The Psychiatric Cell Map Initiative: A Convergent Systems Biological Approach to Illuminating Key Molecular Pathways in Neuropsychiatric Disorders

A. Jeremy Willsey,[1,2,3,*] Montana T. Morris,[2] Sheng Wang,[2] Helen R. Willsey,[1] Nawei Sun,[2] Nia Teerikorpi,[2,4] Tierney B. Baum,[2] Gerard Cagney,[5] Kevin J. Bender,[6] Tejal A. Desai,[3,7] Deepak Srivastava,[8,9,10] Graeme W. Davis,[10,11] Jennifer Doudna,[12,13,14,15,16] Edward Chang,[17] Vikaas Sohal,[1,11] Daniel H. Lowenstein,[6] Hao Li,[3,10] David Agard,[3,10] Michael J. Keiser,[2,3,7,18] Brian Shoichet,[3,18] Mark von Zastrow,[1,3,19] Lennart Mucke,[6,8] Steven Finkbeiner,[6,8,20] Li Gan,[6,8] Nenad Sestan,[21] Michael E. Ward,[22] Ruth Huttenhain,[3,8,19] Tomasz J. Nowakowski[1,23,24]

(Author list continued on next page)

[1]Department of Psychiatry, UCSF Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA 94143, USA
[2]Institute for Neurodegenerative Diseases, UCSF Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA 94143, USA
[3]Quantitative Biosciences Institute (QBI), University of California, San Francisco, San Francisco, CA 94143, USA
[4]Tetrad Graduate Program, University of California, San Francisco, San Francisco, CA 94143, USA
[5]School of Biomolecular and Biomedical Science, Conway Institute, University College Dublin, Dublin 4, Ireland
[6]Department of Neurology, UCSF Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA 94143, USA
[7]Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA 94143, USA
[8]Gladstone Institutes, San Francisco, CA 94158, USA
[9]Department of Pediatrics, University of California, San Francisco, San Francisco, CA 94143, USA
[10]Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA 94143, USA
[11]Kavli Institute for Fundamental Neuroscience, University of California, San Francisco, San Francisco, CA 94143, USA
[12]Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720, USA
[13]Department of Chemistry, University of California, Berkeley, Berkeley, CA 94720, USA
[14]Howard Hughes Medical Institute, University of California, Berkeley, Berkeley, CA, 94720, USA
[15]Innovative Genomics Institute, University of California, Berkeley, Berkeley, CA 94720, USA
[16]MBIB Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[17]Department of Neurological Surgery, UCSF Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA 94143, USA
[18]Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA 94143, USA
[19]Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, CA 94143, USA
[20]Department of Physiology, University of California, San Francisco, San Francisco, CA 94143, USA
[21]Department of Neuroscience and Kavli Institute for Neuroscience, Yale School of Medicine, New Haven, CT 06510, USA
[22]National Institute of Neurological Disorders and Stroke, NIH, Bethesda, MD 20892, USA
[23]Department of Anatomy, University of California, San Francisco, San Francisco, CA 94143, USA

(Affiliations continued on next page)

Although gene discovery in neuropsychiatric disorders, including autism spectrum disorder, intellectual disability, epilepsy, schizophrenia, and Tourette disorder, has accelerated, resulting in a large number of molecular clues, it has proven difficult to generate specific hypotheses without the corresponding datasets at the protein complex and functional pathway level. Here, we describe one path forward—an initiative aimed at mapping the physical and genetic interaction networks of these conditions and then using these maps to connect the genomic data to neurobiology and, ultimately, the clinic. These efforts will include a team of geneticists, structural biologists, neurobiologists, systems biologists, and clinicians, leveraging a wide array of experimental approaches and creating a collaborative infrastructure necessary for long-term investigation. This initiative will ultimately intersect with parallel studies that focus on other diseases, as there is a significant overlap with genes implicated in cancer, infectious disease, and congenital heart defects.

The global burden of mental illness is enormous, whether measured in health care expenditures, lost productivity, or personal suffering. Worldwide, mental and substance use disorders rank number five among the top 10 categories contributing to overall disease burden as measured by global disability-adjusted life years (DALYs) and are the leading cause of non-fatal

Hugo J. Bellen,[25,26] Loren M. Frank,[11,20,27] Mustafa K. Khokha,[28] Richard P. Lifton,[29] Martin Kampmann,[2,3,10,30] Trey Ideker,[31] Matthew W. State,[1,3] and Nevan J. Krogan[3,8,19,32,*]

[24]The Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California, San Francisco, San Francisco, CA 94143, USA

[25]Departments of Molecular and Human Genetics and Neuroscience, Neurological Research Institute at TCH, Baylor College of Medicine, Houston, TX 77030, USA

[26]Howard Hughes Medical Institute, Baylor College of Medicine, Houston, TX 77030, USA

[27]Howard Hughes Medical Institute, University of California, San Francisco, San Francisco, CA 94143, USA

[28]Pediatric Genomics Discovery Program, Departments of Pediatrics and Genetics, Yale University School of Medicine, New Haven, CT 06510, USA

[29]Laboratory of Human Genetics and Genomics, The Rockefeller University, New York, NY 10065, USA

[30]Chan Zuckerberg Biohub, San Francisco, CA 94158, USA

[31]Department of Medicine, University of California, San Diego, La Jolla, CA 92093, USA

[32]Helen Diller Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA 94143, USA

*Correspondence: jeremy.willsey@ucsf.edu (A.J.W.), nevan.krogan@ucsf.edu (N.J.K.)

https://doi.org/10.1016/j.cell.2018.06.016

burden of disease (Whiteford et al., 2013). In the United States, neuropsychiatric disorders (NPDs) as a group account for 6 of the top 30 leading contributors to DALYs, and total costs exceed those of any other area of medicine (Murray et al., 2013; Whiteford et al., 2013). This public health emergency is exacerbated by poor access to care, particularly in much of the developing world, persistent stigma that still plagues those suffering from these conditions, as well as their families, a striking lack of insight into the underlying pathobiology of these syndromes, and a limited armamentarium of efficacious treatments (Krystal and State, 2014).

Recent advances in gene discovery have set the stage for a transformation in our understanding of NPDs. A confluence of high-throughput genomic technologies, team science, and very large patient cohorts has identified dozens of definitive risk alleles and genes for many NPDs (Lehner et al., 2015). The importance of this recent shift in neuropsychiatric genetics—away from unreliable candidate gene studies to highly reproducible exome-wide and genome-wide methodologies—cannot be overstated. For the first time, the scientific community has access to an expanding set of reliable molecular clues to the etiology of NPDs including autism spectrum disorders (ASD), intellectual disability (ID), epilepsy (EP) and epileptic encephalopathies (EE), Tourette disorder (TD), attention-deficit/hyperactivity disorder (ADHD), obsessive-compulsive disorder (OCD), schizophrenia (SCZ), bipolar disorder (BD), major depressive disorder (MDD), and developmental disorders (DDs) as a group.

While there is justifiable excitement about the progress in gene discovery, translating these findings to an understanding of the underlying pathobiology of NPDs remains largely unrealized. This is due to multiple factors. First, risk-associated genetic variants likely affect brain development, an extraordinarily complex process wherein our understanding of molecular, cellular, and circuit level organization is strikingly limited (Figure 1). Furthermore, many of the causal genes likely have pleiotropic biological effects. Consequently, the pertinence of a phenotype downstream of a disease-associated perturbation in a model system is unclear. Second, the proximal cause of NPD is likely alterations in activity patterns in the neural circuits that support mental processes; linking genes to these alterations requires that we under-

stand the relevant underlying biological pathways, the corresponding cell types and neural circuits in which these pathways are present, and the complex interactions between the pathways and the circuits underlying these conditions. Only then will we be able to predict the developmental consequences of risk-associated genetic variants, as well as understand if the resulting changes are actionable later in life. This is a difficult challenge given the functional diversity of genes, cells, and circuits associated with NPDs, and our relatively limited knowledge of pathobiology with which to make sense of this diversity.

One particularly promising approach to addressing this challenge rests on the notion of convergence. A convergent framework posits that multiple diverse biological perturbations carrying risk for a given disorder are likely to converge mechanistically in the path from genome to clinical phenotype (Geschwind and State, 2015; State and Šestan, 2012; Willsey and State, 2015). Importantly, this common pathway can manifest at many levels, from gene and protein networks within a cell to common patterns of neuronal network dysfunction that manifest in the complex distributed networks of the brain. Indeed, it is now clear that across individuals, similar functional networks can be built out of cells that have widely varying patterns of gene and protein expression (Prinz et al., 2004). The converse is also likely to be true: dysfunctional networks may show common modes of dysfunction that arise from very different cellular-level pathways.

Accordingly, parallel investigations of multiple genes and mutations are critical. Ideally, such studies would yield functional data for all NPD risk genes and alleles across multiple levels of investigation, including molecular, cell taxonomic, morphological, and neural circuit. Data would include both spatial and temporal dimensions, with a particular focus on function in human brain development. Such data would likely identify the strongest points of functional convergence, the most relevant pathobiology, and an understanding of the functional connections between risk-associated genes. The utility of this knowledge is exemplified in hypertension: large effect mutations converged on salt handling in the kidney, with vectoral effects predicting direction of blood pressure, greatly informing therapy (Lifton et al., 2001).

To date, functional data of NPD risk genes remains strikingly limited across many levels, including the protein level. While
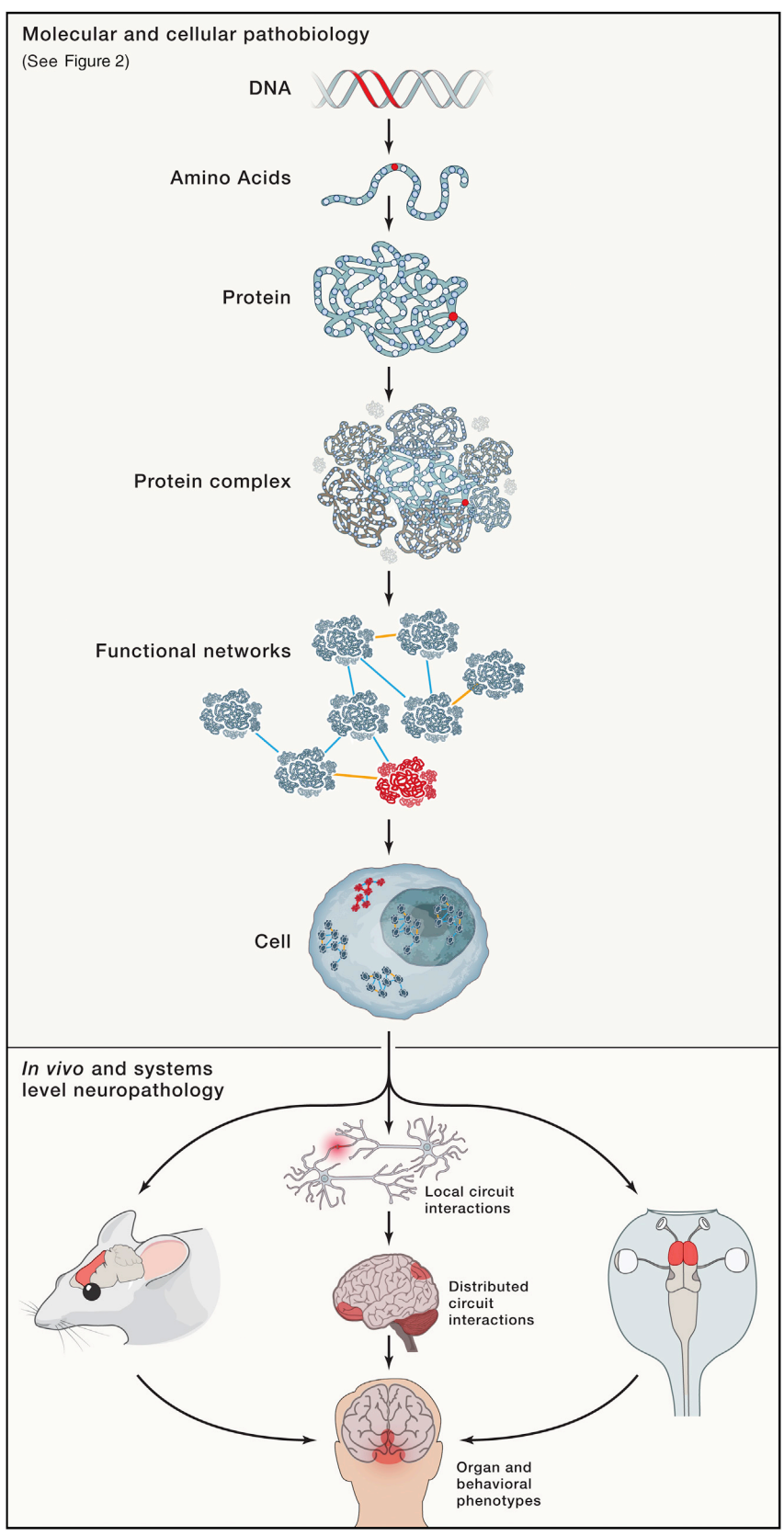
Molecular and cellular pathobiology
(See Figure 2)

