This is a section of doi:10.7551/mitpress/15380.001.0001

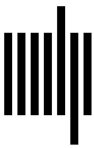
Exploring and Exploiting Genetic Risk for Psychiatric Disorders

Edited by: Joshua A. Gordon, Elisabeth Binder

Citation:

Exploring and Exploiting Genetic Risk for Psychiatric Disorders
Edited by: Joshua A. Gordon, Elisabeth Binder
DOI: 10.7551/mitpress/15380.001.0001
ISBN (electronic): 9780262377423
Publisher: The MIT Press
Published: 2023

The open access edition of this book was made possible by generous funding and support from MIT Press Direct to Open



The MIT Press

2

Delineating Additional Risk Factors

Angelica Ronald, Joshua A. Gordon,
Till F. M. Andlauer, Lukoye Atwoli, Na Cai,
Thomas Lehner, Michel G. Nivard, and Kathryn Roeder

Abstract

This chapter highlights paths, processes, and considerations that become important as we build on the initial success of large genome-wide association studies of psychiatric disorders. As such, it largely focuses on research on common genetic variation and human genetic research. It proposes directing research toward interrogating how genetic variation acts on the developing brain. For this reason, it discusses the potential value and pitfalls of using developmental, circuit-based, and quantitative symptom-based phenotypes in parallel to the traditional approach of reliance on binary diagnoses in genetic research designs. With respect to heterogeneity and co-occurrence present in psychiatric disorders, analytic approaches are outlined that can advance understanding, improve gene discovery, and potentially influence nosology. It argues that increasing cohort diversity is nonnegotiable: it is essential to improve gene discovery, translation, social justice, and research equity. Furthermore, a range of methods that interrogate the processes of environmental risk, gene–environment correlation, and gene–environment interaction enable a more accurate understanding of direct genetic effects and of how environments operate together with genetic risk for psychiatric disorders. Far from being a diversion, these environmentally informed methods are likely to catalyze biological insights. To this end, considerations for optimal future experimental study designs are discussed, outlining their characteristics and the prioritized approaches. The overarching goal is to deliver, through gene discovery research, translational benefits for individuals living with neurodevelopmental conditions and psychiatric disorders.

Background

Over the last two decades, remarkable progress has been made in psychiatric genetics,[1] yet huge knowledge gaps remain, and the delivery of therapeutic

[1] In this chapter, psychiatric genetics is used as an umbrella term to encompass genetic research on both psychiatric and neurodevelopmental conditions; our focus is solely on human genetics.

options created from biological insights is lacking. In this chapter, we contemplate how we can move forward into new scientific territory in our field through a discussion of the following questions:

- What are the advantages and disadvantages of studying (a) quantitative versus binary phenotypes and (b) developmental phenotypes versus diagnostic outcomes?
- How can we best exploit heterogeneity and co-occurrence to improve gene discovery and hypothesis testing?
- What are the costs and benefits of capturing the entirety of the allelic frequency spectrum translationally and biologically, and how does this differ by condition?
- How will increasing diversity in samples contribute to discovery, translation, justice, and equity, and what strategies are needed to pursue this?
- How do we explore the space of environmental risk and gene–environment relationships?

We end with a proposal for an optimal future study of genetic variation in psychiatric conditions. Throughout, the common concerns within our field are featured, including the heterogeneity within and across psychiatric disorders, a discussion of priorities given limited resources, and means to exploit available opportunities where they exist.

Alternative Phenotypes

What are the advantages and disadvantages of studying (a) quantitative versus binary phenotypes and (b) developmental phenotypes versus diagnostic outcomes? While this question pits approaches against each other, as if they are mutually exclusive, the conclusion we reached is that all phenotypic approaches have strengths.

Diagnostic phenotypes arguably offer the greatest specificity for directly studying the phenomena themselves (i.e., psychiatric disorders). Developmental phenotypes enable us to study the site and time of action in which genes, together with any environmental risk, have their etiological effect; this site and time of action being the developing brain. Quantitative trait measures offer a range of strengths and analytical flexibility with complementary confounds to diagnoses. Importantly, given the investments already made in diagnostic approaches and the complexities of measuring novel traits, the field needs to demonstrate that quantitative and developmental traits provide new or better biological information than diagnoses. As outlined in Tables 2.1 and 2.2, respectively, the potential for additional advances using quantitative and developmental phenotypes seems high.

Table 2.1 Advantages and disadvantages of quantitative and binary phenotypes.

	Advantages	Disadvantages
Quantitative phenotypes	Can be collected in nonclinical samples (e.g., general population samples, community samples, research cohort studies). Can be more specific than a compound diagnosis, e.g., by focusing on specific symptoms, age of onset, or progression. This may increase potential for biological discoveries that yield effective treatment targets. Potential for transdiagnostic relevance (e.g., inattentiveness, executive control, impulsivity). Flexibility of measurement and application: often not time- or resource-demanding; amenable to repeated assessments, parametric/hierarchical modeling, and factor analyses. No need for a separate "control" sample. Statistical power gains because no artificial dichotomization necessary.	Known biases inherent in nonclinical samples, including in patterns of participation and attrition (e.g., Martin et al. 2016). Measurement is not always standardized to the same degree as DSM and ICD, though see Strengths and Difficulties Questionnaire for an example of a standardized quantitative measure (Goodman 1997). Distributions can be skewed (often be handled by transformations) and parametric models can require removal of statistical outliers, possibly contradicting the focus in psychiatry on the most severely affected individuals. Rarely show 100% genetic overlap with psychiatric diagnoses; ADHD traits and ADHD are sample exceptions of high genetic overlap between diagnosis and symptoms) (Demontis et al. 2019). Individual phenotypes often studied separately (although multiple single scales are used to create a latent factor) whereas a diagnosis assesses multidimensional symptom data and reduces it to a single outcome or scale. Can be time-consuming and expensive to develop; a rapid, scalable phenotyping protocol requires, e.g., validation and reliability data, language translations, age-specific norms, etc.
Binary phenotypes	There is an established liability threshold model on which they can be modeled. Established statistics and effect measures (e.g., odds ratios, prevalence) enable direct comparisons across studies. Available as yes/no disease status in electronic health records data.	Require a control sample which can introduce ascertainment issues, population stratification, and other unknown confounds. The liability threshold model requires the assumption of an underlying latent liability which is not directly measured. Dichotomization of quantitative data leads to loss of information and statistical power. Because of cross-sectional assessments, binary phenotypes are prone to contain false-positive cases and false negative controls.

Quantitative Traits

As noted in Table 2.1, quantitative phenotypes offer both statistical and biological advantages over binary phenotypes as targets for gene discovery. The possibility that some dimensional phenotypes, if closer to the neurobiological pathways governed by genes than binary phenotypes, might reveal new or more specific genetic information is tantalizing. However, this excitement must be tempered by the need to focus on the most promising quantitative phenotypes and to address issues of measurement and sample biases.

