Transcriptomic correlates of neuron electrophysiological diversity

Shreejoy J. Tripathy*, Lilah Toker, Brenna Li, Cindy-Lee Crichlow, Dmitry Tebaykin, B. Ogan Mancarci, Paul Pavlidis*

Michael Smith Laboratories and Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada

* stripathy@msl.ubc.ca (SJT); paul@msl.ubc.ca (PP)

Abstract

How neuronal diversity emerges from complex patterns of gene expression remains poorly understood. Here we present an approach to understand electrophysiological diversity through gene expression by integrating pooled- and single-cell transcriptomics with intracellular electrophysiology. Using neuroinformatics methods, we compiled a brain-wide dataset of 34 neuron types with paired gene expression and intrinsic electrophysiological features from publically accessible sources, the largest such collection to date. We identified 420 genes whose expression levels significantly correlated with variability in one or more of 11 physiological parameters. We next trained statistical models to infer cellular features from multivariate gene expression patterns. Such models were predictive of gene-electrophysiological relationships in an independent collection of 12 visual cortex cell types from the Allen Institute, suggesting that these correlations might reflect general principles relating expression patterns to phenotypic diversity across very different cell types. Many associations reported here have the potential to provide new insights into how neurons generate functional diversity, and correlations of ion channel genes like *Gabrd* and *Scn1a* (Nav1.1) with resting potential and spiking frequency are consistent with known causal mechanisms. Our work highlights the promise and inherent challenges in using cell type-specific transcriptomics to understand the mechanistic origins of neuronal diversity.

Author summary

Brain cell types have different electrical features, determined by the genes that each cell expresses. By combining data from hundreds of articles studying individual cell types in isolation, we developed a dataset that combines neuron gene expression patterns with their electrical characteristics. We asked if patterns of gene expression could predict a neuron's electrical features; for example, if a neuron that expresses more of a sodium channel also tends to fire action potentials more frequently. We found hundreds of such statistical correlations that also replicated across brain cell types and regions. These relationships provide a starting point for understanding how alterations in the gene expression result in alterations in electrical functioning of neurons and brain circuits.

Introduction

A major goal of neuroscience has been to understand the mechanistic origins of neuronal electrophysiological phenotypes. Such electrical features help define the computational functions of each neuron [1,2], and further, specific electrophysiological deficits contribute to brain disorders such as epilepsy, ataxia, and autism [3–5].

The molecular basis of neuron electrophysiology is complex. There are over 200 mammalian ion channel and transporter genes whose products influence a neuron's electrophysiological phenotype [6–9]. Numerous additional genes regulate channel functional expression through initiating gene transcription and alternative splicing, post-translational modifications, and trafficking channels to and from the membrane surface [10–12]. Even morphological features contribute to cellular electrophysiology [13]. Recent genetic studies in human epileptic and neuropsychiatric patients provide convergent evidence, as mutations in many genes reflecting multiple functional pathways are associated with these disorders [4,14–16]. In light of this complexity, the gold standard employed by neurophysiologists is to use gene knockouts or pharmacology to assay how electrophysiological function changes following protein disruption [7,8]. However, these single-gene focused methods are relatively low-throughput and many potentially relevant genes have yet to be studied for their electrophysiological function.

Cell type-specific transcriptomics, enabling genome-wide assay of quantitative mRNA expression levels, provides a lucrative avenue for discovering novel genes that might contribute to specific aspects of cellular physiology [17,18]. Correlation-based approaches have been proposed that pair single-cell expression profiling with patch-clamp electrophysiology [19–21]. These approaches leverage the biological variability observed across a collection of cells to identify gene expression patterns correlated with cellular phenotypic differences. Generalizing from these studies has proven challenging however, since they typically have been focused on a limited number of cell types. Similarly, and perhaps more critically, there are typically hundreds to thousands of genes correlated with electrophysiological variability[22]. Thus it has been difficult from this data to pin down how individual genes might shape specific cellular phenotypes. Though making use of larger and more diverse collections of cell types could provide a potential solution, collecting such reference data is immensely resource- and labor-intensive.

Here, we present an approach for correlating cell type-specific transcriptomics with neuronal electrophysiological features. We leverage neuroinformatics methods to build a novel reference dataset on brain-wide neuronal gene expression and intrinsic electrophysiological feature diversity. The compiled dataset reflects the neuronal characterization efforts of hundreds of investigators as well as our efforts to compile and normalize these data for unified mega-analysis [23–25]. From this data, we identified hundreds of genes whose expression levels significantly correlate with specific electrophysiological features (e.g., resting potential or maximum spiking frequency). Illustrating the generalizability of these results, we could use these correlations to predict the ephys parameters of an independent neocortex-specific dataset from the Allen Institute. In addition, many of these genes have been further found to directly regulate neuronal electrophysiology, suggesting that some of the correlations reported here likely reflect novel causal relationships. Our findings present a major step for understanding how a multitude of genes contribute to cell type-specific phenotypic diversity.

Results

Our overall approach was to first compile a reference dataset of brain cell type-specific transcriptomes paired with cell type-specific electrophysiological (ephys) profiles. We then assessed the ability of gene expression to statistically explain variance in specific ephys

properties. We next validated whether these gene-ephys relationships generalized using an independent dataset on visual cortex neurons collected by the Allen Institute for Brain Science (AIBS). Lastly we made use of literature review to establish whether any of these gene-ephys correlations had been previously shown to be causal.

Discovery and validation datasets

To construct our primary dataset for gene-ephys correlation analysis, we adapted and combined two databases developed and curated by our group. The first, NeuroExpresso, a database containing microarray-based transcriptomes collected from samples of purified mouse brain cell types under normal conditions [23]. The second, NeuroElectro, a database of rodent neuronal electrophysiological profiles manually curated from the published literature reflecting intracellular ephys characterization of normal, non-treated cells [24,25]. From NeuroElectro's initial publication, we have massively expanded the resource from 331 to 968 articles and have made essential improvements that allow more fine-grained annotation of neuron subtypes and curation of more electrophysiological features.

Given the methodological heterogeneity of the primary data comprising these databases, we applied a number of quality control filtering and cross-laboratory standardization approaches (see Methods and S1 Fig). These include careful re-analysis of neuron type-specific transcriptomes for cellular contamination (e.g., astrocytes, glia) and statistical approaches to normalize ephys measurements for lab-specific experimental conditions (e.g., animal age and slice recording temperatures). We obtained neuron type-specific paired gene expression and ephys data by carefully aligning these databases on cell type identity, making use of our detailed annotations of each sample's specific cell type (Fig 1A, left). This harmonization allows us to merge cell types defined using orthologous criteria, e.g., gene expression data derived from transgenic lines with ephys data collected from cells defined by traditional morpho-electric criteria [26]. The final "discovery" reference dataset is composed of 34 neuron types sampled throughout the brain and reflects cell types with diverse circuit roles, neurotransmitters, and developmental stages (summarized in Table 1 and S2 Table).

For validation we utilized an independent dataset characterizing neurons from adult mouse primary visual cortex collected by the Allen Institute for Brain Science. Here, genetically labeled cells were characterized either for their transcriptomic profiles, using single-cell RNA sequencing (scRNAseq) [27], or their electrophysiological properties, using patch-clamp electrophysiology *in vitro* with standardized protocols (http://celltypes.brain-map.org/). Importantly, for both expression and ephys characterization, the same mouse lines for genetically labeling specific populations of cells were used, making it straightforward to combine samples post-hoc, yielding a final "validation" dataset composed of 12 unique cell types (Table 2). Averaging data across labeled single cells within a mouse line also helps mitigate the influence of cell-to-cell variability and technical "dropouts" in the scRNAseq data [18]. Given the smaller number of cell types present in the AIBS dataset we chose to use these data primarily for validation and generalization of findings made using the discovery dataset. Note that for both the discovery and validation datasets, electrophysiological and gene expression values are from separate cells.

Analysis approach

Our primary analysis focus was to understand how cell type-specific expression of individual genes might statistically explain the variance in electrophysiological parameters observed across cell types (Fig 1A, right). For example, how does *Scn1a* (Na$_v$1.1) expression correlate with neuronal maximum firing rates? Which genes are most correlated with cellular resting
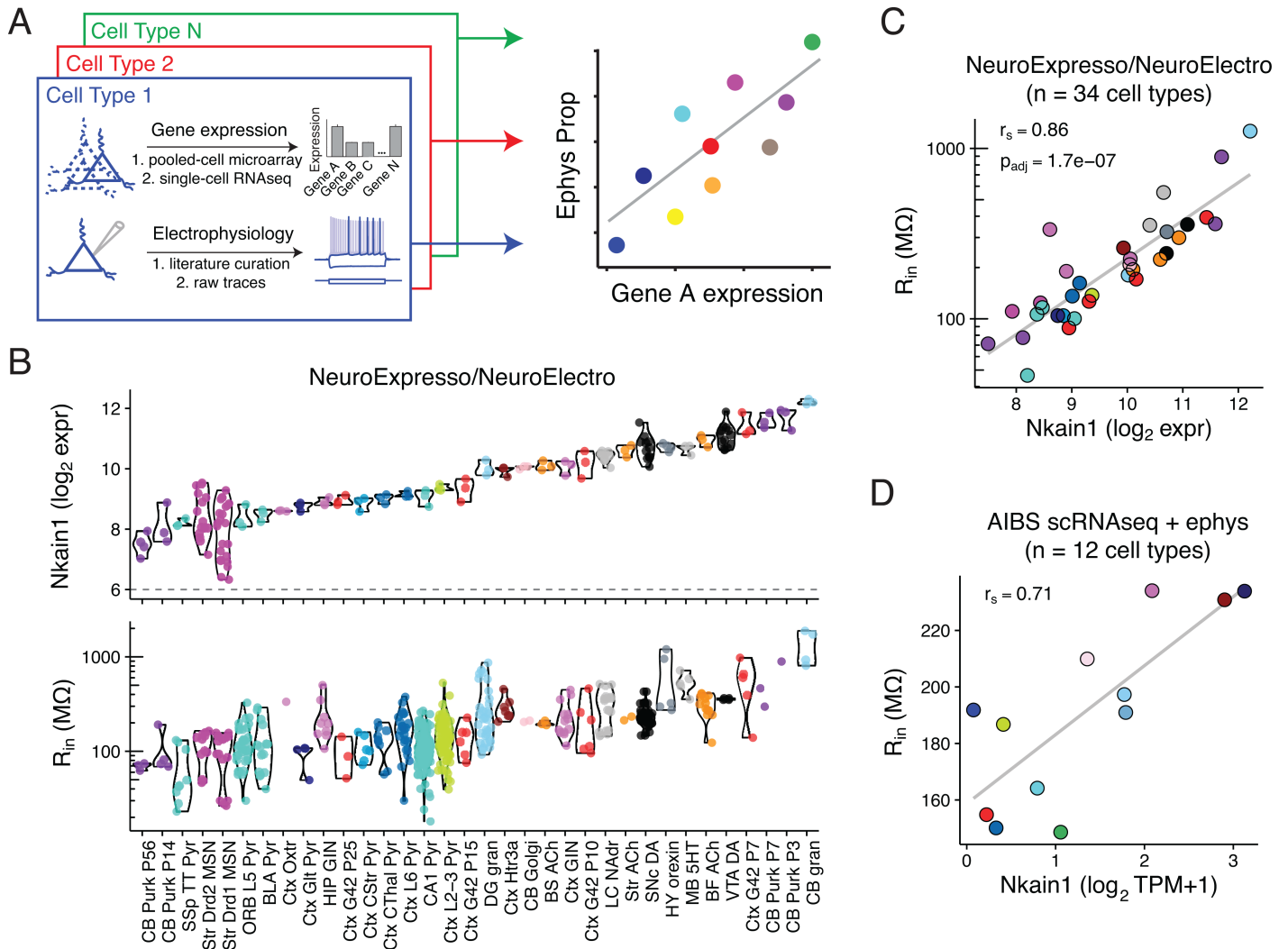
Fig 1. Correlating cell type-specific gene expression with electrophysiological diversity. A) Illustration of transcriptomic and ephys data compilation by cell type (left) and correlation analysis of single gene expression by ephys parameter diversity (right). B) Top row: Gene expression levels of *Nkain1* across 34 neuron types sampled from the combined NeuroExpresso/NeuroElectro dataset. Each dot reflects a unique transcriptomic sample collected from purified cells and y-axis is in units of log2 expression (i.e., each increment reflects a 2-fold change in expression level). Dashed line at 6 indicates approximate level of background expression. Bottom row: Input resistance values for the same cell types in top row. Individual dots reflect population mean electrophysiological values manually curated from individual articles represented in the NeuroElectro database, following experimental condition normalization. C) Same data as in B, but data has been summarized by the mean (expression, x-axis) or median (ephys, y-axis) value within each cell type. r_s indicates Spearman rank correlation and p_{adj} indicates Benjamini Hochberg false discovery rate. Note that cell types with high R_{in}, such as cerebellar granule cells and midbrain dopaminergic cells, express high levels of *Nkain1* whereas cell types with low R_{in}, including neocortical and hippocampal pyramidal cells, express low levels of *Nkain1*. D) Corresponding summary data from the Allen Institute for Brain Science (AIBS) Cell Types dataset. Dots reflect averaged values from 12 individual mouse cre-lines and are detailed in **Table 2**. Expression values are based on single-cell RNAseq (scRNAseq), quantified as Transcripts Per Million (TPM). Ephys values are based on single-cell characterization *in vitro*.

https://doi.org/10.1371/journal.pcbi.1005814.g001

membrane potentials? We primarily chose to employ a single-gene focused approach (utilizing Spearman rank correlations) because of sample size considerations, reasoning that we did not have enough unique cell types in both the discovery or validation datasets to rigorously pursue a combinatorial gene approach. However, as this single-gene focus limits our ability to identify highly combinatorial and/or redundant or degenerate gene-ephys relationships [28,29], we further pursued a machine learning approach where we used sparse, regularized linear models to relate multivariate gene expression to ephys features.