DNA

Amino Acids

Protein

Protein complex

Functional networks

Cell

In vivo and systems level neuropathology

Local circuit interactions

Distributed circuit interactions

Organ and behavioral phenotypes

Figure 1. Levels of Pathogenesis and Analysis

This figure outlines the different levels on which a disorder could manifest or be investigated, starting from a genetic variant (displayed in red). It includes future parallel investigation in both *in vitro* and *in vivo* model systems across levels of complexity (.e.g. human induced pluripotent stem cells, *Xenopus tropicalis* tadpoles, mice, etc.). While much headway has been made in characterizing the genetic architecture of neuropsychiatric disorders, many of the other levels of analysis remain poorly understood. The "bottom-up" approach discussed here targets the basic levels of this hierarchy, building a strong foundation for the translation of genetics to a higher level of biological understanding. "Top-down" and "middle-out" approaches will be critical as well.

emerging datasets exist that allow one to trace the expression of any gene transcribed in the developing and adult human brain and, consequently, to investigate convergence across risk genes (Tebbenkamp et al., 2014; Willsey and State, 2015), we still lack foundational data regarding how the proteins encoded by risk-associated genes interact with other proteins, or DNA. Moreover, we do not understand how the vast majority of risk-associated mutations alter protein structure, function, and physical interactions. Because perturbations of the coding genome strongly contribute to NPD risk (Cappi et al., 2017; Deciphering Developmental Disorders Study, 2017; EuroEPINOMICS-RES Consortium et al., 2014; Fromer et al., 2014; Hamdan et al., 2014; de Ligt et al., 2012; Rauch et al., 2012; Sanders et al., 2015; Satterstrom et al., 2018; Willsey et al., 2017), determining how genetic variations impact these domains should be a high priority.

Fortunately, much of the technology needed to build such resources already exists, and therefore, it is time to establish NPD initiatives similar to the Cancer Cell Mapping Initiative (CCMI; http://www.ccmi.org) (Krogan et al., 2015) and the Host Pathogen Mapping Initiative (HPMI; http://hpmi.ucsf.edu). For example, our group has recently formed the Psychiatric Cell Map Initiative (PCMI; http://pcmi.ucsf.edu), which is focused on a number of NPDs. Each of these initiatives aim to (1) comprehensively map the networks of physical interactions among relevant proteins; (2) map the genetic interactions between risk genes; and (3) establish computational tools to reveal higher order relationships (Figure 2). As described below, we propose starting with ASD and other early-onset NPDs, due to the substantial number of risk genes of large effect already identified. We argue that these efforts are a critical component of an even broader effort that would understand how these interactions alter neuronal circuit function and behavior and how eventual integration with other data sources (e.g., patient data) has the potential to lead to new and rationally designed treatments for NPD.

The Current State of Gene Discovery in Neuropsychiatric Disorders

Unprecedented progress has been made recently in the genetics and genomics of NPDs (Figure 3A). Multiple definitive risk-carrying copy number variations (CNVs), protein-altering mutations, and/or non-coding alleles have been identified in ASD, ID, EP, EE, TD, SCZ, MDD, BD and other disorders such as structural brain abnormalities (Bilgüvar et al., 2010; Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013; Deciphering Developmental Disorders Study, 2017; Wray et al., 2018; Euro EPINOMICS-RES Consortium et al., 2014; Fromer et al., 2014; Hamdan et al., 2014; Hou et al., 2016; Huang et al., 2017; International League Against Epilepsy Consortium on Complex Epilepsies, 2014; Kataoka et al., 2016; de Ligt et al., 2012; Power et al., 2017; Rauch et al., 2012; Sanders et al., 2015; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014; Willsey et al., 2017). Additional progress is imminent in characterizing the underlying genetics of OCD (Cappi et al., 2017; International Obsessive Compulsive Disorder Foundation Genetics Collaborative [IOCDF-GC] and OCD Collaborative Genetics Association Studies [OCGAS], 2017), ADHD

(Garcia-Martínez et al., 2017; Satterstrom et al., 2018), and other NPDs. Much of this progress is due to technological and methodological advances, and parallel changes in the culture of science, such as team science and open data sharing (Lehner et al., 2015). The large-scale studies that have resulted, combined with high-throughput and cost-effective genomic assays, have been the key to reproducible and reliable gene discovery.

Success in identifying specific genetic risk factors has emerged in two broad areas: (1) discovery of rare, large effect *de novo* variants affecting the coding regions of the genome (Figure 3A), and (2) identification of common variants of small effect, often mapping to the non-coding genome. Initial insights from successful genomic investigations suggest that while each disorder involves a wide variety of risk allele types, the specific architecture of different syndromes varies considerably. For example, for ASD, ID, TD, and EE, the lion's share of progress in identifying specific loci has been via the identification of highly penetrant *de novo* protein-altering mutations and genic CNVs (Figure 3A) (EuroEPINOMICS-RES Consortium et al., 2014; Hamdan et al., 2014; Sanders et al., 2015; Willsey et al., 2017). Conversely, for SCZ, MDD, EP, and BD, much of the progress has emerged from genome-wide association studies (GWAS) focusing on common alleles of small individual effect (Wray et al., 2018; Hou et al., 2016; International League Against Epilepsy Consortium on Complex Epilepsies, 2014; Power et al., 2017; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). The trajectory of discovery for OCD, ADHD, post-traumatic stress disorder (PTSD), anxiety, and eating disorders is not yet clear. The enrichment of rare variants with large effect size within certain NPDs (ASD, TD, ID, EE) may be reflective of greater phenotypic severity and earlier onset of these disorders (i.e., in early childhood versus adulthood), the combination of which could contribute to a larger impairment of reproductive fitness. In this situation, common variants of relatively high frequency would be unlikely due to selection pressure. Nonetheless, there is striking overlap of genes with rare *de novo* damaging mutations (Cappi et al., 2017; Fromer et al., 2014; Sanders et al., 2015; Satterstrom et al., 2018) (Figure 3; Tables S2 and S3), genes found in large multigenic CNVs (Sanders et al., 2015), and common variants implicated by GWAS across these conditions (Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium, 2017; Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013; Cross-Disorder Group of the Psychiatric Genomics Consortium et al., 2013). As additional genetic data accumulate in these conditions, a growing list of risk genes will be identified, generating new opportunities to understand individual NPDs as well as the relationships between them. Given the overlap in risks already observed between these conditions, we anticipate discovery of converging pathobiological mechanisms not only within disorders, but also across them.

In ASD, for example, studies leveraging the statistical power afforded by rare *de novo* putatively damaging variants have identified more than 65 strongly associated genes (Sanders et al., 2015). The most deleterious variants (likely gene disrupting or LGD) in the highest confidence subset of these genes (n = 30) as a group confer ∼20-fold increases in risk, with LGD variants in the highest confidence genes carrying even greater risks. To
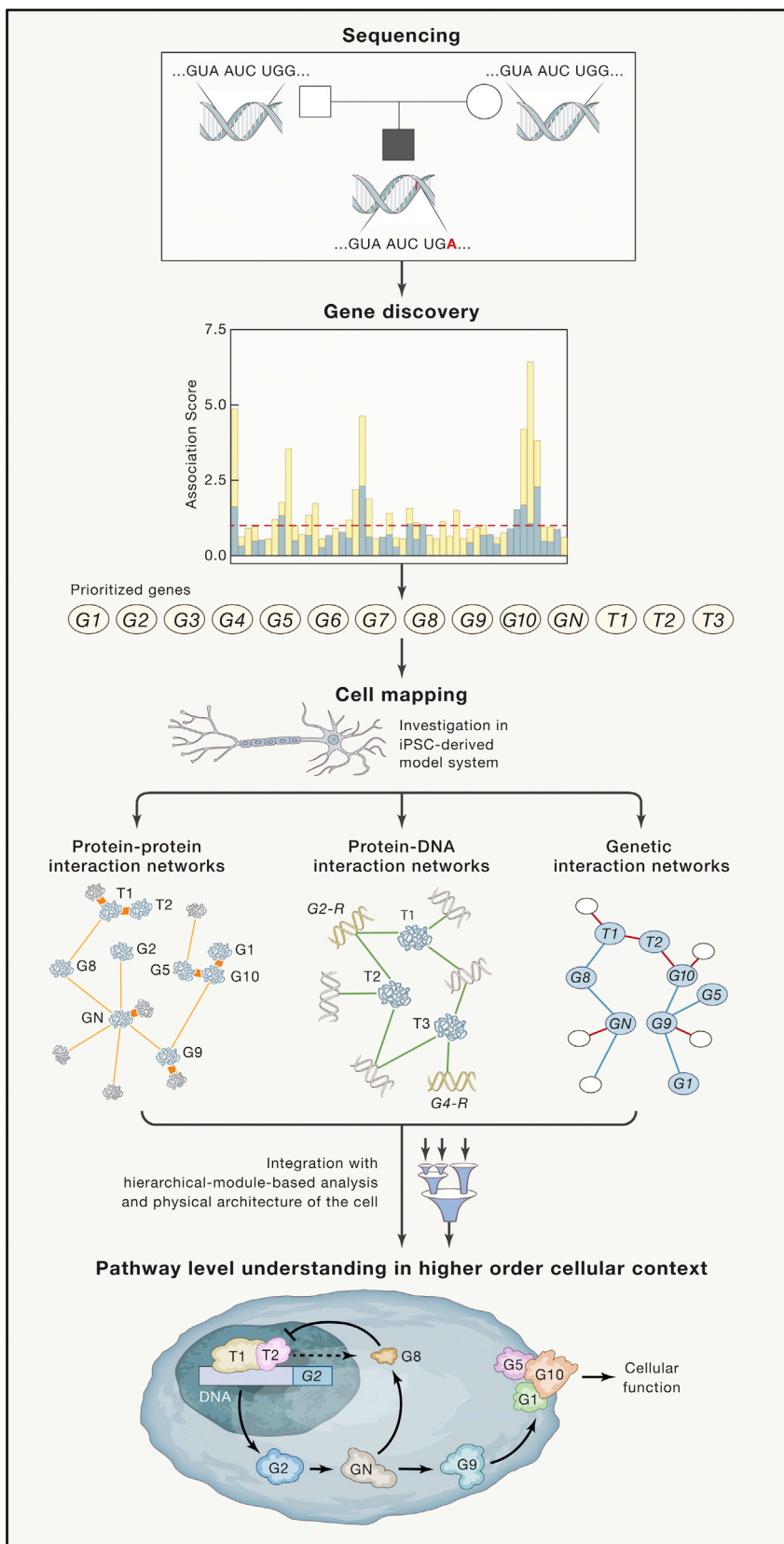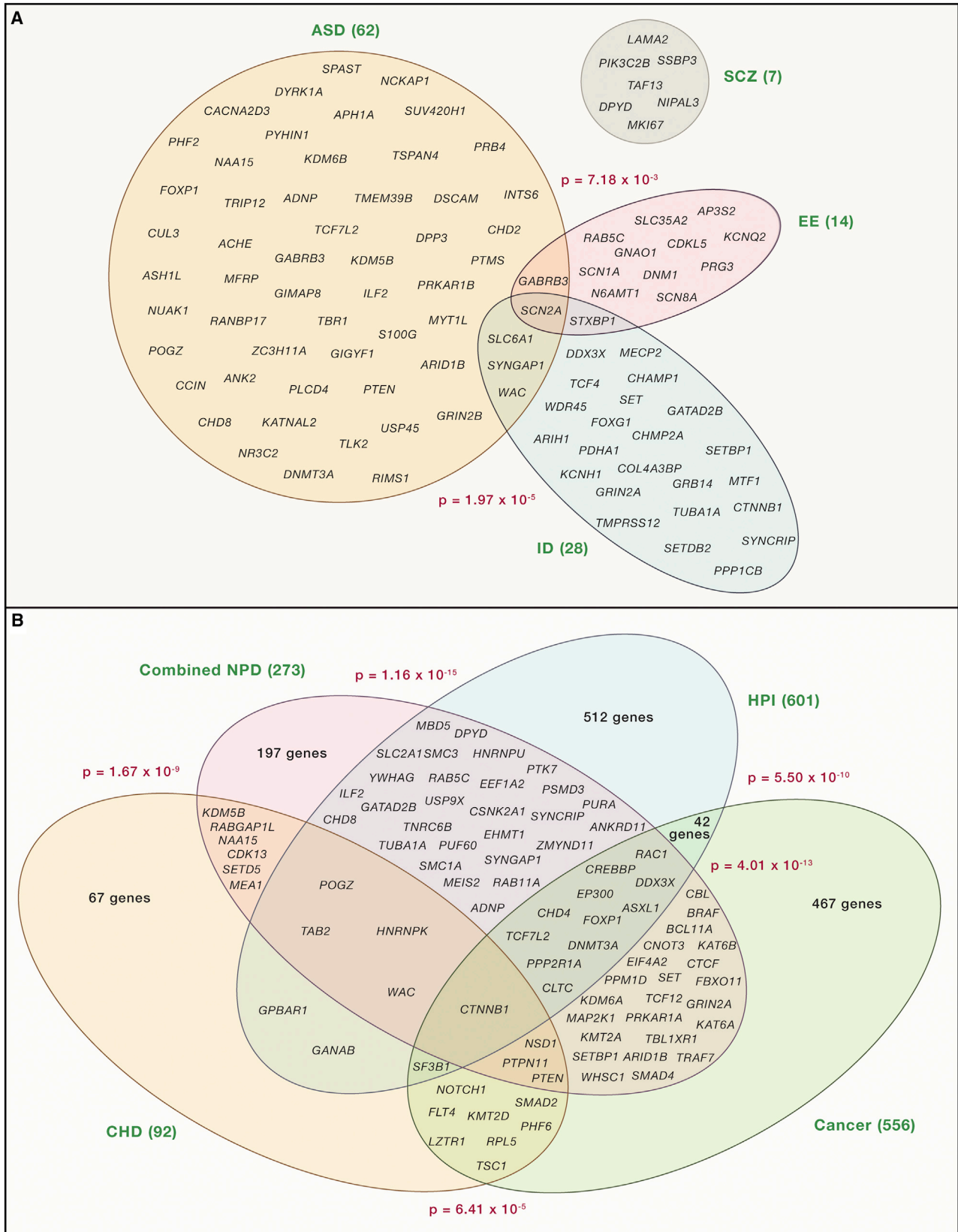
Figure 2. Conceptual Overview of a Bottom-Up Approach

