Understanding Common Variation

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)

8

Common Alleles

Next Steps in the Study of Common Variants

Hyejung Won, Naomi R. Wray, Elisabeth B. Binder, Kristen J. Brennand, Barbara Franke, Michael J. Gandal, Beth Stevens, Thomas Südhof, and Michael J. Ziller

Abstract

Common allele associations provide the starting point for delineating biological pathways that could progress understanding of psychiatric disorders. Multiple genomic approaches have been synergistically used to identify key contributing biological pathways that show evidence of convergence. Crucial next steps entail identification of causal risk variants, identification of causal risk genes, and identification of causal biological pathways. While a small number of genes could be prioritized and studied using the experimental paradigms applied to rare allele (large-effect) associations, this approach is neither feasible nor relevant for common variant associations where each person at high risk of disease carries a unique portfolio of risk variants, and where risk variants are carried by all of us. New experimental paradigms are needed that exploit the natural genetic variation present in populations and seek to understand why these unique portfolios of risk variants converge to disturb biological homeostasis, which lead to a common disease diagnosis. Recommendations for pathways forward are made, including new experimental paradigms that are specifically focused on combinations of risk-associated variants supported by new brain-specific data resources.

Introduction

The challenge of taking common risk allele associations into biological hypotheses and actionable outcomes applies to genome-wide association study (GWAS) results for all common diseases and disorders. Despite the complexity of the challenge, it is estimated that results from human genetic 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)

have contributed to 66% of drugs achieving FDA approval in 2021 (Ochoa et al. 2022). While the route from maps to mechanisms to medicines is never likely to be linear (International Common Disease Alliance 2020), this is particularly true for psychiatric disorders. Studies of brain disorders involve an organ that is much more difficult to access than for nonbrain-related disorders. Accordingly, in this chapter we focus on roadmaps and resources needed to clarify the role of common variation in psychiatry.

Translation of common allele associations requires a different approach from characterization of rare large-effect alleles for the following reasons:

1. Common single nucleotide polymorphisms (SNPs) have correlation structures that make it difficult to pinpoint causal variants based purely on association statistics.
2. Over 90% of common risk alleles significantly associated with psychiatric disorders are in the noncoding space (Watanabe et al. 2019a).
3. Combinations of variants together contribute to the risk of disorders.

In this chapter, we discuss existing knowledge gaps in the translation of GWAS discoveries and recommend experimental approaches, cellular models, and (genomic) resources which could help fill those gaps (see Figures 8.1 and 8.2). Specific recommendations are listed in the final section.

Pathway: Common Risk Allele to Causal Variant

The most significantly associated common risk allele in a genomic region identified in GWAS is not necessarily the functionally relevant variant. When many SNPs are in perfect linkage disequilibrium ($r^2 = 1$), it is not possible from statistical analysis to distinguish between them. Knowing, however, which SNP (or other classes of DNA variants) is functionally relevant is often the first step in understanding the mechanism of risk. The problem of fine mapping a GWAS locus to causal variants is a consideration for any GWAS and is not restricted to psychiatric disorders. Below, we discuss strategies to identify causal variants from the GWAS locus.

Increase GWAS Sample Size

In 2017, it was estimated that at least 80% of GWAS associations were within 33.5 kb of causal variants and that the mapping precision would increase both with larger GWAS sample sizes and larger imputation reference panels (Wu et al. 2017). As sample size increases, haplotypes with recombination events between tightly correlated alleles are increasingly likely to be present in the data set, which allows improved statistical fine mapping. While increased sample sizes will happen organically (see Ronald et al., this volume), current sample sizes are insufficient to infer causal variants purely from association statistics.

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)

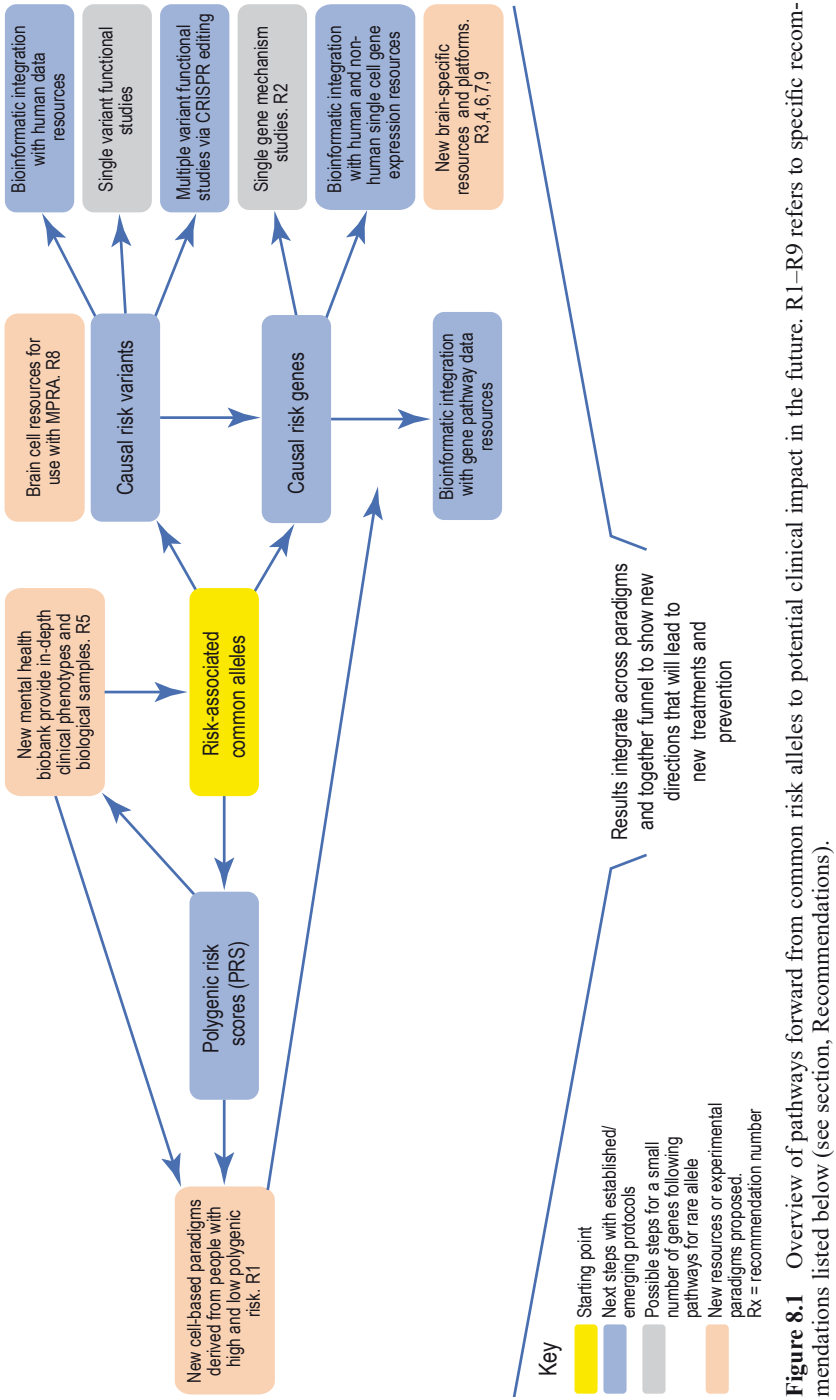

Figure 8.1 Overview of pathways forward from common risk alleles to potential clinical impact in the future. R1–R9 refers to specific recommendations listed below (see section, Recommendations).

