



Published in final edited form as:

Cell. 2018 July 26; 174(3): 505–520. doi:10.1016/j.cell.2018.06.016.

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. Morris², Sheng Wang², Helen R. Willsey¹, Nawei Sun², Nia Teerikorpi², Tierney B. Baum², Gerard Cagney⁴, Kevin J. Bender⁵, Tejal A. Desai⁶, Deepak Srivastava^{7,8,9}, Graeme W. Davis^{9,10}, Jennifer Doudna^{11,12,13,14,15}, Edward Chang¹⁶, Vikaas Sohal^{1,10}, Daniel H. Lowenstein⁵, Hao Li^{3,9}, David Agard^{3,9}, Michael J. Keiser^{2,3,6,17}, Brian Shoichet^{3,17}, Mark von Zastrow^{1,3,18}, Lennart Mucke^{5,7}, Steven Finkbeiner^{5,7,19}, Li Gan^{5,7}, Nenad Sestan²⁰, Michael E. Ward²¹, Ruth Huttenhain^{3,7,18}, Tomasz Nowakowski^{1,22,23}, Hugo J. Bellen^{13,24}, Loren M. Frank^{10,13,19}, Mustafa K. Khokha²⁵, Richard P. Lifton²⁶, Martin Kampmann^{2,3,9,27}, Trey Ideker²⁸, Matthew W. State¹, and Nevan J. Krogan^{3,7,18,29,*}

¹Department of Psychiatry, UCSF Weill Institute for Neurosciences, University of California San Francisco, San Francisco, CA 94143, USA. ²Institute for Neurodegenerative Diseases, UCSF Weill Institute for Neurosciences, University of California San Francisco, San Francisco, CA, 94143, USA. ³Quantitative Biosciences Institute (QBI), University of California San Francisco, San Francisco, CA 94143, USA. ⁴School of Biomolecular and Biomedical Science, Conway Institute, University College Dublin, Dublin 4, Ireland. ⁵Department of Neurology, UCSF Weill Institute for Neurosciences, University of California San Francisco, San Francisco, CA, 94158, USA. ⁶Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA 94143, USA. ⁷Gladstone Institutes, San Francisco, CA 94158, USA. ⁸Department of Pediatrics, University of California San Francisco, San Francisco, CA, 94143, USA. ⁹Department of Biochemistry and Biophysics, University of California San Francisco, San Francisco, CA 94143, USA. ¹⁰Kavli Institute for Fundamental Neuroscience, University of California San Francisco, San Francisco, CA, 94143, USA. ¹¹Department of Molecular and Cell Biology, University of California Berkeley, Berkeley, CA, 94720, USA. ¹²Department of Chemistry,

*Please address correspondence to: jeremy.willsey@ucsf.edu (lead contact; A. J. W.), nevan.krogan@ucsf.edu (N. J. K.).

Author Contributions

Conceptualization, AJW, MTM, NS, MK, TI, MWS, and NJK; Methodology, AJW, MTM, SW, HRW, NS, NT, TBB, JD, DA, MJK, BS, MvZ, SF, LG, NS, MEW, RH, TN, MK, TI, MWS, and NJK; Formal Analysis, AJW, MTM, and SW; Investigation, AJW, MTM, and SW; Resources, AJW, LG, NS, MEW, MKK, MK, TI, MWS, and NJK; Data Curation, MTM and SW; Writing (Original Draft), AJW, MTM, HRW, EC, VS, LM, SF, LG, TN, HJB, LMF, MKK, RPL, MK, TI, MWS, and NJK; Writing (Review & Editing), AJW, MTM, SW, HRW, NS, NT, TBB, KJB, TAD, DS, GWD, JD, EC, VS, DHL, HL, DA, MJK, BS, MvZ, LM, SF, LG, NS, MEW, RH, TN, HJB, LMF, MKK, RPL, MK, TI, MWS, and NJK; Visualization, AJW, MTM, and NJK; Supervision, AJW, MWS, and NJK; Project Administration, AJW, TI, MWS, and NJK; Funding Acquisition, AJW, JD, DA, MJK, BS, MvZ, SF, NS, RH, TN, MK, TI, MWS, and NJK.

Declaration of Interests

The authors declare no competing interests.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

University of California, Berkeley, CA, 94720, USA. ¹³Howard Hughes Medical Institute. ¹⁴Innovative Genomics Institute, University of California, Berkeley, CA, 94720, USA. ¹⁵MBIB Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA. ¹⁶Department of Neurological Surgery, UCSF Weill Institute for Neurosciences, University of California San Francisco, San Francisco, CA, 94143, USA. ¹⁷Department of Pharmaceutical Chemistry, University of California, San Francisco San Francisco, CA 94143, USA. ¹⁸Department of Cellular and Molecular Pharmacology, University of California San Francisco, San Francisco, CA 94143, USA. ¹⁹Department of Physiology, University of California San Francisco, San Francisco, CA 94143, USA. ²⁰Department of Neuroscience and Kavli Institute for Neuroscience, Yale School of Medicine, New Haven, CT, 066510, USA. ²¹National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, 20892, USA. ²²Department of Anatomy, University of California San Francisco, San Francisco, CA, 94143, USA. ²³The Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California San Francisco, San Francisco, CA, 94143, USA. ²⁴Departments of Molecular and Human Genetics and Neuroscience, Neurological Research Institute at TCH, Baylor College of Medicine, Houston, TX 77030, USA. ²⁵Pediatric Genomics Discovery Program, Departments of Pediatrics and Genetics, Yale University School of Medicine, New Haven, CT, 06510, USA. ²⁶Laboratory of Human Genetics and Genomics, The Rockefeller University, New York, New York, 10065, USA. ²⁷Chan Zuckerberg Biohub, San Francisco, CA, 94158, USA. ²⁸Department of Medicine, University of California San Diego, La Jolla, CA 92093, USA. ²⁹Helen Diller Comprehensive Cancer Center, University of California San Francisco, San Francisco, CA 94143, USA.

Summary

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.

