11

Contextualizing Convergent Common Variant Mechanisms through Systems Biology

Michael J. Gandal

Abstract

Psychiatric disorders are highly polygenic, with estimated contributions from hundreds to thousands of causal variants, across the allelic spectrum. Interpretation of such a widely distributed genetic risk architecture is a daunting challenge, as no single locus can explain disease etiology, yet it is also critical for mechanistic understanding and clinical translation. Systems biology can begin to contextualize genetic risk variation within our understanding of the hierarchical organization of the human brain, encompassing its cognate underlying cellular pathways and gene regulatory networks, cell types and states, cell–cell interactions, circuit-level function, and ultimately behavior. This chapter provides an overview of how high-throughput molecular "omic" profiling coupled with network-level inference can provide a framework for biological contextualization of established genetic risk factors to elucidate convergent disease mechanisms. Successes are highlighted leveraging systems biology to prioritize synaptic and chromatin complex genes, and next steps are enumerated to further the translational utility of these approaches.

Introduction

Large-scale genetic and genomic studies have now successfully identified hundreds of genetic loci robustly associated with risk for neuropsychiatric disorders. In schizophrenia (SCZ) and autism spectrum disorder (ASD), there are now well-established genome-wide significant contributions from common variants, recurrent large copy number variants (CNVs), and genes harboring rare protein-disrupting variants (Gandal et al. 2016; Marshall et al. 2017; Satterstrom et al. 2020; Singh et al. 2022; Trubetskoy et al. 2022). As gene discovery continues to move at a rapid pace, fueled by increasing cohort size and decreasing genotyping costs, the slow translation of associated variation into

From "Exploring and Exploiting Genetic Risk for Psychiatric Disorders," edited by Joshua A. Gordon and Elisabeth B. Binder. Strüngmann Forum Reports, vol. 31, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262547383 (paperback) 9780262377423 (pdf)

concrete molecular mechanisms—and ultimately therapeutic targets—remains a critical obstacle. This is particularly challenging in the context of daunting levels of polygenicity and incomplete penetrance, in which two unique affected individuals likely harbor distinct risk variant profiles. Consequently, it becomes imperative to decipher the convergent biological impact of multiple risk variants, both at the population level, to understand genetic pathophysiology more broadly, as well as within a given affected individual, to achieve the promise of "precision" medicine (Gandal et al. 2016).

Here, we address these challenges in the context of two major neuropsychiatric disorders—ASD and SCZ—which have similar estimates of heritability (~70%) and have been relatively well studied in terms of genetic risk architecture, functional genomics, and transcriptomics. Among neuropsychiatric disorders, ASD is the most advanced in terms of rare variant discovery from whole-exome sequencing (WES), whereas SCZ has the most well-powered genome-wide association study (GWAS), and both have well-established associations with recurrent CNVs (Gandal et al. 2016; Sullivan and Geschwind 2019). The principles described here are applicable for interpretation of genetic risk in other disorders, as gene discovery efforts catch up. Finally, discussion is limited to human genetics and functional genomics, as other chapters in this volume are devoted to approaches involving experimental and model systems.

State of Convergence

How can we begin to disentangle the relationship between hundreds to thousands of unique variants on complex brain-level cognitive and behavioral phenotypes? A key insight comes from the observation that for complex polygenic disorders, risk genes and molecules, although dispersed throughout the genome, often coalesce within specific "core" molecular pathways and cellular networks (Gilman et al. 2012; Parikshak et al. 2015). For psychiatric disorders, some of the first such glimpses were observed among rare syndromic forms of ASD, which appeared to converge at the synapse (Zoghbi 2003). As additional high-confidence ASD risk variants were subsequently uncovered through copy number variant profiling and WES, other pathways were implicated, including chromatin remodeling and gene regulation, in addition to synapse formation, neuronal cell adhesion, ubiquitination pathways, and targets of Fragile X mental retardation protein (FMRP), among others (De Rubeis et al. 2014; Glessner et al. 2009; Iossifov et al. 2014; Pinto et al. 2010; Sanders et al. 2015; Satterstrom et al. 2020). In SCZ, although the overall contribution of rare loss-of-function variation is smaller, similar convergence has been observed among synaptic and chromatin gene sets (Figure 11.1), as well as glutamate signaling, and FMRP targets, in particular (Fromer et al. 2014; Purcell et al. 2014; Singh et al. 2022).

From "Exploring and Exploiting Genetic Risk for Psychiatric Disorders," edited by Joshua A. Gordon and Elisabeth B. Binder. Strüngmann Forum Reports, vol. 31, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262547383 (paperback) 9780262377423 (pdf)

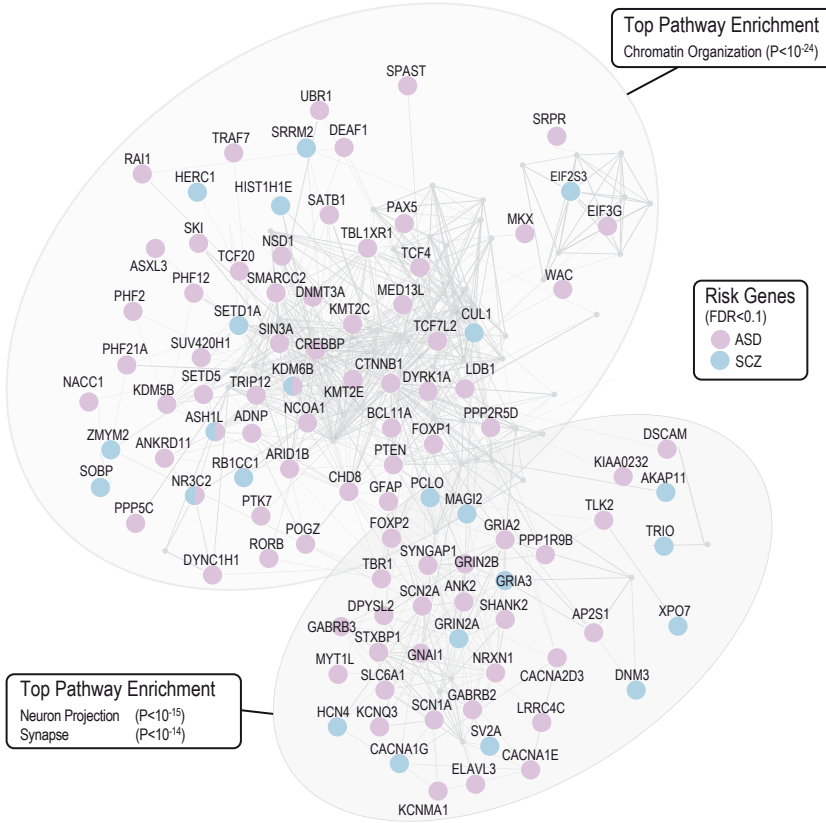


Figure 11.1 An interactome network of autism spectrum disorder (ASD) and schizophrenia (SCZ) risk genes highlights chromatin complex and synaptic gene clusters. A protein–protein interaction network seeded with rare variant implicated risk genes for ASD (n = 102 at FDR < 0.1; Satterstrom et al. 2020) and SCZ (n = 34 at FDR < 0.1; Singh et al. 2022) was built using StringDB (Szklarczyk et al. 2021). Two highly connected subclusters of the interactome network show distinct enrichments for chromatin organization and synapse pathways, as previously identified by (Sanders et al. 2015).

