Similarity among epilepsy genes

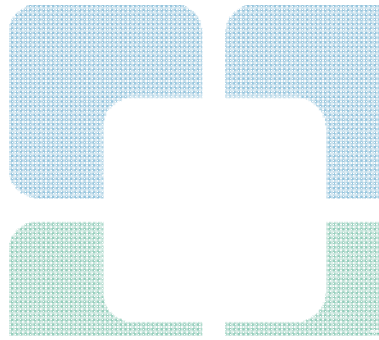
September – 13th

Eduardo Pérez-Palma
PhD

Cleveland Clinic

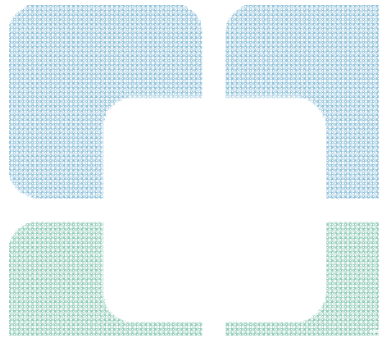
Outline

- Gene similarity and paralog conservation
- Leveraging gene similarity for variant interpretation:
  - Paralog z-score
  - Identification of Pathogenic Enriched Regions (PERs)
  - User-friendly web application
- Summary



Outline

- Gene similarity and paralog conservation
- Leveraging gene similarity for variant interpretation:
  - Paralog z-score
  - Identification of Pathogenic Enriched Regions (PERs)
  - User-friendly web application
- Summary

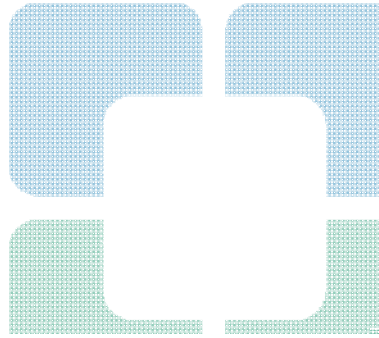


Developmental Epileptic Encephalopathies (DEEs)

Severe & complex

Monogenic

Incidence: ~1:6.000

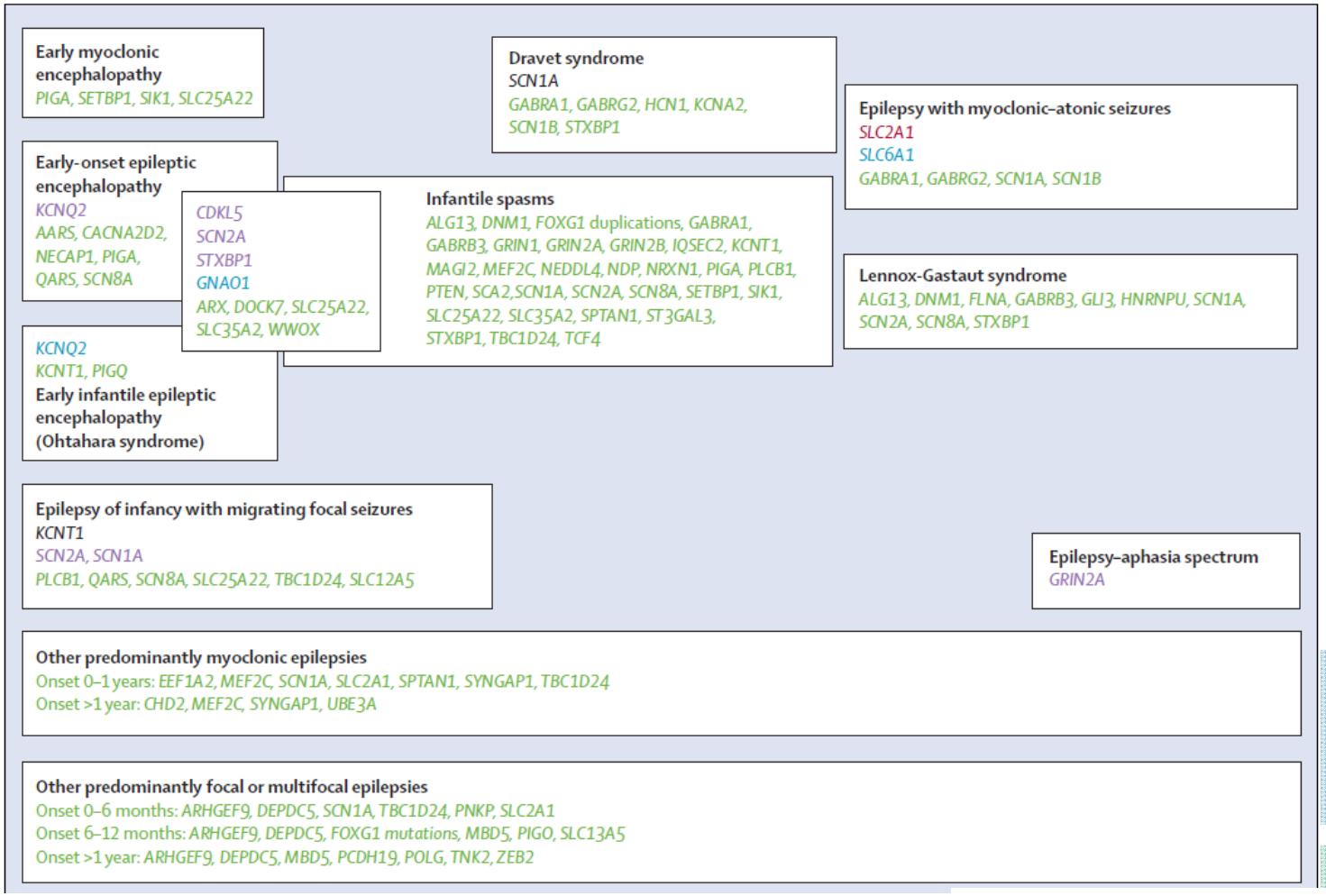


Developmental Epileptic Encephalopathies (DEEs)

Severe & complex

Monogenic

Incidence: ~1:6.000



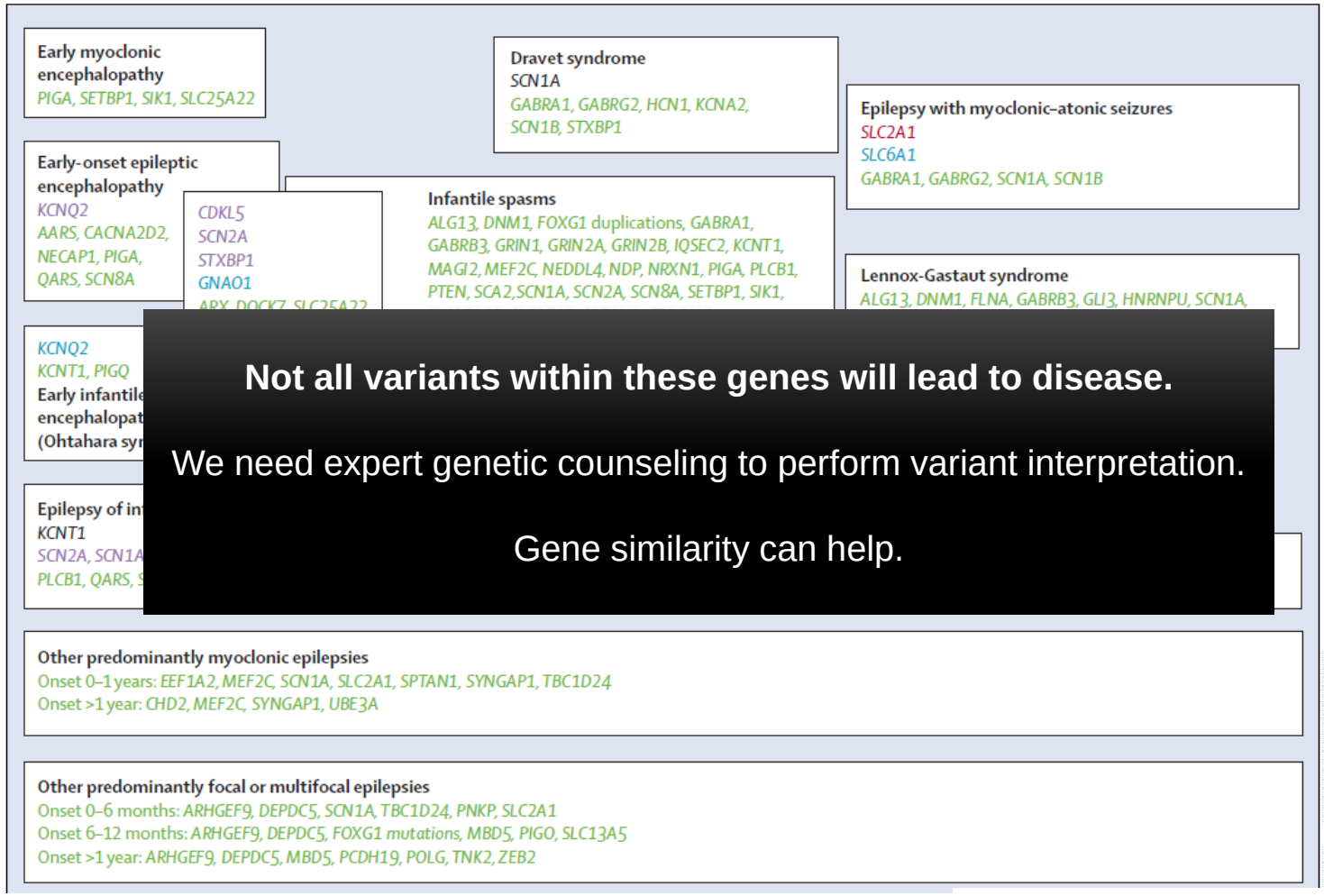
McTague A. *et al,* 2015

Developmental Epileptic Encephalopathies (DEEs)

Severe & complex

Monogenic

Incidence: ~1:6.000



McTague A. *et al,* 2015

Measuring gene similarity

Across species
(Orthologs)

***SCN1A* Human**

***Scn1a* Mice**

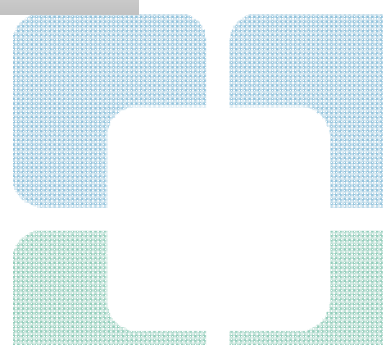
***scn1a* Zebrafish**

Across human related genes
(Paralogs-families)

***SCN1A* Human**

***SCN2A* Human**

***SCN3A* Human**



Ortholog similarity facilitates disease model generation

Across species
(Orthologs)