Measurement

Standardized quantitative trait measures are important for integrating and maximally utilizing data from cohorts. For example, one of the most significant issues in research on clinical characteristics and symptom profiles in major depression lies in the vast numbers of scales on which depression can be measured, and the inconsistency in their usage between studies (Fried 2017). Measurement heterogeneity across cohorts has made inferences challenging; even principled approaches for symptom and patient classification give mixed results depending on assessment instruments (van Loo et al. 2012). Heterogeneity across different measurements of a phenotype can lead to deflated heritability estimates in genetic studies of quantitative traits (Kalman et al. 2021; Wray and Maier 2014; see discussion below on Heterogeneity and Co-occurrence). As such, work to standardize quantitative trait measures that are consistently employed across studies is vital.

Issues also exist for binary phenotypes. Defining cases and controls is not straightforward. Any case/control study implicitly conditions on the selection of controls; for example, whether controls are selected with or without screening may influence genetic findings (Peyrot et al. 2016). This is another reason to avoid a binary diagnosis phenotype as the phenotype of interest relates to what will most help the patient. Case/control studies identify genes associated with a binary disorder (i.e., the biology relevant for the entire diagnostic construct irrespective of individual symptoms, age of onset, or illness progression). There is no guarantee that any of these genes will be functionally relevant for clinical symptoms observed in a patient at the point of diagnosis or treatment. By contrast, understanding the genetics of specific elements of the impairment in patients, such as symptoms or disease progression, could offer more fruitful avenues for translating gene discoveries into targeted therapeutic options (Stein et al. 2021).

Circuit-Based Phenotypes

For quantitative phenotypes to identify novel biology and/or therapeutic targets, they must contribute in a novel way beyond the progress already being

made in gene discovery based on binary diagnoses. This potential is maximized when the quantitative phenotypes used for gene discovery are closer to the underlying biological pathways that genes directly encode. An oft-repeated trope is that genes do not encode mental illnesses; they encode proteins, which form biological pathways governing the behavior of cells, neural circuits, and neural systems, which in turn produce the behaviors that underpin mental illnesses. Mental illnesses are heterogeneous collections of various complex behaviors, several layers removed from basic molecular pathways encoded by genes. A promising approach would be to develop and quantify behaviors that represent the direct output of neural circuits, which lie several steps closer to genes along this pathway.

There are several examples of promising circuit-dependent behaviors ripe for exploitation in this manner. Fear-related behaviors, for instance, are quantifiable and represent the output of well-defined pathways involving an extended circuit that includes the amygdala and midbrain structures. Reward learning and especially the computation of reward prediction error are established, computationally modeled behaviors that map onto circuits, including the dopamine neurons of the ventral tegmental area and their targets in the nucleus accumbens. In both cases, translatable, quantitative measures exist that have the potential for scalable implementation in genetic studies (Haaker et al. 2019; Vassena et al. 2017).

Key outstanding issues are designing and validating these and other measures of circuit-level function and their heritability. Furthermore, to prove the promise of the approach, it would be useful to pilot a handful of such behaviors on a large enough scale to verify their potential to identify novel or more specific genetic findings.

Sample Biases

Though sampling biases are widespread in cohorts collected for genetic studies of psychiatric disorders, the directions and types of biases vary by data collection approach. First, there is a well-established WEIRD bias (western, educated, industrialized, rich, and democratic) in cohorts collected for genetic studies, and in data collected through voluntary participation there is a bias for well-educated and healthy individuals. For example, genome-wide association studies (GWASs) on participation in the mental health questionnaire in the UK Biobank (Fry et al. 2017) showed that the polygenic risk score (PRS) for mental health questionnaire participation is positively correlated with educational attainment and better health, and negatively correlated with psychological distress and schizophrenia (Adams et al. 2020). Similarly, data obtained from paying customers of consumer genomics companies may not be representative of the general population in terms of socioeconomic status and educational attainment (Hyde et al. 2016). Conditional attrition is an additional bias that can create challenges in genetic exploration of longitudinal cohorts (e.g., Martin et al. 2016).

Improved statistical modeling of participation bias, however, can aid valid inference. Careful socioeconomic and demographic reweighting of UK Biobank data to counter participation or healthy volunteer bias reversed spurious participation associations (van Alten et al. 2022). In biobanks where participants are recruited in clinical settings and may be less healthy than the underlying population, comprehensive reweighting based on medical information has proven effective in countering bias in PRS-based health care (Lee et al. 2022).

Developmental Phenotypes

Another underexploited area for future focus in gene discovery is that of developmental phenotypes, which refer to measures capturing infant, child, and adolescent development relevant to psychiatry. Examples include infant temperament, childhood behavior, cognitive constructs such as joint attention or effortful control, developmental milestones, and adolescent-onset traits such as risk taking and emerging mental health traits. Existing cohorts collecting developmental phenotypes that contribute to the psychiatric genetics literature include the Norwegian Mother, Father and Child Cohort Study (Magnus et al. 2016), the Twins Early Development Study (Rimfeld et al. 2019), and the Adolescent Brain Cognitive Development Study (Saragosa-Harris et al. 2022). Existing literature has focused on childhood and adolescence (Pain et al. 2018); to date, no well-powered ($n > 10{,}000$) gene discovery research is available on behavioral phenotypes in the infant years (0–3 years).

The advantage and justification for studying developmental traits is to understand the changes in the brain that occur prior to the onset of a neurodevelopmental or psychiatric condition. A pervasive hypothesis in psychiatry is that many psychiatric conditions have a developmental origin. As such, a complementary approach to studying the genetics of established diagnoses is to conduct gene discovery on pertinent developmental traits that capture the atypical development at the site of action: the developing brain (see Table 2.2).

Longitudinal population samples with reliable phenotyping are essential for unbiased estimates of prevalence, co-occurrence, and the temporal order in which phenotypes present themselves. As with all quantitative traits, gene discovery research on developmental phenotypes needs to reveal new or more specific genetic information relevant to psychiatric disorders to be justified in our field.

Psychiatric Diagnoses

The advantages and disadvantages of psychiatric diagnoses are listed in Table 2.2. At present, psychiatric diagnoses lack objectively measurable diagnostic criteria such as biomarkers. To boost sample sizes, some studies have relied on broad diagnostic definitions, such as allowing self-reported diagnoses instead of requiring diagnostic validation by a trained professional (Hyde et al. 2016).

Some studies exclude participants who have specific co-occurring conditions; it is not until the recent increase in availability of data from electronic health records and medical registries that diagnostic sequences or switches have been studied and accounted for in genetic studies (Krebs et al. 2021). Going forward, developments in methods to incorporate disease trajectories into genetic studies would undoubtedly reshape the research landscape of psychiatric genetics. Despite the disadvantages of psychiatric diagnoses listed in Table 2.2, considerable success has been achieved in gene discovery, likely owing to the relative ease of assembling large cohorts of individuals by diagnosis.

Table 2.2 Advantages and disadvantages of developmental phenotypes and psychiatric diagnoses.