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)

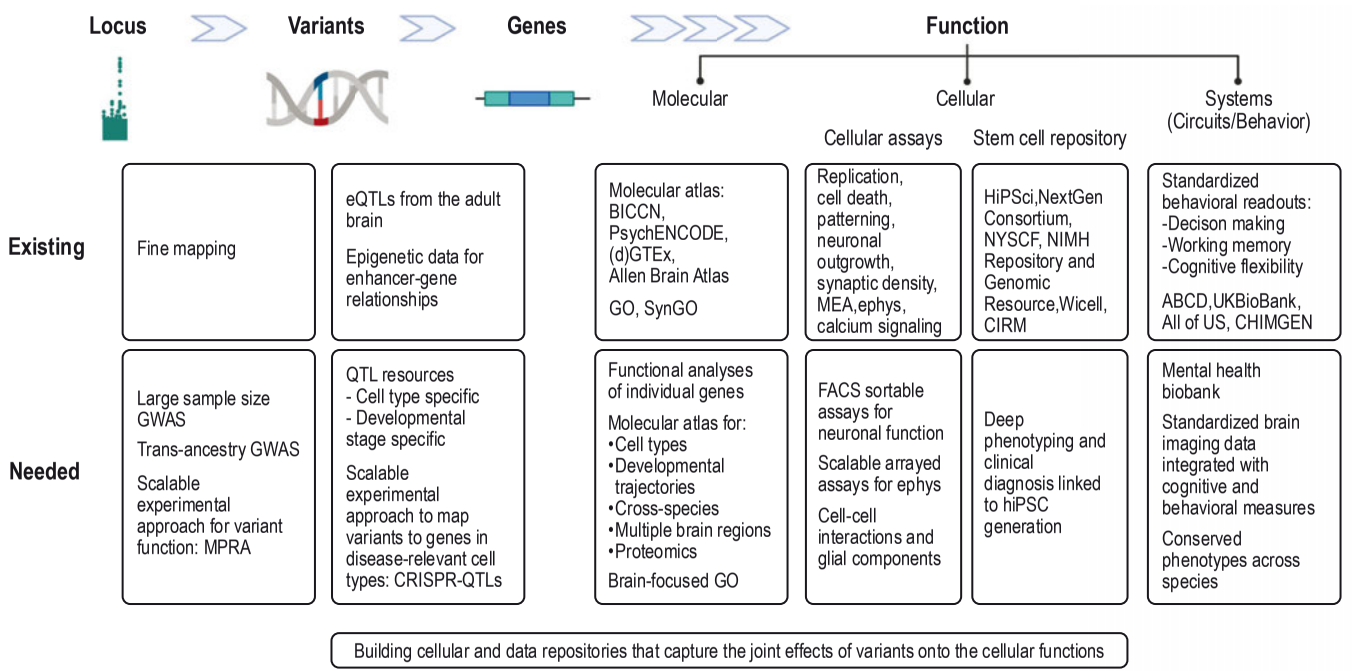

Figure 8.2 Existing and needed resources to distill biological hypotheses from common risk alleles. Electrophysiology (ephys), expression quantitative trait loci (eQTL), fluorescence-activated cell sorting (FACS), gene ontology (GO), genome-wide association studies (GWAS), massively parallel reporter assay (MPRA), multielectrode array (MEA), synaptic gene ontology (SynGO).

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)

Experimental approaches to identify functional variants out of schizophrenia GWAS loci suggest that SNPs with the strongest association signals are functional only in ~10% of loci (McAfee et al. 2022b).

Meta-Analyze GWAS across Ancestries

Common causal variants are likely shared across ancestries, but because ancestries have different population histories, linkage disequilibrium blocks and allele frequencies differ between them. Notably, linkage disequilibrium blocks are much smaller in individuals of African ancestry. Hence, the most associated SNP from cross-ancestry meta-analysis is more likely to be the causal risk variant. The need to increase GWAS sample sizes across ancestries has already been recognized (Martin et al. 2019b; Ronald et al. and Appelbaum, this volume).

Statistical Fine Mapping

A number of statistical fine-mapping tools have been developed which attempt to prioritize the underlying candidate causal variant(s) at a given GWAS locus with some posterior inclusion probability (Schaid et al. 2018). These are statistical predictions, however, and empirical evaluation of results from different fine-mapping algorithms give different sets of credible SNPs (Mah and Won 2020). Challenges can arise, in particular, when fine mapping is performed on aggregated summary statistics across multiple cohorts, when true causal variants are not directly measured and when algorithms assume only one causal variant is present at a given locus. Furthermore, the functionality of these predictions is largely untested. The recent schizophrenia GWAS systematically addressed some of these issues and provided fine-mapping results along with GWAS summary statistics (Trubetskoy et al. 2022). Still, the field needs improved means for prioritizing the most likely candidate causal variant at a given locus, before more costly follow-up experimentation steps are taken. While this issue may ultimately be resolved through trans-population meta-analyses, different fine-mapping approaches are needed in the near term and should be benchmarked against a standard set guided by experimentally validated SNPs.

Experimental Validation of Variant Function

Common SNPs associated with psychiatric disorders are enriched within active regulatory elements and are thought to play a role in gene regulation. Massively parallel reporter assays (MPRA) offer a scalable approach to experimentally test the gene regulatory activity of thousands of elements in a single assay (see Hu and Won as well as Brennand and Kushner, this volume). Such high-throughput screens for regulatory element activity have only recently been applied to measure allelic regulatory activity of variants (Deng et al. 2023; Matoba et al. 2020; McAfee et al. 2022b; Myint et al. 2020; Tewhey et al. 2016) and are

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)

increasingly being applied to evaluate putative causal expression quantitative trait loci (eQTL) that overlap with GWAS loci for complex disorders (Abell et al. 2022). It should be noted, however, that the majority of such experiments have been performed in cancer cell lines rather than brain cells, which makes it difficult to apply the findings to interpret GWAS of psychiatric disorders.

Given the enormous tissue- and cell-type specificity of active regulatory elements in which common SNPs reside, adapting MPRA for use in brain cells is critical if we are to understanding cell type-specific models of regulatory logic in psychiatric disorder risk variants (see section on Recommendations). The successful use of MPRA in human embryonic stem cell-derived neural precursor cells and neurons (Deng et al. 2023; Geller et al. 2019; McAfee et al. 2022b; Uebbing et al. 2021) is a promising step toward their wider application in identifying causal variants from psychiatric GWAS. Particularly exciting is the potential for development of massively parallel sequencing protocols in models based on patient-derived human-induced pluripotent stem cells (hiPSCs) that would enable cell type-specific identification of regulatory elements in polygenic background. Similarly, since hiPSC-derived neurons are already used to model human cortical development, applying these techniques in temporal analyses of hiPSC-derived cells may elucidate early developmental factors involved in the etiology of psychiatric disorders.