SCN1A Human

Scn1a Mice

scn1a Zebrafish

ELSEVIER journal homepage: www.elsevier.com/locate/yebeh

Cannabis constituents reduce seizure behavior in chemically-induced and *scn1a*-mutant zebrafish

Cammi Thornton, Kennedy E. Dickson, Dennis R. Carty, Nicole M. Ashpole, Kristine L. Willett *

Department of BioMolecular Sciences, School of Pharmacy, University of Mississippi, University, MS 38677, United States of America

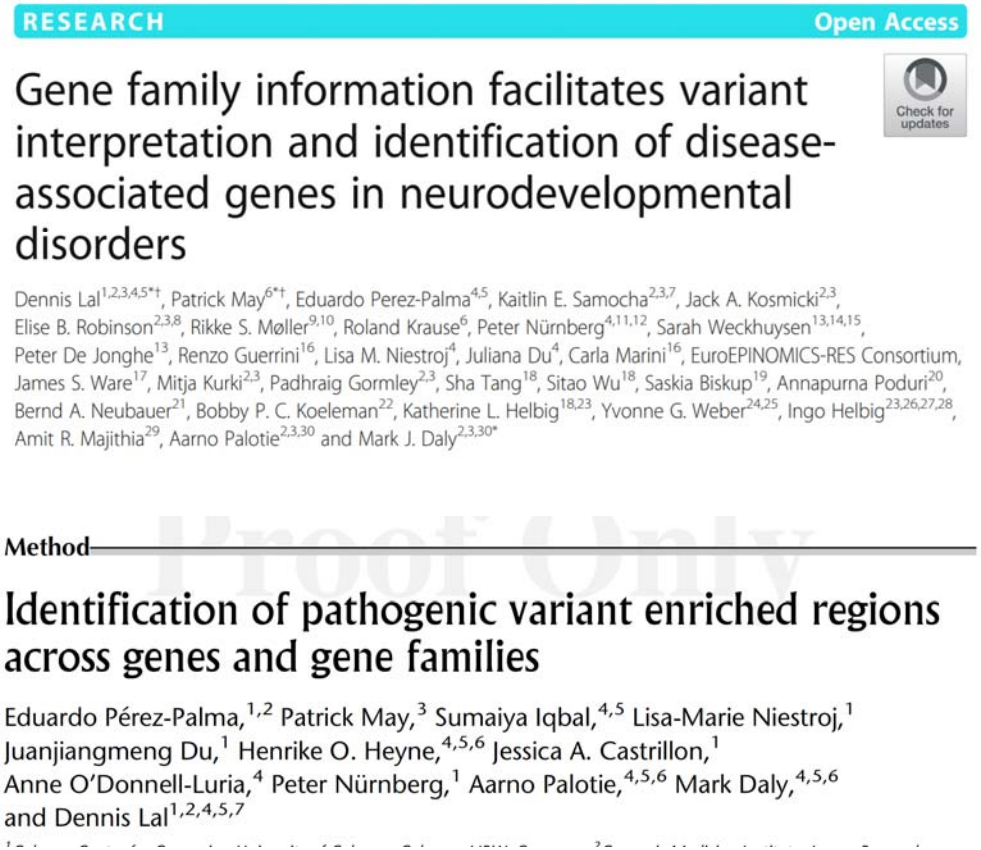
BRIEF COMMUNICATION

Epilepsia

Focal and generalized seizure activity after local hippocampal or cortical ablation of $Na_V1.1$ channels in mice

Nico A. Jansen[1] | Anisa Dehghani[1] | Cor Breukel[1] | Else A. Tolner[1,2] | Arn M. J. M. van den Maagdenberg[1,2]

Paralog similarity facilitates variant interpretation



RESEARCH Open Access

Gene family information facilitates variant interpretation and identification of disease-associated genes in neurodevelopmental disorders

Dennis Lal[1,2,3,4,5*†], Patrick May[6*†], Eduardo Perez-Palma[4,5], Kaitlin E. Samocha[2,3,7], Jack A. Kosmicki[2,3], Elise B. Robinson[2,3,8], Rikke S. Møller[9,10], Roland Krause[6], Peter Nürnberg[4,11,12], Sarah Weckhuysen[13,14,15], Peter De Jonghe[13], Renzo Guerrini[16], Lisa M. Niestroj[4], Juliana Du[4], Carla Marini[16], EuroEPINOMICS-RES Consortium, James S. Ware[17], Mitja Kurki[2,3], Padhraig Gormley[2,3], Sha Tang[18], Sitao Wu[18], Saskia Biskup[19], Annapurna Poduri[20], Bernd A. Neubauer[21], Bobby P. C. Koeleman[22], Katherine L. Helbig[18,23], Yvonne G. Weber[24,25], Ingo Helbig[23,26,27,28], Amit R. Majithia[29], Aarno Palotie[2,3,30] and Mark J. Daly[2,3,30*]

Method

Identification of pathogenic variant enriched regions across genes and gene families

Eduardo Pérez-Palma,[1,2] Patrick May,[3] Sumaiya Iqbal,[4,5] Lisa-Marie Niestroj,[1] Juanjiangmeng Du,[1] Henrike O. Heyne,[4,5,6] Jessica A. Castrillon,[1] Anne O'Donnell-Luria,[4] Peter Nürnberg,[1] Aarno Palotie,[4,5,6] Mark Daly,[4,5,6] and Dennis Lal[1,2,4,5,7]

Across human related genes (Paralogs-families)

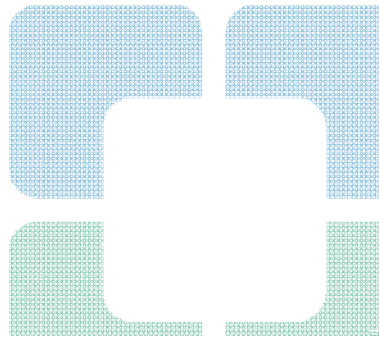
***SCN1A* Human**

***SCN2A* Human**

***SCN3A* Human**

Outline

- Gene similarity and paralog conservation
- Leveraging gene similarity for variant interpretation:
  - Paralog z-score
  - Identification of Pathogenic Enriched Regions (PERs)
  - User-friendly web application
- Summary



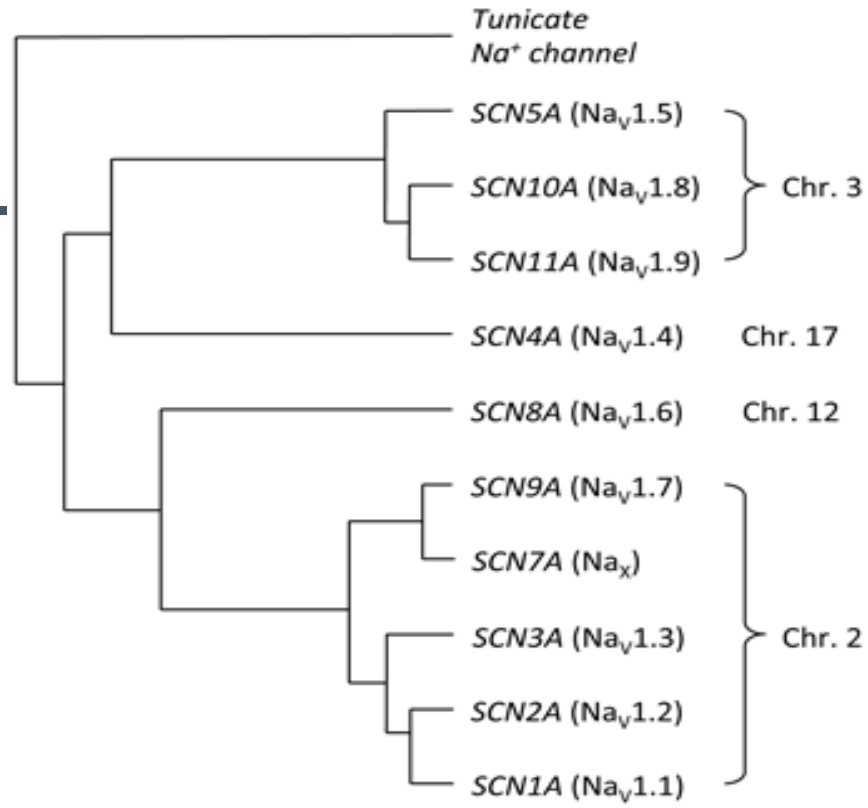
Many DEEs associated genes are related and belong to gene families

List of epilepsy associated and “potentially associated” genes

ALDH7A1	DNM1	KCNB1	PRICKLE2	SLC6A1
ALG13	EEF1A2	KCNC1	PRRT2	SNIP1
ARHGEF9	EPM2A	KCNMA1	RELN	SPTAN1
ARX	GABRA1	KCNQ2	SCARB2	SRPX2
ASAH1	GABRB3	KCNQ3	SCN1A	ST3GAL3
CDKL5	GABRG2	KCNT1	SCN1B	STRADA
CHD2	GNAO1	KCTD7	SCN2A	STX1B
CHRNA2	GOSR2	LGI1	SCN8A	STXBP1
CHRNA4	GRIN1	MEF2C	SCN9A	SYN1
CHRNB2	GRIN2A	NHLRC1	SIAT9	SYNGAP1
CLN8	GRIN2B	PCDH19	SIK1	SZT2
CNTNAP2	HCN1	PLCB1	SLC13A5	TBC1D24
CPA6	HNRNPU	PNKP	SLC25A22	WWOX
CSTB	IER3IP1	PNPO	SLC2A1	
DEPDC5	KCNA2	PRICKLE1	SLC35A2	

