

RESEARCH ARTICLE

Transcriptomic correlates of electrophysiological and morphological diversity within and across excitatory and inhibitory neuron classes

Claire Bomkamp^{1,2}, Shreejoy J. Tripathy^{1,3}, Carolina Bengtsson Gonzales⁴, Jens Hjerling-Leffler⁴, Ann Marie Craig^{1,2}, Paul Pavlidis^{1,3*}

1 Department of Psychiatry, University of British Columbia, Vancouver BC, Canada, **2** Djavad Mowafaghian Centre for Brain Health, University of British Columbia, Vancouver BC, Canada, **3** Michael Smith Laboratories, University of British Columbia, Vancouver BC, Canada, **4** Laboratory of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden

☞ These authors contributed equally to this work.

✉ Current address: Krembil Centre for Neuroinformatics, Centre for Addiction and Mental Health, Toronto ON, Canada and Department of Psychiatry, University of Toronto, Toronto ON, Canada

* paul@msl.ubc.ca



OPEN ACCESS

Citation: Bomkamp C, Tripathy SJ, Bengtsson Gonzales C, Hjerling-Leffler J, Craig AM, Pavlidis P (2019) Transcriptomic correlates of electrophysiological and morphological diversity within and across excitatory and inhibitory neuron classes. *PLoS Comput Biol* 15(6): e1007113. <https://doi.org/10.1371/journal.pcbi.1007113>

Editor: Lyle J. Graham, Université Paris Descartes, Centre National de la Recherche Scientifique, FRANCE

Received: January 24, 2019

Accepted: May 18, 2019

Published: June 18, 2019

Copyright: © 2019 Bomkamp et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The Bengtsson Gonzales PatchSeq dataset is available via GEO, accession number GSE130950. Processed data derived from the AIBS dataset are available at https://github.com/PavlidisLab/transcriptomic_correlates

Funding: PP was funded by Kids Brain Health Network (<http://kidsbrainhealth.ca/>), Natural Sciences and Engineering Research Council

Abstract

In order to further our understanding of how gene expression contributes to key functional properties of neurons, we combined publicly accessible gene expression, electrophysiology, and morphology measurements to identify cross-cell type correlations between these data modalities. Building on our previous work using a similar approach, we distinguished between correlations which were “class-driven,” meaning those that could be explained by differences between excitatory and inhibitory cell classes, and those that reflected graded phenotypic differences within classes. Taking cell class identity into account increased the degree to which our results replicated in an independent dataset as well as their correspondence with known modes of ion channel function based on the literature. We also found a smaller set of genes whose relationships to electrophysiological or morphological properties appear to be specific to either excitatory or inhibitory cell types. Next, using data from Patch-Seq experiments, allowing simultaneous single-cell characterization of gene expression and electrophysiology, we found that some of the gene-property correlations observed across cell types were further predictive of within-cell type heterogeneity. In summary, we have identified a number of relationships between gene expression, electrophysiology, and morphology that provide testable hypotheses for future studies.

Author summary

The behavior of neurons is governed by their electrical properties, for example how readily they respond to a stimulus or at what rate they are able to send signals. Additionally, neurons come in different shapes and sizes, and their shape defines how they can form

Discovery grant RGPIN-2016-05991 (http://www.nserc-crsng.gc.ca/index_eng.asp), and NIH grant MH111099 (<https://www.nih.gov/>). SJT was funded by a Canadian Institute for Health Research Post-doctoral Fellowship (<http://www.cihr-irsc.gc.ca/e/193.html>). AMC was supported by CIHR FDN-143206 (<http://www.cihr-irsc.gc.ca/e/193.html>) and Canada Research Chair (<http://www.chairs-chaires.gc.ca/home-accueil-eng.aspx>). JH-L was funded by the Swedish Research Council (Vetenskapsrådet, award 2014-3863, <https://www.vr.se/english.html>), StratNeuro (<https://ki.se/en/research/about-stratneuro>), and the Swedish Brain Foundation (Hjärnfonden, <https://www.hjarnfonden.se/om-hjarnfonden/about-hjarnfonden/>). CBG was funded by the NIH-KI doctoral program (<https://www.nimh.nih.gov/abs-at-nimh/scientific-director/office-of-fellowship-and-training/nih-karolinska-institute-graduate-program/index.shtml>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

connections with specific partners and thus function within the complete circuit. We know that these properties are governed by genes, acting acutely or during development, but we do not know which specific genes underlie many of these properties. Understanding how gene expression changes the properties of neurons will help in advancing our overall understanding of how neurons, and ultimately brains, function. This can in turn help to identify potential treatments for brain-related diseases. In this work, we aimed to identify genes whose expression showed a relationship with the electrical properties and shape measurements of different types of neurons. While our analysis does not identify causal relationships, our findings provide testable predictions for future research.

Introduction

Two prominent features that distinguish neurons from other cells are their electrical activity and their characteristic morphology. The specific pattern of electrophysiological activity displayed by a given neuron is a core property of its identity as one type of neuron or another. Similarly, different cell types often show striking differences in their size, branching complexity, and other morphological features. Neuronal cell types defined according to their electrophysiological or morphological characteristics show substantial correspondence with one another as well as with those defined using classification schemes based on transcriptomic criteria [1]. Electrophysiological characteristics of neurons, as well as their connectivity patterns, give rise to the computational properties of a given circuit [2,3]. Additionally, modeling studies show that morphological changes in simulated neurons can critically change their signaling capabilities [4–6]. Thus, understanding the origins of neuronal electrophysiology and morphology is an important step in understanding the mechanisms of brain function, both in the context of basic research and in the search for treatments for neuropsychiatric disorders.

A comprehensive understanding of the mechanisms that give rise to electrophysiological or morphological diversity must necessarily include a catalogue of the genes whose products contribute to these properties. Many genes have been shown experimentally to influence neuronal electrophysiology through a variety of mechanisms, including but not limited to ion channel activity, protein trafficking, and transcription factor activity [7–9]. Processes such as axon guidance and the development of dendrite morphology are also known to be under genetic control [10]. Despite this, our understanding of the relationship between gene expression and electrophysiological or morphological properties is quite limited.

Correlational approaches looking at the relationship between gene expression and neuronal properties in the absence of experimental manipulation can be a valuable method of prioritizing genes for future experimental study. For example, multiplex RT-PCR has been used to examine relationships between binary expression of a small group of genes and various electrophysiological parameters, allowing for reliable prediction of electrophysiological phenotypes based on combinatorial gene expression [11]. In addition, the recently developed Patch-Seq methodology, in which electrophysiology, morphology, and gene expression are analyzed in a single cell, has been used to examine relationships between gene expression and morphological or electrophysiological properties. However these analyses have thus far been limited to either predicting neuronal properties based on combinatorial gene expression profiles, examining relationships between cell types, or showing relationships predicted *a priori* based on known gene function [12–15].

In previous work [16], we combined publicly accessible electrophysiological and gene expression datasets in order to examine the relationship between gene expression and

electrophysiological properties. By matching groups of cells inferred to be similar based on multiple information sources, such as the transgenic reporter line and the brain region cells were isolated from, we were able to combine separate datasets containing gene expression and electrophysiological data to generate lists of genes which were correlated with one of several electrophysiological properties (as outlined in Fig 1A). The goal of this approach was to identify candidate genes that could be further studied using knockout or knockdown approaches in order to determine whether a causal relationship was present.

One caveat in our prior study is that the gene-electrophysiology correlations we identified may have been confounded by overall differences between broad cell classes. Across multiple datasets and cellular characterization methods, including gene expression [17–20], and electrophysiology and morphology [1], clustering cellular phenotypes in an unbiased manner

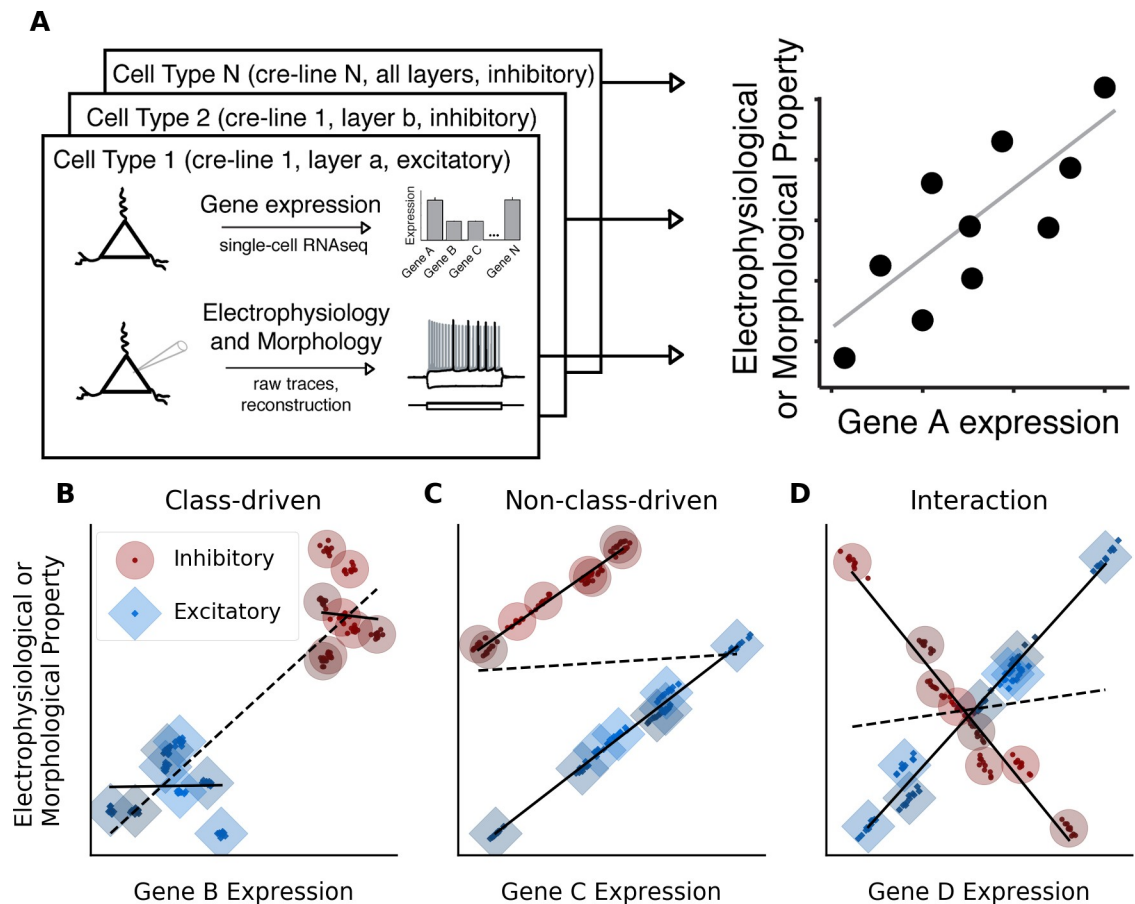


Fig 1. Methods for modeling relationships between gene expression and electrophysiological or morphological properties with respect to cell class (A) Schematic for defining cell types from single-cell transcriptomic or electrophysiological and morphological data. We divided cells into types based on Cre-driver expression as well as cortical layer and excitatory/inhibitory identity (left). Right panel shows summarization of cellular features by cell type for a hypothetical gene and property, where each point in the scatter plot represents each cell type’s mean gene expression (x-axis) and the mean value of an electrophysiological or morphological property (y-axis). (B) A hypothetical class-driven relationship between a gene and an electrophysiological or morphological property, in which neither cell class (excitatory or inhibitory) shows a relationship between gene expression and the property (solid lines), but an overall relationship appears because of systematic cross-class differences in both data modalities (dashed line). For B-D, small points represent individual cells and larger circles or diamonds represent cell type averages. (C) A hypothetical example of a non-class-driven relationship, where the gene-property relationship appears within each major cell class (solid lines), but would be obscured if modeled in a class-independent manner (dashed line). (D) A hypothetical example of a gene-property relationship exhibiting an interaction with cell class. Here, expression of the gene is positively correlated with the property in excitatory cell types but negatively correlated in inhibitory types (solid lines).

<https://doi.org/10.1371/journal.pcbi.1007113.g001>

reveals the major taxonomic difference between neurons to be between projecting and non-projecting neurons [18], or in the case of those cell types present in the cortex or hippocampus, excitatory and inhibitory neurons [17,19,20]. Thus, the commonly held view that a neuron's identity is first and foremost defined by its excitatory or inhibitory identity [21] is corroborated across multiple data sources and experimental modalities.

Therefore, we reasoned that the dataset we used previously was potentially susceptible to this confounding effect of cell class, since it contained a mixture of cells from different broad cell classes. In this work, we will use the term “cell type” to refer to narrowly-defined cell types, and “cell class” to refer to those which are broadly-defined (excitatory versus inhibitory or projecting versus non-projecting). We refer to correlations between gene expression and electrophysiological or morphological properties that are explained by differences between cell classes as “class-driven,” (e.g. Fig 1B) and to those that exist based on graded differences within broad cell classes as “non-class-driven” (e.g. Fig 1C). We reason that gene-property relationships that are non-class-driven would be more likely to be potential causal regulators of the associated property. Although some class-driven correlations likely do reflect true relationships between genes and properties which distinguish excitatory from inhibitory cells, separating these relationships from instances where one cell class has a higher value of a property and coincidentally higher or lower expression of a gene without additional sources of data is not possible. Effectively, such situations are analogous to attempting to draw conclusions about correlations with only two data points.

Due to limitations in available data, we were unable to address the effect of cell class in our previous work [16]. Since then, the RNA-seq and electrophysiology datasets from the Allen Institute for Brain Science (AIBS) (which we originally used as validation data) have expanded greatly, with more cells and more transgenic lines represented. This increase in size, together with the fact that the AIBS data were collected using standardized protocols, suggests that this dataset might prove valuable for discovering genes correlated with electrophysiological and morphological properties. In addition, the growing use of the PatchSeq methodology [12], allowing transcriptomic, electrophysiological, and morphological characterization of the same single cell, also affords an opportunity to test gene-property correlations.

Leveraging the larger size of the new AIBS dataset, we were able to address limitations of our previous study related to excitatory versus inhibitory cell class by employing statistical methods to help mitigate the effects of cell class. These methods, together with the larger number of cell types represented in the new dataset, allowed us to identify novel electrophysiological and morphological property-related gene sets which are potentially more likely to represent meaningful biological relationships.

Results

Primary dataset

The primary dataset we used combined groups of cells from mouse visual cortex characterized by the Allen Institute for Brain Science (AIBS; <http://celltypes.brain-map.org/>), where multiple Cre-driver lines were used to target cells for characterization. Standard electrophysiological protocols were used to characterize cells *in vitro*, with a subset of these cells further undergoing detailed morphological characterization [1]. In addition, a separate group of cells were subjected to deep single-cell RNA-sequencing to characterize cellular transcriptomes [20]. Because the same Cre-lines were used to characterize cells along multiple modalities of neuronal function, we were able to summarize these data to the “cell type” level (reflecting Cre-line, cortical layer, and major neurotransmitter; shown in S1 Table) by pooling and combining cellular characterization data across different animals and data modalities. We acknowledge that

the cell type classification used here is somewhat crude and that there is overlap between cell types as we have defined them, as well as variability within types, but chose to define cell types in this way because it allows us to combine data from the RNA-seq and electrophysiology/morphology datasets. The definition of multiple cell types within one Cre-line based on cortical layer and major neurotransmitter is supported by cross-layer differences in gene expression [20] and in electrophysiological properties (S1 Fig).

