BRAIN
ORIGINAL ARTICLE

Conserved patterns across ion channels correlate with variant pathogenicity and clinical phenotypes

Tobias Brünger,[1] Eduardo Pérez-Palma,[2] Ludovica Montanucci,[3] Michael Nothnagel,[1,4] Rikke S. Møller,[5] Stephanie Schorge,[6] Sameer Zuberi,[7,8] Joseph Symonds,[7,8] Johannes R. Lemke,[9,10] Andreas Brunklaus,[7,8] Stephen F. Traynelis,[11] Patrick May[12] and Dennis Lal[1,3,13,14]

Clinically identified genetic variants in ion channels can be benign or cause disease by increasing or decreasing the protein function. As a consequence, therapeutic decision-making is challenging without molecular testing of each variant. Our biophysical knowledge of ion-channel structures and function is just emerging, and it is currently not well understood which amino acid residues cause disease when mutated.

We sought to systematically identify biological properties associated with variant pathogenicity across all major voltage and ligand-gated ion-channel families. We collected and curated 3049 pathogenic variants from hundreds of neurodevelopmental and other disorders and 12 546 population variants for 30 ion channel or channel subunits for which a high-quality protein structure was available. Using a wide range of bioinformatics approaches, we computed 163 structural features and tested them for pathogenic variant enrichment. We developed a novel 3D spatial distance scoring approach that enables comparisons of pathogenic and population variant distribution across protein structures. We discovered and independently replicated that several pore residue properties and proximity to the pore axis were most significantly enriched for pathogenic variants compared to population variants. Using our 3D scoring approach, we showed that the strongest pathogenic variant enrichment was observed for pore-lining residues and alpha-helix residues within 5Å distance from the pore axis centre and not involved in gating. Within the subset of residues located at the pore, the hydrophobicity of the pore was the feature most strongly associated with variant pathogenicity. We also found an association between the identified properties and both clinical phenotypes and functional in vitro assays for voltage-gated sodium channels (SCN1A, SCN2A, SCN8A) and *N*-methyl-D-aspartate receptor (GRIN1, GRIN2A, GRIN2B) encoding genes. In an independent expert-curated dataset of 1422 neurodevelopmental disorder pathogenic patient variants and 679 electrophysiological experiments, we show that pore axis distance is associated with seizure age of onset and cognitive performance as well as differential gain versus loss-of-channel function.

In summary, we identified biological properties associated with ion-channel malfunction and show that these are correlated with *in vitro* functional readouts and clinical phenotypes in patients with neurodevelopmental disorders. Our results suggest that clinical decision support algorithms that predict variant pathogenicity and function are feasible in the future.

1 Cologne Center for Genomics, University of Cologne, 50931 Cologne, Germany
2 Centro de Genética y Genómica, Facultad de Medicina Clínica Alemana, Universidad de Desarrollo, Santiago 7590943, Chile
3 Lerner Research Institute Cleveland Clinic, Genomic Medicine Institute, Cleveland, OH 44195, USA
4 University Hospital Cologne, 50937 Cologne, Germany
5 Department of Epilepsy Genetics and Personalized Treatment, the Danish Epilepsy Center, DK 4293 Dianalund, Denmark

Received February 22, 2022. Revised July 29, 2022. Accepted August 09, 2022. Advance access publication August 29, 2022

6 Department of Neuroscience, Physiology and Pharmacology, UCL, London WC1E 6BT, UK
7 The Paediatric Neurosciences Research Group, Royal Hospital for Children, Glasgow, UK
8 Institute of Health and Wellbeing, University of Glasgow, UK
9 Institute of Human Genetics, University of Leipzig Medical Center, 04103 Leipzig, Germany
10 Center for Rare Diseases, University of Leipzig Medical Center, 04103 Leipzig, Germany
11 Department of Pharmacology, Emory University School of Medicine, Rollins Research Center, Atlanta, GA 30322-3090, USA
12 Luxembourg Centre for Systems Biomedicine, University of Luxembourg, L-4362 Esch-sur-Alzette, Luxembourg
13 Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
14 Epilepsy Center, Neurological Institute, Cleveland Clinic, Cleveland, OH 44195, USA

Correspondence to: Dennis Lal
9500 Euclid Ave., NE-50, Cleveland, OH 44195, USA
E-mail: lald@ccf.org

Keywords: genetics; epilepsy; neurodevelopmental disorder; ion channel; bioinformatics

Introduction

Ion channels are membrane proteins that act as gated pathways for the movement of ions across cell membranes. They play essential roles in the physiology of all cells. At least 130 'channelopathies' have been associated with genetic variants causing ion-channel dysfunctions across the human body, including diseases in the nervous system (e.g. epilepsy, familial hemiplegic migraine), cardiovascular system (e.g. long QT syndrome, Brugada syndrome), respiratory system (e.g. cystic fibrosis) and endocrine system (e.g. neonatal diabetes mellitus).[1,2]

Ion channels malfunction due to missense variants that alter the protein sequence can cause a wide spectrum of disorders.[1,3] Missense variants even within the same gene can produce different molecular effects in ion channels: increased or decreased ion permeation; dysregulation of gating elements; changes in opening and closing probabilities; altered kinetics and protein trafficking.[4–9] Notably, not all variants that cause a molecular change from the wild-type cause necessarily a disease.[10,11] The location of missense variants on protein structure has been associated with variant pathogenicity.[12–14] In addition, several recent studies showed that among pathogenic variants, the variant position in specific functional units correlates with the patient subphenotype.[15–22] However, the knowledge about microdomain to phenotype associations is sparse and typically only available to experts of specific channels and cannot easily be adapted by a wider audience. Variant interpretation of less studied and recently identified channelopathies represents a major challenge.

In this study, we sought to identify biological properties associated with variant pathogenicity that are shared among all ion channels. These properties could be incorporated into future algorithms for predicting variant pathogenicity and, thus, provide insights into genotype–phenotype correlations in established and newly identified channelopathies regardless of the ion-channel class. To accomplish this goal, we performed a series of association analyses for biological properties with 12 546 population and 3049 pathogenic variants across 30 ion-channel structures. This work led to the development of a bioinformatic structure-based framework that showed that the distance of a residue from the central pore axis and pore hydrophobicity are the most pathogenicity-associated features shared across all ion channels. We applied our framework to two independent data sets comprising >1400 patient variants and ∼700 functionally tested variants, for the NMDA receptor and voltage-gated sodium channels to demonstrate how correlations between the localization of variants and their associated molecular effect and clinical outcome can be captured.

Materials and methods

Collection of ion channels and protein structure selection

We collected all voltage and ligand-gated ion channels from IUPHAR[23] ($n = 172$ genes). Out of these channels, we included those in our study that harbour at least one known pathogenic variant in the ClinVar[24] database and whose encoded protein structures have been determined at <5Å resolution in the protein data bank[25] and form a pore. For each channel, the protein structure with the largest protein sequence coverage was chosen (Supplementary Table 1). All selected ion channels were divided into evolutionarily derived channel families according to HGNC[26] criteria.

Missense variant annotation

Canonical transcripts for proteins encoded by the ion-channel genes in our cohort were accessed from the UniProt[27] database. For these transcripts, variants were collected from multiple databases. In particular, protein-coding missense variants from the general population were retrieved from the genome aggregation Database[28] (https://gnomad.broadinstitute.org/, public release 2.0.2) including all ethnic backgrounds. Variant call format files were downloaded for all available exomes of ion-channel genes to access the variants. These 'population variants' served as control dataset in subsequent enrichment and association analyses. The extraction of missense variants (Filter = 'PASS') was performed with vcftools (v.0.1.12b) using the pre-annotated 'CSQ' field. ClinVar[24] variants were accessed from the ftp site (ftp://ftp.ncbi.nlm.nih.gov/pub/clinvar/, accessed July 2020). Only variants with a clinical consequence (annotated as 'Pathogenic'/'Likely Pathogenic') were extracted. Next, we obtained missense variants from the Human Gene Mutation Database (HGMD[29]) and filtered them for high-confidence calls (hgmd_confidence = 'HIGH' flag) as well as disease-causing states (hgmd_variant_Type = 'DM' flag) (HGMD, July 2020). Variants from ClinVar and HGMD databases

were combined into a pathogenic variant data set containing unique amino acid substitutions. We applied randomized sampling to create separate subsets of the collected variants to obtain discovery and validation cohorts each containing 70 and 30% of the pathogenic and population variants, respectively.

