Computational analysis of the phenotypic spectrum of SCN2A-related disorders

Shridhar Parthasarathy, Julie Xian, Katherine Crawford, Katherine Helbig, Peter Galer, David Lewis-Smith, Michael Kaufman, Eryn Fitch, Shiva Ganesan, Margaret O'Brien, Veronica Codoni, Colin Ellis, Laura Conway, Deanne Taylor, Roland Krause, Ingo Helbig

Introduction

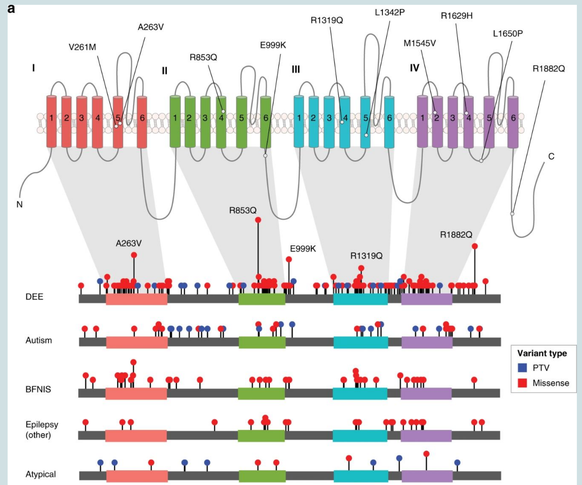
Over 100 genetic etiologies are known for neurodevelopmental disorders. Pathogenic variants in *SCN2A* ($Na_V1.2$, neuronally expressed type II voltage-gated sodium channel gene) are commonly implicated in benign familial infantile seizures (BFNIS), autism spectrum disorders (ASD), and developmental and epileptic encephalopathies (DEE). Current understanding of *SCN2A*-related disease phenotypes is limited by overlapping presentations, sparse quantitative data, and heterogeneous and limited clinical descriptions. Simple associations between gain- or loss-of-function variants and broad phenotypes do not exist.

The Human Phenotype Ontology (HPO) is a standardized phenotypic dictionary that is useful for quantitative high-throughput phenotypic analysis. Here, clinical phenotypes are computationally analyzed in 413 individuals with pathogenic *SCN2A* mutations using over 10,000 HPO phenotypic annotations to elucidate the spectrum of *SCN2A*-related disorders.

Methods

- Literature review for *SCN2A*-related disorder cases 2001-2019
- Included 21 unreported individuals within the Epilepsy Genetics Research Project, CHOP
- Manual annotation of HPO terms, including negative terms denoting explicit absence of phenotype
- Assignment of gain- or loss-of-function labels to variants with available functional experimental results
- Automated data harmonization with more general phenotypes
- Determination of information content, frequency, distribution of each phenotypic term
- Association analysis with the ion channel topology and variant types
- Phenotypic similarity analysis: generated null distributions of 100,000 bootstrapped samples of size *n* to determine if a group of *n* individuals is significantly similar to each other in the cohort

Phenotypic Annotation



b

HPO	DEE	Autism	BFNIS	Epilepsy	Atypical	Definition
HP:0001250	1	0.09	1	1	0.11	Seizures
HP:0012759	0.81	0.65	0.06	0	0.67	Neurodevelopmental abnormality
HP:0001311	0.68	0.07	0.14	0.17	0.06	Abnormal nervous system electrophysiology
HP:0002353	0.68	0.07	0.14	0.17	0.06	EEG abnormality
HP:0030178	0.68	0.07	0.14	0.17	0.06	Abnormality of CNS electrophysiology
HP:0000708	0.30	1	0.04	0	0.28	Behavioral abnormality
HP:0000729	0.23	1	0	0	0	Autistic behavior
HP:0000717	0.18	0.93	0	0	0	Autism
HP:0012759	0.81	0.65	0.06	0	0.67	Neurodevelopmental abnormality
HP:0001249	0.59	0.28	0	0	0.50	Intellectual disability
HP:0001250	1	0.09	1	1	0.11	Seizures
HP:0410280	0.42	0.05	0.57	0.22	0.06	Pediatric onset
HP:0003593	0.23	0.03	0.14	0.11	0.06	Infantile onset
HP:0002197	0.59	0.03	0.43	0.19	0	Generalized-onset seizure
HP:0007359	0.42	0.02	0.43	0.22	0	Focal-onset seizure
HP:0001250	1	0.09	1	1	0.11	Seizures
HP:0003623	0.35	0.02	0.14	0.11	0	Neonatal onset
HP:0001311	0.68	0.07	0.14	0.17	0.06	Abnormal nervous system electrophysiology
HP:0002353	0.68	0.07	0.14	0.17	0.06	EEG abnormality
HP:0030178	0.68	0.07	0.14	0.17	0.06	Abnormality of CNS electrophysiology
HP:0012759	0.81	0.65	0.06	0	0.67	Neurodevelopmental abnormality
HP:0001249	0.59	0.28	0	0	0.50	Intellectual disability
HP:0000708	0.30	1	0.04	0	0.28	Behavioral abnormality
HP:0012758	0.31	0.28	0.06	0	0.33	Neurodevelopmental delay
HP:0100753	0	0	0	0	0.28	Schizophrenia

(a) The $Na_V1.2$ channel (above) and gene (below), highlighting a selection of recurrent variants. (b) frequency of phenotypic features within categorized phenotypic subgroups: developmental and epileptic encephalopathy (n = 255), autism (ASD, n = 60), benign familial neonatal–infantile seizures (BFNIS, n = 53), Other epilepsy (n = 27), and atypical *SCN2A*-related phenotypes (n = 18). Boxed frequencies indicate the five most frequent Human Phenotype Ontology (HPO) terms within each respective phenotypic subgroup. CNS - central nervous system, EEG - electroencephalogram, PTV - protein-truncating variant.