The final combined dataset is composed of 34 inhibitory GABAergic and 14 excitatory glutamatergic types (48 total) with electrophysiological data, and 30 inhibitory and 13 excitatory types (43 total) with morphological data. The increased size of this dataset is a considerable advance over our prior analysis [16], which employed an older version of the same dataset (only 12 cell types) [19]. This was made possible in part because of more Cre-lines available for analysis and finer cortical layer dissections for the transcriptomic data. For each cell type thus defined, we computed the mean log expression value for each gene represented in the RNA-seq dataset and the mean value of each of sixteen electrophysiological and six morphological properties (described in S2 Table). Some of the electrophysiological and morphological properties were log-transformed prior to taking the mean (see Methods). In order to account for the existence of correlations among properties, we additionally performed principal component analysis (PCA) on either the electrophysiological or the morphological properties, and assessed the correspondence of each of the resulting principal components (E_PC1-E_PC3 and M_PC1-M_PC3) to gene expression.

Analysis approach

Our goal was to identify, for each electrophysiological or morphological property, genes that were correlated with the property (Fig 1A). However, overall differences between excitatory and inhibitory cell classes can make the interpretation of such relationships more complicated in several ways. For example, Fig 1B shows an example of a gene-property correlation that appears almost entirely **class-driven**, meaning that although no relationship appears *within* either cell class, the apparent relationship is entirely driven by differences *between* cell classes. In this case, inhibitory cell types show higher expression of the gene and a greater value of the property compared to excitatory cell types. In contrast, Fig 1C shows a **non-class-driven** relationship, meaning one that manifests in both cell classes, but which may be obscured by baseline differences when the cell classes are grouped. In this example, a correlation that appears within both classes independently is obscured by a higher value of the property in inhibitory compared to excitatory cell types. Although this obscuring effect is present in this particular example, it is not required for a relationship to be considered non-class-driven; we expected to see some relationships that were consistent both within each class as well as among all cell types. Both class-driven correlations and non-class-driven correlations that are obscured by the effects of class can be viewed as instances of Simpson's paradox, in which the direction of an overall relationship is the opposite of the relationship within relevant subsets of the data [22,23]. We hypothesized that non-class-driven, but not class-driven, relationships would be consistent with those relationships existing across individual cells within a cell type (small points, representing individual cells within a type, in Fig 1B–1D).

In order to computationally account for these possibilities, we evaluated each combination of gene and property using a statistical model that assesses the predictive value of the gene on the property while controlling for the effects of cell class. We termed this model the **class-conditional model** ($P \sim G + C$, where P is the value of the property, G is the gene expression level, and C is cell class). This model would be expected to identify a significant relationship between the gene and the property when a non-class-driven relationship is present (Fig 1C), but would

not identify relationships that are class-driven (Fig 1B). The class-conditional model is by no means the only method that could be used to control for the effects of cell class, and other methods such as subsetting the dataset to look at only one class at a time would likely produce useful results as well. We chose to use the class-conditional approach because it allowed us to use the data from both excitatory and inhibitory cell types, whereas subsetting the dataset to include only one class would have led to our throwing out data from many relevant cell types. For comparison, we modeled the same gene-property pairs using a **class-independent model** ($P \sim G$), which assesses the predictive value of the gene on the property irrespective of cell class. This model is similar in principle to the correlational method used in our previous work [16] and would be expected to produce a significant result in cases showing class-driven relationships (such as Fig 1B) but might miss some instances of non-class-driven relationships (such as Fig 1C).

Another possible gene-property relationship is one where there is an interaction between gene and class, meaning that the gene-property relationship is different in excitatory and inhibitory cell types. An interaction could indicate either that excitatory and inhibitory cell types both show a correlation between the gene and property, but the slopes are in opposite directions (as in the example in Fig 1D), or that the gene is correlated with the property only in one cell class. To detect such situations, we introduced a third model, the **interaction model** ($P \sim G + C + G * C$), which tested whether the relationship between gene expression and the property in question was significantly different between excitatory and inhibitory cell types. In summary, the three models are designed to answer three different questions:

Class-independent model: Is expression of the gene a significant predictor of the property if we assume that cell class is not a factor?

Class-conditional model: After accounting for cell class, is the gene's expression a significant predictor of the property?

Interaction model: Is the relationship between the gene's expression and the property statistically different in inhibitory and excitatory cells?

Accounting for cell class results in the identification of a distinct but overlapping set of genes

We first set out to understand how accounting for cell class identity (excitatory or inhibitory) affects the interpretation of gene-property relationships. We found 12,225 genes with expression levels passing our basic filtering criteria (see Methods for details on how expression-level filtering was performed) and assessed each gene in combination with each electrophysiological or morphological property. We modeled each relationship with or without including an indicator variable for cell class, using the class-conditional or class-independent models described above. For most properties, we found that the degree of overlap between the sets of genes identified in the two models (at a false discovery rate (FDR) of 0.1) was substantial but far from a complete intersection (Fig 2A, Venn diagrams, and S2 Table). For example, for after-hyperpolarization (AHP) amplitude, we found ~6000 significantly-associated genes in the class-independent model and ~6500 in the class-conditional model; out of these, ~3700 genes were shared between models. Thus, accounting for cell class results in the identification of a substantially different set of candidate genes, which suggests that many of the genes identified in our previous work [16] might reflect class-driven gene-property relationships.

We found that most of the electrophysiological properties showed some degree of correlation with E_PC1, as well as varying degrees of overlap with the genes showing a significant result in the class-conditional model, whereas E_PC2 showed substantial overlap with only sag and AP threshold (S2A and S2B Fig). The morphological properties showed poorer overlap

with the principal components, likely because M_PC1 and M_PC2 were primarily correlated with properties for which no significant genes were identified in the class-conditional model (S2D and S2E Fig). For both sets of features, excitatory and inhibitory cell types were fairly well separated by the combination of PC1 and PC2 (S2C and S2F Fig).

We next asked how overall differences in morphological and electrophysiological properties between excitatory and inhibitory cells affect gene-property relationships. To this end, we used a linear model to estimate the effect of cell class on each property. For most properties, there was a significant ($p < 0.05$) effect of cell class. The features of action potential (AP) threshold, input resistance, sag, rheobase, branchiness, soma surface, and bifurcation angle, as well as E_PC3, M_PC2, and M_PC3, are exceptions to this. The existence of a significant difference in most properties between excitatory and inhibitory cell types highlights the importance of taking cell class into account when attempting to relate these properties to gene expression. The properties without a significant difference are likely to be less susceptible to class-driven effects, but the class-independent model still might miss potentially interesting relationships due to differences in gene expression between classes, resulting in genes which are identified by the class-conditional model only.

We compared the strength and direction of the relationship in both the class-independent and class-conditional models by directly comparing the slopes derived from each model for each gene-property relationship (where slope indicates the change in the property per 2-fold change in gene expression; shown for AHP amplitude in Fig 2B). While there is broad agreement between the class-independent and class-conditional models ($r_{\text{Spearman}} = 0.52$), a substantial number of gene-property relationships are significant in one model but not the other (at FDR = 0.1). In other words, these relationships are either class-driven (significant in the class-independent model only) or non-class-driven and obscured by class (significant in the class-conditional model only). For example, the relationship between the gene *Gprasp1* and AHP amplitude illustrates an example of a class-driven relationship where the apparent relationship is entirely due to broad differences in excitatory and inhibitory classes (Fig 2C). The gene *Camk2g* shows a non-class-driven relationship with the same property that is obscured in the class-independent model by higher AHP amplitude values in inhibitory cell types (Fig 2D). However, many genes, such as *Xxylt1*, are identified using either model (Fig 2E).

Divergent gene-property relationships in inhibitory versus excitatory cell classes

We next wondered whether some gene-property relationships might be potentially different within, or specific to, excitatory or inhibitory cell types. To test this, we incorporated an interaction term between gene expression and excitatory versus inhibitory cell class to assess whether the gene-property relationships (i.e. slopes) were different within each cell class. For nearly all properties, there were fewer genes with a significant interaction term in the interaction model compared to the number with a significant gene term in the class-conditional model (Fig 3A, Venn diagrams, and S3 Table). For example, out of the ~6600 genes significantly associated with AHP amplitude in the class-conditional model, ~2300 also show interactions, and there are an additional ~700 which show an interaction but are not significant in the class-conditional model. This could indicate that “true” interactions are comparatively rare, but this finding is also likely partly explained by differences in statistical power. In addition, these interactions do not appear to be merely the result of low or no gene expression within one cell class but not the other; we did not observe strong correlations for any property between the interaction term coefficient and the average difference in expression levels between inhibitory and excitatory cell types (S3 Fig).

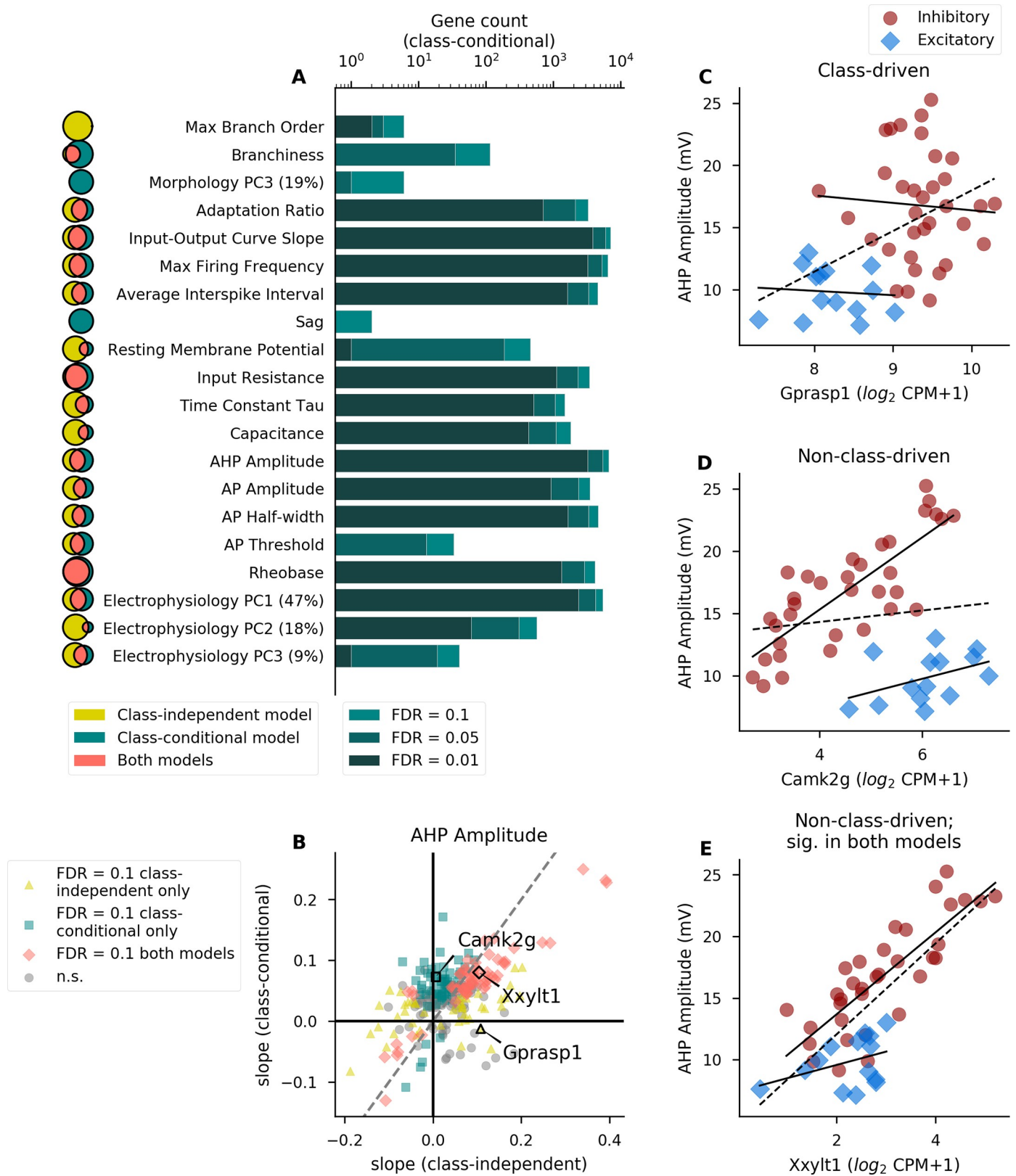


Fig 2. Different sets of genes are associated with electrophysiological and morphological properties after correcting for cell class. (A) Number of genes significantly associated with each property in the class-conditional model at various levels of significance (only properties with significant genes in this model are

shown). Darkness of the bar represents the significance level of each group of genes. Venn diagrams to the left indicate the extent of overlap (pink; middle) between the gene sets identified by the class-independent (gold; left) and class-conditional (teal; right) models, where the area of each segment is proportional to the significant gene count at a threshold of FDR = 0.1. Venn diagrams for different properties are not to scale with one another. Percentages next to principal components (PCs) indicate the percent variability explained by that component. See [S2 Table](#) for descriptions of electrophysiological and morphological properties analyzed here, as well as gene counts for all properties. (B) Comparison of model-based slopes from the class-independent and class-conditional models. Each point represents a single gene's relationship with the electrophysiological property AHP amplitude and is colored according to whether the relationship is significant in one or both models (at FDR = 0.1). Example genes in C-E are indicated. For clarity of visualization, only a random subset of genes (2% total) are shown to mitigate over-plotting. Dashed line indicates identity. (C-E) Examples of genes showing significant associations with AHP amplitude that are class-driven (C; significant in class-independent model only), non-class-driven (D; significant in class-conditional model only), or non-class-driven but significant by either model (E). Solid lines indicate linear fits within excitatory or inhibitory cell classes only and dashed line indicates a linear fit including all cell types. Gene expression is quantified as counts per million (CPM).

<https://doi.org/10.1371/journal.pcbi.1007113.g002>

For all properties, we found that the slopes of the gene-property relationships within excitatory cell types were poorly correlated with those within inhibitory cell types (example features maximum branch order and AHP amplitude shown in [Fig 3B and 3C](#)). By definition, the genes with significant interaction terms were those where the slopes calculated within excitatory and inhibitory classes were very different from each other (pink and purple points in [Fig 3B and 3C](#)). If the majority of gene-property relationships are shared between excitatory and inhibitory cell types, as one might expect if neuronal properties are regulated using largely consistent mechanisms even across cell classes, one might expect a positive correlation between slopes calculated in inhibitory and excitatory cell types. However, such a correlation may be lacking in this analysis because we would expect most genes to have no relationship to a given property and thus for the deviations of their slopes from zero to be essentially random.

The properties maximum branch order and sag are unusual in that they show few significant genes using the class-conditional model, but many (1914 and 1109, respectively) with significant interactions ([Fig 3A](#), Venn diagrams, and [S3 Table](#); slopes for maximum branch order plotted in [Fig 3B](#)). We hypothesize that this might be because these properties are under stronger (or otherwise more readily identified) genetic control in excitatory compared to inhibitory cell types, together with the uneven numbers of excitatory and inhibitory cell types (see [Discussion](#)).

[Fig 3D and 3E](#) show examples of genes with significant interaction terms for AHP amplitude. The class-conditional model also shows a significant relationship between AHP amplitude and expression of *Man1c1* ([Fig 3E](#)) but not *Nrxn3* ([Fig 3D](#)). In other words, the interaction model identified a potentially interesting relationship in the case of *Nrxn3* which was missed by the class-conditional model. For *Man1c1*, the interaction model does not reveal a new relationship, but instead highlights the fact that this gene-property relationship, if real, is potentially more complicated than would be assumed based on the class-conditional model alone. *Man1c1* is an enzyme involved in the maturation of N-linked oligosaccharides [24], and is thus a plausible regulator of AHP amplitude, since N-linked glycosylation of voltage-gated potassium channels or their auxiliary subunits is known to regulate both surface trafficking and channel function [25,26]. The apparent class-specificity of this relationship could result from class-specific co-expression of certain potassium channels or other enzymes involved in glycan synthesis or maturation.