Table 1. Descriptions for neuron types composing the NeuroExpresso/NeuroElectro discovery dataset. References for individual transcriptomic and electrophysiological samples are available in **S2 Table**.

Neuron Type	Abbreviation
Basal forebrain cholinergic cells	BF ACh
Basolateral amygdala pyramidal cells	BLA Pyr
Brain stem cholinergic cells	BS ACh
Cerebellum Golgi cells	CB Golgi
Cerebellum granule cells	CB gran
Cerebellum Purkinje cells, P14	CB Purk P14
Cerebellum Purkinje cells, P3	CB Purk P3
Cerebellum Purkinje cells, P56	CB Purk P56
Cerebellum Purkinje cells, P7	CB Purk P7
Dentate gyrus granule cells	DG gran
Frontal cortex layer 5 pyramidal cells	ORB L5 Pyr
Hippocampus CA1 pyramidal cells	CA1 Pyr
Hippocampus GIN (SST) interneurons	HIP GIN
Hypothalamus hypocretinergic cells	HY orexin
Locus cereuleus noradrenergic cells	LC NAdr
Midbrain serotonergic cells	MB 5HT
Neocortex corticostratial pyramidal cells	Ctx CStr Pyr
Neocortex corticothalamic pyramidal cells	Ctx CThal Pyr
Neocortex G42 (PV) interneurons, P10	Ctx G42 P10
Neocortex G42 (PV) interneurons, P15	Ctx G42 P15
Neocortex G42 (PV) interneurons, P25	Ctx G42 P25
Neocortex G42 (PV) interneurons, P7	Ctx G42 P7
Neocortex GIN (SST) interneurons	Ctx GIN
Neocortex Glt25d2-expressing pyramidal cells	Ctx Glt Pyr
Neocortex Htr3a-expressing cells	Ctx Htr3a
Neocortex layer 2–3 pyramidal cells	Ctx L2-3 Pyr
Neocortex layer 6 pyramidal cells	Ctx L6 Pyr
Neocortex Oxtr-expressing cells	Ctx Oxtr
Somatosensory cortex layer 5 pyramidal cells	SSp TT Pyr
Striatum cholinergic cells	Str ACh
Striatum Drd1-expressing medium spiny neurons	Str Drd1 MSN
Striatum Drd2-expressing medium spiny neurons	Str Drd2 MSN
Substantia nigra pars compacta dopaminergic cells	SNc DA
Ventral tegmental area dopaminergic cells	VTA DA

https://doi.org/10.1371/journal.pcbi.1005814.t001

Correlation of neuronal transcriptomics with electrophysiological properties

For each of the 34 neuron types in the NeuroExpresso/NeuroElectro discovery dataset, we obtained a gene expression profile for 11,509 genes and 5–11 intrinsic electrophysiological properties (mean = 9 +/- 2 ephys properties per cell type; described in **S1 Table**). We first asked whether there are individual genes whose quantitative mRNA expression levels correlate with systematic ephys diversity in both the discovery and AIBS validation datasets. Using the discovery dataset, after first filtering for genes with sufficiently high and variable expression across cell types (see Methods), we found a total of 653 genes (of 2694 tested) correlated with at least 1 of the 11 ephys properties at $p_{adj} < 0.05$ (p_{adj} indicates Benjamini-Hochberg false

Table 2. Descriptions for neuron types composing the Allen Institutes for Brain Sciences cell types validation dataset. Mouse line indicates cre-driver lines used to label specific populations of cells in the adult mouse visual cortex. N cells indicates number of cells assayed per cre-line via single-cell RNAseq or patch-clamp electrophysiology. Color indicates cell type color used within this manuscript.

Mouse line (cre-driver)	N cells (scRNAseq)	N cells (ephys)	Color
Ctgf	13	12	midnightblue
Cux2	122	55	olivedrab1
Gad2	69	11	thistle1
Htr3a	123	81	firebrick4
Nr5a1	48	62	blue2
Ntsr1	90	37	deepskyblue
Pvalb	88	141	firebrick2
Rbp4	173	61	mediumseagreen
Rorb	51	106	skyblue3
Scnn1a.Tg2	19	28	cyan
Scnn1a.Tg3	99	52	lightskyblue
Sst	105	107	orchid

discovery rate adjusted p-value). 1095 genes were identified at $p_{adj} < 0.1$ and 217 genes were identified at $p_{adj} < 0.01$.

As an illustrative example of one gene-ephys correlation, we found that expression levels of the gene *Nkain1* correlated with input resistance (R_{in}) values across cell types in the discovery dataset (**Fig** 1B and 1C; Spearman correlation, $r_s = 0.86$; $p_{adj} = 1.7*10^{-7}$). We also saw this trend recapitulated when only considering within-cell type changes observed across cortical basket cell and Purkinje cell development, with *Nkain1* expression and R_{in} decreasing dramatically as these cells mature (**S2 Fig**). In the AIBS validation dataset, after summarizing the single-cell data to the level of cell types, we further found a consistent *Nkain1*- R_{in} correlation amongst adult visual cortex cell types (**Fig** 1D; $r_s = 0.71$). Little is known about *Nkain1* protein function, except that it interacts with the Na^+/K^+ pump β-subunit and likely modulates the pump's function and membrane localization [30]. Intriguingly, the Na+/K+ pump has a known role in establishing cellular volumes and input resistance [31].

We provide a summary of the total number of genes identified as significantly correlated with each of the 11 ephys properties in Fig 2A and the full list of gene-ephys correlations in **S3 Table**. We initially noticed that different ephys properties were significantly correlated with varying numbers of genes. For example, at the somewhat conservative threshold of $p_{adj} < 0.05$, we found no genes correlated with action potential threshold voltage (AP_{thr}), despite there being many genes previously implicated with this feature [5,32]. In contrast, there were over 200 genes significantly correlated with either V_{rest} or AHP_{amp}. However, we consider it unlikely that all of these genes reflect a direct causal relationship, as gene-gene correlations driven by gene co-regulation create ambiguity.

We note that in the discovery dataset, not all ephys properties were available for each cell type, with 19–34 cell types quantified per ephys property. Furthermore, since correlation p-values are in part related to sample size, we found a positive relationship between the total number of genes associated with each ephys property and the number of cell types where the ephys property was quantified ($R^2 = 0.30$; **S3 Fig**). Next, given that ephys properties tend to be correlated with one another [21,25], we asked if pairs of correlated ephys properties also tend to share associated genes. For example, cellular measurements of membrane capacitance (C_m) and R_{in} are highly anti-correlated ($r_s = -0.69$ in the discovery dataset); furthermore, of the 80 genes significantly associated with C_m, 36 were also associated with R_{in}. Though some pairs of
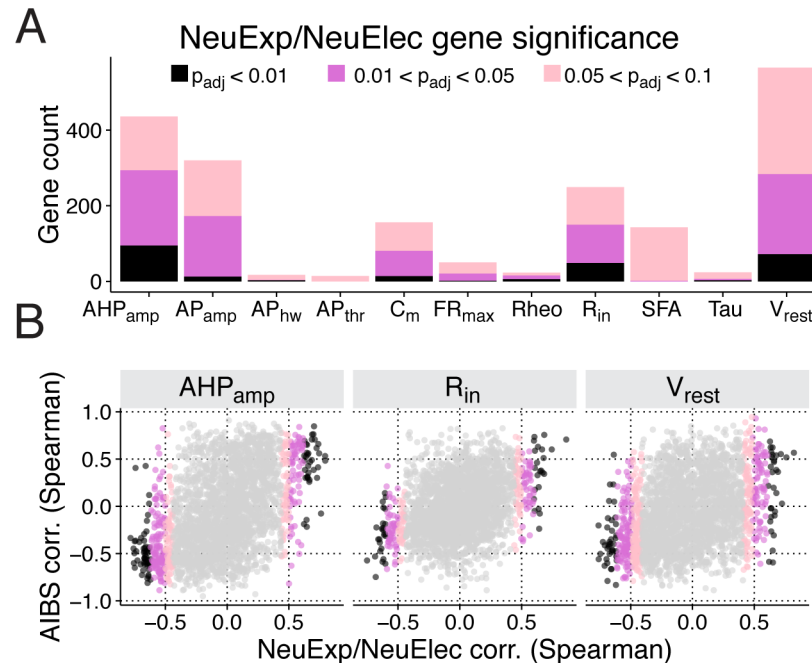
Fig 2. Identification and validation of transcriptomic—electrophysiological correlations. A) Count of genes significantly correlated with various electrophysiological properties, broken down by statistical significance of Benjamini-Hochberg FDR-adjusted correlation p-values (p_{adj}). Names and descriptions of ephys properties are provided in **S1 Table**. B) Comparison of correlations calculated using NeuroExpresso/NeuroElectro discovery dataset (NeuExp/NeuElec, x-axis) versus correlations calculated using Allen Institute validation dataset (AIBS, y-axis). Dots reflect correlation values of individual genes. Subpanels indicate correlations computed across various electrophysiological properties and p-values are provided in **Table 3**.

ephys properties share common biophysical mechanisms and could be thus regulated via common genes (e.g., C_m and R_{in} are both dependent in part on cell size), correlations between ephys properties likely limit the specificity of the relationships reported here.

We next used the AIBS dataset to validate the significant correlations observed in the discovery dataset. We predicted that gene-ephys correlations discovered in our brain-wide dataset should generalize to the transcriptomic and electrophysiological diversity among adult visual cortex cell types. Because of the limited number of cell types available in the validation dataset relative to the discovery dataset, we were generally underpowered to identify statistically significant relationships using the AIBS dataset alone for most electrophysiological properties (**S3 Table** and **S4 Table**). We therefore chose to compare results between the discovery and validation datasets as: 1) overall consistency, defined by the global rank correlation between results from the two datasets (Fig 2B); and 2) consistency for the subset of gene-ephys relationships meeting our threshold for significance in the discovery dataset ($p_{adj} < 0.05$). Overall, we found positive, but modest, agreement between the two datasets, with most ephys properties showing a positive correlation (**Table 3**). However, AP_{thr}, Rheo, and Tau are notable exceptions and might reflect challenges in normalizing these ephys features from the cross-study NeuroElectro database [25]. Focusing specifically on significant gene-ephys correlations identified in the discovery dataset, we found that the majority of these, 61.2%, reflecting 420 individual genes, were consistent in the validation dataset, with consistency defined as a matching correlation direction and with an absolute value of $r_s > 0.3$ (**Table 3**).

The degree of consistency between the NeuroExpresso/NeuroElectro and AIBS datasets is encouraging given their dissimilarity in design and content. For example, the AIBS cell types

Table 3. Consistency of gene-electrophysiological property correlations between NeuroExpresso/NeuroElectro discovery and AIBS validation datasets. Overall AIBS consistency indicates overall Spearman rank correlation between the full set of gene-electrophysiological correlations calculated in both the discovery and validation datasets, as shown in *Fig 2*B. P-values based on 1000 random reshuffles of cell type labels in the AIBS validation dataset. Discovered genes, p_{adj} < 0.05 reflects count of genes significantly correlated with each ephys property with in discovery dataset (only includes genes that are also present in AIBS scRNAseq dataset). AIBS consistency, $|r_s|$> 0.3 reflects count and percentage of discovered genes that further show a consistent relationship in the AIBS validation dataset. P-value also based on 1000 shuffled samples of cell type labels in the validation dataset.

Ephys Property	Overall AIBS consistency		Discovered genes; p_{adj} < 0.05	AIBS consistency; $|r_s|$ > 0.3		
	Spearman corr.	*p-value*	*count*	*count*	*%*	*p-value*
AHP_{amp}	0.45	0.009	285	204	72	0.005
AP_{amp}	0.404	<0.001	169	119	70	0.006
AP_{hw}	0.04	0.323	4	3	75	0.056
AP_{thr}	-0.146	0.877	0	- - -	- - -	- - -
C_m	0.384	0.037	80	55	69	0.015
FR_{max}	0.209	0.074	21	7	33	0.159
Rheo	-0.049	0.649	15	5	33	0.162
R_{in}	0.346	0.004	144	68	47	0.029
SFA	0.298	0.01	2	1	50	0.277
Tau	-0.106	0.713	6	5	83	0.007
V_{rest}	0.332	0.029	279	148	53	0.025

https://doi.org/10.1371/journal.pcbi.1005814.t003

dataset is sampled from a single brain region (visual cortex) at one developmental stage (adult). Moreover, there are considerable technical differences between the datasets, such as transcriptome quantification via single-cell RNAseq vs pooled-cell microarrays or between standardized versus heterogeneous ephys data collection.