Genes of the same gene family have the same ancestral gene

Ancestral *GENE*



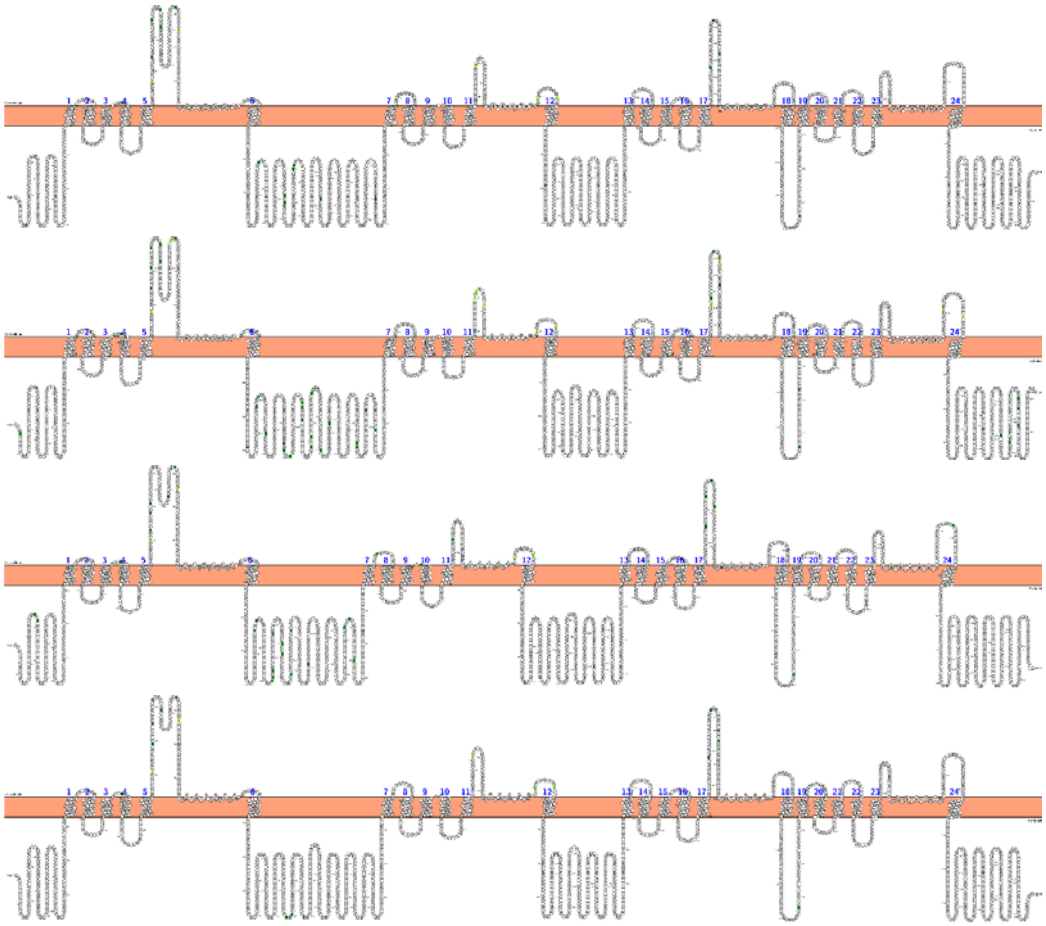
Paralogs have high sequence similarity

SCN1A

SCN2A

SCN5A

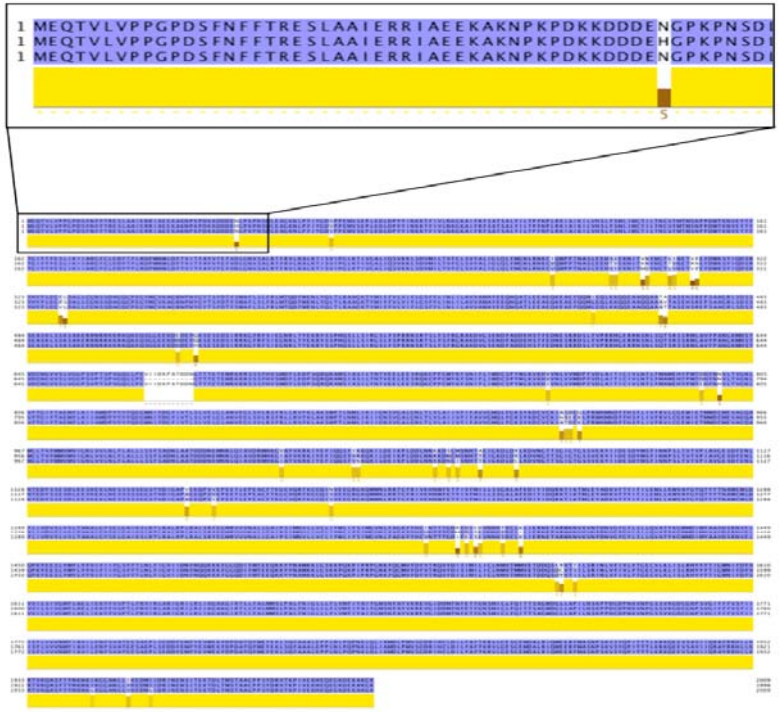
SCN8A



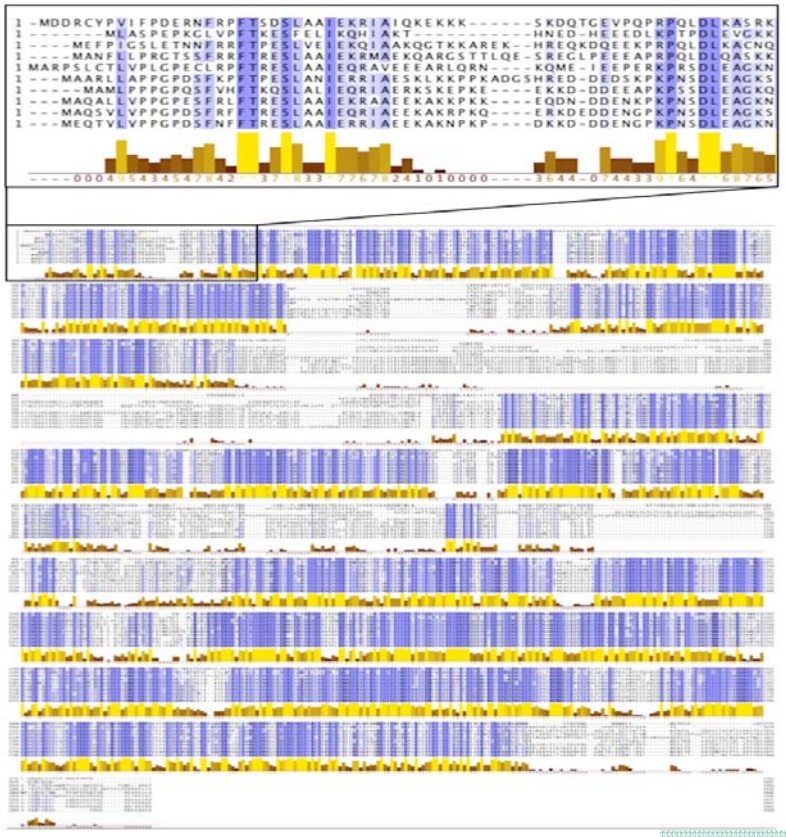
Omasits U. *et al*, 2014

Paralog conservation is very different than Ortholog conservation

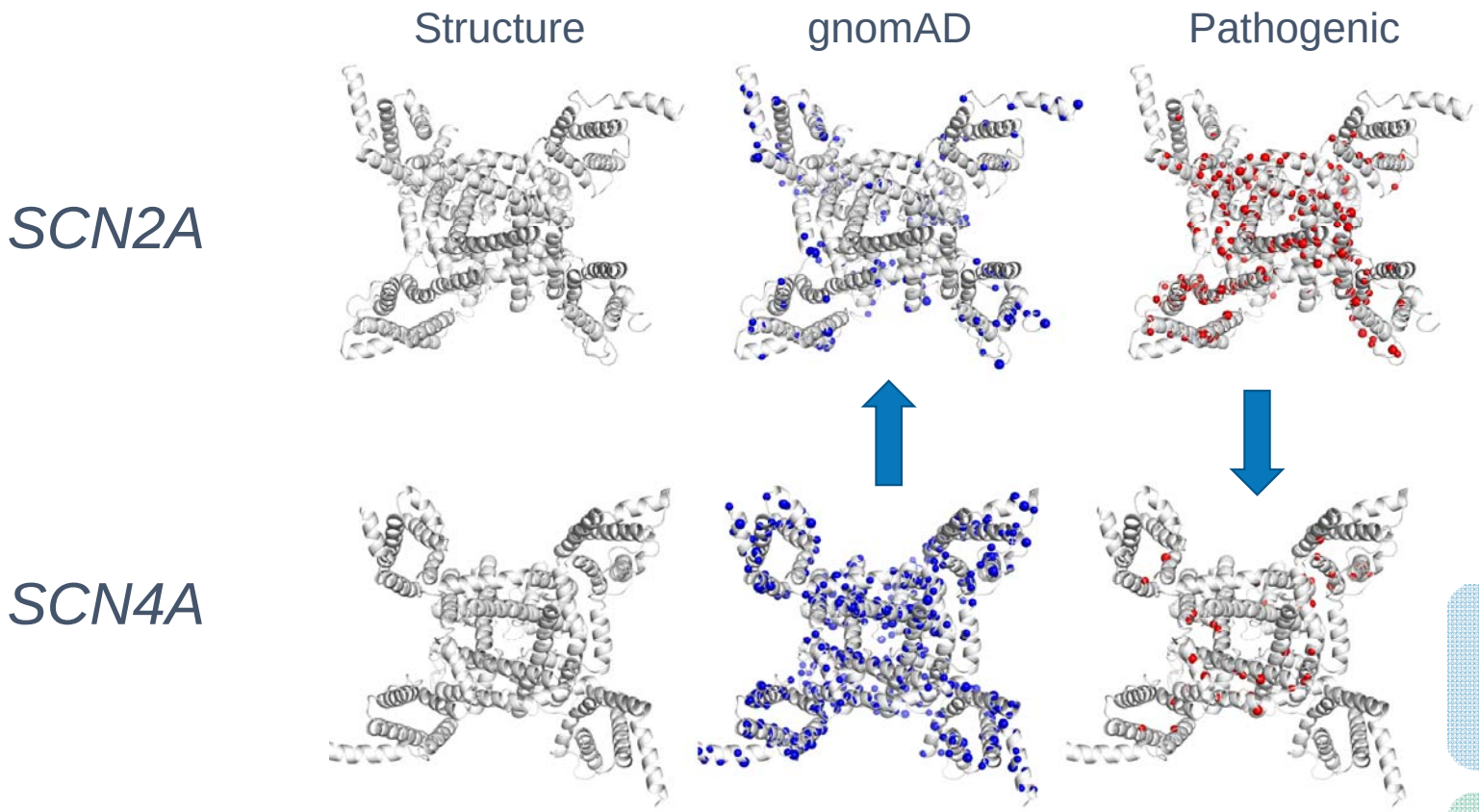
Alignment of Homo Sapiens, Mus Musculus and Bos Taurus *SCN1A* sequence



Alignment of Homo Sapiens *SCN1A-SCN11A* sequences



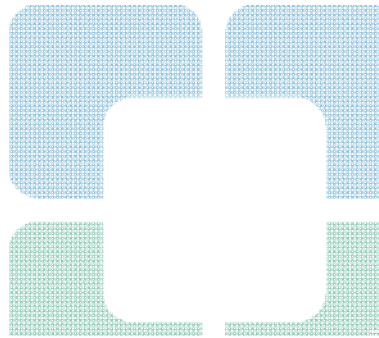
Paralogs have higher structural similarity