Results from the class-conditional model are more likely to validate using independent methods

We next asked how the gene-property relationships from the class-independent and class-conditional models, based on our analysis of the AIBS cortical cell types dataset, might generalize to other datasets. We first compared the results reported here to those from our earlier NeuroElectro/NeuroExpresso (NE) literature-based dataset [16], after subsetting these data to

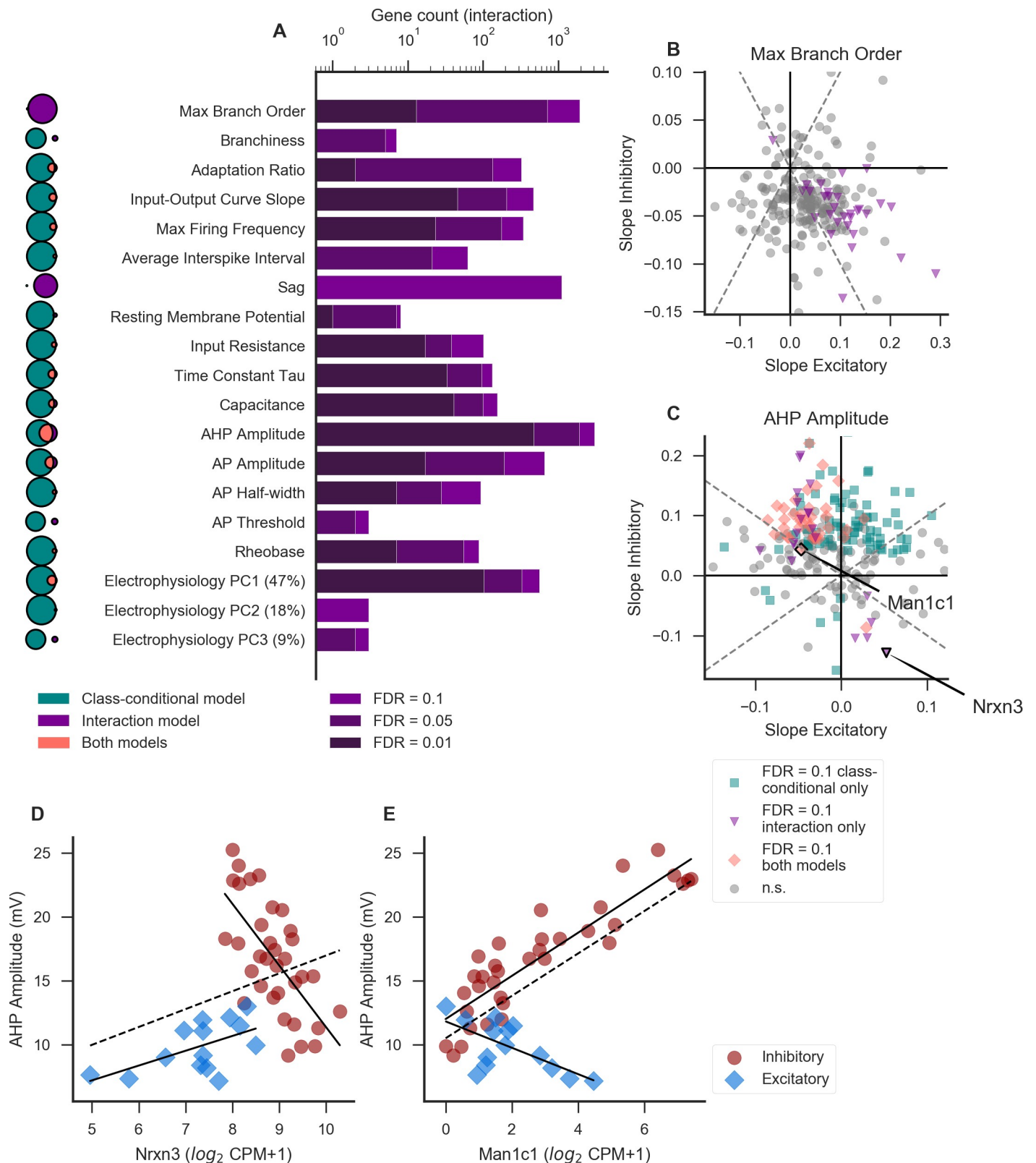


Fig 3. Identification of divergent gene-property relationships in excitatory versus inhibitory cell classes (A) Number of genes showing a significant interaction effect between gene and class for each property. Darkness of the bar represents the significance level of each group of genes. Venn diagrams to the left indicate the extent of overlap (pink; middle) between the class-conditional (teal; left) and interaction (purple; right) models, where the area of each segment is proportional to the

significant gene count at a threshold of FDR = 0.1. Venn diagrams for different properties are not to scale with one another. Percentages next to principal components (PCs) indicate the percentage of variability explained by that component. (B-C) Slope values within excitatory cell types (x axis) plotted against the slope values for the same set of genes in inhibitory cell types (y axis). Each point represents a single gene's relationship to the morphological property maximum branch order (B) or electrophysiological property AHP amplitude (C), and is colored according to its significance in one or both models (see inset legend). Example gene-property relationships highlighted in D-E are marked in panel C. For clarity of visualization, only a random 2% subset of the total number of genes are plotted. Dashed lines indicate positive and negative unity lines. (D) Example of a gene with a significant interaction term which is not significant in the class-conditional model. For D and E, solid lines indicate linear fits including only excitatory or only inhibitory cell types, and dashed line indicates a linear fit including all cell types. (E) Example of a gene which is significant in both the class-conditional and interaction models.

<https://doi.org/10.1371/journal.pcbi.1007113.g003>

include only non-projecting cell types (reflecting 19 cell types in total sampled throughout the brain, described in detail in the Methods). We chose to use non-projecting cell types in the NE dataset, as these were recently described by a mouse brain-wide transcriptomic survey as corresponding to a single broad cell class [18]. To this end, we used the same linear model-based approach as in the AIBS dataset to assess the relationships between genes and electrophysiological properties in the NE dataset. This approach was identical to the class-independent model, except that the dataset contained only one cell class. Next, for gene-property relationships from both the class-independent and class-conditional models, we assessed their aggregate consistency with those from the NE dataset. Here, we defined “consistency” for a given model (i.e. class-independent or class-conditional) and property as the Spearman correlation between gene-property slopes calculated from the AIBS dataset with the slopes for the same set of gene-property relationships in the NE dataset (illustrated in Fig 4B).

In Fig 4A we show a comparison of the gene/electrophysiology slopes from the AIBS and NE datasets [16]. We found that for seven out of the eleven electrophysiological properties shared between the datasets, both AIBS dataset-based statistical models were consistent with analogous gene-property relationships based on the NE dataset (95% confidence intervals above zero based on 100 bootstrap resamples). For six out of the eleven features, we found that the class-conditional model was considerably more consistent ($p < 0.05$ based 100 bootstrap resamples) than the class-independent model with relationships in the NE dataset. For only one feature, membrane time constant (τ), was the class-independent model more consistent ($p < 0.05$) than the class-conditional with the NE dataset. Fig 4B shows an example of how consistency was measured for AP half-width. The relationship between *Atp2a2* expression and AP half-width is shown in Fig 4C and 4D as an example of a gene-property relationship which is consistent between the NE ($r = -0.742$) and AIBS datasets for the class-conditional ($\beta = -0.099 \pm 0.024$; $q = 0.002$, where q indicates the lowest FDR threshold at which the gene would be considered significant, also sometimes called the adjusted p-value) but not the class-independent ($\beta = -0.024 \pm 0.034$; $q = 0.62$) model.

Assessing within-cell type correlations using PatchSeq datasets

We next wondered whether these between-cell type gene-property relationships might be predictive of cell-to-cell heterogeneity within a given cell type. We reasoned that the recently developed PatchSeq methodology, allowing morphological, electrophysiological, and transcriptomic characterization from the same single cell, presents a unique opportunity to test this possibility [12]. While these data at present are limited by relatively modest sample sizes and technical factors such as inefficient mRNA capture and potential off-target cellular mRNA contamination [27], we nonetheless sought to use these data to assess the nature of within-cell type gene-property relationships.

To this end, we performed an integrated analysis of 5 PatchSeq datasets, with each dataset characterizing transcriptomic and electrophysiological diversity of mouse forebrain inhibitory cells from the neocortex, hippocampus, and striatum (Table 1). We chose to focus on

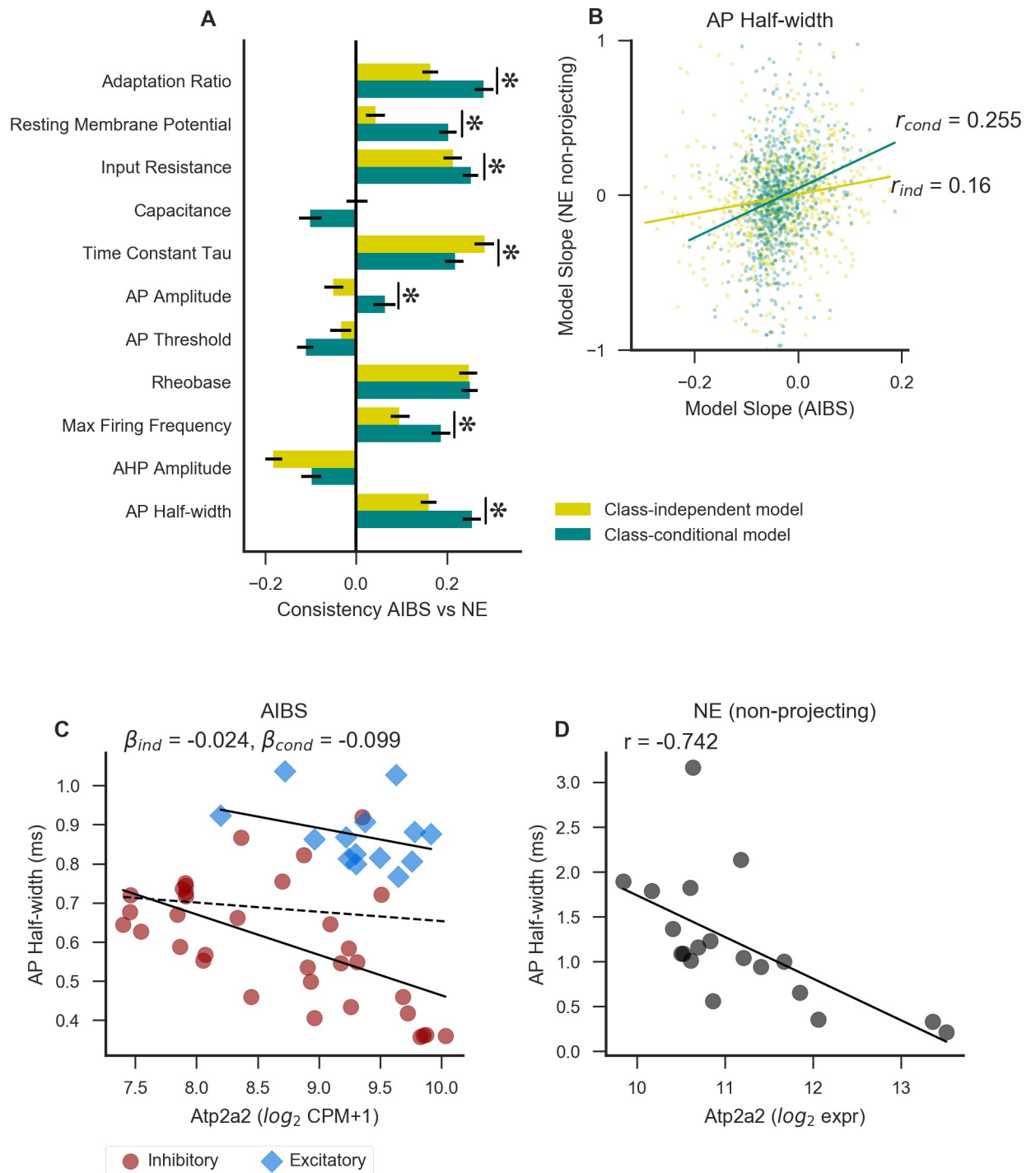


Fig 4. Modeling gene/electrophysiology relationships using the class-conditional model is more predictive than the class-independent model of model slopes in an independent dataset containing non-projecting cell types only (A) Aggregate gene-property relationship consistency between AIBS and NeuroExpresso/NeuroElectro (NE) datasets. Error bars indicate a 95% confidence interval, and asterisk indicates a significant ($p < 0.05$) difference in the consistency metric between the class-independent and class-conditional models, calculated using 100 bootstrap resamples of the original values (indicated only for properties where both values are positive). (B) Direct comparison of gene-property relationships between the AIBS and NE datasets. Each point represents the relationship between a single gene and the property AP half-width. The model slope from the AIBS dataset is plotted on the x axis (with the class-independent model (ind) slopes in gold, and the class-conditional model (cond) slopes in teal), and that for the same set of genes in the NE dataset on the y axis. For clarity of visualization only 10% of the total number of genes are plotted. Lines indicate a linear fit for each set of points. The correlation within each set of points is used as a measure of cross-dataset consistency (plotted for all properties in panel A). (C-D) Example of a gene showing consistent results between the NE dataset and

the AIBS dataset using the class-conditional model, but not the class-independent model. C shows the relationship within the AIBS dataset, and D shows the same gene and property in the NE dataset. Solid lines indicate a linear fit including only types belonging to one cell class, and dashed line indicates a linear fit including all cell types.

<https://doi.org/10.1371/journal.pcbi.1007113.g004>

inhibitory neurons here as these were present in each of the PatchSeq datasets. Our analysis includes one novel dataset of 19 Pvalb-Cre positive interneurons recorded in region CA1 of the mouse hippocampus, reported here for the first time. Cells in this dataset (referred to as the Bengtsson Gonzales dataset), were characterized as described in [15].

To jointly analyze these PatchSeq datasets, we first mapped PatchSeq sampled cells to the subclass level (i.e., major cell type level; e.g., Pvalb, Sst, Vip; defined in S4 Table) as defined by as defined by cellular dissociation-based single-cell RNAseq reference atlases from the cortex and striatum (see Methods) [15,20]. Next, for each major cell type, we identified genes that are highly variable in their expression levels within cells of the same type. We reasoned that these highly-variable genes might be those most likely to drive or appear correlated with electrophysiological heterogeneity within each cell type. Lastly, we performed a joint analysis across PatchSeq datasets to assess the strength of gene-property relationships within cell types where the gene was highly variable. Here, we used a mixed-effects regression model, with gene expression as a fixed effect and dataset and cell type as random effects and with cells weighted by their estimated transcriptome quality (see Methods).

Despite the limitations of the PatchSeq data, we found a small number of genes whose expression levels were significantly associated with cell-to-cell electrophysiological heterogeneity within cell types (FDR = 0.1; Fig 5A). For example, we found that expression of *Kcna1*, which encodes the potassium channel Kv1.1, was inversely correlated with AP half-width (Fig 5B; $\text{Beta}_{\text{PatchSeq}} = -0.0484 \pm 0.0106$, $q_{\text{PatchSeq}} = 0.0683$) within hippocampal *Pvalb* and striatum *Pthlh* cells (the only cell types in which the variability in *Kcna1* expression met our threshold for analysis). Importantly, there was also a significant relationship with the same directionality for *Kcna1* and AP half-width in the AIBS dataset ($\text{Beta}_{\text{class-conditional}} = -0.048 \pm 0.011$, $q_{\text{class-conditional}} = 0.001$). Moreover, the relationship between *Kcna1*/Kv1.1 expression and action potential width has been experimentally reported previously [28].

As another example, we saw an inverse correlation between *Fxyd6* expression and AHP amplitude, based on cortical Lamp5- and striatum Th- cells (Fig 5C, $\text{Beta}_{\text{PatchSeq}} = -0.695 \pm 0.118$, $q_{\text{PatchSeq}} = 0.00841$). We also saw a similar relationship in the AIBS dataset ($\text{Beta}_{\text{class-conditional}} = -0.021 \pm 0.003$, $q_{\text{class-conditional}} = 0.00001$). Intriguingly, *Fxyd6* encodes

Table 1. Description of PatchSeq datasets re-analyzed in this study. Depending on the dataset, RNA amplification was performed using variations on single-cell-tagged reverse transcription (STRT) [31] or Switching Mechanism At the end of the 5'-end of the RNA Transcript (SMART) [32]. The Bengtsson Gonzales dataset reflects a novel dataset reported here for the first time.