Keywords

Proteomics; mass spectrometry; molecular; pathway; systems biology; physical interactions; protein-protein interactions; protein-DNA interactions; genetics; genetic interactions; functional; interactome; convergence; network; co-expression; CRISPR; psychiatric cell map initiative; PCMI; gene discovery; ASD; autism spectrum disorders; intellectual disability; epilepsy; Tourette disorder; obsessive compulsive disorder; neurodevelopmental disorder; psychiatry; psychiatric disorder

Introduction

The global burden of mental illness is enormous, whether measured in health care expenditures, lost productivity, or personal suffering. Worldwide, neuropsychiatric disorders (NPDs) rank number five among the top 10 categories contributing to overall disease burden as measured by global disability-adjusted life years (DALYs) as well as the leading cause of non-fatal burden of disease (Whiteford et al., 2013). In the United States, these disorders 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 understand 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. This utility of this knowledge is exemplified in hypertension: large effect mutations converged on salt handling in the kidney, with vectorial 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 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. Since 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 (CCMI¹) (Krogan et al., 2015) and the Host Pathogen Mapping Initiative (HPMI²) (Shah et al., 2015). For example, our group has recently formed the Psychiatric Cell Map Initiative (PCMI³), which is focused on a number of NPDs. Each of these initiatives aim to (1)

¹<http://www.ccmi.org>

²<http://hpmi.ucsf.edu>

³<http://pcmi.ucsf.edu>

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 (Bilguvar et al., 2010; Cross-Disorder Group of the PGC, 2013; Deciphering Developmental Disorders Study, 2017; eQTLGen et al., 2018; EuroEPINOMICS-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 PGC, 2014; Willsey et al., 2017). Additional progress is imminent in characterizing the underlying genetics of OCD (Cappi et al., 2017; IOCDF-GC and 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 risk alleles 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 (eQTLGen 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 PGC, 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, Table S2 and Table 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 PGC, 2017; Cross-Disorder Group of the PGC, 2013; Cross-Disorder Group of the PGC et al., 2013). As additional genetic data accumulates 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 approximately 20-fold increases in risk, with LGD variants in the highest confidence genes carrying even greater risks. To date, there is very little evidence for a multi-hit hypothesis vis-à-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 genome-wide ($p = 7.8 \times 10^{-12}$, 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 100's to 1000's 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 confounds due to 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 neurons, particularly deep layer (layer (L) 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 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.⁴ 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 (Figure 1 and Figure 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 at once, 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,

⁴See NIH FOA at <https://grants.nih.gov/grants/guide/pa-files/PAR-17-176.html> for a detailed explanation.

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), 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; Jager 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). 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—for example, regulatory complexes may control the expression of genes encoding interacting synaptic proteins.

Mapping Genetic & 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’ which 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 which 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 (The Gene Ontology Consortium 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 (Chang, 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 (Chang, 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 (entürk and Bellen, 2017). 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 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 (entürk and Bellen, 2017; Wangler et al., 2017).

Zebrafish is an alternative *in vivo* model organism which 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 percent (NCBI Homologene). In comparison, mice have 64 of 65 with 93 percent identity and zebrafish have 62 of 65 with 71 percent 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 × 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 Resource⁵), progenitor biology (e.g. Pax6-GFP (Hirsch et al., 2002)), and synapse formation (e.g. PSD95-GFP⁶) *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

⁵<http://www.mbl.edu/xenopus/tropicalis-stocks>

⁶Nicolas Marsh-Armstrong, unpublished data

genes, $p = 4.01 \times 10^{-13}$; Figure 3B, Table S2 and Table 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), though 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 ($p = 1.16 \times 10^{-15}$, $n = 43/273$ 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) 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 ($p = 1.67 \times 10^{-9}$, $n = 14/273$ NPD genes, 92 total CHD genes; Table S2 and Table 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 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.

Methods

Selection of de novo variants and Identification of risk genes for NPDs and for CHD

We manually curated de novo variants from several large scale studies, representative of the current state of the field of whole-exome sequencing in NPDs (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; Willsey et al., 2017) as well as in CHD (Jin et al., 2017) (Table S1). We omitted targeted sequencing studies in order to generate a fair “exome-wide” comparison across disorders. We focused on de novo coding variants only as genes affected by these variants are good initial targets for bottom-up approaches. As has been done previously, we focused on de novo likely gene disrupting variants (LGD; insertion of premature stop codon, disruption of canonical splice site, or frameshift insertion-deletion variant) or probably damaging missense (Mis3) variants (Polyphen 2 score > 0.909 (Adzhubei et al., 2013)). We then performed a basic TADA analysis (He et al., 2013) for each of the respective disorders (ASD, EE, ID, TD, OCD, SCZ, and CHD), with 1000 permutations per disorder, utilizing established false discovery rate (FDR) thresholds (De Rubeis et al., 2014; He et al., 2013) to identify “high confidence” genes (FDR < 0.1 ; Table S1). For each disorder, we estimated the effect sizes for LGD and Mis3 variants as in He et al. (2013), and set the number of risk genes at 500, which fits well with the current literature (Cappi et al., 2017; Deciphering Developmental Disorders Study, 2017; EuroEPINOMICS-RES Consortium et al., 2014; Fromer et al., 2014; He et al., 2013; Jin et al., 2017; Sanders et al., 2015; Willsey et al., 2017). We also identified risk genes based on the 2017 Deciphering Developmental Disorders (DDD) study (Deciphering Developmental Disorders Study, 2017), which examined mutations in developmental disorders as a group. Therefore, for the purposes of generating a list of all NPD risk genes (Figure 3), we used the union of the DDD genes with the risk genes from the other NPDs (Table S1).

Identification of Cancer and Host-Pathogen genes

We utilized the Tier 1 genes in the Cancer Gene Census (Futreal et al., 2004) as our list of cancer genes. This Tier is reserved for genes known to play a causative role in cancer. This results in a list of 556 cancer genes after trimming to genes with mutation rates estimated for TADA (Table S1). We prioritized host-pathogen interacting (HPI) genes based on the database HPIDB 2.0 (Ammari et al., 2016). This database summarizes all known (experimentally validated) interactions between proteins from host species (animal, plant) and proteins from pathogenic species. Each interaction is assigned a confidence score. To prioritize critical HPI genes, therefore, for each of the host proteins, we generated a ‘weighted’ connectivity metric (“gene degree”), by summing the confidence scores for each unique interaction that protein has. In the case of multiple reported interactions with the same pathogen protein, we used the maximum confidence score reported. We prioritized HPI proteins based on their weighted connectivity, keeping only those in the upper quartile for all reported host proteins. This results in a list of 601 proteins after trimming to genes covered by TADA (Table S1).