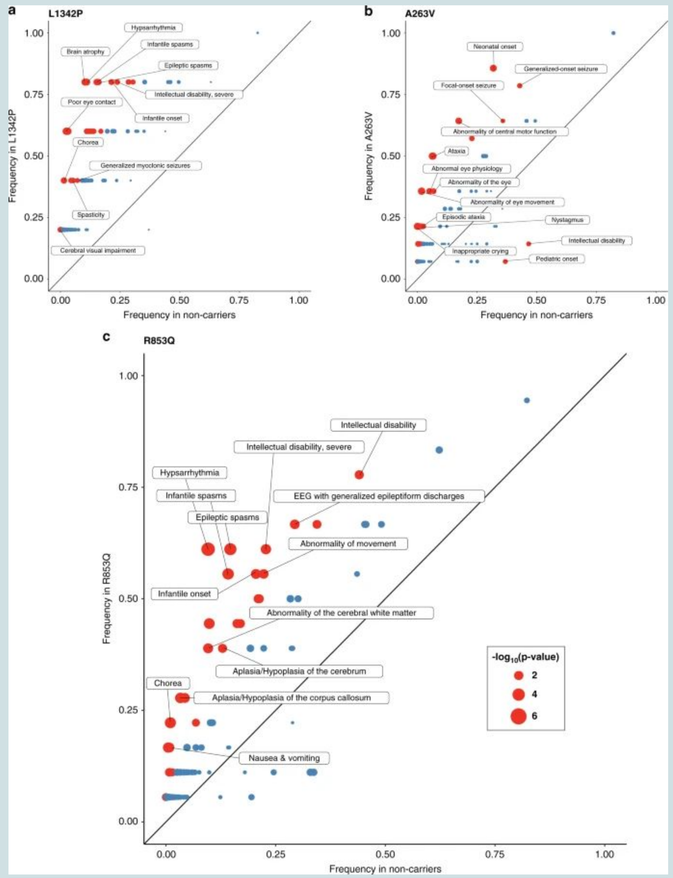
Phenotypic Annotations Distributed Through $Na_V1.2$

- 2,935 phenotypic terms encoded ("base" terms)
- 10,860 total terms ("propagated" terms), 562 unique

Identification of Subgroups

Similar Sugroups Identified by Phenotypic Similarity Analysis

- Individuals with any missense variant (p=0.01, n=341)
- Individuals with missense variants in S1 segment (n=18)
  - Driven by status epilepticus, EEG w/ abnormally slow frequencies
- Individuals with missense variants in domain IV (n=65)
  - Driven by polymicrogyria, spastic tetraplegia, tremor, EEG abnormality
- Individuals with 8/62 recurrent variants including



Frequency of phenotypic terms in individuals with variant (a) p.L1342P, (b) p.A263V, (c) p.R853Q vs remainder of the cohort. Red points indicate terms with uncorrected p value <0.05, blue points indicate terms with uncorrected p value ≥0.05.

Phenotypic Associations

Variant Type Associations

- Missense variants strongly positively associated with neonatal onset, seizures, epileptic spasms, and 19 other phenotypes, as well as absence of intellectual disability
- Protein-truncating variants (PTV) strongly positively associated with behavioral abnormality, autism, autistic behavior, and 10 others, as well as absence of seizures

Topological Associations

- Missense variants in the S5-S6 selectivity filter of the channel pore-forming loop domain strongly positively associated with autism and negatively with seizures

Discussion

Conclusions

- We mapped all known data on *SCN2A*-related diseases to $Na_V1.2$ using a standardized phenotypic framework
- 562 distinct phenotypes were evaluated and analyzed computationally through over 10,000 annotations
- We found that individuals with missense variants are more likely to present with seizures and abnormal EEG , and less likely autism, than those with PTV
- Missense mutations in S5-S6 are more likely to result in autism and no seizures than others, more closely resembling PTV
- We identified novel genotype-phenotype correlations, including grouping of individuals with any missense variant, S1 segment missense variants, domain IV missense variants, and eight recurrent variants into significantly similar subgroups that reliably produce distinct phenotypes

Limitations and Future Directions

- Using HPO topology alone without external data limits the quantity of inferrable phenotypic relationships
- Future study using data-enriched vectorization of HPO
- Further functional experiments can also elucidate mechanisms relating variants to specific phenotypes

Acknowledgements

Helbig Lab and the Epilepsy Genetics Research Project at Children's Hospital of Philadelphia, Philadelphia, PA