The objective of a "bottom-up" approach, such as the one depicted here, is to translate genetic findings to specific insights about pathobiology of neuropsychiatric disorders. The first step is the identification of specific genes. To date, the majority of gene-level discoveries have been made through the study of high effect-size *de novo* variants (top). An example association graph is displayed in the second panel, with contributions from different classes of variants shown per gene (yellow versus blue; e.g., missense versus nonsense variants). G1–G10 represent generic genes or the proteins they encode, and T1–T3 transcription factors. "GN" represents a novel gene implicated through network analyses. There are three main types of interaction networks that could be mapped: protein-protein interaction (PPI), protein-DNA interaction (PDI), and genetic interaction (GI) networks (middle panels). The PPI networks will identify protein complexes (thick edges) as well as relationships between complexes (thin edges). Similarly, PDI networks would identify connections that correspond to putative regulatory relationships between proteins and DNA regulatory regions of other genes (e.g., "G4-R"). GI networks would identify functional relationships between genes with directionality (red indicates positive interactions, blue negative interactions). These orthogonal datasets should ultimately be integrated, in the context of a cell, to generate a pathway-level understanding (bottom panel). These hypotheses should then be followed up with cellular phenotyping and validation experiments.

A

ASD (62)

SPAST NCKAP1
DYRK1A
CACNA2D3 APH1A SUV420H1
PYHIN1
PHF2 KDM6B TSPAN4 PRB4
NAA15
FOXP1 TRIP12 ADNP TMEM39B DSCAM INTS6
CUL3 TCF7L2 DPP3 CHD2
ACHE
ASH1L GABRB3 KDM5B PTMS
MFRP
NUAK1 GIMAP8 ILF2 PRKAR1B
RANBP17 TBR1 MYT1L
POGZ S100G
ZC3H11A GIGYF1 ARID1B
CCIN ANK2 PLCD4 PTEN
CHD8 KATNAL2 GRIN2B
NR3C2 TLK2 USP45
DNMT3A RIMS1

GABRB3
SCN2A

SLC6A1
SYNGAP1
WAC

STXBP1

SCZ (7)

LAMA2
PIK3C2B SSBP3
TAF13
DPYD NIPAL3
MKI67

p = 7.18 x 10^{-3}

EE (14)

SLC35A2 AP3S2
RAB5C CDKL5 KCNQ2
GNAO1
SCN1A DNM1 PRG3
N6AMT1 SCN8A

ID (28)

DDX3X MECP2
TCF4 CHAMP1
WDR45 SET GATAD2B
FOXG1
ARIH1 CHMP2A
PDHA1 SETBP1
KCNH1 COL4A3BP GRB14 MTF1
GRIN2A CTNNB1
TMPRSS12 TUBA1A
SETDB2 SYNCRIP
PPP1CB

p = 1.97 x 10^{-5}

B

Combined NPD (273) p = 1.16 x 10^{-15} HPI (601)

512 genes

197 genes

MBD5 DPYD
SLC2A1 SMC3 HNRNPU
YWHAG RAB5C PTK7
ILF2 EEF1A2 PSMD3
GATAD2B USP9X PURA
CHD8 CSNK2A1 SYNCRIP
TNRC6B EHMT1 ANKRD11
TUBA1A PUF60 ZMYND11
SMC1A SYNGAP1
MEIS2 RAB11A
ADNP

p = 1.67 x 10^{-9}

KDM5B
RABGAP1L
NAA15
CDK13
SETD5
MEA1

67 genes

POGZ

TAB2 HNRNPK

WAC

GPBAR1

GANAB

p = 5.50 x 10^{-10}

42
genes

p = 4.01 x 10^{-13}

RAC1
CREBBP DDX3X CBL
EP300 ASXL1 BRAF
CHD4 FOXP1 BCL11A
TCF7L2 CNOT3 KAT6B
DNMT3A EIF4A2 CTCF
PPP2R1A PPM1D SET FBXO11
CLTC KDM6A TCF12 GRIN2A
MAP2K1 PRKAR1A KAT6A
NSD1 KMT2A TBL1XR1
PTPN11 SETBP1 ARID1B TRAF7
PTEN WHSC1 SMAD4

CTNNB1

467 genes

SF3B1

NOTCH1 SMAD2
FLT4 KMT2D PHF6
LZTR1 RPL5
TSC1

CHD (92)

Cancer (556)

p = 6.41 x 10^{-5}

(legend on next page)

date, there is very little evidence for a multi-hit hypothesis vis-á-vis these highly damaging *de novo* coding mutations. Indeed, a single highly protein-disrupting mutation appears sufficient to confer these very large risks, almost certainly operating in conjunction with a polygenic background of common alleles. The strong effects imparted by heterozygous LGD variants implies that haploinsufficiency is a predominant mechanism of pathogenesis. Indeed, in more than half of the top 65 ASD-associated genes heterozygous LGD mutations are predicted to produce haploinsufficient phenotypes based on the paucity of these mutations in large databases (34 of the 57 represented in the dataset or 52.31%). This represents a strong enrichment compared to the rate of 18.65% (2,984 of 15,998 genes) observed genomewide (p = 7.8 × 10^{-12}, hypergeometric test) (Cassa et al., 2017). A lack of recurrent variants at the same position also makes it unlikely that these variants are contributing risk through gain of function.

Challenges of Moving from Genes to Pathobiology

A driving motivation for gene discovery is the hope that existing knowledge about newly identified genes will inform our understanding of the underlying pathobiology of a given disorder. However, there are numerous challenges in translating genetic findings into an actionable understanding. These challenges differ somewhat based on the types of genetic changes that have been discovered. First, the number of genes involved in each disorder is quite large. For example, with respect to rare mutations of large effect there are 100s to 1,000s in ASD and TD (Sanders et al., 2015; Willsey et al., 2017). Second, disruption of each gene may impart multiple, context-dependent effects (pleiotropy) that manifest at multiple levels of organization of the brain. These effects may also show variable expressivity, particularly in the setting of dosage effects, resulting in varied phenotypic consequence. As a result, it can be difficult to pinpoint the relevant pathobiology among many potential downstream perturbations. Third, the human brain is a tremendously complex organ, and the inaccessibility of this tissue coupled with incomplete knowledge of its molecular-, cellular-, and circuit-level organization further complicates understanding the roles that specific genes play in these various levels. Together, these are substantial barriers in generating and testing pathobiological hypotheses and, ultimately, developing rational, targeted

therapeutics. Investigation of common variants of small effect suffers from all of these issues and further adds extreme polygenicity, difficulty in identifying main versus interactive effects, and linking (mostly non-coding) variants to specific genes.

Neurodegenerative diseases exemplify the challenge of translating genetic findings into specific biological underpinnings and corresponding therapeutic interventions. Tremendous progress has been made in identifying genetic risk factors, including autosomal dominant genetic determinants like *HTT* in Huntington disease (The Huntington's Disease Collaborative Research Group, 1993) or *APP*, *PSEN1*, and *PSEN2* in Alzheimer's disease (Goate et al., 1991; Levy-Lahad et al., 1995; Sherrington et al., 1995) and powerful, semi-dominant risk factors such as *APOE4* (Corder et al., 1993; Strittmatter et al., 1993). However, the development of effective treatments during the past 30 years has eluded the field of neurodegeneration despite considerable progress in understanding these conditions and the relatively simple underlying genetic architecture (Noble and Burns, 2010).