Overlap analysis

We estimated the significance of gene overlap across individual NPDs in R with the hypergeometric function *phyper*. More specifically, *phyper(q, m, n, k, lower.tail = FALSE)* where q = number of overlapping genes - 1, m = gene count for set testing overlap with, n = number of genes in corrected TADA background - m , k = number of genes in primary gene set. For example, for the overlap of ASD with ID, the function used was *phyper(4-1, 28, 18572*0.9-28, 62, lower.tail = F)*, which is $p = 3.29 \times 10^{-6}$. We also assessed overlap between NPD genes as a group and CHD, Cancer, or HPI genes (as well as each of their overlaps) with this function (Table S2 & Table S3). We used the list of genes from TADA (He et al., 2013) as the background for each test, and excluded all genes/proteins not present in this population. Moreover, to account for differences in sequencing coverage across the various disorders (we do not expect all genes to have sufficient sequencing coverage), the total number of background genes used was reduced by 10%, resulting in a slightly more stringent test.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported directly or indirectly by the NIH/NIMH (grants U01 MH105575 (MWS), R01 MH110928 (AJW, MWS), and R01 MH109901 (AJW, MWS)); the NIH/NIGMS (New Innovator Award to MK); the NIH/NICHD/NHLBI (R01 to MKK); Edward Mallinckrodt, Jr. Foundation (MKK is a Mallinckrodt Scholar); the QBI at UCSF (Bold & Basic Grant to AJW, BS, MvZ); the UCSF Weill Institute for Neurosciences (Trailblazer 2017 Award to AJW); the Taube/Koret Center (SF); the Paul G. Allen Family Foundation (Distinguished Investigator Award to MJK); the Human Frontiers in Science Program (postdoctoral fellowship to RH); the Intramural Research Program of the NIH, NINDS (MEW); the Overlook International Foundation (MWS and AJW); the Department of Psychiatry at UCSF and the UCSF Chancellor's Fund for Precision Medicine (seed funding for the PCMI to AJW, MWS, and NJK). We thank all of the PCMI investigators as well as the members of the Willsey, Ideker, State, and Krogan labs for their detailed input and support. Sarah Pyle for graphic design. Finally, we thank all of the patients and their family members who have made this research possible by donating their time and biomaterials.

References

- Adzhubei I, Jordan DM, and Sunyaev SR (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 MB, Severi KE, Macklin JJ, et al. (2013). Genetically encoded calcium indicators for multi-color neural activity imaging and combination with optogenetics. *Front. Mol. Neurosci* 6, 2. [PubMed: 23459413]
- Ammari MG, Gresham CR, McCarthy FM, and Nanduri B (2016). HPIDB 2.0: a curated database for host-pathogen interactions. *Database* 2016.
- Autism Spectrum Disorders Working Group of The PGC (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. [PubMed: 28540026]
- Baraban SC, Dinday MT, and Hortopan GA (2013). Drug screening in *Scn1a* zebrafish mutant identifies clemizole as a potential Dravet syndrome treatment. *Nat. Commun* 4, 2410. [PubMed: 24002024]
- Bassik MC, Kampmann M, Lebbink RJ, Wang S, Hein MY, Poser I, Weibezahn J, Horlbeck MA, Chen S, Mann M, et al. (2013). A systematic mammalian genetic interaction map reveals pathways underlying ricin susceptibility. *Cell* 152, 909–922. [PubMed: 23394947]

- Beltrao P, Cagney G, and Krogan NJ (2010). Quantitative genetic interactions reveal biological modularity. *Cell* 141, 739–745. [PubMed: 20510918]
- Ben-Shalom R, Keeshen CM, Berrios KN, An JY, Sanders SJ, and Bender KJ (2017). Opposing Effects on NaV1.2 Function Underlie Differences Between SCN2A Variants Observed in Individuals With Autism Spectrum Disorder or Infantile Seizures. *Biol. Psychiatry* 82, 224–232. [PubMed: 28256214]
- Bilguvar K, Ozturk AK, Louvi A, Kwan KY, Choi M, Tatli B, Yalnizoglu D, Tuysuz B, Caglayan AO, Gokben S, et al. (2010). Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. *Nature* 467, 207–210. [PubMed: 20729831]
- Blackiston DJ, and Levin M (2012). Aversive training methods in *Xenopus laevis*: general principles. *Cold Spring Harb. Protoc* 2012.
- Blomström Å, Karlsson H, Gardner R, Jørgensen 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. [PubMed: 26303935]
- Boettcher M, Tian R, Blau JA, Markegard E, Wagner RT, 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. [PubMed: 29334369]
- Brohée S, and van Helden J (2006). Evaluation of clustering algorithms for protein-protein interaction networks. *BMC Bioinformatics* 7, 488. [PubMed: 17087821]
- Cappi C, Oliphant ME, Peter Z, Zai G, Sullivan CAW, Gupta AR, Hoffman EJ, Virdee M, Jeremy Willsey A, Shavitt RG, et al. (2017). De novo damaging coding mutations are strongly associated with obsessive-compulsive disorder and overlap with autism.
- Cassa CA, Weghorn D, Balick DJ, Jordan DM, Nusinow D, Samocha KE, O'Donnell-Luria A, MacArthur DG, Daly MJ, Beier DR, et al. (2017). Estimating the selective effects of heterozygous protein-truncating variants from human exome data. *Nat. Genet* 49, 806–810. [PubMed: 28369035]
- Chang EF (2015). Towards large-scale, human-based, mesoscopic neurotechnologies. *Neuron* 86, 68–78. [PubMed: 25856487]
- Cho KKA, Hoch R, Lee AT, Patel T, Rubenstein JLR, and Sohal VS (2015). Gamma rhythms link prefrontal interneuron dysfunction with cognitive inflexibility in *Dlx5/6(+/-)* mice. *Neuron* 85, 1332–1343. [PubMed: 25754826]
- Chong JX, Buckingham KJ, Jhangiani SN, Boehm C, Sobreira N, Smith JD, Harrell TM, McMillin MJ, Wiszniewski W, Gambin T, et al. (2015). The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. *Am. J. Hum. Genet* 97, 199–215. [PubMed: 26166479]
- Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, and Pericak-Vance MA (1993). Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 261, 921–923. [PubMed: 8346443]
- Cotney J, Muhle RA, Sanders SJ, Liu L, Willsey AJ, 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. [PubMed: 25752243]
- Cross-Disorder Group of the PGC (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* 381, 1371–1379. [PubMed: 23453885]
- Cross-Disorder Group of the PGC, Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, Perlis RH, Mowry BJ, Thapar A, Goddard ME, et al. (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet* 45, 984–994. [PubMed: 23933821]
- Davis GW (2006). Homeostatic control of neural activity: from phenomenology to molecular design. *Annu. Rev. Neurosci* 29, 307–323. [PubMed: 16776588]
- Davis ZH, Verschuere E, Jang GM, Kleffman K, Johnson JR, Park J, Von Dollen J, Maher MC, 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. [PubMed: 25544563]
- Deciphering Developmental Disorders Study (2017). Prevalence and architecture of de novo mutations in developmental disorders. *Nature* 542, 433–438. [PubMed: 28135719]