	Advantages	Disadvantages
Developmental phenotypes	Developmental change and continuity can be accounted for through repeated measures or trajectory phenotypes. Closer reflection of nature to the extent that all individuals continually develop over time. Capture the developing brain, at a time when early interventions or preventive strategies would be appropriate. Potential to offer insights into diagnostic subtypes based on age of onset, trajectory subtype or prodromal features. High potential for transdiagnostic relevance.	The genetic correlation between a developmental phenotype and a psychiatric diagnosis is not always known. No guaranteed specificity to disorder of interest. The time-lag involved in prospective studies between time of data collection and transition of affected individuals to illness can impact feasibility. There are, however, strategic solutions in terms of design and methods.
Psychiatric diagnoses	Measurement typically relies on standardized diagnostic systems, including DSM and ICD; decades of research, clinical use and reiterations have refined these systems. Structured clinical interviews designed to handle a multidimensional symptomatology space and to dichotomize information with maximum reliability. Studying the same phenotype that is intended to benefit from the research.	Diagnoses are arguably not as objective as a biomarker. Some studies mix clinician-based best-estimate diagnoses with self-reported diagnoses, potentially adding heterogeneity and increasing error of measurement. Clinical samples susceptible to a range of biases, including over-representation of more severe cases and help-seeking characteristics in individuals. Heterogeneity in diagnostic practice across sites. Heavy reliance on combining data across multiple sites due to low numbers of cases at any single site. Cross-sectional assessments are prone to contain false-positive cases and false negative controls.

Summary

To expand research on quantitative and developmental phenotypes, concerted effort is needed to identify standardized measures. A definition of common standards and benchmarking of diverse measures across different cohorts, biobanks, and electronic health records is important to facilitate harmonization and to reduce heterogeneity between them. In the meantime, a phenotypic reference panel is one solution to model phenotypic heterogeneity if meta-analyzing across cohorts that have relied on different quantitative trait measures (Luningham et al. 2020; Luningham et al. 2019). We further note that phenotypes should be assessed for specificity and relevance to psychiatric outcomes. Though no sample is free of all bias, careful consideration of the range of biases that are present in clinical and nonclinical samples as described can help to ensure maximal specificity and relevance. A range of neurodevelopmental conditions have benefited from genetic research that focuses on both quantitative traits alongside the more traditional case-control design (Demontis et al. 2019; Pain et al. 2018). The next decade will reveal whether the merits of developmental and quantitative phenotypes will pay off in terms of biological innovations that provide translational benefits.

Heterogeneity and Co-occurrence

Heterogeneity and co-occurrence are rife in psychiatric illness, and the genetic data suggest that these phenomena are important and could be useful in understanding the causes of mental illness and developing targets for new treatments. How do we best exploit heterogeneity and co-occurrence to improve gene discovery and hypothesis testing?

Heterogeneity

There is heterogeneity in psychiatric genetics; accepting this statement as universal in psychiatry will enable our field to be constructive about heterogeneity in ways that have so far not been achieved. Here, we discuss how heterogeneity manifests in psychiatric disorders and the methods that can be used to account for this heterogeneity.

Why is tackling heterogeneity so important in psychiatric genetics? In short, heterogeneity in samples will reduce genetic signal and impede gene discovery. In both empirical data and simulations (Cai et al. 2020b; Dahl et al. 2020), heritability estimates based on single nucleotide polymorphisms (SNPs) are reduced when two phenotypes with heterogeneous genetic architectures are analyzed as one. For instance, the estimate of SNP-based heritability is deflated when different definitions of major depressive disorder that have different genetic architectures are analyzed as a single entity; major depressive disorder

with and without prior severe stress exposures are both found to have higher SNP heritabilities than when they were analyzed together (Peterson et al. 2018).

Given mounting evidence that many psychiatric disorders may be the common outcome of heterogeneous pathways, it is becoming clear that treating each disorder as an entity and studying it at the level of a binary diagnosis or quantitative symptom total score is not the only or best solution. We suggest two alternative ways forward:

1. Examine the biological mechanisms behind the putative subtypes of psychiatric disorders delineated by the diagnostic (e.g., DSM-5) specifiers. Using major depressive disorder as an example, these specifiers came from decades of clinical experience and patients' own accounts; specifiers for major depressive disorder such as atypical, melancholic, and anxious depression have been proposed in addition to those based on developmental timing (Harrington et al. 1996), treatment resistance (Fagiolini and Kupfer 2003), and recurrence (Merikangas et al. 1994). Furthermore, studies have found that typical and atypical major depressive disorder subtypes show different patterns of associations with other PRSs for other traits or diseases (Badini et al. 2022; Milaneschi et al. 2016, 2017, 2020).
2. Refocus genetic discovery efforts on more granular phenotypes with higher validity and reliability than binary diagnoses or sum scores, such as individual symptoms or clinical characteristics (Fried 2015; Persons 1986). Staying with the example of major depressive disorder, its symptoms are genetically correlated with each other in the range of 0.6–0.9 (Howard et al. 2020; Jermy et al. 2022). Further, although disorders are often defined through sum scores of symptoms, genetic effects captured through symptoms may not account for all genetic risks for their corresponding disorders; the average genetic correlation between a specific major depressive disorder symptom and the disorder is 0.6 (Jermy et al. 2022). Looking into the genetics of individual symptoms and other clinical characteristics is likely a complementary approach to studying binary diagnoses, with the potential to generate new hypotheses and discoveries.

Approaches to Handling Heterogeneity

Covariates. To perform a GWAS, we regress the phenotype on the genotype, often including additional covariates. Appropriate analysis, however, depends upon whether the covariates are confounders or colliders. Confounding occurs when the phenotype and the genotype have a shared common cause that is not controlled for, thereby inducing a false association (e.g., ancestry). In contrast, a collider is a third variable that is influenced by both the phenotype and the genotype. Including a collider induces a false association between genotype

and phenotype, which is called collider bias. Collider bias is a major statistical challenge that prevents the inclusion of a third variable (which may be a source of heterogeneity) from being "controlled for" in GWASs via a covariate approach. There are covariate selection methods, however, that can be used to avoid collider bias (Aschard et al. 2017; Dahl et al. 2019). Such methods have been used in smaller omics data sets to increase power for discovery (Gallois et al. 2019). In a GWAS design, these methods need to be applied for each individual SNP and, as such, need to be made more tractable for use in large sample sizes in psychiatric genetics.

Polygenic risk scores can be used to differentiate genetically defined psychiatric subgroups. For example, bipolar disorder type II was found to be most strongly associated with major depressive disorder PRSs whereas bipolar disorder type I was most strongly associated with schizophrenia PRSs (Stahl et al. 2019). Conducting separate GWASs by subgroups inevitably reduces power in terms of sample size. Nevertheless, the definition of disorder subtypes using PRS profiles is promising. This approach does not require further phenotypic data and should work even in small samples given the high statistical power of PRS analyses.

