A tiered strategy for variant selection in patients with epilepsy

Alexandra Fonzi[1], Lorenzo Scalzini[2], Simon WM Eng[3], Kevin CH Ha[3], Rokshana Stephny Geread[3], Riccardo Ottalevi[2], Michela Luciano[1], Isabella Baldini[1], Riccardo Paone[1], Generoso Ianniciello[1], Juan Caceres[2]

[1]Dante Genomics, Research and Development, L'Aquila, Italy [2]Dante Genomics, Research & Development - Software and Bioinformatics, L'Aquila, Italy [3]BioSymetrics, Research and Development, Toronto, Canada

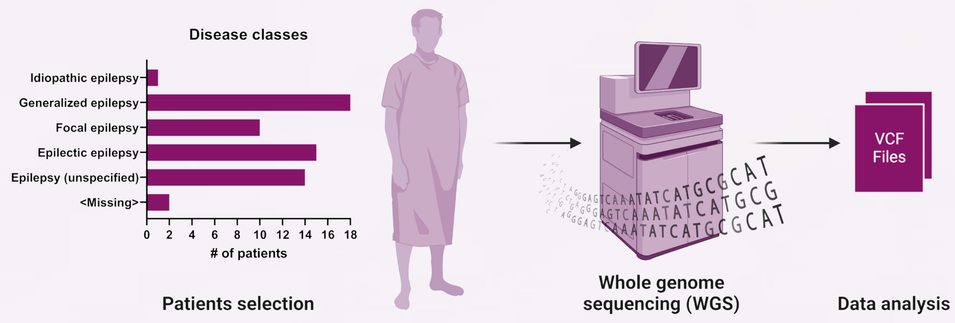
Introduction

Epilepsy is one of the most common neurological conditions often characterized by recurrent episodes of seizures. The etiology is heterogeneous, however genetic factors are thought to play a major role *(1)*. In the past decade, Next Generation Sequencing (NGS) resulted in the identification of novel sequence variants associated with susceptibility to epilepsy *(2)*. This set of new knowledge has already been used to identify personalized therapies for affected patients *(3,4)*.

Objective

In this study, we sought to discover novel variants associated with epilepsy with the aim to pinpoint candidate genes as therapeutic targets.

Methods



58 patients with different phenotype of epilepsy were selected for this study.

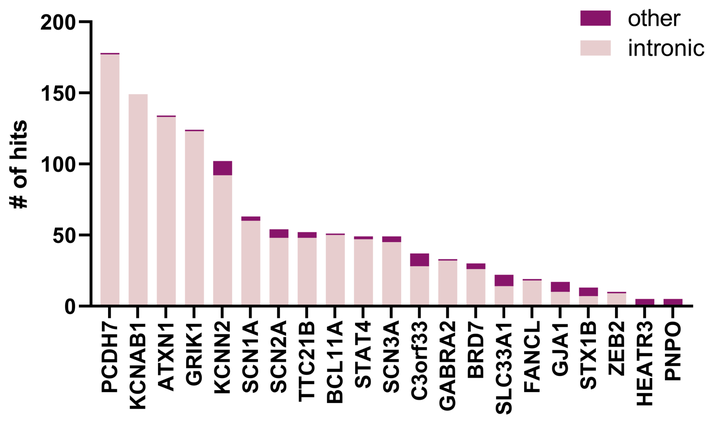
DNA samples were sequenced (WGS) and variant analysis was performed using Variant Call Files (VCFs) computed with DRAGEN (Illumina) on the human reference genome GRCh37.

A three-tiered strategy was used to identify epilepsy-associated variants in novel genetic loci.

- In Tier 1, we analyzed variants in known epilepsy-associated genes (*SCN1A*) and from a medical diagnosis of a patient in the cohort (*WDR45*) to serve as validation of observable signals within the population.
- In Tier 2, we focused on identification of mutations within 16 epilepsy-associated genetic loci, encompassing 21 genes, known from prior literature (2).
- In Tier 3, we conducted a more exploratory approach by expanding to unblinding prospective bioinformatic assessment of all sequence variants in order to identify novel candidate genes.