- De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, Kou Y, Liu L, Fromer M, Walker S, et al. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 515, 209–215. [PubMed: 25363760]
- Du D, Roguev A, Gordon DE, Chen M, Chen S-H, Shales M, Shen JP, Ideker T, Mali P, Qi LS, et al. (2017). Genetic interaction mapping in mammalian cells using CRISPR interference. *Nat. Methods* 14, 577–580. [PubMed: 28481362]
- Duncan AR, and Khokha MK (2016). *Xenopus* as a model organism for birth defects-Congenital heart disease and heterotaxy. *Semin. Cell Dev. Biol* 51, 73–79. [PubMed: 26910255]
- Dutkowski J, Kramer M, Surma MA, Balakrishnan R, Cherry JM, Krogan NJ, and Ideker T (2013). A gene ontology inferred from molecular networks. *Nat. Biotechnol* 31, 38–45. [PubMed: 23242164]
- 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. [PubMed: 21475194]
- eQTLGen, 23andMe, the Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui, A, Adams MJ, et al. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet* 34, 119.
- EuroEPINOMICS-RES Consortium, Epilepsy Phenome/Genome Project, and Epi4K Consortium (2014). De novo mutations in synaptic transmission genes including DNM1 cause epileptic encephalopathies. *Am. J. Hum. Genet* 95, 360–370. [PubMed: 25262651]
- Faulkner RL, Wishard TJ, Thompson CK, Liu H-H, and Cline HT (2015). FMRP regulates neurogenesis in vivo in *Xenopus laevis* tadpoles. *eNeuro* 2, e0055. [PubMed: 25844398]
- Finkbeiner S, Frumkin M, and Kassner PD (2015). Cell-based screening: extracting meaning from complex data. *Neuron* 86, 160–174. [PubMed: 25856492]
- Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P, Georgieva L, Rees E, Palta P, Ruderfer DM, et al. (2014). De novo mutations in schizophrenia implicate synaptic networks. *Nature* 506, 179–184. [PubMed: 24463507]
- Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, and Stratton MR (2004). A census of human cancer genes. *Nat. Rev. Cancer* 4, 177–183. [PubMed: 14993899]
- García-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. [PubMed: 28710364]
- Geschwind DH, and State MW (2015). Gene hunting in autism spectrum disorder: on the path to precision medicine. *Lancet Neurol.* 14, 1109–1120. [PubMed: 25891009]
- Goate A, Chartier-Harlin MC, Mullan M, Brown J, Crawford F, Fidani L, Giuffra L, Haynes A, Irving N, and James L (1991). Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature* 349, 704–706. [PubMed: 1671712]
- Guo W, Allan AM, Zong R, Zhang L, Johnson EB, Schaller EG, Murthy AC, Goggin SL, Eisch AJ, Oostra BA, et al. (2011). Ablation of *Fmrp* in adult neural stem cells disrupts hippocampus-dependent learning. *Nat. Med* 17, 559–565. [PubMed: 21516088]
- Hamdan FF, 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. [PubMed: 25356899]
- He X, Sanders SJ, Liu L, De Rubeis S, Lim ET, Sutcliffe JS, Schellenberg GD, Gibbs RA, Daly MJ, Buxbaum JD, et al. (2013). Integrated model of de novo and inherited genetic variants yields greater power to identify risk genes. *PLoS Genet.* 9, e1003671. [PubMed: 23966865]
- Hellsten U, Harland RM, Gilchrist MJ, Hendrix D, Jurka J, Kapitonov V, Ovcharenko I, Putnam NH, Shu S, Taher L, et al. (2010). The genome of the Western clawed frog *Xenopus tropicalis*. *Science* 328, 633–636. [PubMed: 20431018]
- Hirsch N, Zimmerman LB, Gray J, Chae J, Curran KL, Fisher M, Ogino H, and Grainger RM (2002). *Xenopus tropicalis* transgenic lines and their use in the study of embryonic induction. *Dev. Dyn* 225, 522–535. [PubMed: 12454928]

- Hoffman EJ, Turner KJ, Fernandez JM, Cifuentes D, Ghosh M, Ijaz S, Jain RA, Kubo F, Bill BR, Baier H, et al. (2016). Estrogens Suppress a Behavioral Phenotype in Zebrafish Mutants of the Autism Risk Gene, CNTNAP2. *Neuron* 89, 725–733. [PubMed: 26833134]
- Homsy J, Zaidi S, Shen Y, Ware JS, Samocha KE, Karczewski KJ, DePalma SR, 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. [PubMed: 26785492]
- Horwitz R, and Johnson GT (2017). Whole cell maps chart a course for 21st-century cell biology. *Science* 356, 806–807. [PubMed: 28546174]
- Hou L, Bergen SE, Akula N, Song J, Hultman CM, Landén M, Adli M, Alda M, Ardu 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. [PubMed: 27329760]
- Huang AY, Yu D, Davis LK, Sul JH, Tsetsos F, Ramensky V, Zelaya I, Ramos EM, Osiecki L, Chen JA, et al. (2017). Rare Copy Number Variants in NRXN1 and CNTN6 Increase Risk for Tourette Syndrome. *Neuron* 94, 1101–1111 e7. [PubMed: 28641109]
- Huttlin EL, Ting L, Bruckner RJ, Gebreab F, Gygi MP, 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. [PubMed: 26186194]
- Ideker T, Dutkowski J, and Hood L (2011). Boosting signal-to-noise in complex biology: prior knowledge is power. *Cell* 144, 860–863. [PubMed: 21414478]
- 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. [PubMed: 25087078]
- IOCDF-GC and OCGAS (2017). Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Mol. Psychiatry*
- Jager S, Cimermancic P, Gulbahce N, Johnson JR, McGovern KE, Clarke SC, Shales M, Mercenne G, Pache L, Li K, et al. (2011). Global landscape of HIV-human protein complexes. *Nature* 481, 365–370. [PubMed: 22190034]
- 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. [PubMed: 27287966]
- Jin SC, Homsy J, Zaidi S, Lu Q, Morton S, DePalma SR, Zeng X, Qi H, Chang W, Sierant MC, et al. (2017). Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. *Nat. Genet* 49, 1593–1601. [PubMed: 28991257]
- Jun JJ, Steinmetz NA, Siegle JH, Denman DJ, Bauza M, Barbarits B, Lee AK, Anastassiou CA, Andrei A, Aydın Ç, et al. (2017). Fully integrated silicon probes for high-density recording of neural activity. *Nature* 551, 232–236. [PubMed: 29120427]
- Kampmann M, Bassik MC, and Weissman JS (2013). Integrated platform for genome-wide screening and construction of high-density genetic interaction maps in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A* 110, E2317–E2326. [PubMed: 23739767]
- 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. [PubMed: 27217147]
- Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis AP, et al. (2006). Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* 440, 637–643. [PubMed: 16554755]
- Krogan NJ, Lippman S, Agard DA, Ashworth A, and Ideker T (2015). The cancer cell map initiative: defining the hallmark networks of cancer. *Mol. Cell* 58, 690–698. [PubMed: 26000852]
- Krystal JH, and State MW (2014). Psychiatric disorders: diagnosis to therapy. *Cell* 157, 201–214. [PubMed: 24679536]
- Lamb TM, Knecht AK, Smith WC, Stachel SE, Economides AN, Stahl N, Yancopoulos GD, and Harland RM (1993). Neural induction by the secreted polypeptide noggin. *Science* 262, 713–718. [PubMed: 8235591]