While MPRA offers a scalable approach to screen regulatory activity of noncoding variants at an unparalleled scale across diverse cell types of the brain, it does not provide information on the impact of the endogenous loci or identify the genes of action. This limitation can be mitigated by combining the MPRA strategies with other complementary approaches, such as CRISPR-mediated genome editing. The Impact of Genomic Variation on Function (IGVF) Consortium aims to validate experimentally the functional impact of SNPs via the combination of MPRA and CRISPR engineering in more physiologically relevant cell types (see section on Recommendations).

Given the lack of evolutionary conservation of noncoding elements (Han et al. 2018), hiPSCs provide a unique system to study the impact of noncoding variation on cellular features. CRISPR editing in hiPSC-derived cell types can resolve the molecular and cellular impacts of perturbing individual SNPs. Of the 108 schizophrenia GWAS loci identified initially by the Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014), 19 harbored colocalized GWAS and eQTL signals in postmortem brain RNA sequencing (Fromer et al. 2016). Of the five loci predicted to involve a single protein-coding gene, only one locus (*FURIN*) was the most significant GWAS SNP (rs4702) and also the most significant eQTL-SNP. Fine-mapping analysis strongly suggests this to be a single putative causal *cis*-eQTL (posterior probability = 0.94) (Schrode et al. 2019). CRISPR-based allelic conversion resulted in decreased *FURIN* expression in hiPSC-derived glutamatergic neurons accompanied with reduced neurite length and altered population-wide neuronal activity patterns (Schrode et al. 2019). This example

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)

illustrates the power of hiPSC-derived cell types in studying molecular and cellular impacts of individual noncoding variants, although the mechanistic link to disease remains unclear.

Pathway: Causal Risk Variant to Gene

Mapping a causal risk variant in the noncoding space of the genome to the gene it modulates remains a major challenge in the field. This difficulty is rooted in the large number of parameters that determine a variant's effect and the distinct mechanisms of action an individual variant can have. Noncoding genetic variants can disrupt the function of noncoding gene regulatory elements that modulate distinct genes to different extents (Fulco et al. 2019). Roughly 80% of noncoding variants also regulate "non-nearest" genes (Sey et al. 2020), suggesting that variant-gene relationships cannot be simply captured in the context of the linear genome. Moreover, the effects of noncoding variants on genes have been shown to be mediated through multiple distinct mechanisms of action, such as modulating gene expression levels, splicing patterns, chromatin accessibility, transcription factor and microRNA binding as well as alternative polyadenylation (GTEx Consortium 2020; Huan et al. 2015; Li et al. 2021b; Wang et al. 2018). Adding another layer of complexity, noncoding variants frequently operate in a highly cell-type-, developmental stage, and condition-specific manner. Thus, tailored high-throughput assays and context-specific cellular model systems are required to systematically associate individual variants with their downstream targets. The resulting data sets require advanced computational approaches to facilitate discovery and integration of respective results. Below, we summarize existing and urgently needed resources, methods, and assays to start tackling these questions, relevant to both common and rare associated alleles (see also Bearden et al., this volume).

Bioinformatic Integration with QTL Resources

Systematic integration of GWAS with experimentally defined quantitative trait loci (QTL), for molecular features profiled within the human brain, provides an important way to prioritize the candidate causal gene at a given GWAS locus. This can be achieved without identifying the causal risk variant. Large-scale efforts have been undertaken to identify gene expression and splicing QTLs in the adult human brain. As sample size and tissue are the most critical factors for QTL discovery, this has largely been done in bulk adult cortex samples through mega- and meta-analysis (de Klein et al. 2023; Wang et al. 2018; Zeng et al. 2022). Despite these concerted efforts, brain eQTLs alone cannot link all GWAS loci to targetable genes (Zeng et al. 2022). Hu and Won (this volume) and Robinson et al. (this volume) describe currently available QTL resources in both adult and developing human brains as well as critical next steps, such

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)

as developing cell type-specific QTL resources and expanding splicing QTL and isoform-level resources. We recommend that more brain-specific gene expression resources be generated (see section on Recommendations).

Bioinformatic Integration with Other Genomic Resources

Compared to current eQTL resources, which lack cellular and developmental resolution, epigenomic profiling provides an alternative resource to address the cell type and nature of variant-gene relationships specific to development. Multi-omic data sets are particularly useful to link variants to genes. Existing multi-omic resources and approaches to link enhancers (and variants) to genes are described in detail by Hu and Won (this volume). One key benefit of using epigenomic assays to infer target genes of variants is that enhancer-gene relationships are robust to linkage disequilibrium (Whalen and Pollard 2019). Furthermore, a growing number of epigenetic data sets are being generated at cellular resolution through the advancement of single-cell genomics, which allows cell type-specific target gene identification. To facilitate growth in this area, brain-specific omics resources and ear-marked funding for data integration would be beneficial (see section on Recommendations).

Experimental Validation of Variant-Gene Relationship

Inferences regarding variant-gene relationships from transcriptomic or multi-omic approaches need to be functionally validated. One approach to confirm the effects of variants in regulatory elements involves CRISPR screening. Coupling CRISPR editing with single-cell RNA sequencing readouts can validate the regulatory impact of GWAS SNPs on predicted target genes at cellular resolution, yielding rich and high-dimensional phenotypes (Gasperini et al. 2019). To date, however, many CRISPR screens have been conducted in cancerous cell lines. Similar assays will need to be conducted in brain cell types. In particular, CRISPR screens that target noncoding variants in brain cell types across multiple differentiation time points will allow us to systematically interrogate the variant-gene connection in a cell type and developmental context-specific manner. Current CRISPR screens are designed to measure gene expression changes upon perturbing a regulatory element rather than a single nucleotide change. Therefore, CRISPR screens alone may not provide the directionality of effects of the variants. As discussed earlier, CRISPR screens can be complemented by other experimental validation assays: MPRA provides causal variants and directionality of effects, whereas CRISPR provides target genes of variants.

Beyond Gene Expression Readouts

The functional impact of noncoding variants could involve multiple mechanisms (e.g., splicing regulation, UTR function, RNA localization), yet most

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)

experimental assays focus on enhancer function. There has been a considerable progress in the development of high-throughput functional genomic strategies to measure the impact of individual genetic variants on splicing (Cheung et al. 2019; Rosenberg et al. 2015), mRNA degradation (Rabani et al. 2017), UTR function (Griesemer et al. 2021), and subcellular RNA localization (Mikl et al. 2022). Adapting such high-throughput assays for distinct molecular layers will enable comprehensive understanding of the functional impact of variants that exert their effects through different molecular mechanisms. In contrast, advances in the identification of *trans*-mediated effects of individual genetic variants have been much more limited.