Collection of missense variants with detailed clinical phenotype or known molecular effect

We collected missense variants with known clinical phenotypes for three sodium channels (*SCN1A*, *SCN2A*, *SCN3A*) from the SCN Portal[30] (https://scn-portal.broadinstitute.org, accessed 1 December 2021). Missense variants with known clinical phenotype for three genes that encode ionotropic glutamate receptors (*GRIN1A*, *GRIN2A*, *GRIN2B*) were obtained from the GRIN Portal[31] (https://grin-portal.broadinstitute.org, access 1 December 2021). In addition, we obtained missense variants across all voltage-gated sodium channel encoding genes (*SCN1A*-*SCN11A*) with a reported molecular effect from Brunklaus *et al.*[22] Variants were categorized as either gain-of-function (GoF), loss-of-function (LoF) or mixed-function (mixed) function depending on the electrophysiological readouts. In addition, missense variants with a functional effect in any of the GRIN genes had been obtained (CFERV,[32] http://functionalvariants.emory.edu/, GRIN Portal[31]).

Collection and annotation of features

We annotated features that describe the localization of residues in functional regions of the protein as well as their protein structure (Supplementary Fig. 1). Therefore, we first mined the UniProt[27] database (https://www.uniprot.org/, accessed 2 May 2021) to collect the location of the amino acids inside the protein (intracellular, membrane-spanning or extracellular) and whether a residue is part of a specific unit (voltage sensor, allosteric-, agonist or other binding sites). The secondary structure of amino acid residues was calculated using the Dictionary of Protein Secondary Structure[33] program. Protein–protein interactions were obtained from the PDBsum[34] database (accessed 17 May 2021). We further considered two differently derived pore annotations to capture residues that are predicted to be located at the pore. First, a residue was considered to be part of the protein according to its annotation in the UniProt database. Second, we predicted the pore using the Mole2.0 webserver[35] (https://mole.upol.cz/, accessed 3 April 2021) and obtained the annotation of those residues that were predicted to be pore lining.

Statistical analysis

If not otherwise stated, we used a two-by-two contingency table to calculate the fold enrichment of pathogenic over population variants across different sets of residues, using the odds ratio (OR), augmented by the 95% confidence interval, as an estimate for the enrichment. We used a two-sided Fisher's exact test to test for the association while adjusting for multiple testing using Bonferroni correction. For the identification of important features (single features or feature combinations that describe a set of residues being enriched for pathogenic variants), a randomly subsampled set maintaining 70% of the population and pathogenic variants was considered as a discovery dataset. The remaining set consisting of each 30% population and pathogenic variants formed the validation dataset that was used to confirm the important single features and feature combinations identified in the discovery dataset.

A principal component analysis (PCA) was performed with the stats package (v.3.6.2) to aggregate the variance of the seven different physiochemical pore features across the pore residues. All statistical analyses were performed in R v.4.1.2.

Data availability

Residue-wise feature and distance annotations for all ion channels included in this study can be obtained from Supplementary Tables 2 and 3 and our GitHub repository (https://github.com/LalResearchGroup/Channel_Distances).

Results

Identification of pathogenic variant associated residue features across ion channels

We sought to identify features associated with variant pathogenicity in all ion channels (Fig. 1) Therefore, we collected 2411 pathogenic and 638 probably-pathogenic classified variants from ClinVar[24] and HGMD[29] and 12 546 population missense variants from the gnomAD database.[28] Of these, 2543 (83.1%) pathogenic variants and 5967 (47.6%) population variants could be mapped onto the protein structure of 30 ion-channel proteins (see 'Materials and methods' section for details). We calculated the fold enrichment of patient variants over population variants that served as controls in different residue sets.

For each amino acid residue, we computed 13 features that captured its structural and functional context (Supplementary Fig. 1 and Supplementary Table 2, also available in our GitHub repository: https://github.com/LalResearchGroup/Channel_Distances). These features are mainly independent of each other and include the secondary protein structure, the localization within a functional site (e.g. residues located at a ligand-binding site), and within structurally defined protein regions (e.g. residues located at the pore or inside the membrane) (Supplementary Figs 1A and 2). We then used these features to identify residue characteristics that are associated with pathogenic genetic variants compared to population variants (Fig. 1) across all channel proteins. We did observe a strong correlation across effective sizes for enrichment analyses of pathogenic versus likely-pathogenic variants in comparison with population variants (Supplementary Fig. 3 and Supplementary Table 4; Pearson correlation $\rho = 0.97$, $P = 1.7 \times 10^{-06}$). For this reason and to increase statistical power, pathogenic and likely-pathogenic variants were combined and from here on forward referred to as 'pathogenic' in the manuscript. Groups of residues that share the same features (either one or multiple features) are here referred to as 'residue sets' (Supplementary Fig. 1B, Supplementary material). For example, one residue set contains all residues that form helices (10 500 residues). Furthermore, a set of residues can be also described by a combination of features. For example, 104 residues are all (i) located at the pore; (ii) form helices; and (iii) form protein–protein interactions. The combinatorial assignment of the 13 features resulted in the identification of 163 different residue sets. We then investigated if there is an enrichment of pathogenic variants in each of the 163 residue sets. In the discovery screen, where we used a randomly assigned subset containing each 70% of the curated pathogenic and population variants, we found that 52 residue sets were enriched for pathogenic compared to population variants (Fisher's exact test, Bonferroni-adjusted $P < 0.05$ for