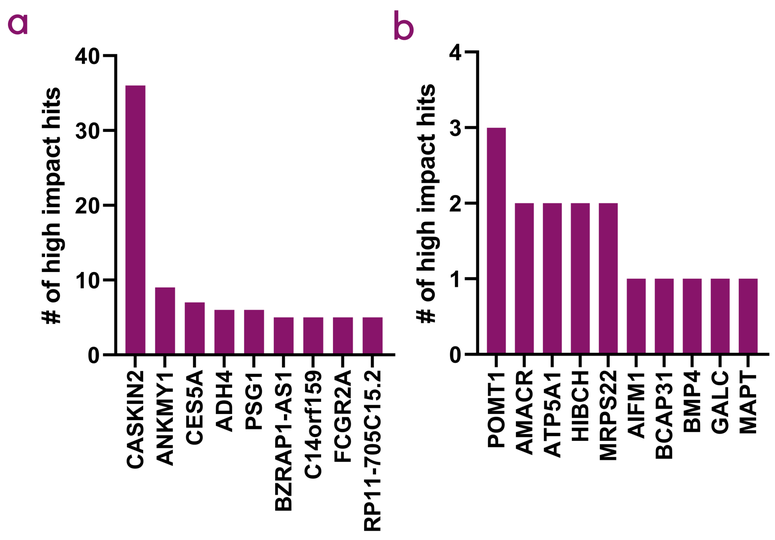
Variant annotation was performed with Ensembl's Variant Effect Predictor (VEP), version 103.1, for the human genome assembly GRCh37. Candidate genes in Tier 1 and 2 were validated by overlap with high scoring candidates in literature references. In Tier 3, novel associations were predicted using different metrics and database annotations, such as: maximum allele frequency, SIFT score, PolyPhen score, clinical significance, and impact. Biological pathways and genes associated with the hits were identify using the g:GOSt module of g:Profiler.

Results



1. All International League of Associations for Epilepsy (ILAE) genes represented in hits.

Fig 1. Hits in known epilepsy-associted genes. *Number of hits (y-axis) per gene identified in ILAE, Nat Communications (2018) (x-axis), regardless of consequence. These hits were largely driven by intronic variants.*

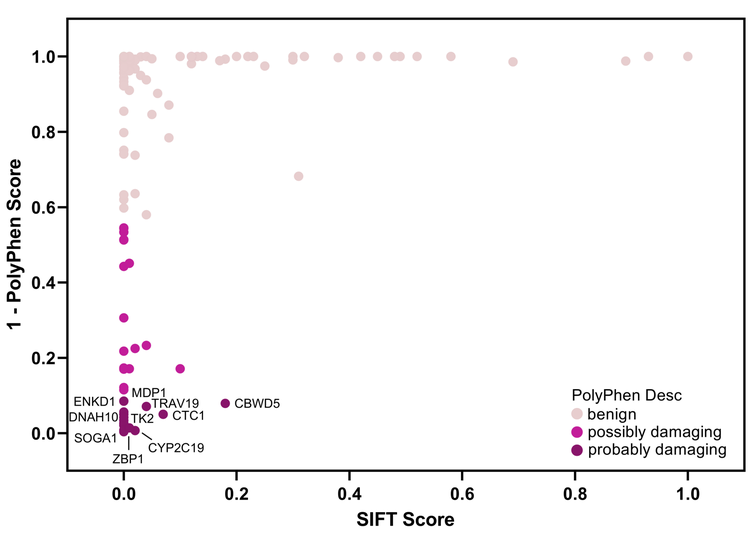


2. Potentially novel candidate genes based on impact rating of hits

Fig 2. a) High-impact hits as annotated by VEP. *9 genes with the highest number of high IMPACT hits out of 50 are shown. CASKIN2, a paralog of CASKIN1, which interacts with calcium/calmodulin-dependent serine protein kinase (CASK), showed up with highest number of significant high impact hits, with 36.*
b) High-impact hits in genes identified from literature review. *POMT1, an O-mannosyltransferase related to establishing synapses in the retina in a mouse model, had the highest number of significant high impact hits.*

3. High-impact pathogenic variants grouped by gene

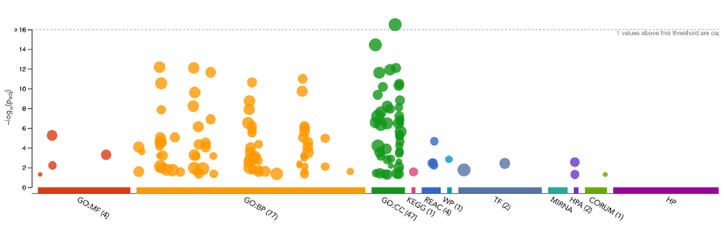
Fig 3. Clinical significance og high impact variants. *Severity of the variant was computed on ClinVar. 11 genes containing a pathogenic or likely pathogenic term out of 58 are shown. Majority of genes have at most one high impact pathogenic variant.*