Leveraging convergent biology is critical for translating a list of genes into biological insights. The goal is to integrate a complex set of many observations from one (typically "lower") level of analysis (e.g., risk-associated genetic variants) in order to generate simpler hypotheses on another (often "higher") level of analysis (e.g., cellular processes perturbed in NPDs; Figure 1). Central to this concept is the idea that pleiotropy can be minimized by focusing on points of strong convergence, which are expected to indicate the biology most relevant to the disorder in question. For example, a superficial analysis suggests convergence of many of the 65 most strongly associated ASD genes into three distinct cellular roles: (1) chromatin modification/regulation, (2) synaptic function (De Rubeis et al., 2014), and (3) ubiquitination (unpublished data). Additionally, ASD-associated genes also show enrichment of protein-protein interactions (De Rubeis et al., 2014) and protein-DNA regulatory relationships (Cotney et al., 2015), further suggesting converging mechanisms. Finally, recent developments in systems biology approaches have identified convergence of ASD genetic findings along the spatial, temporal, and cellular axes of human brain development (Willsey and State, 2015). ASD-associated genes appear to be highly co-expressed during specific developmental epochs and in specific brain regions and cell types, most notably in the mid-fetal prefrontal cortex and glutamatergic excitatory

Figure 3. Neuropsychiatric Disorder Genes Overlap with Other Disorders and Initiatives
(A) Risk genes, identified from de novo variants, overlap across NPDs. We utilized a common pipeline to identify high confidence genes (false discovery rate [FDR] <0.1) for autism spectrum disorders (ASD), intellectual disability (ID), epileptic encephalopathies (EE), schizophrenia, (SCZ), Tourette disorder (TD), and obsessive-compulsive disorder (OCD) (Table S1). We considered de novo variants from whole exome sequencing studies only and leveraged TADA to estimate FDRs (He et al., 2013). We omitted targeted sequencing studies in order to generate a fair "exome-wide" comparison across disorders and focused on de novo coding variants only as non-coding variants are outside of the scope of this review. NPDs with 5 or more high confidence risk genes are shown (i.e., TD and OCD are excluded), and the sizes of the ovals are proportional to the number of genes identified for each disorder (total number in parentheses). ASD and ID as well as ASD and EE strongly overlap. Each p value is corrected for multiple comparisons. Some of these studies excluded probands with co-morbid NPDs. For example, in the EE studies, probands with moderate-to-severe developmental impairment, or diagnosis of autism or pervasive developmental disorder before the onset of seizures were excluded. Therefore, the observed overlaps may underestimate the extent of shared genetic etiology. Similarly, we identified these genes solely based on de novo (rare) variants. Hence, the well-documented genetic overlap at the common variant level is not reflected here.
(B) NPD genes, identified from de novo variants, overlap with other disorders and initiatives. Using the same methods as in (A), we identified risk genes for congenital heart disease (CHD). We also generated a combined list of NPD genes, corresponding to the unique set of genes from pooling (A) with the set of genes identified from analyzing the deciphering developmental disorders (DDD) studies (Table S1), which examine developmental disorders collectively. The DDD genes were omitted from (A) because many of the NPDs are represented in this cohort and therefore the relevance of overlap is unclear. Finally, we derived a list of cancer genes from the Cancer Gene Census (Futreal et al., 2004) and Host-Pathogen Interaction (HPI) genes from the HPIDB 2.0 database (Ammari et al., 2016) (Table S1; Data S1). High confidence NPD genes significantly overlap with each of the other three gene sets (Tables S2 and S3) suggesting strong synergy between NPD initiatives and the Cancer Cell Map Initiative (CCMI), the Host Pathogen Map Initiative (HPMI), and similar efforts in CHD.

neurons, particularly deep layer (layer 5/6) projection neurons (Willsey and State, 2015). While estimates of the number of genes involved in ASD risk range from hundreds to thousands (Sanders et al., 2015), these findings suggest that a large number of risk genes in ASD will converge on a smaller number of biological pathways, cell types, and modes of network dysfunction. If true, understanding the underlying pathobiology and developing effective therapeutics may be broadly feasible in ASD. In a similar vein, despite different pathological hallmarks, converging biology has emerged across various neurodegenerative diseases associated with distinct genetic variants. These common mechanisms include protein quality control and degradation mechanisms, RNA metabolism and stress granules, and innate immunity (Wang et al., 2018). However, identification of more specific pathobiological pathways is crucial for translational biology in these disorders.

The resolving power of these types of broad, annotation-focused approaches is often constrained or led astray by the assumption that proteins have mostly one function. Many physiological proteins and, perhaps even more so, pathogenic proteins, have a diversity of actions and interactions, most of which still remain unknown. Therefore, we propose that it is imperative to generate hypothesis-free data (based on hypothesis-free genetic findings) characterizing the physical and genetic interactions among these genes in cell types present in the human brain across both spatial/anatomical and temporal/developmental dimensions and, furthermore, to integrate findings across such contexts.

A Convergent Systems Biological Approach to Understanding Neuropsychiatric Disorders

The National Institute of Mental Health (NIMH) defines convergent neuroscience as an approach that aims to establish directional bridges across different levels of analysis (e.g., genetic, molecular, cellular, circuit) in order to fully explain emergent phenomena, and ultimately, pathobiology (see NIH FOA at https://grants.nih.gov/grants/guide/pa-files/PAR-17-176.html for a detailed explanation). In this framework, we view a bottom-up convergent approach as generating the foundational functional data and insights necessary to begin to bridge the knowledge gap between the genome and the clinic (Figures 1 and 2). In the initial phase, we suggest that one key area of focus should be risk genes for early-onset NPDs since rare, large effect, coding variants offer substantive advantages over the investigation of CNVs, non-coding variants, or common variants in proteomic and other functional studies. This focus avoids the challenges of interpreting multiple genes with unclear individual contributions, intergenic loci, and variants of small effect, respectively. ASD is an attractive starting point as there is an extensive list of ASD genes with rare coding variants (Figure 3A). To concretize what further steps might look like, we describe critical goals below, separated into two major categories: (1) data and insights and (2) higher-level integrations. Furthermore, we discuss potential synergies with initiatives focused on other disease areas.

Data and Insights

Interactomes vary among cell type, developmental stage, and physiological context. However, it is difficult to choose the contexts in which to characterize risk gene function in a given NPD. One attractive approach would therefore be to generate low resolution data across a diversity of cell types of the human brain (e.g., iPSC-derived neural progenitor cells, cortical excitatory neurons, inhibitory interneurons, microglia, and astrocytes), and then focus on the points (cell types) of strongest convergence (e.g., the cell type with the most significant enrichment for protein-protein interactions between risk genes). We hypothesize these will be the most pathologically relevant cell type(s). Other data can also inform the choice of cell type(s). Co-expression networks and specific expression analyses are powerful tools for identifying the context(s) most relevant to a given NPD (Willsey and State, 2015). Gene expression profiling of bulk tissue and single cells from the developing human brain, model organisms, primary human neuronal cell culture, or iPSC-derived cells will therefore be crucial data for continuing to identify important biological contexts enriched for NPD genes.

Mapping Physical Interaction Networks

A convergent bottom-up approach aims to group risk genes into pathways and subsequently understand the relationships between these pathways (Figure 2). Systematically mapping high-resolution physical interaction networks of risk genes—by characterizing the protein-protein interactions (PPI) and, where relevant, protein-DNA interactions (PDI)—will elucidate the pairwise relationships between these genes and other human genes. Understanding these specific relationships, and how they are grouped into higher-order cellular systems, will help place these genes into their specific functional roles. This level of understanding will likely generate testable hypotheses for downstream experiments as we can predict the effects of mutations in a particular gene, or the outcome of pharmaceutical manipulation of a specific pathway. Given the confounds discussed earlier, generating these data in a hypothesis-free manner, searching for the strongest points of convergence, and minimizing the extent to which this process is based on existing knowledge/assumptions is critical. Importantly, recent gene discovery studies combine data from both LGD and missense mutations, setting the stage for studies of allelic series that include both complete and partial loss of function mutations.

Indeed, missense variants could be particularly useful for characterizing shifts in interactions or structural changes in key protein complexes, which may be more illuminating than complete null alleles. Ion channels that harbor both LGDs and missense mutations are particularly attractive in this regard, as missense effects on channel function can be quantified and compared to LGD variants (e.g., SCN2A) (Ben-Shalom et al., 2017). Knowledge of physical interactions would have an additional benefit in improving gene discovery as integration of network-level data identifies additional risk genes based on their relationship to currently known risk genes (Cotney et al., 2015; Liu et al., 2014).

Protein-Protein Interaction Networks

While there are extensive PPI databases (e.g., Intomics [Li et al., 2017], Bioplex [Huttlin et al., 2015], and STRING [Szklarczyk et al., 2015]), data from human-brain-specific cells are limited. One approach to generate these data is to leverage affinity tag-purification mass spectrometry (AP-MS) to systematically map PPIs in iPSC-derived cells, followed by validation of a

subset of these interactions in animal model systems and in human primary tissue and cells. We have successfully used this system in other contexts (Davis et al., 2015; Jäger et al., 2011; Krogan et al., 2006; Mirrashidi et al., 2015). These newly generated data could be integrated with existing data to further improve discovery and to understand similarities and differences between interactions in cells of the brain versus other tissues.

Protein-DNA Interaction Networks
ASD risk genes tend to function in chromatin modification and transcriptional regulation. For example, a protein encoded by strongly associated ASD gene, CHD8, appears to regulate other ASD risk genes during a critical mid-fetal developmental window in the prefrontal cortex (Cotney et al., 2015). Chromatin immunoprecipitation sequencing (ChIP-seq)-based characterization of NPD-associated regulatory networks through systematic evaluation of implicated chromatin modifiers, transcriptional regulators, and other DNA-binding proteins in iPSC-derived cell types would therefore identify potential downstream pathways likely relevant to pathobiology. Integrating these data with PPI data would be synergistic, as we hypothesize that many of the DNA-binding proteins will associate into chromatin-modifying or transcriptional complexes (De Rubeis et al., 2014). We also anticipate that PDIs may functionally relate to PPIs (e.g., regulatory complexes may control the expression of genes encoding interacting synaptic proteins).

Mapping Genetic and Chemogenetic Interaction Networks
AP-MS captures physical interactions between proteins but misses functional relationships between network nodes that do not interact physically (Beltrao et al., 2010) (Figure 2). For example, enzymes in a metabolic pathway may have strong functional interaction without interacting physically. Furthermore, PPIs do not provide the directionality necessary to define a pathway. While PDIs do have directionality and provide important information about regulatory relationships between genes, they may be too broad to point directly to specific biological pathways. Genetic and chemogenetic interaction mapping are therefore natural complements to physical interaction datasets as they facilitate interpretation of gene-gene relationships in terms of epistasis and pathway hierarchy.

Genetic Interactions
Large-scale systematic genetic interaction (GI) maps were pioneered in model organisms such as *Saccharomyces cerevisiae* (yeast) (Beltrao et al., 2010), and have recently been implemented in mammalian cells (Bassik et al., 2013; Roguev et al., 2013), including approaches using CRISPR and CRISPRi (Boettcher et al., 2018; Du et al., 2017; Kampmann et al., 2013; Shen et al., 2017). Beyond detecting GIs between individual gene pairs, large-scale GI maps make it possible to group genes into clusters based on the similarity of their genetic interaction patterns with other genes. One can also identify additional interactions that were not evident in the primary physical interaction data. For example, GI maps have defined functionally distinct protein complexes that contain a large number of common protein subunits that had eluded definition by PPI methods (Bassik et al., 2013; Roguev et al., 2013). Thus, GI and PPI maps could provide complementary insights and can be used to recursively refine functional models.

Chemogenetic Interactions
Combining a small molecule as a perturbagen with other small molecules or with genetic perturbation is a powerful way to gain functional insight behind physical and genetic interaction maps and may suggest novel therapeutic strategies. The effects of small molecules on these systems can be considered in two main ways. First, systematic perturbation with selected small molecules and then organization of molecules by the signatures they induce (e.g., changes in gene expression, shifts in physical interactions). Leveraging system-wide shifts in co-expression or physical interactions may help to characterize previously unknown off-target effects and diverse mechanisms of action, which could otherwise confound these experiments.

Second, small molecules, targeting a specific protein, can be used as agents that are finely tunable in dosage, timing, and combination—as compared to genetic manipulations (Zhao et al., 2011). Again, careful consideration of the full spectrum of effects would be critical to reduce confounds.

Higher-Level Integrations
Hierarchical Modular Analysis
Many approaches to organize network data are based on clustering or community detection (Brohée and van Helden, 2006; Vandin et al., 2012), yielding lists of gene clusters or "modules" that roughly correspond to protein complexes or pathways. However, cells are not merely modular (e.g., a list of functions or complexes) but also hierarchical and multi-scale, in which these subsystems nest within larger units that nest within pathways and organelles. Previously, it has been shown that molecular networks embed such hierarchical structure, such that analysis of these networks can capture, to large degree, the hierarchy of biological processes and components encoded by the cell (Dutkowski et al., 2013). Indeed, hierarchies of cell components formulated entirely from systematic physical and genetic interaction networks parallel the hierarchies of cell biology built through manual curation approaches such as the Gene Ontology (GO) project (Ashburner et al., 2000), but are potentially less biased and can identify previously undocumented subsystems. Additionally, leveraging the GO as a standard reference that can be aligned to these novel hierarchies has proven to be a powerful approach to extracting meaning (Dutkowski et al., 2013).