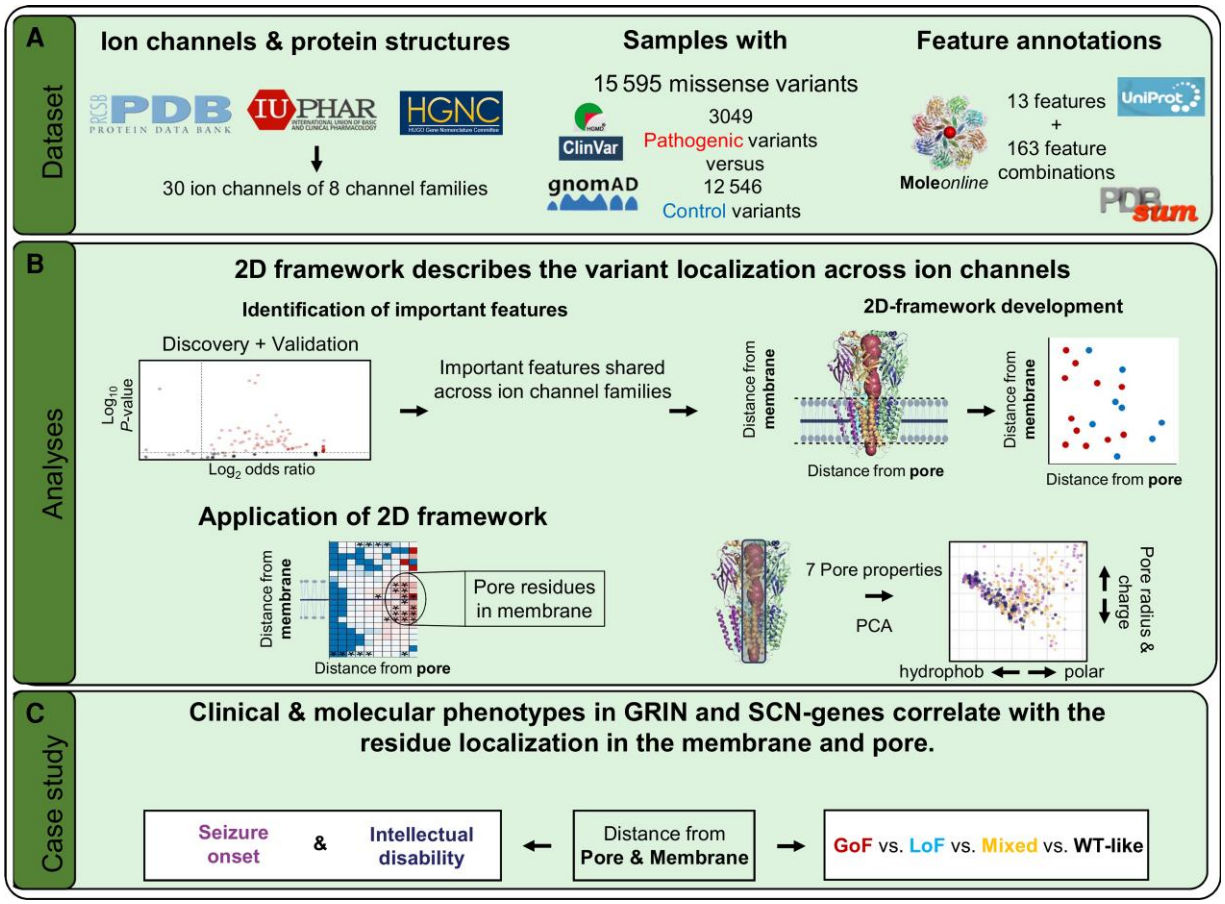

Figure 1 Graphical summary of the study. (**A**) Dataset. First, a subset of ion channels was selected from IUPHAR and screened for available protein structures. Second, missense variants were assembled from three different databases: gnomAD, ClinVar and HGMD. Third, structural features were annotated to describe the location and secondary structure of all residues comprising the ion channels. (**B**) Analyses. Sets of residues described by a single feature or a combination of features that were enriched for pathogenic variants were identified across all ion channels. A subset of these features was identified to show enrichment for pathogenic variants across all considered channel families. On the basis of these shared features a two-dimensional reference framework was developed for describing the location of a residue with respect to the distance from the membrane and the pore. Using this framework, the highest variant burden was observed in pore residues that were located inside the membrane. Finally, within these pore residues, the impact of the biophysical pore environment on the variant burden was investigated. (**C**) Case studies. Our framework identified correlations between functional and clinical phenotypes and the localization of missense variants in the protein structure. GRIN = *GRIN1*, *GRIN2A*, *GRIN2B* genes; SCN = *SCN1A*, *SCN2A*, *SCN8A* genes; Mixed = electrophysiological readouts showed conflicting functional changes; PCA = Principal component analysis; WT = wild-type.

163 tests; Fig. 2A). Of those 52 residue sets, 31 (60%) were enriched for pathogenic variants also in the validation screen, comprising the remaining 30% of variants (Fig. 2B). The highest variant enrichment of pathogenic variants was observed for the set of residues that combines 'pore lining', 'located in the membrane', 'in an alpha-helix secondary structure' and 'being involved in protein–protein interactions' ($OR = 34.5$, $P = 3.6 \times 10^{-19}$). This residue set contains residues from 11 out of the 30 channel genes (37%), including genes encoding for gamma-aminobutyric acid (GABA) receptors, glycine receptors, *N*-methyl-D-aspartate (NMDA)-receptors, potassium channels and transient receptor potential cation channels.

After identification of the residue sets most strongly associated with pathogenic variants across the full dataset of 30 ion channels, we sought to identify the subset of residue sets that are consistently associated with pathogenicity across all ion channels. Since the pathogenicity of homologous residues is similar across gene families,[36] we categorized our set of ion channels into eight evolutionarily derived families and assessed the enrichment of pathogenic versus population variants for each residue set in every channel family. As to residue sets, the lowest number of residue sets that were enriched for pathogenic variants was found in the glycine receptor family ($n = 9$), while the highest number was found in the glutamate receptor family ($n = 28$). We identified four features that contributed, individually or in combination, to all residues sets enriched for pathogenic variants in all eight ion-channel families, namely 'alpha-helix secondary structure', 'localization at the pore', 'alpha-helix secondary structure' and 'coiled secondary structure'.

Spatial distance to pore is associated with variant pathogenicity across all ion channels

The previous analysis identified four features consistently and prominently contributing to variant pathogenicity across all ion-channel families. Out of these, we observed the highest enrichment of pathogenic variants in residues that are located in the membrane ($OR = 4.8$, $P = 1.8 \times 10^{-273}$) and in residues that are pore lining ($OR = 3.8$, $P = 1.8 \times 10^{-28}$). To study the spatial variant localization in association with variant pathogenicity in more detail, we

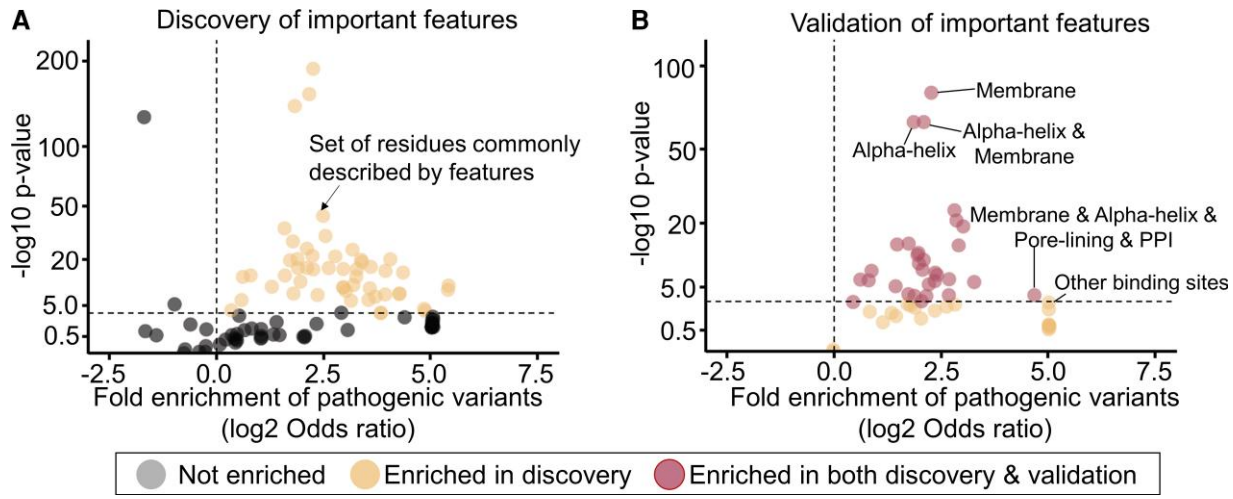

Figure 2 Identification of single features and feature combinations that are enriched for pathogenic variants. (**A**) Scatterplot of the odds ratio (x-axis) versus the *P*-value (y-axis) of the variant burden analysis of pathogenic versus population variants in each of the 163 different sets of residues described by one feature or a feature combination in the discovery cohort (Supplementary Fig. 1). Residue sets with significant enrichment of pathogenic variants after multiple-testing corrections (Fisher's exact test, OR > 1, $P < 0.05$) are displayed above the dotted line. (**B**) Scatterplot of variants in the validation cohort. Only the subset of the 55 features and feature combinations that were found to be enriched for pathogenic variants in the discovery cohort is shown. Residue sets above the dotted line are enriched for pathogenic variants in both the discovery and validation cohort.

developed a two-dimensional framework that normalizes the localization of each residue for each ion channel relative to its distance from the pore axis and the membrane centre (Fig. 3A, Supplementary material and Supplementary Table 3; also available in our GitHub repository: https://github.com/LalResearchGroup/Channel_Distances). This approach enabled us to compare the localization of variants in relation to pore and membrane distance across all ion channels, independently of their structural and sequence similarities. We identified in the combined ion-channel data set the most significant pathogenic variant enrichment in regions close to the pore axis and to the membrane centre, respectively (OR = 6.2, $P = 8.8 \times 10^{-08}$; Fig. 3B). To quantify the added value of our pore distance approach, we compared the enrichment of pathogenic variants with groups of residues defined by the normalized distances to the pore axis and their presence inside the membrane with the transmembrane (TM) domain residue annotation from UniProt. Residues closely located to the pore show a higher enrichment for pathogenic variants compared to population variants than residues covered by the UniProt TM annotation (Supplementary Fig. 4). For each ion-channel class, we observed a strong correlation between a greater relative number of pathogenic versus population variants with a closer distance to the pore axis (Pearson correlation $\rho = -0.98$, $P = 8.0 \times 10^{-07}$; Fig. 3C). In contrast, no significant correlation was found between the greater relative number of pathogenic versus population variants and the localization of the variant with respect to the membrane centre (Pearson correlation $\rho = -0.20$, $P = 1$).

The highest enrichment of pathogenic variants was observed closest to the pore axis. We next investigated the enrichment of pathogenic versus population variants in more refined subsets of residues that were located at or close to the pore, referred to as pore residues (see Supplementary material for details). We observed enrichment for pathogenic variants in all subsets of pore residues that were located inside the membrane and that form a helix or a coiled secondary structure (Fig. 3D). Among those, the highest enrichment of pathogenic variants was observed in helix forming residues that were close to the pore surface and close to the pore axis (OR = 5.6, $P = 1.44 \times 10^{-24}$). However, a similar enrichment of pathogenic variants was observed in membrane-located pore-lining residues forming alpha helices (OR = 5.5, $P = 2.6 \times 10^{-13}$; Fig. 3D), indicating that channel residues close to the pore may be similarly prone to disease-causing variants than pore-lining residues. Interestingly, the enrichment for pathogenic variants in these pore residues differed highly between channel families, having the lowest enrichment in sodium channels (OR = 2.25, $P = 2.1 \times 10^{-05}$) and the highest one in GABA receptors (OR = 54.5, $P = 3.2 \times 10^{-11}$).