Using gene similarity within families to facilitate variant interpretation

Gene family alignment

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Paralog 1	N	L	P	F	V	I	T	L	D	-	G	-	N	L	P	K	N	-	G	V
Paralog 2	D	G	P	F	V	I	T	L	T	L	G	-	D	Q	A	-	N	L	G	V
Paralog 3	G	S	-	F	V	I	T	L	T	L	G	-	D	Q	A	-	N	L	G	V
Paralog 4	I	D	-	F	V	I	T	S	N	L	G	D	N	L	P	K	N	L	G	V
Paralog 5	D	Q	A	F	V	I	T	C	N	G	G	D	D	G	P	K	N	G	G	V
Paralog 6	N	L	P	F	V	I	T	L	D	L	G	D	G	S	-	K	N	L	G	V
Paralog 7	D	G	P	F	V	V	T	L	T	L	G	D	D	G	P	K	N	L	G	V
Paralog 8	G	S	-	F	V	I	T	L	T	L	G	-	G	S	-	K	N	L	G	V
Paralog 9	I	D	-	F	V	V	T	S	N	L	G	-	I	D	-	Q	N	L	G	V
Paralog 10	D	Q	A	F	V	I	T	C	N	L	G	-	I	D	-	Q	N	L	G	V



Using gene similarity within families to facilitate variant interpretation

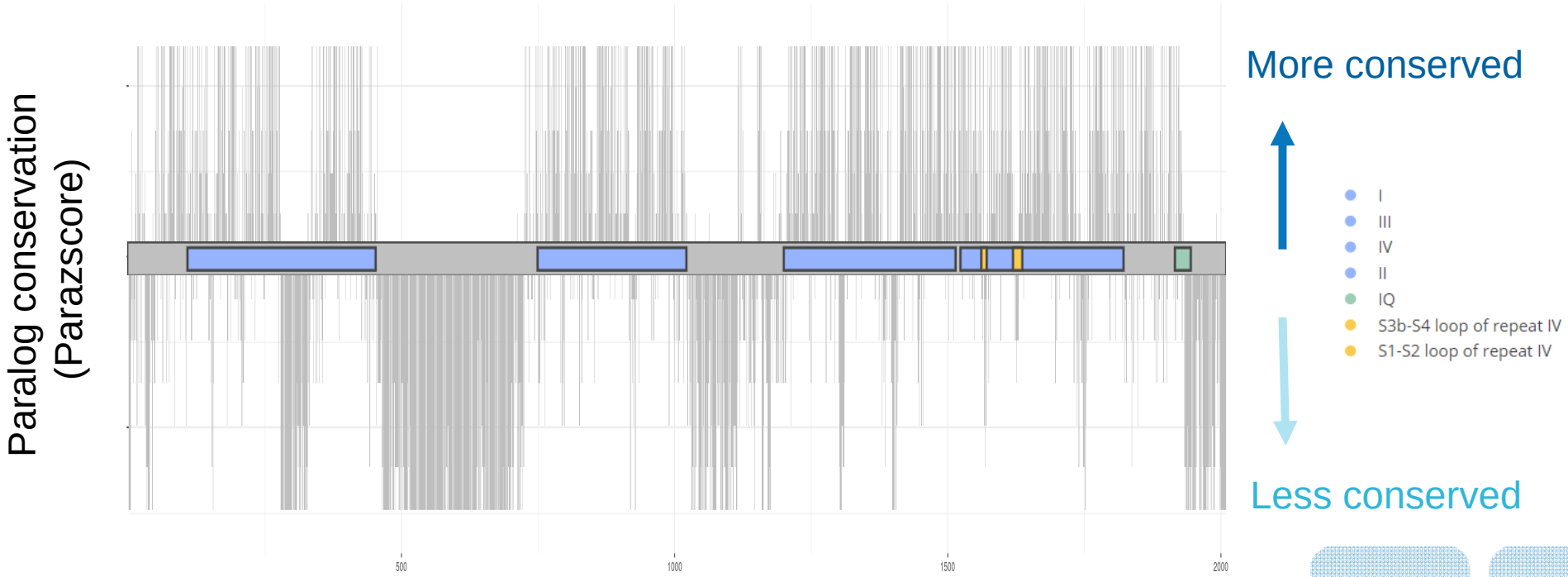
Gene family alignment

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Paralog 1	N	L	P	F	V	I	T	L	D	-	G	-	N	L	P	K	N	-	G	V
Paralog 2	D	G	P	F	V	I	T	L	T	L	G	-	D	Q	A	-	N	L	G	V
Paralog 3	G	S	-	F	V	I	T	L	T	L	G	-	D	Q	A	-	N	L	G	V
Paralog 4	I	D	-	F	V	I	T	S	N	L	G	D	N	L	P	K	N	L	G	V
Paralog 5	D	Q	A	F	V	I	T	C	N	G	G	D	D	G	P	K	N	G	G	V
Paralog 6	N	L	P	F	V	I	T	L	D	L	G	D	G	S	-	K	N	L	G	V
Paralog 7	D	G	P	F	V	V	T	L	T	L	G	D	D	G	P	K	N	L	G	V
Paralog 8	G	S	-	F	V	I	T	L	T	L	G	-	G	S	-	K	N	L	G	V
Paralog 9	I	D	-	F	V	V	T	S	N	L	G	-	I	D	-	Q	N	L	G	V
Paralog 10	D	Q	A	F	V	I	T	C	N	L	G	-	I	D	-	Q	N	L	G	V

Conserved regions are more likely to hold key functional features

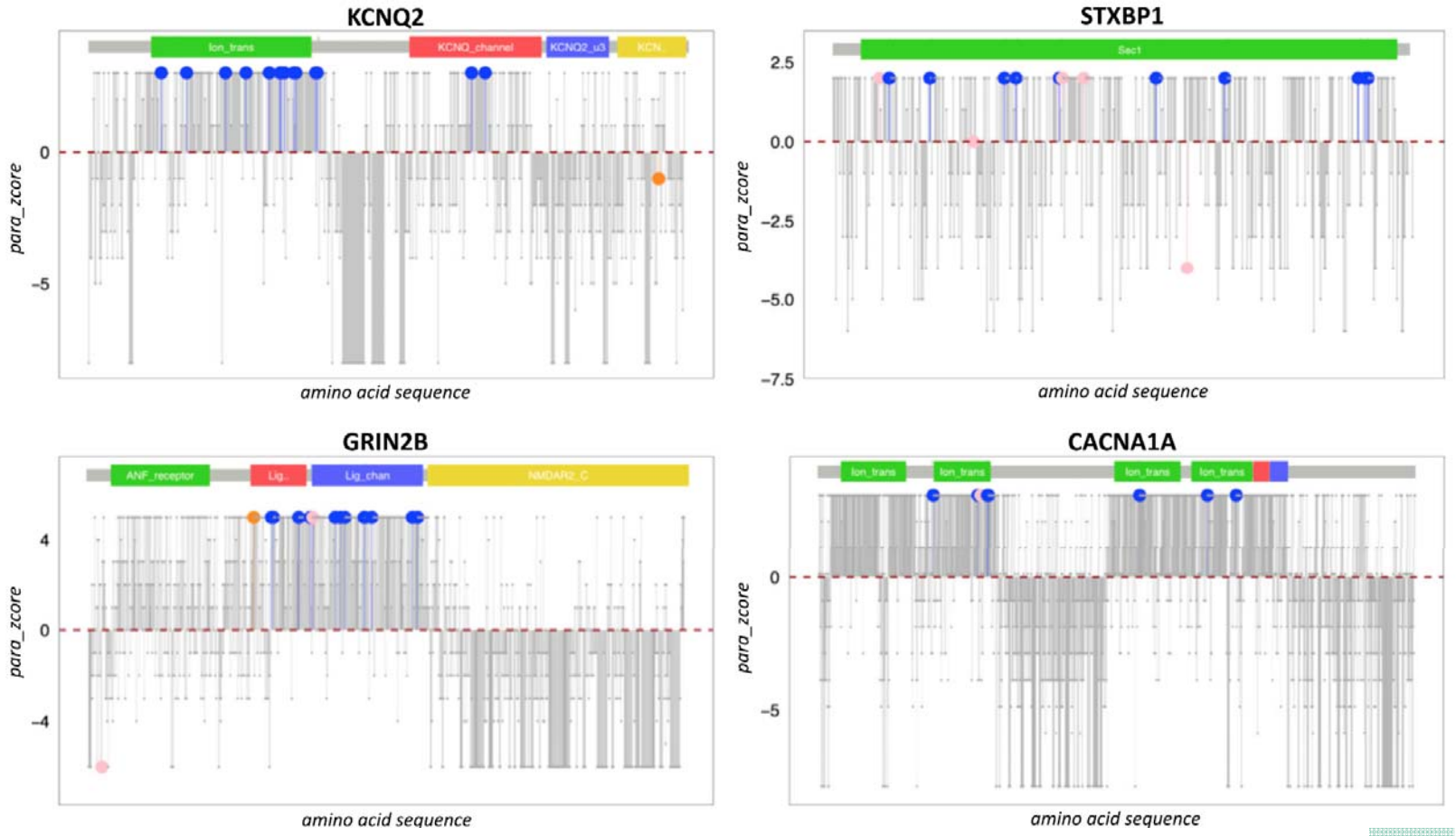
Non conserved regions are less biologically relevant

We developed score for every amino acid of gene-family members



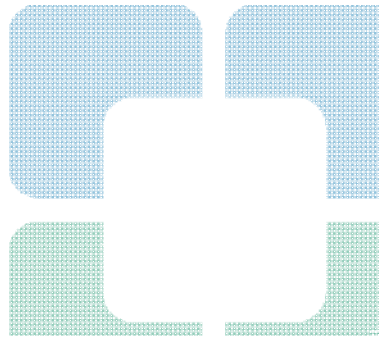
Paralog conservation scores are available for ~9,000 and ~2,800 gene families.

Paralog conservation vs patient variants



Outline

- Gene similarity and paralog conservation
- Leveraging gene similarity for variant interpretation:
  - Paralog z-score
  - Identification of Pathogenic Enriched Regions (PERs)
  - User-friendly web application
- Summary



Where would patient and gnomAD variants fall?