- Lee P-T, Zirin J, Kanca O, Lin W-W, Schulze KL, Li-Kroeger D, Tao R, Devereaux C, Hu Y, Chung V, et al. (2018). A gene-specific T2A-GAL4 library for *Drosophila*. *Elife* 7.
- Lehner T, Senthil G, and Addington AM (2015). Convergence of Advances in Genomics, Team Science, and Repositories as Drivers of Progress in Psychiatric Genomics. *Biol. Psychiatry* 77, 6–14. [PubMed: 24503471]
- Levy-Lahad E, Wasco W, Poorkaj P, Romano DM, Oshima J, Pettingell WH, Yu CE, Jondro PD, Schmidt SD, and Wang K (1995). Candidate gene for the chromosome 1 familial Alzheimer's disease locus. *Science* 269, 973–977. [PubMed: 7638622]
- Li T, Wernersson R, Hansen RB, Horn H, Mercer J, Slodkowitz G, Workman CT, Rigina O, Rapacki K, Starfeldt HH, et al. (2017). A scored human protein-protein interaction network to catalyze genomic interpretation. *Nat. Methods* 14, 61–64. [PubMed: 27892958]
- Lifton RP, Gharavi AG, and Geller DS (2001). Molecular mechanisms of human hypertension. *Cell* 104, 545–556. [PubMed: 11239411]
- de Ligt J, Willemsen MH, van Bon BW, Kleefstra T, Yntema HG, Kroes T, Vulto-van Silfhout AT, Koolen DA, de Vries P, Gilissen C, et al. (2012). Diagnostic exome sequencing in persons with severe intellectual disability. *N. Engl. J. Med* 367, 1921–1929. [PubMed: 23033978]
- Liu L, Lei J, Sanders SJ, Willsey AJ, Kou Y, Cicek AE, 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. [PubMed: 24602502]
- Luo Y, Shan G, Guo W, Smrt RD, Johnson EB, Li X, Pfeiffer RL, Szulwach KE, Duan R, Barkho BZ, et al. (2010). Fragile × mental retardation protein regulates proliferation and differentiation of adult neural stem/progenitor cells. *PLoS Genet.* 6, e1000898. [PubMed: 20386739]
- Luongo FJ, Horn ME, and Sohal VS (2016). Putative Microcircuit-Level Substrates for Attention Are Disrupted in Mouse Models of Autism. *Biol. Psychiatry* 79, 667–675. [PubMed: 26022075]
- Marsh-Armstrong N, Huang H, Berry DL, and Brown DD (1999). Germ-line transmission of transgenes in *Xenopus laevis*. *Proc. Natl. Acad. Sci. U. S. A* 96, 14389–14393. [PubMed: 10588715]
- Mirrashidi KM, Elwell CA, Verschueren E, Johnson JR, Frando A, Von Dollen J, Rosenberg O, Gulbahce N, Jang G, Johnson T, et al. (2015). Global Mapping of the Inc-Human Interactome Reveals that Retromer Restricts Chlamydia Infection. *Cell Host Microbe* 18, 109–121. [PubMed: 26118995]
- Murray CJ, Atkinson C, Bhalla K, Birbeck G, Burstein R, Chou D, Dellavalle R, Danaei G, Ezzati M, Fahimi A, et al. (2013). The state of US health, 1990–2010: burden of diseases, injuries, and risk factors. *JAMA* 310, 591–608. [PubMed: 23842577]
- Nelson SB, and Valakh V (2015). Excitatory/Inhibitory Balance and Circuit Homeostasis in Autism Spectrum Disorders. *Neuron* 87, 684–698. [PubMed: 26291155]
- Noble W, and Burns MP (2010). Challenges in neurodegeneration research. *Front. Psychiatry* 1, 7. [PubMed: 21423419]
- Parker L, Howlett IC, Rusan ZM, and Tanouye MA (2011). Seizure and epilepsy: studies of seizure disorders in *Drosophila*. *Int. Rev. Neurobiol* 99, 1–21. [PubMed: 21906534]
- Power RA, Tansey KE, Buttenschön HN, Cohen-Woods S, Bigdeli T, Hall LS, Kutalik Z, Lee SH, Ripke S, Steinberg S, et al. (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. [PubMed: 27519822]
- Prinz AA, Bucher D, and Marder E (2004). Similar network activity from disparate circuit parameters. *Nat. Neurosci* 7, 1345–1352. [PubMed: 15558066]
- Ramoni RB, Mulvihill JJ, Adams DR, Allard P, Ashley EA, Bernstein JA, Gahl WA, Hamid R, Loscalzo J, McCray AT, et al. (2017). The Undiagnosed Diseases Network: Accelerating Discovery about Health and Disease. *Am. J. Hum. Genet* 100, 185–192. [PubMed: 28157539]
- Rauch A, Wieczorek D, Graf E, Wieland T, Ende 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. [PubMed: 23020937]