Notably, we observed that the fold enrichment for pathogenic compared to population variants is up to five times higher in a subset of ion-channel genes that are depleted for missense variants ($n = 13$ genes, missense-z score[37] > 3.09; Fig. 3E). This subset comprises pore residues that are close to both the pore axis and the membrane surface, but which are not pore lining (OR = 27.3, $P = 1.9 \times 10^{-07}$). Within the set of constraint ion-channel genes, we identified a total of 116 pathogenic variants (7.1% out of 1641) in pore residues that are located inside the membrane whereas only 13 population variants (0.3% out of 4179) were found at these residue positions.

Hydrophobic pore sections contain a higher density of pathogenic versus population variants

Next, after observing that the enrichment of pathogenic versus population variants correlates with the distance from the pore, we asked whether the pathogenicity of a variant also correlates with more detailed biophysical properties of the pore. We computed seven biophysical pore properties, such as hydrophobicity, for each pore region and assigned these properties to each pore residue (see Supplementary material for details, Fig. 4A). Given the high correlation between the seven biophysical pore properties (Supplementary Fig. 5), we performed a PCA to transform them into a set of equivalent but non-correlated variables (the principal components, PCs). The first two PCs discriminated majorly hydrophobic and polar pore sections (PC1, 60.4% variance explained) and the

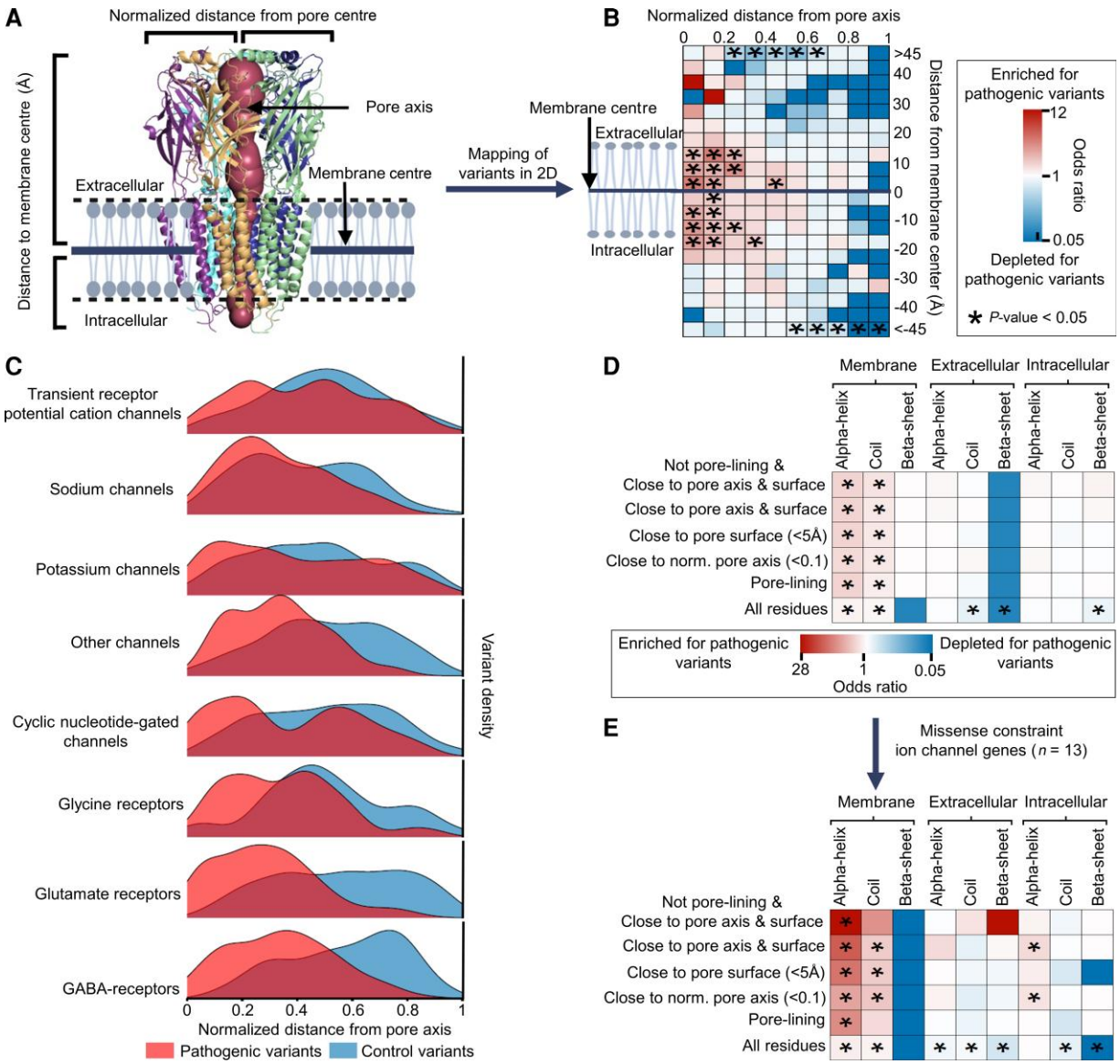

Figure 3 Residue distance from pore and membrane correlates with the pathogenic burden, harbouring the highest burden at membrane-spanning pore residues. (**A**) Graphical representation of the framework that describes the localization of each residue in ion-channel proteins on two dimensions (2D). The localization of each residue was depicted by the gene-wise normalized distance from the pore axis and the distance from the membrane centre, thereby allowing a comparison of the variant distribution across ion channels. (**B**) Table of the combined enrichment or depletion of pathogenic variants across all ion-channel proteins summarized over 200 different 2D regions. Each bin represents a distinct localization in the protein structure that is described by the normalized distance from the pore axis (*x*-axis) and the distance from the membrane centre (*y*-axis). Significant enrichments or depletions of pathogenic variants are indicated with a star. (**C**) The variant densities of pathogenic variants and population variants are displayed along with the normalized distance from the pore axis, separately for each channel class ($n = 8$). (**D**) Table of the enrichment or depletion of pathogenic variants in different sets of pore residues. Each set of pore residues was defined by their location (membrane, extracellular, intracellular) and secondary structure (*x*-axis), together with a description of which residues were considered as pore residues (*y*-axis). Significant values are indicated by a star. (**E**) Table as in **D** but based on a subset of channel genes that are constrained for variants in the general population ($n = 13$ genes).

pore radius and charge (PC2, 14.7% variance explained). Among the seven biophysical properties, hydrophobicity, hydropathy and water solubility made the highest contribution to the first two PCs (as indicated by the length of the projection of the arrow on the *x*- and *y*-axes in Fig. 4B). Hence, small PC1 values describe hydrophobic pore sections whereas larger PC1 values characterize a polar and water-soluble pore environment. We mapped the pathogenic and population variants along PC1 and PC2 and observed enrichment of pathogenic variants compared to population variants at hydrophobic pore sections (median PC1 pathogenic variants: –0.12, median PC1 population variants: 0.14, $P = 1.9 \times 10^{-06}$; Fig. 4C), indicating a correlation between the hydrophobicity of the pore and the pathogenicity of the pore region.

Localization of missense variants correlates with functional effect and clinical phenotype

Next, we explored whether the localization of missense variants can inform the functional effect or clinical phenotype in two independent and structurally different examples. We

