**“Bioinformatics NGS Prioritization and Functional Annotation of Genes causing Seizures and Learning Disability.”**

**INTRODUCTION**

Intellectual Disability and Seizures are two different yet often interrelated conditions. Seizures are rapid disruptions in the brain which results in change of movements or loss of consciousness. Intellectual disability or Learning Disability is a condition which affect a person’s ability to learn or process any kind of information. Seizures accompany intellectual disability in almost half of the syndromes caused by mutation of genes on the X-chromosome. [[i]](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3531238/) Utilizing a comprehensive selection of 15 distinct variants, this intricate study delves in prioritizing and annotating them, to ascertain which variant is the cause of the phenotype by unlocking new insights and understanding.

**NEXT GENERATION SEQUENCE**

Next generation sequence brought in evolution in the field of genomics by allowing the identification of genetic variations of insertions, deletions, polymorphisms, and other structural variations which is quicker and cost-effective for generation of massive amount of DNA sequences. Involving multiple stages that must be completed to turn raw data into easily understandable information from the genome. The process includes Raw Sequencing and Quality Control based on the quality metrics alongside Mapping Reads. Furthermore, needed are Variant Calling and Annotation together with the Variant Prioritization to ensure high quality results.

The FASTQ file with Quality Scores for each base is the first result of raw sequencing, based on which the readings are trimmed and filtered during Quality Control. Mapping and Alignment is performed, to learn the difference between aligned reads and reference genomes. A Variant Call File (VCF) is created by variant calling to track variations in the sample which include functional information by Variant Annotation. Prioritisation ranks a gene depending on how frequently it is involved in the disease. NGS Data Analysis is an essential tool in determining the origin of complicated disorders.

NGS platforms provide vast quantities of data, but the associated error rates (~0.1–15%) are higher and the read lengths generally shorter (35–700 bp for short-read approaches) than those of traditional [Sanger sequencing](https://www.nature.com/articles/nrg.2016.49#Glos2) platforms, requiring careful examination of the results, particularly for variant discovery and clinical applications. Although long-read sequencing overcomes the length limitation of other NGS approaches, it remains considerably more expensive and has lower throughput than other platforms, limiting the widespread adoption of this technology in favour of less-expensive approaches. [[viii]](https://doi.org/10.1038/nrg.2016.49)

**PRIORITIZING THE VARIANTS**

Prioritisation is the process of identifying and ranking genetic variants that is most likely to be the cause of the phenotype by applying different filters like inheritance pattern, variant population frequency, pathway filter, annotation or pathogenicity. There are multiple ways to detect genes responsible for causing the phenotype like Targeted Exome and Custom Panel along with Whole Exome Sequencing and Whole Genome Sequencing. In spite of whichever method opted, the means to recognize the potentially disease-causing variants from others? Prioritization has become imperative now. At the beginning of the twenty-first century, many computational strategies for prioritising putative casual genes of a particular phenotype were presented due to the rapid accumulation of various types of genomic data. These techniques heavily rely on the similarities between disease genes' traits, such as sequence-based properties, expression patterns, and functional annotation data. Although these methods work well, they have several intrinsic drawbacks, such as insufficient and false-positive disease-causal gene data, imprecise disease boundaries, and a high degree of illness heterogeneity. (Zhu, J., Qin, Y., Liu, T. *et al.,* 2013) [[ii]](https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-14-S5-S5). Prioritizing becomes difficult if there is only single sample or multiple unrelated affected samples, whereas it is normal to use if the samples are siblings, trios or quartet. Prioritising the variants that will ultimately aid in the efficient and effective analysis of genetic data can be done by Inheritance Patterns like Recessive, Compound Heterozygous, De novo and Autosomal Dominant.

Diagram

Description automatically generated

Figure 1: Family Trio Pedigree

The variants discovered by WES is ranked based on criteria like Inheritance Pattern: homozygous variant, Variant Population Frequency: if gnomAD MAF% is less than 1% and Predicted Protein Pathogenicity: Damaging or Deleterious, for a family pedigree where proband and sibling have phenotypic symptoms while parents are asymptomatic. Variants that meet all 3 criteria are designated as having the highest priority coded in green, followed by those that satisfy 2 are second priority with blue color code, those that meet any one requirement is orange and those that meet none of the criteria is designated as having least priority coded in yellow.