On the common variant side, the use of psychiatric GWAS to perform heritability enrichments (Finucane et al. 2015) and functional/proximity mapping (Watanabe et al. 2017) largely implicates similar pathways. Psychiatric GWASs are predominately enriched for brain tissues as well as neuronal cell types (Finucane et al. 2018; Gandal et al. 2018a; Skene et al. 2018; Xu et al. 2014). The most recent SCZ GWAS, which has the greatest power of any individual psychiatric GWAS, showed enrichment for gene sets that were largely synaptic and relatively nonspecific, such as "postsynaptic specialization" and "ion channel complex" (Trubetskoy et al. 2022). Cross-disorder

From "Exploring and Exploiting Genetic Risk for Psychiatric Disorders," edited by Joshua A. Gordon and Elisabeth B. Binder. Strüngmann Forum Reports, vol. 31, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262547383 (paperback) 9780262377423 (pdf)

psychiatric GWAS analyses further implicate synaptic, immune, and histone gene sets, as well as chromatin regulation in the developing human brain (Cross-Disorder Group of the Psychiatric Genomics Consortium 2019; Network Pathway Analysis Subgroup of Psychiatric Genomics Consortium 2015; Schork et al. 2019).

While these findings of "genetic convergence" provide a starting context for understanding pathobiology and a foothold for hypothesis-driven experimental dissection, they only begin to scratch the surface in terms of pinpointing overarching biological risk mechanisms with any degree of specificity. Convergence is almost always investigated at the population level, rather than within an individual where such findings are potentially much more actionable. These gene sets largely reflect manually curated or annotated pathways, often defined outside of the nervous system, although many are known to have distinct functional roles in the brain. Finally, gene set enrichments are typically viewed in isolation, rather than within the dynamic, diverse, and interconnected network that defines the CNS. Here, we discuss how systems biology can provide an organizing framework to interpret genetic convergence for psychiatric disorders and to provide modular, data-driven annotations for gene functions within the CNS.

Molecular Hierarchies and the Human Brain

The development of the human brain is under precise molecular genetic control, and the underlying molecular, cellular, genetic, and epigenetic regulatory landscape exhibits tight spatiotemporal regulation (Silbereis et al. 2016). As such, characterizing the dynamic patterns of gene expression and epigenetic changes in the brain throughout development can provide insights into the underlying regulatory logic defining neuronal cell types, states, as well as their maturation into functional neural circuits (Kang et al. 2011; Li et al. 2018). Along these lines, spurred by the rapid advent of high-throughput multi-omic profiling technologies, consortium-level efforts have been successful in mapping the whole-tissue transcriptome, epigenome, and proteome across developmental stages, brain regions, and sexes in neurotypical and psychiatric disease samples (Carlyle et al. 2017; Fromer et al. 2016; Gandal et al. 2018b; GTEx Consortium 2020; Hawrylycz et al. 2012; Jaffe et al. 2016, 2018; Kang et al. 2011; Li et al. 2018; Miller et al. 2014; Wang et al. 2018); see Figure 11.2.

The human brain has been most extensively genomically profiled at the level of the transcriptome, first with microarrays followed by bulk RNA sequencing and now with single-cell and spatial transcriptomic technologies. Initial seminal work from BrainSpan used microarrays to map gene expression trajectories across 1,340 tissue samples spanning 16 brain regions and 15 developmental timepoints, from 6 weeks postconception to 82 years (Kang et al. 2011). Later extended with RNA sequencing as part of PsychENCODE, these data

From "Exploring and Exploiting Genetic Risk for Psychiatric Disorders," edited by Joshua A. Gordon and Elisabeth B. Binder. Strüngmann Forum Reports, vol. 31, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262547383 (paperback) 9780262377423 (pdf)

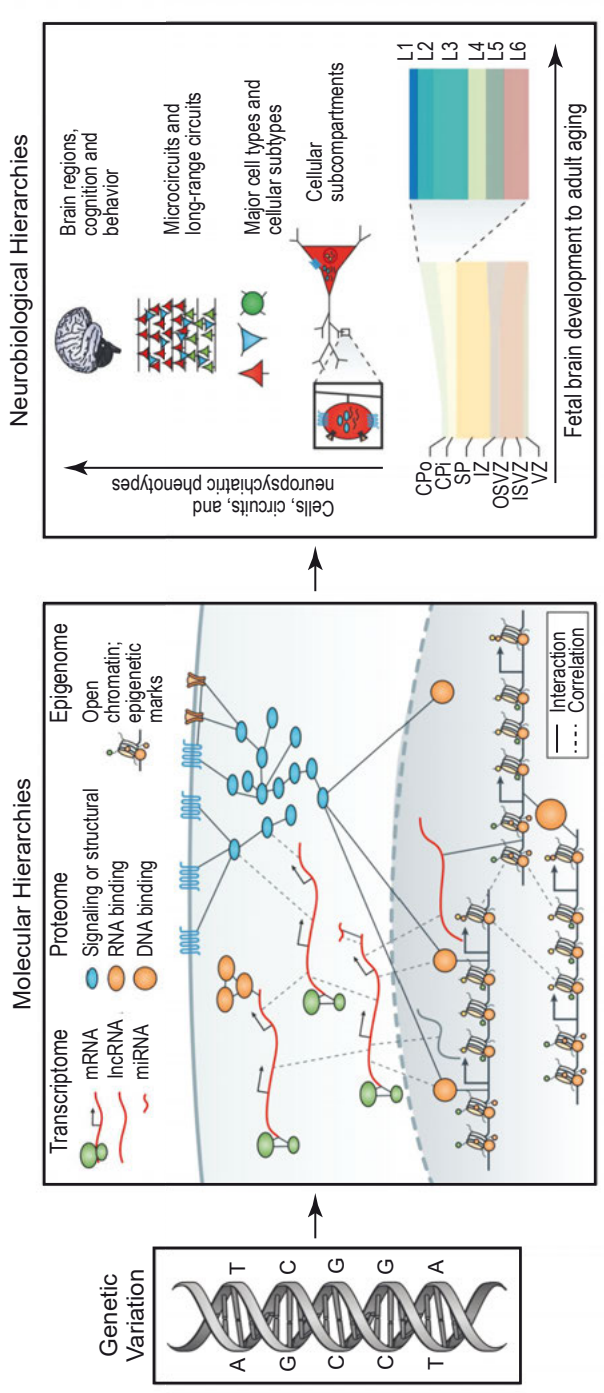


Figure 11.2 Molecular hierarchies and the human brain. Large-scale multi-omic profiling of human brain tissue samples has been undertaken to connect single nucleotide polymorphism-level genetic variation with transcriptomic, epigenomic, and proteomic changes. Reproduced with permission from Parikshak et al. (2015).