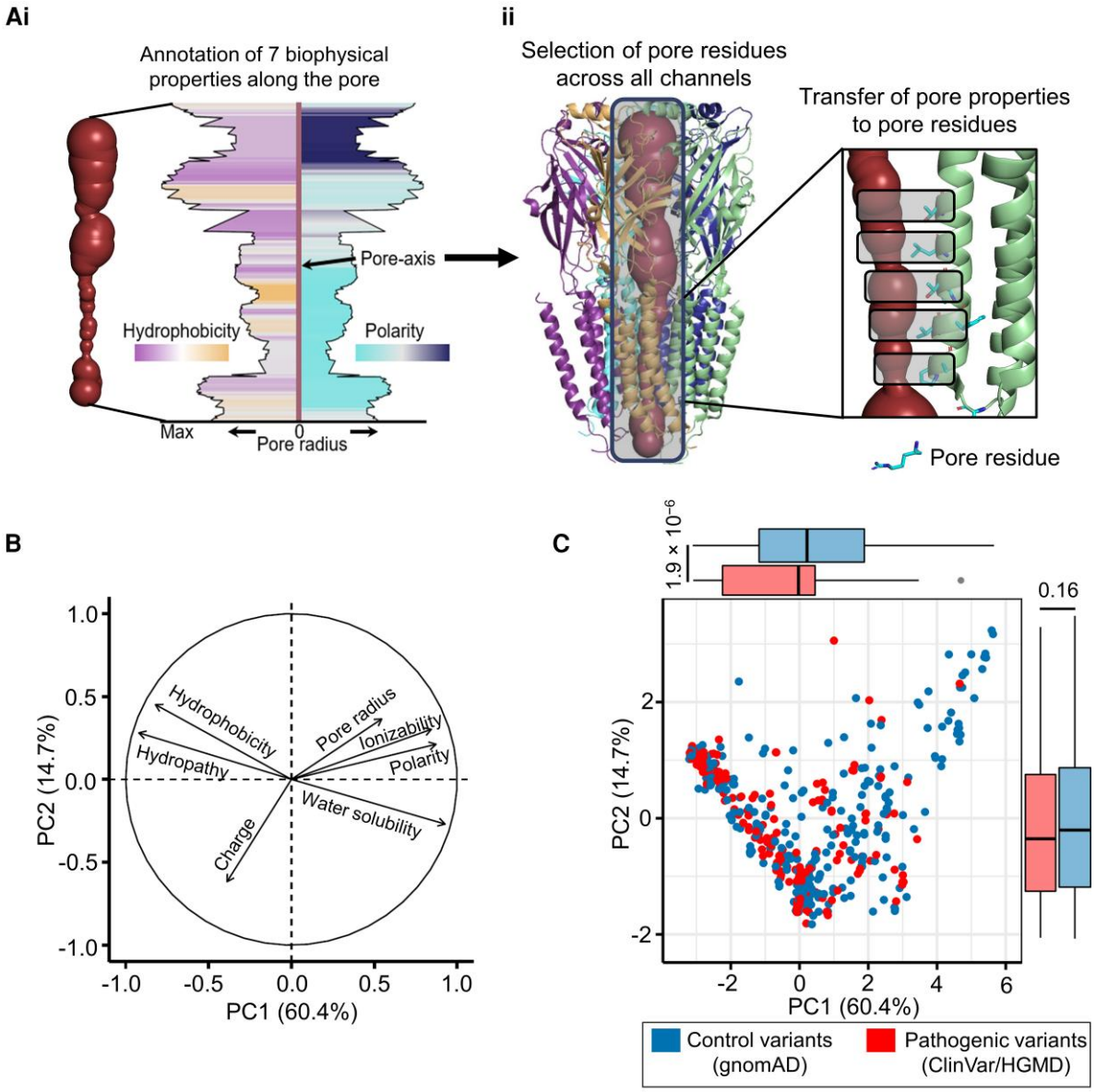

Figure 4 Biophysical pore properties discriminate pathogenic and population variants and identify residues that are most likely to harbour pathogenic variants in the hydrophobic pore sections. [**A(i)**] Graphical representation of the different biophysical pore properties for each pore section along the pore axis. [**A(ii)**] Each pore residue is assigned to the biophysical properties of its closest pore section. (**B**) Contribution of the seven considered biophysical pore properties to the first (*x*-axis, PC1) and second (*y*-axis, PC2) dimension of the PCA that was performed on the pore properties together for all ion-channel proteins. (**C**) Scatterplot along the first two dimensions of the PCA (PC1 and PC2). Each dot represents a pore residue where a pathogenic variant (ClinVar/HGMD) or population variant (gnomAD) was observed. Pore residues where population and pathogenic variant were observed are not shown.

collected 1104 and 314 pathogenic missense variants with curated clinical data across three voltage-gated sodium channel encoding genes (SCN genes: *SCN1A*, *SCN2A*, *SCN8A*) and three genes that encode NMDA receptors (GRIN genes: *GRIN1*, *GRIN2A*, *GRIN2B*), respectively. Most of the residues in sodium channels are located inside the membrane, whereas NMDA receptors have two large extracellular domains on top of their membrane-spanning part. Both ion-channel families have been associated with clinical and molecular heterogeneous neurodevelopmental disorders with epilepsy, intellectual disability and other neurological disorders.[21,38] For the six selected channels, the patient variants clustered within and close to the pore whereas population variants tended to be located at the surface of the protein (Fig. 5A and D). Pathogenic variants in the selected GRIN genes that are associated with an early seizure onset were localized closer to the pore axis than variants associated with a late onset ($P = 0.001$; Fig. 5F). Pathogenic variants in SCN genes that were observed in patients reporting intellectual disability (ID) tend to be closer to the pore ($P = 3.8 \times 10^{04}$; Fig. 5B) than pathogenic variants observed in patients without ID. A similar pattern was observed for the GRIN genes ($P = 5.8 \times 10^{-04}$; Fig. 5E).

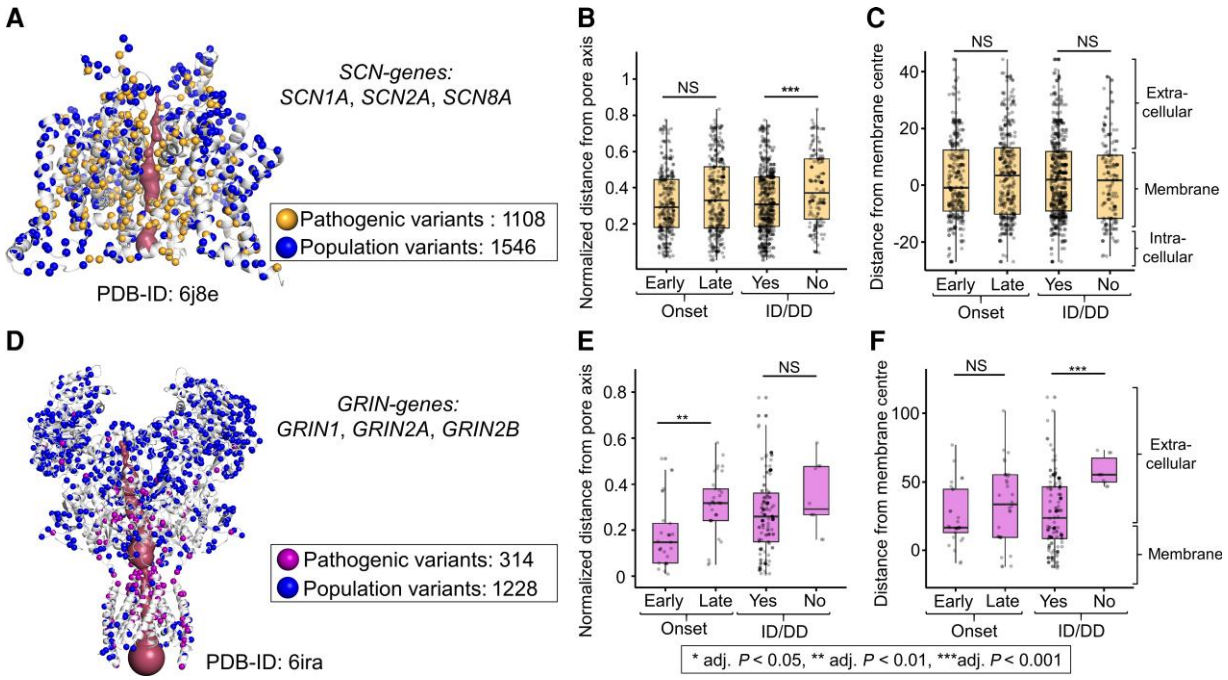

Figure 5 Distances to membrane and pore correlate with the clinical representation in sodium channels and NMDA receptors. (**A**) Nav1.2 protein structure (PDB-ID: 6j8e) encoded by *SCN2A* showing patient variants and population variants observed in *SCN1A*, *SCN2A* and *SCN8A* (SCN genes) that were aligned to *SCN2A*. Pathogenic SCN variants were curated from the SCN Portal, population variants from gnomAD. (**B**) Boxplot of the distribution of patients' variants grouped by their clinical phenotypes along with the normalized distance from the pore centre. Boxes represent patient variants that were associated with an early seizure onset (seizure onset < median onset) or a late seizure onset (seizure onset > median onset) or patient variants associated with intellectual disability (ID) or developmental delay (DD). (**C**) Boxplots showing the same groups as in **C**, but along the distance to the membrane centre. (**D**) Heterotetrameric protein complex consisting of two Glu1N and two Glu2NA subregions (PDB-ID: 6ira) encoded by *GRIN1* and *GRIN2A*, respectively. Patient and population variants observed in *GRIN1*, *GRIN2A* and *GRIN2B* (GRIN genes) are visualized on the structure. Patient variants were recruited from our internal variant database, population variants from gnomAD. Variants in *GRIN2B* were aligned to *GRIN2A* and were visualized on the GluN2A subregions. (**E** and **F**) Boxplots as in **B** and **C** of the clinical phenotypes observed in the GRIN patient cohort.