Hence, by organizing a collection of systematic network data with these tools, constructing a complete hierarchy of neurodevelopmental processes and relating them to phenotype may be quite plausible (Yu et al., 2016). Clearly, a comprehensive hierarchical model of neurodevelopment would represent a major resource with potential to impact a broad cross-section of research. Moreover, such work would provide important proof-of-concept for automated construction of gene ontologies for a wide range of cellular processes.

Cellular Phenotypes
Hierarchical modular analysis, or other similar approaches, would help group genes (proteins) based on their functional relationships to each other. This would in turn generate testable hypotheses and anchor investigations of higher-order phenotypes (Yu et al., 2016); for example, high-resolution cellular phenotyping coupled to machine learning (Finkbeiner et al., 2015)

to investigate cell-level consequences of mutations in a functionally related group of genes. Sophisticated computational approaches such as deep learning offer exciting and relatively unbiased ways to discover novel phenotypes and reference cell maps will be critical reference datasets (e.g., Cell Atlas) (Horwitz and Johnson, 2017; Thul et al., 2017). Longitudinal single cell approaches that can capture the dynamic changes in networks and the order of events in a cascade would likely be necessary to understand the temporal dynamics of development of pathobiology. These data would help resolve changes in networks that are primary pathogenic events from those that might be adaptive or maladaptive responses (Finkbeiner et al., 2015). These experiments could be conducted in cells derived from isogenic iPSCs with induced perturbations and from patient-derived iPSCs, as well as in higher-level model systems such as 3D iPSC culture, bioengineered tissue models, primary slice tissue culture, and model organisms. The development of *in vitro* engineered models that recapitulate the microenvironmental context of cell-cell interactions would accelerate these efforts.

Interfacing with Clinical Efforts

Eventually, comparing patient-derived and typically developing tissue will be necessary to fully understand pathobiology and may also yield strategies to stratify patients into phenotypic subgroups. Appropriate clinical data will be key in order to relate molecular modules to higher level phenotypes (Figure 1). Therefore, it will be crucial to work with clinicians to collect these data during recruitment and to be able to re-contact patients.

Neural Circuitry and Systems-Level Investigation

Higher-level cell-cell interactions, as well as systems-level neural circuitry represent a promising avenue of convergent investigation, particularly for extending insights gained from a convergent bottom-up approach (Figure 1). Neuronal circuits exhibit emergent behavior that reflects the collective properties of multiple cell types, likely distributed across multiple brain regions. Furthermore, homeostatic mechanisms operating at the circuit level may alter the properties of one cell type in order to regulate activity in a distinct population (Davis, 2006; Nelson and Valakh, 2015). As a result, circuits represent logical loci for the manifestation of convergence, in which changes in diverse genes, protein networks, cell types, or developmental stages may elicit similar changes in circuit function. Identifying such convergence requires the development of specific assays of circuit function; currently, these are being developed at both the microscopic and mesoscopic scales. At the microscopic scale, rapid advances in genetically encoded calcium indicators and multielectrode recording technology have made it possible to measure simultaneous activity in many neurons at once (Akerboom et al., 2013; Jun et al., 2017). This, in turn, makes it possible to obtain a relatively unbiased screen of many aspects of circuit function that can be examined to discover conserved circuit changes (resulting from distinct etiological perturbations), which elicit common phenotypes (Luongo et al., 2016). At the mesoscopic scale, translational biomarkers observed in mouse models and human patients, e.g., synchronized oscillations within electroencephalographic (EEG) or magnetoencephalographic (MEG) recordings (Cho et al., 2015; Senkowski and Gallinat, 2015), could link dysfunction of specific cell types to behavioral phenotypes (Cho et al., 2015; Sohal, 2012). *In vitro*

systems that allow for precise patterning of cell circuits may also facilitate useful controlled studies of neuronal activity and communication. Other technologies, such as electrocorticography (ECoG), hold similar potential for phenotyping as well (Cho et al., 2015). Taken together, these emerging methodologies represent an exciting opportunity to gain further, high-level insight into the pathobiology of NPDs.

In Vivo Models

Identifying the molecular networks underlying NPD risk genes in human *in vitro* models would provide a powerful framework for understanding cell-intrinsic processes, but may lack non-autonomous, circuit-level, and developmental context. The most utilized *in vivo* model for studying NPDs is mice (*Mus musculus*) (Razafsha et al., 2013). However, hypothesis-free characterization of many genes in parallel is cost-ineffective in mouse and other mammalian systems, especially because it is largely unclear which cell types, brain regions, or time points to investigate. Therefore, we propose an iterative investigation of *in vitro* and *in vivo* systems, starting with more simple and higher-throughput model organisms (Figure 1). Recently resources have been developed to integrate data related to human genes and their variants across most model organisms (Wang et al., 2017).

The fruit fly, *Drosophila melanogaster*, offers an attractive and relatively simple *in vivo* model to tackle neurologic disease associated genes (Şentürk and Bellen, 2018). High throughput behavioral genetic screens have allowed the isolation of numerous genes including those that govern diurnal rhythmicity (Sehgal, 2017) and cause epilepsy (Parker et al., 2011). A large-scale screen for neurodevelopmental and neurodegenerative mutations by Yamamoto et al. (2014), led to the isolation of 165 fly genes, 30% of which were associated with human disease genes in the Online Mendelian Inheritance in Man (OMIM) database in 2014. This number has now risen to 55% in 2018 because of sequencing efforts including those of the Center for Mendelian Genetics (Chong et al., 2015) and the Undiagnosed Disease Network (Ramoni et al., 2017). In fruit flies, a simple and powerful strategy consists of introducing an artificial exon containing a Splice Acceptor-T2A-GAL4-polyA tail in introns of homologs of human genes using CRISPR. This nearly always creates strong loss of function mutations because of transcriptional arrest at the polyA tail. Expression of transactivator protein GAL4 by the endogenous regulatory elements of the tagged gene can then be used to assess the expression pattern of the gene with UAS-GFP (Lee et al., 2018) and rescue the mutant phenotype with UAS-homologous human-cDNA, humanizing the fly gene. This approach has already proven successful, and has provided an assessment of patient-derived variants and identified new human diseases. These and other sophisticated rescue strategies allow dissection of the basic biological functions of the human genes and can probe into molecular mechanisms using secondary screens based on unbiased protein interactions revealed by IP-MS and metabolomics (Şentürk and Bellen, 2018; Wangler et al., 2017).

Zebrafish is an alternative *in vivo* model organism that has recently emerged as a powerful system for identifying behavioral phenotypes downstream of NPD risk gene perturbation. Zebrafish exhibit complex social behaviors which can be easily assessed in the laboratory (Stewart et al., 2014). For example,

high-throughput behavioral screening identified night time hyperactivity as a phenotypic outcome following loss of the ASD risk gene *cntnap2* (Hoffman et al., 2016). In this study, large-scale drug screening then identified estrogen receptor agonists to have a behavioral signature anti-correlated with *cntnap2* knockout, with some compounds able to rescue this phenotype. A similar high-throughput screening approach using *scn1a* mutations in zebrafish identified clemizole as a potential treatment for Dravet syndrome, a common epileptic encephalopathy (Baraban et al., 2013). These kinds of large-scale screening efforts are not possible in mammalian systems, and highlight the promise of this aquatic species to reveal new behavioral insights and potential therapeutics. However, their duplicated genome complicates genetic and proteomic analyses.

The frog (*Xenopus tropicalis*) is a model organism that boasts all the advantages of zebrafish while being 100+ million years evolutionarily closer to humans and possessing a diploid genome (Hellsten et al., 2010; Wheeler and Brändli, 2009). Of the top 65 ASD-associated genes, 64 have well-annotated orthologs in *X. tropicalis*, with an average protein sequence identity to the human gene of 78% (NCBI Homologene). In comparison, mice have 64 of 65 with 93% identity and zebrafish have 62 of 65 with 71% identity (calculated with closest homolog). Importantly, the basic features of vertebrate brain development are conserved between frogs and humans. In fact, our understanding of neural induction comes from experiments in *Xenopus*, in which Noggin was identified as the key factor that acts by inhibiting BMP signaling (Lamb et al., 1993; Zimmerman et al., 1996). Many current protocols for inducing neural differentiation in human stem cells *in vitro* rely on this foundational knowledge. Furthermore, self-organizing structures, such as the eye, induced from human stem cells were first demonstrated in *Xenopus* (Eiraku et al., 2011; Thomsen et al., 1990). More recently, findings in *Xenopus* related to NPDs have proven translatable to other model systems, such as mice. For example, loss of FMRP (Fragile X Mental Retardation Protein) in *Xenopus* tadpoles leads to defects in neurogenesis that have also been observed in FMRP null mice (Faulkner et al., 2015; Guo et al., 2011; Luo et al., 2010).

Another advantage of studying brain development in *Xenopus* is an ability to make animals in which only half of the brain is targeted by CRISPR reagents. The other half of the brain serves as an internal control, and comparison between sides allows for the identification of subtle phenotypes that may be difficult to detect in fully mutant animals. These animals can then be observed longitudinally over the course of brain development to assay neuroanatomy (e.g., b-tubulin-GFP) (Marsh-Armstrong et al., 1999), neural activity (e.g., GcAMP6s) (National *Xenopus* Resource; http://www.mbl.edu/xenopus/tropicalis-stocks), progenitor biology (e.g., Pax6-GFP) (Hirsch et al., 2002), and synapse formation (e.g., PSD95-GFP) *in vivo*. Furthermore, like in zebrafish, drug screening is straightforward and easily scaled (Wheeler and Brändli, 2009). Finally, high-throughput behavioral analyses have recently been developed for tadpoles, which show sophisticated and predictable responses to a wide variety of external cues (Blackiston and Levin, 2012; Rothman et al., 2016).

While these models allow for high-throughput screening in a model amenable to genetics, they lack the complex circuitry of the mouse or human brain and will not capture human- or mammalian-specific interactions. Therefore, these higher throughput models could be used to build hypothesis-free, foundational knowledge regarding cell types, brain regions, and time points that are important for pathobiology. Findings from these systems could then inform targeted experiments in human iPSCs, organoids, mouse models, and human brain slice cultures. Reciprocally, hypotheses formed from experiments in the other systems could be tested quickly in *Xenopus* or *Drosophila* without major investment. Comparing physical and genetic interaction networks in human cell lines to cell lines from these model organisms, and relating these comparisons to *in vivo* phenotypes in these systems may also be a powerful approach.

Synergy with Other Initiatives

Excitingly, information flow between the PCMI and other established mapping initiatives has the potential to create exciting and productive synergies, including the repurposing of therapeutics. Many of the NPD genes harboring germline mutations are somatically mutated and drivers of cancer (38 genes are shared among 273 NPD genes and 556 Tier 1 Cancer genes, p = 4.01×10^{-13}; Figure 3B; Tables S2 and S3). This suggests that understanding the molecular networks underlying cancer could also improve our understanding of specific aspects of NPDs, and vice-versa. With respect to the HPMI, there is mounting interest in a potential link between maternal immune activation during pregnancy and the development of NPDs, including ASD and schizophrenia (Blomström et al., 2016; Jiang et al., 2016; Scola and Duong, 2017), although there is conflicting evidence (Zerbo et al., 2017). This putative link is interesting given that 43 NPD genes have also been implicated as central to virus-related host-pathogen interactions (p = 1.16×10^{-15}, n = 43/273 NPD genes, 601 total HPI genes; Figure 3B; Table S2). Thus, data and insights gained from the HPMI could be critical to understanding this potential phenomenon.