In addition, the dependence of the molecular consequence of a single genetic variant on the genetic background of a specific individual remains unclear. To study the variant effects in the context of a polygenic background, new experimental and computational strategies are required to better identify *trans*-effects of individual genetic variants beyond simply acquiring larger cohorts of individuals profiled for gene expression. Moreover, systematic efforts are needed to evaluate the impact of genetic background on single variant-gene association by performing the assays discussed earlier in cellular contexts from distinct genetic backgrounds and ideally across species (see section on Recommendations).

How to Prioritize Genes for Validation

Once we identify target genes implicated by the common risk variants, we must understand their functional role in the brain. This step can follow the same experimental paradigms used to investigate the functional role of genes identified from rare-variant, large-effect associations (see Bearden et al., this volume). Gandal (this volume) provides an example of how the common allele association at the major histocompatibility complex locus led to a detailed analysis of the function of the *C4A* gene. Despite this success, the majority of psychiatric GWAS loci have not yielded testable biological hypotheses. Unlike rare variation that disrupts protein function, the exact mechanisms through which common SNPs act on the genes remain unknown. Moreover, GWAS results implicate hundreds of genes that may act together in a combinatorial fashion within the polygenic background. Therefore, not all genes implicated by GWAS may reveal clear mechanistic insights into disease when investigated individually (see section on Recommendations).

Pathway: Gene to Function and Phenotype

Once variants are mapped to genes, the next steps are to identify disease-relevant biochemical pathways with which to generate and test mechanistic biological hypotheses. This is where the strategies to study rare-variant associations

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)

and common variant associations clearly diverge. For most common risk alleles, it is simply not feasible to generate biological hypotheses from individual variants, since the role of risk variants likely depends on the genetic context in which they are present. New strategies are needed to identify the biological convergence associated with the heterogeneous set for risk variants (unique to each individual) which are associated with high risk of disease. To achieve this, we need new experimental paradigms and new omics reference data sets that can be integrated with GWAS results, supplemented with a platform that facilitates multimodal data integration (see section on Recommendations). This, in turn, requires a better understanding of the cell types and circuits as well as the relevant developmental periods. Some of this information may emerge from ongoing (single-cell) transcriptomic studies of human and animal model brain tissue. However, diverse cellular and animal-based systems are needed to study gene function and candidate disease mechanisms (Figures 8.1 and 8.2).

Molecular Function and Biological Pathways

The functional impact of genes targeted by common SNPs can be interrogated within biological contexts by bioinformatic integration of disease-associated gene lists with annotated gene sets (e.g., gene ontology, GO). Most available gene sets, however, are not curated with respect to brain function, which makes it challenging to decipher brain-relevant disease biology from psychiatric variants. Whereas some data sets already exist through initiatives such as the BRAIN Initiative Cell Census Network (BICCN) (2021), the GTEx Consortium (2020), the Allen Brain Atlas, and the PsychEncode Consortium et al. (2015), integrating multimodal data sets acquired from different initiatives can be challenging. Concerted efforts aimed at systematic annotation of genes and pathways can be powerful, as illustrated by SynGO, an approach to integrating data in synapse biology (Koopmans et al. 2019). Its success argues strongly for the establishment of a more comprehensive brain ontology consortium which would extend the SynGO approach to a central "brain-centric GO" repository (see section on Recommendations). This would host not only manually curated brain-relevant biological pathways but also the growing list of gene sets derived from perturbation experiments (e.g., CRISPR screens, drug treatment response).

Cellular Function

Given the lack of evolutionary conservation of noncoding elements (Han et al. 2018), animal models may not provide the necessary context to explore the variant-gene-function continuum. However, hiPSCs provide a unique system to study the impact of noncoding variation on cellular features. As described above, CRISPR editing can be used to investigate the function of individual risk variants. To understand the impact of unique portfolios of thousands of

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)

risk variants, multiple biologically functional variants need to be simultaneously introduced, and it is not straightforward to scale via CRISPR editing. One potential alternative mode to measure the joint effects of variants on cellular phenotypes would be to create a large collection of hiPSCs representing a range of genetic backgrounds and disease conditions. There are several ongoing initiatives relevant to fast-track this approach. The Human Induced Pluripotent Stem Cells Initiative (HipSci) includes more than 700 hiPSC lines from ~300 individuals, the vast majority of which are from control donors of British ancestry (Kilpinen et al. 2017). The NextGen Consortium (a US-based team) generated a large collection of hiPSCs in parallel; although not focused on neuronal cell types, their focus on polygenic metabolic disorders may provide insights relevant to neuropsychiatric disease (Warren et al. 2017). The New York Stem Cell Foundation (NYSCF) has likewise assembled a large trans-ancestry collection of hiPSCs with a focus on Parkinson disease. Individual researchers are contributing their hiPSC collections to a variety of repositories, such as the NIMH Repository and Genomics Resource, and WiCell, yielding a bank of patient-specific hiPSCs generated across a variety of laboratories via an assortment of methodologies. The California Institute of Regenerative Medicine (CIRM) hiPSC collection is one of the largest US-based single-derived collections of genotyped hiPSCs (1,618 donors), generated by a standardized, non-integrating episomal reprogramming approach in a single production facility. Scalable approaches to culture and differentiate hiPSC lines from different donors are emerging for systematic investigation of variant-cellular function relationship. Village-in-a-dish approaches mix hiPSCs from dozens of donors together for transcriptomic (single-cell RNA sequencing) and/or phenotypic (FACS-based) assays in pools (Jerber et al. 2021; Mitchell et al. 2020; Neavin et al. 2021b), making it possible to test the genotype-dependent effects at scale.

Since the pioneering study by Dobrindt et al. (2021; see also Brennand and Kushner, this volume), which utilized a high versus low polygenic risk score (PRS) design, another study has extended the paradigm to compare 13 high PRS neuronal cell lines derived from people diagnosed with schizophrenia against 15 neurotypical individuals with low PRS (Page et al. 2022). The latter study identified altered Na+ channel function, action potential interspike interval, and GABAergic neurotransmission as being associated with high PRS. While these early results need further validation, the ability to identify basic neuronal physiological properties that can be related to core clinical characteristics of illness may be a critical step in understanding mechanism of polygenic disease and generating leads for novel therapeutics. Widespread adoption of a common set of hiPSC lines with extreme schizophrenia PRS will not only help evaluate the reproducibility of functional genomic studies, it will also facilitate the integration of data sets generated by different laboratories, revealing any convergent impacts of independent schizophrenia-associated variants. This leads to our recommendations for a high/low PRS cell-line study conducted at scale and for the establishment of a mental health biobank to generate the

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)

cell-line resource linked to in-depth clinical, longitudinal phenotypes (see section on Recommendations).