Furthermore, we investigated the relationship between the functional effects of variants that were tested in 689 electrophysiological experiments[6,22] in association with their localization in GRIN and SCN genes (Fig. 6A and D). On the basis of electrophysiological readouts variants were classified according to https://grin-portal.broadinstitute.org/ to have a GoF (combined: likely/potential-GoF), LOF (combined: likely/potential LoF), mixed (complex) or no effect. Variants in GRIN genes that were classified as GoF variants were located closer to the pore compared to variants classified as LoF (Wilcoxon rank-sum test: GRIN genes: $P = 8.0 \times 10^{-06}$; Fig. 6B and E). In contrast, no significant difference was observed in variants classified as GoF and LoF in SCN genes (Wilcoxon rank-sum test: SCN genes: $P = 0.075$). Similarly, in both gene families, GoF-associated variants were closer located to the membrane centre than variants classified as LoF (Wilcoxon rank-sum test: SCN genes: $P = 9.8 \times 10^{-11}$; GRIN genes: $P = 1.2 \times 10^{-05}$; Fig. 6C and F). In addition, functionally tested variants that showed no difference in their activity compared to the wild-type were only available for the set of GRIN genes and characterized by a larger distance from the membrane centre ($P = 6.9 \times 10^{-14}$) and pore axis ($P = 6.3 \times 10^{-13}$) than variants that were annotated with any functional difference (Fig. 6E and F). For functionally tested variants located at the pore, we further explored whether the variants of different functional effects cluster in distinct biophysical pore environments. Whereas the variants in GRIN genes were scattered along with the PCs of the biophysical pore properties (Supplementary Fig. 6A), GoF classified variants in SCN genes cluster at low PC1 values (median = −2.7) and high PC2 values (median = 1.0) indicating a hydrophobic pore environment, whereas LoF variants were predominantly present at a less hydrophobic pore and charged pore environment (PC1 values median = −0.2; PC2 values median = −0.1; Supplementary Fig. 6B).

Discussion

We performed the first approach to systematically determine the most functionally essential structural features of residues across disease-associated ion channels. Some of the identified features are conserved across evolutionary diverse ion channels. In a proof-of-concept analysis of six neurodevelopmental disorder-associated channelopathies, we show that the newly identified conserved features are correlated with variant pathogenicity, molecular function and clinical phenotype.

Spatial constraints have been shown as a powerful tool to predict pathogenicity in ion channels.[40] We observe that residues that are close to the pore, are located in the membrane, and form an alpha-helix, are most enriched for pathogenic variants across ion channels. Our results are in line with previous molecular biological studies on small sets of genes, which reported ion-channel residues at the pore to be prone to disease-causing mutations and correlate with molecular mutations.[16,41–43] For example, clusters of pathogenic variants had been found in the pore region in some specific ion channels, including voltage-gated sodium channels,[44] NMDA receptors,[41] potassium channels[16] and GABA receptors.[45] Here, we systematically show that across 163 feature combinations characterizing amino acid residues, the pore

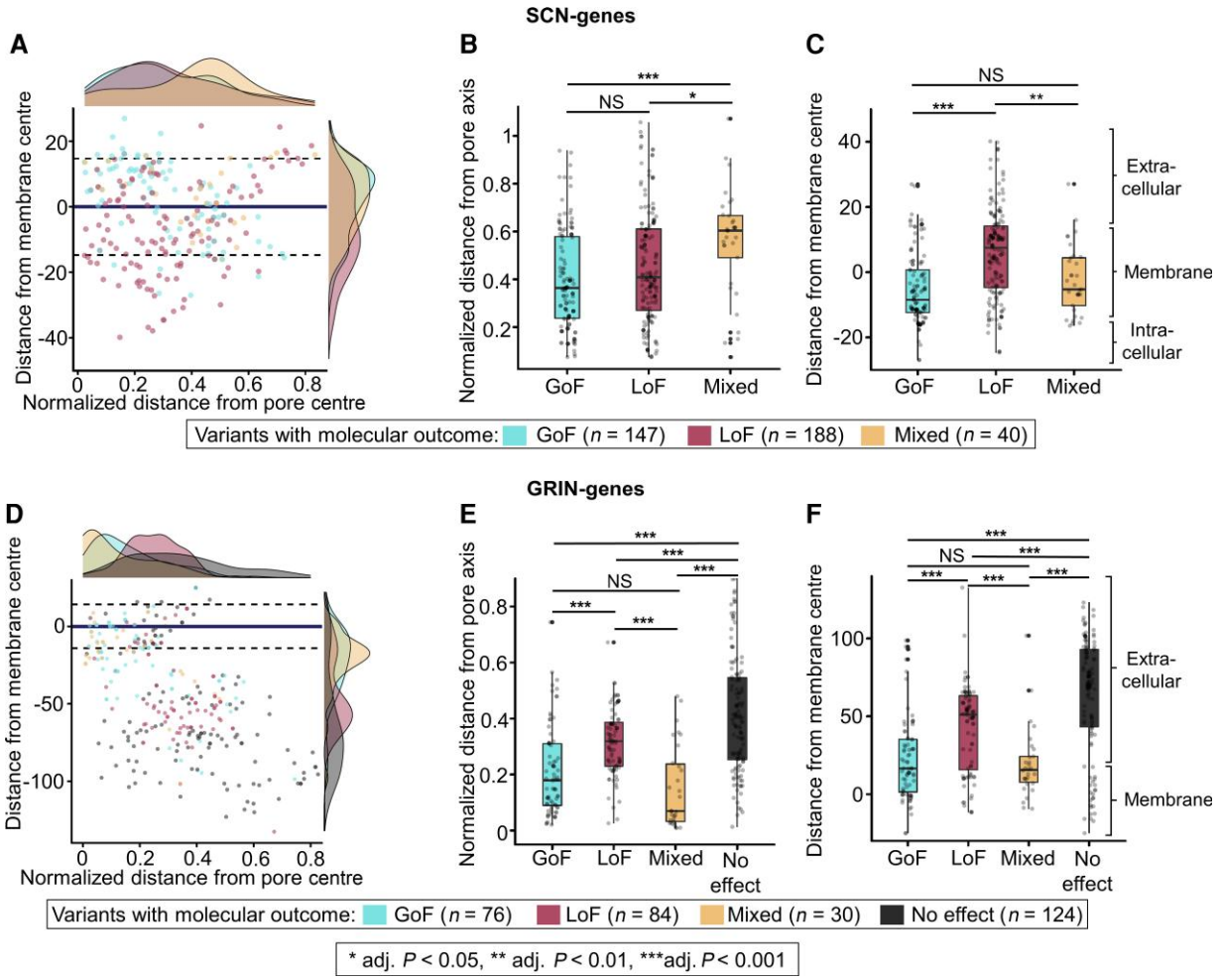

Figure 6 Distances to membrane and pore correlate with the molecular effect in voltage-gated sodium channels and NMDA receptors. (**A**) Scatterplot of variants associated with a molecular effect plotted with respect to the normalized distance to the pore axis (*x*-axis) and the distance from the membrane centre (*y*-axis). All readouts were assigned to the complementary *SCN2A* protein position based on a multiple sequence alignment. Variant with a mixed effect had contrary effects in different electrophysiological measurements.[39] (**B**) Box plot of the molecular effects scattered across the normalized distance from the pore axis in SCN genes. (**C**) Box plot of the molecular effects scattered across the distance from the membrane centre in SCN genes. (**D**) Scatterplot as in **A** visualizing variants associated with a molecular effect in GRIN genes. (**E** and **F**) Box plots as in **B** and **C** show the variants classified with a molecular effect along with the normalized distance from the pore axis and membrane centre in GRIN genes.

features are most important. We were able to map these most important pore features and identify for the first time a spatially conserved pattern across eight different channel families. Although there are well characterized functional domains, such as the voltage sensor in sodium channels, that are distant from the pore and known to be enriched for pathogenic variants,[42,46,47] the overall observation that variants which are spatially more distant from the pore axis are less often pathogenic, remained stable across all eight channel families. The core function of the pore-forming channel subunits is to allow a passive ion flow through the pore.[3] Residues comprising the water-filled pore have stringent requirements in terms of their precise location relative to other residues to maintain pore dimensions and allow channel opening and closing to proceed with normal stimuli.[48,49] The pore region is under strong evolutionary selection, and pathogenic variants probably alter the side chain size or characteristics of the pore and interfere with the pore function.[41,50] Our observation that pathogenic variants compared to populations variants are enriched at residues close to the pore could thus be explained by the increased likelihood of the mutations interfering with the pore function. We make our generated distance annotations available in Supplementary Table 3 as well as our GitHub repository (https://github.com/LalResearchGroup/Channel_Distances) and discuss how our annotations can be used in variant classification or research projects (see Supplementary material).