Case by case genome-wide association studies. Pooling data across disorders is an effective strategy for increasing statistical power for the discovery of loci related to psychiatric disorders (Grotzinger et al. 2022). Where heterogeneity is the specific research interest, the opposite might be effective. To this end, case by case GWASs, performed for one disorder versus those for other disorders, can identify loci that specifically distinguish between disorders (Peyrot and Price 2021), providing tractable biological leads that may aid in understanding one specific form of heterogeneity: differences between highly comorbid disorders.

Use genomic structural equation modeling to combine two disorder genome-wide association studies. Genomic structural equation modeling, or other tools suited for further analysis based on GWAS summary data, can aid in delineating distinctions between different aspects of a single phenotype (Grotzinger et al. 2019). An example from sociogenetics is the effort to disentangle the cognitive and noncognitive contributions to success in school. Demange et al. (2021) used the genome-wide association studies of cognition and education and split the GWAS signal for education into a cognitive component and a component not related to cognition. The specific noncognitive component related genetically to conscientiousness and delay of gratification and negatively to psychiatric traits such as bipolar disorder and schizophrenia. The ability to model both shared and trait-specific signals could be applied to study the relationships between psychiatric disorder subtypes. This would allow for the study of traits that have not themselves been directly subjected to gene discovery studies.

Relevance for Nosology

Heterogeneity in manifestations and etiologies within current diagnostic categories is a result of our operationalization of neurodevelopmental conditions and psychiatric disorders (Cai et al. 2020a; Ronald et al. 2011). Without knowing the biological mechanisms underlying them, we have created diagnostic categories that do not necessarily line up with etiological pathways. As a result, it is not surprising that many such diagnostic categories contain heterogeneous etiologies, some of which may be indexed by heterogeneous manifestations (through symptoms and other clinical characteristics). The process of developing testable hypotheses on disorder subtypes, and the knowledge gained through this research might ultimately improve diagnostic category operationalization.

Summary

Heterogeneity is the norm in psychiatry and addressing it is likely to pay off from a gene discovery point of view. Heterogeneity can be studied in several ways. Two options are to investigate diagnostic specifiers and focus on specific symptom profiles. Here we considered a range of methodological options to address heterogeneity. This area of future research may also lead to clinical impact by influencing nosology via refined biological understanding.

Co-occurrence

Co-occurrence of conditions is high in psychiatry, both within psychiatric disorders and between psychiatric and nonpsychiatric conditions. Ignoring co-occurrence means ignoring a fundamental feature of psychiatric disorders, impacting genetic analyses. By embracing transdiagnostic features of related disorders, however, we can surpass traditional diagnostic categories. Indeed, high co-occurrence (Kessler et al. 2005) and pleiotropy (Brainstorm Consortium et al. 2018; Gandal et al. 2018; Lee et al. 2019) between psychiatric disorders have motivated attempts to identify common genetic factors and implicated molecular pathways shared by multiple psychiatric disorders (Maier et al. 2015; Schork et al. 2019). Here, we focus on three approaches for exploiting co-occurrence to improve gene discovery and hypothesis testing.

Longitudinal Co-occurrence

One aspect of co-occurences that has not received sufficient attention is the temporal relationships between the multiple conditions. Any cross-sectional assessment, whether obtained through research, electronic health records, or nonclinical cohort studies, may only capture concurrently occurring diagnoses at the time of the survey. However, co-occurring conditions may have different

times of onset, exhibit different peaks of severity, or even show completely nonoverlapping phases of manifestation, with underlying genetic factors acting pleiotropically. Importantly, co-occurring disorders share a common etiology, and/or one of the disorders could increase the risk of another over time. Therefore, more nuanced approaches to co-occurrence that aim to capture co-occurrence longitudinally rather than cross-sectionally may yield additional insight. Alternatively, co-occurring disorders might constitute a different disease entity altogether. These and other scenarios can only be examined by determining the longitudinal sequence of occurrence, accounting for the relative time of onset of the co-occurring conditions. Gathering such longitudinal information would facilitate modeling disorders as longitudinal occurrences, testing such models for their explanatory power, and mapping them onto the underlying genetics. For example, there might be a different genetic etiology for individuals whose co-occurring mood and anxiety disorders begin with major depression and progress to panic disorder, compared to those whose begin with panic disorder and progress to major depression. Moreover, longitudinal modeling might enable personalized risk profiles for psychiatric disorders including better predictions of disease progression.

Exploiting Nonpsychiatric Disorder Biology via Co-occurrence

Many individuals with psychiatric disorders have co-occurring nonpsychiatric disorders with potentially partly shared underlying genetic risk factors. For example, depression is more frequent in patients diagnosed with some nonpsychiatric disorders than in the general population (Boeschoten et al. 2017; Garrido et al. 2017). Conducting genetic analyses in patients showing co-occurrence of psychiatric with nonpsychiatric disorders for which the biology is well known might facilitate unraveling the etiology of psychiatric disorders.

Concepts and Methods to Model Co-occurrence on a Latent Level

Various models have been proposed to categorize features of psychopathology in hierarchical models. For example, the p factor (Caspi et al. 2014), derived from bifactor models, represents an underlying general liability for psychiatric conditions. Preliminary evidence suggests a genetic basis of the p factor (Selzam et al. 2018), but the utility of this construct continues to be scrutinized (Grotzinger et al. 2022).

Genomic structural equation modeling (SEM) is a method to perform modeling and hypothesis testing on complex etiological models after genome-wide association studies that is well suited for modeling psychiatric co-occurrence. In a specific application of genomic SEM sharing features with Mendelian randomization, Grotzinger et al. (2022) evaluated various forms of the latent p factor using genome-wide association studies of eleven psychopathologies. Genomic SEM analyses indicated a model where pleiotropy or co-occurrence

was best explained by a set of correlated factors, each influencing two to three traits, rather than an overarching p factor. It remains to be seen whether alternate taxonomies of psychopathology, such as HiTOP (Conway et al. 2022), are consistent with these results.

Summary

The field of psychiatric genetics has mainly studied individual conditions, with some exceptions (Smoller et al. 2013). Moving forward, co-occurrence could be exploited by considering longitudinal co-occurrence, the biology of co-occurring nonpsychiatric disorders, by modeling co-occurrence latently, and by applying methods like genomic SEM.

Allelic Frequency Spectrum

What are the costs and benefits of capturing the entirety of the allelic frequency spectrum translationally and biologically, and how does this differ by condition? After a brief description of previous work from the psychiatric genetics community, including the Psychiatric Genomics Consortium, our discussion focuses on ways to advance future research. We explain what we mean by the entirety of the allelic frequency spectrum and which parts are currently not well studied. Examples are provided of work that has achieved findings on the "missing" part of the allelic frequency spectrum from other fields. We conclude that a likely by-product of the initiation and prioritization of genetic studies in diverse populations will be greater knowledge regarding the missing parts of the allelic frequency spectrum.