From "Exploring and Exploiting Genetic Risk for Psychiatric Disorders," edited by Joshua A. Gordon and Elisabeth B. Binder. Strüngmann Forum Reports, vol. 31, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262547383 (paperback) 9780262377423 (pdf)

highlighted substantial spatiotemporal gene expression changes distinguishing pre- and postnatal brain samples, as well as a notable late-fetal transition (Li et al. 2018). The Allen Brain Institute generated comprehensive regional gene expression atlases of both adult and fetal brains. Gene expression was characterized across ~900 precise subregions in the adult brain (Hawrylycz et al. 2012) and ~300 regions in the mid-fetal human brain (Miller et al. 2014). The fetal atlas identified notable transcriptomic laminar differences capturing cellular maturation between proliferative and postmitotic layers, as well as clear gradients in cortical patterning along a frontotemporal axis. The adult atlas captured enormous transcriptomic variation by anatomic location, and although the neocortex exhibited relatively homogeneous transcriptomic signatures, there were again clear gradients along frontal-occipital axes. Notably, these typical gradients in cortical arealization have been shown to be attenuated in postmortem studies of ASD and SCZ (Haney et al. 2020; Parikshak et al. 2016; Roussos et al. 2012; Voineagu et al. 2011). Additional large-scale efforts from GTEx, CommonMind, BrainSeq, and PsychENCODE consortia, among others, have been undertaken to connect brain gene expression with single nucleotide polymorphism (SNP)-level genetic variation through *cis*-expression quantitative trait loci (*cis*-eQTL) (Fromer et al. 2016; GTEx Consortium 2020; Jaffe et al. 2018; Wang et al. 2018). Similar, albeit smaller, efforts have also mapped *cis*-eQTLs in the developing human brain (O'Brien et al. 2018; Walker et al. 2019; Werling et al. 2020). Finally, the advent of single-cell/single nucleus (sc/sn) RNA sequencing has enabled the bottom-up categorization of the underlying neural cell types and their developmental trajectories in the fetal (Nowakowski et al. 2017; Polioudakis et al. 2019) and adult human brain (Lake et al. 2018; Li et al. 2018). Leveraging these transcriptomic approaches and data sets, nearly all well-powered psychiatric genetic association studies show strong functional enrichment for brain genomic annotations, including brain-expressed genes and *cis*-eQTLs, with enrichment among cell types defined by sc/snRNA sequencing largely implicating neuronal lineages in genetic risk for SCZ and ASD (Calderon et al. 2017; Finucane et al. 2018; Satterstrom et al. 2020; Skene et al. 2018). Intriguingly, cell types defined by open chromatin regions (scATAC sequencing) captured substantially more disease heritability than those defined from scRNA sequencing (Kim et al. 2021b).

Genome-wide profiling has also begun to illuminate the molecularly defined epigenetic landscape of the human brain, which has been particularly critical for functional interpretation of the noncoding, regulatory regions of the genome where the majority of GWAS hits occur (PsychEncode Consortium et al. 2015). Girdhar et al. (2018) used ChIP-Seq to profile histone marks of active promoters (H3K4me3) and enhancers (H3K27ac) across two cortical regions from 157 neuronal (NeuN+), neuron-depleted (NeuN–), and bulk-tissue samples. Neuronal enhancers from the adult brain showed the strongest enrichment for SCZ GWAS signal, more so than bulk tissue or nonneurons. DNA

From "Exploring and Exploiting Genetic Risk for Psychiatric Disorders," edited by Joshua A. Gordon and Elisabeth B. Binder. Strüngmann Forum Reports, vol. 31, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262547383 (paperback) 9780262377423 (pdf)

methylation is another epigenetic signature exhibiting dynamic patterns in the human brain across development, regions, and cell types (Hannon et al. 2016; Jaffe et al. 2016; Rizzardi et al. 2019). Patterns of CpG methylation—particularly those that varied across brain regions, in neuronal (vs. nonneuronal) cell types, and fetal developmental periods—overlap substantially with GWAS loci for SCZ and other psychiatric disorders. Neuronal CpG sites exhibiting brain regionally variable methylation patterns captured the greatest degree of psychiatric heritability, but overlapped substantially with other measures of neuron-specific epigenetic regulation (Rizzardi et al. 2019). As discussed further below, these types of direct head-to-head comparisons between distinct molecular readouts are critical to build an integrative understanding of their biological interrelationships as well as to distill nonredundant insights into underlying genetic risk mechanisms.

Analyses of the alternative splicing landscape in the brain, which is particularly extensive compared with other tissues (Garrido-Martín et al. 2021; GTEx Consortium 2020), point to a strong contribution of genetic risk variants to splicing dysregulation, relatively orthogonal to effects on gene expression (Gandal et al. 2018b; Li et al. 2016; Takata et al. 2017). RNA-binding proteins coordinate many aspects of posttranscriptional regulation, including splicing, as well as subcellular localization, RNA stabilization and translational control, which are particularly important in neurons where transcripts are trafficked for long distances (Darnell 2013). Indeed, a number of RNA-binding proteins are themselves strong neurodevelopmental disorder risk genes, including *FMR1*, *RBFOX1*, *CHD8*, and *CELF4*, and their experimentally defined targets (Van Nostrand et al. 2020) exhibit strong enrichment for psychiatric GWAS signals (Park et al. 2021).

Advances in mass spectrometry have now begun to enable high-throughput proteomic profiling, although the sensitivity and dynamic range remains relatively limited compared with other genomic readouts. Large-scale, bottom-up proteomic profiling has been conducted across regions and postnatal timepoints in the human brain (Carlyle et al. 2017), as well as in SCZ case/control cohorts (MacDonald et al. 2020). Indeed, profiling of the synaptic proteome in auditory cortex from individuals with SCZ and controls (n = 48/group) identified significant alterations in >100 synaptosomal and homogenate protein levels, with a weak—but significant—correlation in effect size compared with transcriptomic changes observed in SCZ cortex (MacDonald et al. 2020). Further, "target capture"-based proteomics can be used to identify protein–protein interactions (PPIs), uncovering, for example, specific macromolecular complexes and interacting sub-networks among high-confidence ASD risk genes in cultured human neurons (Li et al. 2015; Pintacuda et al. 2021). While less scalable, such approaches complement existing PPI databases, which generally lack cell type and tissue (especially brain) specificity. Finally, although the genetic control of the human brain proteome has only recently begun to be elucidated through pQTL profiling (Robins et al. 2021), it has recently been

From "Exploring and Exploiting Genetic Risk for Psychiatric Disorders," edited by Joshua A. Gordon and Elisabeth B. Binder. Strüngmann Forum Reports, vol. 31, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262547383 (paperback) 9780262377423 (pdf)

integrated with psychiatric GWAS, for example prioritizing novel candidate risk proteins underlying depression GWAS loci, several of which were not captured at the transcriptome level (Wingo et al. 2021).

Finally, single-cell and cell type-specific genomic profiling are now feasible at a cost and scale necessary for capturing population-level allelic effects. Jaffe et al. (2020) profiled the granule-cell layer of the dentate gyrus, along with bulk hippocampal tissue, from 263 postmortem donors to generate cell type-specific QTL maps. Cell type-specific QTLs, a substantial fraction (15%) of which were not detectable in bulk tissue profiled from the same individuals, were used to prioritize *GRM3* and *CACNA1C* as risk genes within SCZ GWAS loci. Similar approaches, now being undertaken using snRNA sequencing from human brain samples, are beginning to uncover context-specific gene regulation across a wider range of cell types, with SCZ colocalization observed most substantially in excitatory neurons (Bryois et al. 2022).