By considering the *CACNA1S* variant gene which has heterozygous father – mother and homozygous proband; additionally, the gnomAD MAF % is 0.0021; finally, the pathogenicity shows damaged which ultimately satisfies all the criteria and is highly prioritized. Likewise, the *GRIN1* satisfies all 3 criteria and is marked green. Whereas, the *SCN9A* gene meets Inheritance pattern and Population Frequency, hence is coded blue with second priority. Further, *TMEM192, GLRA3, KCNN2, BRAF, CFTR, SLC52A2, PDE6G, BRCA1, ZNF57, HAUS8* and *KCNH2* when analyzed results meeting one of the criterions and hence marked third priority with orange color coded. Hindmost, *PKD1L2* gene which meets none has the least priority. The prioritized gene table is created in this manner with color coded genes as shown below.

Graphical user interface, application, table, Excel

Description automatically generated

Figure 2: Resultant table of prioritized genes

**FUNCTIONAL ANNOTATION**

Functional Annotation is describing all the biologically related data such as Function, Names and Taxonomy, Subcellular Location, Disease and Variants, PTM/ Processing, Expression, Interaction, Structure, Family and Domains, Sequence and Isoform which helps in understanding the contribution of different genes to phenotypes or diseases or any biological process. To annotate the prioritised genes, websites like PubMed, Uniprot, OMIM, GeneCards, Expression Atlas, Human Protein Atlas, STRING can be used. To ascertain the Protein, Function, Expression, Associated Phenotypes, Interactions, Animal Models, Phylogenetic Conservation, Protein Domains and Family Association, the gene name is inputted in the UniProt Database. Observation is done in the prioritised pattern.

*CACNA1S –* This encode voltage-gated calcium channels, the role of the calcium channels and tissues is the reason why mutations in CACNA1S are not associated with seizures as it is usually in skeletal muscle by regulating muscle contractions or disorders like hypokalaemia periodic paralysis and malignant hyperthermia.[[iii]](https://pubmed.ncbi.nlm.nih.gov/15001631/) Hence, CACNA1S is opted out from the list of variants causing the phenotype.

*GRIN1* - Heterozygous variants of GRIN1, encoding the GluN1 subunit of the NMDA receptor, have been reported in patients with neurodevelopmental disorders including epileptic encephalopathy, severe intellectual disability, and movement disorders. [[iv]](https://pubmed.ncbi.nlm.nih.gov/28051072/) This gives the evidence that developmental and epileptic encephalopathy-101 (DEE101) is caused by homozygous mutation in the GRIN1 gene ([138249](https://www.omim.org/entry/138249)) on chromosome 9q34. The Disease & Variants Annotation of the GRIN1 in the UniProt shows that this variant may be the cause of the phenotype as it mentions both Intellectual Disability (Learning Disability) and Seizures.

Table

Description automatically generated

Figure 3: Table showing Epilepsy symptoms in GRIN1

*SCN9A –* The observation in UniProt proclaims that their symptoms of generalized epilepsy with febrile seizures-plus, hence it can be the cause of the phenotype as well. In a genomic region on human chromosome 2 known to harbour the febrile seizure SCN1A sodium channel gene, we now report a disease-causing mutation in the adjacent gene, SCN9A (Nav1.7), in a large family with febrile seizures.[[v]](https://pubmed.ncbi.nlm.nih.gov/19763161/)

Graphical user interface, text, application, email

Description automatically generated

Figure 4: UniProt showing the Seizure Disease annotation of SCN9A

*KCNN2 -* An autosomal dominant disorder characterized by motor and language developmental delay, intellectual disability often associated with early-onset movement disorders comprising cerebellar ataxia and/or extrapyramidal symptoms. Other variable features include autism spectrum disorder or autistic features and epilepsy. Patients with KCNN2 variants had motor and language developmental delay, intellectual disability often associated with early-onset movement disorders comprising cerebellar ataxia and/or extrapyramidal symptoms. [[vi]](https://research.vumc.nl/en/publications/variants-in-the-sk2-channel-gene-kcnn2-lead-to-dominant-neurodeve) In addition, by UniProt annotations, we can conclude this variant can be a cause of the phenotype.

*BRAF -* The annotation observation is that the variant causes intellectual disability which is categorised under the Noonan Syndrome 7(NS7). This syndrome includes various disorders like Motor delay, variable intellectual deficits, skeletal defects, deafness. [[vii]](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4028130/) Therefore, BRAF gene might be the cause of the disease as well.

Similarly, annotations on *TMEM192, GLRA3, CFTR, SLC52A2, PDE6G, BRCA1, ZNF57, HAUS8, KCNH2* and *PKD1L2* demonstrates no symptoms of Seizures or Learning disability in any of the genes.

This, therefore, shows that amongst the 15 genes, SCN9A, GRIN1, KCNN2 and BRAF are the genes exhibiting phenotypic symptoms.