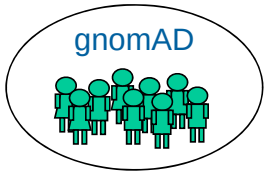
Gene family alignment

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Paralog 1	N	L	P	F	V	I	T	L	D	-	G	-	N	L	P	K	N	-	G	V
Paralog 2	D	G	P	F	V	I	T	L	T	L	G	-	D	Q	A	-	N	L	G	V
Paralog 3	G	S	-	F	V	I	T	L	T	L	G	-	D	Q	A	-	N	L	G	V
Paralog 4	I	D	-	F	V	I	T	S	N	L	G	D	N	L	P	K	N	L	G	V
Paralog 5	D	Q	A	F	V	I	T	C	N	G	G	D	D	G	P	K	N	G	G	V
Paralog 6	N	L	P	F	V	I	T	L	D	L	G	D	G	S	-	K	N	L	G	V
Paralog 7	D	G	P	F	V	V	T	L	T	L	G	D	D	G	P	K	N	L	G	V
Paralog 8	G	S	-	F	V	I	T	L	T	L	G	-	G	S	-	K	N	L	G	V
Paralog 9	I	D	-	F	V	V	T	S	N	L	G	-	I	D	-	Q	N	L	G	V
Paralog 10	D	Q	A	F	V	I	T	C	N	L	G	-	I	D	-	Q	N	L	G	V

Non conserved regions are less biologically relevant

Conserved regions are more likely to hold key functional features

Hypothesis



Gene family alignment

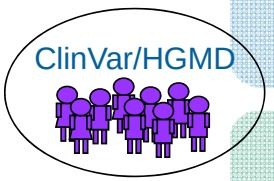
Aligment	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Paralog 1	N	L	P	F	V	I	T	L	D	-	G	-	N	L	P	K	N	-	G	V
Paralog 2	D	G	P	F	V	I	T	L	T	L	G	-	D	Q	A	-	N	L	G	V
Paralog 3	G	S	-	F	V	I	T	L	T	L	G	-	D	Q	A	-	N	L	G	V
Paralog 4	I	D	-	F	V	I	T	S	N	L	G	D	N	L	P	K	N	L	G	V
Paralog 5	D	Q	A	F	V	I	T	C	N	G	G	D	D	G	P	K	N	G	G	V
Paralog 6	N	L	P	F	V	I	T	L	D	L	G	D	G	S	-	K	N	L	G	V
Paralog 7	D	G	P	F	V	V	T	L	T	L	G	D	D	G	P	K	N	L	G	V
Paralog 8	G	S	-	F	V	I	T	L	T	L	G	-	G	S	-	K	N	L	G	V
Paralog 9	I	D	-	F	V	V	T	S	N	L	G	-	I	D	-	Q	N	L	G	V
Paralog 10	D	Q	A	F	V	I	T	C	N	L	G	-	I	D	-	Q	N	L	G	V

Non conserved regions are less biologically relevant

Conserved regions are more likely to hold key functional features

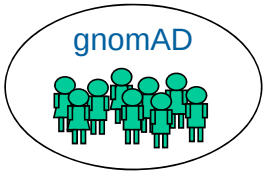
Patients missense variants

Population missense variants



Missense variant mapping: binary annotation

Population missense variants

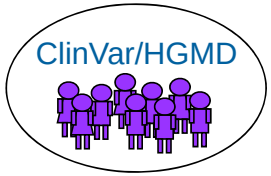


Aligment	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Paralog 1	N	L	P	F	V	I	T	L	D	-	G	-	N	L	P	K	N	-	G	V
Paralog 2	D	G	P	F	V	I	T	L	T	L	G	-	D	Q	A	-	N	L	G	V
Paralog 3	G	S	-	F	V	I	T	L	T	L	G	-	D	Q	A	-	N	L	G	V
Paralog 4	I	D	-	F	V	I	T	S	N	L	G	D	N	L	P	K	N	L	G	V
Paralog 5	D	Q	A	F	V	I	T	C	N	G	G	D	D	G	P	K	N	G	G	V
Paralog 6	N	L	P	F	V	I	T	L	D	L	G	D	G	S	-	K	N	L	G	V
Paralog 7	D	G	P	F	V	V	T	L	T	L	G	D	D	G	P	K	N	L	G	V
Paralog 8	G	S	-	F	V	I	T	L	T	L	G	-	G	S	-	K	N	L	G	V
Paralog 9	I	D	-	F	V	V	T	S	N	L	G	-	I	D	-	Q	N	L	G	V
Paralog 10	D	Q	A	F	V	I	T	C	N	L	G	-	I	D	-	Q	N	L	G	V

Is the aminoacid site mutated at least once?

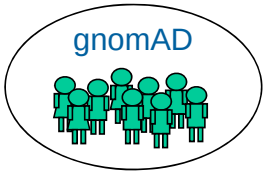
YES=1
NO=0

Patients missense variants



Missense variant mapping: binary annotation

Population missense variants

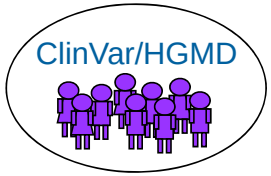


Aligment	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Paralog 1	N	L	P	F	V	I	T	L	D	-	G	-	N	L	P	K	N	-	G	V
Paralog 2	D	G	P	F	V	I	T	L	T	L	G	-	D	Q	A	-	N	L	G	V
Paralog 3	G	S	-	F	V	I	T	L	T	L	G	-	D	Q	A	-	N	L	G	V
Paralog 4	I	D	-	F	V	I	T	S	N	L	G	D	N	L	P	K	N	L	G	V
Paralog 5	D	Q	A	F	V	I	T	C	N	G	G	D	D	G	P	K	N	G	G	V
Paralog 6	N	L	P	F	V	I	T	L	D	L	G	D	G	S	-	K	N	L	G	V
Paralog 7	D	G	P	F	V	V	T	L	T	L	G	D	D	G	P	K	N	L	G	V
Paralog 8	G	S	-	F	V	I	T	L	T	L	G	-	G	S	-	K	N	L	G	V
Paralog 9	I	D	-	F	V	V	T	S	N	L	G	-	I	D	-	Q	N	L	G	V
Paralog 10	D	Q	A	F	V	I	T	C	N	L	G	-	I	D	-	Q	N	L	G	V

Population burden

	4.3	4.2	4	3.8	3.1	2.4	1.7	1.6	1.7	2.2	3.1	3.8	4	4.2	4.3	4.3		
Paralog 1	1	1	1	1	1	1	0	1	1	0	0	1	1	1	1	1	1	1
Paralog 2	1	1	1	1	1	1	1	0	0	0	0	1	1	0	1	1	1	1	0	1
Paralog 3	1	1	1	1	0	1	1	1	0	0	0	1	0	1	0	1	1	1	1	1
Paralog 4	0	1	1	1	1	1	1	0	0	0	0	0	1	1	1	0	1	1	1	1
Paralog 5	1	1	1	1	1	1	0	1	0	0	0	0	1	0	1	1	1	1	1	1
Paralog 6	1	0	1	0	1	0	1	1	0	0	0	1	1	1	1	1	0	1	1	1
Paralog 7	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	0	1	1	1
Paralog 8	1	1	0	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	1	1
Paralog 9	1	1	1	1	0	1	1	1	0	0	0	0	1	0	1	1	1	1	0	1
Paralog 10	1	1	1	1	0	1	0	1	1	0	1	1	0	1	1	1	1	1	1	0

Patients missense variants



Patient burden

	0.1	0.1	0.1	0.3	0.5	1	1.4	1.6	1.4	1.1	0.7	0.3	0.1	0.1	0.2	0.1		
Paralog 1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0
Paralog 2	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
Paralog 3	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0
Paralog 4	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Paralog 5	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Paralog 6	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
Paralog 7	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
Paralog 8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Paralog 9	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
Paralog 10	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0

Missense burden analysis

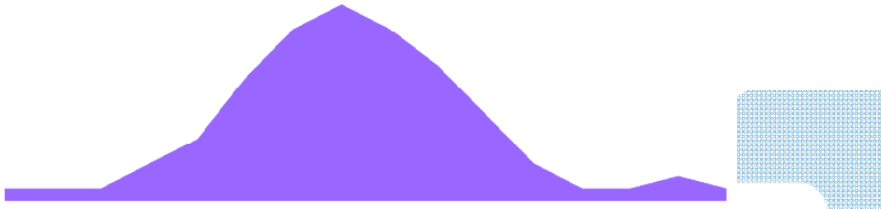
Population burden

	4.3	4.2	4	3.8	3.1	2.4	1.7	1.6	1.7	2.2	3.1	3.8	4	4.2	4.3	4.3	
Paralog 1	1	1	1	1	1	0	1	1	0	0	1	1	1	1	1	1	1	1
Paralog 2	1	1	1	1	1	1	0	0	0	0	1	1	0	1	1	1	1	0	1
Paralog 3	1	1	1	1	0	1	1	1	0	0	0	1	0	1	0	1	1	1	1	1
Paralog 4	0	1	1	1	1	1	1	0	0	0	0	0	1	1	1	0	1	1	1	1
Paralog 5	1	1	1	1	1	1	0	1	0	0	0	0	1	0	1	1	1	1	1	1
Paralog 6	1	0	1	0	1	0	1	1	0	0	0	1	1	1	1	1	0	1	1	1
Paralog 7	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	0	1	1	1
Paralog 8	1	1	0	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	1	1
Paralog 9	1	1	1	1	0	1	1	1	0	0	0	0	1	0	1	1	1	1	0	1
Paralog 10	1	1	1	1	0	1	0	1	1	0	1	1	0	1	1	1	1	1	1	0

Patient burden

	0.1	0.1	0.1	0.3	0.5	1	1.4	1.6	1.4	1.1	0.7	0.3	0.1	0.1	0.2	0.1		
Paralog 1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0
Paralog 2	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
Paralog 3	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0
Paralog 4	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Paralog 5	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Paralog 6	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
Paralog 7	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
Paralog 8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Paralog 9	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
Paralog 10	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0

Burden analysis



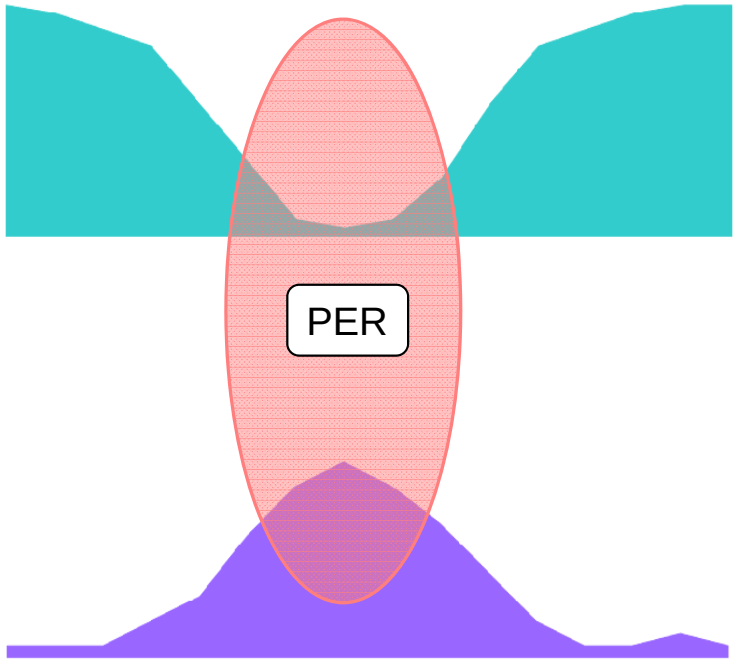
Missense burden analysis: Identification of PERs