Dataset	Description	RNA amplification	Number of cells	Accession
Cadwell [12]	Cortical layer 1 interneurons	Smart-seq2	57	E-MTAB-4092
Fuzik [13]	Cortical layer 1/2 interneurons and pyramidal cells	STRT-C1 (with unique molecule identifiers)	80	GSE70844
Földy [14]	Hippocampal CA1 and subiculum pyramidal cells and regular- and fast-spiking interneurons	SMARTer	93	GSE75386
Muñoz-Manchado [15]	Striatum interneurons	STRT-C1 (with unique molecule identifiers)	99	GSE119248
Bengtsson Gonzales	Hippocampal CA1 Pvalb-Cre interneurons	STRT-C1 (with unique molecule identifiers)	19	GSE130950

<https://doi.org/10.1371/journal.pcbi.1007113.t001>

phosphohippolin, a regulator of Na⁺/K⁺ ATPase activity [29] and is thus plausibly involved in the AHP and action potential repolarization. Intriguingly, in a separate single-cell RNA-seq study of CA1 interneurons, *Fxyd6* was found to be more highly expressed cells known to spike more slowly [30].

In general, we found that when a gene-property relationship was statistically significant in both the PatchSeq and AIBS class-conditional analyses (at FDR = 0.1), this relationship was usually in the same direction in both analyses (Fig 5A; 10 out of 13 gene-property relationships total). Results were similar in the class-independent model, except with a smaller set of gene/phys relationships matching between both (7 out of 9 relationships were in a consistent direction). All of the genes which were consistent between the class-independent and PatchSeq analyses were also consistent in the class-conditional model. While our analyses of these PatchSeq datasets should be considered preliminary (pending the availability of larger and higher-quality datasets), we find the correspondence with our earlier analysis encouraging. Namely, this analysis suggests that some of the same genes that appear to drive large differences across cortical cell types might also be defining more subtle within-cell type heterogeneity.

The expected relationship between voltage-gated potassium channels and AHP amplitude is apparent only after accounting for cell class

We next asked whether we see a relationship between an electrophysiological feature and a category of genes which are known regulators of that feature. Voltage-gated potassium channels are known to be involved in producing the after-hyperpolarization following an action potential [33,34] (AHP amplitude; illustrated by the dashed arrow in Fig 6A). We thus hypothesized that for many of these genes, higher expression levels would be associated with larger AHP amplitudes (although not all voltage-gated potassium channels necessarily contribute directly to AHP amplitude). We further hypothesized that this relationship would be more apparent after accounting for cell class, in part because AHP amplitudes differ considerably between excitatory and inhibitory cell classes (Fig 6B–6D). Indeed, our previous work found a spurious negative correlation between expression of the *Kcnb1* gene and AHP amplitude which resulted from higher expression of *Kcnb1* in excitatory cell types compared to others [16].

We evaluated model slopes between each of 29 voltage-gated potassium channel genes [35] and AHP amplitude in the AIBS dataset for each of the class-independent and class-conditional statistical models (examples shown in Fig 6B–6D and summary in Fig 6E).

Examples of voltage-gated potassium channel genes associated with AHP amplitude include *Kcnh3* (Fig 6B) in a class-driven and *Kcnh7* and *Kcnc2* in a non-class-driven manner (Fig 6C and 6D). In total, the class-independent model identified 17 significant genes (at a stringent threshold of FDR = 0.01), with 8 of these genes having positive slopes and 9 negative. In contrast, there were 13 genes that were significantly associated with AHP amplitude in the class-conditional model at the same statistical threshold, and 11 of these genes had slopes in the positive direction. Thus the results obtained using the class-conditional model are consistent with our *a priori* hypothesis that expression levels of voltage-gated potassium channel genes are more likely to show positive than negative relationships with AHP amplitude, whereas the results obtained using the class-independent approach do not appear to support this conclusion.

Evidence of causal support for specific gene-property relationships

To further validate the gene-property correlations found in the AIBS dataset, we asked whether any of the same relationships showed direct support in the literature. In some cases we found that previously published work showed that manipulation of the gene of interest caused electrophysiological effects in line with what would be predicted by our analysis.

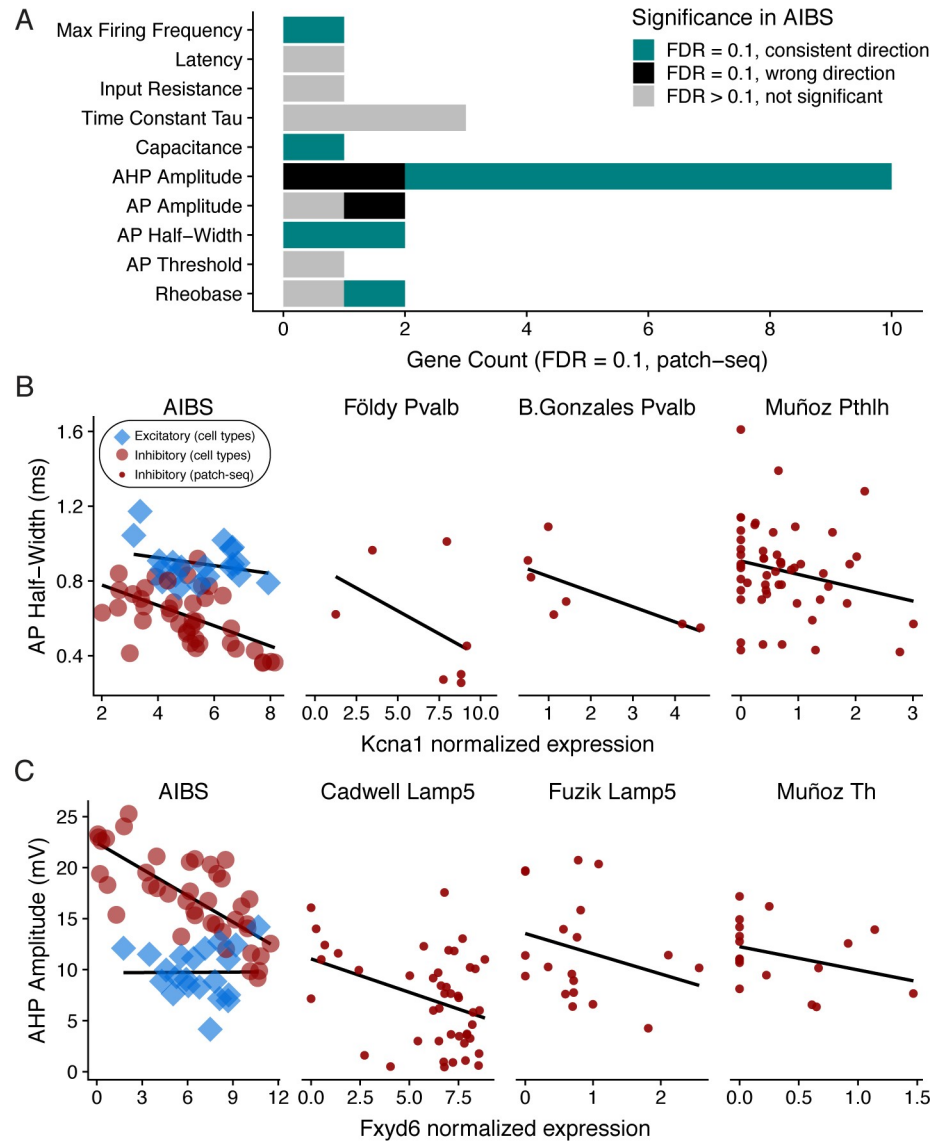


Fig 5. Assessing gene-property relationships within cell subclasses using PatchSeq (A) Number of genes associated with each electrophysiological property based on a joint cross-laboratory analysis of 5 PatchSeq datasets. Genes shown are significant at FDR = 0.1, based on a mixed-effects regression model, treating gene expression as a fixed effect and dataset identity and cell type as random effects. Bar color denotes overlap of PatchSeq based gene-property relationships with analogous relationships from the AIBS class-conditional model analysis. Note that analysis of gene-property relationships in the PatchSeq datasets are independent from those in the AIBS cell types analysis. (B, C) Examples of genes showing significant associations with electrophysiological features in the class-conditional analysis of the AIBS dataset (left-most panel) and the mixed-effects analysis of the PatchSeq datasets (other panels). Dataset name and cell type is shown in the subpanel title and solid lines indicate linear fits within cell classes (AIBS) or fits within each PatchSeq dataset and cell type, after weighting cells by transcriptome-quality (see Methods). Based on differences in mRNA quantification, x-axis units for AIBS, Cadwell, and Földy datasets are log₂ (CPM+1), and for Bengtsson Gonzales, Muñoz, and Fuzik datasets are log₂ normalized molecule counts (normalized to 2000 unique molecules per cell).

<https://doi.org/10.1371/journal.pcbi.1007113.g005>

Kcna1, a voltage-gated potassium channel, is significantly related to E_PC1 along with a number of electrophysiological features in our analysis, including maximum firing frequency ($q = 0.0002$; Fig 7A). This finding of a relationship between *Kcna1* expression and maximum firing frequency is consistent with a published study on the same gene. Kopp-Scheinflug et al.

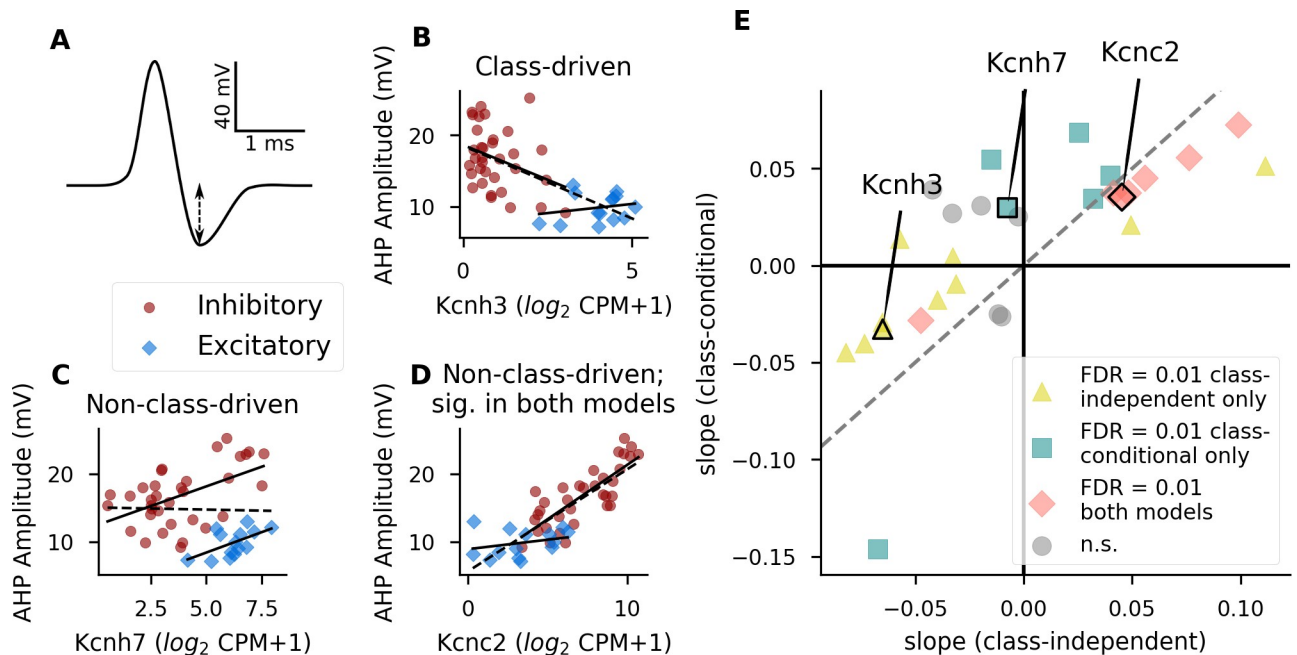


Fig 6. Accounting for cell class changes the interpretation of the relationship between potassium channel expression and after-hyperpolarization amplitude (A) Schematic view of an action potential trace, with the dashed line representing the AHP amplitude value. (B-D) Examples of voltage-gated potassium channel genes significantly associated with AHP amplitude in the class-independent model (B), the class-conditional model (C), or both (D) at a threshold of FDR = 0.01. Solid lines indicate a linear fit including only excitatory or only inhibitory cell types, and dashed line indicates a linear fit including all cell types. (E) Comparison of class-independent and class-conditional approaches for detecting associations between voltage-gated potassium channels and AHP amplitude. Each point indicates a single gene, and x and y axes are the slopes from the class-independent and class-conditional models, respectively. Labeled points are the example genes shown in B-D. Dashed line indicates identity.

<https://doi.org/10.1371/journal.pcbi.1007113.g006>

(2003) examined mice with a knockout of the *Kcna1* gene and found that firing rates in auditory neurons were reduced in the knockouts only at high intensities of an auditory stimulus, and that this difference was more robust in the inhibitory neurons of the medial nucleus of the trapezoid body (MNTB) compared to excitatory ventral cochlear nucleus (VCN) bushy cells [7].

Expression of *Scn1b*, a voltage-gated sodium channel subunit, shows a negative relationship with action potential half-width in the class-conditional model ($q = 0.0008$; Fig 7B), and is also related to a number of other properties including *E_PC1* and *E_PC2*. This relationship is obscured in the class-independent model due to overall longer half-widths in excitatory cell types. Consistent with the idea that *Scn1b* might function to shorten AP half-widths, layer 5 cortical pyramidal neurons from mice lacking the *Scn1b* gene show longer half-widths than controls, due to changes in protein stability of voltage-gated potassium channels [36].

Interestingly, the *Lrrk2* gene, mutations in which contribute to Parkinson's disease [37], is positively correlated with neurite branchiness (number of branch points per μm) in the class-conditional model, but not the class-independent model ($q = 0.049$; Fig 7C). *Lrrk2* has been shown by several studies to regulate neurite outgrowth and branching in cultures [38–41].