Systems Level: Circuits and Behavior

Brain Circuits

Having the ability to link molecular and cellular phenotypes to brain structure, function, and connectivity as well as cognitive and behavioral phenotypes would enable the translation of polygenic effects and enhance our understanding of psychiatric disease. Early studies of circuit-relevant phenotypes were plagued by small sample size, but now the field of genetic imaging has seen immense progress through the formation of large-scale consortia, such as ENIGMA (Thompson et al. 2022) and CHARGE (Psaty et al. 2009), with an emphasis on the development of open-source protocols and resources to improve standardization of data analysis workflows in the field of magnetic resonance imaging (MRI) and electroencephalogram (EEG). Similarly, efforts such as the UK Biobank and the ABCD Study have provided large data sources with standardized neuroimaging data. The brain measures used in these studies are moderately to highly heritable, and while obtained largely from individuals without disease, they show associations with psychiatric disease phenotypes, although the genetic overlap between them is small (Grasby et al. 2020). The need for large sample sizes has so far limited the range of phenotypes assessed to relatively crude measures of global and regional morphology, brain activity, and structural and functional connectivity. While the costs of generating omics data are expected to decline, the cost of generating imaging data is likely to remain high. Thus, due to cost considerations, we are unable to recommend the purpose-driven generation of MRI data sets of the scale required (Marek et al. 2022) for gene discovery. If generated in the context of other large-scale efforts, however, they may be useful for integrative analyses of common variant associations.

Cognitive and Behavioral Phenotyping

Large-scale data collections of cognitive and behavioral phenotypes are becoming available through population-wide initiatives (e.g., the UK Biobank, the ABCD Study, All of US Research Program), but the depth of psychiatry-relevant phenotypes is, to date, insufficient. It is well-recognized that large-scale population, volunteer-based initiatives underrepresent people with psychiatric disorders. Notably, most cohorts provide limited longitudinal measures even though the importance of longitudinal trajectories of these measures is increasingly recognized (Shah et al. 2020). The human connectome project for early psychosis is a good example of a cohort (N = 200 with early psychosis and

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)

N = 100 controls) with in-depth multimodal phenotyping and potential for a hiPSC resource (Demro et al. 2021).

Clearly, the field needs initiatives directed at defining the most disease-relevant imaging and behavioral traits. While progress is being made (e.g., the Psychiatric Genomics Consortium conducts analyses of secondary phenotypes including cognitive measures), sample sizes are much smaller than for case/control analyses. New data collection is urgently needed, including in-depth, longitudinal phenotyping across multiple dimensions and biobanking of samples, not only for DNA and other omics analyses, but processed to allow future generation of hiPSCs. We recommend that a mental health biobank be initiated to provide longitudinal information on individuals with psychiatric disorders (see also section on Recommendations).

Phenotypes that Transcend Experimental Paradigms

Since cognitive, behavioral, and imaging phenotypes (unlike psychiatric disorders) can be measured in animal models, efforts to increase translatability of phenotypes across species are important. At the brain imaging level, development of whole brain function ultrasound in awake-behaving animals may permit generation of phenotypes that parallel human fMRI (Brunner et al. 2020; Brunner et al. 2021), including brain network activity changes at rest or under specific tasks that emerge as highly relevant for psychiatric disorders. Such translatability across species could also be achieved with EEG-based measures, with increased availability of wireless EEG in rodents, including sleep EEG (Karamihalev et al. 2019), where headband-based ambulatory EEG methods are now increasingly available for humans. At the level of behavioral and physiological assessments, current technical developments in human and animal phenotyping (e.g., touch screen, actimetry, startle responses) allow better cross-species alignment.

With increased research focus on cell-based models, phenotypes that can be measured *in vivo* an *in vitro* are of particular interest. For instance, measures captured at the electrophysiological level in hiPSC-derived neuronal models (Page et al. 2022) can be related to the *in vivo* context in animals using next-generation electrodes (e.g., Neuropixels probes) that are increasingly translatable to EEG or even MEG in humans (e.g., Schulte et al. 2021). Similarly, circadian rhythm phenotypes measured in cell lines have been associated with circadian rhythms of bipolar disorder patients from whom the cell lines were derived (Sanghani et al. 2021). Likewise, synaptic density, a feature associated with psychiatric disorders in postmortem brains and *in vitro*, can now be measured *in vivo* through positron emission tomography imaging of synaptic density using synaptic vesicle glycoprotein 2A (SV2A) in animal models and humans (Cai et al. 2019; Toyonaga et al. 2022). We recommend specific focus on phenotypes that can be meaningfully measured across multiple experimental paradigms (see section on Recommendations).

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)

Pathway: Context

Genes and genetic variants may have different functions under different biological and environmental contexts. For example, variant and gene function may vary in different cell and tissue types, across development, aging, and in response to stressors. Having a map of context-dependent annotations will be invaluable in guiding our understanding of variant to function connections. Context changes could unmask previously unrecognized variant to gene, gene to function, or polygene to function connections that are not apparent in a baseline state. Context needs to be a key consideration in all experimental approaches, and challenges represent any stimuli acting as stressors that would deviate cells/organisms from maintaining homeostasis. For instance, MPRA screens (discussed above) used for functional validation of risk variants should be conducted in different cell types as well as after exposing cell lines to stimuli. Supporting this claim, MPRA in the context of activating the glucocorticoid receptor (a key mediator of the stress response) unveiled a novel function of genetic variants associated with psychiatric disorders specific to stress hormone exposure (Penner-Goeke et al. 2022). In designing perturbation experiments, "stressors" can be summarized into major categories of metabolic stress, oxidative stress, action of hormones (e.g., glucocorticoids), and inflammatory stress. Other stimuli particularly relevant to neuronal function could be neuronal activation as well as the action of certain neurotransmitters. Incorporation of gene–environment interactions may be especially critical in studies of neuropsychiatric disorders, which may require or include specific stressors (e.g., posttraumatic stress disorder) or exposures (e.g., substance use disorder) among diagnostic criteria, or may count exposure to illicit substances or extreme stress as among the most critical risk factors for disease development (e.g., schizophrenia). Toward deciphering complex gene–environment interactions, there is an urgent need to explore the additive impact of environmental stressors on the effects of risk variants and genes within individuals, and extending PRS approaches to include the gene–environment interactions at the population level. Importantly, sex-specific effects and sex-environment/perturbation interactions are likely to be needed. While increasing the complexity of the proposed experiments, they are essential for ensuring generalizability of results (see section on Recommendations).

A systematic interrogation of gene–environment interactions could be achieved by initiatives such as Connectivity Map (CMAP). This platform provides a comprehensive catalog of molecular and cellular signatures elicited across multiple cell lines by systematic perturbation of cell lines by pharmacological perturbations. CMAP serves as a database resource to investigate the modes of action in a wide range of drug perturbations. Researchers can purchase the cell lines used in CMAP to conduct perturbation experiments and can benchmark gene expression perturbations induced by new compounds against

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 CMAP perturbation experiments. CMAP, however, is not well curated to address gene–environment interactions in the context of psychiatric disorders. It only includes a small portfolio of drugs relevant to psychiatric disorders and does not cover brain cell lines.