In the remainder of the manuscript, we focus on incorporating multivariate methods and further characterizing the significant gene-ephys correlations from the discovery dataset that have evidence for further validating in the AIBS dataset.

Predicting cell type-specific electrophysiological values from gene expression

Given the relatively high correlation between the expression of single genes and specific ephys properties, we next wondered if we could construct statistical models to predict ephys parameters from gene expression patterns. Using the discovery dataset, we trained sparse, regularized statistical models to predict cell type-specific ephys values from multivariate gene expression (using a consensus set of 2603 genes with high variance in the discovery dataset that were also available in the AIBS validation dataset). Across the set of 11 ephys properties, we used leave-one-out cross-validation (LOOCV) to evaluate how well gene expression patterns can predict the ephys parameters of cell types not used for model training. For most ephys properties, such as action potential amplitude (**Fig** *3*A, R^2_{LOOCV} = 0.63) and maximum firing rate (**Fig** *3*C, R^2_{LOOCV} = 0.58), we found considerable predictive power between cell type-specific gene expression and ephys (summarized results across ephys properties shown in (**Fig** *3*E). We further noted that, qualitatively, ephys properties with relatively poor predictive performance also tended to be those with fewer genes identified as significantly correlated with that feature, such as AP_{thr} and AP_{hw} (**Table** 3).

Next, we asked if the statistical models that were originally trained on the discovery dataset could further be used to predict the ephys properties of the cell types in the AIBS validation dataset, even though technical differences would likely limit the accuracy of such cross-dataset prediction. We first applied simple normalizations to help align the RNAseq-based expression values and ephys measurements to those from the discovery dataset (see Methods). After using
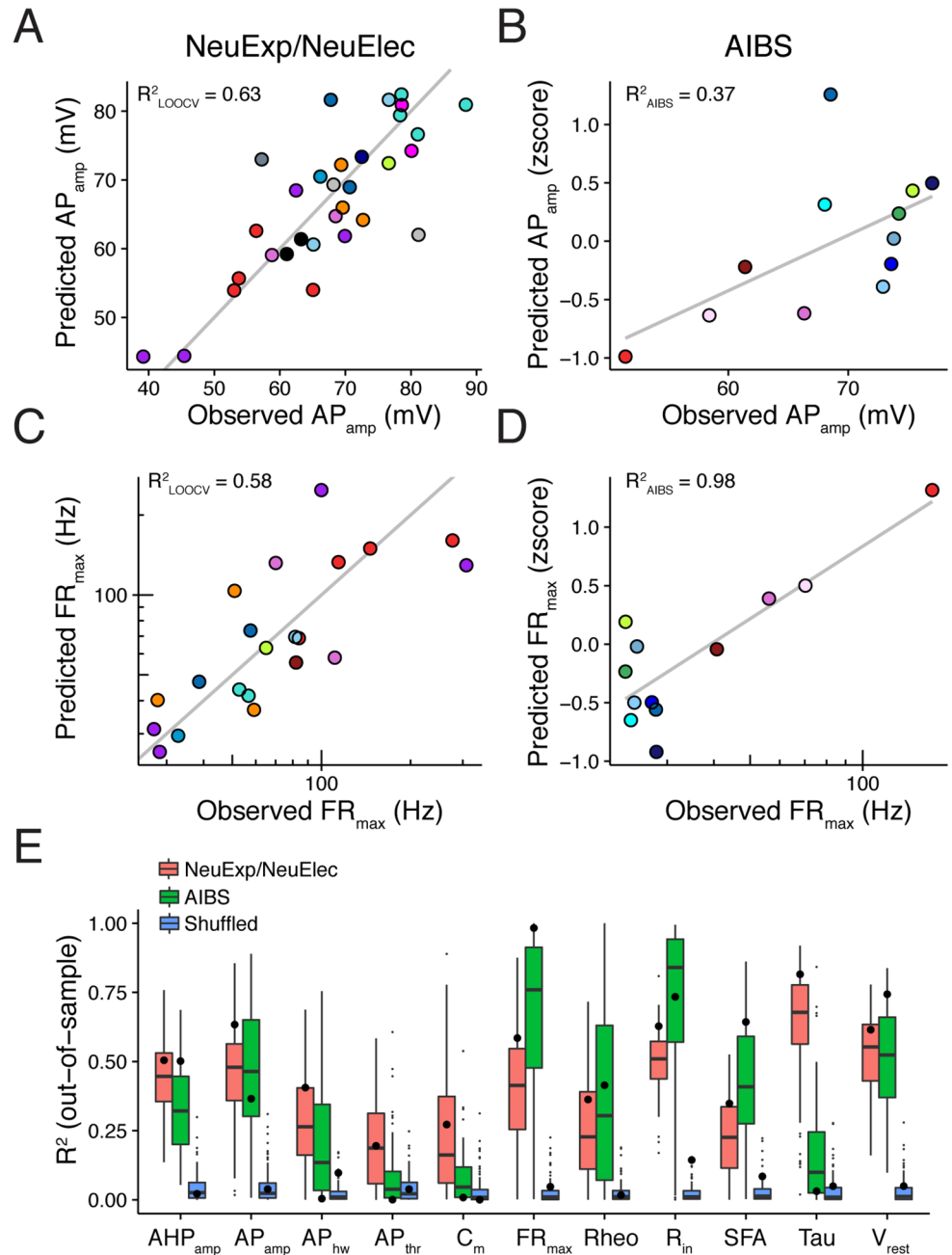
Fig 3. Multivariate gene expression can predict cell type-specific electrophysiological parameters. A) Comparison of observed action potential amplitudes (AP_{amp}; x-axis) to predicted values (y-axis) using gene expression-based statistical models trained using the NeuroExpresso/NeuroElectro discovery dataset. The y-value of each point (a cell type) is based on leave-one-out cross-validation (LOOCV). R^2_{LOOCV} indicates the calculated R^2 across the set of cell type predictions and grey line indicates the unity line. B) Same as A, but observed and predicted values are based on the AIBS validation dataset. Ephys predictions on y-axis are made by applying the discovery dataset-based models (as in A) to the AIBS-dataset multivariate gene expression profiles. R^2_{AIBS} is calculated across the set of predictions made for the AIBS cell types and grey line indicates best linear fit. C,D) Same as A and B, but for maximum firing rate (FR_{max}). E) Summarized performance of gene expression-based statistical models for predicting ephys parameters. Large dots indicate the R^2_{LOOCV} from the NeuExp/NeuElec discovery dataset (pink), R^2_{AIBS} values from the validation dataset (green), and R^2_{LOOCV} values on a version of the NeuExp/NeuElec discovery dataset where cell type labels were randomly shuffled (blue). Boxplots are based on 100 bootstrap resamples of the discovery dataset and small dots indicate boxplot outliers.

https://doi.org/10.1371/journal.pcbi.1005814.g003

the models to predict AIBS ephys values from the single cell-based gene expression patterns, we found good accuracy for some ephys properties, such as APamp (**Fig** 3B, $R^2_{AIBS} = 0.37$) and FRmax (**Fig** 3D, $R^2_{AIBS} = 0.98$). We tended to find similar generalization performance between the discovery and validation datasets for a number of ephys properties, with membrane time constant (Tau) and cellular capacitance (C_m) being notable outliers (**Fig** 3E). While individual poorly predicted ephys properties and cell types should be investigated further, these results speak to the generalizability of the gene expression-ephys relationships described here. Such findings suggest that these relationships could be used to potentially inform on cellular phenotypes when only expression data are available.

Causal relationships between discovered gene-electrophysiological correlations

A key question is whether any of the univariate gene-ephys correlations we observed are due to direct causal relationships supported by specific evidence. To this end, we made use of the existing literature on gene-ephys relations. We focused on ion channel genes (**Fig** 4A), reasoning that these would be most likely to have been directly tested for electrophysiological function. We manually searched the literature for such experiments, since at present this data is not reflected within a comprehensive database (the current NeuroElectro database reflects experiments done under standard or control conditions, not genetic or pharmacological manipulations).

We present a brief summary of our gene-centered literature search alongside highlights from our correlation-based analysis below, with the complete results provided in **S5 Table**. Of 31 significant and validated ion channel-ephys correlations, we found 17 had been directly tested through genetic manipulations or channel-specific pharmacology (reflecting 12 unique ion channel genes). To compare our correlations to individual results from direct experiments, we first mapped our correlations to predicted causal effects; for example, knocking out a gene whose expression is positively correlated with maximum firing rate should tend to lower firing rates, all else being equal. We found that of 17 total tested ion channel-ephys correlations, 11 were consistent with literature evidence, 2 showed mixed evidence, 1 showed no effect on the ephys property, and 3 were inconsistent. Here, we defined inconsistent evidence as those where a predicted increase (or decrease) in an ephys property was reflected by a change in the opposite direction in the literature; mixed evidence were those where some manipulations were consistent but others were inconsistent (e.g., pharmacology versus gene knockout). Below, we provide specific illustrative examples from this literature search.

Scn1a, encoding the sodium channel Nav1.1, was positively correlated with maximum firing rate (**Fig** 4B; NeuExp/NeuElec $r_s = 0.86$, AIBS $r_s = 0.36$), with the highest *Scn1a* expression observed in adult cortical PV interneurons and Purkinje cells. In a mouse model of Dravet syndrome with a hemizygous gene deletion (i.e., *Scn1a* +/-), it was observed that fast-spiking PV interneurons cells could no longer fire at their characteristically high frequencies (**Fig** 4C), with a smaller but significant effect also observed in Sst-expressing Martinotti cells [5]. However, the same change was not seen in layer 5 pyramidal cells, which express ~3–4 fold less *Scn1a* relative to PV cells (in NeuroExpresso and AIBS), potentially suggesting that total expression levels might mediate the effect of hemizygous *Scn1a* deletion. Intriguingly, in a haploinsufficiency model of Dravet syndrome, directly upregulating *Scn1a* expression using long non-coding RNAs rescued the firing phenotype in PV cells and lowered seizure number and duration [36].