Population burden

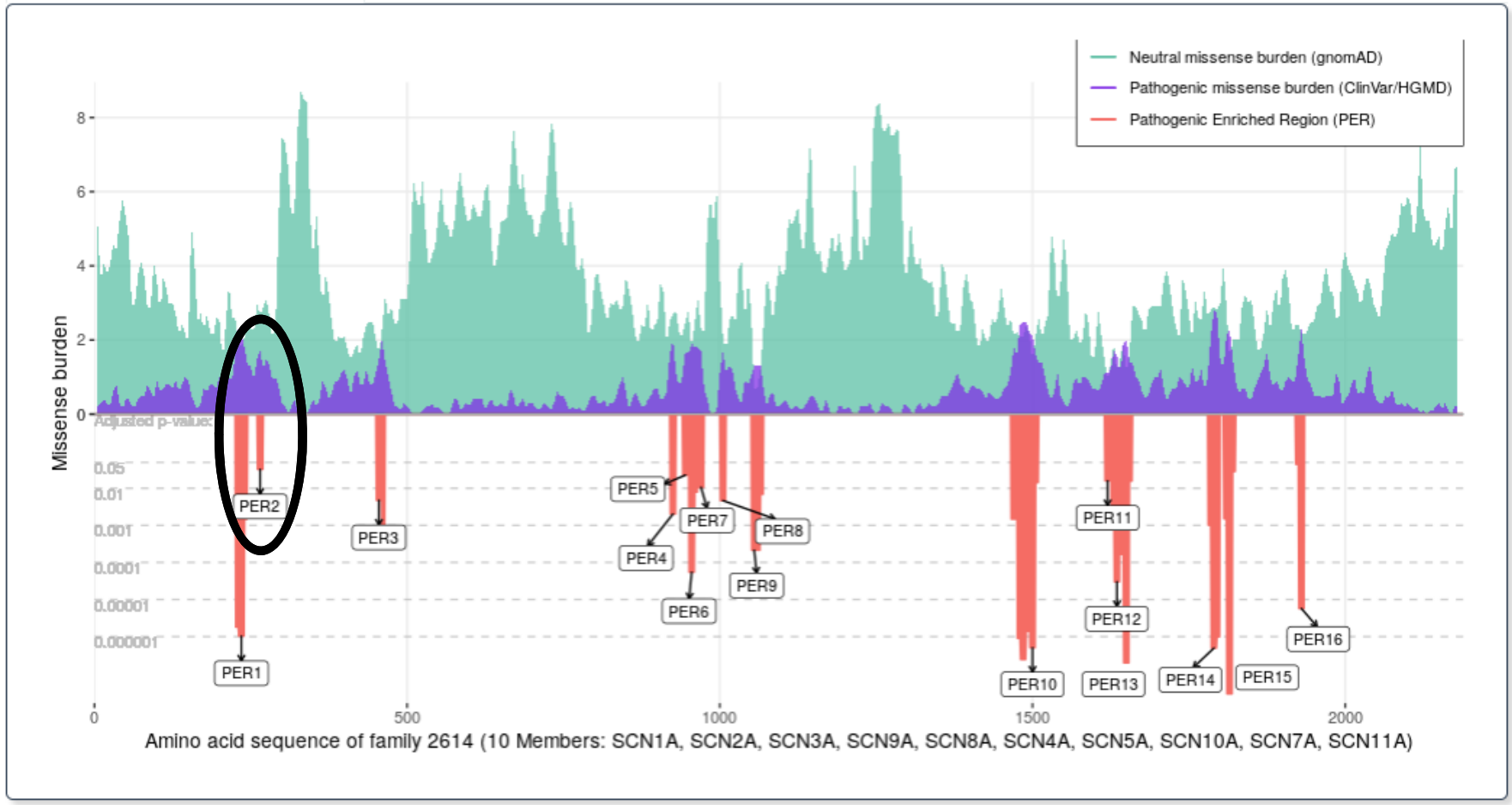
	4.3	4.2	4	3.8	3.1	2.4	1.7	1.6	1.7	2.2	3.1	3.8	4	4.2	4.3	4.3	
Paralog 1	1	1	1	1	1	0	1	1	0	0	1	1	1	1	1	1	1	1
Paralog 2	1	1	1	1	1	1	0	0	0	0	1	1	0	1	1	1	1	0	1
Paralog 3	1	1	1	1	0	1	1	1	0	0	0	1	0	1	0	1	1	1	1	1
Paralog 4	0	1	1	1	1	1	1	0	0	0	0	0	1	1	1	0	1	1	1	1
Paralog 5	1	1	1	1	1	1	0	1	0	0	0	0	1	0	1	1	1	1	1	1
Paralog 6	1	0	1	0	1	0	1	1	0	0	0	1	1	1	1	1	0	1	1	1
Paralog 7	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	0	1	1	1
Paralog 8	1	1	0	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	1	1
Paralog 9	1	1	1	1	0	1	1	1	0	0	0	0	1	0	1	1	1	1	0	1
Paralog 10	1	1	1	1	0	1	0	1	1	0	1	1	0	1	1	1	1	1	1	0

Patient burden

	0.1	0.1	0.1	0.3	0.5	1	1.4	1.6	1.4	1.1	0.7	0.3	0.1	0.1	0.2	0.1		
Paralog 1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0
Paralog 2	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
Paralog 3	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0
Paralog 4	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Paralog 5	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Paralog 6	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
Paralog 7	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
Paralog 8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Paralog 9	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
Paralog 10	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0



Example: The voltage gate sodium Family



Example: The voltage gate sodium Family (Zoom to PER 2)

Align	SCN1A	SCN2A	SCN3A	SCN9A	SCN8A	SCN4A	SCN5A	SCN10A	SCN7A	SCN11A	PER	Gene:Disease
455	I_415	I_417	V_416	I_394	V_403	I_439	V_405	V_389	A_386	I_392	Neutral	N/A
456	N_416	N_418	N_417	N_395	N_404	N_440	N_406	N_390	S_387	N_393	PER2	**SCN9A**:Erythermalgia primary,Erythermalgia primary,Primary erythromelalgia,not provided**;SCN4A**:Myotonia,Hyperkalemic Periodic Paralysis Type 1;**SCN5A**:Long QT syndrome,Long QT syndrome,Brugada syndrome,Congenital long QT syndrome|not provided
457	L_417	L_419	L_418	L_396	L_405	L_441	L_407	L_391	L_388	L_394	PER2	N/A
458	I_418	I_420	I_419	I_397	I_406	I_442	I_408	I_392	F_389	T_395	PER2	N/A
459	L_419	L_421	L_420	L_398	L_407	L_443	L_409	L_393	L_390	L_396	PER2	**SCN8A**:Intellectual disability and epilepsy,not provided,Early infantile epileptic encephalopathy 13;**SCN5A**:Long QT syndrome,Long QT syndrome;**SCN11A**:Neuropathy, hereditary sensory and autonomic, type VII|not provided

- Gene-family sequence alignment shows for all members the sequence and mutations
- Pathogenic enriched regions (PERs) significantly accumulates patients' mutations
- PERs sites can help the interpretation of variants in related genes

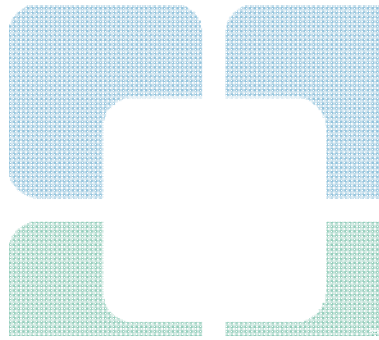
Example: The voltage gate sodium Family

Align	SCN1A	SCN2A	SCN3A	SCN9A	SCN8A	SCN4A	SCN5A	SCN10A	SCN7A	SCN11A	PER	Gene:Disease
455	I_415	I_417	V_416	I_394	V_403	I_439	V_405	V_389	A_386	I_392	Neutral	N/A
456	N_416	N_418	N_417	N_395	N_404	N_440	N_406	N_390	S_387	N_393	PER2	**SCN9A**:Erythermalgia primary,Erythermalgia primary,Primary erythromelalgia,not provided**;SCN4A**:Myotonia,Hyperkalemic Periodic Paralysis Type 1;**SCN5A**:Long QT syndrome,Long QT syndrome,Brugada syndrome,Congenital long QT syndrome|not provided
457	L_417	L_419	L_418	L_396	L_405	L_441	L_407	L_391	L_388	L_394	PER2	N/A
458	I_418	I_420	I_419	I_397	I_406	I_442	I_408	I_392	F_389	T_395	PER2	N/A
459	L_419	L_421	L_420	L_398	L_407	L_443	L_409	L_393	L_390	L_396	PER2	**SCN8A**:Intellectual disability and epilepsy,not provided,Early infantile epileptic encephalopathy 13;**SCN5A**:Long QT syndrome,Long QT syndrome;**SCN11A**:Neuropathy, hereditary sensory and autonomic, type VII|not provided