Work from the Psychiatric Genomics Consortium

Founded in 2008, the explicit goal of the Psychiatric Genomics Consortium was to perform large genome-wide association studies on neuropsychiatric disorders and to delineate their genetic and phenotypic architecture. Organized around 14 working groups to study 11 psychiatric disorders and cross-disorder genetics, the consortium has published more than 320 articles, including papers on the common variant risk architecture of schizophrenia, bipolar disorder, major depressive disorder, and autism spectrum disorder, as well as their genetic sharing and phenotypic overlap. By integrating discovery cohorts and generating genotypes for meta-analysis involving over 40 countries, and more than 800 investigators from over 150 institutions, the Psychiatric Genomics Consortium has become the largest collaboration in psychiatric genetics.

To date, the Psychiatric Genomics Consortium is focused on array-based common variants and array-derived structural variant analysis. While these methods capture many common variants with small effects on liability to

disease, the approaches seldom capture low-frequency variants due to lack of power (except for rare variants with large effects). As such, current efforts do not capture the full allelic frequency spectrum of disease. Furthermore, though the consortium is international in nature, until recently the integrated cohorts were almost exclusively European in genetic ancestry and offered little diversity and limited genetic admixture.

What Is the Allelic Frequency Spectrum and What Is Missing?

It is estimated that any given human genome is different from that of another by about four to five million loci, many of which are SNPs (1000 Genomes Project Consortium et al. 2015). Variants lie on an allelic frequency spectrum largely determined by forces of population history (population size bottle necks, drift, and natural selection), which tend to limit effect sizes of common variants. Allelic frequency refers to the frequency with which a given variant is found in the population; a minor allelic frequency of 5% means that 5% of chromosomes carry that particular allele. Common variants are typically defined as those with a minor allele frequency of greater than 1%. Many of these are located in the noncoding regions of the genome and other places where genetic variation can be tolerated without catastrophic effects. Each associated common variant by itself confers a very small increase in risk. A PRS is commonly estimated as the aggregate genomic risk from the total of such variants in an individual. Since power for detection of a risk allele is a function of the variance explained (i.e., $2p(1 - p)b^2$, where p is allele frequency and b is effect size) and sample size, rare-variant associations (minor allele frequency under 0.5%) are potentially detectable when they have large effect sizes. Hence, very large cohorts, such as UK Biobank, are powered to extend discovery to a minor allele frequency of $\geq 0.01\%$ and to find alleles with smaller effect sizes, but only for common diseases and quantitative traits (as case numbers for less common disorders are lower). Sequencing studies of large case-control cohorts have identified rare genetic variants associated with schizophrenia with bigger effects (Martin et al. 2022). Individual variants may have frequencies as low as 0.005% but their association is established through gene-level burden tests. Between these two extremes exists an area of the allele frequency spectrum where variants occur with minor allele frequencies between 0.005% and 0.01%. Current experimental designs do not typically have the power to detect alleles with minor allele frequencies in this range (in Singh et al. 2022, see Fig. 6.).

Three potential reasons for directing efforts to find this missing part of the allele frequency spectrum (and effect size) space are as follows:

1. New biological pathways may be uncovered.
2. New gene-gene, gene–environment, and gene–sex interactions may be discovered.

3. There might be a combination of allele frequency and effect size that enables novel paradigms of neurobiological inquiry.

Existing research outside of psychiatry shows that alleles in this "missing" spectrum can be identified, but only with extremely large sample sizes. For example, genome-wide association studies on height identified variants in this space (Marouli et al. 2017; Yengo et al. 2022) as well as studies on neurodevelopmental disorders (Stoll et al. 2013) and inflammatory bowel disease (Luo et al. 2017).

Capturing the Entire Allelic Frequency Spectrum versus Other Research Priorities

Assuming a finite budget for data collection, obtaining sufficient sample sizes to fill out the missing regions of the allelic frequency spectrum for each condition must be weighed against other priorities, including the competing goal of expanding samples to encompass full global representation. With the information we have at hand and considering the current research landscape, we emphasize the importance of research efforts and resources on diverse ancestry populations over attempting to capture the "missing" part of the allele frequency spectrum. The missing part of the frequency spectrum will fill organically with larger sample sizes as they accumulate.

Focusing on Diversity

Historically, most gene discoveries have been conducted in European ancestry individuals. How will increasing diversity in samples contribute to discovery, translation, justice, and equity, and what strategies are needed to pursue this?

The failure to recruit individuals from diverse ancestry into genetic samples has significant repercussions in at least three areas. First, without diversity, the knowledge gained from genetics will be inequitable. It is challenging or impossible to apply genetic knowledge from existing data sets to risk prediction in non-European populations, and therapeutics developed based on this knowledge might not be globally applicable (Martin et al. 2017, 2019). Second, in some ancestral populations, additional risk variants may be present, or the frequency of associated variants might differ from other populations. Including diverse ancestries may therefore identify a greater variety of biological or therapeutic pathways potentially applicable to all ancestries. Third, due to the larger linkage disequilibrium blocks present in more recently diverged ancestries compared to ancestrally older populations, fine mapping of risk loci will be improved through genetic research on diverse samples. Recently, there has been a marked shift toward broader global representation in existing and new consortia-based studies.

In this section, we discuss strategies employed in such diversification efforts, and the lessons that have emerged from them. Subsequently, we address how increasing cohort diversity will contribute to gene discovery, translation, social justice, and research equity and consider optimal strategies to pursue this goal.

Strategies and Lessons Learned

In all attempts to increase diversity in global mental health research, there are problems inherent in researchers from high-income countries conducting research in low- and middle-income countries. These are largely fueled by different resources in terms of finances, human resource capacity, equipment availability, and institutional support. Kumar et al. (2022) examined this problem and have proposed recommendations for consideration by high-income country institutions as they establish collaborations with low- and middle-income country institutions. These recommendations include (a) the need to devolve global health research centers to where the health challenges being addressed are located and (b) to invest more in researchers from low- and middle-income countries. It is imperative that funding proposals include capacity development to ensure that local scientists are able to continue with similar work past the funding cycle.

An example of an attempt to design collaborations based on equity can be found in the Academic Model Providing Access to Health (AMPATH) collaboration between a consortium of North American institutions and Moi University in western Kenya. This relationship might provide a framework for broad-based collaborations of this kind, cutting across research, education, and health service delivery. Figure 2.1 provides a summary of the practical strategies that have grown out of this collaboration, based on principles articulated by Melby et al. (2016) and described in detail by Turissini et al. (2020).

A specific example of the application of these principles in the same setting, from a broad research perspective, is the Neuropsychiatric Genetics of African Populations Psychosis (NeuroGAP Psychosis) study (Martin et al. 2022; Stevenson et al. 2019). This collaboration brings researchers from Kenya, Uganda, Ethiopia, and South Africa together with collaborators from the United States. To date, this study has collected over 37,000 samples and is developing the capacity of local scientists to continue with the work beyond the funding cycle. Lessons learned from this work reemphasize the need for including local collaborators right from the beginning, designing clear and detailed collaboration agreements that keep equity in mind, and creating flexible funding mechanisms that recognize the resource and capacity differentials and intentionally set out to address them.