Network-Level Inference

Given that so many features and layers of molecular regulation in the human brain contribute to psychiatric risk, how can we organize and integrate them into a coherent and interpretable set of functional units? Here, systems-level network biology provides a powerful organizing framework, which has been extensively leveraged particularly for brain transcriptomic data sets, as well as for protein interactomes (McGillivray et al. 2018; Parikshak et al. 2015). Network models are ubiquitous in biology, depicting connections between nodes (e.g., genes, proteins, regulatory elements) with edges defined by biologically measured or inferred relationships (e.g., co-expression or co-regulation, physical binding). Network topology can subsequently be characterized by patterns of connectivity and modularity. For example, many biological networks, including gene co-expression, exhibit scale-free topology, in which there are a few highly interconnected "hub" genes and many genes/nodes with few connections. Within this framework, nodes can be clustered into a small, discrete set of interconnected modules capturing the major axes of variation. Within each module, the most interconnected "hub" genes can be used to infer biological function, characterized based on enrichment for known cell types, gene ontology pathways, protein complexes, transcription factor binding sites, or other types of regulatory relationships (Kuleshov et al. 2016).

To date, network approaches have been most extensively leveraged in the context of large, bulk-tissue gene expression compendia, using techniques like Weighted Gene Correlation Network Analysis (WGCNA) and others (Langfelder and Horvath 2008). These approaches have revealed a robust, hierarchical organization to the human brain transcriptome, with co-expression modules recapitulating the unique cell types, subcellular organelles, and region-, sex-, and developmentally regulated processes (Gandal et al. 2018a, b;

From "Exploring and Exploiting Genetic Risk for Psychiatric Disorders," edited by Joshua A. Gordon and Elisabeth B. Binder. Strüngmann Forum Reports, vol. 31, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262547383 (paperback) 9780262377423 (pdf)

Hawrylycz et al. 2012; Kang et al. 2011; Oldham et al. 2008). Co-expression networks provide a natural, data-driven means for organizing long lists of observed differentially expressed genes into modules, with presumed shared biological regulation and/or function. Indeed, co-expression modules built from large, human bulk brain data sets have been shown to capture all major CNS cell classes, with hub genes consisting of well-established cell type markers (Gandal et al. 2018a; Kelley et al. 2018). In practice, this enables *in silico* dissection of cell type-specific expression and/or cell proportion changes across conditions, as discussed further below. As proteomic profiling catches up with next-generation sequencing technologies, in terms of sensitivity and dynamic range, network-based analyses of the human brain proteome will provide important additional biological layers. Indeed, protein co-regulation networks from SCZ and control brain samples identified SCZ-downregulated modules representing synaptic mitochondria, very similar to what was observed at the gene expression level.

PPI-based networks, which define the interaction between two proteins (nodes) with binary edges representing physical binding interactions, have also been extensively characterized and leveraged in genomic analyses. These networks generally leverage literature-curated PPI databases, which compile results from experiments like yeast two-hybrid screens and immunoprecipitation followed by proteomics. As such, these networks can be sparse and incomplete, with established biases toward well-studied proteins, and generally lack tissue (e.g., brain) specificity (Corominas et al. 2014). Nevertheless, PPI networks more directly define macromolecular complexes than the guilt-by-association framework of co-expression. In addition, networks defined by PPIs show significantly increased co-expression (and vice versa), indicating concordance (Gandal et al. 2018a; Parikshak et al. 2015; Sakai et al. 2011).

Finally, gene regulatory networks are directional networks that map connecting genes and their regulators. Typically, gene regulatory networks integrate hierarchical relationships to predict gene expression by linking transcription factors to target *cis*-regulatory elements (e.g., enhancers and promoters) and target genes through experimentally defined binding site motifs. Typically, gene regulatory networks leverage the tissue and/or cell type specificity of such interactions, which can be inferred from comprehensive databases of epigenomic annotations (Marbach et al. 2016). In a comparison with PPI and gene co-expression networks, tissue-specific gene regulatory networks captured greater enrichment for SCZ and cross-disorder psychiatric GWAS signal, which most strongly implicated striatal and cortical tissues (Marbach et al. 2016). More recently as part of PsychENCODE, Wang et al. (2018) developed a comprehensive regulatory network for the human brain, linking 42,681 enhancers to target genes via eQTL and 3D chromatin conformation contacts. Approximately 43,000 transcription factor to target gene connections were then incorporated based on transcription factor binding site compatibility and regularized (elastic net) regression, which related transcription factor

From "Exploring and Exploiting Genetic Risk for Psychiatric Disorders," edited by Joshua A. Gordon and Elisabeth B. Binder. Strüngmann Forum Reports, vol. 31, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262547383 (paperback) 9780262377423 (pdf)

expression with that of the predicted target gene. Integrating this gene regulatory network with SCZ GWAS results prioritized specific transcription factors in disease risk, including *SOX7*, and further implicated excitatory neurons. Building true cell type-specific gene regulatory networks (Aibar et al. 2017) to leverage the wealth of emerging snRNA sequencing and multi-omic data from the human brain will be critical as these regulatory relationships are known to be highly context specific.

What Has Systems Biology Taught Us about Psychiatric Genetics?

Cellular-Spatial-Temporal Context

Leveraging the emerging wealth of data-driven functional annotations for the human brain coupled with network-based contextualization, several early papers demonstrated the promise of such "convergent" approaches to localizing the context in which diverse risk genes overlap. To pinpoint convergence at a molecular level, early exome-sequencing studies in ASD integrated results within protein interactome networks, identifying highly connected clusters including specific chromatin remodeling complexes (De Rubeis et al. 2014; Li et al. 2015; O'Roak et al. 2012b), WNT/β-catenin signaling (O'Roak et al. 2012b), synaptic genes (De Rubeis et al. 2014), and FMRP targets (De Rubeis et al. 2014; Iossifov et al. 2012). Further, given that many individual high-confidence ASD risk genes (e.g., *CHD8*) encode transcriptional regulators, gene regulatory networks built from the experimentally defined genomic targets for 26 such ASD-associated regulatory proteins showed strong, convergent enrichment for additional genetic signal (Satterstrom et al. 2020). Of note, while these direct (e.g., *cis*) regulatory targets implicate additional risk genes, stronger enrichments have been observed among the indirect (e.g., *trans*) targets—those genes that are downregulated upon *CHD8* knockdown but which are not direct targets of *CHD8* (Sugathan et al. 2014).

To place risk genes within a relevant spatiotemporal context, several studies leveraged BrainSpan (Kang et al. 2011) to build gene co-expression networks from neurotypical brains spanning early developmental epochs, finding that high-confidence ASD risk genes showed convergence within mid-fetal, prefrontal cortex networks and glutamatergic neurons (Ben-David and Shifman 2013; Parikshak et al. 2013; Willsey et al. 2018). Such findings were replicated using larger, updated high-confidence rare variant-implicated ASD risk genes, and, with the incorporation of scRNA sequencing data from the developing human brain, now have resolution to detect strongest enrichment among both early excitatory neuron and striatal interneuron lineages, among others (Li et al. 2018; Satterstrom et al. 2020). Similar findings were initially reported in SCZ, linking rare *de novo* variants to co-expression networks in

From "Exploring and Exploiting Genetic Risk for Psychiatric Disorders," edited by Joshua A. Gordon and Elisabeth B. Binder. Strüngmann Forum Reports, vol. 31, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262547383 (paperback) 9780262377423 (pdf)

the mid-fetal prefrontal cortex (Gulsuner et al. 2013). The more recent SCZ rare variant sequencing study from SCHEMA, however, did not observe a prenatal expression bias for the ten high-confidence SCZ risk genes (Singh et al. 2022).