Recommendations

The biological robustness implied by the polygenic architecture of psychiatric disorders provides a major challenge in translating genetic association results into meaningful outcomes for those whose lives are, or will be in the future, affected by psychiatric disorders. Common risk allele associations provide multiple new directions for research. In describing the pathways forward (Figure 8.1) we have uncovered the need for data resources and experimental paradigms, many of which must be undertaken at scale as well as collaboratively across groups and even nations (Table 8.1). Guiding principles for building such resources include transparency, accessibility (e.g., open source), strong emphasis on quality control, and community building. In this section, we provide nine recommendations (R1–R9) for future research priorities. Given our focus on pathways forward for common alleles, we have ordered the recommendations around this goal. We believe the results from these proposed studies, regardless of actual outcomes, will propel the field forward and expose

Table 8.1 List of existing and needed resources.

	Existing Resources	Resources Required
Molecular atlas	• Brain Initiative Cell Census Network (BICCN) • PsychENCODE consortium • Genotype-Tissue Expression (GTEx) • Developmental GTEx • Allen Brain Atlas	• Cross-species • Multiple brain regions beyond cortex • Inclusion of data beyond expression and proteomics
Gene ontology	• GO • SynGO	• Brain ontology consortium
Stem cell resources	• HipSci • NextGen consortium • NYSCF • NIMH Repository and Genomics Resource • WiCell • CIRM	• Deep phenotyping • Clinical diagnosis linked to the cell lines
Longitudinal behavioral phenotyping	• ABCD Study • UK Biobank • All of US Research Program	• Mental health biobank

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)

new avenues of research that will ultimately drive progress to the key goals of improved prevention, diagnosis, and treatments.

R1: A New Experimental Paradigm That Can Capture Polygenicity

We need an experimental paradigm that goes beyond the analysis of individual genetic variants and assesses instead the aggregated impact of naturally occurring combinations of common risk variants that jointly act, integrating over *cis*- and *trans*-effects, and results in high (or low) risk of disease (Figure 8.3). hiPSC-based model systems provide the unique opportunity to identify consistently dysregulated genes and pathways associated with diagnostic groups. To maximize power for a given budget, we propose a high/low polygenic risk experimental design that has already been piloted on a small scale (Dobrindt et al. 2021; Page et al. 2022). Although the scale of study we propose stretches current technical limits, scalability in hiPSC-based model systems is a fast-moving field and will be driven further by need. The power of the design increases with the strength of selection on polygenic risk. Currently, there is a 39-fold difference (95% confidence 29–53) in risk of schizophrenia between the top and bottom centile of polygenic risk, but only a 16-fold difference (95% confidence 15–17) between the top and bottom decile (Trubetskoy et al. 2022). To achieve an extreme PRS design with sufficiently large N, a very large cohort of participants needs to be collected with genotype data to measure PRS. Unfortunately, recontact with participants, necessary to establish hiPSC models, is likely to be more difficult in psychiatry compared to other branches of medicine; hence, biobanking of blood processed to enable generation of cell lines must be done early in study participation. This links to the need for a mental health biobank, discussed below (see R5). The experimental approach we propose will pave the way for precision medicine in psychiatry, similar to approaches already pioneered in cancer, heart, or kidney disease: optimal treatments personalized to an individual will be determined through responses to perturbations applied to person-specific cell lines or organoids. The model is described in Figure 8.3. Table 8.2 provides a justification for the underlying experimental design, and Table 8.3 presents a SWOT analysis. Although the design is set up for schizophrenia (SCZ) because the genetic discovery to date is highest (PRS explain 10% of variation in liability), the design could be implemented for any disease.

- *Vision*: To establish a novel and well-powered experimental paradigm that uses cellular measures to identify biological signatures that correlate with diagnosis. A cellular platform is a simplification of real biological processes because different sets of risk loci likely to have biological impacts in multiple cell types and at multi-developmental time points in response to biological environments. We believe this platform will provide a tractable model in which different sets of risk

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)

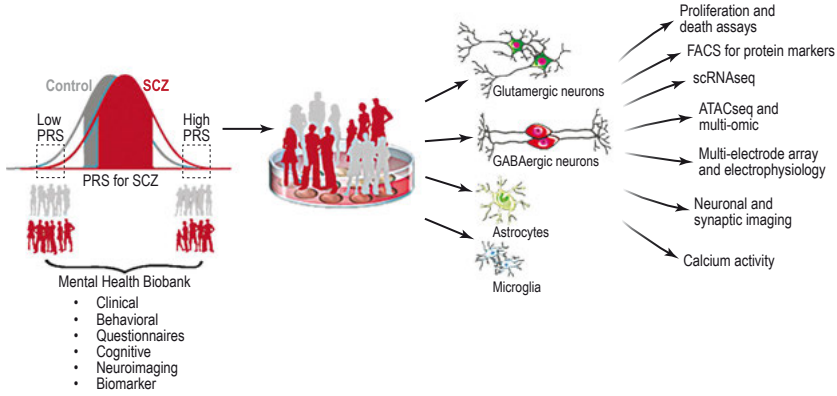

Figure 8.3 Experimental model system to identify biological correlates of genetic risk and diagnosis.

variants generate convergent biological readouts. This paradigm would allow us not just to better understand relationships between risk variants, function, and phenotype, and to causally query these hypotheses across disease relevant contexts.

- *Major goals:* To translate polygenicity into biological and cellular pathophysiology; to establish protocols for repeatable cellular phenotypes that can inform disease mechanisms; to provide protocols for repeatable perturbation studies applied to at scale that expose convergent biological responses; and to provide a framework for precision medicine in psychiatry.
- *Design:* Generate hiPSC lines using a 2×2 design of high and low PRS with/without SCZ diagnosis, justified in Table 8.2. The extreme PRS design is cost-effective and the 2×2 design generates combinations of comparisons that together are more informative than simpler designs.
- *Which cell type*: Variant function can diverge in different cell types, so identification of the disease-relevant cell type will be essential for the success of the proposed experimental system. Analyses of GWAS and emerging single-cell transcriptomic data nominate excitatory neurons for prioritization (Sey et al. 2020; Skene et al. 2018). Hence, initial focus could be on excitatory neurons and later expanded to other cell types, as well as more complex organoid models.
- *Which model*: The hiPSC village-in-a-dish model can increase scalability, minimize variability, and might be more affordably implemented across laboratories. This model is well suited to sequencing measures as individual cells can be identified through genomic data. For other

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)

Table 8.2 Justification of the 2 × 2 design: high vs. low polygenic risk score (PRS), with or without a schizophrenia (SCZ) diagnosis.