We do find some examples in which the reported direction of a gene-property relationship in the literature is inconsistent with the direction predicted based on our class-conditional model. For example, we found a positive correlation between expression of the potassium channel subunit gene *Kcnab2* and maximum firing frequency ($q = 1.8 \times 10^{-5}$; Fig 7D). In contrast to this, mice with a knockout of the same gene exhibit hyperexcitability in neurons of the amygdala [42]. This discrepancy could reflect a role of *Kcnab2* in limiting spiking within fast-

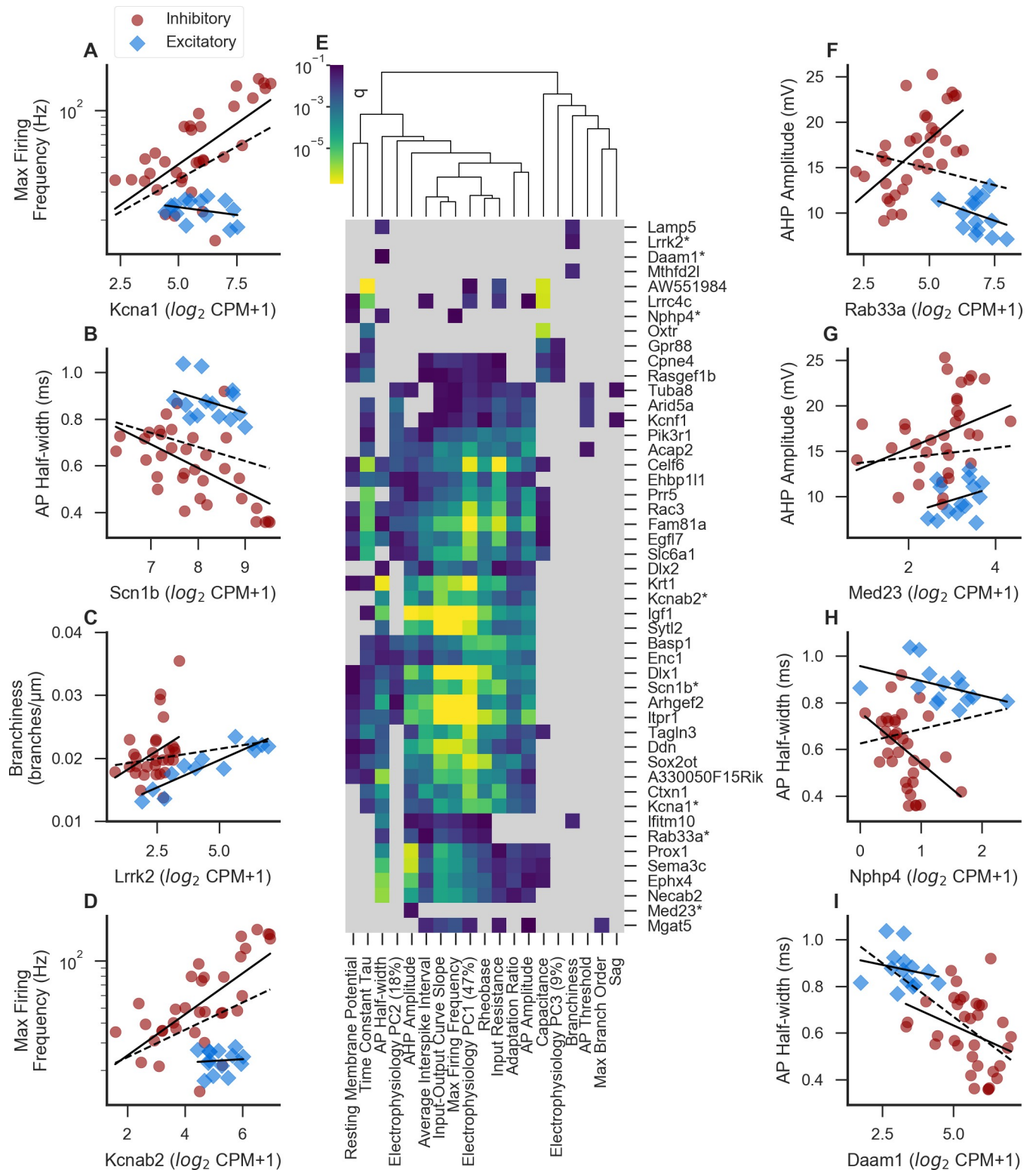


Fig 7. Examples of experimentally supported or otherwise potentially interesting genes (A-D) Examples of genes showing statistically-significant gene-property relationships in the class-conditional model (at FDR = 0.1) that also have experimental support for their causal regulation of the property in the literature. Solid lines indicate linear fits including only excitatory or only inhibitory cell types, and dashed line indicates a linear fit including all cell types (also applies to F-I). (E) Heatmap showing a subset of the most significant genes for each property in the class-conditional model, sorted along both axes by similarity. Dendrogram represents cross-property similarity between the significance levels for the genes shown here; properties appearing closely linked in the

dendrogram are those which are strongly associated with the same genes in our analysis. For each property, up to 3 top genes were chosen that were significant (at FDR = 0.1) in the class-conditional model, and also non-significant (at FDR = 0.2) in both the class-independent and interaction models for the same property. In addition, genes marked by asterisks are shown here based on their known function based on the literature in addition to at least one significant result in the class-conditional model, shown as scatterplots in A-D and F-I. Light grey indicates a non-significant result in the class-conditional model (at FDR = 0.1). (F-I) Examples of under-studied but plausibly causal genes showing significant results in the class-conditional model (see text).

<https://doi.org/10.1371/journal.pcbi.1007113.g007>

spiking cells. In other words, perhaps certain cell types express high levels of *Kcnab2* because they are fast-spiking, rather than the other way around. Since our analysis is by definition correlational, we are unable to distinguish the direction of causation, if any, between gene expression and neuronal properties.

Not only do the genes discussed here provide important validation for our method, but the existence of a smooth correlation between these genes and their associated properties is potentially interesting. The previous studies cited above provide causal evidence for gene-property relationships via gain- and loss-of-function approaches, which are likely more reminiscent of pathological states than of natural variability between cell types. Our results suggest that these genes could additionally play an instructive role in setting the precise levels of electrophysiological or morphological properties between cell types under normal physiological conditions. In addition, since morphological features are in part established due to developmental gene expression patterns [43], such features may show poor correlations with mRNA sampled from adult cells. Of course, the correspondence between the causal effects of gene expression as measured by loss-of-function studies may not always correspond neatly to the correlational relationship observed between naturally-occurring cell types, as illustrated above for *Kcnab2*.

Novel gene-property relationships

In addition to those discussed above, we identified many genes whose function in regulating neuronal electrophysiology or morphology is less well characterized. These present testable hypotheses for future study. In Table 2, we list some of the top significant genes from the class-conditional model for each property, chosen based on significance levels and/or previous studies into their cellular function (also shown in Fig 7E).

One notable feature from this analysis is that many of these genes, like *Kcna1* and *Scn1b* discussed above, are significantly associated with several or many different properties. For example, maximum firing frequency, input-output curve slope, and average interspike interval show a similar pattern in the strength of their association with this set of genes. These features all measure similar aspects of neuronal function (broadly speaking, whether a neuron tends to fire rapidly or not), so it would be surprising if they did not show correlations with the same genes. Two more properties that closely share associated genes are AP half-width and AHP amplitude, which measure distinct aspects of the action potential waveform, but might share genetic underpinnings related to rapid channel opening and closing [44]. The genes most strongly associated with various electrophysiological properties tend not to show significant associations with the morphological properties of branchiness and max branch order. However, some of the genes associated with these morphological properties do show some (generally weak) associations with some electrophysiological properties (for example *Mgat5* and *Ifitm10*).

Several of the genes for which we were unable to find conclusive loss-of-function studies in the current literature (Fig 7F–7I) seem particularly intriguing, given what is known about their cellular function. In the discussion, we briefly speculate about how these genes might function as regulators of the properties with which they are associated in our analysis. However, further study will be needed to determine what role, if any, these genes play in regulating electrophysiological or morphological properties.

Table 2. Top correlated genes for each electrophysiological property. Genes marked with asterisks are significantly associated (at FDR = 0.1) with the indicated property in the class-conditional model, and selected based on their reported function in the literature. All other genes are significant (at FDR = 0.1) in the class-conditional model and non-significant (at FDR = 0.2) in both the class-independent and interaction models for the indicated property. "Direction" indicates the direction of the model slope; for example, high expression of *Daam1* in a cell type predicts a low value of AP half-width and vice versa.

Property	Gene	Gene Name	q	Direction
Electrophysiology PC3 (9%)	<i>Rasgef1b</i>	RasGEF domain family, member 1B	0.029	+
Electrophysiology PC3 (9%)	<i>Gpr88</i>	G-protein coupled receptor 88	0.036	+
Electrophysiology PC3 (9%)	<i>Cpne4</i>	copine IV	0.041	+
Electrophysiology PC2 (18%)	<i>Arid5a</i>	AT rich interactive domain 5A (MRF1-like)	0.00088	+
Electrophysiology PC2 (18%)	<i>Basp1</i>	brain abundant, membrane attached signal protein 1	0.0026	-
Electrophysiology PC2 (18%)	<i>Pik3r1</i>	phosphatidylinositol 3-kinase, regulatory subunit, polypeptide 1 (p85 alpha)	0.0026	-
Electrophysiology PC1 (47%)	<i>Igf1</i>	insulin-like growth factor 1	9.4 * 10 ⁻¹¹	-
Electrophysiology PC1 (47%)	<i>Itp1</i>	inositol 1,4,5-trisphosphate receptor 1	1.3 * 10 ⁻¹⁰	+
Electrophysiology PC1 (47%)	<i>Arhgef2</i>	rho/rac guanine nucleotide exchange factor (GEF) 2	1.4 * 10 ⁻⁸	+
Rheobase	<i>Dlx1</i>	distal-less homeobox 1	9.5 * 10 ⁻⁷	-
Rheobase	<i>Dlx2</i>	distal-less homeobox 2	0.00052	-
Rheobase	<i>Slc6a1</i>	solute carrier family 6 (neurotransmitter transporter, GABA), member 1	0.00085	+
AP Threshold	<i>Arid5a</i>	AT rich interactive domain 5A (MRF1-like)	0.013	+
AP Threshold	<i>Kcnf1</i>	potassium voltage-gated channel, subfamily F, member 1	0.013	-
AP Threshold	<i>Tuba8</i>	tubulin, alpha 8	0.019	+
AP Half-width	<i>Krt1</i>	keratin 1	2.4 * 10 ⁻⁷	+
AP Half-width	<i>Necab2</i>	N-terminal EF-hand calcium binding protein 2	1.6 * 10 ⁻⁶	+
AP Half-width	<i>Ephx4</i>	epoxide hydrolase 4	2.2 * 10 ⁻⁶	-
AP Amplitude	<i>Itp1</i>	inositol 1,4,5-trisphosphate receptor 1	2.9 * 10 ⁻⁶	-
AP Amplitude	<i>Rac3</i>	RAS-related C3 botulinum substrate 3	4.5 * 10 ⁻⁵	+
AP Amplitude	<i>Acap2</i>	ArfGAP with coiled-coil, ankyrin repeat and PH domains 2	5.3 * 10 ⁻⁵	-
AHP Amplitude	<i>Igf1</i>	insulin-like growth factor 1	6.7 * 10 ⁻⁹	-
AHP Amplitude	<i>Prox1</i>	prospero homeobox 1	4.8 * 10 ⁻⁷	-
AHP Amplitude	<i>Sema3c</i>	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3C	6.1 * 10 ⁻⁷	-
Capacitance	<i>AW551984</i>	expressed sequence AW551984	4.6 * 10 ⁻⁷	+
Capacitance	<i>Lrrc4c</i>	leucine rich repeat containing 4C	0.00000049	-
Capacitance	<i>Oxtr</i>	oxytocin receptor	0.00000081	+
Time Constant Tau	<i>Celf6</i>	CUGBP, Elav-like family member 6	1.2 * 10 ⁻⁶	+
Time Constant Tau	<i>Prr5</i>	proline rich 5 (renal)	3.7 * 10 ⁻⁶	+
Time Constant Tau	<i>Fam81a</i>	family with sequence similarity 81, member A	4.3 * 10 ⁻⁶	-
Input Resistance	<i>Ctxn1</i>	cortixin 1	7.7 * 10 ⁻⁶	+
Input Resistance	<i>Enc1</i>	ectodermal-neural cortex 1	7.7 * 10 ⁻⁵	+
Input Resistance	<i>A330050F15Rik</i>	RIKEN cDNA A330050F15 gene	0.00025	-
Resting Membrane Potential	<i>Egfl7</i>	EGF-like domain 7	0.012	+
Resting Membrane Potential	<i>Ehbp1l1</i>	EH domain binding protein 1-like 1	0.012	+
Resting Membrane Potential	<i>Tagln3</i>	transgelin 3	0.013	+
Sag	<i>Tuba8</i>	tubulin, alpha 8	0.064	-
Sag	<i>Kcnf1</i>	potassium voltage-gated channel, subfamily F, member 1	0.064	+
Average Interspike Interval	<i>Igf1</i>	insulin-like growth factor 1	2.8 * 10 ⁻⁷	+
Average Interspike Interval	<i>Arhgef2</i>	rho/rac guanine nucleotide exchange factor (GEF) 2	3.97 * 10 ⁻⁶	-

(Continued)

Table 2. (Continued)

Property	Gene	Gene Name	q	Direction
Average Interspike Interval	<i>Ddn</i>	dendrin	$1.1 * 10^{-5}$	-
Max Firing Frequency	<i>Igf1</i>	insulin-like growth factor 1	$4.6 * 10^{-12}$	-
Max Firing Frequency	<i>Itp1</i>	inositol 1,4,5-trisphosphate receptor 1	$1.9 * 10^{-9}$	+
Max Firing Frequency	<i>Arhgef2</i>	rho/rac guanine nucleotide exchange factor (GEF) 2	0.00000017	+
Input-Output Curve Slope	<i>Igf1</i>	insulin-like growth factor 1	$1.6 * 10^{-13}$	-
Input-Output Curve Slope	<i>Itp1</i>	inositol 1,4,5-trisphosphate receptor 1	$3.9 * 10^{-10}$	+
Input-Output Curve Slope	<i>Syt12</i>	synaptotagmin-like 2	0.00000034	+
Adaptation Ratio	<i>Sox2ot</i>	SOX2 overlapping transcript (non-protein coding)	0.00021	-
Adaptation Ratio	<i>Igf1</i>	insulin-like growth factor 1	0.00023	-
Adaptation Ratio	<i>Dlx1</i>	distal-less homeobox 1	0.0003	-
Branchiness	<i>Mthfd2l</i>	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 2-like	0.023	-
Branchiness	<i>Ifitm10</i>	interferon induced transmembrane protein 10	0.024	-
Branchiness	<i>Lamp5</i>	lysosomal-associated membrane protein family, member 5	0.024	+
Max Branch Order	<i>Mgat5</i>	mannoside acetylglucosaminyltransferase 5	0.017	-
Max Firing Frequency	<i>Kcna1*</i>	potassium voltage-gated channel, shaker-related subfamily, member 1	0.00018	+
AP Half-width	<i>Scn1b*</i>	sodium channel, voltage-gated, type I, beta	0.00079	-
Branchiness	<i>Lrrk2*</i>	leucine-rich repeat kinase 2	0.048	+
Max Firing Frequency	<i>Kcnab2*</i>	potassium voltage-gated channel, shaker-related subfamily, beta member 2	$1.8 * 10^{-5}$	+
AHP Amplitude	<i>Rab33a*</i>	RAB33A, member RAS oncogene family	0.004	+
AHP Amplitude	<i>Med23*</i>	mediator complex subunit 23	0.057	+
AP Half-width	<i>Nphp4*</i>	nephronophthisis 4 (juvenile) homolog (human)	0.036	-
AP Half-width	<i>Daam1*</i>	dishevelled associated activator of morphogenesis 1	0.096	-

<https://doi.org/10.1371/journal.pcbi.1007113.t002>

Discussion

In this work we presented a series of correlations between gene expression and electrophysiological or morphological properties, each representing a testable hypothesis for future studies. Our key insight here is to introduce cell class (i.e., excitatory and inhibitory cell type identity) as an indicator variable when modeling the relationship between genes and properties. This has the advantage of 1) avoiding the identification of class-driven correlations, 2) helping identify a subset of non-class-driven correlations that might have been obscured by overall differences between excitatory and inhibitory cell types, and 3) revealing instances where gene-property relationships might be different for excitatory versus inhibitory cell types.

Although the idea that non-class-driven correlations would have a higher chance of reflecting a meaningful relationship between a specific gene and property compared to class-driven ones seems straightforward, we evaluated this prediction through a number of specific empirical tests. First, we found better correspondence between gene-property relationships from the class-conditional model with those derived from the non-projecting cell type subset of our prior NeuroExpresso/NeuroElectro dataset. Second, we observed consistency between the class-conditional model and gene-property relationships derived from five independently-collected PatchSeq datasets, suggesting that the relationships described here might be predictive of gene-property relationships within narrowly-defined cell types, consistent with the hypothesis presented in Fig 1B–1D. Third, our analysis of the relationship between action potential after-hyperpolarization (AHP) amplitude and voltage-gated potassium channel genes suggests that genes and electrophysiological features showing a significant result in the class-conditional model are more likely to reflect known functions of those genes.