We found 4 (of 5 total) ion channel genes correlated with V_{rest} that were consistent with literature evidence. *Hcn3*, encoding a slow HCN channel variant [6], was positively correlated
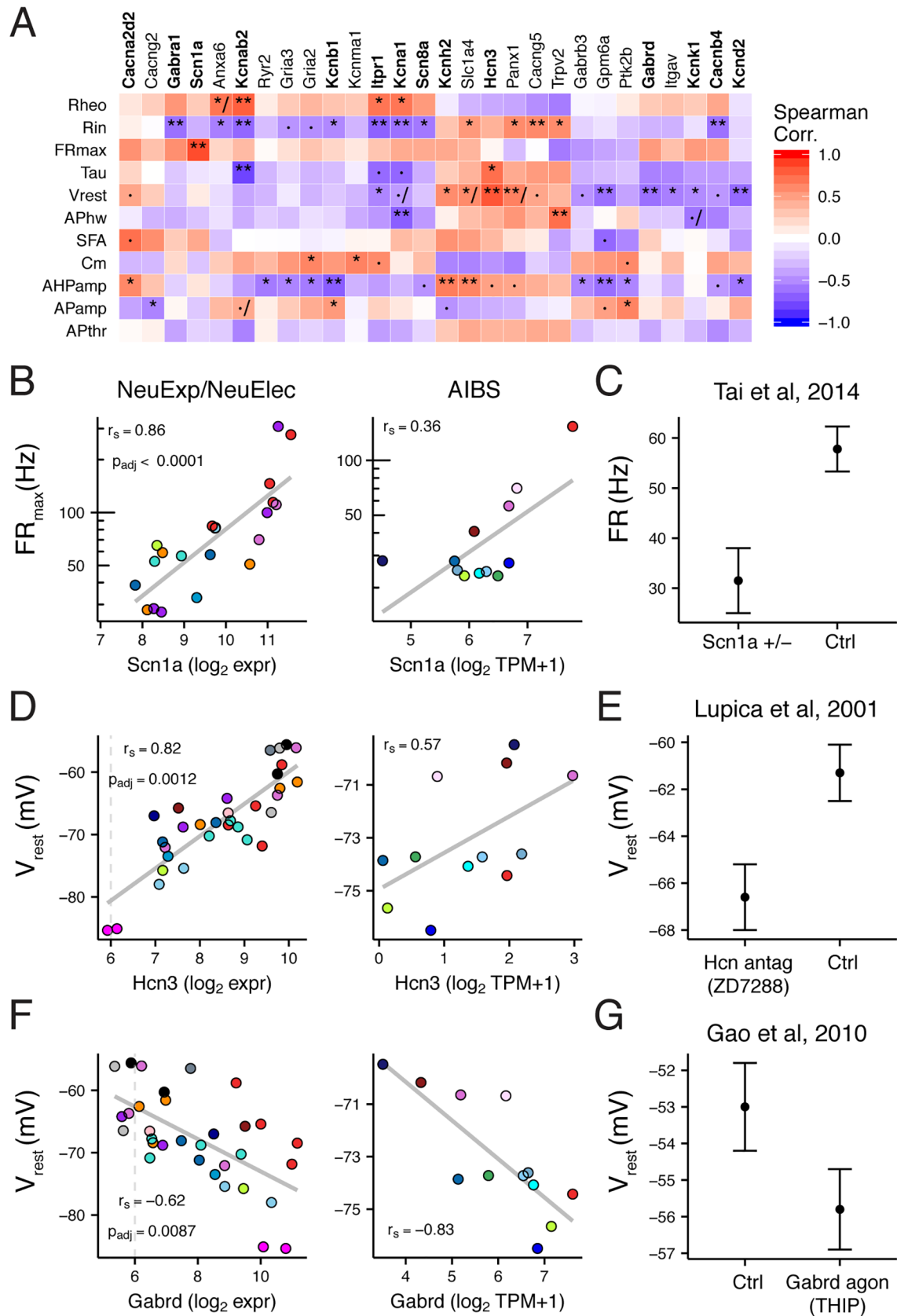
Fig 4. Ion channel specific gene-electrophysiological correlations and literature evidence for causal regulation. A) Heatmap showing NeuExp/NeuElec dataset gene-ephys correlations for ion channel genes. Genes filtered for those with at least one significant ephys correlation ($p_{adj} < 0.05$) and with validation supported in AIBS dataset. Gene names in bold indicate those we found to be previously studied for specific predicted ephys properties, based on our literature search. Symbols within heatmap: ·, $p_{adj} < 0.1$; *, $p_{adj} < 0.05$; **, $p_{adj} < 0.01$; /, indicates inconsistency between discovery and AIBS validation dataset. B) Correlation between cell type-specific *Scn1a* ($Na_v1.1$) gene expression and maximum firing rate (FR_{max}) from discovery dataset (NeuExp/NeuElec, left) and Allen Institute dataset (AIBS, right). Grey trend lines indicate linear fit. C) Replotted data from [33], showing evoked firing rates at 300 pA current injection for parvalbumin positive interneurons in control and Scn1a heterozygous mice (Scn1a +/-). Data plotted as mean +/- SEM. D) Same as B, but for *Hcn3* and resting membrane potential (V_{rest}). E) Replotted data from [34], where V_{rest} from CA1 OLM interneurons was measured before and after the application of ZD7288, a selective antagonist of HCN channels. F) Same as B, but for *Gabrd* and V_{rest}. G) Replotted data from [35], showing V_{rest} recorded from dorsal motor nucleus of vagus neurons after application of THIP, a selective agonist of *Gabrd*-mediated tonic inhibition.

with V_{rest} (**Fig** 4D; NeuExp/NeuElec $r_s = 0.82$, AIBS $r_s = 0.57$). Blocking HCN-current using ZD7288 across multiple cell types consistently made V_{rest} more hyperpolarized (**Fig** 4E) [34,37]. *Gabrd*, *Kcnk1*, and *Itpr1*, were each negatively correlated with V_{rest} and each gene reflects a different mechanistic route towards V_{rest} hyperpolarization (**Fig** 4F and **S4 Fig**). For example, *Gabrd* encodes the δ-subunit of the $GABA_A$ receptor and mediates extrasynaptic tonic inhibition, effectively turning the $GABA_A$ receptor into a chloride channel[38]. Thus, increased *Gabrd* expression, or pharmacologically increasing its activity (**Fig** 4F and 4G)[35] would tend to hyperpolarize cells through the chloride reversal potential (median E_{Cl} = -72 mV, based on reported internal and external solutions). Similarly, *Kcnk1*, encoding the $K_{2P}1.1$ 2-pore potassium channel, hyperpolarizes V_{rest} through the potassium reversal potential ($E_K \sim$ -100 mV) [39]. *Itpr1* activity releases calcium from intracellular stores and hyperpolarizes V_{rest} through calcium-activated potassium channels [40,41]. Taken together, each of these genes reflect distinct and potentially degenerate routes towards modulating cellular V_{rest}.

We found evidence for two ion channel subunits, *Kcna1* and *Kcnab2*, regulating multiple distinct electrophysiological properties (**S4 Fig**). For example, *Kcna1*, encoding the delayed rectifier potassium channel Kv1.1, was negatively correlated with action potential half width (NeuExp/NeuElec $r_s = -0.70$, AIBS $r_s = -0.52$) and positively correlated with rheobase (NeuExp/NeuElec $r_s = 0.69$, AIBS $r_s = 0.66$). These correlations were corroborated by *Kcna1* genetic knockouts or pharmacological block in auditory brainstem neurons and are consistent with known mechanistic insight about Kv1.1 function [42–44].

While the previous examples are encouraging, not all of our findings were concordant with previous literature. For example, we saw that *Kcnb1*, encoding the Kv2.1 channel, was negatively correlated with spike afterhyperpolarization amplitude (AHP_{amp}) (**S5A and S5B Fig**; NeuExp/NeuElec $r_s = -0.70$, $p_{adj} = 0.0033$; AIBS $r_s = -0.62$). Based on this correlation, we would expect that decreasing $K_v2.1$ functional expression should increase AHP_{amp} values. However, convergent genetic and pharmacological evidence suggests the opposite: decreasing $K_v2.1$ activity or expression decreases AHP_{amp} values [45,46]. Delving deeper, the *Kcnb1*-AHP_{amp} correlation appears driven in part by gross differences between excitatory and non-excitatory cell types, with excitatory cells strongly expressing Kcnb1 and also having small AHP_{amp} relative to non-excitatory cell types (**S5C Fig**). Thus though there is likely some mechanistic explanation for why excitatory cells tend to express more *Kcnb1*, this does not appear to be directly related to AHP_{amp} per-se. This example suggests that caution is needed before interpreting each correlation reported here as a direct causal relationship.

To summarize, we found multiple examples of direct regulation of specific ephys properties by individual genes identified through our correlation-based methodology. In the remainder of the results, we highlight additional genes that may be of relevance in future studies.

Further analysis of specific gene-electrophysiology correlations

Encouraged that many of the univariate ion channel gene-ephys associations discovered through our analysis were consistent with previous experimental manipulations, we next expanded our attention to other classes of genes. From the larger list of correlations identified in our analysis (**S3 Table**), we have highlighted below a small number of individual gene-ephys correlations.

Multiple genes known to regulate ion channel functional expression and localization were identified in our analysis (**Fig** 5A and 5B). For example, two genes regulating the localization of sodium channels, *L1cam* and *Fgf14*, were correlated with V_{rest} in our analysis and the direction of correlation was further supported by previous experiments [47,48]. Along this theme, our analysis identified novel associations between *Nedd4l* and *Slmap* with V_{rest}, *Ank1* with maximum firing frequency, and *Nkain1* with R_{in} (as shown in **Fig** 1). *Nedd4l*, identified as an epilepsy gene through whole-exome sequencing [14], ubiquitinates voltage-gated sodium and potassium channels [49]; *Slmap*, associated with Brugada syndrome, controls the trafficking and surface expression of voltage-gated sodium channels in cardiac and muscle cells but remains unstudied in neurons [50]. *Ank1*, a member of the ankyrin family, has recently been shown to coordinate the localization of specific Na_v subunits to nodes of Ranvier [51]. Though we found the highest expression of *Ank1* in fast-spiking cells, including Purkinje and PV interneurons, its function remains completely uncharacterized in these cells.

We noted several transcription factors in our list of associated genes, including some that have known roles in the nervous system that are compatible with possible, but unknown, roles in the regulation of cellular ephys (**Fig** 5C). For example, we found *Zbtb18* (a.k.a., RP58, Zfp238) to be negatively correlated with V_{rest}. Though *Zbtb18* has yet to be studied for its potential electrophysiological effects, this gene has been shown to be required for the normal development of neocortical glutamatergic cells [52,53] and its human homolog has recently been identified as a causative gene for autism and neurodevelopmental disorders [54]. As another example, *Zscan21* (a.k.a., Zipro1 or Zfp38) positively correlated with input resistance here and has been shown to be involved in the normal proliferation of progenitor cells into cerebellar granule cells [55].

Among genes correlated with membrane capacitance and input resistance, we noticed that many of these were cytoskeletal proteins or otherwise associated with regulating neuronal differentiation and dendritic morphology, including *Cap2*, *Chn1*, *Stmn4*, *Bex1*, and *Tpm4* (**S6 Fig**).

In summary, this analysis presents suggestive evidence for many novel gene-ephys relationships. Though we do not expect all of these novel associations to reflect direct causal relationships, by focusing on gene classes that are compatible with possible regulation of ephys, we can further hone the list of associated genes to those that might be of further interest for follow-up investigation.

Discussion

The relationship between gene expression and cellular phenotypes like electrophysiology or morphology is complex and largely unknown. Here, we have enumerated a subset of potential gene-electrophysiology relationships by identifying genes whose expression significantly correlates with specific electrophysiology parameters across a brain-wide collection of neuron types. The majority of these relationships generalized in an independent sample of visual cortex cell types and further allow the prediction of ephys features from multivariate gene expression patterns. Beyond correlation, some of these genes, such as *Scn1a*/Na_v1.1 and *Gabrd*, have been experimentally shown to be causally responsible for specific ephys properties. The majority of genes discussed here, such as *Nkain1* and *Slmap*, have yet to be investigated in the
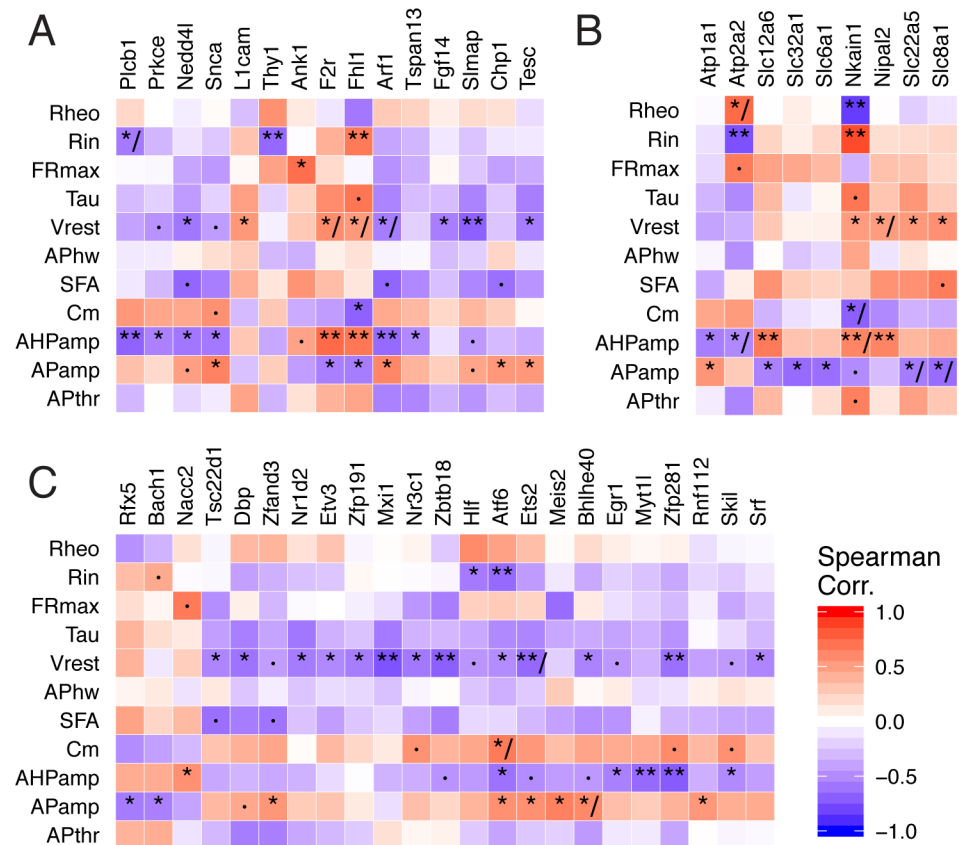
Fig 5. Summary of gene-ephys correlations for selected functional gene sets. A) Genes regulating ion channels and transporter function. B) Ion transporters. C) Transcription factors. Genes filtered for those with at least one statistically significant correlation with an ephys property ($p_{adj} < 0.05$) and validating in AIBS dataset. Symbols within heatmap: ·, $p_{adj} < 0.1$; *, $p_{adj} < 0.05$; **, $p_{adj} < 0.01$; /, indicates inconsistency between discovery and AIBS dataset.

https://doi.org/10.1371/journal.pcbi.1005814.g005

context of neuronal intrinsic electrophysiology. These genes present opportunities for further study and potential avenues for targeted manipulation of electrophysiological features.