4. Potentially novel gene candidates identified by PolyPhen and SIFT scores

Fig 5. High impact variants annotated by SIFT and PolyPhen (maximum allele frequency < 5%) *Comparison of 85 high impact variants that have both SIFT and PolyPhen scores. PolyPhen scores are shown as 1 - PolyPhen so that it follows the same trend as SIFT scores (y-axis). Lower values indicate a more damaging variant, therefore the variants labelled in the bottom left corner represent those predicted to be damaging by both PolyPhen and SIFT.*

a



b

Gene	Full name	# significant terms	Identified in literature review?	Function
NRG1	Neuregulin 1	71		Neuronal development
NRXN1	Neurexin 1	104	√	Neuronal cell adhesion
PTPRD	Protein tyrosine phosphatase receptor type D	73		
DCC	DCC netrin 1 receptor	71		
LRRC4C	Leucin rich repeat containing 4C	71		
CNTNAP2	Contactin associate protein 2	58	√	
CTNND2	Catenin delta 2	52		
CTNNA2	Catenin alpha 2	52		
GRID2	Glutamate ionotropic receptor delta type subunit 2	90		Synapse-related functions
NLGN4X	Neuroligin 4 X-linked	75		

5. Potentially novel gene candidates identified by enrichment analysis of hits

Fig 4. a) g:Profiler results for the 200 genes with the highest number of hits. *The enrichment analysis identified numerous terms within the biological process (GO:BP) and cellular components (GO:CC) subdomain of GO. Each colored horizontal bar represents a collection of terms whose width scales with the number of terms in that collection. Each significant term is a circle. Their p-values are shown on the y-axis.*
b) Top 10 genes by number of significant terms sorted by function. *Hand-curated function related to the epilepsy phenotype of genes with highest number of significant terms.*

Legend

Hit = variant with observed alternative allele frequency that is significantly different from 1,000 Genomes. Determined using binomial tests.
Impact rating identifies the severity of a variant's impact on the protein.
High-impact= variant assumed to have high (disruptive) impact on the protein, probably causing protein truncation, loss of function or triggering nonsense mediated decay.
A **SIFT score** between 0-0.05, indicates that the variant is predicted to be probably damaging.
A **PolyPhen score** between 0.4-1.0, indicates a variant is predicted to be possibly to probably damaging.

Summary

We were able to identify:

- clinically relevant genetic variants in known epilepsy-associated genes, including *SCN1A, and WDR45 (data not shown)*;
- high number of hits in genes included in a list previously identified from literature *(Fig 1)*;
- novel disease-gene relationships by prioritizing genes with high number of hits associated with significant high impact rating *(Fig 2)*, classified as pathogenic *(Fig 3),* or probably damaging *(Fig 4)*;
- candidates involved with neuronal development, neuronal cell adhesion, and synapse-related functions *(Fig 5)*.

Conclusions

This study identifies novel genetic variants in patients with confirmed epileptic conditions, extending scopes of clinical diagnosis and providing the groundwork for future functional studies to characterize the exact molecular mechanisms of the genes identified in the etiopathogenesis of the disease and pinpoint candidate targets for new therapeutic approaches.

References

1. *Perucca P, Bahlo M, Berkovic SF. The Genetics of Epilepsy. Annu Rev Genomics Hum Genet. 2020;21:205-30.*
2. *International League Against Epilepsy Consortium on Complex Epilepsies. Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies. Nat Commun. 2018;9(1):5269.*
3. *Møller RS, Hammer TB, Rubboli G, Lemke JR, Johannesen KM. From next-generation sequencing to targeted treatment of non-acquired epilepsies. Expert Rev Mol Diagn. 2019;19(3):217-28.*
4. *Møller RS, Dahl HA, Helbig I. The contribution of next generation sequencing to epilepsy genetics. Expert Rev Mol Diagn. 2015;15(12):1531-8.*

Visit our website framing the qr code

Authors declare no conflict of interests