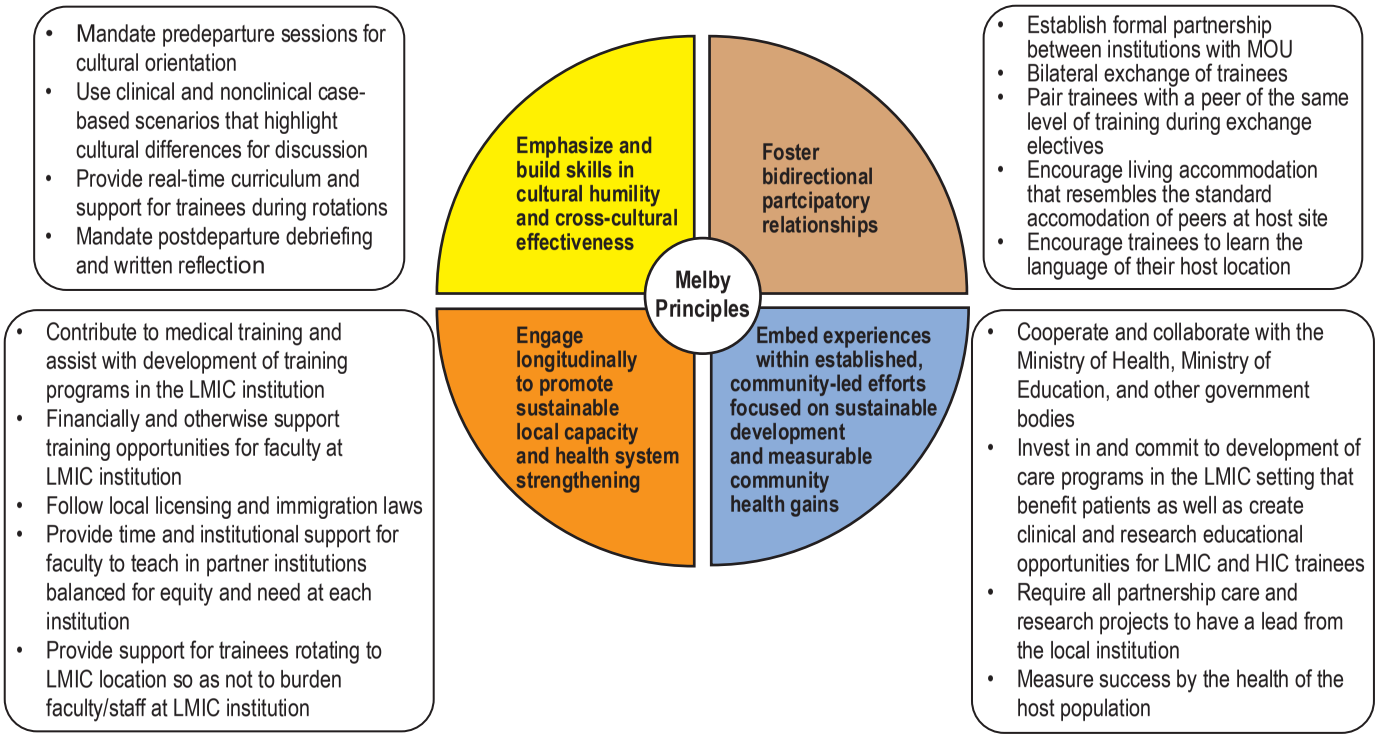


Figure 2.1 Practical strategies for implementation of Melby principles from the AMPATH experience. MOU: memorandum of understanding; LMIC: low- or middle-income country; HIC: high-income country.

Sample Diversity and Gene Discovery

Collection of diverse ancestry samples in psychiatric genetics will contribute to gene discovery and post-discovery follow-up. Due to migration patterns out of Africa, a substantial portion of allelic diversity was lost in ancestrally newer populations (1000 Genomes Project Consortium et al. 2015), and allelic diversity can only be determined by studying ancestrally older populations (Bentley et al. 2020). One example is the discovery of variants in *PCSK9*, which dramatically reduce low-density lipoprotein cholesterol concentrations. Discoveries such as this can identify key genes for drug therapy, which has implications for all populations.

Another advantage of studying diverse ancestries for gene discovery is that allelic frequencies vary by population, and power is greater when alleles are more common. Discovering specific alleles can thus be easier, due to higher allele frequencies, in some ancestries over others. For instance, the first two genome-wide significant loci for recurrent major depressive disorder in a Han Chinese population have high-risk allele frequencies (45% and 26%, respectively) in the Han Chinese population, but much lower allele frequencies in the European population (3% and 8%, respectively) (Converge Consortium 2015). Both have been replicated in an independent Han Chinese cohort, but not in European cohorts (Converge Consortium 2015).

Beyond gene discovery efforts, sample diversity will enable PRSs to be created that are optimized for all ancestries. It has been shown that even in examples where shared variants are discovered across ancestries, PRSs are not always portable across them. The PRSs generated from European genome-wide association studies on schizophrenia were 45% as accurate in predicting schizophrenia in samples with East Asian descent as in European ancestry samples (Lam et al. 2019). Ancestrally younger populations have considerably longer blocks of linkage disequilibrium (1000 Genomes Project Consortium et al. 2015). Long blocks of linkage disequilibrium complicate fine-mapping efforts and impair the portability of PRSs.

Sample Diversity and Fine Mapping

Statistical fine mapping is used to identify the credible set of causal variants at each GWAS-associated locus. Identifying these variants is typically the prerequisite for uncovering the associated gene and the mode of association (e.g., increased expression or alternate splicing).

When near one another, SNPs tend to be correlated due to linkage disequilibrium—a limiting factor for fine mapping. For example, when realizations of a "causal" SNP are perfectly correlated ($r = 1$) with one or more noncausal SNPs, no fine-mapping approach can distinguish between these variants. Fine-mapping intervals would be smaller if they were applied to a population with shorter correlated blocks (e.g., sub-Saharan African ancestry compared to

European ancestry). A large sample is required to detect each independent signal, and current samples of non-European populations are not large enough for this purpose. Clearly, increasing the sample sizes or incorporating information from both populations would yield tighter intervals for fine mapping.

Sample Diversity and Translation

Increasing diversity in samples is also essential for translation. Polygenic risk scores hold potential for clinical application to facilitate, for example, diagnosis and predict risk and progression. A PRS developed from samples of European ancestry, however, associates less strongly with the same phenotype in non-European samples (Curtis 2018; Yang and Zhou 2022). Notably, a PRS for schizophrenia was shown to be more strongly associated with ancestry than with schizophrenia (Curtis 2018). The challenge is largely expected to be due to linkage disequilibrium patterns and variant frequencies varying across ancestries, although differences in phenotyping also need careful consideration.