The combined NeuroExpresso/NeuroElectro reference dataset is a first-of-its-kind resource of cell type-specific transcriptomes paired with electrophysiological profiles across a large collection of neuron types. The community resource directly reflects the efforts of hundreds of investigators to characterize the rich diversity of neuron types throughout the brain. It further reflects our considerable neuroinformatics-focused efforts in curating and standardizing this heterogeneous data [23–25]. The dataset includes cell type-specific samples from a wide range of cell types varying in sub-threshold and spiking patterns, morphologies, and developmental stages. We have made the combined dataset available here, as it could be a useful resource and benchmark for future analyses. Moreover, our cell type-based integration approach could be expanded to incorporate additional cellular phenotypes, like neuronal morphology or synaptic physiology, and newer genomic data sources including from RNA-seq, epigenomics, or proteomics [56–58].

In our framework, a causal gene-ephys relationship implies that a consistent change in a gene's expression would result in a corresponding change in an ephys phenotype, all else being equal. Based on the diversity of cell types present here, we hypothesize that these gene-ephys relationships might further be relatively independent of cell type identity. Indeed, we found examples during our literature search where the specific experiment to confirm a causal gene-

ephys relationship was performed in a cell type not present in either the discovery or AIBS datasets, including auditory and autonomic brainstem neurons (**Fig 4**, **S4 Fig**). Not only do these examples provide direct support for the gene-ephys relation, but we also infer the same causal relationship in other cell types, beyond those tested directly. Though additional experiments are needed to determine whether these relationships are truly cell type-independent, this possibility is exciting as it suggests that there could be some genes that contribute to similar ephys functions across very different cell types.

Every novel correlation reported here presents a specific, testable causal prediction. The results from our ion channel-focused literature search are encouraging, as 13 of 17 tested gene-ephys relationships showed some evidence for direct experimental support. However, it is overly optimistic to conclude that most novel ephys-correlated genes reported here will prove causal. Instead, we advocate further in-depth analysis of gene function when prioritizing individual genes for future experiments. For example, the correlation between *Nkain1* and input resistance (R_{in}) is plausibly causal because the *Nkain1* protein interacts with the Na^+/K^+ pump complex [30] and the pump's activity regulates R_{in} through helping maintain cellular volumes [31]. Similarly, the correlation between *Ank1* and FR_{max} is intriguing because *Ank1*, an isoform of the autism gene *Ank3*, helps coordinate the localization of Na_v subunits to the nodes of Ranvier [51]. Though we found *Ank1* to be highly expressed in adult PV and Purkinje cells here, its function in these cells has yet to be characterized. Specific transcription factors identified might regulate the expression of downstream genes relevant to ephys. For example, *Zbtb18*, correlated with resting potential here, is required for normal glutamatergic cell development and has recently been implicated in human neurodevelopmental disorders through genome sequencing [52–54]. Ultimately, these genes could provide novel means for manipulating cellular ephys in the context of disease. For example, upregulating *Scn1a* expression using anti-sense RNA approaches has been shown to be an effective means of reducing seizures in a model of Dravet syndrome [36].

Limitations and caveats

The results presented here are restricted to a limited range of situations. First, we can only identify genes where mRNA, as measured in dissociated cells [59], is an adequate readout of a gene's functional activity at the protein level. Future datasets employing RNA-seq, proteomics, or techniques to capture non-somatic mRNA will likely be able to identify more genes where alternative splicing and post-translational modifications are essential for understanding gene function [10–12].

Second, the univariate approach that forms the majority of our study assumes a gene's contribution to electrophysiology is similar and monotonic across cell types. This single-gene focused analysis likely misses genes that contribute to complex ephys features in ways that are biologically degenerate and are highly non-linear or combinatorial [28,29]. For example, K_v3-family ion channels, including *Kcnc1*/K_v3.1, have been implicated in helping fast-spiking cells maintain narrow spike widths [32,60], but we did not identify *Kcnc1* as correlated with AP width in our analysis. Further utilizing multivariate approaches (like shown in **Fig 3**) and incorporating other information sources, such as how proteins interact to form functional complexes, might reveal additional signals and help mitigate spurious correlations. However, pursuing such approaches will likely necessitate larger datasets than are currently available.

Third, the focus of our analysis is to explain how ephys differences across cell types emerge through gene expression. It remains an open question whether the same genes driving large *across* cell type differences would also be the same genes that are defining subtler *within* cell type differences, like amongst olfactory bulb mitral cells or CA1 pyramidal cells [1,2,58]. As

the patch-seq methodology, enabling transcriptomic and ephys characterization from the same single-cell [19,20], is further developed and applied, we eagerly anticipate testing these hypotheses. However, small changes in expression of individual genes, as expected within a single cell type, are difficult to reliably detect using current technologies, in part, due to relatively limited sample sizes and technical challenges like "dropouts" [18]. Indeed, while these patch-seq studies have demonstrated their utility in classifying individual cells into types [19,20], how variance in expression of specific genes gives rise to within cell type ephys differences remains largely unaddressed.

Fourth, ephys property correlations and gene co-expression limits the potential specificity of any causal prediction made here. For example, some pairs of ephys properties, like AHP_{amp} and R_{in}, are correlated but probably do not share common biophysical underpinnings (**S3B Fig**). Because of this common correlation, genes significantly associated with one ephys feature are more likely to be also associated with other ephys features, potentially spuriously. Similarly, many pairs of genes show correlated expression across samples (i.e., gene co-expression). Gene co-expression often reflects biologically meaningful signals, such as co-regulation by common transcription factors or shared membership in biological pathways and cellular compartments [61]. However, co-expression makes interpreting individual gene-ephys associations difficult and likely contributes to why we found many more genes for some ephys properties than we would naively expect, such as V_{rest} and AHP_{amp}. Future analysis approaches that explicitly consider co-expression might prove useful [62].

Lastly, the heterogeneous nature of the compiled NeuroExpresso/NeuroElectro dataset [23,25,59] might limit our power to see possible biologically relevant signals and could explain our failure to find genes for some ephys features. For example, because data in NeuroElectro are compiled from different studies collected in the absence of standards for how some ephys properties are defined [24,63], this likely limits our downstream attempts at normalization. Similarly, the cell types reflected in the aggregated dataset are likely composed of multiple transcriptomic or morphologically-defined subtypes [27,64]. However, the overall consistency with the AIBS Cell Types dataset, where data were collected using standardized conditions and protocols, suggests that the results shown here are not entirely the result of technical artefacts due to data compilation.

Future directions

Our findings suggest a number of directions for future study. Can specific gene-ephys relationships be used as biomarkers to detect electrophysiological changes in a disease or treatment context? For example, if *Scn1a*/Na_v1.1 is upregulated in a cell type, does that serve as a reliable indicator of hyper-excitability? Given the relative ease and growing popularity of single-cell transcriptomics on dissociated cells and nuclei [18,27], could the multivariate gene expression-based statistical models we developed be useful in imputing ephys phenotypes from transcriptomic signatures alone? Lastly, are the gene-ephys correlations reported here predictive of cell-to-cell variability reported within the same cell type?

In summary, our results suggest that large-scale transcriptomics can prove useful in helping elucidate the biophysical basis for the rich electrophysiological diversity seen amongst neuron types throughout the brain.

Methods

NeuroExpresso database description

To obtain neuron type-specific transcriptomic data, we made use of the NeuroExpresso database (neuroexpresso.org), described previously [23]. Briefly, the database contains

transcriptomic studies collected from mouse brain cell types sampled under normal conditions. We specifically utilized the microarray-specific subset of NeuroExpresso. These samples were collected using purified, pooled-cell microarrays with transcriptomes quantified using the Affymetrix Mouse Expression 430A Array (GPL339) or Mouse Genome 430 2.0 Array (GPL1261). We further only used probesets that were shared between both platforms. Transcriptomic samples were quality controlled and manually curated for cell type identity and basic sample metadata, including animal age, array platform, and purification method. Transcriptomic samples are from adult mice unless explicitly mentioned. The samples were subjected to RMA normalization and an additional round of quantile normalization in order to obtain a uniform distribution of signals across samples. When a single gene was represented by multiple probesets, the probeset with highest variability across samples was chosen to represent the gene. We note that we have re-annotated the cell type labels used here from those used in the NeuroExpresso database and web resource.

For the purpose of obtaining a large corpus of cell types, we made use of a small number of cell type-specific transcriptomic samples excluded from analysis in the original NeuroExpresso publication (e.g., developmentally immature samples). Specifically, for two major cell types with transcriptomic data collected at varying ages, cortical parvalbumin-positive (PV) interneurons labelled by the G42 mouse line and cerebellar Purkinje cells [22,65], we kept samples collected at different ages separate and used of samples collected from animals aged less than P14. We further included data representing cortical *Htr3a-* and *Oxtr*-expressing cells from Gene Expression Omnibus (GEO) accession GSE56996 [66] and layer 2–3 and layer 6 pyramidal cells from GSE69340 [67]. The complete listing of transcriptomic samples, annotated cell types, and references is provided in **S2 Table**.

Gene filtering and sample summarization

Following data compilation, we filtered genes to retain only those with 1) high mean expression; and 2) highly variable expression across cell types in the combined dataset. Specifically, for each gene, g, we calculated its expression mean, μ_g, and standard deviation, σ_g, across the collection of 34 cell types in the combined discovery dataset. Next, we calculated a global mean, μ_{global} defined as $mean(\mu_{g1:gN})$, and standard deviation, σ_{global} defined as $mean(\sigma_{g1:gN})$ across the total set of genes. Here, $\mu_{global} = 7.5$ and $\sigma_{global} = 0.75$; for context, background expression levels were approximately ~6.0 ($\log_2$ expression units). We filtered genes where $\mu_g > \mu_{global}$ and $\sigma_g > \sigma_{global}$, leaving 2694 from 11667 total genes quantified. Lastly, we summarized each cell type by the mean expression per gene across samples.

NeuroElectro database description and normalization

To obtain neuron type-specific electrophysiological measurements, we used an updated version of the NeuroElectro database (neuroelectro.org), originally described in [24,25]. Briefly, we populate the NeuroElectro database using manual curation to extract information on electrophysiological measurements such as resting membrane potential and input resistance (described in **S1 Table**) from the results sections of published papers using intracellular electrophysiology. These ephys features were chosen because they were frequently reported across articles and were calculated using relatively consistent criteria from article to article. Curators also annotate a set of relevant methodological information, including species, animal age, electrode type, preparation type, recording temperature, and use of liquid junction potential correction.

NeuroElectro database. We note the following major improvements to the NeuroElectro database, beyond an increase in the overall database size (from 331 to 968 articles as of December 2016).

First, we have now curated and manually standardized a greater number of electrophysiological properties, including after hyperpolarization amplitude (AHP$_{amp}$), maximum spiking frequency (FRmax), and spike frequency adaptation (SFA). For example, in the process of data curation we have standardized electrophysiological properties for the use of different baselines, for example, AHP amplitude reported as an absolute voltage as opposed to amplitude relative to spike threshold (e.g., -70 mV vs 10 mV). We note that because of raw data unavailability, we do not recalculate measurements in NeuroElectro from raw ephys traces. Thus, we could not ensure that ephys properties such as SFA or AHP$_{amp}$ were calculated using a consistent stimulation protocol across different studies. These differences where present would tend to contribute to study-to-study variability.

Second, when curating specific neuron subtypes reported in the literature, we now take care to manually annotate the specific features the authors used to define each cell subtype (e.g., the mouse line used, brain region, gene or protein expression, firing pattern, etc.); for example, "barrel cortex layer 2–3 somatostatin-expressing interneuron from the GIN mouse line" or "hypothalamus orexin-expressing cell". This level of fine-grained cell type curation allows us to better harmonize relevant electrophysiological to transcriptomic datasets post hoc.

NeuroElectro data preprocessing. Electrophysiological data was filtered for: 1) recordings from acute brain slices *in vitro* (thus removing *in vivo* recordings and from slice and cell cultures); *2)* from mice, rats, or guinea pigs; 3) with an animal age greater than 2 days old. Animal ages, when reported as a range (e.g., P14-P20), were summarized using the geometric mean. When animal age or recording temperature was not reported, we used median imputation to fill in missing values (which typically was rare). To address the correction of liquid junction potential (LJP), we manually removed or "uncorrected" the correction of LJP when it had previously been performed and when the original authors provided the explicit voltage correction value used (i.e., LJP offset). We then used a custom LJP metadata field denoted 'PostCorrected' to define these cases.

Experimental condition-based data normalization. Building on the approach described previously, we used statistical regression models to normalize ephys data for study-to-study differences in experimental methodologies [25]. Here, we used elastic-net penalized regression, implemented using the *cv.glmnet* function within the R *glmnet* package [68] with an alpha value of .99 and nlambda = 100. The regression model for each ephys parameter (EphysProp) was fit using the following formula:

$$\text{EphysProp} = \text{NeuronType} + \text{Species} + \text{JxnPotential} + \text{ElectrodeType} + \text{bs}(\log_{10}(\text{AnimalAge})) + \text{bs}(\text{RecTemp})$$

where bs indicates the use of bsplines with 5 degrees of freedom. Here, NeuronType, Species, JxnPotential, and ElectrodeType each indicate nominal metadata types. AnimalAge and RecTemp refer to animal age and slice recording temperature and reflect continuous parameters. For example, ElectrodeType indicates the use of patch-clamp, perforated patch, or sharp electrodes whereas JxnPotential indicates whether the liquid junction potential was explicitly corrected, not corrected, or unmentioned within the article's methods section. The ephys properties, R$_{in}$, Tau, AP$_{hw}$, C$_m$, Rheo, FR$_{max}$, were log$_{10}$-transformed prior to metadata modeling.