PRS with SCZ	PRS no SCZ	Justification of comparison
High	Low	Since all people carry risk alleles for SCZ, differences at the biological level between cases and controls are expected to be subtle. This design utilizes PRS data to accentuate the biological differences between case and control groups. Based on current PRS, there is a 39-fold difference in risk based on top vs. bottom centiles. In this comparison we expect to observe or construct perturbations that generate strong differences in cellular measures between the two groups. The observed differences may reflect general case/control status or something more specific to polygenicity. Results from the pairwise comparisons will help differentiate these scenarios.
High	High	Here, all individuals have high PRS for SCZ, so differences in cellular measures will provide information about functional pathways associated with SCZ over and above those associated with the identified polygenic risk. A parsimonious hypothesis is that nonidentified genetic risk variants (i.e., those with high PRS and SCZ compared to those with high PRS but without SCZ) will impact the same biological pathways in the same way as currently identified risk variants. Empirical evidence is needed to support or reject this hypothesis.
Low	Low	This comparison is underpinned by parsimonious hypotheses: Those with low PRS and SCZ must be enriched for genetic risk variants not yet identified (see exclusions). Nonidentified genetic risk variants will impact the same biological pathways as currently identified risk variants. Cellular measures that differ between low PRS with/without SCZ should be the same as cellular measures that differ between high PRS with/without SCZ. Empirical evidence is needed to support or reject these hypotheses.
Low	High	This comparison is underpinned by parsimonious hypotheses: Those with low PRS and SCZ must be enriched for genetic risk variants not yet identified, which have a stronger biological signal than those with high PRS but no disease diagnosis. Cellular measures that differ between low PRS with SCZ and high PRS without SCZ should be the same as cellular measures that differ in at least one of the other pairwise combinations. Empirical evidence is needed to support or reject these hypotheses.
Exclusions		In all comparisons, those carrying large-effect rare-variant alleles will be excluded. Such alleles are often associated with syndromic phenotypes and/or are nonspecific to SCZ. We propose that they should be studied in add-on cell-line comparisons and excluded from the baseline 2 × 2 design.

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)

Table 8.3 SWOT (strengths/weaknesses/opportunities/threats) for the 2 × 2 cell-based experimental paradigm designed to understand the molecular mechanisms associated with common allele polygenic risk.

Strengths	Weaknesses
• Captures polygenic background in the absence of confounds • The 2 × 2 design provides multiple comparisons which together will generate more robust conclusions • Regardless of results, new biological insights will be obtained • Village-in-a-dish model improves scalability of the approach • Systematic platform to test interplays between genes and environmental challenges	• Scaling to sample sizes desired will be challenging • Relevant readouts are not yet available • Readouts can be noisy and difficult to interpret • Lack of tools for analyzing and integrating different phenotypic modalities • Will be challenging to set up to be representative of all ancestries; other ancestry-specific studies will be needed to inform this
Opportunities	**Threats**
• Multimodal integration • Provides possibility for development of diagnostic biomarkers in psychiatry • Provides possibility of linking cellular phenotypes to clinical phenotypes (e.g., electrophysiology translating to EEG) • Motivates collection of a mental health biobank future-proofed to benefit from outcomes • New methods will be developed which will be valuable in many research domains • Globally shared platform • Fosters collaboration	• Sample sizes may be too small given actual effects sizes and multiple comparisons • Identification of individuals with extreme PRS with/without disease requires large cohorts genotyped and with biological samples for hiPSC • Ethical concerns associated with hiPSC research

technologies, the participant identity is integral if cell lines are held individually.

- *What to measure*: Phenotypic assays need to be carefully selected so that they are orthogonal and allow high-throughput characterizations:
  1. Genomics: single-cell RNA sequencing and single-cell sequencing assay for transposase-accessible chromatin
  2. Morphology: High-content imaging for neurite outgrowth and synapse density
  3. Electrophysiological property: multielectrode array, optical electrophysiology

Measures need to be repeatable (which could be achieved through testing of protocols across multiple labs) and heritable (i.e., phenotypic variation is associated with genetic variation), which could be verified in parent/offspring or sibling designs.

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)

- *What perturbations to apply*: In the first instance, we propose a small number of perturbations which could be informed by CMAP type experiments (see below, R4).
- *Sample size:* Pilot studies demonstrate that significant differences can be identified from sample sizes of less than 20 in each group (Dobrindt et al. 2021; Page et al. 2022). The power of the study also depends on the strength of selection that can be applied to the PRS. Our expectation is that the platform will be used to test multiple hypotheses. Moreover, our knowledge of common disease establishes a baseline expectation of complexity. Hence, suggestions for the number of samples required for each comparison are unknown. Since we expect the strongest differences in cellular measures from the high PRS with SCZ/low PRS without SCZ, this comparison could be tested in a pilot phase that develops protocols to determine which cell types to generate, and which cellular measures and which perturbations generate reproduceable and repeatable results. Ultimately, we anticipate that a design with N = 500 per group balanced by sex, is the minimum sample size required.

R2: New Scalable Platforms for Functional Analyses

Complementing R1, this recommendation is based on the need to understand the functional implications of common genetic variants associated with neuropsychiatric disorders in a well-controlled and reproducible approach. We acknowledge that the functions of most protein-coding genes as well as of most noncoding genomic loci are incompletely understood. Even for genes with apparently clear-cut functions and involvement in psychiatric disorders (Dai et al. 2019; Singh et al. 2022), such as those encoding subunits and regulators of NMDA-type glutamate receptors (NMDARs), the field does not yet understand the synaptic and extra-synaptic functions of these genes. We do not know, for example, what the various types of NMDAR subunits do, how they are regulated, and how the interplay of different types of glutamate receptors (AMPA, NMDA, and kainite receptors) informs neural circuits. Functional analyses with a range of relevant readouts are needed for studying the functional effects of individual common genetic variants and combinations of common variants in multiple genetic backgrounds. This can be achieved by having shared cell lines each of known genetic background, to allow comparisons across laboratories.

To meet this challenge, new scalable platforms for functional analyses are necessary that are currently not available. The challenge is made difficult because several key processes in brain, such as various synaptic functions (receptor composition, release probability, short- and long-term plasticity), neuronal excitability, myelination, and microglial immune responses, are not (yet) amenable to high-throughput assays. For example, current scalable approaches, such as multielectrode arrays or calcium imaging, provide an excellent readout

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)

of neuronal activity but do not provide insight into either any synaptic function or neuronal excitability. In view of these needs and limitations, it is recommended to invest in tools that achieve scalable analyses either by developing completely new high-throughput assays or enabling a medium-scale analysis of various brain processes. Such tools, for example, could (for synaptic functions) consist of automated measurements of miniature synaptic responses using robotic patching or optical measurements of specifically pre- and postsynaptic calcium transients using dual color calcium sensors. The proposal complements the Brain Initiative's focus on scalable cellular/molecular approaches (BRAIN Initiative Cell Census Network 2021).

R3: Facilitate Data Integration Research

Many brain-omics data sets have been, and will be, generated through a wide range of research projects funded internationally. Progress in understanding biological mechanisms associated with polygenic disease can be made through intelligent bioinformatic integration of data sets. The underlying omics data sets can be expensive to generate, so the integrative analyses will be integral to maximize the utility of the data at relatively low (but not zero) additional cost. To facilitate research integrating data sets, a platform should be established that brings together multimodal data sets. This will allow researchers to query and contribute to data sets. An example platform for hosting multimodal data sets is the Alzheimer Disease Forum (AlzForum). An example platform for comparing machine learning algorithms or statistical models to standardized data sets is Kipoi. Similar platforms need to be established to underpin research in psychiatric disorders.