Phenotype differences between cohorts across ancestries further exacerbate the low portability of genetic findings. Phenotype differences can emerge as a result of study design, self-selection biases in participation, and other cultural differences in disease diagnosis. Concordance in phenotypes across ancestries also varies based on the specific psychiatric disorder. For example, while the genetic correlation for schizophrenia between East Asian and European populations is 0.98 (Lam et al. 2019) and the equivalent for bipolar disorder is 0.68, that of major depressive disorder between the same two populations is lower, at 0.33 (Bigdeli et al. 2017). In fact, just 11% of depression risk loci robustly identified in Europeans are associated with depression in East Asia (Giannakopoulou et al. 2021).

Moreover, efficacy of a treatment and risk of adverse effects may be dependent on genetic variants influencing drug metabolism or treatment-associated complications (e.g., neutralizing antibodies against biopharmaceuticals). The frequencies of these variants are ancestry dependent (Andlauer et al. 2020) and, hence, management of effective treatment and risk of adverse effects requires genetic analyses on diverse ancestries.

Increasing Diversity in Samples

The need for social justice and equity provides a strong argument for inclusion of diverse communities, both in the genetic samples we study and in the research workforce that conducts research on these samples. The social justice and equity arguments for diversifying our genetic samples derive from the need to ensure that all global citizens benefit from the advances that arise from gene discovery. Genetics provides opportunities for clinical advances in two independent ways. First, it is believed that genetic information about risk and resilience will be useful clinically at both the population and individual levels.

As theory predicts and empirical studies verify, the application of genetic information to risk prediction only works within a given genetic background. It is thus imperative for genetic studies to include diverse samples so that individuals from non-European ancestries can benefit from predictive knowledge. Second, the biological pathways elucidated from these studies can identify novel treatment targets. If we fail to study genetic risk in non-European populations, we may miss important clues to novel treatments that may be specific to, or especially important in, addressing mental illness in individuals from these populations.

Arguments for diversifying the genetic research workforce are equally compelling. We cannot hope to understand the nature and needs of global communities if those leading the research are not from those communities. Moreover, the benefits, both material and inspirational, of contributing to advances gained through genetics would be unfairly constrained without such participation. Our field of psychiatric genetics requires both sample and workforce diversity for it to reach its goals and contribute to better health across the globe.

Environmental Risk and Gene–Environment Relationships

How do we explore the space of environmental risk and gene–environment relationships? Environmental risk can impact disease development, without any contributing genetic factors. However, many environments are related to our genotypes. Only some rare, stochastic environments (e.g., experiencing a tsunami) do not correlate in any way with our genotype or the genotype of our close relatives. Most environmental influences are correlated with genetic influence (gene–environment correlation), or their effect is contingent on genetic influence (gene–environment interaction).

The classic twin design, the adoption design, and the monozygotic twin design have traditionally been used to parse out environmental variance from heritable effects. These remain powerful and well-tested designs. Still, new approaches in psychiatric genetics can be used to accelerate progress in understanding environmental risk and gene–environment relationships. These approaches measure and capture environmental variables known to have a large effect on psychiatric phenotypes—a key example being childhood trauma (Nelson et al. 2020)—and to incorporate them into genetic designs where possible. In particular, we need to strengthen the ability of psychiatric genetic studies to address and incorporate environmental risk by identifying environmental parameters that can be efficiently measured, are feasible to collect, and will generalize across studies.

Gene–Environment Correlation: Behavior Genetic Designs

Models using twin data can estimate the presence of gene–environment correlation by estimating the heritability of environmental measures and the degree of shared genetic influences between a measured "environmental" variable and a psychiatric phenotype (e.g., Shakoor et al. 2016). The advent of genome-wide complex trait analysis and genome-based restricted maximum likelihood (Yang et al. 2011) enables estimation of the SNP-based heritability of any environment measured in unrelated individuals (Plomin 2014).

Genome-Wide Association Studies of "Environments"

A natural next step from estimating the heritability of an environmental variable is to conduct gene discovery via GWAS to find genetic variants influencing measured environments. Early on, a GWAS of a measure of childhood family environment was conducted (Butcher and Plomin 2008); more recently, UK Biobank data have been used, for example, to identify SNPs associated with social deprivation and household income (Hill et al. 2016). This literature moves the field forward from hypothesizing about gene–environment correlation to identifying its underlying biology; that is, which genetic variants play a role in influencing the environments in which people live.

Polygenic Risk Score Associations with Environments

Another form of evidence for gene–environment correlation is through PRS associations. Polygenic risk scores should index a summed additive genetic signal of a phenotype. If a PRS is associated with a measured environment, it suggests some shared variance. For example, in a systematic review of all studies using the latest PRS for attention deficit hyperactivity disorder (ADHD) (Demontis et al. 2019), the ADHD PRS was consistently associated with lower socioeconomic status (Ronald et al. 2021).

Indeed, the signal in PRSs is thought to include direct genetic effects, but also to be inflated by a range of indirect effects including gene–environment correlations. Partitioning these indirect effects from the direct genetic effects, within-family analysis designs has been proposed (Selzam et al. 2019). Ultimately, this type of work helps to quantify the extent of genetic effects.

Within-Family versus Between-Family Genetic Association

Within-family designs, whether in GWASs or for PRS associations, enable direct genetic effects to be isolated from indirect effects. Indirect effects can include gene–environment correlation, population stratification, and assortative mating. For example, we can conduct analyses with PRSs using a within-family design (e.g., within sibling pairs) and compare effect sizes to those found

for analyses with PRSs using a between-family design (i.e., unrelated individuals). This comparison enables us to gauge the extent of direct genetic effects relative to indirect effects (Selzam et al. 2019). When the within-family design involves comparing siblings, it controls for a wide range of environmental factors that are shared by the siblings.

Another within-family methodological development focuses on genes that are shared and not shared between parents and offspring. The traditional view is that within families, children's outcomes are influenced by the child's own genotype and the home environment shared with their parents. However, parents also have their own genotypes, some of which are shared with their offspring and some of which are not. Polygenic risk scores in parents can be partitioned into the alleles transmitted to their offspring and the parents' alleles that were not transmitted to their offspring but nevertheless may have influenced the child's environment. The latter process has been termed "genetic nurture" (Kong et al. 2018). These nontransmitted alleles may play a role in the offspring's outcome phenotypes through gene–environment correlation. This approach helps to unravel the causal pathways, both genetic and environmental, that contribute to risk for psychiatric disorders.

Gene–Environment Interaction

Twin Designs

Gene–environment interaction refers to the effect of environments being contingent on genetic influence, or vice versa. Twin models can test for gene–environment interaction (e.g., whether heritability varies as a function of environmental severity), while controlling for any gene–environment correlation. A recent example of gene–environment interaction in a twin design found that psychotic experiences are less heritable in individuals who have experienced greater environmental adversity and are more heritable for individuals who have experienced less environmental adversity, after controlling for any gene–environment correlation (Taylor et al. 2022).