We used the filtered NeuroElectro dataset to fit regression models to model study-to-study variability in ephys measurements. After fitting these models, we then used the models to adjust ephys data for the influence of major differences in experimental conditions between studies.

To summarize electrophysiological measurements per each unique cell type, we first took the mean of measurements reported within a single paper and then calculated the median ephys value across the multiple papers characterizing each cell type.

Harmonizing cell types across NeuroExpresso and NeuroElectro

Because it was uncommon for a single study to characterize both a cell type's transcriptomic and electrophysiological parameters, we developed a neuroinformatics-based strategy for pairing gene expression and ephys datasets from different studies based on common cell type identity.

We first manually re-annotated the cell type identity of each transcriptomic sample from NeuroExpresso using a descriptive semantic label (shown in **S2 Table**), defined by a minimally sufficient number of defining features (including brain region and marker gene expression or projection pattern [69]). For example, the transcriptomic samples corresponding to cerebellar granule cells in NeuroExpresso were purified using the L10a-Neurod1 mouse line, where GFP is specifically expressed in the ribosomes of these cells [70]. Here, we merely annotated these samples using the label, "cerebellar granule cells" (CB gran). We next identified all curated electrophysiological data within NeuroElectro corresponding to this same major cell type, making use of the manual annotations for each electrophysiological sample's cell type identity (n = 9 articles for CB granule cells). We note that subtle differences between how CB granule cells are labelled in the L10a-Neurod1 mouse line and how CB granule cells are targeted by lamina and morphology for ephys recordings would tend not to be preserved after this data harmonization step. Lastly, we note that these cell types reflect broad cellular classes and likely encompass multiple morpho-electric or transcriptomic subtypes [27,64].

To pair transcriptomic to ephys datasets explicitly defined by different ages (e.g., P7 and P25), we matched animal ages +/- 2.5 days. For example, the samples corresponding to "Ctx G42 P15" reflect neocortical parvalbumin-positive interneurons labeled by GFP in the G42 mouse line aged P15 +/- 2.5 days. Because we tended to have fewer data points after subsetting the cortical G42 cells into different age groups, for one ephys property, AP_{thr}, we excluded AP_{thr} values from these cells since they varied widely (~10mV) across studies from the same time point.

Allen Institute for Brain Sciences cell types dataset

Single cell transcriptomic samples.   We made use of an Allen Institute for Brain Sciences (AIBS) Cell Types dataset employing single-cell RNAseq to characterize diversity of cells in adult mouse visual cortex labelled by different mouse cre-lines. Specifically, we obtained data originally reported in [27] from GSE71585, representing data from 1809 single-cells. We made use of the summary data file where expression for each gene was summarized as reads per kilobase sequenced per million (TPM) with 24,057 genes quantified per cell.

Single cell electrophysiological samples.   We made use of the AIBS Cell Types dataset employing *in vitro* patch clamp electrophysiology to characterize mouse visual cortex cellular intrinsic electrophysiology using standardized protocols. For each cell in the AIBS Cell Types database (http://celltypes.brain-map.org/), representing 847 single cells as of December 2016, we downloaded its corresponding raw and summarized ephys data (summary measurements included input resistance and resting potential). For all spiking measurements except maximum firing rate and spike frequency adaptation, we used the voltage trace corresponding to the first spike at rheobase stimulation level. For a few ephys properties, like action potential half width, we calculated these from the raw ephys traces, as these were not available in the pre-calculated summarized data. Membrane capacitance was defined as the ratio of the

membrane time constant to the membrane input resistance. Maximum firing rate and spike frequency adaptation were calculated using the voltage trace corresponding to the current injection eliciting the greatest number of spikes. Spike frequency adaptation (SFA) was defined as the ratio between the first and mean inter-spike intervals during this maximum spike-eliciting trace (i.e., neurons with greater SFA will show values closer to 0).

Data summarization and harmonization. We summarized single cell transcriptomic and ephys data to the level of cell types by averaging measurements within the same cre-line (i.e., defining cell types by unique cre-lines). We filtered cre-lines that were sampled by at least 10 cells in each of the transcriptomic and ephys data, leaving a total of 12 cell types / cre-lines. We also filtered single cell transcriptomic samples to include only those corresponding to neuronal cells (i.e., removing glial cells erroneously labelled by the cre-line). We did not further attempt to make use of the novel transcriptomics-based cellular subtypes as defined in [27], since we cannot make a correspondence between these subtypes (defined on the basis of multivariate gene expression in the absence of ephys or morphological characterization) with individual cells sampled in the ephys data. We matched genes across the AIBS and NeuroExpresso/NeuroElectro datasets using NCBI entrez gene identifiers. Of the total 2694 genes present in the discovery dataset after expression level-based filtering, there were 2603 total genes in common with the AIBS scRNAseq dataset.

Data availability

The harmonized and processed cell type-specific data for the discovery and validation datasets has been made publically available at http://hdl.handle.net/11272/10485.

Statistical analysis and methodology

Gene-electrophysiological property correlation analysis. For each gene in the filtered NeuroExpresso/NeuroElectro data matrix, we calculated its Spearman rank correlation and uncorrected p-value (two-sided test) with each the 11 ephys properties, using the function *cor.test* from the R *stats* package, with 'method = "spearman"'. We also calculated the Spearman correlation (r_s) for each gene and ephys property in the AIBS validation dataset. We chose to use the Spearman correlation here to mitigate the impact of outliers and the undue influence of genes highly expressed in one or a small number of cell types.

Corrections for multiple comparisons. We used the Benjamini-Hochberg correction for False Discovery Rate (FDR) to correct for comparisons performed across multiple genes[71], implemented using the function *p.adjust* from the R *stats* package. Here, for ease of interpretation, we refer to the Benjamini-Hochberg FDR as p_{adj}. Because of ephys property correlations, we did not further correct for multiple comparisons across ephys properties.

Comparing results across discovery and validation datasets. To evaluate the consistency between discovery and validation datasets, we defined two separate measures. First, to obtain a measure of the overall consistency per ephys property, we calculated the rank correlation across the set of 2603 genes in common to both datasets (after filtering genes for expression levels based on the discovery dataset). Second, to specifically focus on gene-ephys correlations meeting our threshold for significance in the discovery dataset ($p_{adj} < 0.05$), we defined consistent correlations as those with matching correlation directions and also with the absolute value of the gene-ephys rank correlation in the validation dataset exceeding 0.3 (i.e., $|r_{s, validation}| > 0.3$). For both criteria, we obtained p-values through randomly shuffling cell type labels in the validation dataset between ephys and gene expression data. We obtained an expected p-value null distribution through performing 1000 random shuffles and recalculating gene-ephys correlations per shuffle. Our final list of gene-ephys correlations are those that are significant in

the discovery dataset (i.e., $p_{adj, discovery} < 0.05$) that further validated in the AIBS dataset ($|r_{s, validation}| > 0.3$).

Modeling ephys properties using multivariate gene expression. We trained statistical models to model the relationship between each ephys property and multivariate patterns of gene expression. We first normalized the gene expression values from the discovery dataset using z-score normalization and $\log_{10}$-transformed the ephys properties R_{in}, Tau, AP_{hw}, C_m, Rheo, FR_{max}, prior to model training. We used elastic-net penalized regression to model univariate ephys properties as a function of the expression of multiple genes (using the complete set of 2603 genes as input). Penalized regression was implemented using the *cv.glmnet* function within the R *glmnet* package [68] with an alpha value of 0.99 and nlambda = 100 (identical to how we modeled ephys properties as a function of experimental condition parameters). Following the approach outlined in [19], models were fit in two stages, where the first stage was used to decide the optimal amount of regularization (using nested cross-validation to decide the L1 regularization parameter lambda with the lowest prediction error) and which set of genes to use for prediction. In the next stage, we refit the model using only this set of selected genes. To evaluate model accuracy in the discovery dataset, we used leave-one-out cross-validation (LOOCV), where each cell type was iteratively left out and then predicted using a model constructed without that cell type. We evaluated model accuracy by calculating the R^2_{LOOCV} using the set of ephys values from all predicted cell types. As an explicit null-comparison, we repeated these steps on a version of the discovery dataset where cell type labels had been shuffled randomly between the ephys and expression data. In addition, for the purpose of obtaining variance estimates, we further used bootstrap resampling where we randomly sampled with replacement from the underlying NeuroElectro and NeuroExpresso datasets before constructing the final combined cell types dataset used for model training. We implemented this bootstrapping procedure to ensure that the full set of 34 cell types were present prior to model training. Lastly, we fit a final model for each ephys property that uses the full set of cell types in the discovery dataset.

To apply the statistical models originally trained on the discovery dataset to the AIBS validation dataset, we first $\log_2$-transformed the AIBS cell type-summarized expression data (quantified as TPM+1) and subsequently normalized these to z-scores, putting them on a similar scale to the discovery dataset-based expression data. Similarly, because ephys data from the discovery and AIBS datasets were collected and normalized using different methods, we $\log_{10}$-transformed R_{in}, Tau, AP_{hw}, C_m, Rheo, FR_{max}, and next z-score transformed all ephys properties to help reconcile some of these methodological discrepancies. After these normalization steps, we predicted cell type-specific ephys values using the discovery dataset-based models and normalized expression values from the AIBS dataset. We evaluated generalization accuracy by calculating the R^2 value across this set of predicted ephys values (termed R^2_{AIBS}).

Gene lists

To obtain specific gene sets, we made use of Gene Ontology annotations (as of August 2016). We used the GO term 0005216 corresponding to "ion channel activity" to identify ion channels; the term 0015075 corresponding to "ion transmembrane transporter activity" in addition to *Nkain1* to identify ion transporters; the term 0007010 corresponding to "cytoskeleton organization" to identify cytoskeletal genes; the term 0007399 corresponding to "nervous system development" to identify developmental genes; and the term 0034765 to identify "regulation of ion transport" in addition to the genes *L1cam*, *Slmap*, and *Ank1*. To obtain a comprehensive manually curated listing of transcription factors, we used the Transcription Factor Checkpoint resource [72].

Ion channel focused literature search

Literature search methodology. We performed a systematic literature search to identify causal experiments consistent or inconsistent with the individual gene-ephys correlations reported here. Specifically, we started with a set of 23 ion channel genes identified by our analysis (defined by GO term 0005216) that further validated in the AIBS dataset.

For each gene, we manually searched for articles where these genes had been perturbed, either using genetic approaches to knockout or knockdown the gene's expression or using channel-specific pharmacology. When searching for individual genes, we made use of common gene name synonyms, for example, that $K_v1.1$ is a synonym for the gene *Kcna1*. We further searched for papers where the individual ephys properties suggested by our correlative analysis (e.g., AP_{hw}, rheobase) had been explicitly measured. To this end, we used Google Scholar with the gene name or gene name synonym and the associated ephys property as search terms. When the name of a pharmacological blocker of an ion channel was known it was included in search terms. We also checked the top 40 papers related to a gene on its NCBI Gene page for those in which the gene was manipulated and ephys properties of interest were measured. For some widely studied ion channel genes, such as *Kcna1*/$K_v1.1$ and *Kcnd2*/$K_v4.2$, we did not attempt to systematically review each article studying these genes and typically ended our search after 3–5 relevant articles were identified. We further limited our assessment to perturbations involving mammalian neurons.

When our search yielded pertinent articles, we annotated relevant information, including: the kind of manipulation (e.g., genetic manipulation and type; pharmacological compound used, etc.); cell type; and direction and magnitude of effect. Quantitative values from each group comparison were extracted manually from either the article text or digitized from **Fig**s. To categorize effects, we assessed whether the perturbation resulted in an increase or decrease in the value of the ephys property and whether this change was further either statistically significant or non-significant. In a small number of cases, there was effectively no change or a negligible change between the control and perturbed condition that were curated as "negligible changes".

When scoring whether an individual gene-ephys correlation was either consistent or inconsistent with literature evidence, we assessed the direction effect. For example, for an ion channel gene that our analysis found as positively correlated with V_{rest}, we would expect that knocking out the gene would make V_{rest} to become more negative and more hyperpolarized, all else being equal. Similarly, applying an agonist of the ion channel should make V_{rest} more positive and depolarized. In cases with multiple lines of evidence linking specific ion channel perturbations to ephys changes (e.g., both pharmacological and genetic changes), we aggregated these along the following categories: consistent, inconsistent, mixed, and no effect. Gene-ephys correlations supported by both consistent and inconsistent literature evidence were marked as "mixed". Those with consistent evidence and also some evidence for a negligible change but no inconsistent evidence were marked as "consistent", and similarly for inconsistent evidence.