- Razafsha M, Behforuzi H, Harati H, Wafai RA, Khaku A, Mondello S, Gold MS, and Kobeissy FH (2013). An updated overview of animal models in neuropsychiatry. *Neuroscience* 240, 204–218. [PubMed: 23473749]
- Roguev A, Talbot D, Negri GL, Shales M, Cagney G, Bandyopadhyay S, Panning B, and Krogan NJ (2013). Quantitative genetic-interaction mapping in mammalian cells. *Nat. Methods* 10, 432–437. [PubMed: 23407553]
- Rothman GR, Blackiston DJ, and Levin M (2016). Color and intensity discrimination in *Xenopus laevis* tadpoles. *Anim. Cogn* 19, 911–919. [PubMed: 27146661]
- Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, Murtha MT, Bal VH, Bishop SL, Dong S, et al. (2015). Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* 87, 1215–1233. [PubMed: 26402605]
- Satterstrom FK, Walters RK, Singh T, Wigdor EM, Lescai F, Demontis D, Kosmicki JA, Grove J, Stevens C, Bybjerg-Grauholm J, et al. (2018). ASD and ADHD have a similar burden of rare protein-truncating variants.
- Schizophrenia Working Group of the PGC (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511, 421–427. [PubMed: 25056061]
- Scola G, and Duong A (2017). Prenatal maternal immune activation and brain development with relevance to psychiatric disorders. *Neuroscience* 346, 403–408. [PubMed: 28153689]
- Sehgal A (2017). Physiology Flies with Time. *Cell* 171, 1232–1235. [PubMed: 29195066]
- 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. [PubMed: 25847179]
- entürk M, and Bellen HJ (2017). Genetic strategies to tackle neurological diseases in fruit flies. *Curr. Opin. Neurobiol* 50, 24–32. [PubMed: 29128849]
- Shah PS, Wojcechowskyj JA, Eckhardt M, and Krogan NJ (2015). Comparative mapping of host-pathogen protein-protein interactions. *Curr. Opin. Microbiol* 27, 62–68. [PubMed: 26275922]
- Shen JP, Zhao D, Sasik R, Luebeck J, Birmingham A, Bojorquez-Gomez A, Licon K, Klepper K, Pekin D, Beckett AN, et al. (2017). Combinatorial CRISPR-Cas9 screens for de novo mapping of genetic interactions. *Nat. Methods* 14, 573–576. [PubMed: 28319113]
- Sherrington R, Rogaev EI, Liang Y, Rogaeva EA, 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. [PubMed: 7596406]
- Sohal VS (2012). Insights into cortical oscillations arising from optogenetic studies. *Biol. Psychiatry* 71, 1039–1045. [PubMed: 22381731]
- State MW, and Šestan N (2012). The emerging biology of autism spectrum disorders. *Science* 337, 1301–1303. [PubMed: 22984058]
- Stewart AM, Nguyen M, Wong K, Poudel MK, and Kalueff AV (2014). Developing zebrafish models of autism spectrum disorder (ASD). *Prog. Neuropsychopharmacol. Biol. Psychiatry* 50, 27–36. [PubMed: 24315837]
- Strittmatter WJ, Saunders AM, Schmechel D, Pericak-Vance M, Enghild J, Salvesen GS, and Roses AD (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. U. S. A* 90, 1977–1981. [PubMed: 8446617]
- Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43, D447–D452. [PubMed: 25352553]
- Tebbenkamp ATN, Willsey AJ, State MW, and Šestan N (2014). The developmental transcriptome of the human brain: implications for neurodevelopmental disorders. *Curr. Opin. Neurol* 27, 149–156. [PubMed: 24565942]
- The Gene Ontology Consortium, Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, et al. (2000). Gene ontology: tool for the unification of biology. *Nat. Genet* 25, 25–29. [PubMed: 10802651]

- 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. [PubMed: 8458085]
- Thomsen G, Woolf T, Whitman M, Sokol S, Vaughan J, Vale W, and Melton DA (1990). Activins are expressed early in *Xenopus* embryogenesis and can induce axial mesoderm and anterior structures. *Cell* 63, 485–493. [PubMed: 2225062]
- Thul PJ, Åkesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, Alm T, Asplund A, Björk L, Breckels LM, et al. (2017). A subcellular map of the human proteome. *Science* 356.
- Vandin F, Clay P, Upfal E, and Raphael BJ (2012). Discovery of mutated subnetworks associated with clinical data in cancer. *Pac. Symp. Biocomput* 55–66. [PubMed: 22174262]
- Wang C, Telpoukhovskaia MA, Bahr BA, Chen X, and Gan L (2018). Endo-lysosomal dysfunction: a converging mechanism in neurodegenerative diseases. *Curr. Opin. Neurobiol* 48, 52–58. [PubMed: 29028540]
- Wang J, Al-Ouran R, Hu Y, Kim S-Y, Wan Y-W, Wangler MF, Yamamoto S, Chao H-T, Comjean A, Mohr SE, et al. (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. [PubMed: 28502612]
- Wangler MF, Yamamoto S, Chao H-T, Posey JE, Westerfield M, Postlethwait J, Members of the Undiagnosed Diseases Network (UDN), Hieter P, Boycott KM, Campeau PM, et al. (2017). Model Organisms Facilitate Rare Disease Diagnosis and Therapeutic Research. *Genetics* 207, 9–27. [PubMed: 28874452]
- Wheeler GN, and Brändli AW (2009). Simple vertebrate models for chemical genetics and drug discovery screens: lessons from zebrafish and *Xenopus*. *Dev. Dyn* 238, 1287–1308. [PubMed: 19441060]
- Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, Charlson FJ, Norman RE, Flaxman AD, 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. [PubMed: 23993280]
- Willsey AJ, and State MW (2015). Autism spectrum disorders: from genes to neurobiology. *Curr. Opin. Neurobiol* 30, 92–99. [PubMed: 25464374]
- Willsey AJ, Fernandez TV, Yu D, King RA, Dietrich A, Xing J, Sanders SJ, Mandell JD, Huang AY, Richer P, et al. (2017). De Novo Coding Variants Are Strongly Associated with Tourette Disorder. *Neuron* 94, 486–499 e9. [PubMed: 28472652]
- Yamamoto S, Jaiswal M, Charng W-L, Gambin T, Karaca E, Mirzaa G, Wiszniewski W, Sandoval H, Haelterman NA, Xiong B, et al. (2014). A drosophila genetic resource of mutants to study mechanisms underlying human genetic diseases. *Cell* 159, 200–214. [PubMed: 25259927]
- Yu MK, Kramer M, Dutkowski J, Srivas R, Licon K, Kreisberg J, Ng CT, Krogan N, Sharan R, and Ideker T (2016). Translation of Genotype to Phenotype by a Hierarchy of Cell Subsystems. *Cell Syst* 2, 77–88. [PubMed: 26949740]
- Zaidi S, Choi M, Wakimoto H, Ma L, Jiang J, Overton JD, Romano-Adesman A, Bjornson RD, Breitbart RE, Brown KK, et al. (2013). De novo mutations in histone-modifying genes in congenital heart disease. *Nature* 498, 220–223. [PubMed: 23665959]
- Zerbo O, Qian Y, Yoshida C, Fireman BH, Klein NP, and Croen LA (2017). Association Between Influenza Infection and Vaccination During Pregnancy and Risk of Autism Spectrum Disorder. *JAMA Pediatr.* 171, e163609. [PubMed: 27893896]
- 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. [PubMed: 22219721]
- Zimmerman LB, De Jesús-Escobar JM, and Harland RM (1996). The Spemann organizer signal noggin binds and inactivates bone morphogenetic protein 4. *Cell* 86, 599–606. [PubMed: 8752214]