Work by the Pediatric Cardiac Genetics Consortium (PCGC) has demonstrated striking overlap of genes harboring *de novo* damaging mutations in congenital heart disease (CHD) and in NPDs, and that CHD and NPDs show a high prevalence of co-occurrence (Homsy et al., 2015; Jin et al., 2017; Zaidi et al., 2013). For example, 19 genes have *de novo* LGD mutations in both CHD and NPDs and 48 have damaging mutations (LGD + probably damaging missense variants [Adzhubei et al., 2013]) in both. These events are highly unlikely by chance (p < 10^{-6}). Similarly, we observe here that established NPD risk genes overlap significantly with CHD risk genes (p = 1.67×10^{-9}, n = 14/273 NPD genes, 92 total CHD genes; Tables S2 and S3). Genes that are mutated in both CHD and NPDs tend to be highly expressed in both developing heart and brain (Homsy et al., 2015). Moreover, CHD probands with *de novo* LGD mutations in the overlapping gene set (Figure 3B) are markedly more likely to be diagnosed with NPD, and those with LGD mutations in chromatin modifiers are particularly prone to develop NPDs (87%) (Jin et al., 2017). Importantly, because children with CHD come to clinical attention in the newborn period, if not *in utero*, those with mutations that

impart high likelihood of subsequently developing ASD or other NPDs can be identified at birth, affording the opportunity for both observational studies of probands before a clinical presentation of NPD, as well as efforts to mitigate neurodevelopmental consequences via early intervention. *Xenopus* has also proven to be a powerful model system for studying overlapping NPD genes in the context of CHD (Duncan and Khokha, 2016), again suggesting synergistic overlap in understanding the pathobiology of both classes of disorders. Notably, CHD genes also overlap with genes from the other initiatives (Figure 3B).

Conclusions and Outlook

The goal of the PCMI is to uncover new molecular and functional interaction data and pathway-level insights as they relate to NPD risk genes, which would have a variety of important applications. First, these data would likely facilitate a higher-order understanding of NPDs at the molecular, cellular, and circuit level. Second, they would likely reveal new targets for future therapeutics development. There are numerous examples of pathway-level insights translating to biological mechanisms and therapies, including in cancer research (Krogan et al., 2015), infectious diseases (Shah et al., 2015), and hypertension (Lifton et al., 2001). Identification of *in vitro* biomarkers through improved pathway-level understanding would greatly help therapeutics discovery, and can enable high-throughput phenotypic screening with large-scale small molecule libraries. However, the identification of clinically reliable and relevant biomarkers will likely require additional preclinical work in model systems. Third, these data would likely feed back into gene discovery by identifying physical and genetic interaction partners not yet revealed by genomic methods. This has been explored in ASD through approaches layering genetic association data with gene co-expression and regulatory networks (Cotney et al., 2015; Ideker et al., 2011; Liu et al., 2014). Fourth, findings could be utilized in conjunction with phenotypic data to explore disorder subtypes (Yu et al., 2016).

The approach described here is not envisioned to be the only avenue of investigation of NPDs, as there are limitations. First, there are concerns about context; it is difficult to determine the appropriate biological context from the outset (e.g., cell types and time points). Furthermore, *in vitro* monoculture methods do not fully reflect the *in vivo* cellular environment in which NPDs develop, particularly given the complex nature of neural tissues. Parallel *in vivo* investigation in more simple model systems (e.g., *Drosophila*, *Xenopus*, etc) followed by testing specific hypotheses in more complex *in vivo* (e.g., mice) and *in vitro systems* (e.g., organoids, bioengineered tissue models) will help address concerns about predictive validity of models. Finally, analyses will ultimately have to include a comparison of cases and controls, but the most critical (causal) differences between these groups likely emerge during very early, preclinical stages (including those *in utero*), making access to truly informative human samples a huge challenge that remains to be addressed. As mentioned earlier, children with CHD come to clinical attention very early on, and therefore, those with mutations that substantially increase the risk for ASD or other NPDs can be identified at birth; perhaps, this may partially mitigate this limitation by facilitating ascertainment and study of probands before a clinical presentation of NPD. Alternatively, if iPSC-

derived models turn out to have good predictive validity, they may help overcome this obstacle. Nevertheless, we frame this approach as an important foundational tool for beginning to unravel NPDs, the results of which will have to be interpreted in context and confirmed in the clinic. A second set of concerns relates to scope. Investigations at the pathway level may not readily translate to higher levels of complexity, such as circuit level phenotypes or behavior. This would be especially true if strong convergence is not observed at the molecular level. Linking bottom-up approaches to other efforts to study biological convergence at higher levels—for example, "top-down" and "middle-out" *in vitro* and *in vivo* electrophysiology and systems-neuroscience—will therefore be critical.

Despite these challenges, the field is at a potential tipping point in the endeavor to understand the pathobiology of NPDs. Hypothesis-free, large-scale, genome-wide analyses are identifying risk genes and loci at an astounding rate. We have the ability to generate diverse cell types of the human brain and to model certain periods of human brain development *in vitro* and *in vivo*, CRISPR-based genetics approaches are revolutionizing the modeling of patient-derived mutations, technologies for generating high-resolution molecular data are available, and our ability to extract complex insights from these data with machine learning approaches are advancing rapidly. As a result, we now have an unprecedented opportunity to bridge the gap between gene discovery and translational biology.

AUTHOR CONTRIBUTIONS

Conceptualization, A.J.W., M.T.M., N.Sestan, M.K., T.I., M.W.S., and N.J.K.; Methodology, A.J.W., M.T.M., S.W., H.R.W., N.Sun, N.T., T.B.B., J.D., D.A., M.J.K., B.S., M.v.Z., S.F., L.G., N.Sestan, M.E.W., R.H., T.N., M.K., T.I., M.W.S., and N.J.K.; Formal Analysis, A.J.W., M.T.M., and S.W.; Investigation, A.J.W., M.T.M., and S.W.; Resources, A.J.W., L.G., N.Sestan, M.E.W., M.K.K., M.K., T.I., M.W.S., and N.J.K.; Data Curation, M.T.M. and S.W.; Writing – Original Draft, A.J.W., M.T.M., H.R.W., E.C., V.S., L.M., S.F., L.G., T.N., H.J.B.,

L.M.F., M.K.K., R.P.L., M.K., T.I., M.W.S., and N.J.K.; Writing – Review & Editing, A.J.W., M.T.M., S.W., H.R.W., N.S., N.T., T.B.B., G.C., K.J.B., T.A.D., D.S., G.W.D., J.D., E.C., V.S., D.H.L., H.L., D.A., M.J.K., B.S., M.v.Z., L.M., S.F., L.G., N.S., M.E.W., R.H., T.N., H.J.B., L.M.F., M.K.K., R.P.L., M.K., T.I., M.W.S., and N.J.K.; Visualization, A.J.W., M.T.M., and N.J.K.; Supervision, A.J.W., M.W.S., and N.J.K.; Project Administration, A.J.W., T.I., M.W.S., and N.J.K.; Funding Acquisition, A.J.W., J.D., D.A., M.J.K., B.S., M.v.Z., S.F., N.Sestan, R.H., T.N., M.K., T.I., M.W.S., and N.J.K..

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Adzhubei, I., Jordan, D.M., and Sunyaev, S.R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. Curr. Protoc. Hum. Genet. *Chapter 7*, Unit7.20.

Akerboom, J., Carreras Calderón, N., Tian, L., Wabnig, S., Prigge, M., Tolö, J., Gordus, A., Orger, M.B., Severi, K.E., Macklin, J.J., et al. (2013). Genetically encoded calcium indicators for multi-color neural activity imaging and combination with optogenetics. Front. Mol. Neurosci. *6*, 2.

Ammari, M.G., Gresham, C.R., McCarthy, F.M., and Nanduri, B. (2016). HPIDB 2.0: a curated database for host-pathogen interactions. Database (Oxford). https://doi.org/10.1093/database/baw103.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al.; The Gene Ontology Consortium (2000). Gene ontology: tool for the unification of biology. Nat. Genet. *25*, 25–29.

Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium (2017). Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. Mol. Autism *8*, 21.

Baraban, S.C., Dinday, M.T., and Hortopan, G.A. (2013). Drug screening in Scn1a zebrafish mutant identifies clemizole as a potential Dravet syndrome treatment. Nat. Commun. *4*, 2410.

Bassik, M.C., Kampmann, M., Lebbink, R.J., Wang, S., Hein, M.Y., Poser, I., Weibezahn, J., Horlbeck, M.A., Chen, S., Mann, M., et al. (2013). A systematic mammalian genetic interaction map reveals pathways underlying ricin susceptibility. Cell *152*, 909–922.

Beltrao, P., Cagney, G., and Krogan, N.J. (2010). Quantitative genetic interactions reveal biological modularity. Cell *141*, 739–745.

Ben-Shalom, R., Keeshen, C.M., Berrios, K.N., An, J.Y., Sanders, S.J., and Bender, K.J. (2017). Opposing effects on Na$_V$1.2 function underlie differences between SCN2A variants observed in individuals with autism spectrum disorder or infantile seizures. Biol. Psychiatry *82*, 224–232.

Bilgüvar, K., Oztürk, A.K., Louvi, A., Kwan, K.Y., Choi, M., Tatli, B., Yalnizoğlu, D., Tüysüz, B., Cağlayan, A.O., Gökben, S., et al. (2010). Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. Nature *467*, 207–210.

Blackiston, D.J., and Levin, M. (2012). Aversive training methods in Xenopus laevis: general principles. Cold Spring Harb. Protoc. *2012* https://doi.org/10.1101/pdb.top068338.

Blomström, Å., Karlsson, H., Gardner, R., Jörgensen, L., Magnusson, C., and Dalman, C. (2016). Associations between maternal infection during pregnancy, childhood infections, and the risk of subsequent psychotic disorder–a Swedish cohort study of nearly 2 million individuals. Schizophr. Bull. *42*, 125–133.

Boettcher, M., Tian, R., Blau, J.A., Markegard, E., Wagner, R.T., Wu, D., Mo, X., Biton, A., Zaitlen, N., Fu, H., et al. (2018). Dual gene activation and knockout screen reveals directional dependencies in genetic networks. Nat. Biotechnol. *36*, 170–178.

Brohée, S., and van Helden, J. (2006). Evaluation of clustering algorithms for protein-protein interaction networks. BMC Bioinformatics *7*, 488.

Cappi, C., Oliphant, M.E., Peter, Z., Zai, G., Sullivan, C.A.W., Gupta, A.R., Hoffman, E.J., Virdee, M., Jeremy Willsey, A., Shavitt, R.G., et al. (2017). De novo damaging coding mutations are strongly associated with obsessive-compulsive disorder and overlap with autism. bioRxiv. https://doi.org/10.1101/127712.

Cassa, C.A., Weghorn, D., Balick, D.J., Jordan, D.M., Nusinow, D., Samocha, K.E., O'Donnell-Luria, A., MacArthur, D.G., Daly, M.J., Beier, D.R., and Sunyaev, S.R. (2017). Estimating the selective effects of heterozygous protein-truncating variants from human exome data. Nat. Genet. *49*, 806–810.

Cho, K.K., Hoch, R., Lee, A.T., Patel, T., Rubenstein, J.L., and Sohal, V.S. (2015). Gamma rhythms link prefrontal interneuron dysfunction with cognitive inflexibility in Dlx5/6(+/-) mice. Neuron *85*, 1332–1343.

Chong, J.X., Buckingham, K.J., Jhangiani, S.N., Boehm, C., Sobreira, N., Smith, J.D., Harrell, T.M., McMillin, M.J., Wiszniewski, W., Gambin, T., et al.; Centers for Mendelian Genomics (2015). The genetic basis of Mendelian phenotypes: discoveries, challenges, and opportunities. Am. J. Hum. Genet. *97*, 199–215.

Corder, E.H., Saunders, A.M., Strittmatter, W.J., Schmechel, D.E., Gaskell, P.C., Small, G.W., Roses, A.D., Haines, J.L., and Pericak-Vance, M.A. (1993). Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science *261*, 921–923.

Cotney, J., Muhle, R.A., Sanders, S.J., Liu, L., Willsey, A.J., Niu, W., Liu, W., Klei, L., Lei, J., Yin, J., et al. (2015). The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment. Nat. Commun. *6*, 6404.