Several tools have been developed to characterize GWAS enrichment patterns among specific cell types, tissues, brain regions, developmental time points, and co-expressed gene sets (Calderon et al. 2017; Finucane et al. 2018; Pers et al. 2015; Skene et al. 2018; Watanabe et al. 2017, 2019b; Xu et al. 2014). For both ASD and SCZ, there is evidence implicating GWAS signal within developmental—particularly mid-fetal—timepoints, although some genetic risk factors clearly also act postnatally, and the current largest ASD GWAS remains relatively underpowered to detect strong enrichments (Calderon et al. 2017; Grove et al. 2019; Parikshak et al. 2013; Walker et al. 2019). Spatially, genetic risk for these disorders appears to be distributed brain-wide, including across the cortex and cerebellum, rather than exhibiting any strong regional specificity (Haney et al. 2020; Hartl et al. 2021; Krishnan et al. 2016). Cell type-specific enrichment patterns for SCZ largely parallel those seen with rare variants in ASD, with clear enrichment for neuronal lineages but limited specificity beyond that (Skene et al. 2018). Intriguingly, greater specificity will likely be achieved with more detailed cell type specifications leveraging single-cell multi-omic profiling, which highlights a more prominent contribution to bipolar disorder GWAS risk from deep layer excitatory neurons (Luo et al. 2022).

Network-Informed Discovery and Interpretation of Risk Genes

Prioritizing candidate psychiatric risk genes from both WES and GWAS remains, in many cases, a major challenge. Index SNPs from GWAS typically fall within noncoding regions of the genome and often tag large haplotype blocks, obscuring both the identity of the true "causal" variant(s) as well as their target gene. In WES, interpreting the pathogenicity of identified variants, in particular missense variants which comprise the majority of coding mutations, remains difficult. In both cases, network-based approaches that incorporate functional genomic annotations have been successfully leveraged to improve prioritization and contextualization of disease-associated mutations as well as to increase power (Leiserson et al. 2013).

Among the most powerful demonstrations of network-informed risk gene discovery is the DAWN framework, which uses hidden Markov random fields to prioritize disease-associated gene clusters that exhibit strong patterns of co-expression in a relevant disease context (Liu et al. 2014). Leveraging the established convergence of ASD genetic risk within mid-fetal cortex gene networks, DAWN substantially boosts power in rare variant-sequencing studies of ASD, prioritizing dozens of additional risk genes (De Rubeis et al. 2014), many of which have since been replicated in subsequent, larger sequencing studies

From "Exploring and Exploiting Genetic Risk for Psychiatric Disorders," edited by Joshua A. Gordon and Elisabeth B. Binder. Strüngmann Forum Reports, vol. 31, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262547383 (paperback) 9780262377423 (pdf)

(Satterstrom et al. 2020). Along similar lines, Krishnan et al. (2016) showed that incorporating machine learning with a brain-specific Bayesian gene-interaction network (comprised of gene expression, PPI, and regulatory-sequence based data sets) enhanced prediction of ASD risk genes as well as subsequent functional characterization of associated pathways. The strongest enrichments were observed for postsynaptic density genes and FMRP targets, and clustering further implicated pathways underlying genetic risk for ASD, including chromatin remodeling, WNT/β-catenin signaling and mRNA splicing, among others. Finally, Chen et al. (2018) leveraged an interactome network-based approach to facilitate interpretation of ASD-associated missense mutations, the pathogenicity of which can be difficult to decipher, by prioritizing mutations that affect the binding interfaces of hub proteins within a PPI network. These missense mutations further clustered with previously identified genes harboring *de novo* protein-truncating variants in ASD as well as with other relevant gene sets, including FMRP targets, chromatin modifiers, genes in the PSD, and genes expressed early in development.

Defining Molecular Pathology

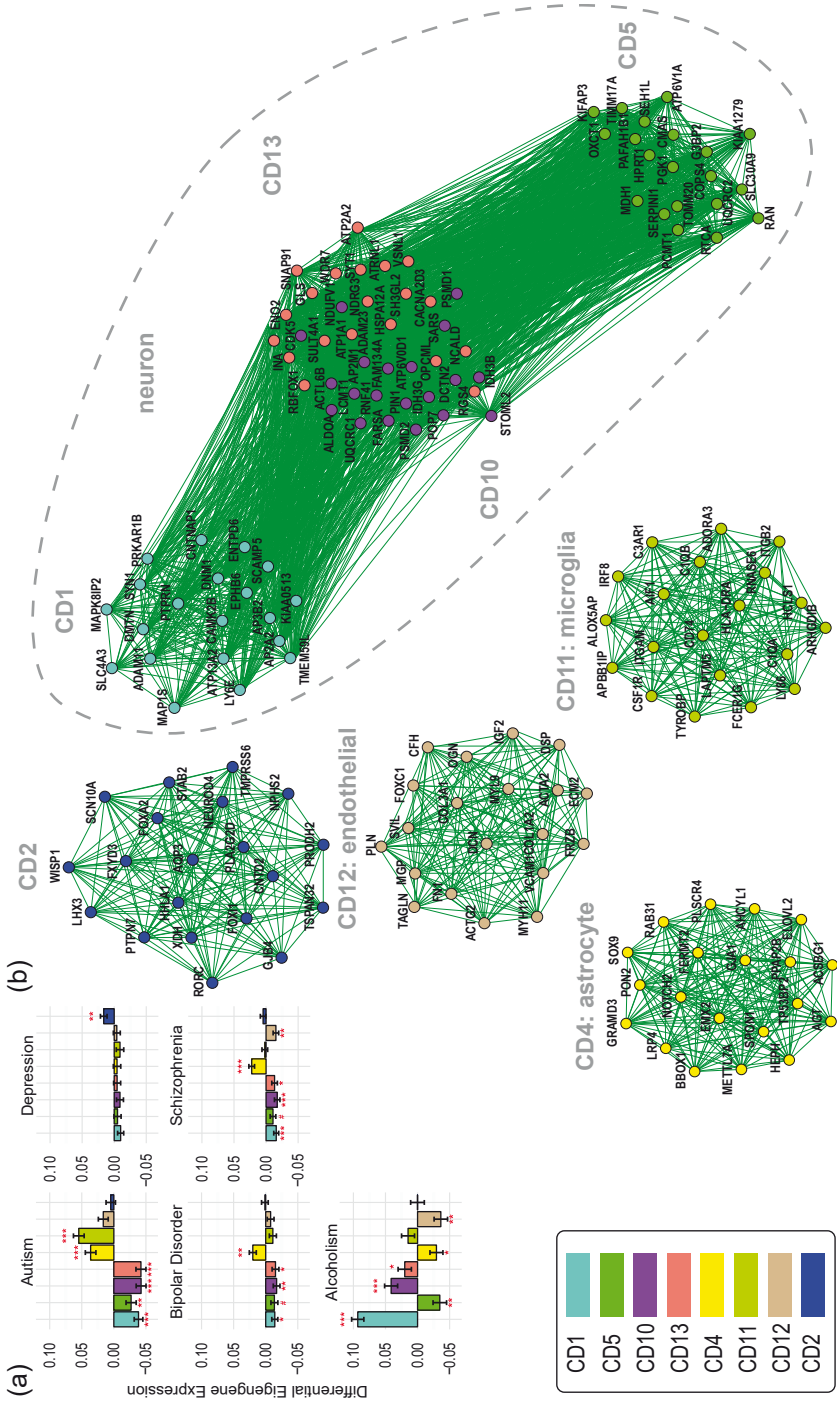
Network-based approaches have proved a useful organizing framework to interpret results from molecular profiling studies of case/control cohorts. Although psychiatric disorders lack a clearly defined neuroanatomic or cellular pathology—in contrast with many neurologic conditions—large-scale, genome-wide transcriptomic profiling has now established a characteristic brain-level molecular pathology for several psychiatric disorders, including ASD and SCZ (Fromer et al. 2016; Gandal et al. 2018a; Parikshak et al. 2016; Voineagu et al. 2011). Initial postmortem human gene expression profiling studies of psychiatric case/control cohorts were small and often reported variable results, particularly at the individual gene level; this was likely due to cohort heterogeneity, analytic or methodologic differences, or statistical noise (Hernandez et al. 2021). Nevertheless, several key findings emerged that have since been extensively replicated in larger mega- and meta-analytic studies, including a downregulation of synaptic, interneuron, and mitochondrial related genes in SCZ and ASD cortices as well as a concordant upregulation of neural-immune and inflammatory gene expression signatures (Gandal et al. 2018a, b; Horváth and Mirnics 2015; Parikshak et al. 2016; Voineagu et al. 2011). Furthermore, ASD cases showed similar, but more extreme changes, in these gene expression patterns compared with SCZ. Importantly, these early studies paved the way for large-scale consortium-level efforts including BrainSeq (Collado-Torres et al. 2019; Jaffe et al. 2018), CommonMind (Fromer et al. 2016), and PsychENCODE (Gandal et al. 2018b; Li et al. 2018; Wang et al. 2018) to perform transcriptome profiling at sufficient scale to generate reproducible results.