Supporting information

S1 Fig. Cartoon of data collection, curation, and normalization. *Top row*: Schematic of construction of NeuroExpresso database. As originally described in [23], following characterization and public depositing of cell-type specific expression datasets, raw transcriptomics datasets were obtained and QCed before being quantile normalized and summarized at the level of individual cell types by gene expression. *Bottom row*: Schematic of construction of NeuroElectro database. As originally described in [24,25], following characterization and

publication of neuron-type specific electrophysiological summary data, data were systematically curated and normalized for methodological differences before summarization at the level of cell types and electrophysiological properties.
(EPS)

S2 Fig. Example of cell type-specific transcriptomic and electrophysiological changes across development. A) Gene expression levels of *Nkain1* across development of cortical G42 parvalbumin-expressing interneurons. Dots reflect unique transcriptomic samples. B) Same as A, but for cerebellar Purkinje cells. C) Values of input resistance sampled from cortical G42 parvalbumin-expressing interneurons at various points in development. Individual dots reflect population means from individual articles represented in the NeuroElectro database and lines are based on loess smoothing. D) Same as C, but for cerebellar Purkinje cells.
(EPS)

S3 Fig. Factors affecting numbers of genes identified as significantly correlated with different electrophysiological properties. A) Scatterplot illustrating the relationship between the numbers of genes identified as significantly correlated with each ephys property ($p_{adj} < 0.05$) versus the number of cell types with ephys data in the NeuroExpresso/NeuroElectro dataset. B) Pairwise correlations between electrophysiological properties, based on cell types in combined NeuroExpresso/NeuroElectro sample. Heatmap colors indicate the absolute value of measured Spearman correlations between ephys property pairs. Inset values indicate the number of significant genes shared between each pair of ephys properties ($p_{adj} < 0.05$). Numbers in parentheses on y-axis and values along diagonal indicate number of significant genes identified for each ephys property (i.e., as in y-axis in A).
(EPS)

S4 Fig. Further evidence for causal regulation of specific gene-ephys correlations. A) Correlation between cell type-specific *Kcnk1* ($K_2P1.1$/TWIK1) gene expression and resting membrane potential (V_{rest}) from discovery dataset (NeuExp/NeuElec, left) and Allen Institute dataset (AIBS, right). B) Replotted data from [39], showing effects of siRNA-induced knockdown of *Kcnk1* expression in dentate gyrus granule cells. C, E, I, G, K) Same as A but shown for specific ephys properties and genes. D) Replotted data from [40], showing effects of antagonizing *Itpr1* function through the use of 2-APB. F, H) Replotted data from [42], showing effects of knocking out *Kcna1* ($K_v1.1$) on action potential half width (AP_{hw}) and rheobase (Rheo) as measured in auditory brainstem neurons. J, L) Replotted data from [44], showing effects of knocking out *Kcnab2* ($K_v$beta2) on rheobase and input resistance (R_{in}) as measured in lateral amygdala pyramidal neurons.
(EPS)

S5 Fig. Specific evidence for gene-electrophysiology correlation not implying causation. A) Correlation between cell type-specific *Kcnb1* ($K_v2.1$) gene expression and action potential after-hyperpolarization amplitude (AHP_{amp}) from discovery dataset (NeuExp/NeuElec, left) and Allen Institute dataset (AIBS, right). B) Replotted data from [46], showing measured AHP_{amp} values from entorhinal cortex pyramidal neurons during control and under perfusion of Guangxitoxin-1E, a specific blocker of K_v2-family currents. Data illustrates that effect of $K_v2.1$ blockade results in increased AHP_{amp}, the opposite of expected result based on correlations shown in A. C) Same data shown in A, but broken down by major cell types, illustrating that *Kcnb1*-AHP_{amp} correlation is in part related to major differences in *Kcnb1* expression and AHP_{amp} values between excitatory glutamatergic and non-excitatory cell types.
(EPS)

S6 Fig. Summary of gene-ephys correlations for additional functional gene sets. Top: Nervous system development genes. Bottom: Cytoskeletal organization genes. Genes filtered for those with at least one statistically significant correlation with an ephys property ($p_{adj} < 0.05$) and validating in AIBS dataset. Symbols within heatmap: ·, $p_{adj} < 0.1$; *, $p_{adj} < 0.05$; **, $p_{adj} < 0.01$; /, indicates inconsistency between discovery and AIBS dataset.
(EPS)

S1 Table. Description of electrophysiological properties used in this study.
(CSV)

S2 Table. Description of cell types composing the combined NeuroExpresso/NeuroElectro dataset.
(CSV)

S3 Table. List of significant gene-electrophysiological correlations. Column headers are as follows: EphysProp refers to the electrophysiology property, GeneSymbol, GeneName, GeneEntrezID all refer to information about the gene tested and DiscProbeID indicates the Affymetrix probe ID used in the discovery dataset. DiscCorr refers to the gene-ephys Spearman correlation calculated in the NeuroExpresso/NeuroElectro discovery dataset and DiscFDR and DiscUncorrPval refers to the Benjamini-Hochberg FDR and uncorrected p-value based on this correlation. AIBSCorr, AIBSUncorrPval, and AIBSFDR refer to the gene-ephys rank correlation, uncorrected p-value, and Benjamini-Hochberg FDR calculated in the AIBS replication sample. AIBSMeanExpr (log2 TPM+1) indicates the mean expression values in the AIBS dataset. AIBSConsistent refers to consistency of correlation direction between the discovery and replication datasets with an absolute value of $r_s > 0.3$ in the AIBS dataset.
(CSV)

S4 Table. Summarized counts of gene-ephys significance in discovery and AIBS datasets. Counts of genes significantly associated with individual electrophysiological properties at various statistical thresholds (indicated by FDR) for Discovery and AIBS datasets and the count of genes in common between these (Overlap).
(XLSX)

S5 Table. Complete dataset of literature search for ion channels predicted to be significantly correlated with electrophysiological diversity.
(XLSX)

Acknowledgments

Author Contributions

Conceptualization: Shreejoy J. Tripathy, Paul Pavlidis.

Data curation: Shreejoy J. Tripathy, Lilah Toker, Brenna Li, Cindy-Lee Crichlow, Dmitry Tebaykin, B. Ogan Mancarci.

References

1. Padmanabhan K, Urban NN. Intrinsic biophysical diversity decorrelates neuronal firing while increasing information content. Nat Neurosci. 2010; 13: 1276–1282. https://doi.org/10.1038/nn.2630 PMID: 20802489

2. Tripathy SJ, Padmanabhan K, Gerkin RC, Urban NN. Intermediate intrinsic diversity enhances neural population coding. Proc Natl Acad Sci. 2013; 110: 8248–8253. https://doi.org/10.1073/pnas.1221214110 PMID: 23630284

3. Bernard G, Shevell MI. Channelopathies: A Review. Pediatr Neurol. 2008; 38: 73–85. https://doi.org/10.1016/j.pediatrneurol.2007.09.007 PMID: 18206787

4. Klassen T, Davis C, Goldman A, Burgess D, Chen T, Wheeler D, et al. Exome sequencing of ion channel genes reveals complex profiles confounding personal risk assessment in epilepsy. Cell. 2011; 145: 1036–1048. https://doi.org/10.1016/j.cell.2011.05.025 PMID: 21703448

5. Tai C, Abe Y, Westenbroek RE, Scheuer T, Catterall WA. Impaired excitability of somatostatin- and parvalbumin-expressing cortical interneurons in a mouse model of Dravet syndrome. Proc Natl Acad Sci. 2014; 111: E3139–E3148. https://doi.org/10.1073/pnas.1411131111 PMID: 25024183

6. Biel M, Wahl-Schott C, Michalakis S, Zong X. Hyperpolarization-activated cation channels: from genes to function. Physiol Rev. 2009; 89: 847–885. https://doi.org/10.1152/physrev.00029.2008 PMID: 19584315

7. Catterall WA, Goldin AL, Waxman SG. International Union of Pharmacology. XLVII. Nomenclature and Structure-Function Relationships of Voltage-Gated Sodium Channels. Pharmacol Rev. 2005; 57: 397–409. https://doi.org/10.1124/pr.57.4.4 PMID: 16382098

8. Coetzee WA, Amarillo Y, Chiu J, Chow A, Lau D, McCORMACK T, et al. Molecular Diversity of K+ Channels. Ann N Y Acad Sci. 1999; 868: 233–255. https://doi.org/10.1111/j.1749-6632.1999.tb11293.x PMID: 10414301

9. Ranjan R, Khazen G, Gambazzi L, Ramaswamy S, Hill SL, Schürmann F, et al. Channelpedia: an integrative and interactive database for ion channels. Front Neuroinformatics. 2011; 5: 36. https://doi.org/10.3389/fninf.2011.00036 PMID: 22232598

10. Jung H, Yoon BC, Holt CE. Axonal mRNA localization and local protein synthesis in nervous system assembly, maintenance and repair. Nat Rev Neurosci. 2012; 13: 308–324. https://doi.org/10.1038/nrn3210 PMID: 22498899

11. Schulz DJ, Temporal S, Barry DM, Garcia ML. Mechanisms of voltage-gated ion channel regulation: from gene expression to localization. Cell Mol Life Sci CMLS. 2008; 65: 2215–2231. https://doi.org/10.1007/s00018-008-8060-z PMID: 18408887

12. Shipston MJ. Ion Channel Regulation by Protein Palmitoylation. J Biol Chem. 2011; 286: 8709–8716. https://doi.org/10.1074/jbc.R110.210005 PMID: 21216969

13. Mainen ZF, Sejnowski TJ. Influence of dendritic structure on firing pattern in model neocortical neurons. Nature. 1996; 382: 363–366. https://doi.org/10.1038/382363a0 PMID: 8684467

14. Epi4K Consortium, Epilepsy Phenome/Genome Project. De novo mutations in epileptic encephalopathies. Nature. 2013; 501: 217–221. https://doi.org/10.1038/nature12439 PMID: 23934111

15. Lal D, Ruppert A- K, Trucks H, Schulz H, de Kovel CG, Kasteleijn-Nolst Trenité D, et al. Burden Analysis of Rare Microdeletions Suggests a Strong Impact of Neurodevelopmental Genes in Genetic

Generalised Epilepsies. PLoS Genet. 2015; 11: e1005226. https://doi.org/10.1371/journal.pgen.1005226 PMID: 25950944

16. O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat Genet. 2011; 43: 585–589. https://doi.org/10.1038/ng.835 PMID: 21572417

17. Nelson SB, Sugino K, Hempel CM. The problem of neuronal cell types: a physiological genomics approach. Trends Neurosci. 2006; 29: 339–345. https://doi.org/10.1016/j.tins.2006.05.004 PMID: 16714064

18. Poulin J-F, Tasic B, Hjerling-Leffler J, Trimarchi JM, Awatramani R. Disentangling neural cell diversity using single-cell transcriptomics. Nat Neurosci. 2016; 19: 1131–1141. https://doi.org/10.1038/nn.4366 PMID: 27571192

19. Cadwell CR, Palasantza A, Jiang X, Berens P, Deng Q, Yilmaz M, et al. Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. Nat Biotechnol. 2015; https://doi.org/10.1038/nbt.3445 PMID: 26689543

20. Fuzik J, Zeisel A, Máté Z, Calvigioni D, Yanagawa Y, Szabó G, et al. Integration of electrophysiological recordings with single-cell RNA-seq data identifies neuronal subtypes. Nat Biotechnol. 2016; 34: 175–183. https://doi.org/10.1038/nbt.3443 PMID: 26689544

21. Toledo-Rodriguez M, Blumenfeld B, Wu C, Luo J, Attali B, Goodman P, et al. Correlation maps allow neuronal electrical properties to be predicted from single-cell gene expression profiles in rat neocortex. Cereb Cortex N Y N 1991. 2004; 14: 1310–1327. https://doi.org/10.1093/cercor/bhh092 PMID: 15192011

22. Okaty BW, Miller MN, Sugino K, Hempel CM, Nelson SB. Transcriptional and electrophysiological maturation of neocortical fast-spiking GABAergic interneurons. J Neurosci Off J Soc Neurosci. 2009; 29: 7040–7052. https://doi.org/10.1523/JNEUROSCI.0105-09.2009 PMID: 19474331

23. Mancarci BO, Toker L, Tripathy S, Li B, Rocco B, Sibille E, et al. NeuroExpresso: A cross-laboratory database of brain cell-type expression profiles with applications to marker gene identification and bulk brain tissue transcriptome interpretation. bioRxiv. 2016; 89219. https://doi.org/10.1101/089219

24. Tripathy SJ, Savitskaya J, Burton SD, Urban NN, Gerkin RC. NeuroElectro: a window to the world's neuron electrophysiology data. Front Neuroinformatics. 2014; 8: 40. https://doi.org/10.3389/fninf.2014.00040 PMID: 24808858

25. Tripathy SJ, Burton SD, Geramita M, Gerkin RC, Urban NN. Brain-wide analysis of electrophysiological diversity yields novel categorization of mammalian neuron types. J Neurophysiol. 2015; jn.00237.2015. https://doi.org/10.1152/jn.00237.2015 PMID: 25810482