Cross-Disorder Group of the Psychiatric Genomics Consortium (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet *381*, 1371–1379.

Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee, S.H., Ripke, S., Neale, B.M., Faraone, S.V., Purcell, S.M., Perlis, R.H., Mowry, B.J., Thapar, A., Goddard, M.E., Witte, J.S., et al. (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat. Genet. *45*, 984–994.

Davis, G.W. (2006). Homeostatic control of neural activity: from phenomenology to molecular design. Annu. Rev. Neurosci. *29*, 307–323.

Davis, Z.H., Verschueren, E., Jang, G.M., Kleffman, K., Johnson, J.R., Park, J., Von Dollen, J., Maher, M.C., Johnson, T., Newton, W., et al. (2015). Global mapping of herpesvirus-host protein complexes reveals a transcription strategy for late genes. Mol. Cell *57*, 349–360.

de Ligt, J., Willemsen, M.H., van Bon, B.W., Kleefstra, T., Yntema, H.G., Kroes, T., Vulto-van Silfhout, A.T., Koolen, D.A., de Vries, P., Gilissen, C., et al. (2012). Diagnostic exome sequencing in persons with severe intellectual disability. N. Engl. J. Med. *367*, 1921–1929.

De Rubeis, S., He, X., Goldberg, A.P., Poultney, C.S., Samocha, K., Cicek, A.E., Kou, Y., Liu, L., Fromer, M., Walker, S., et al.; DDD Study; Homozygosity Mapping Collaborative for Autism; UK10K Consortium (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. Nature *515*, 209–215.

Deciphering Developmental Disorders Study (2017). Prevalence and architecture of de novo mutations in developmental disorders. Nature *542*, 433–438.

Du, D., Roguev, A., Gordon, D.E., Chen, M., Chen, S.-H., Shales, M., Shen, J.P., Ideker, T., Mali, P., Qi, L.S., and Krogan, N.J. (2017). Genetic interaction mapping in mammalian cells using CRISPR interference. Nat. Methods *14*, 577–580.

Duncan, A.R., and Khokha, M.K. (2016). Xenopus as a model organism for birth defects-Congenital heart disease and heterotaxy. Semin. Cell Dev. Biol. *51*, 73–79.

Dutkowski, J., Kramer, M., Surma, M.A., Balakrishnan, R., Cherry, J.M., Krogan, N.J., and Ideker, T. (2013). A gene ontology inferred from molecular networks. Nat. Biotechnol. *31*, 38–45.

Eiraku, M., Takata, N., Ishibashi, H., Kawada, M., Sakakura, E., Okuda, S., Sekiguchi, K., Adachi, T., and Sasai, Y. (2011). Self-organizing optic-cup morphogenesis in three-dimensional culture. Nature *472*, 51–56.

EuroEPINOMICS-RES Consortium; Epilepsy Phenome/Genome Project; Epi4K Consortium (2014). De novo mutations in synaptic transmission genes including DNM1 cause epileptic encephalopathies. Am. J. Hum. Genet. *95*, 360–370.

Faulkner, R.L., Wishard, T.J., Thompson, C.K., Liu, H.-H., and Cline, H.T. (2015). FMRP regulates neurogenesis *in vivo* in *Xenopus laevis* tadpoles. eNeuro *2*, e0055.

Finkbeiner, S., Frumkin, M., and Kassner, P.D. (2015). Cell-based screening: extracting meaning from complex data. Neuron *86*, 160–174.

Fromer, M., Pocklington, A.J., Kavanagh, D.H., Williams, H.J., Dwyer, S., Gormley, P., Georgieva, L., Rees, E., Palta, P., Ruderfer, D.M., et al. (2014). De novo mutations in schizophrenia implicate synaptic networks. Nature *506*, 179–184.

Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M.R. (2004). A census of human cancer genes. Nat. Rev. Cancer *4*, 177–183.

Garcia-Martínez, I., Sánchez-Mora, C., Soler Artigas, M., Rovira, P., Pagerols, M., Corrales, M., Calvo-Sánchez, E., Richarte, V., Bustamante, M., Sunyer, J., et al. (2017). Gene-wide association study reveals RNF122 ubiquitin ligase as a novel susceptibility gene for attention deficit hyperactivity disorder. Sci. Rep. *7*, 5407.

Geschwind, D.H., and State, M.W. (2015). Gene hunting in autism spectrum disorder: on the path to precision medicine. Lancet Neurol. *14*, 1109–1120.

Goate, A., Chartier-Harlin, M.C., Mullan, M., Brown, J., Crawford, F., Fidani, L., Giuffra, L., Haynes, A., Irving, N., James, L., et al. (1991). Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. Nature *349*, 704–706.

Guo, W., Allan, A.M., Zong, R., Zhang, L., Johnson, E.B., Schaller, E.G., Murthy, A.C., Goggin, S.L., Eisch, A.J., Oostra, B.A., et al. (2011). Ablation of Fmrp in adult neural stem cells disrupts hippocampus-dependent learning. Nat. Med. *17*, 559–565.

Hamdan, F.F., Srour, M., Capo-Chichi, J.-M., Daoud, H., Nassif, C., Patry, L., Massicotte, C., Ambalavanan, A., Spiegelman, D., Diallo, O., et al. (2014). De novo mutations in moderate or severe intellectual disability. PLoS Genet. *10*, e1004772.

He, X., Sanders, S.J., Liu, L., De Rubeis, S., Lim, E.T., Sutcliffe, J.S., Schellenberg, G.D., Gibbs, R.A., Daly, M.J., Buxbaum, J.D., et al. (2013). Integrated model of de novo and inherited genetic variants yields greater power to identify risk genes. PLoS Genet. *9*, e1003671.

Hellsten, U., Harland, R.M., Gilchrist, M.J., Hendrix, D., Jurka, J., Kapitonov, V., Ovcharenko, I., Putnam, N.H., Shu, S., Taher, L., et al. (2010). The genome of the Western clawed frog Xenopus tropicalis. Science *328*, 633–636.

Hirsch, N., Zimmerman, L.B., Gray, J., Chae, J., Curran, K.L., Fisher, M., Ogino, H., and Grainger, R.M. (2002). Xenopus tropicalis transgenic lines and their use in the study of embryonic induction. Dev. Dyn. *225*, 522–535.

Hoffman, E.J., Turner, K.J., Fernandez, J.M., Cifuentes, D., Ghosh, M., Ijaz, S., Jain, R.A., Kubo, F., Bill, B.R., Baier, H., et al. (2016). Estrogens suppress a behavioral phenotype in zebrafish mutants of the autism risk gene, CNTNAP2. Neuron *89*, 725–733.

Homsy, J., Zaidi, S., Shen, Y., Ware, J.S., Samocha, K.E., Karczewski, K.J., DePalma, S.R., McKean, D., Wakimoto, H., Gorham, J., et al. (2015). De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. Science *350*, 1262–1266.

Horwitz, R., and Johnson, G.T. (2017). Whole cell maps chart a course for 21st-century cell biology. Science *356*, 806–807.

Hou, L., Bergen, S.E., Akula, N., Song, J., Hultman, C.M., Landén, M., Adli, M., Alda, M., Ardau, R., Arias, B., et al. (2016). Genome-wide association study of 40,000 individuals identifies two novel loci associated with bipolar disorder. Hum. Mol. Genet. *25*, 3383–3394.

Huang, A.Y., Yu, D., Davis, L.K., Sul, J.H., Tsetsos, F., Ramensky, V., Zelaya, I., Ramos, E.M., Osiecki, L., Chen, J.A., et al. (2017). Rare copy number variants in NRXN1 and CNTN6 increase risk for Tourette syndrome. Neuron *94*, 1101–1111.

Huttlin, E.L., Ting, L., Bruckner, R.J., Gebreab, F., Gygi, M.P., Szpyt, J., Tam, S., Zarraga, G., Colby, G., Baltier, K., et al. (2015). The BioPlex network: a systematic exploration of the human interactome. Cell *162*, 425–440.

Ideker, T., Dutkowski, J., and Hood, L. (2011). Boosting signal-to-noise in complex biology: prior knowledge is power. Cell *144*, 860–863.

International League Against Epilepsy Consortium on Complex Epilepsies. (2014). Genetic determinants of common epilepsies: a meta-analysis of genome-wide association studies. Lancet Neurol. *13*, 893–903.

International Obsessive Compulsive Disorder Foundation Genetics Collaborative (IOCDF-GC) and OCD Collaborative Genetics Association Studies (OCGAS) (2017). Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. Mol. Psychiatry *23*, 1181–1188.

Jäger, S., Cimermancic, P., Gulbahce, N., Johnson, J.R., McGovern, K.E., Clarke, S.C., Shales, M., Mercenne, G., Pache, L., Li, K., et al. (2011). Global landscape of HIV-human protein complexes. Nature *481*, 365–370.

Jiang, H.Y., Xu, L.L., Shao, L., Xia, R.M., Yu, Z.H., Ling, Z.X., Yang, F., Deng, M., and Ruan, B. (2016). Maternal infection during pregnancy and risk of autism spectrum disorders: A systematic review and meta-analysis. Brain Behav. Immun. *58*, 165–172.

Jin, S.C., Homsy, J., Zaidi, S., Lu, Q., Morton, S., DePalma, S.R., Zeng, X., Qi, H., Chang, W., Sierant, M.C., et al. (2017). Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. Nat. Genet. *49*, 1593–1601.

Jun, J.J., Steinmetz, N.A., Siegle, J.H., Denman, D.J., Bauza, M., Barbarits, B., Lee, A.K., Anastassiou, C.A., Andrei, A., Aydın, Ç., et al. (2017). Fully integrated silicon probes for high-density recording of neural activity. Nature *551*, 232–236.

Kampmann, M., Bassik, M.C., and Weissman, J.S. (2013). Integrated platform for genome-wide screening and construction of high-density genetic interaction maps in mammalian cells. Proc. Natl. Acad. Sci. USA *110*, E2317–E2326.

Kataoka, M., Matoba, N., Sawada, T., Kazuno, A.-A., Ishiwata, M., Fujii, K., Matsuo, K., Takata, A., and Kato, T. (2016). Exome sequencing for bipolar disorder points to roles of de novo loss-of-function and protein-altering mutations. Mol. Psychiatry *21*, 885–893.

Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P., et al. (2006). Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature *440*, 637–643.

Krogan, N.J., Lippman, S., Agard, D.A., Ashworth, A., and Ideker, T. (2015). The cancer cell map initiative: defining the hallmark networks of cancer. Mol. Cell *58*, 690–698.

Krystal, J.H., and State, M.W. (2014). Psychiatric disorders: diagnosis to therapy. Cell *157*, 201–214.

Lamb, T.M., Knecht, A.K., Smith, W.C., Stachel, S.E., Economides, A.N., Stahl, N., Yancopolous, G.D., and Harland, R.M. (1993). Neural induction by the secreted polypeptide noggin. Science *262*, 713–718.

Lee, P.-T., Zirin, J., Kanca, O., Lin, W.-W., Schulze, K.L., Li-Kroeger, D., Tao, R., Devereaux, C., Hu, Y., Chung, V., et al. (2018). A gene-specific *T2A-GAL4* library for *Drosophila*. eLife *7*, e35574.

Lehner, T., Senthil, G., and Addington, A.M. (2015). Convergence of advances in genomics, team science, and repositories as drivers of progress in psychiatric genomics. Biol. Psychiatry *77*, 6–14.

Levy-Lahad, E., Wasco, W., Poorkaj, P., Romano, D.M., Oshima, J., Pettingell, W.H., Yu, C.E., Jondro, P.D., Schmidt, S.D., Wang, K., et al. (1995). Candidate gene for the chromosome 1 familial Alzheimer's disease locus. Science *269*, 973–977.

Li, T., Wernersson, R., Hansen, R.B., Horn, H., Mercer, J., Slodkowicz, G., Workman, C.T., Rigina, O., Rapacki, K., Stærfeldt, H.H., et al. (2017). A scored human protein-protein interaction network to catalyze genomic interpretation. Nat. Methods *14*, 61–64.