From "Exploring and Exploiting Genetic Risk for Psychiatric Disorders," edited by Joshua A. Gordon and Elisabeth B. Binder. Strüngmann Forum Reports, vol. 31, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262547383 (paperback) 9780262377423 (pdf)

Network-based approaches, like WGCNA, have yielded important insights into the molecular pathology of these disorders. For example, in our cross-disorder paper which reported on 700 postmortem human brain samples, including from subjects with ASD, SCZ, and bipolar disorder (Figure 11.3), co-expression modules clearly captured major CNS cell types, with hub genes consisting of canonical cell type markers, enabling an *in silico* dissection (Gandal et al. 2018a). Modules representing gene expression patterns within neurons (and synaptic mitochondria) were downregulated across the three disorders, whereas reciprocal upregulation was observed for gene expression signatures of astrocytes. Notably, a microglial module was specifically upregulated only in ASD. It remains unclear whether such findings reflect changes in underlying cell proportion and/or cell type-specific gene expression patterns, although emerging analyses using snRNA sequencing seem to indicate subtle, at best, cell fraction shifts in ASD (Velmeshev et al. 2019).

Postmortem brain transcriptomics characterize the current, reactive state of a biological sample and therefore cannot differentiate causal from reactive or compensatory effects. Integration of transcriptomic results with directional genetic anchors, through GWAS and rare variant enrichments, can provide orthogonal evidence for pathophysiological versus compensatory or reactive changes. For example, the CommonMind paper built unsigned co-expression networks from bulk RNA sequencing profiling of >500 prefrontal cortex brain samples and identified a module (M2c; 1411 genes) that captured genes differentially expressed in SCZ as well as were enriched for multiple classes of genetic risk variation, including signals from GWAS, recurrent CNVs, and rare variants (Fromer et al. 2016). This module was enriched for neuronal markers and relevant pathways including ARC and NMDA-receptor signaling, PSD genes, and FMRP targets. Further, module hub genes included the NMDAR subunit *GRIN2A* and the GABA-B receptors *GAB BR2*, both of which have subsequently been prioritized by the latest GWAS and rare variant-sequencing studies in SCZ (Singh et al. 2022; Trubetskoy et al. 2022). Similarly, in the cross-disorder paper described above, the neuron/synaptic module (CD1) downregulated in ASD and SCZ showed convergent enrichment for GWAS signal, non-synonymous *de novo* variation, and recurrent psychiatric CNVs (Gandal et al. 2018b). Finally, we expanded this approach in the PsychENCODE data set of >1,300 samples across the lifespan to build co-expression modules using transcript-isoform (along with gene-level) expression measures. Isoform-level networks captured the same processes as gene networks but added biological specificity and showed greater genetic enrichments overall. Here, an isoform-level module comprised of oligodendrocyte markers and neuron projection pathways exhibited the greatest overall GWAS enrichment for SCZ and was downregulated in ASD and SCZ (Gandal et al. 2018b).

From "Exploring and Exploiting Genetic Risk for Psychiatric Disorders," edited by Joshua A. Gordon and Elisabeth B. Binder. Strüngmann Forum Reports, vol. 31, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262547383 (paperback) 9780262377423 (pdf)



From "Exploring and Exploiting Genetic Risk for Psychiatric Disorders," edited by Joshua A. Gordon and Elisabeth B. Binder. Strüngmann Forum Reports, vol. 31, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262547383 (paperback) 9780262377423 (pdf)