The PatchSeq and voltage-gated potassium channel analyses highlighted distinct advantages of the class-conditional model. The class-conditional model revealed higher overlap between the PatchSeq and AIBS datasets, compared to the class-independent model, where most shared relationships (for both models) were in a consistent direction. This indicates that the class-conditional model might be more sensitive to certain relationships, and our analysis of the PatchSeq datasets argues in favor of the biological relevance of these same relationships. In contrast, the main advantage of the class-conditional model in the voltage-gated potassium channel analysis was primarily to avoid class-driven correlations. In other words, the class-conditional model exhibits increased specificity, an important factor when considering that these results might be used to help prioritize genes for experimental study.

In this work, we have operationalized the concepts of class-driven and non-class-driven correlations as those which produce a significant result in the class-independent model only or in the class-conditional model, respectively. This is a simplification, since both effects can exist simultaneously to differing degrees (for example, *Daam1* and AP half-width, Fig 7I) and our ability to distinguish them with confidence is limited by the number and composition of cell types in the dataset. It should be emphasized that, since these categories are defined based on significance thresholds, the distinction between, for example, a non-class-driven relationship which is obscured by class and one which is significant in either model is not meaningful in a statistical sense and should not be interpreted as being directly informative about the underlying biology. Bearing this in mind, the distinction may be useful in practice for prioritizing genes for further examination. Thus, we have shown that thresholding the set of all genes based on one model or the other results in the identification of a distinct but overlapping set of genes, meaning that the choice of model is consequential.

A novel feature of our analysis is the investigation of gene-property relationships that are divergent within excitatory and inhibitory cell types. Using the interaction model, we found a small subset of genes showing significant associations in the class-conditional model that also have a significant interaction term, indicating that their relationship with the property in question is dependent on cell class. We also found another small set of gene-property relationships that have a significant term in the interaction but not the class-conditional model. In contrast to all other properties analyzed, for the properties sag and maximum branch order, the interaction model identified many more genes compared to the class-conditional model. One possible explanation is that for both of these features, the absolute slopes in excitatory cells tend to be higher than those in inhibitory cells (shown in Fig 3B for maximum branch order), suggesting either that these features might be under stronger genetic control in excitatory types compared to inhibitory, or that the genes associated with them in excitatory cell types are more readily identified by our analysis. Since this dataset contains more inhibitory than excitatory types, an inhibitory-specific relationship may be identified in the class-conditional model by virtue of the number of cell types, but an excitatory-specific relationship would likely be “diluted” by the larger number of inhibitory cell types not showing the relationship. It is also possible that, in the case of maximum branch order, this effect is partially explained by methodological differences in the dataset, since inhibitory but not excitatory morphological reconstructions contain axons in addition to dendrites [1].

Novel putative gene/electrophysiology relationships

Our primary motivation for comparing gene expression to neuronal properties is to identify candidate genes that might influence those properties. While directly testing the functional relevance of specific gene-property predictions is beyond the scope of this work, we have highlighted below some of our potentially novel findings that might be of greatest interest for further follow up.

Rab33a expression is positively correlated in the AIBS dataset with AHP amplitude with a significant interaction (Fig 7F), and also shows significant positive correlations with E_PC1, input-output curve slope, maximum firing frequency, and rheobase, and significant negative correlations with AP half-width and average interstimulus interval (ISI). *Rab33a* is a small GTPase thought to be involved in regulation of vesicle trafficking, likely at stages prior to plasma membrane docking [45,46]. One hypothesis for how *Rab33a* could regulate AHP amplitude and/or AP half-width is that *Rab33a* might facilitate the transport and/or insertion of vesicles containing voltage-gated potassium channels, or regulators thereof, into the axonal membrane, leading to narrower action potentials and larger AHPs. Our analysis of the AIBS data suggests that any effects of *Rab33a* expression on AHP amplitude would be present only in inhibitory cell types.

Med23 (also known as *Crsp3*), a subunit of the mediator complex which acts as a transcriptional co-activator for RNA polymerase II [47,48], shows a positive correlation with AHP amplitude (Fig 7G). Although the complete set of roles played by *Med23* are incompletely understood, it has been shown to modulate signaling by the BMP, Ras/ELK1, and RhoA/MAL pathways [49,50]. Thus it has the potential to regulate a variety of genes, including potentially voltage-gated potassium channels or interacting proteins thereof. Given *Med23*'s role in regulating transcription through a variety of signaling pathways, it is notable that our analysis showed only one feature with which it was convincingly associated. It is also interesting to note that mutations in *Med23* have previously been associated with intellectual disability, in some cases with a predisposition to seizures [51,52].

Expression of *Nphp4* encoding the cytoskeletal-associated protein nephrocystin-4 was negatively correlated with AP half-width (Fig 7H) as well as with resting membrane potential and maximum firing frequency. Although *Nphp4* is primarily understood for its function in the kidney, *Nphp4* mutations often cause co-morbid deficits in the nervous system [53]. Furthermore, *Nphp4* has been shown to regulate actin networks via its interaction with the polarity protein *Inturned* and with the formin *Daam1* [54]. *Daam1* is also negatively correlated with AP half-width (Fig 7I), and not significantly correlated with any other features. The actin network in the axon forms a highly regular lattice structure which includes regularly interspersed voltage-gated sodium channels [55]. A similar relationship between the actin network and other voltage-gated ion channels has not been tested, but seems plausible. A potential mechanism through which *Nphp4* and *Daam1* could regulate the shape of the action potential might involve the organization of the axonal actin network structure, which might change the local levels or relative positioning of voltage-gated ion channels, especially potassium channels, or their regulators.

Limitations and caveats

We note that the gene-property relationships reported here are by definition correlational. Demonstrating that any specific gene is involved in regulation of any electrophysiological or morphological property is beyond the scope of this work. Our goal in this study was to generate testable hypotheses which, together with the current body of published literature, will help guide future experiments. We expect that this list of putative relationships contains some proportion of causal genes, and based on our analyses expect that this proportion may be higher than that in our previous work [16]. However, causality can only be determined for a given gene and property using direct experimental methods.

Additionally, as in our prior work [16], we have limited our analyses to models in which expression levels of a single gene predict downstream properties in an approximately linear fashion, and in which that gene is regulated primarily at the transcriptional level. Some

instances of mechanisms involving interactions between multiple genes, or those involving a non-linear relationship between log-gene expression and an electrophysiological or morphological property, are likely to have been missed here. In addition, for mechanisms through which electrophysiological or morphological properties are controlled at the translational or post-translational level, our analysis is unlikely to provide insight into the gene whose product directly controls the property. However, this analysis has the power to identify transcripts whose products are involved in the translation, modification, or trafficking of proteins which in turn regulate electrophysiology or morphology.

Furthermore, the generalizability of the gene-property relationships reported here might be limited by the fact that the AIBS dataset only reflects cells sampled from the adult mouse primary visual cortex. Therefore, the relevance of our results to other brain regions depends on the assumption that many of the same genes regulate electrophysiological or morphological properties in different cell types. This assumption of generalizability across brain areas appears to be appropriate in the case of *Kcna1* and maximum firing frequency (Fig 7A and [7]). Additionally, this assumption is supported by our comparisons with the NeuroExpresso/NeuroElectro dataset and PatchSeq datasets, both of which contain cells sampled from other brain regions. However, some relationships may not generalize across brain regions due to differences in expression of other genes or the presence of post-translational modifications which modify the consequences of expressing a given gene.

An additional caveat in the interpretation of our analysis is the likely existence of what we might call “subclass-driven” relationships. In the same way that differences between excitatory and inhibitory cell types can drive apparent correlations, we might expect that differences between other categories, such as Sst and Pvalb interneurons, would have a similar effect. However, we expect this effect to be small relative to the effects of excitatory/inhibitory cell class (see Zeisel et al., 2018).

Another potential confounding factor in our reliance on the AIBS datasets is the uneven balance in the count of inhibitory versus excitatory cell types. The practical consequence of this is that the results from the class-conditional model are likely biased towards explaining gene-property relationships within inhibitory cell types, and might be missing relationships that are specific to excitatory cell types. Additionally, our validation of the class-conditional model using the NeuroExpresso/NeuroElectro and PatchSeq datasets was limited to interneurons for reasons of data availability. Even in the absence of a significant interaction term, gene-property relationships may have stronger evidence in one cell class than the other. An example of this is *Lrrk2* and branchiness (Fig 7C), where despite very similar slopes between classes and no statistical evidence of an interaction, the correlation among excitatory cells is much tighter than that among inhibitory cells. For this reason, when prioritizing genes for future study, we strongly recommend making a plot of gene, property, and cell class before concluding that the overall result is likely to apply to both classes.

Future directions

The primary goal of this project was to produce a list of genes which we can recommend for future study based on their correlations with electrophysiological and morphological properties in the AIBS dataset. We believe that some of the genes we identified are promising candidates for future study.

In order to facilitate the use of our results by others in prioritizing genes for investigation, we are providing a Jupyter Notebook file to facilitate exploration of the data (available at https://github.com/PavlidisLab/transcriptomic_correlates). We have endeavored to make this easy to use for researchers with little or no coding experience. The notebook runs using Binder

[56], which allows for interactive exploration of the data within the user's web browser without the need for any software installation. We encourage those who are interested in a particular electrophysiological or morphological property, gene, or set of genes, to explore the data and to make their own judgements as to which genes are worth following through on experimentally and which measures should be prioritized for recording. Our recommendation is to use the gene list in conjunction with other sources of information about gene function, such as Gene Ontology annotations [57,58] and previously published literature, in prioritizing genes for future study.

Materials and methods

AIBS dataset

The RNA-seq dataset from [20] was accessed via the Allen Institute for Brain Science's Cell Types database (<http://celltypes.brain-map.org/>) on June 19, 2018, and contains 15,413 cells isolated by microdissection and fluorescence-activated cell sorting from primary visual cortex of mice expressing tdTomato under the control of various Cre driver lines. Electrophysiological and morphological data were also accessed via the Allen Institute for Brain Science Cell Types database on June 21, 2018. The dataset includes electrophysiological recordings from 1920 cells, of which 1815 are reporter-positive, from the visual cortex of mice also expressing tdTomato driven by Cre, many of which are from the same lines represented in the RNA-seq dataset. A subset of these cells (509, of which 471 are reporter-positive) have morphological reconstruction data available. Cells in both the electrophysiology/morphology and RNA-seq datasets are annotated according to the cortical layer they reside in (for electrophysiology/morphology this is always a single layer, and for RNA-seq may be a single layer, subset of layers, or all layers), their Cre-line, and whether they express the reporter.

Filtering and matching datasets

Single-cell RNA-sequencing data, summarized as counts per million reads sequenced (CPM), were log₂-transformed prior to combining with electrophysiological and morphological data. Cells from the RNA-seq dataset were excluded if they were annotated as having failed quality control checks, if they were negative for expression of tdTomato, or if they were labeled as non-neuronal or unclassified. Cells in the electrophysiology/morphology dataset were excluded if they were negative for expression of tdTomato.

Electrophysiological and morphological measures

Electrophysiological data were downloaded from <http://celltypes.brain-map.org/> and summarized as described previously [16] except for the features response frequency versus stimulus intensity (input-output) curve slope, average interstimulus interval (ISI), and sag, which we did not use previously as they were not represented in the NE dataset. All three of these new features were pre-computed in the downloaded dataset. In order to include only sag values which could be meaningfully compared, any cells having a value of "vm-for-sag" (the membrane voltage at which sag values were measured) not between -90 and -110 mV, or having a resting membrane potential lower than -80 mV, were excluded from analyses of sag, but were used for analyses of other electrophysiological features. The morphological features "average_bifurcation_angle_local", "max_branch_order", "soma_surface", "total_length", and "total_volume" were pre-computed in the dataset. We defined "branchiness" according to the pre-computed feature "number_branches" divided by "total_length" as a measure of how often a given cell produces branches per unit of neurite length. For the features input resistance, tau,

capacitance, rheobase, maximum firing frequency, AHP amplitude, adaptation ratio, input-output curve slope, latency, branchiness, max branch order, total length, and total volume, values were log10-transformed prior to use in order to mitigate underlying skew or non-normality in these data values. Scaling and principal component analysis (PCA) were performed using scikit-learn [59] for both the electrophysiological and morphological property sets. The first four principal components for each set of properties were extracted, and the first three were used in downstream analyses.

Defining cell types

Cell types in the AIBS dataset were defined according to the Cre-line they were isolated from, whether they were excitatory or inhibitory, and in most cases either a single cortical layer or a range of layers. Where multiple layer dissections containing a sufficient number of cells were present for a Cre-line in the RNAseq data, we decided on whether and how to combine layers based on the following criteria: 1) producing the maximum number of cell types, 2) producing the most homogenous cell types possible, and 3) producing cell types containing sufficiently large numbers of cells in both the RNA-seq and electrophysiology or morphology datasets. The first two criteria favored splitting layers more finely, whereas the last favored combining layers. Only cell types where both datasets contained at least 6 cells (for the electrophysiology analysis) or at least 3 cells (for the morphology analysis) were included in the final analysis. Cell type definitions, along with the numbers of cells meeting the criteria for each type, are shown in [S1 Table](#).

Splitting cells from certain Cre-lines into multiple types based on their layer location and their identity as excitatory or inhibitory allowed us to increase the number of types in our analysis. Splitting cell types in this way makes biological sense in that cells isolated from the same Cre-line but different layers often belong to different transcriptomically-defined cell types. For example, cells isolated from the upper cortical layers of Sst-Cre mice primarily belong to the Sst Cbln4 type, whereas the majority of cells from lower layers belong to either the Sst Myh8 or Sst Th types [19]. We have further justified this decision based on the fact that there are frequently electrophysiological differences between cells from the same Cre-line but from different layers (examples of three electrophysiological properties are shown in [S1 Fig](#)).

After the two datasets were matched, the combined dataset contained 1359 cells belonging to 48 types with electrophysiological data, 369 cells belonging to 43 types with morphological data, and 4403 cells belonging to 50 types with RNA-seq data ([S1 Table](#)). The remaining cells in the original datasets were those whose types could not be matched, either because the Cre-line or layer they were isolated from was not sampled in the other datasets, or because the number of cells belonging to that type was below our threshold for the number of cells per type required.

Modeling the relationship between gene expression and electrophysiology/morphology

Mean expression values for each gene and mean values for each electrophysiological or morphological property were calculated for each cell type as defined above. Only genes which were expressed at a level of 1 CPM or higher in at least ten cell types were included. Out of all genes represented in the RNA-seq dataset, ~26% passed this thresholding step. For the remaining genes, and for each electrophysiological or morphological property, we fit one or more linear models relating the property (P) to expression of the gene (G) and/or cell class (C). Model 1 (P~G; “class-independent model”) attempted to explain the property based on only expression

of the gene. For genes which were expressed in both excitatory and inhibitory types, we fit three additional models. Model 2 (P~C) related property to cell class, model 3 related the electrophysiological parameter to the gene and cell class (P~G+C; “class-conditional model”), and model 4 related the electrophysiological parameter to gene, cell class, and an interaction term between gene and cell class (P~G+C+G*C; “interaction model”). Models 2 and 3, as well as models 3 and 4, were compared to one another using an ANOVA in order to determine p-values for the class-conditional relationship between gene and property and for the gene-class interaction, respectively. Additionally, we ran a version of model 1 in which we only considered inhibitory cell types, as an alternative method of accounting for cross-class differences. Beta coefficients from models 1, 3, and 4 (separately for each cell type) were recorded, as well as p-values from model 1 and from both ANOVAs. Prior to filtering for significantly-correlated genes, false discovery rate (FDR) correction was performed using the Python package `statsmodels.stats.multitest.fdr correction`, which calculates the FDR-adjusted p-value (q-value) based on the Benjamini-Hochberg method [60] using the formula $q = p * (i/m)$, followed by taking the cumulative minimum starting from the highest p-value, where p is the uncorrected p-value, i is the rank of that p-value (the most significant having a rank of 1), and m is the number of comparisons. Model 2 was also used directly to test for significant differences between cell classes in the value of each property.