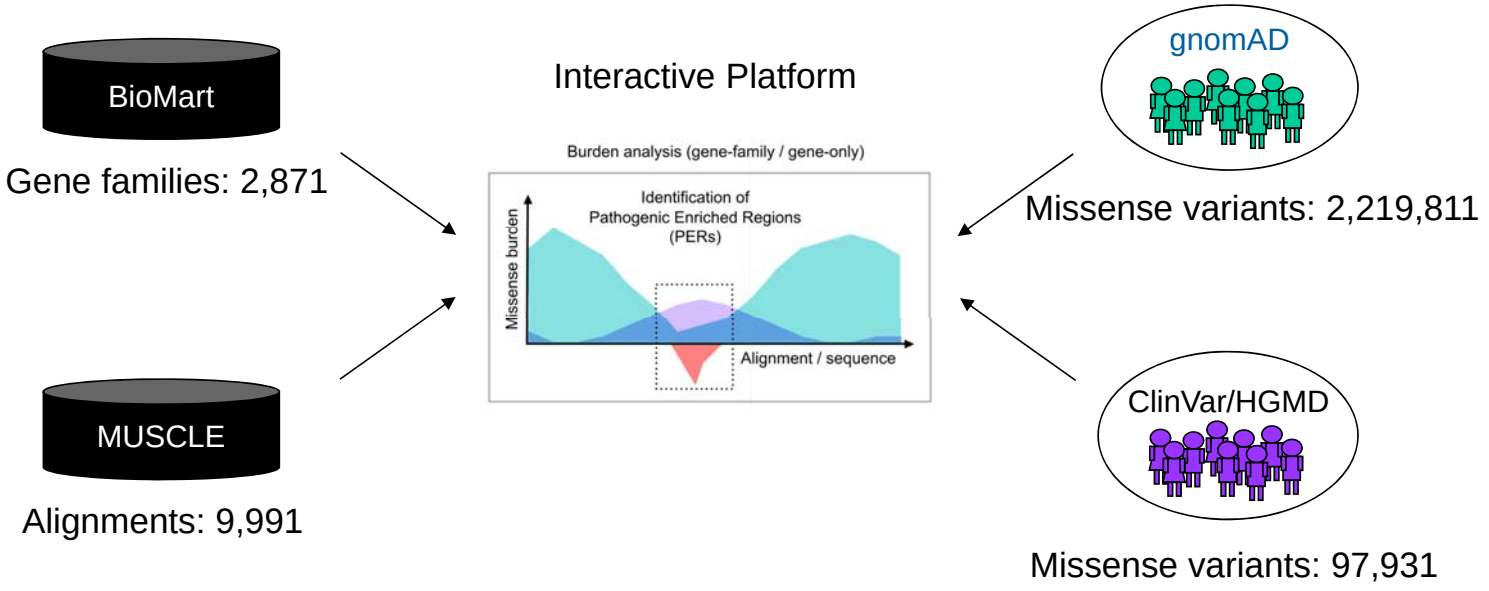
- Gene-family sequence alignment shows for all members the sequence and mutations
- Pathogenic enriched regions (PERs) significantly accumulates patients mutations
- PERs sites can help the interpretation of variants in related genes
- PERs capture sites never observed in patients surrounded by pathogenic variants

Outline

- Gene similarity and paralog conservation
- Leveraging gene similarity for variant interpretation:
  - Paralog z-score
  - Identification of Pathogenic Enriched Regions (PERs)
  - User-friendly web application
- Summary

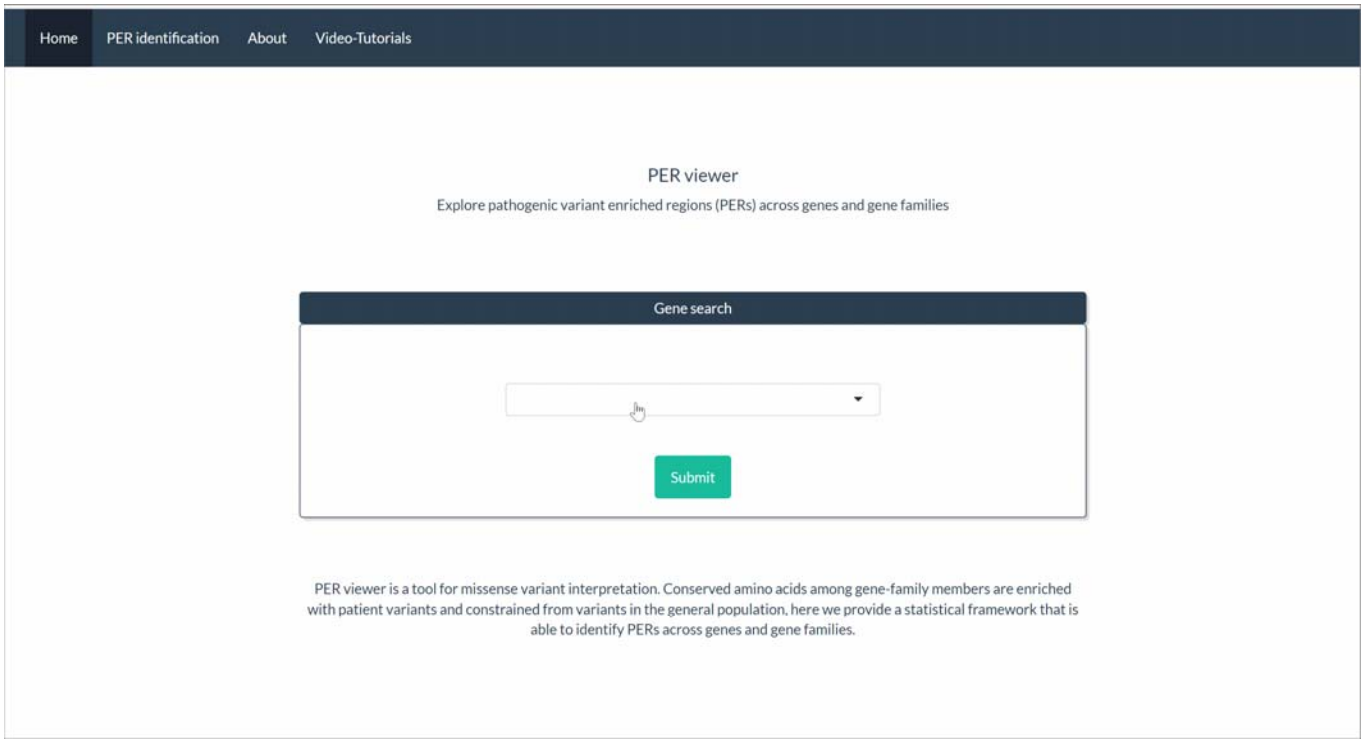


The PER viewer: http://per.broadinstitute.org



We deployed our all our results into a user-friendly online tool

The PER viewer: http://per.broadinstitute.org



- Input any Gene
- If the gene belongs to a family, results will be displayed for the family
- Main Plot shows Burden analysis results and PER identified.

The PER viewer: http://per.broadinstitute.org

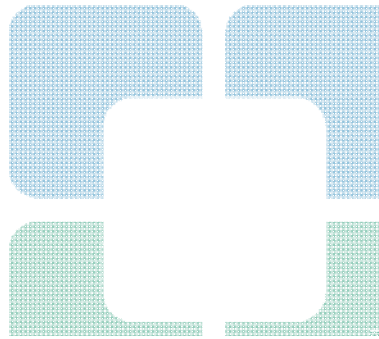
- The user can swicht between family wise and gene wise analysis
- Paralog conservation is also provide alongside PERs
- The user can Zoom in regions of interest

The PER viewer: http://per.broadinstitute.org

- Below main plot, table mode shows family and mutation aligments
- The user can explore the sequence and mutations observed
- Sites without known patient variants can be observed surrounded by pathogenic mutations.

Outline

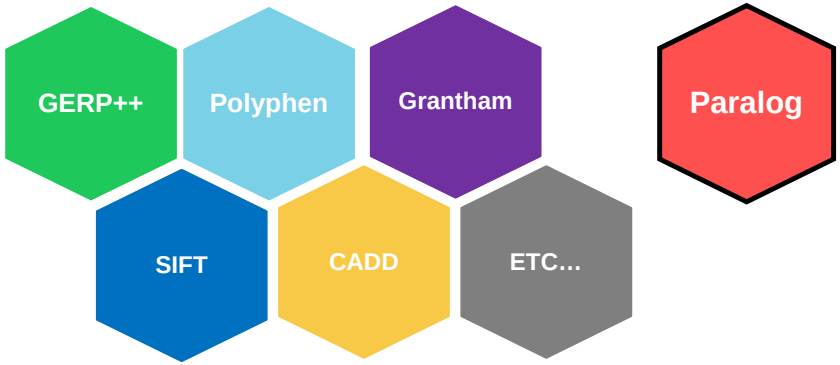
- Gene similarity and paralog conservation
- Leveraging gene similarity for variant interpretation:
  - Paralog z-score
  - Identification of Pathogenic Enriched Regions (PERs)
  - User-friendly web application
- Summary



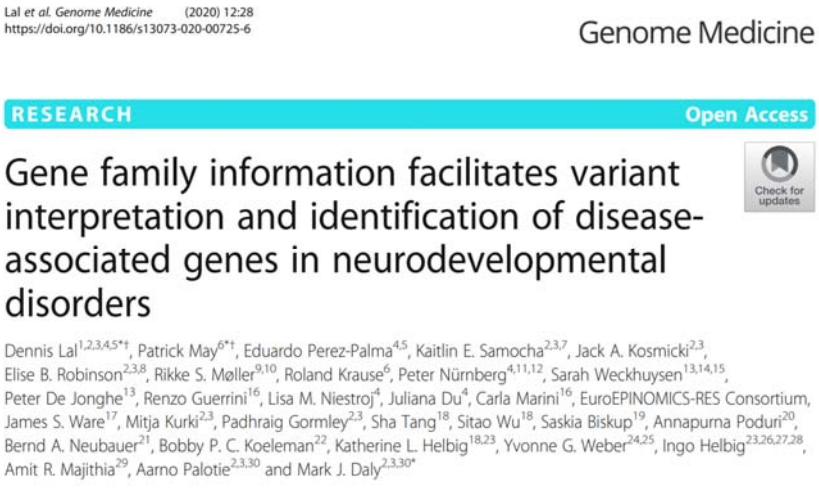
Integration of gene similarity to the ACMG guidelines for variant interpretation

Moderate ▼ PM1: Located in a mutational hot spot and/or critical and well-established functional domain (e.g., active site of an enzyme) without benign variation

Supporting ▼ PP3: Multiple lines of computational evidence support a deleterious effect



More information



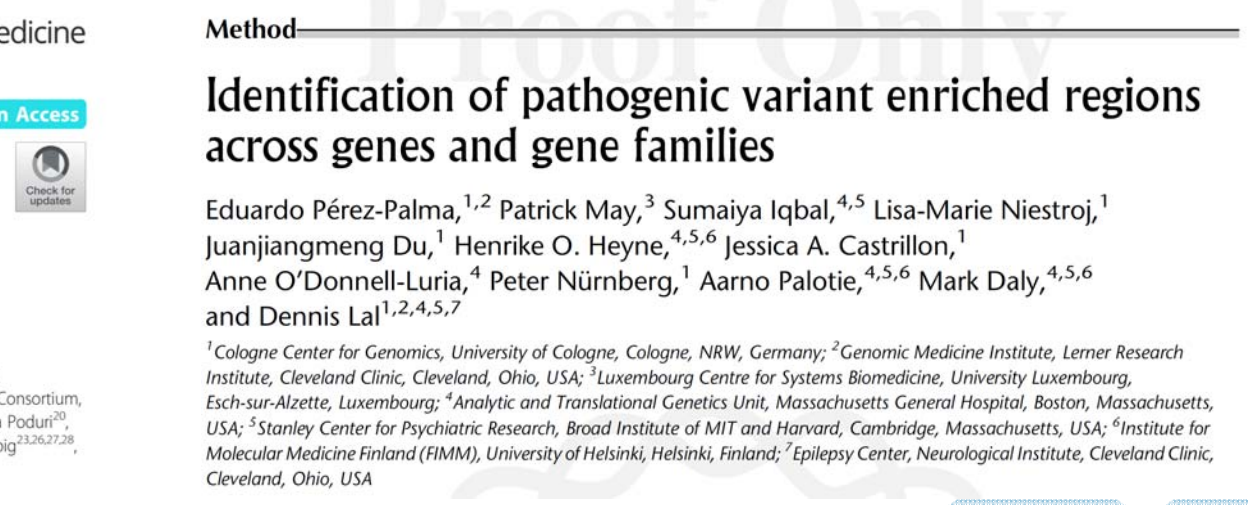
Lal et al. Genome Medicine (2020) 12:28
https://doi.org/10.1186/s13073-020-00725-6