Structure-based bioinformatic methods that score physicochemical residue properties at the pore, such as hydrophobicity, have been shown to improve the prediction of the channel conformational states and to provide insights into the channel gating processes.[51] In our study, we observed that hydrophobicity showed the strongest association among biophysical properties with variant pathogenicity for pore residues across all investigated ion channels. Previous studies of the voltage-gated sodium channel $Na_V1.1$[44] and $Na_V1.7$[52] identified a correlation between pore hydrophobicity and variant pathogenicity. It has been suggested that a hydrophobic pore section can present an energetic barrier to ion permeation interrupting the channel gating, without having the pore physically closed due to de-wetting of the hydrophobic environment, a mechanism known as hydrophobic gating.[53,54] As a

consequence, the variants we identified to be located at the hydrophobic pore sections may not only affect the channel activity through a possible change in the pore size but also the alterations of the hydrophobic gating.[55]

We investigated the identified pathogenicity features associated with pore axis and membrane centre distance in molecular and clinical datasets. Prediction of phenotypes will guide patients and families and enable a more individualized prognosis. We showed the variant position in voltage-gated sodium channel genes (SCN genes) and NMDA receptors (GRIN genes), which are both associated with early-onset epilepsies and neurodevelopmental disorders,[47,56–59] correlates with the functional effect and clinical phenotype. Overall, our spatial distance scoring approach agrees with previously reported observations that have been made on association with specific domains. For example, we find that LoF variants in SCN genes are close to the pore axis in the extracellular side of the membrane, which is the region where the selectivity filter is located, whereas GoF variants cluster in the intracellular site of the membrane.[42] Such knowledge of the molecular functional consequences of a variant may inform treatment decisions.[60–63] In contrast to the previous analysis, which compared the functional effect across predefined protein regions such as the voltage sensor, the S5 and S6 Linker, or the selectivity filter,[42] our approach has a higher spatial resolution for association analysis. Our approach can be applied to proteins in different conformational states and to ion-channel proteins where the function of domains is not well studied.

Still, our study has several limitations. First, while hundreds of genes that encode voltage- and ligand-gated ion channels have so far been identified in humans, we limited our study to only those 30 disease-associated ion channels for which 3D structural data describing proteins and protein complexes were available. Nevertheless, our study represents the most comprehensive assessment of important features across ion channels originating from eight different protein families.[26] This diversity of ion channels that are captured in our dataset suggests that our findings may also be generalizable to other ion channels, such as aquaporins, chloride channels or piezo channels that were not part of the present study. Second, to validate generalizability and to explore gene-specific important protein features, future gene-level association analyses should be performed. Given the exponential increase in sequencing data generation, gene-level analysis should be sufficiently powered in the next few years. Third, ion channels can be observed in different conformational states, namely open, closed or inactivated conformation,[64] and this affects the location of the residues in 3D. We selected a variety of protein structures with different conformational changes, indicating that our results are probably valid for a spectrum of conformational changes. However, once human protein structures of many ion channels will be available in several conformations, our approach can be used to systematically study the structural changes and effects of the variant localization in different conformations.

We introduced a new potentially powerful approach to identify pathogenic enriched regions and identify molecular and clinical correlates across diverse ion channels. In the foreseeable future, high-quality data on protein structures and protein complexes will be available for most ion channels in the light of recent improvements in *in silico* structure prediction tools,[65] and this will allow a wider application of our approach. Further, with the availability of upcoming high-throughput mutagenesis screens for ion channels, we will refine our analysis by providing high-quality assessments of the effect of variants. Finally, our method could be expanded to other protein classes, where two features represent a horizontal and vertical axis along with the protein structure, such as membrane-located transporter, to study pathogenicity and molecular and clinical phenotypes in context variant localization.

Funding

Funding for this work was provided from the *Bundesministrium für Bildung und Forschung* (German Federal Ministry for Education and Research Treat-ION, BMBF 01GM1907D) to D.L., T.B. and P.M., the Fonds National de la Recherche Luxembourg (Research Unit FOR-2715, FNR grant INTER/DFG/21/16394868 MechEPI2) to P.M., the *Bundesministrium für Bildung und Forschung* (German Federal Ministry for Education and Research Treat-Ion2, 01GM2210B) to P.M., the Agencia Nacional de Investigación y Desarrollo (ANID, PAI77200124) of Chile to E.P., the Familie SCN2A foundation 2020 Action Potential (website: FamilieSCN2A Foundation) Grant to E.P., the Dravet Syndrome Foundation (grant number, 272016) to D.L. and the National Institute of Neurological Disorders and Stroke (NIH NINDS) (Channelopathy-Associated Epilepsy Research Center, 5-U54-NS108874) to D.L.

Competing interests

The authors report no competing interests.

Supplementary material

Supplementary material is available at *Brain* online.

References

1. Kim JB. Channelopathies. *Korean J Pediatr*. 2014;57:1–18.
2. Bajaj S, Ong ST, Chandy KG. Contributions of natural products to ion channel pharmacology. *Nat Prod Rep*. 2020;37:703–16.
3. Nayak S, Batalov S, Jegla T, Zmasek C. Evolution of the human ion channel set. *Comb Chem High Throughput Screen*. 2009;12:2–23.
4. Addis L, Virdee JK, Vidler LR, Collier DA, Pal DK, Ursu D. Epilepsy-associated GRIN2A mutations reduce NMDA receptor trafficking and agonist potency—Molecular profiling and functional rescue. *Sci Rep*. 2017;7:66.
5. Rivaud MR, Delmar M, Remme CA. Heritable arrhythmia syndromes associated with abnormal cardiac sodium channel function: Ionic and non-ionic mechanisms. *Cardiovasc Res*. 2020;116:1557–70.
6. Hansen KB, Wollmuth LP, Bowie D, *et al*. Structure, function, and pharmacology of glutamate receptor ion channels. *Pharmacol Rev*. 2021;73:298–487.
7. Allen NM, Weckhuysen S, Gorman K, King MD, Lerche H. Genetic potassium channel-associated epilepsies: Clinical review of the Kv family. *Eur J Paediatr Neurol*. 2019;24:105–16.
8. Chuang SH, Reddy DS. Genetic and molecular regulation of extrasynaptic GABA-A receptors in the brain: Therapeutic insights for epilepsy. *J Pharmacol Exp Ther*. 2018;364:180–97.
9. Catterall WA. Sodium channel mutations and epilepsy. In: Noebels JL, Avoli M, Rogawski MA, Olsen RW, Delgado-Escueta AV, eds. *Jasper's basic mechanisms of the epilepsies*. 4th edn. National Center for Biotechnology Information (US); 2012.