Non-projecting class-specific correlations in the NeuroElectro/NeuroExpresso dataset

The NeuroElectro and NeuroExpresso datasets were described previously [16,61–63]. In order to limit the dataset to only non-projecting cell types [18], we chose cells whose major type was annotated as anything other than “Pyramidal,” “Glutamatergic,” or “MSN”. Cells of the types “Ctx Htr3a” and “Ctx Oxtr” were excluded due to their lower transcriptomic quality compared to others in the dataset [63]. After subsetting, 19 cell types remained. Average values were calculated for gene expression and electrophysiological properties across cells within a type, and linear models relating the property to expression of the gene were calculated for each combination of gene and electrophysiological property.

In order to assess cross-dataset consistency, we calculated a Spearman correlation between the beta coefficients (slopes) calculated based on the NE dataset versus those resulting from the class-independent or class-conditional model in the AIBS. If there was a significant positive correlation between the AIBS slope and the NE slope, we concluded that the results of the two analyses were consistent (although this does not imply that they were highly consistent). For those comparisons which were consistent, we considered one method to be “more consistent” than the other if the AIBS/NE correlation value was higher ($p < 0.05$ based on 100 bootstrap resamples) than that derived using the second method.

Data analysis and visualization

All statistical analyses and data visualization were performed using Jupyter Notebook [64] and Python 2.7, and the following packages: `scipy.stats`, `numpy` [65], `pandas` [66], `matplotlib` [67], `mpl_toolkits`, `matplotlib_venn`, `seaborn`, `statsmodels.stats.multitest.fdr correction`, `mygene` [68,69], `sklearn.decomposition.PCA` [59], `sklearn.preprocessing.StandardScaler`.

Bootstrapped confidence intervals and significance between models for correlations between the NE and AIBS datasets were calculated as follows: Starting with the list of paired beta coefficients for a given electrophysiological feature and model (class-independent or class-conditional), in which each pair represented a single gene and each value in that pair was calculated using one of the two datasets, a new list of paired correlation values of the same

length was calculated by resampling with replacement. A new Spearman correlation was then calculated based on the resampled list. The resampling procedure was repeated 100 times, and the upper and lower ends of the confidence intervals were calculated by finding the values at the 2.5th and 97.5th percentiles. Significance was determined by finding the difference between each pair of resampled correlations from the two models, and then again finding the values at the 2.5th and 97.5th percentiles. If this interval did not contain zero, the two consistency metrics were said to be significantly different from one another at $p < 0.05$.

Hierarchical clustering in Fig 7D was performed using the `seaborn.clustermap` tool using the “average” (UPGMA) method and the euclidean metric [70,71].

Analysis of PatchSeq datasets

Overview of datasets used. Our analysis of the PatchSeq datasets builds on our analysis described previously [27]. Here, we made use of four previously published PatchSeq datasets that have characterized interneurons of the mouse forebrain, described in detail in Table 1. (“Cadwell,” “Földy,” “Fuzik,” “Muñoz”; [12–15]). Our analysis also includes one novel dataset of 19 Pvalb-Cre positive interneurons recorded in region CA1 of the mouse hippocampus, reported here for the first time. Cells in this dataset (referred to as the Bengtsson Gonzales dataset), were treated, processed, and analyzed using the same methodology as described in [15].

Datasets were processed and normalized as described in [27] with a small number of exceptions. First, datasets employing unique molecule identifiers (UMIs), including the Fuzik, Muñoz and Bengtsson Gonzales datasets, were normalized to a total library size of two thousand UMIs per cell. Similarly, the Cadwell and Földy datasets were normalized to counts per million (CPM), to be more directly comparable with how we have normalized the AIBS datasets here. Second, because PatchSeq sampled cells varied considerably in amount of mitochondrial and other non-coding mRNAs, when normalizing cells to the total count of reads detected in each cell, we only quantified reads mapping to protein coding genes, as defined by `biomaRt` [72]. Furthermore, we used `biomaRt` to help reconcile gene names between PatchSeq datasets.

Assigning PatchSeq single cells to transcriptomically-defined cell types. We implemented a nearest-centroid classifier to map PatchSeq transcriptomes to transcriptomically defined clusters, as defined in the Tasic 2018 cortical and Muñoz-Manchado 2018 striatum reference atlases (e.g., Sst Cbln4 and Sst Myh8). Specifically, for each transcriptomically-defined cluster in these reference datasets, we first calculated the mean expression level across all cells assigned to the cluster. Next, using the two thousand most variable genes amongst inhibitory cell types in the Tasic dataset (described in the section below), we calculated the Spearman correlation of each PatchSeq cell to every cluster in the dissociated cell dataset and assigned cells to the cluster that they were most correlated with (we compared all PatchSeq datasets except the striatum Muñoz dataset to the Tasic cortical dataset). For cortical and hippocampal cell types, to increase the number of cells defined per transcriptomic type, we utilized the “subclass” mappings of each transcriptomically defined cluster using mappings provided in the Tasic 2018 dataset, mapping neurons to the Pvalb, Sst, Vip, Lamp5, and Sncg major interneuron cell types. For example, a neuron mapped to the Sst Cbln4 would belong to the Sst subclass. To estimate transcriptome quality we used the “quality score” metric from our prior analysis, using the full set of “on” and “off” marker genes.

Identifying highly variable genes per cell type. We used the ‘`decomposeVar`’ function from the ‘`scran`’ R package [73] to identify highly variable genes in each subclass in the Tasic 2018 dataset and each cell type in the Muñoz-Manchado reference datasets.

Mixed effects statistical model to identify gene-property relationships in PatchSeq cell types. We used a mixed effects model of the following form with gene expression as a fixed effect and dataset and cell type as random effects:

$$m1 = \text{ephys_prop} \sim \text{Beta} * \log_2(\text{norm_gene_expr}) + (1|\text{dataset} * \text{cell_type})$$

where we used an anova to test for the significance of the beta associated with the gene expression term by comparison to an equivalent statistical model without the gene expression term. We used the quality score as a weight in the regression analysis, and normalized these across datasets. We used the 'lmer' function within the 'lme4' R package for fitting mixed-effects models. We performed this analysis on the top 250-most variable genes per cell type and for genes that were highly variable in at least one cell type across at least 2 (of the 5 total) PatchSeq datasets used here. In addition, we did not use PatchSeq cell types where gene expression was detected in fewer than 33% of cells and with fewer than 5 cells expressing the gene.

Supporting information

S1 Fig. Justification for cell type definitions in the AIBS dataset. Cell types defined based on the same Cre line but different layers and/or excitatory/inhibitory identity show differences in electrophysiological features. Data are represented as mean \pm SEM.

(TIFF)

S2 Fig. Correspondence between principal components and raw neuronal properties. (A) Spearman correlations between the first 3 principal components (PC) calculated based on the set of electrophysiological properties (E_PC1-3) versus the properties themselves. (B) Percent overlap between each electrophysiological property with each PC. In other words, out of the genes which are significantly associated with the property at FDR = 0.1, the percentage which are also significantly associated with the indicated PC at FDR = 0.1. Properties for which there were no significant genes are not shown. (C) Vector map showing the strength of the association between each electrophysiological property and the first two PCs. Points represent individual cells types plotted according to their values of the first two PCs. (D-F) Same as A-C, but for morphological properties.

(TIFF)

S3 Fig. Interactions do not result primarily from low gene expression in one cell class.

Between-class differences in gene expression plotted against differences in gene-property slope in the interaction model for the property AHP amplitude. Each point represents a single gene; grey points do not have a significant interaction and others are colored according to their significance level in the interaction model. For clarity of visualization only a random subset of the data (10% of the total number of genes) are plotted.

(TIFF)

S1 Table. Criteria used for defining cell types from the AIBS dataset according to the cre line and layer they were isolated from as well as excitatory/inhibitory identity. For each cell type, the number of cells meeting the criteria which were profiled for each of the three data modalities are indicated. For electrophysiology and morphology, blank cells indicate that not enough cells meeting the criteria were present in that dataset, so that cell type was not included in the analysis.

(CSV)

S2 Table. Overlap between class-independent and class-conditional models. Comparison of the number of genes showing a significant result (at FDR = 0.1) for each electrophysiological

or morphological property in the class-independent or class-conditional model, and extent of overlap between these two sets of genes. Definitions of electrophysiological properties are reproduced from [16], except for input-output curve slope, latency, ISI CoV, average ISI, and sag, which are described based on the Allen Cell Types database (<http://celltypes.brain-map.org/>). Morphological features are described based on [1].
(CSV)

S3 Table. Overlap between class-conditional and interaction models. Comparison of the number of genes showing a significant result (at FDR = 0.1) for each electrophysiological or morphological property in the class-conditional or interaction model, and extent of overlap between these two sets of genes.
(CSV)

S4 Table. Listing of subclasses defined by dissociated cell single-cell RNAsequencing datasets used for mapping in PatchSeq analysis. “Muñoz-Manchado” refers to the dissociated cell dataset [15] which was used as a reference atlas to define the cell types in the PatchSeq dataset from the same work. The Allen Institute dataset [20] was used as the reference atlas for all other PatchSeq datasets, which were obtained from neocortical or hippocampal cell types.
(DOCX)

Acknowledgments

Some of the ideas explored in this work resulted from the Allen Institute for Brain Science 2017 Summer Workshop on the Dynamic Brain. We would like to thank the organizers, instructors, and participants in this event for their guidance and support. In particular, we would like to thank Agata Budzillo and Nathan Gouwens for their helpful advice and guidance, and Annie Vogel-Cierna for her invaluable input at the early stages of the project, as well as for providing feedback on the manuscript. We would like to thank Cindy-Lee Crichlow for her assistance with literature searches. We would also like to acknowledge the investigators responsible for collecting the data represented in the AIBS Cell Types Database as well as the PatchSeq datasets re-analyzed in this study. We would like to thank the Eukaryotic Single Cell Genomics facility at Science for Life Laboratory.

Author Contributions

Conceptualization: Claire Bomkamp, Shreejoy J. Tripathy, Paul Pavlidis.

Formal analysis: Claire Bomkamp, Shreejoy J. Tripathy.

Investigation: Carolina Bengtsson Gonzales.

Supervision: Jens Hjerling-Leffler, Ann Marie Craig, Paul Pavlidis.

Writing – original draft: Claire Bomkamp.

Writing – review & editing: Shreejoy J. Tripathy, Carolina Bengtsson Gonzales, Jens Hjerling-Leffler, Ann Marie Craig, Paul Pavlidis.

References

1. Gouwens NW, Sorensen SA, Berg J, Lee C, Jarsky T, Ting J, et al. Classification of electrophysiological and morphological types in mouse visual cortex. 2018; Preprint. Available from: <https://www.biorxiv.org/content/biorxiv/early/2018/07/17/368456.full.pdf>

2. Padmanabhan K, Urban NN. Intrinsic biophysical diversity decorrelates neuronal firing while increasing information content. *Nat Neurosci* [Internet]. 2010; 13(10):1276–82. Available from: <https://doi.org/10.1038/nn.2630> PMID: 20802489
3. Markram H, Muller E, Ramaswamy S, Reimann MW, Defelipe J, Hill SL, et al. Reconstruction and Simulation of Neocortical Microcircuitry. *Cell* [Internet]. 2015; 163:456–92. Available from: <https://doi.org/10.1016/j.cell.2015.09.029> PMID: 26451489
4. Connors BW, Regehr WG. Neuronal firing: Does function follow form? *Curr Biol* [Internet]. 1996; 6(12):1560–2. Available from: [https://doi.org/10.1016/S0960-9822\(02\)70771-9](https://doi.org/10.1016/S0960-9822(02)70771-9) PMID: 8994812
5. Chklovskii DB. Synaptic connectivity and neuronal morphology: Two sides of the same coin. *Neuron* [Internet]. 2004; 43(5):609–17. Available from: <https://doi.org/10.1016/j.neuron.2004.08.012> PMID: 15339643
6. Stiefel KM, Sejnowski TJ. Mapping Function Onto Neuronal Morphology. *J Neurophysiol* [Internet]. 2007; 98(1):513–26. Available from: <https://doi.org/10.1152/jn.00865.2006> PMID: 17428904
7. Kopp-Scheinflug C, Fuchs K, Lippe WR, Tempel BL, Ru R. Decreased Temporal Precision of Auditory Signaling in Kcna1-Null Mice: An Electrophysiological Study In Vivo. *J Neurosci* [Internet]. 2003; 23(27):9199–207. Available from: <https://doi.org/10.1523/JNEUROSCI.23-27-09199.2003> PMID: 14534254
8. Qin X, Jiang Y, Tse YC, Wang Y, Wong TP, Paudel HK. Early growth response 1 (Egr-1) regulates N-methyl-D-aspartate receptor (NMDAR)-dependent transcription of PSD-95 and alpha-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid receptor (AMPA) trafficking in hippocampal primary neurons. *J Biol Chem* [Internet]. 2015; 290(49):29603–16. Available from: <https://doi.org/10.1074/jbc.M115.668889> PMID: 26475861
9. Marcotti W, Corns LF, Goodyear RJ, Rzadzinska AK, Avraham KB, Steel KP, et al. The acquisition of mechano-electrical transducer current adaptation in auditory hair cells requires myosin VI. *J Physiol* [Internet]. 2016; 13:3667–81. Available from: <https://doi.org/10.1113/JP272220>
10. Santiago C, Bashaw GJ. Transcription factors and effectors that regulate neuronal morphology. *Development* [Internet]. 2014; 141(24):4667–80. Available from: <https://doi.org/10.1242/dev.110817> PMID: 25468936
11. 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* [Internet]. 2004; 14:1310–27. Available from: <https://doi.org/10.1093/cercor/bbh092> PMID: 15192011
12. 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* [Internet]. 2016; 34(2):199–203. Available from: <https://doi.org/10.1038/nbt.3445> PMID: 26689543
13. 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* [Internet]. 2016; 34(2):175–83. Available from: <https://doi.org/10.1038/nbt.3443> PMID: 26689544
14. Földy C, Darmanis S, Aoto J, Malenka RC, Quake SR, Südhof TC. Single-cell RNAseq reveals cell adhesion molecule profiles in electrophysiologically defined neurons. *Proc Natl Acad Sci* [Internet]. 2016; 113(35):E5222–31. Available from: <https://doi.org/10.1073/pnas.1610155113> PMID: 27531958
15. Muñoz-Manchado AB, Bengtsson Gonzales C, Zeisel A, Munguba H, Bekkouche B, Skene NG, et al. Diversity of Interneurons in the Dorsal Striatum Revealed by Single-Cell RNA Sequencing and Patch-Seq. *Cell Rep* [Internet]. 2018; 24(8):2179–90. Available from: <https://doi.org/10.1016/j.celrep.2018.07.053> PMID: 30134177
16. Tripathy SJ, Toker L, Li B, Crichlow C, Tebaykin D, Mancarci BO, et al. Transcriptomic correlates of neuron electrophysiological diversity. *PLoS Comput Biol* [Internet]. 2017; 13(10):1–28. Available from: <https://doi.org/10.1371/journal.pcbi.1005814>
17. Zeisel A, Muñoz-Manchado AB, Codeluppi S, Lönnerberg P, Manno G La, Juréus A, et al. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* [Internet]. 2015; 347(6226):1138–42. Available from: <https://doi.org/10.1126/science.aaa1934>
18. Zeisel A, Hochgerner H, Ernfors P, Lo P, Marklund U, Linnarsson S, et al. Molecular Architecture of the Mouse Nervous System. *Cell* [Internet]. 2018; 174:999–1014. Available from: <https://doi.org/10.1016/j.cell.2018.06.021> PMID: 30096314
19. 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* [Internet]. 2016; 19(2):335–46. Available from: <https://doi.org/10.1038/nn.4216> PMID: 26727548
20. Tasic B, Yao Z, Smith KA, Graybeck L, Nguyen TN, Bertagnolli D, et al. Shared and distinct transcriptomic cell types across neocortical areas. *Nature* [Internet]. 2018; 563:72–8. Available from: <https://doi.org/10.1038/s41586-018-0654-5> PMID: 30382198