**CONCLUSION**

NGS Gene Prioritization and Annotation techniques are a great aid and has paved a quicker and easier way for Genome Analysis by prioritizing the genes based on Inheritance pattern, Population frequency and Pathogenicity. Additionally, Functional Annotation using tools which provide accurate and reliable results for identifying the disorder-causing gene by providing the underlying insights of each gene. The utilization of these tools and techniques can help improve the treatment and understanding of complex genetic disorders. This concludes that *KCNN2* and *GRIN1* is causative gene for both Seizures and Learning Disability, while *SCN9A* is the causative gene for Seizures and *BRAF* is for Intellectual Disability.

**REFERENCES**

1. Seizures and X-linked intellectual disability - [Eur J Med Genet.](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3531238/) Author manuscript; available in PMC 2013 May 1. *Published in final edited form as:* [Eur J Med Genet. 2012 May; 55(5): 307–312.](https://www.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&retmode=ref&cmd=prlinks&id=22377486)Published online 2012 Feb 8. doi: [10.1016/j.ejmg.2012.01.017](https://doi.org/10.1016%2Fj.ejmg.2012.01.017)
2. Zhu, J., Qin, Y., Liu, T. *et al.* Prioritization of candidate disease genes by topological similarity between disease and protein diffusion profiles. *BMC Bioinformatics* **14** (Suppl 5), S5 (2013). <https://doi.org/10.1186/1471-2105-14-S5-S5>
3. Kung AW, Lau KS, Fong GC, Chan V. Association of novel single nucleotide polymorphisms in the calcium channel alpha 1 subunit gene (Ca(v)1.1) and thyrotoxic periodic paralysis. J Clin Endocrinol Metab. 2004 Mar;89(3):1340-5. doi: 10.1210/jc.2003.030924 . <https://pubmed.ncbi.nlm.nih.gov/15001631/> PMID: 15001631.
4. Rossi M, Chatron N, Labalme A, Ville D, Carneiro M, Edery P, des Portes V, Lemke JR, Sanlaville D, Lesca G. Novel homozygous missense variant of GRIN1 in two sibs with intellectual disability and autistic features without epilepsy. Eur J Hum Genet. 2017 Feb;25(3):376-380. doi: 10.1038/ejhg.2016.163. Epub 2017 Jan 4. PMID: 28051072; PMCID: PMC5315503. <https://pubmed.ncbi.nlm.nih.gov/28051072/>
5. Singh NA, Pappas C, Dahle EJ, Claes LR, Pruess TH, De Jonghe P, Thompson J, Dixon M, Gurnett C, Peiffer A, White HS, Filloux F, Leppert MF. A role of SCN9A in human epilepsies, as a cause of febrile seizures and as a potential modifier of Dravet syndrome. PLoS Genet. 2009 Sep;5(9):e1000649. doi: 10.1371/journal.pgen.1000649. Epub 2009 Sep 18. PMID: 19763161; PMCID: PMC2730533. <https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1000649>
6. Mochel F, Rastetter A, Ceulemans B, Platzer K, Yang S, Shinde DN, Helbig KL, Lopergolo D, Mari F, Renieri A, Benetti E, Canitano R, Waisfisz Q, Plomp AS, Huisman SA, Wilson GN, Cathey SS, Louie RJ, Gaudio DD, Waggoner D, Kacker S, Nugent KM, Roeder ER, Bruel AL, Thevenon J, Ehmke N, Horn D, Holtgrewe M, Kaiser FJ, Kamphausen SB, Abou Jamra R, Weckhuysen S, Dalle C, Depienne C. Variants in the SK2 channel gene (KCNN2) lead to dominant neurodevelopmental movement disorders. Brain. 2020 Dec 1;143(12):3564-3573. doi: 10.1093/brain/awaa346. Erratum in: Brain. 2023 Feb 07;: PMID: 33242881. <https://research.vumc.nl/en/publications/variants-in-the-sk2-channel-gene-kcnn2-lead-to-dominant-neurodeve>
7. Sarkozy A, Carta C, Moretti S, Zampino G, Digilio MC, Pantaleoni F, Scioletti AP, Esposito G, Cordeddu V, Lepri F, Petrangeli V, Dentici ML, Mancini GM, Selicorni A, Rossi C, Mazzanti L, Marino B, Ferrero GB, Silengo MC, Memo L, Stanzial F, Faravelli F, Stuppia L, Puxeddu E, Gelb BD, Dallapiccola B, Tartaglia M. Germline BRAF mutations in Noonan, LEOPARD, and cardiofaciocutaneous syndromes: molecular diversity and associated phenotypic spectrum. Hum Mutat. 2009 Apr;30(4):695-702. doi: 10.1002/humu.20955. PMID: 19206169; PMCID: PMC4028130. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4028130/>
8. Goodwin, S., McPherson, J. & McCombie, W. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* **17**, 333–351 (2016). <https://doi.org/10.1038/nrg.2016.49>