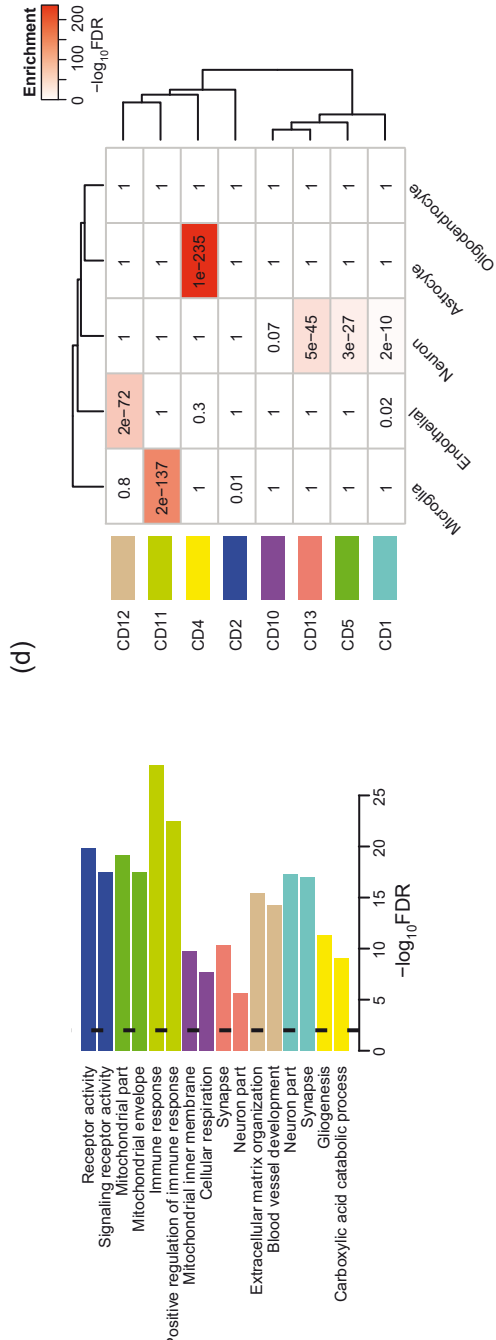


Figure 11.3 Network-level contextualization of the transcriptomic molecular pathology across major psychiatric disorders. Co-expression networks generated from 700 human brain samples enabled *in silico* dissection of shared and distinct cell type-specific gene expression alterations across disorders, with downregulation of neuronal and synaptic gene expression profiles across ASD, SCZ, and BD. The downregulated neuronal CD1 module exhibited significant, convergent enrichment for disease-associated common and rare genetic variation. Adapted from Gandal et al. (2018a).

From "Exploring and Exploiting Genetic Risk for Psychiatric Disorders," edited by Joshua A. Gordon and Elisabeth B. Binder. Strüngmann Forum Reports, vol. 31, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262547383 (paperback) 9780262377423 (pdf)

Improved Annotation of Brain-Level Pathways

A deep understanding of the key pathways mediating psychiatric genetic risk necessitates a clear definition of the genes involved in these pathways and their interrelationships. While a number of essential expert- and literature-curated databases organize genes within ontological pathways, existing resources largely fail to capture the complexity and specificity of gene functions within the nervous system. To address this gap, data-driven network approaches can leverage guilt-by-association to generate improved functional genomic and pathway annotations, particularly for tissues with highly complex and less well-characterized regulatory dynamics, such as the human brain (Parikshak et al. 2015). As an example, the most significant GWAS signal for SCZ lies within the major histocompatibility complex region, which was subsequently fine mapped to the *C4* locus and shown to reflect, in part, increased copy number and upregulation of *C4A* (Sekar et al. 2016). Although *C4A* has been well characterized as a key component of the complement cascade and the innate immune system, its functional role in the human brain remains much less understood. To address this gap, Kim et al. (2021a) characterized brain co-expression partners of *C4A* that were either positively or negatively correlated with *C4A* expression across varying *C4A* genomic copy numbers, annotating their cell type and pathway contributions as well as their relation to established SCZ genetic risk factors. This type of "seeded" network or top-down approach can provide an unbiased functional annotation for a poorly understood gene by capturing coherent biological processes that covary across samples (Parikshak et al. 2015). This work identified a putative transcriptomic signature of *C4A*-mediated synaptic pruning, reinforcing the idea that over-pruning likely contributes to SCZ pathogenesis and/or progression (Feinberg 1982). Further, this study found that negatively co-expressed genes with *C4A* are overrepresented for synapse-related pathways, which in turn were enriched for convergent SCZ genetic signals, indicating that synaptic pathways are the key biological link between genetic dysregulation of complement signaling and SCZ pathophysiology.

How Do We Move Systems Analyses of Psychiatric Genetics Forward?

As detailed above, network biology has begun to provide a coherent, organizing framework for functional interpretation of diverse genetic and genomic changes associated with psychiatric disorders. As genomic associations continue to multiply, however, much more needs to be done to enhance specificity and to move toward mechanistic insight and therapeutic target prediction within an individual. Here, I outline some critical next steps.

From "Exploring and Exploiting Genetic Risk for Psychiatric Disorders," edited by Joshua A. Gordon and Elisabeth B. Binder. Strüngmann Forum Reports, vol. 31, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262547383 (paperback) 9780262377423 (pdf)

Connecting Major Findings

Among the most striking unanswered questions that have arisen from network-level insights is what relationship, if any, exists between risk genes that cluster within the two most enriched psychiatric risk pathways (chromatin complex and synaptic gene sets)? These are among the most strongly and broadly implicated pathways, yet it remains unclear whether these reflect distinct disease subtypes or have an underlying biological connection. They exhibit subtle but distinct functional differences: chromatin complex genes exhibit earlier developmental expression patterns and synaptic genes peak more postnatally (Satterstrom et al. 2020). One proposed hypothesis is that chromatin modifiers somehow regulate the later expression of synaptic genes. However, when empirically tested, Satterstrom et al. (2020) failed to find significant overlap in terms of PPIs, co-expression networks, or known regulatory targets of chromatin complex (e.g., "GER") and synaptic (e.g., "NC") genes. Likewise, understanding how FMRP targets connect with other observed pathways and levels of convergence remains an open question (Clifton et al. 2020). Here, there are more potential direct connections, as FMRP is known to regulate the translation of long, brain-expressed genes, which include many known synaptic proteins, in an activity-dependent manner (Clifton et al. 2020).

Integration across Functional Hierarchies

More broadly, integrative efforts need to better connect distinct levels across the molecular hierarchy. While each omic measurement offers a unique snapshot of a complex system, molecular layers are often highly interdependent and reflect a common, latent underlying biological process. Although multi-omic integration has been an active area of research for some time, particularly in the cancer field (Huang et al. 2017; Ritchie et al. 2015), substantial methodological challenges remain in terms of analysis and visualization, which are further complicated by the scarcity of human brain tissue available for profiling. Among the unsupervised methods, similarity network fusion performs sample level clustering on each molecular feature independently and then fuses resulting similarity networks to identify subclusters with shared changes across multiple axes of molecular readouts (Wang et al. 2014). Ramaswami et al. (2020) leveraged this approach to integrate gene and microRNA expression (and co-expression), along with DNA methylation and histone acetylation changes across ASD and control brain samples and found convergent patterns of dysregulation across the transcriptome and epigenome. The predicted model based on these results suggested that ASD genetic risk factors were acting to downregulate neuronal gene expression, with DNA methylation and some histone acetylation changes acting in a secondary or compensatory fashion. As demonstrated, this integrative approach can be powerful and interpretable but requires relatively large numbers of samples with shared measurements.

From "Exploring and Exploiting Genetic Risk for Psychiatric Disorders," edited by Joshua A. Gordon and Elisabeth B. Binder. Strüngmann Forum Reports, vol. 31, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262547383 (paperback) 9780262377423 (pdf)

Supervised and semi-supervised methods leverage *a priori* knowledge about sample identities and biological relationships between features to build connections. A notable example comes from PsychENCODE, in which Wang et al. (2018) implemented a multilevel, generative deep learning model called a Deep Structured Phenotype Network (DSPN). This DSPN model linked common genetic variation (input) with psychiatric phenotypes (output), through a series of visible layers comprised of gene regulatory linkages, cell fractions, and co-expression modules, as well as a series of intermediate hidden layers. Interrogation of the latent nodes within this framework prioritized key contributory pathways—including synaptic activity, splicing, immune response and chromatin modification, among others—as well as the "best" positive paths connecting these nodes with both SNP genotypes and psychiatric traits. A major advantage of generative models centers on the ability to interrogate these hidden layers; however, far too often, integrative efforts to connect multiple molecular profiles have a "black box" feel, without the ability to visualize or directly interpret such connections. The utility of integrative models will ultimately rest on (a) whether the key underlying predicted biological connections can lead to new experimentally testable insights into disease mechanisms, and/or (b) whether these models can lead to biologically informative individual-level predictions in independent clinical populations.