26. Ascoli GA, Alonso-Nanclares L, Anderson SA, Barrionuevo G, Benavides-Piccione R, Burkhalter A, et al. Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. Nat Rev Neurosci. 2008; 9: 557–568. https://doi.org/10.1038/nrn2402 PMID: 18568015

27. Tasic B, Menon V, Nguyen TN, Kim TK, Jarsky T, Yao Z, et al. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat Neurosci. 2016; 19: 335–346. https://doi.org/10.1038/nn.4216 PMID: 26727548

28. Prinz AA, Bucher D, Marder E. Similar network activity from disparate circuit parameters. Nat Neurosci. 2004; 7: 1345–1352. https://doi.org/10.1038/nn1352 PMID: 15558066

29. Marder E, O'Leary T, Shruti S. Neuromodulation of Circuits with Variable Parameters: Single Neurons and Small Circuits Reveal Principles of State-Dependent and Robust Neuromodulation. Annu Rev Neurosci. 2014; 37: 329–346. https://doi.org/10.1146/annurev-neuro-071013-013958 PMID: 25032499

30. Gorokhova S, Bibert S, Geering K, Heintz N. A novel family of transmembrane proteins interacting with β subunits of the Na,K-ATPase. Hum Mol Genet. 2007; 16: 2394–2410. https://doi.org/10.1093/hmg/ddm167 PMID: 17606467

31. Armstrong CM. The Na/K pump, Cl ion, and osmotic stabilization of cells. Proc Natl Acad Sci. 2003; 100: 6257–6262. https://doi.org/10.1073/pnas.0931278100 PMID: 12730376

32. Bean BP. The action potential in mammalian central neurons. Nat Rev Neurosci. 2007; 8: 451–465. https://doi.org/10.1038/nrn2148 PMID: 17514198

33. Tai C, Abe Y, Westenbroek RE, Scheuer T, Catterall WA. Impaired excitability of somatostatin- and parvalbumin-expressing cortical interneurons in a mouse model of Dravet syndrome. Proc Natl Acad Sci U S A. 2014; https://doi.org/10.1073/pnas.1411131111 PMID: 25024183

34. Lupica CR, Bell JA, Hoffman AF, Watson PL. Contribution of the Hyperpolarization-Activated Current (I h) to Membrane Potential and GABA Release in Hippocampal Interneurons. J Neurophysiol. 2001; 86: 261–268. PMID: 11431507

35. Gao H, Smith BN. Tonic GABAA Receptor-Mediated Inhibition in the Rat Dorsal Motor Nucleus of the Vagus. J Neurophysiol. 2010; 103: 904–914. https://doi.org/10.1152/jn.00511.2009 PMID: 20018836

36. Hsiao J, Yuan TY, Tsai MS, Lu CY, Lin YC, Lee ML, et al. Upregulation of Haploinsufficient Gene Expression in the Brain by Targeting a Long Non-coding RNA Improves Seizure Phenotype in a Model of Dravet Syndrome. EBioMedicine. 2016; 9: 257–277. https://doi.org/10.1016/j.ebiom.2016.05.011 PMID: 27333023

37. Gasparini S, DiFrancesco D. Action of the hyperpolarization-activated current (I h) blocker ZD 7288 in hippocampal CA1 neurons. Pflüg Arch. 1997; 435: 99–106. https://doi.org/10.1007/s004240050488

38. Zheleznova NN, Sedelnikova A, Weiss DS. Function and modulation of delta-containing GABA(A) receptors. Psychoneuroendocrinology. 2009; 34 Suppl 1: S67–73. https://doi.org/10.1016/j.psyneuen.2009.08.010 PMID: 19766404

39. Yarishkin O, Lee DY, Kim E, Cho C-H, Choi JH, Lee CJ, et al. TWIK-1 contributes to the intrinsic excitability of dentate granule cells in mouse hippocampus. Mol Brain. 2014; 7: 80. https://doi.org/10.1186/s13041-014-0080-z PMID: 25406588

40. Hagenston AM, Rudnick ND, Boone CE, Yeckel MF. 2-Aminoethoxydiphenyl-borate (2-APB) increases excitability in pyramidal neurons. Cell Calcium. 2009; 45: 310–317. https://doi.org/10.1016/j.ceca.2008.11.003 PMID: 19100621

41. Stutzmann GE, LaFerla FM, Parker I. Ca2+ Signaling in Mouse Cortical Neurons Studied by Two-Photon Imaging and Photoreleased Inositol Triphosphate. J Neurosci. 2003; 23: 758–765. PMID: 12574404

42. Brew HM, Hallows JL, Tempel BL. Hyperexcitability and reduced low threshold potassium currents in auditory neurons of mice lacking the channel subunit Kv1.1. J Physiol. 2003; 548: 1–20. https://doi.org/10.1113/jphysiol.2002.035568 PMID: 12611922

43. Gittelman JX, Tempel BL. Kv1.1-Containing Channels Are Critical for Temporal Precision During Spike Initiation. J Neurophysiol. 2006; 96: 1203–1214. https://doi.org/10.1152/jn.00092.2005 PMID: 16672305

44. Perkowski JJ, Murphy GG. Deletion of the Mouse Homolog of KCNAB2, a Gene Linked to Monosomy 1p36, Results in Associative Memory Impairments and Amygdala Hyperexcitability. J Neurosci. 2011; 31: 46–54. https://doi.org/10.1523/JNEUROSCI.2634-10.2011 PMID: 21209188

45. Guan D, Armstrong WE, Foehring RC. Kv2 channels regulate firing rate in pyramidal neurons from rat sensorimotor cortex. J Physiol. 2013; 591: 4807–4825. https://doi.org/10.1113/jphysiol.2013.257253 PMID: 23878373

46. Hönigsperger C, Nigro MJ, Storm JF. Physiological roles of Kv2 channels in entorhinal cortex layer II stellate cells revealed by Guangxitoxin-1E. J Physiol. 2017; 595: 739–757. https://doi.org/10.1113/JP273024 PMID: 27562026

47. Goldfarb M, Schoorlemmer J, Williams A, Diwakar S, Wang Q, Huang X, et al. Fibroblast Growth Factor Homologous Factors Control Neuronal Excitability through Modulation of Voltage-Gated Sodium Channels. Neuron. 2007; 55: 449–463. https://doi.org/10.1016/j.neuron.2007.07.006 PMID: 17678857

48. Valente P, Lignani G, Medrihan L, Bosco F, Contestabile A, Lippiello P, et al. Cell adhesion molecule L1 contributes to neuronal excitability regulating the function of voltage-gated Na+ channels. J Cell Sci. 2016; 129: 1878–1891. https://doi.org/10.1242/jcs.182089 PMID: 26985064

49. Ekberg JA, Boase NA, Rychkov G, Manning J, Poronnik P, Kumar S. Nedd4-2 (NEDD4L) controls intracellular Na+-mediated activity of voltage-gated sodium channels in primary cortical neurons. Biochem J. 2014; 457: 27–31. https://doi.org/10.1042/BJ20131275 PMID: 24152020

50. Ishikawa T, Sato A, Marcou CA, Tester DJ, Ackerman MJ, Crotti L, et al. A Novel Disease Gene for Brugada SyndromeClinical Perspective. Circ Arrhythm Electrophysiol. 2012; 5: 1098–1107. https://doi.org/10.1161/CIRCEP.111.969972 PMID: 23064965

51. Chang K-J, Zollinger DR, Susuki K, Sherman DL, Makara MA, Brophy PJ, et al. Glial ankyrins facilitate paranodal axoglial junction assembly. Nat Neurosci. 2014; 17: 1673–1681. https://doi.org/10.1038/nn.3858 PMID: 25362471

52. Ohtaka-Maruyama C, Hirai S, Miwa A, Heng JI-T, Shitara H, Ishii R, et al. RP58 Regulates the Multipolar-Bipolar Transition of Newborn Neurons in the Developing Cerebral Cortex. Cell Rep. 2013; 3: 458–471. https://doi.org/10.1016/j.celrep.2013.01.012 PMID: 23395638

53. Okado H, Ohtaka-Maruyama C, Sugitani Y, Fukuda Y, Ishida R, Hirai S, et al. The transcriptional repressor RP58 is crucial for cell-division patterning and neuronal survival in the developing cortex. Dev Biol. 2009; 331: 140–151. https://doi.org/10.1016/j.ydbio.2009.04.030 PMID: 19409883

54. Deciphering Developmental Disorders Study. Prevalence and architecture of de novo mutations in developmental disorders. Nature. 2017; 542: 433–438. https://doi.org/10.1038/nature21062 PMID: 28135719

55. Yang XW, Wynder C, Doughty ML, Heintz N. BAC-mediated gene-dosage analysis reveals a role for Zipro1 (Ru49/Zfp38) in progenitor cell proliferation in cerebellum and skin. Nat Genet. 1999; 22: 327–335. https://doi.org/10.1038/11896 PMID: 10431235

56. Sharma K, Schmitt S, Bergner CG, Tyanova S, Kannaiyan N, Manrique-Hoyos N, et al. Cell type- and brain region-resolved mouse brain proteome. Nat Neurosci. 2015; 18: 1819–1831. https://doi.org/10.1038/nn.4160 PMID: 26523646

57. Mo A, Mukamel EA, Davis FP, Luo C, Henry GL, Picard S, et al. Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain. Neuron. 2015; 86: 1369–1384. https://doi.org/10.1016/j.neuron.2015.05.018 PMID: 26087164

58. Cembrowski MS, Bachman JL, Wang L, Sugino K, Shields BC, Spruston N. Spatial Gene-Expression Gradients Underlie Prominent Heterogeneity of CA1 Pyramidal Neurons. Neuron. 2016; 89: 351–368. https://doi.org/10.1016/j.neuron.2015.12.013 PMID: 26777276

59. Okaty BW, Sugino K, Nelson SB. A Quantitative Comparison of Cell-Type-Specific Microarray Gene Expression Profiling Methods in the Mouse Brain. PLoS ONE. 2011; 6: e16493. https://doi.org/10.1371/journal.pone.0016493 PMID: 21304595

60. Porcello DM, Ho CS, Joho RH, Huguenard JR. Resilient RTN Fast Spiking in Kv3.1 Null Mice Suggests Redundancy in the Action Potential Repolarization Mechanism. J Neurophysiol. 2002; 87: 1303–1310. PMID: 11877504

61. Lee HK, Hsu AK, Sajdak J, Qin J, Pavlidis P. Coexpression analysis of human genes across many microarray data sets. Genome Res. 2004; 14: 1085–94. https://doi.org/10.1101/gr.1910904 PMID: 15173114

62. Horvath S, Dong J. Geometric Interpretation of Gene Coexpression Network Analysis. PLoS Comput Biol. 2008; 4: e1000117. https://doi.org/10.1371/journal.pcbi.1000117 PMID: 18704157

63. Teeters JL, Godfrey K, Young R, Dang C, Friedsam C, Wark B, et al. Neurodata Without Borders: Creating a Common Data Format for Neurophysiology. Neuron. 2015; 88: 629–634. https://doi.org/10.1016/j.neuron.2015.10.025 PMID: 26590340

64. Jiang X, Shen S, Cadwell CR, Berens P, Sinz F, Ecker AS, et al. Principles of connectivity among morphologically defined cell types in adult neocortex. Science. 2015; 350: aac9462. https://doi.org/10.1126/science.aac9462 PMID: 26612957

65. Paul A, Cai Y, Atwal GS, Huang ZJ. Developmental Coordination of Gene Expression between Synaptic Partners During GABAergic Circuit Assembly in Cerebellar Cortex. Front Neural Circuits. 2012; 6: 37. https://doi.org/10.3389/fncir.2012.00037 PMID: 22754500

66. Nakajima M, Görlich A, Heintz N. Oxytocin Modulates Female Sociosexual Behavior through a Specific Class of Prefrontal Cortical Interneurons. Cell. 2014; 159: 295–305. https://doi.org/10.1016/j.cell.2014.09.020 PMID: 25303526

67. Shrestha P, Mousa A, Heintz N. Layer 2/3 pyramidal cells in the medial prefrontal cortex moderate stress induced depressive behaviors. eLife. 2015; 4. https://doi.org/10.7554/eLife.08752 PMID: 26371510

68. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw. 2010; 33: 1–22. PMID: 20808728

69. Richardet R, Chappelier JC, Tripathy S, Hill S. Agile text mining with Sherlok. 2015 IEEE International Conference on Big Data (Big Data). 2015. pp. 1479–1484. 10.1109/BigData.2015.7363910

70. Doyle JP, Dougherty JD, Heiman M, Schmidt EF, Stevens TR, Ma G, et al. Application of a Translational Profiling Approach for the Comparative Analysis of CNS Cell Types. Cell. 2008; 135: 749–762. https://doi.org/10.1016/j.cell.2008.10.029 PMID: 19013282

71. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Methodol. 1995; 57: 289–300.

72. Chawla K, Tripathi S, Thommesen L, Lægreid A, Kuiper M. TFcheckpoint: a curated compendium of specific DNA-binding RNA polymerase II transcription factors. Bioinformatics. 2013; 29: 2519–2520. https://doi.org/10.1093/bioinformatics/btt432 PMID: 23933972