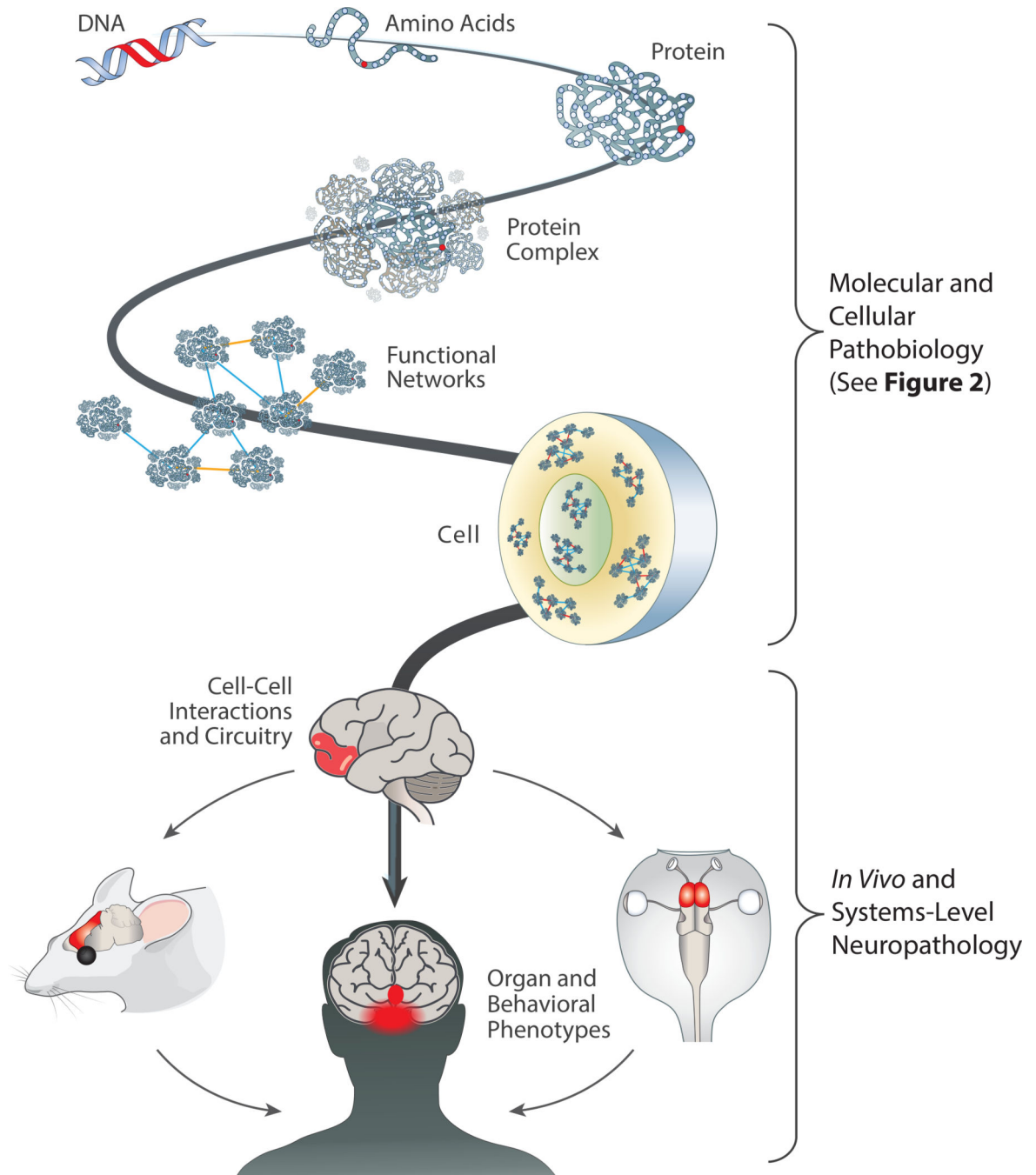


Figure 1 - Levels of Pathogenesis and Analysis. This figure outlines the different levels on which a disorder could manifest or be investigated, starting from a genetic variant. While much headway has been made in characterizing the genetic architecture of neuropsychiatric disorders, many of the other 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.