Genome Medicine

RESEARCH Open Access

Gene family information facilitates variant interpretation and identification of disease-associated genes in neurodevelopmental disorders

Dennis Lal[1,2,3,4,5*†], Patrick May[6*†], Eduardo Perez-Palma[4,5], Kaitlin E. Samocha[2,3,7], Jack A. Kosmicki[2,3], Elise B. Robinson[2,3,8], Rikke S. Møller[9,10], Roland Krause[6], Peter Nürnberg[4,11,12], Sarah Weckhuysen[13,14,15], Peter De Jonghe[13], Renzo Guerrini[16], Lisa M. Niestroj[4], Juliana Du[4], Carla Marini[16], EuroEPINOMICS-RES Consortium, James S. Ware[17], Mitja Kurki[2,3], Padhraig Gormley[2,3], Sha Tang[18], Sitao Wu[18], Saskia Biskup[19], Annapurna Poduri[20], Bernd A. Neubauer[21], Bobby P. C. Koeleman[22], Katherine L. Helbig[18,23], Yvonne G. Weber[24,25], Ingo Helbig[23,26,27,28], Amit R. Majithia[29], Aarno Palotie[2,3,30] and Mark J. Daly[2,3,30*]



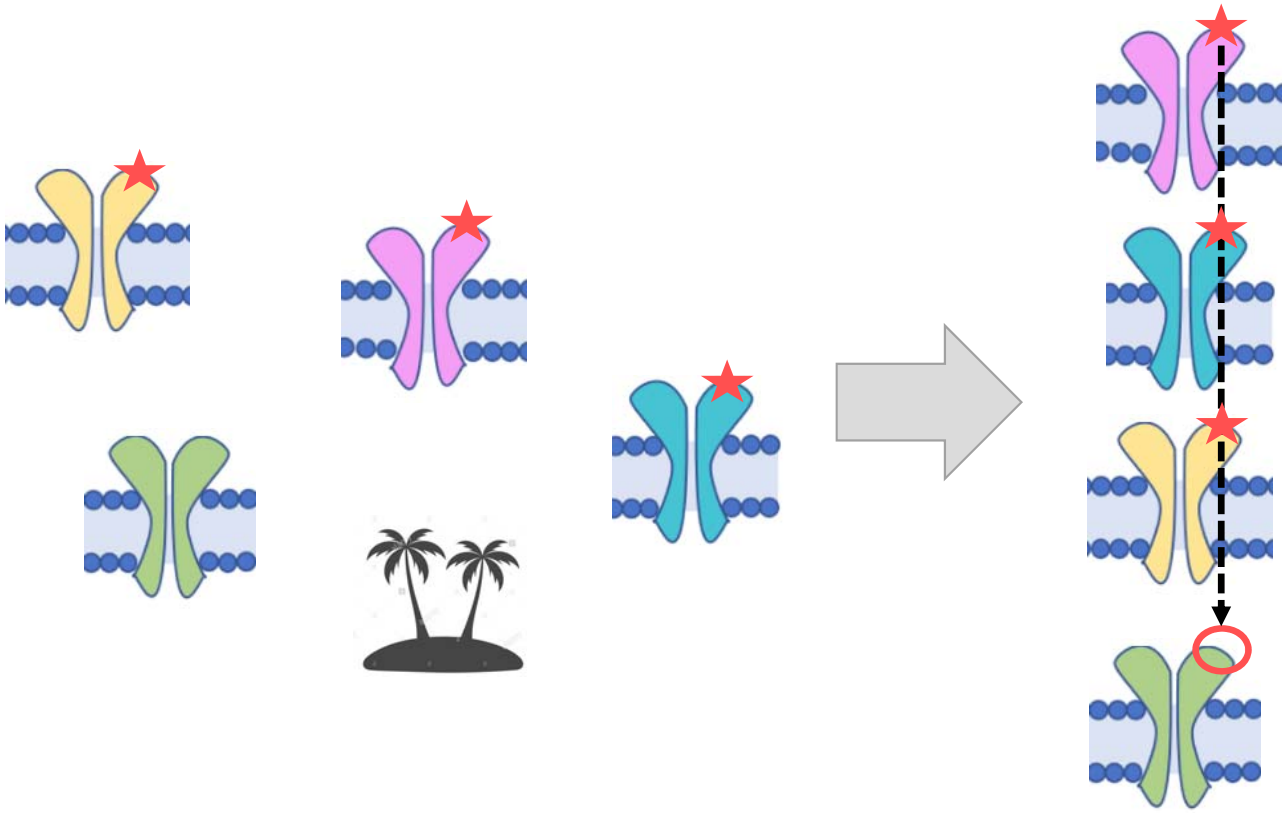
Method

Identification of pathogenic variant enriched regions across genes and gene families

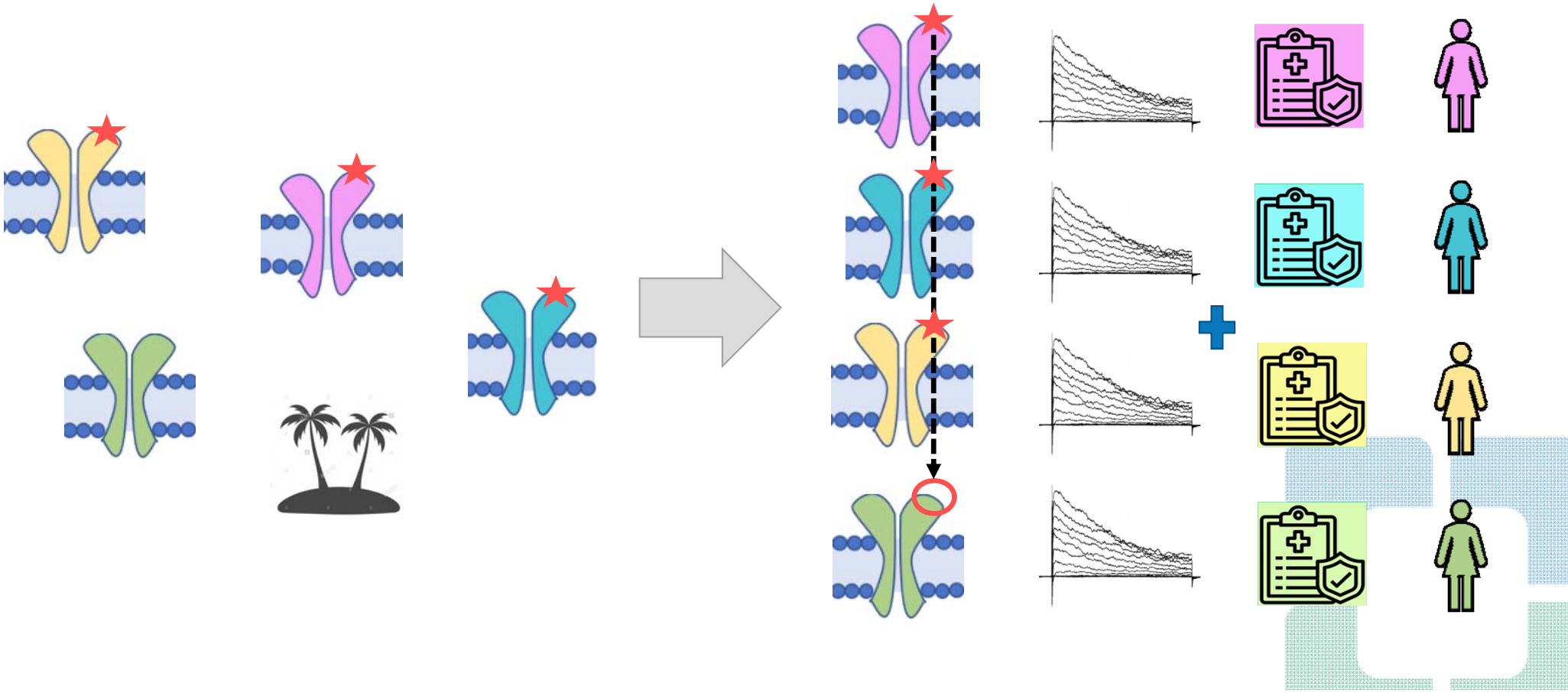
Eduardo Pérez-Palma,[1,2] Patrick May,[3] Sumaiya Iqbal,[4,5] Lisa-Marie Niestroj,[1] Juanjiangmeng Du,[1] Henrike O. Heyne,[4,5,6] Jessica A. Castrillon,[1] Anne O'Donnell-Luria,[4] Peter Nürnberg,[1] Aarno Palotie,[4,5,6] Mark Daly,[4,5,6] and Dennis Lal[1,2,4,5,7]

[1]Cologne Center for Genomics, University of Cologne, Cologne, NRW, Germany; [2]Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio, USA; [3]Luxembourg Centre for Systems Biomedicine, University Luxembourg, Esch-sur-Alzette, Luxembourg; [4]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA; [5]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; [6]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; [7]Epilepsy Center, Neurological Institute, Cleveland Clinic, Cleveland, Ohio, USA

The collective study of similar genes can help us increase our understanding of individually rare DEEs



The collective study of related genes may help us increase our understanding of individually rare DEEs



Acknowledgements

Cleveland Clinic, OH, US

Arthur Stefanski
Charis Eng
Costin Leu
Imad Najm
Javier Lopez
Ignacio Mata
Elif Irem Sarihan
Dennis Lal
Victoria Smuk

Stanley Center & Broad Institute, US

Sumaiya Iqbal
Arthur J. Campbell
Mark Daly

University of Cologne, Germany

Lisa-Marie Niestroj
Marie Gramm
Tobias Bruenger
Peter Nürnberg

International Collaborators

Patrick May, University of Luxembourg, Luxembourg
Rikke S Møller, Danish Epilepsy Center, Denmark
Andreas Brunklaus, Royal Hospital for Children, Glasgow, UK

eperezpalma@gmail.com

@EdoPerezP