Lifton, R.P., Gharavi, A.G., and Geller, D.S. (2001). Molecular mechanisms of human hypertension. Cell *104*, 545–556.

Liu, L., Lei, J., Sanders, S.J., Willsey, A.J., Kou, Y., Cicek, A.E., Klei, L., Lu, C., He, X., Li, M., et al. (2014). DAWN: a framework to identify autism genes and subnetworks using gene expression and genetics. Mol. Autism 5, 22.

Luo, Y., Shan, G., Guo, W., Smrt, R.D., Johnson, E.B., Li, X., Pfeiffer, R.L., Szulwach, K.E., Duan, R., Barkho, B.Z., et al. (2010). Fragile x mental retardation protein regulates proliferation and differentiation of adult neural stem/progenitor cells. PLoS Genet. 6, e1000898.

Luongo, F.J., Horn, M.E., and Sohal, V.S. (2016). Putative microcircuit-level substrates for attention are disrupted in mouse models of autism. Biol. Psychiatry 79, 667–675.

Marsh-Armstrong, N., Huang, H., Berry, D.L., and Brown, D.D. (1999). Germline transmission of transgenes in Xenopus laevis. Proc. Natl. Acad. Sci. USA 96, 14389–14393.

Mirrashidi, K.M., Elwell, C.A., Verschueren, E., Johnson, J.R., Frando, A., Von Dollen, J., Rosenberg, O., Gulbahce, N., Jang, G., Johnson, T., et al. (2015). Global mapping of the lnc-human interactome reveals that retromer restricts chlamydia infection. Cell Host Microbe 18, 109–121.

Murray, C.J., Atkinson, C., Bhalla, K., Birbeck, G., Burstein, R., Chou, D., Dellavalle, R., Danaei, G., Ezzati, M., Fahimi, A., et al.; U.S. Burden of Disease Collaborators (2013). The state of US health, 1990-2010: burden of diseases, injuries, and risk factors. JAMA 310, 591–608.

Nelson, S.B., and Valakh, V. (2015). Excitatory/inhibitory balance and circuit homeostasis in autism spectrum disorders. Neuron 87, 684–698.

Noble, W., and Burns, M.P. (2010). Challenges in neurodegeneration research. Front. Psychiatry 1, 7.

Parker, L., Howlett, I.C., Rusan, Z.M., and Tanouye, M.A. (2011). Seizure and epilepsy: studies of seizure disorders in Drosophila. Int. Rev. Neurobiol. 99, 1–21.

Power, R.A., Tansey, K.E., Buttenschøn, H.N., Cohen-Woods, S., Bigdeli, T., Hall, L.S., Kutalik, Z., Lee, S.H., Ripke, S., Steinberg, S., et al.; CONVERGE Consortium, CARDIoGRAM Consortium, GERAD1 Consortium (2017). Genome-wide association for major depression through age at onset stratification: Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. Biol. Psychiatry 81, 325–335.

Prinz, A.A., Bucher, D., and Marder, E. (2004). Similar network activity from disparate circuit parameters. Nat. Neurosci. 7, 1345–1352.

Ramoni, R.B., Mulvihill, J.J., Adams, D.R., Allard, P., Ashley, E.A., Bernstein, J.A., Gahl, W.A., Hamid, R., Loscalzo, J., McCray, A.T., et al.; Undiagnosed Diseases Network (2017). The Undiagnosed Diseases Network: accelerating discovery about health and disease. Am. J. Hum. Genet. 100, 185–192.

Rauch, A., Wieczorek, D., Graf, E., Wieland, T., Endele, S., Schwarzmayr, T., Albrecht, B., Bartholdi, D., Beygo, J., Di Donato, N., et al. (2012). Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet 380, 1674–1682.

Razafsha, M., Behforuzi, H., Harati, H., Wafai, R.A., Khaku, A., Mondello, S., Gold, M.S., and Kobeissy, F.H. (2013). An updated overview of animal models in neuropsychiatry. Neuroscience 240, 204–218.

Roguev, A., Talbot, D., Negri, G.L., Shales, M., Cagney, G., Bandyopadhyay, S., Panning, B., and Krogan, N.J. (2013). Quantitative genetic-interaction mapping in mammalian cells. Nat. Methods 10, 432–437.

Rothman, G.R., Blackiston, D.J., and Levin, M. (2016). Color and intensity discrimination in Xenopus laevis tadpoles. Anim. Cogn. 19, 911–919.

Sanders, S.J., He, X., Willsey, A.J., Ercan-Sencicek, A.G., Samocha, K.E., Cicek, A.E., Murtha, M.T., Bal, V.H., Bishop, S.L., Dong, S., et al.; Autism Sequencing Consortium (2015). Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. Neuron 87, 1215–1233.

Satterstrom, F.K., Walters, R.K., Singh, T., Wigdor, E.M., Lescai, F., Demontis, D., Kosmicki, J.A., Grove, J., Stevens, C., Bybjerg-Grauholm, J., et al. (2018). ASD and ADHD have a similar burden of rare protein-truncating variants. bioRxiv. https://doi.org/10.1101/277707.

Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature 511, 421–427.

Scola, G., and Duong, A. (2017). Prenatal maternal immune activation and brain development with relevance to psychiatric disorders. Neuroscience 346, 403–408.

Sehgal, A. (2017). Physiology flies with time. Cell 171, 1232–1235.

Senkowski, D., and Gallinat, J. (2015). Dysfunctional prefrontal gamma-band oscillations reflect working memory and other cognitive deficits in schizophrenia. Biol. Psychiatry 77, 1010–1019.

Şentürk, M., and Bellen, H.J. (2018). Genetic strategies to tackle neurological diseases in fruit flies. Curr. Opin. Neurobiol. 50, 24–32.

Shah, P.S., Wojcechowskyj, J.A., Eckhardt, M., and Krogan, N.J. (2015). Comparative mapping of host-pathogen protein-protein interactions. Curr. Opin. Microbiol. 27, 62–68.

Shen, J.P., Zhao, D., Sasik, R., Luebeck, J., Birmingham, A., Bojorquez-Gomez, A., Licon, K., Klepper, K., Pekin, D., Beckett, A.N., et al. (2017). Combinatorial CRISPR-Cas9 screens for de novo mapping of genetic interactions. Nat. Methods 14, 573–576.

Sherrington, R., Rogaev, E.I., Liang, Y., Rogaeva, E.A., Levesque, G., Ikeda, M., Chi, H., Lin, C., Li, G., Holman, K., et al. (1995). Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. Nature 375, 754–760.

Sohal, V.S. (2012). Insights into cortical oscillations arising from optogenetic studies. Biol. Psychiatry 71, 1039–1045.

State, M.W., and Šestan, N. (2012). Neuroscience. The emerging biology of autism spectrum disorders. Science 337, 1301–1303.

Stewart, A.M., Nguyen, M., Wong, K., Poudel, M.K., and Kalueff, A.V. (2014). Developing zebrafish models of autism spectrum disorder (ASD). Prog. Neuropsychopharmacol. Biol. Psychiatry 50, 27–36.

Strittmatter, W.J., Saunders, A.M., Schmechel, D., Pericak-Vance, M., Enghild, J., Salvesen, G.S., and Roses, A.D. (1993). Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. Proc. Natl. Acad. Sci. USA 90, 1977–1981.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 43, D447–D452.

Tebbenkamp, A.T.N., Willsey, A.J., State, M.W., and Sestan, N. (2014). The developmental transcriptome of the human brain: implications for neurodevelopmental disorders. Curr. Opin. Neurol. 27, 149–156.

The Huntington's Disease Collaborative Research Group (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72, 971–983.

Thomsen, G., Woolf, T., Whitman, M., Sokol, S., Vaughan, J., Vale, W., and Melton, D.A. (1990). Activins are expressed early in Xenopus embryogenesis and can induce axial mesoderm and anterior structures. Cell 63, 485–493.

Thul, P.J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Björk, L., Breckels, L.M., et al. (2017). A subcellular map of the human proteome. Science 356, eaal3321.

Vandin, F., Clay, P., Upfal, E., and Raphael, B.J. (2012). Discovery of mutated subnetworks associated with clinical data in cancer. Pac. Symp. Biocomput. 2012, 55–66.

Wang, J., Al-Ouran, R., Hu, Y., Kim, S.-Y., Wan, Y.-W., Wangler, M.F., Yamamoto, S., Chao, H.-T., Comjean, A., Mohr, S.E., et al.; UDN (2017). MARRVEL: integration of human and model organism genetic resources to facilitate functional annotation of the human genome. Am. J. Hum. Genet. 100, 843–853.

Wang, C., Telpoukhovskaia, M.A., Bahr, B.A., Chen, X., and Gan, L. (2018). Endo-lysosomal dysfunction: a converging mechanism in neurodegenerative diseases. Curr. Opin. Neurobiol. 48, 52–58.

Wangler, M.F., Yamamoto, S., Chao, H.-T., Posey, J.E., Westerfield, M., Postlethwait, J., Hieter, P., Boycott, K.M., Campeau, P.M., and Bellen, H.J.; Members of the Undiagnosed Diseases Network (UDN) (2017). Model organisms facilitate rare disease diagnosis and therapeutic research. Genetics 207, 9–27.

Wheeler, G.N., and Brändli, A.W. (2009). Simple vertebrate models for chemical genetics and drug discovery screens: lessons from zebrafish and Xenopus. Dev. Dyn. *238*, 1287–1308.

Whiteford, H.A., Degenhardt, L., Rehm, J., Baxter, A.J., Ferrari, A.J., Erskine, H.E., Charlson, F.J., Norman, R.E., Flaxman, A.D., Johns, N., et al. (2013). Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. Lancet *382*, 1575–1586.

Willsey, A.J., and State, M.W. (2015). Autism spectrum disorders: from genes to neurobiology. Curr. Opin. Neurobiol. *30*, 92–99.

Willsey, A.J., Fernandez, T.V., Yu, D., King, R.A., Dietrich, A., Xing, J., Sanders, S.J., Mandell, J.D., Huang, A.Y., Richer, P., et al. (2017). De novo coding variants are strongly associated with Tourette disorder. Neuron *94*, 486–499.

Wray, N.R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E.M., Abdellaoui, A., Adams, M.J., Agerbo, E., Air, T.M., Andlauer, T.M.F., et al. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat. Genet. *50*, 668–681.

Yamamoto, S., Jaiswal, M., Charng, W.-L., Gambin, T., Karaca, E., Mirzaa, G., Wiszniewski, W., Sandoval, H., Haelterman, N.A., Xiong, B., et al. (2014). A

Drosophila genetic resource of mutants to study mechanisms underlying human genetic diseases. Cell *159*, 200–214.

Yu, M.K., Kramer, M., Dutkowski, J., Srivas, R., Licon, K., Kreisberg, J., Ng, C.T., Krogan, N., Sharan, R., and Ideker, T. (2016). Translation of genotype to phenotype by a hierarchy of cell subsystems. Cell Syst. *2*, 77–88.

Zaidi, S., Choi, M., Wakimoto, H., Ma, L., Jiang, J., Overton, J.D., Romano-Adesman, A., Bjornson, R.D., Breitbart, R.E., Brown, K.K., et al. (2013). De novo mutations in histone-modifying genes in congenital heart disease. Nature *498*, 220–223.

Zerbo, O., Qian, Y., Yoshida, C., Fireman, B.H., Klein, N.P., and Croen, L.A. (2017). Association between influenza infection and vaccination during pregnancy and risk of autism spectrum disorder. JAMA Pediatr. *171*, e163609.

Zhao, X.-M., Iskar, M., Zeller, G., Kuhn, M., van Noort, V., and Bork, P. (2011). Prediction of drug combinations by integrating molecular and pharmacological data. PLoS Comput. Biol. *7*, e1002323.

Zimmerman, L.B., De Jesús-Escobar, J.M., and Harland, R.M. (1996). The Spemann organizer signal noggin binds and inactivates bone morphogenetic protein 4. Cell *86*, 599–606.