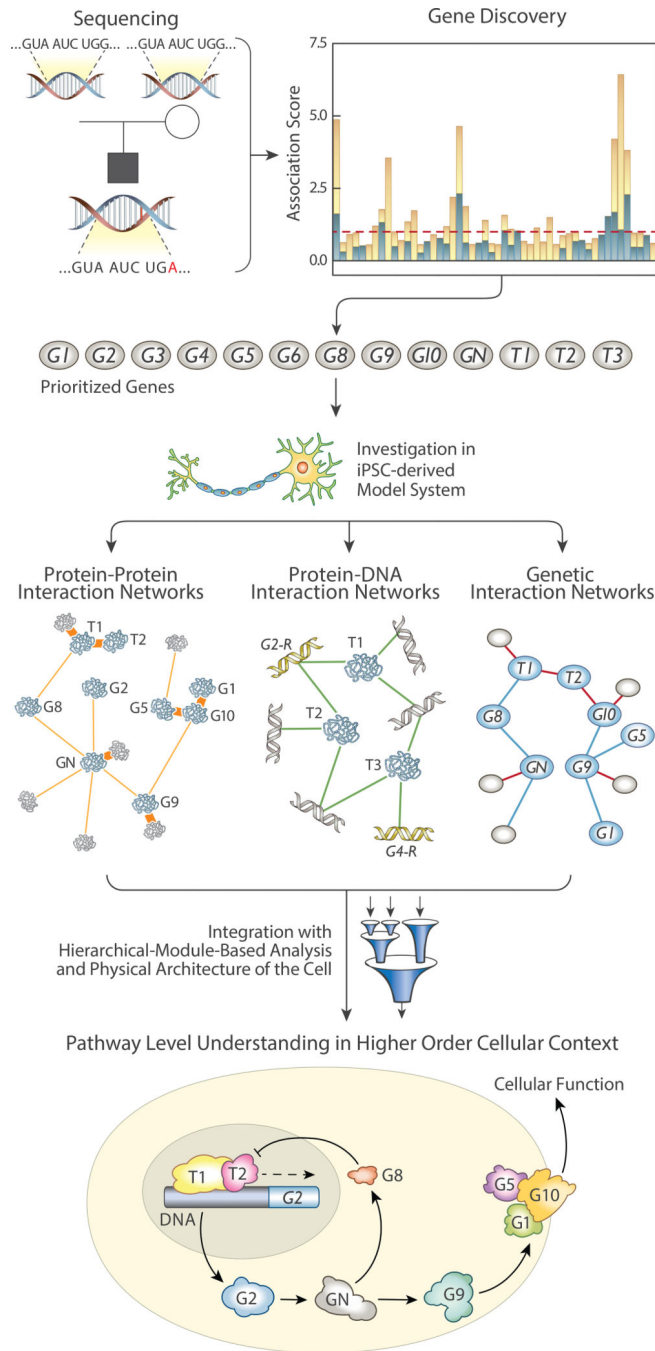


Figure 2 - Conceptual Overview of a Bottom-up Approach. The objective of a ‘bottom-up’ approach, such as the one depicted here, is to translate genetic findings to specific insights about pathobiology of neuropsychiatric disorders. The first step is the identification of specific genes. To date, the majority of gene-level discoveries have been made through the study of high effect-size de novo variants (top left). An example association graph is displayed at the top right, with contributions from different classes of variants shown per gene (yellow versus blue; e.g. missense versus nonsense variants). G1–10 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. 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. These hypotheses should be followed up cellular phenotyping and validation experiments.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

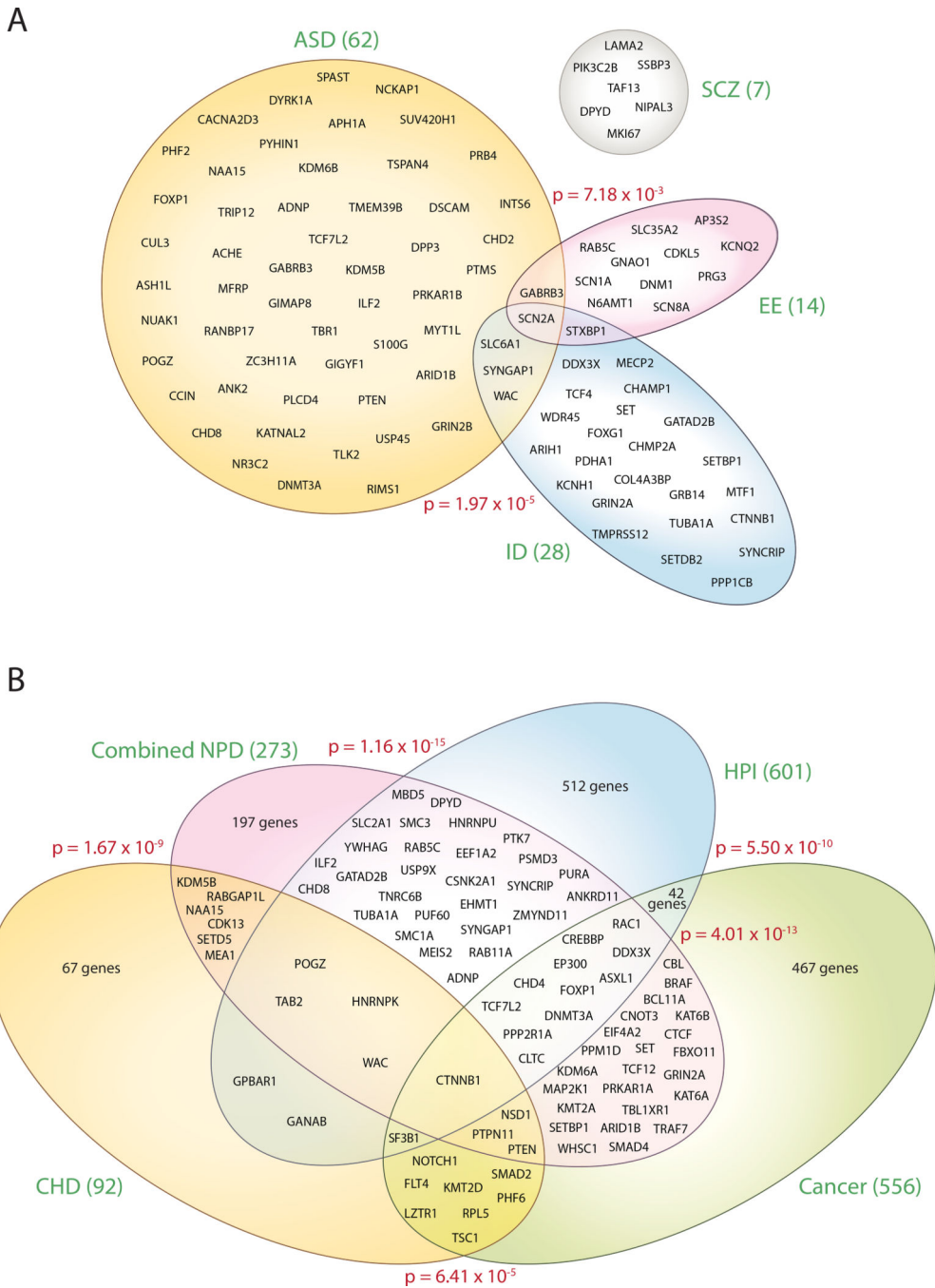


Figure 3 - Neuropsychiatric Disorder Genes Overlap with Other Disorders and Initiatives. **(A)** Risk genes for neuropsychiatric disorders (NPDs) overlap. We utilized a common pipeline to identify high confidence genes (false discovery rate (FDR) < 0.1) for autism spectrum disorders (ASD), intellectual 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 comorbid 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 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. Finally, we derived a list of Cancer genes from the Cancer Gene Census (Futreal et al., 2004) and HPI genes from the HPIDB 2.0 database (Ammari et al., 2016) (Methods and Table S1). High confidence NPD genes significantly overlap with each of the other three gene sets (Table S2 and Table 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.