variants were labeled according to the ACMG recommendations for standards for the interpretation and reporting of sequence variations (Richards et al, 2015). The variants were classified into six categories: 1) pathogenic, 2) likely pathogenic, 3) variant of uncertain significance with probable damaging effect (VUSD), 4) variant of uncertain significance (VUS), 5) likely benign, and 6) benign.

**Copy number variation (CNV) Analysis**

In addition to SNVs and small indels, copy number analysis was performed to identify large deletions or duplications ranging from a single exon, multi-exons, whole gene, to multiple genes (continuous). The CNV analysis was performed by comparing normalized reads across 23 other samples from the same run (usually) by using Strand NGS (Mannan et al., 2016). For each sample, potential CNV changes in the genes of interest were identified by manual interpretation based on the following cutoffs, single exon cases, deletions: CN<1.2, Z-score>7; multi exon cases, deletions: CN<1.2, Z-score>5; duplications: CN>2.8, Z-score>5 , multi-gene cases, deletions: CN<1.3, Z-score>4; duplications: CN>2.7, Z-score>4 in the gene of interest.

The CNV calls which did not meet above mentioned CN and Z-score cut-off values, were validated by quantitative polymerase chain reaction (qPCR). For quantitative polymerase chain reaction (qPCR), primers spanning atleast two independent regions within the identified deletion/duplication were designed (primer sequence are available on request). The qPCR was done using a Mx3005P apparatus (Thermo Fisher Scientific, USA) using SYBR Fast qPCR kit (KAPA, USA) as per manufacture’s instruction. The deletions and duplications were determined using the comparative threshold cycle method (delta delta CT).

**Split read analysis for the identification of break points:**

Split read alignment was performed on the reads that did not align with an alignment score >95%, as described previously [3]. In brief, the input reads were split into two segments, which were then independently mapped to the reference genome and assessed for the split segments, which are the reads that aligned uniquely to two regions, with an alignment score of at least 97%. StrandNGS v2.6 was used to assess the split read alignment scores, which was used to call out large deletions, insertions, inversions, and translocation events.

**REFERENCES**

1. Mannan AU, Singh J, Lakshmikeshava R, et al (2016) Detection of high frequency of mutations in a breast and/or ovarian cancer cohort: Implications of embracing a multi-gene panel in molecular diagnosis in India. J Hum Genet 61:515–522. https://doi.org/10.1038/jhg.2016.4

2. Laboratories KD, Genetics M, Health O, et al (2015) HHS Public Access. Genet Med 17:405–424. https://doi.org/10.1038/gim.2015.30.Standards

3. Singh J, Mishra A, Pandian AJ, et al (2016) Next-generation sequencing-based method shows increased mutation detection sensitivity in an Indian retinoblastoma cohort. Mol Vis 22:1036–47

4. Mu W, Lu HM, Chen J, et al (2016) Sanger Confirmation Is Required to Achieve Optimal Sensitivity and Specificity in Next-Generation Sequencing Panel Testing. J Mol Diagnostics 18:923–932. https://doi.org/10.1016/j.jmoldx.2016.07.006

5. Baudhuin LM, Lagerstedt SA, Klee EW, et al (2015) Confirming variants in next-generation sequencing panel testing by sanger sequencing. J Mol Diagnostics 17:456–461. https://doi.org/10.1016/j.jmoldx.2015.03.004

**Figure S1:** This supplementary file illustrates the age distribution of the 1012 patients referred for neurological disorders.