R4: Establish a Brain CMAP

The CMAP study has proven to be a useful resource for investigating mode of action of drugs, with 4,435 Google cites to the primary paper (Lamb et al. 2006). We recommend that this study be extended to generate gene expression changes that result from a catalog of psychiatry-relevant cell types, cellular assays, and pharmacological agents, as we believe this could reveal new layers of convergence which, in turn, could inform biology of psychiatric disease. The resulting data base and resource would be relevant to many research studies of the brain. Brain CMAP perturbation could inform which types of stimuli would most likely unmask relevant functional differences in brain cells with different polygenic risk background.

R5: Establish a Mental Health Biobank

Our high/low PRS design is dependent on relatively large samples with and without recorded diagnosis (e.g., SCZ) together with genotype data to generate

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)

PRS and biological samples to allow generation of cell lines. The impact of the design is maximized if cellular phenotypes can be correlated with in-person phenotypes, because it prepares for a future cell-based precision medicine platform. To support this design, we recommend that a mental health biobank be established to provide longitudinal clinical phenotypes and matched biological samples of people diagnosed with psychiatric disorders. Such a biobank would have considerable impact across the breadth of research in psychiatry. Common disease research has benefited massively from the UK Biobank project, where 500,000 people have been measured for hundreds of phenotypes as well as genome-wide genotypes. Such volunteer biobanks, however, underrepresent people with psychiatric disorders. Thus, the biobank needs to be established in collaboration with participants and other relevant stakeholders to ensure engagement with such an initiative. Avoiding pitfalls of stigmatization, the mental health community deserves the same opportunities to benefit from technological advances applied in other branches of medicine.

R6: Establish a Cross-Species Brain Gene Expression Resource

Having a comprehensive gene expression atlas will benefit the field if it captures multiple brain regions (e.g., prefrontal cortex, hippocampus, basal ganglia, cerebellum) at cellular resolution (e.g., neurons, oligodendrocytes, microglia, astrocytes), across developmental time points (e.g., embryonic, postnatal, adolescence, adult), and matched across multiple species (e.g., human, marmoset, mouse) (Figure 8.4). Some research questions cannot be addressed in human cellular models, and different species provide different experimental paradigms for the investigation of mechanisms. The resource could be designed to include gene expression changes resulting from specific perturbations that can be imposed in more than one species, in conjunction with phenotypes that transcend species. A matched resource for gene expression across species would support integration of results across paradigms and ensure early establishment of the direct relevance of results from animal models for human disease.

R7: Establish a Human Brain-QTL Resource

Extending the GTEx resource to include brain cell types brings another general benefit to the field. The SNP-gene eQTL associations from GTEx are a key resource for integration with GWAS SNP-trait associations. GTEx includes data from 11 brain regions, although the associations are at the level of bulk tissue rather than individual cell type. GTEx has demonstrated that highly significant associations can be identified from relatively small sample sizes (N ~100). Although postmortem brains are a scarce resource, the power of a brain-QTL study could be enhanced by sourcing different brain regions from different individuals so that gene expression measures from different brain regions are not correlated; this would increase the specificity with which cell type-specific

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)

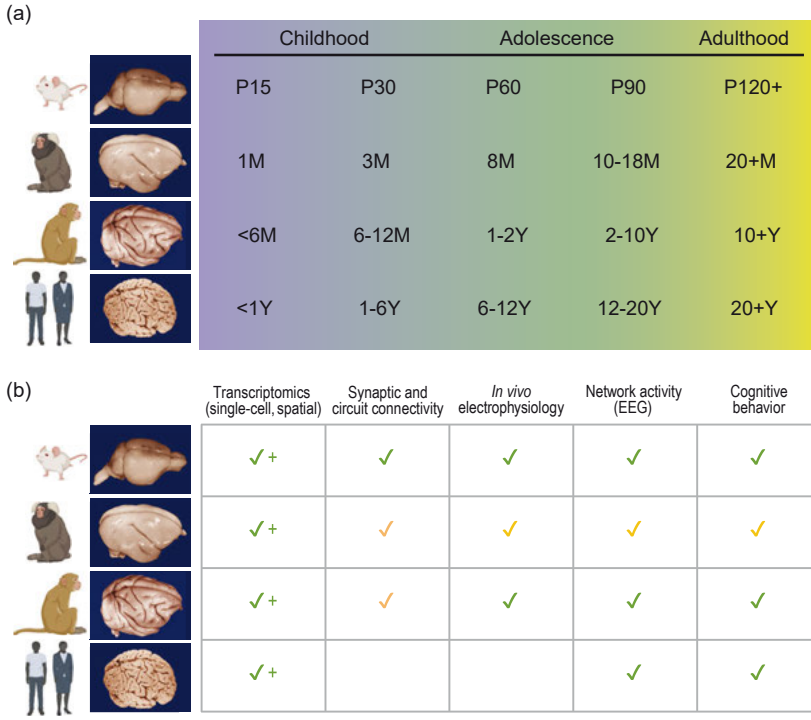

Figure 8.4 Neurodevelopmental timing in mouse, marmoset, macaque, and human. Approximate cross-species alignment of postnatal developmental ages is based on brain transcriptomics (Kang et al. 2011; Zhu et al. 2018), structural neuroimaging (Sawiak et al. 2018), and cellular neuroanatomy (Charvet 2020; Charvet and Finlay 2018; Charvet et al. 2022). Species icons from Biorender.com; brain images from brainmuseum.org. Figure prepared by Matthew Johnson.

QTL could be identified. The same resource could be used for expression and chromatin accessibility QTL. For a discussion of the existing gaps in brain-QTL resources, see Hu and Won (this volume).

R8: Establish a Brain-IGVF

We recommend that the scope of the Impact of Genomic Variation on Function (IGVF) Consortium be broadened to include a focus on brain cell types. IGVF was launched (a) to interrogate the functional impact of variants via; perturbation experiments such as MPRA and CRISPR engineering, (b) to establish cell type-specific regulatory networks, and (c) to develop a computational model to predict the variant function on phenotypes. The initiative, however, is not focused on generating brain-centric data sets. Thus, resulting data may not be readily applicable to risk variants of psychiatric disorders. We propose a Brain-IGVF consortium in which the functional impact of variants associated with

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 disorders can be studied in physiologically relevant cell types under pharmacological and genetic perturbations relevant to psychiatric disorders.

R9: Establish a Brain Ontology Consortium

We recommend that a brain ontology consortium be established to bring core groups together (a) to identify gaps and coordinate and synergize efforts, (b) to share data and protocols, and (c) to provide quality control and process data in a unified format. This will ensure that a rich, high-quality data set is generated to serve as a resource for interrogating the mechanisms which underpin psychiatry disorders. In addition to the manually curated GO, gene sets curated for brain function will facilitate interpretation of psychiatric GWAS. SynGO provides a clear example of how concerted efforts result in systematic annotation of genes and pathways involved in synapse biology (Koopmans et al. 2019). We propose extending SynGO, as a central repository of "brain-centric GO," which hosts manually curated brain-relevant biological pathways and the growing list of gene sets derived from perturbation experiments (e.g., CRISPR screens, drug treatment response). Ultimately, this repository could be further expanded to serve as a central data platform where multimodal data is held and accessible for integrative analyses.

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)