21. Petilla Interneuron Nomenclature Group, Ascoli GA, Alonso-Nanclares L, Anderson SA, Barrionuevo G, Benavides-Piccione R, et al. Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nat Rev Neurosci* [Internet]. 2008 Jul 1; 9:557. Available from: <https://doi.org/10.1038/nrn2402> PMID: 18568015
22. Simpson EH. The Interpretation of Interaction in Contingency Tables. *J R Stat Soc Ser B* [Internet]. 1951; 13(2):238–41. Available from: <https://www.jstor.org/stable/2984065>
23. Blyth CR. On Simpson's Paradox and the Sure-Thing Principle. *J Am Stat Assoc* [Internet]. 1972; 67(338):364–6. Available from: <https://doi.org/10.1080/01621459.1972.10482387>
24. Herscovics A. Structure and function of Class I α 1,2-mannosidases involved in glycoprotein synthesis and endoplasmic reticulum quality control. *Biochimie* [Internet]. 2001; 83(8):757–62. Available from: [https://doi.org/10.1016/S0300-9084\(01\)01319-0](https://doi.org/10.1016/S0300-9084(01)01319-0) PMID: 11530208
25. Huang ZG, Liu HW, Yan ZZ, Wang S, Wang LY, Ding JP. The glycosylation of the extracellular loop of β 2 subunits diversifies functional phenotypes of BK Channels. *Channels* [Internet]. 2017; 11(2):156–66. Available from: <https://doi.org/10.1080/19336950.2016.1243631> PMID: 27690717
26. Vicente PC, Kim JY, Ha JJ, Song MY, Lee HK, Kim DH, et al. Identification and characterization of site-specific N-glycosylation in the potassium channel Kv3.1b. *J Cell Physiol* [Internet]. 2018; 233(1):549–58. Available from: <https://doi.org/10.1002/jcp.25915> PMID: 28322444
27. Tripathy SJ, Tokar L, Bomkamp C, Mancarci BO, Belmadani M, Pavlidis P. Assessing Transcriptome Quality in Patch-Seq Datasets. *Front Mol Neurosci* [Internet]. 2018; 11:363. Available from: <https://doi.org/10.3389/fnmol.2018.00363> PMID: 30349457
28. Brew HM, Hallows JL, Tempel BL. Hyperexcitability reduced low threshold potassium currents auditory neurons of mice lacking the channel subunit Kv1.1. *J Physiol*. 2003; 548(1):1–20.
29. Delprat B, Schaer D, Roy S, Wang J, Puel J-L, Geering K. FXyD6 Is a Novel Regulator of Na,K-ATPase Expressed in the Inner Ear. *J Biol Chem* [Internet]. 2007; 282(10):7450–6. Available from: <https://doi.org/10.1074/jbc.M609872200> PMID: 17209044
30. Harris KD, Hochgerner H, Skene NG, Magno L, Katona L, Bengtsson Gonzales C, et al. Classes and continua of hippocampal CA1 inhibitory neurons revealed by single-cell transcriptomics. *PLoS Biol* [Internet]. 2018; 16(6):e2006387. Available from: <https://doi.org/10.1371/journal.pbio.2006387> PMID: 29912866
31. Islam S, Kjällquist U, Moliner A, Zajac P, Fan JB, Lönnerberg P, et al. Highly multiplexed and strand-specific single-cell RNA 5' end sequencing. *Nat Protoc* [Internet]. 2012; 7(5):813–28. Available from: <https://doi.org/10.1038/nprot.2012.022> PMID: 22481528
32. Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD. Reverse transcriptase template switching: A SMARTTM approach for full-length cDNA library construction. *Biotechniques* [Internet]. 2001; 30(4):892–7. Available from: <https://doi.org/10.2144/01304pf02> PMID: 11314272
33. Hodgkin AL, Huxley AF. Currents carried by sodium and potassium ions through the membrane of the giant axon of Loligo. *J Physiol* [Internet]. 1952; 116(4):449–72. Available from: <https://doi.org/10.1113/jphysiol.1952.sp004717> PMID: 14946713
34. Stuhmer W, Stocker M, Sakmann B, Seeburg P, Baumann A, Grube A, et al. Potassium channels expressed from rat brain cDNA have delayed rectifier properties. *FEBS Lett* [Internet]. 1988; 242(1):199–206. Available from: [https://doi.org/10.1016/0014-5793\(88\)81015-9](https://doi.org/10.1016/0014-5793(88)81015-9) PMID: 2462513
35. Alexander SPH, Striessnig J, Kelly E, Marrion N V., Peters JA, Faccenda E, et al. The concise guide to pharmacology 2017/18: Voltage-gated ion channels. *Br J Pharmacol* [Internet]. 2017; 174:S160–94. Available from: <https://doi.org/10.1111/bph.13884> PMID: 29055033
36. Marionneau C, Carrasquillo Y, Norris AJ, Townsend RR, Isom LL, Link AJ, et al. The Sodium Channel Accessory Subunit Nav 1 Regulates Neuronal Excitability through Modulation of Repolarizing Voltage-Gated K⁺ Channels. *J Neurosci* [Internet]. 2012; 32(17):5716–27. Available from: <https://doi.org/10.1523/JNEUROSCI.6450-11.2012> PMID: 22539834
37. Paisán-Ruiz C. LRRK2 gene variation and its contribution to Parkinson disease. *Hum Mutat* [Internet]. 2009; 30(8):1153–60. Available from: <https://doi.org/10.1002/humu.21038> PMID: 19472409
38. MacLeod D, Dowman J, Hammond R, Leete T, Inoue K, Abeliovich A. The Familial Parkinsonism Gene LRRK2 Regulates Neurite Process Morphology. *Neuron* [Internet]. 2006; 52(4):587–93. Available from: <https://doi.org/10.1016/j.neuron.2006.10.008> PMID: 17114044
39. Dächsel JC, Behrouz B, Yue M, Beevers JE, Melrose HL, Farrer MJ. A comparative study of Lrrk2 function in primary neuronal cultures Justus. *Park Relat Disord* [Internet]. 2010; 16(10):650–5. Available from: <https://doi.org/10.1016/j.parkreldis.2010.08.018>
40. Häbig K, Gellhaar S, Heim B, Djuric V, Giesert F, Wurst W, et al. LRRK2 guides the actin cytoskeleton at growth cones together with ARHGEF7 and Tropomyosin 4. *Biochim Biophys Acta—Mol Basis Dis* [Internet]. 2013; 1832(12):2352–67. Available from: <https://doi.org/10.1016/j.bbadis.2013.09.009>

41. Borgs L, Peyre E, Alix P, Hanon K, Grobarczyk B, Godin JD, et al. Dopaminergic neurons differentiating from LRRK2 G2019S induced pluripotent stem cells show early neuritic branching defects. *Sci Rep* [Internet]. 2016; 6:33377. Available from: <https://doi.org/10.1038/srep33377> PMID: 27640816
42. 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* [Internet]. 2011; 31(1):46–54. Available from: <https://doi.org/10.1523/JNEUROSCI.2634-10.2011> PMID: 21209188
43. Barnat M, Le Fric J, Benstaali C, Humbert S. Huntingtin-Mediated Multipolar-Bipolar Transition of Newborn Cortical Neurons Is Critical for Their Postnatal Neuronal Morphology. *Neuron* [Internet]. 2017; 93(1):99–114. Available from: <https://doi.org/10.1016/j.neuron.2016.11.035> PMID: 28017473
44. Bean BP. The action potential in mammalian central neurons. *Nat Rev Neurosci* [Internet]. 2007; 8:451–65. Available from: <https://doi.org/10.1038/nrn2148> PMID: 17514198
45. Tsuboi T, Fukuda M. Rab3A and Rab27A cooperatively regulate the docking step of dense-core vesicle exocytosis in PC12 cells. *J Cell Sci* [Internet]. 2006; 119(11):2196–203. Available from: <https://doi.org/10.1242/jcs.02962>
46. Nakazawa H, Sada T, Toriyama M, Tago K, Sugiura T, Fukuda M, et al. Rab33a Mediates Anterograde Vesicular Transport for Membrane Exocytosis and Axon Outgrowth. *J Neurosci* [Internet]. 2012; 32(37):12712–25. Available from: <https://doi.org/10.1523/JNEUROSCI.0989-12.2012> PMID: 22972995
47. Kelleher RJ, Flanagan PM, Kornberg RD. A Novel Mediator Between Activator Proteins and the RNA Polymerase II Transcription Apparatus. *Cell* [Internet]. 1990; 61:1209–15. Available from: [https://doi.org/10.1016/0092-8674\(90\)90685-8](https://doi.org/10.1016/0092-8674(90)90685-8) PMID: 2163759
48. Ryu S, Zhou S, Ladurner AG, Tjian R. The transcriptional cofactor complex CRSP is required for activity of the enhancer-binding protein Sp1. *Nature* [Internet]. 1999; 397(6718):446–50. Available from: <https://doi.org/10.1038/17141> PMID: 9989412
49. Yin J, Liang Y, Park JY, Chen D, Yao X, Xiao Q, et al. Mediator MED23 plays opposing roles in directing smooth muscle cell and adipocyte differentiation. *Genes Dev* [Internet]. 2012; 26(19):2192–205. Available from: <https://doi.org/10.1101/gad.192666.112> PMID: 22972934
50. Zhu W, Yao X, Liang Y, Liang D, Song L, Jing N, et al. Mediator Med23 deficiency enhances neural differentiation of murine embryonic stem cells through modulating BMP signaling. *Development* [Internet]. 2015; 142(3):465–76. Available from: <https://doi.org/10.1242/dev.112946> PMID: 25564654
51. Hashimoto S, Boissel S, Zarhrate M, Rio M, Munnich A, Egly J, et al. MED23 Mutation Links Intellectual Disability to Dysregulation of Immediate Early Gene Expression. *Science* [Internet]. 2011; 333:1161–4. Available from: <https://doi.org/10.1126/science.1206638> PMID: 21868677
52. Trehan A, Brady JM, Maduro V, Bone WP, Huang Y, Golas GA, et al. MED23-associated intellectual disability in a non-consanguineous family. *Am J Med Genet Part A* [Internet]. 2015; 167(6):1374–80. Available from: <https://doi.org/10.1002/ajmg.a.37047> PMID: 25845469
53. Hoefele J, Sudbrak R, Reinhardt R, Lehrack S, Hennig S, Imm A, et al. Mutational analysis of the NPHP4 gene in 250 patients with nephronophthisis. *Hum Mutat* [Internet]. 2005; 25(4):411. Available from: <https://doi.org/10.1002/humu.9326>
54. Yasunaga T, Hoff S, Schell C, Helmstädter M, Kretz O, Kuechlin S, et al. The polarity protein intuned links NPHP4 to daam1 to control the subapical actin network in multiciliated cells. *J Cell Biol* [Internet]. 2015; 211(5):963–73. Available from: <https://doi.org/10.1083/jcb.201502043> PMID: 26644512
55. Xu K, Zhong G, Zhuang X. Actin, Spectrin, and Associated Proteins Form a Periodic Cytoskeletal Structure in Axons. *Science* [Internet]. 2013; 339:452–6. Available from: <https://doi.org/10.1126/science.1232251> PMID: 23239625
56. Project Jupyter, Bussonnier M, Forde J, Freeman J, Granger B, Head T, et al. Binder 2.0—Reproducible, interactive, sharable environments for science at scale. In: Akici F, Lippa D, Niederhut D, Pacer M, editors. *Proceedings of the 17th Python in Science Conference* [Internet]. 2018. p. 113–20. Available from: <https://doi.org/10.25080/Majora-4af1f417-011>
57. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: Tool for the unification of biology. *Nat Genet* [Internet]. 2000; 25(1):25–9. Available from: <https://doi.org/10.1038/75556> PMID: 10802651
58. Carbon S, Dietze H, Lewis SE, Mungall CJ, Munoz-Torres MC, Basu S, et al. Expansion of the gene ontology knowledgebase and resources: The gene ontology consortium. *Nucleic Acids Res* [Internet]. 2017; 45(D1):D331–8. Available from: <https://doi.org/10.1093/nar/gkw1108> PMID: 27899567
59. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. *J Mach Learn Res* [Internet]. 2011; 12:2825–30. Available from: <http://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>

60. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* [Internet]. 1995; 57(1):289–300. Available from: <https://www.jstor.org/stable/2346101>
61. Tripathy SJ, Savitskaya J, Burton SD, Urban NN, Gerkin RC. NeuroElectro: a window to the world's neuron electrophysiology data. *Front Neuroinform* [Internet]. 2014; 8(April):1–11. Available from: <https://doi.org/10.3389/fninf.2014.00040>
62. Tripathy SJ, Burton SD, Geramita M, Gerkin XRC, Urban NN. Brain-wide analysis of electrophysiological diversity yields novel categorization of mammalian neuron types. *J Neurophysiol* [Internet]. 2015; 113:3474–89. Available from: <https://doi.org/10.1152/jn.00237.2015> PMID: 25810482
63. Mancarci BO, Toker L, Tripathy SJ, Li B, Rocco B, Sibille E, et al. Cross-Laboratory Analysis of Brain Cell Type Transcriptomes with Applications to Interpretation of Bulk Tissue Data. *Eneuro* [Internet]. 2017; 4(6):e0212–17.2017. Available from: <https://doi.org/10.1523/ENEURO.0212-17.2017>
64. Kluyver T, Ragan-Kelley B, Pérez F, Granger B, Bussonnier M, Frederic J, et al. Jupyter Notebooks—a publishing format for reproducible computational workflows. In: *Proceedings of the 20th International Conference on Electronic Publishing* [Internet]. 2016. p. 87–90. Available from: <https://doi.org/10.3233/978-1-61499-649-1-87>
65. van der Walt S, Colbert SC, Varoquaux G. The NumPy Array: A Structure for Efficient Numerical Computation. *Comput Sci Eng* [Internet]. 2011; 13:22–30. Available from: <https://doi.org/10.1109/MCSE.2011.37>
66. McKinney W. Data Structures for Statistical Computing in Python. *Proc 9th Python Sci Conf* [Internet]. 2010;51–6. Available from: <http://conference.scipy.org/proceedings/scipy2010/mckinney.html>
67. Hunter JD. Matplotlib: A 2D Graphics Environment. *Comput Sci Eng* [Internet]. 2007; 9:90–5. Available from: <https://doi.org/10.1109/MCSE.2007.55>
68. Wu C, Macleod I, Su AI. BioGPS and MyGene.info: organizing online, gene-centric information. *Nucleic Acids Res* [Internet]. 2013; 41:561–5. Available from: <https://doi.org/10.1093/nar/gks1114>
69. Xin J, Mark A, Afrasiabi C, Tsueng G, Juchler M, Gopal N, et al. High-performance web services for querying gene and variant annotation. *Genome Biol* [Internet]. 2016; 17(91):1–7. Available from: <https://doi.org/10.1186/s13059-016-0953-9>
70. Bar-Joseph Z, Gifford DK, Jaakkola TS. Fast optimal leaf ordering for hierarchical clustering. *Bioinformatics* [Internet]. 2001; 17(SUPPL. 1):22–9. Available from: https://doi.org/10.1093/bioinformatics/17.suppl_1.S22
71. Müllner D. Modern hierarchical, agglomerative clustering algorithms. *Arxiv* [Internet]. 2011;1–29. Available from: <http://arxiv.org/abs/1109.2378>
72. Smedley D, Haider S, Durinck S, Pandini L, Provero P, Allen J, et al. The BioMart community portal: An innovative alternative to large, centralized data repositories. *Nucleic Acids Res* [Internet]. 2015; 43(W1):W589–98. Available from: <https://doi.org/10.1093/nar/gkv350> PMID: 25897122
73. Lun ATL, McCarthy DJ, Marioni JC. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Research* [Internet]. 2016; 5:2122. Available from: <https://doi.org/10.12688/f1000research.9501.2> PMID: 27909575