10. Klassen T, Davis C, Goldman A, *et al*. Exome sequencing of ion channel genes reveals complex profiles confounding personal risk assessment in epilepsy. *Cell*. 2011;145:1036–48.
11. Baez-Nieto D, Allen A, Akers-Campbell S, *et al*. Analysing an allelic series of rare missense variants of CACNA1I in a Swedish schizophrenia cohort. *Brain*. 2022;145:1839–53.
12. Silk M, Pires DEV, Rodrigues CHM, *et al*. MTR3D: Identifying regions within protein tertiary structures under purifying selection. *Nucleic Acids Res*. 2021;49:W438–45.
13. Kamburov A, Lawrence MS, Polak P, *et al*. Comprehensive assessment of cancer missense mutation clustering in protein structures. *Proc Natl Acad Sci USA*. 2015;112:E5486–95.
14. Sivley RM, Dou X, Meiler J, Bush WS, Capra JA. Comprehensive analysis of constraint on the spatial distribution of missense variants in human protein structures. *Am J Hum Genet*. 2018; 102:415–26.
15. Reynolds C, King MD, Gorman KM. The phenotypic spectrum of SCN2A-related epilepsy. *Eur J Paediatr Neurol*. 2020;24:117–22.
16. Goto A, Ishii A, Shibata M, Ihara Y, Cooper EC, Hirose S. Characteristics of KCNQ2 variants causing either benign neonatal epilepsy or developmental and epileptic encephalopathy. *Epilepsia*. 2019;60:1870–80.
17. Johannesen KM, Liu Y, Koko M, *et al*. Genotype-phenotype correlations in SCN8A-related disorders reveal prognostic and therapeutic implications. *Brain*. 2022;145:2991–3009.
18. Strehlow V, Heyne HO, Vlaskamp DRM, *et al*. GRIN2A-related disorders: Genotype and functional consequence predict phenotype. *Brain*. 2019;142:80–92.
19. Kelly M, Park M, Mihalek I, *et al*. Spectrum of neurodevelopmental disease associated with the GNAO1 guanosine triphosphate-binding region. *Epilepsia*. 2019;60:406–18.
20. Wu RH, Tang WT, Qiu KY, *et al*. Identification of novel CSNK2A1 variants and the genotype–phenotype relationship in patients with Okur–Chung neurodevelopmental syndrome: A case report and systematic literature review. *J Int Med Res*. 2021;49: 03000605211017063.
21. Brunklaus A, Du J, Steckler F, *et al*. Biological concepts in human sodium channel epilepsies and their relevance in clinical practice. *Epilepsia*. 2020;61:387–99.
22. Brunklaus A, Feng T, Brünger T, *et al*. Gene variant effects across sodium channelopathies predict function and guide precision therapy. *Brain*. Published online 17 January 2022. doi:10.1093/brain/awac006
23. Armstrong JF, Faccenda E, Harding SD, *et al*. The IUPHAR/BPS guide to pharmacology in 2020: Extending immunopharmacology content and introducing the IUPHAR/MMV guide to malaria pharmacology. *Nucleic Acids Res*. 2020;48:D1006–21.
24. Landrum MJ, Lee JM, Benson M, *et al*. Clinvar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. 2018;46:D1062–7.
25. Berman HM, Westbrook J, Feng Z, *et al*. The Protein Data Bank. *Nucleic Acids Res*. 2000;28:235–42.
26. Yates B, Gray KA, Jones TEM, Bruford EA. Updates to HCOP: The HGNC comparison of orthology predictions tool. *Briefings in Bioinformatics*. 2021;22:bbab155.
27. Consortium TU. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res*. 2019;47:D506–15.
28. Karczewski KJ, Francioli LC, Tiao G, *et al*. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581:434–43.
29. Stenson PD, Ball EV, Mort M, *et al*. Human Gene Mutation Database (HGMD): 2003 update. *Hum Mutat*. 2003;21:577–81.
30. SCN Portal. Accessed 1 December 2021. https://scn-portal.broadinstitute.org/
31. GRIN Portal. Accessed 1 December 2021. https://grin-portal.broadinstitute.org/
32. CFERV. Accessed 1 December 2021. http://functionalvariants.emory.edu/
33. Touw WG, Baakman C, Black J, *et al*. A series of PDB-related databanks for everyday needs. *Nucleic Acids Res*. 2015;43:D364–8.
34. Laskowski RA, Chistyakov VV, Thornton JM. PDBsum more: New summaries and analyses of the known 3D structures of proteins and nucleic acids. *Nucleic Acids Res*. 2005;33(Suppl 1): D266–8.
35. Sehnal D, Svobodová Vařeková R, Berka K, *et al*. MOLE 2.0: Advanced approach for analysis of biomacromolecular channels. *J Cheminform*. 2013;5:39.
36. Lal D, May P, Perez-Palma E, *et al*. Gene family information facilitates variant interpretation and identification of disease-associated genes in neurodevelopmental disorders. *Genome Med*. 2020;12:28.
37. Lek M, Karczewski KJ, Minikel EV, *et al*. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536:285–91.
38. Benke TA, Park K, Krey I, *et al*. Clinical and therapeutic significance of genetic variation in the GRIN gene family encoding NMDARs. *Neuropharmacology*. 2021;199:108805.
39. Brunklaus A, Schorge S, Smith AD, *et al*. SCN1A variants from bench to bedside—Improved clinical prediction from functional characterization. *Human Mutat*. 2020;41:363–74.
40. Li B, Roden DM, Capra JA, *et al*. The 3D mutational constraint on amino acid sites in the human proteome. *Nature Communications*. 2022;13:3273.
41. Li J, Zhang J, Tang W, *et al*. De novo GRIN variants in NMDA receptor M2 channel pore-forming loop are associated with neurological diseases. *Human Mutat*. 2019;40:2393–413.
42. Heyne HO, Baez-Nieto D, Iqbal S, *et al*. Predicting functional effects of missense variants in voltage-gated sodium and calcium channels. *Sci Transl Med*. 2020;12:eaay6848.
43. Glazer AM, Wada Y, Li B, *et al*. High-throughput reclassification of SCN5A variants. *Am J Hum Genet*. 2020;107:111–23.
44. Zuberi SM, Brunklaus A, Birch R, Reavey E, Duncan J, Forbes GH. Genotype-phenotype associations in SCN1A-related epilepsies. *Neurology*. 2011;76:594–600.
45. Hernandez CC, Zhang Y, Hu N, *et al*. GABA A receptor coupling junction and pore GABRB3 mutations are linked to early-onset epileptic encephalopathy. *Sci Rep*. 2017;7:15903.
46. Skrenkova K, Song JM, Kortus S, *et al*. The pathogenic S688Y mutation in the ligand-binding domain of the GluN1 subunit regulates the properties of NMDA receptors. *Sci Rep*. 2020;10:18576.
47. Endele S, Rosenberger G, Geider K, *et al*. Mutations in GRIN2A and GRIN2B encoding regulatory subunits of NMDA receptors cause variable neurodevelopmental phenotypes. *Nat Genet*. 2010;42:1021–6.
48. Schlegel AM, Haswell ES. Charged pore-lining residues are required for normal channel kinetics in the eukaryotic mechanosensitive ion channel MSL1. *Channels (Austin)*. 2020;14:310–25.
49. Liao M, Cao E, Julius D, Cheng Y. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature*. 2013;504:107–12.
50. Pan X, Li Z, Jin X, *et al*. Comparative structural analysis of human Nav1.1 and Nav1.5 reveals mutational hotspots for sodium channelopathies. *Proc Natl Acad Sci USA*. 2021;118:e2100066118.
51. Rao S, Klesse G, Stansfeld PJ, Tucker SJ, Sansom MSP. A heuristic derived from analysis of the ion channel structural proteome permits the rapid identification of hydrophobic gates. *Proc Natl Acad Sci USA*. 2019;116:13989–95.
52. Xenakis MN, Kapetis D, Yang Y, *et al*. Hydropathicity-based prediction of pain-causing NaV1.7 variants. *BMC Bioinformatics*. 2021;22:212.

53. Beckstein O, Biggin PC, Sansom MSP. A hydrophobic gating mechanism for nanopores. *J Phys Chem B*. 2001;105:12902–5.
54. Bertil H. *Ionic channels of excitable membranes*. 3rd edn. Sinauer; 2001.
55. Yazdani M, Jia Z, Chen J. Hydrophobic dewetting in gating and regulation of transmembrane protein ion channels. *J Chem Phys*. 2020;153:110901.
56. Claes L, Del-Favero J, Ceulemans B, Lagae L, Van Broeckhoven C, De Jonghe P. *De novo* mutations in the sodium-channel gene SCN1A cause severe myoclonic epilepsy of infancy. *Am J Hum Genet*. 2001;68:1327–32.
57. Veeramah KR, O'Brien JE, Meisler MH, *et al*. *De novo* pathogenic SCN8A mutation identified by whole-genome sequencing of a family quartet affected by infantile epileptic encephalopathy and SUDEP. *Am J Hum Genet*. 2012;90:502–10.
58. Matalon D, Goldberg E, Medne L, Marsh ED. Confirming an expanded spectrum of SCN2A mutations: A case series. *Epileptic Disord*. 2014;16:13–8.
59. Hamdan FF, Gauthier J, Araki Y, *et al*. Excess of de novo deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability. *Am J Hum Genet*. 2011; 88:306–16.
60. Krey I, von Spiczak S, Johannesen KM, *et al*. L-serine treatment is associated with improvements in behavior, EEG, and seizure frequency in individuals with GRIN-related disorders due to null variants. *Neurotherapeutics*. 2022;19:334–41.
61. Soto D, Olivella M, Grau C, *et al*. L-Serine dietary supplementation is associated with clinical improvement of loss-of-function GRIN2B-related pediatric encephalopathy. *Sci Signal*. 2019;12: eaaw0936.
62. Pierson TM, Yuan H, Marsh ED, *et al*. GRIN2A mutation and early-onset epileptic encephalopathy: Personalized therapy with memantine. *Ann Clin Transl Neurol*. 2014;1:190–8.
63. Wolff M, Johannesen KM, Hedrich UBS, *et al*. Genetic and phenotypic heterogeneity suggest therapeutic implications in SCN2A-related disorders. *Brain*. 2017;140:1316–36.
64. Yoder N, Yoshioka C, Gouaux E. Gating mechanisms of acid sensing ion channels. *Nature*. 2018;555:397–401.
65. Jumper J, Richard E, Pritzel A, *et al*. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596:583–9.