Finally, it is important to note that recent large-scale integrative effects connecting functional genomic readouts with common variation (e.g., GTEx, PsychENCODE) to build maps of expression and splicing QTLs have largely focused on regulation within individual loci (e.g., *cis* effects) to maximize power. Moving forward, efforts to connect genetics directly with network-defined phenotypes (e.g., *trans*-regulatory effects) will almost certainly uncover relevant biology, particularly in the context of environmental stimulation (Kolberg et al. 2020).

Improved Specificity of Gene Annotations, Particularly at the Synapse

Given the strong convergence of neuropsychiatric risk genes at the "synapse," and the observation that synaptic dysfunction is strongly implicated in nearly all such disorders, we must gain more comprehensive and specific insights into the genetic underpinnings of synaptic architecture and diversity. The human brain is estimated to comprise hundreds of distinct synaptic types, with diverse structural and functional properties, including neurotransmitter specialization, neuromodulatory activity, and release probability (Südhof 2018). Yet, it remains unclear whether some of these are more predominantly impacted in specific disorders or by specific genetic risk profiles, which could provide important insights as to which specific circuits may be affected. For example, it is tantalizing to speculate that the recent identification of glutamate receptors *GRIN2A* and *GRIA3*—two of the top ten rare variant-associated

From "Exploring and Exploiting Genetic Risk for Psychiatric Disorders," edited by Joshua A. Gordon and Elisabeth B. Binder. Strüngmann Forum Reports, vol. 31, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262547383 (paperback) 9780262377423 (pdf)

SCZ risk genes (Singh et al. 2022)—provides credence to the decades-old NMDA-receptor hypofunction model (Olney et al. 1999). Yet, to address this question rigorously, we need a much more comprehensive understanding and catalog of the molecular machinery underlying diverse synaptic architectures. Historically, proteomic profiling, in combination with biochemical techniques such as subcellular fractionation and/or immunoprecipitation, has been a critical tool for high-throughput interrogation of molecular complexes at the synapse. To provide an organizing framework for these types of data, the SynGO consortium built an online knowledge base of expert-curated annotations for 1,112 synaptic genes with respect to protein locations and synaptic functions (Koopmans et al. 2019). While this is an important start, the resulting synaptic ontologies remain sparse, limited perhaps by the substantial efforts required for manual curation. Technological advances in high-throughput single-cell and spatial transcriptomic profiling coupled with viral tracing or barcoding strategies (Muñoz-Castañeda et al. 2021), as well as approaches that integrate molecular and physiological readouts within single cells (e.g., Patch-seq; Cadwell et al. 2016), have the potential to expand our understanding of such synaptic ontologies greatly. Further, the recent advent of proximity ligation-based approaches now enables proteomic profiling of synaptic structures with exquisite temporal and spatial resolution, including those difficult to isolate biochemically (Loh et al. 2016; Uezu et al. 2016). Altogether, integration of such high-throughput readouts across a wide array of cellular connections, developmental timepoints, and disease-relevant genetic perturbations is likely to greatly facilitate future systems-level dissection of psychiatric genetic risk mechanisms.

Synaptic genes are 2.6-fold longer with 1.7-fold more transcript-isoforms than nonsynaptic brain-expressed genes, indicative of a substantially expanded capacity and complexity of alternative splicing (Koopmans et al. 2019). Furthermore, isoform-level quantifications have been shown to capture greater differential expression effect sizes and GWAS enrichments than at the gene level in SCZ, particularly with neurons (Gandal et al. 2018b). As such, one approach that will very likely improve the specificity of our understanding of synaptic gene ontologies is to incorporate specific transcript-isoforms into these annotations, particularly as such data becomes increasingly available with long-read sequencing technologies. For example, the presynaptic cell adhesion molecular and high-confidence psychiatric risk gene *NRXN1* has now been shown to encode thousands of unique isoforms, many of which are regulated independently and thought to mediate diverse functions at the synapse (Treutlein et al. 2014). Furthermore, patient-specific *NRXN1* mutations have been shown to disrupt splicing, resulting in mutant isoforms that themselves are capable of disrupting neuronal activity in culture (Flaherty et al. 2019). Finally, Corominas et al. (2014) developed a novel protein interaction network by experimentally cataloging interactions between brain-expressed alternatively spliced isoforms of ASD risk genes, many of which localize to

From "Exploring and Exploiting Genetic Risk for Psychiatric Disorders," edited by Joshua A. Gordon and Elisabeth B. Binder. Strüngmann Forum Reports, vol. 31, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262547383 (paperback) 9780262377423 (pdf)

the synapse. The resulting "autism spliceform interaction network" showed that splicing altered ~50% of detected interactions and uncovered a substantial proportion of new PPIs. Altogether, these results highlight the potential added benefit of an isoform-centric approach for enhancing our understanding of synaptic architecture, particularly as it relates to psychiatric genetics.

Moving Beyond Enrichment

Although measures of population-level genetic convergence have gotten us this far, we must begin to now pivot toward interpretation—and *prediction*—of convergent risk within an individual. Personalized gene regulatory networks have been proposed as one approach to integrate distinct risk profiles within an individual and move toward precision medicine (van der Wijst et al. 2018). Polygenic risk scores will become more impactful as they incorporate multiple classes of genetic variation (SNPs, CNVs, and rare variants) along with improved identification of the underlying causal variant(s) across populations. As polygenic risk scores become more powerful, this will create new opportunities to incorporate network biology and pathway-level knowledge. For example, do affected individuals with synaptic risk gene profiles show distinct phenotypic trajectories or outcomes from those with risk variants affecting chromatin biology? Could targeted interventions be tailored to these distinct underlying risk profiles? Such predictions will require rigorous statistical validation in independent data sets to avoid overfitting.

Conclusions

The highly complex risk factors contributing to psychiatric disorders like ASD and SCZ, with polygenic contributions from variants across the genome, indicate that no simple parsimonious model will likely ever be able to fully explain mechanistic etiopathogenesis with any degree of generalizability. However, network-level embedding of complex genetic risk variation can begin to elucidate key underlying cellular and molecular pathways, providing a tractable framework for future experimental dissection, biological contextualization, and the potential for enhanced predictive modeling.

Acknowledgments

I would like to thank members of the Gandal Lab for helpful discussions and feedback for this chapter, especially Minsoo Kim, Leanna Hernandez, Pan Zhang, Michael Margolis, and Connor Jops.

From "Exploring and Exploiting Genetic Risk for Psychiatric Disorders," edited by Joshua A. Gordon and Elisabeth B. Binder. Strüngmann Forum Reports, vol. 31, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262547383 (paperback) 9780262377423 (pdf)