Individual Loci and Polygenic Risk Scores

It is possible to identify interactions between multiple measured environments with specific genetic loci. First, iSet is a method based on linear mixed models that tests for interactions between sets of variants and environmental states or other contexts (Casale et al. 2017). It jointly tests for gene–environment interaction across multiple contexts and sets of adjacent variants in genomic loci across the genome and, as such, allows for characterizing the local architecture of gene–environment interactions. Second, the structured linear mixed model (StructLMM)—a variance component test to identify and characterize gene–environment interactions between individual SNPs and multiple

environments—allows the identification of interactions that are simultaneously driven by multiple environments (Moore et al. 2019). Similarly, the linear environment mixed model analysis (LEMMA) uses a Bayesian approach and estimates a linear combination of environmental variables that interacts with genetic variants (Kerin and Marchini 2020).

Polygenic gene–environment interaction can be tested using extensions to genome-wide genomic restricted maximum likelihood (used to estimate SNP-based heritability from genome-wide data) under the mixed model for gene–environment interaction (GxEMM) framework, by allowing for environment-specific genetic variance and noise (Dahl et al. 2020). Importantly, GxEMM is further able to accommodate quantitative and multiple environments, an extension from previous models, and has already been used to show polygenic interactions with environmental stress indices for major depression (Dahl et al. 2020).

Using Results from Genome-Wide Association Studies to Test for Environmental Causality

Mendelian randomization is a form of instrumental variable analysis that uses genetic variants as instrumental variables (Davey Smith and Hemani 2014). It is now an established method for deriving a form of evidence of causality, which leverages genetic information (and yet does not focus on genetic influence itself) to estimate the causal effect of an exposure on an outcome. It has the potential to evaluate the causal effect of environments on psychiatric phenotypes more cheaply and in a complementary manner than randomized control trials. Mendelian randomization has a range of assumptions which, to some degree, are managed by a range of complementary methods (because different Mendelian randomization methods have differing assumptions). It is a method that exploits results from GWASs to study nongenetic processes; namely, the causality of environmental modifiers.

Summary

There is a range of methods that enable environmental risk and gene–environment correlation and gene–environment interaction to be tested empirically using the twin design, GWASs, and post-GWAS approaches. Far from being a diversion in our field, these methods are likely to catalyze biological insights because they enable a more accurate understanding of direct genetic effects and of how environments operate together with genetic risk. As such, the opportunities and benefits offered by these approaches should not be underestimated and very much add to the scoreboard of contributions made by psychiatric genetic genome-wide research in completing the picture of why psychiatric disorders develop.

Future Experimental Design

To integrate the various threads of our discussions, let us consider some key principles of experimental design that will optimize future discovery and advance our understanding of the genetic architecture of neurodevelopmental conditions and psychiatric disorders and their neurobiology, while accounting for equity and inclusion. Given finite time, resources, and consents, considering these principles will help guide future gene discovery studies to ensure maximum returns.

The most pertinent design features and the key approaches that offer the greatest potential for such advancements are summarized in Table 2.3. We consistently identified the need for well-powered global cohorts with a focus on diverse ancestral representation as the feature that should receive the highest priority. Appropriately designed studies that also account for local genetic variation will facilitate fine mapping, gene identification, and risk prediction.

Neurodevelopmental conditions and psychiatric disorders are developmental disorders and shedding light on developmental trajectories and their underlying genetic architecture will aid in defining the neurobiology of these disorders. Thus, study designs that allow for the collection of neurodevelopmental

Table 2.3 Key features of experimental design in psychiatric genetics.

Priorities	Approaches
Diversity	Aim for diverse ancestral samples on a global scale. Engage local researchers in sampling decisions. Enable diversity in the psychiatric genetics workforce where possible.
Development	Include collection of appropriate and feasible developmental data (e.g., birth records, milestones) collected through electronic health records.
Environment	Link by location (e.g., using the American Community Survey). Include siblings in a subset of the sampling frame to enable within-family designs. Measure known environmental influences for stratification (e.g., via birth records where available).
Heterogeneity	Incorporate heterogeneity into the study design and plan for subgroup analyses. Collect detailed information and make it available to analysts. Make future contact part of the consent.
Quantitative traits	Supplement diagnoses and symptom trackers with quantitative phenotypes and phenotypes closer to biology (e.g., circuit-based behaviors).
Co-occurrence	Be explicit about disorder exclusions. Collect detailed information and make it available to analysts. Consider the relative timing of onset of a co-occurring condition, beyond cross-sectional absence/presence.
Genetic platform	Aim to sequence at the highest practical coverage and resolution. Sequencing methods capture more diversity in the genome due to less reliance on imputation.

phenotypes (e.g., as a minimum standard, birth records and electronic health records data) are advantageous whenever possible.

Another crucial co-factor in the risk architecture of mental disorders is the environment. Study designs that allow for the collection of a minimum set of environmental exposure data (e.g., geolocation, trauma) would enable analytical tools to evaluate environmental factors and gene–environment interactions.

Historically, psychiatric genetics has relied on binary clinical diagnoses as outcome variables—an approach that has led to successes in gene discovery. There is, however, a vast distance between the different levels of analysis needed in psychiatry: from genes to molecules to cells to pathways to circuits and behavior (Figure 2.1). It is widely accepted that the binary phenotypes used in psychiatry are complex and heterogeneous and do not necessarily map well to their underlying biology. For this reason, heterogeneity is a core feature of neurodevelopmental conditions and psychiatric disorders and must be accounted for in study design and analysis. In addition to the collection of clinical diagnoses, it is thus important for heritable, disease-related, quantitative, and developmental phenotypes to be measured and analyzed.

Neurodevelopmental conditions and psychiatric disorders often do not occur in isolation and are frequently accompanied by the co-occurrence of other disorders of the brain or organ systems. The co-occurrence of other disorders reflects the underlying genetic and phenotypic heterogeneity and their complexity. Designs that permit all relevant phenotypes and enable analyses that include them are, therefore, advantageous. The co-occurrence of non-brain diseases should be leveraged as a possible pointer to the underlying biology and incorporated into the analysis.

Finally, it is important for genetic variation present in the cohort to be assessed on contemporary sequencing platforms, at the highest practical sequence coverage and resolution. The sequencing effort should be commensurate to the effort that goes into ascertainment, consent, exposure measurements, and collection of binary and quantitative phenotype data. Based on the continuing technological advancements in sequencing technologies, and the accompanying reduction in the cost of generating whole genomes, array-based technologies carry fewer advantages for most new studies. By using cost-effective sequencing methods, including low-pass whole-genome sequencing (Li et al. 2021; Martin et al. 2021), major shortcomings of microarrays, in particular poor imputation quality in diverse ancestries and a preselection of known variants, can be avoided.

In an era of limited resources, we realize that the strategies discussed in this chapter and summarized above may be difficult to implement concurrently in a single study. We believe, however, that through team-based national and international collaborations that are developed with community engagement and explicit emphasis on equity, these strategies are achievable and worthwhile. The study of non-European and older ancestral populations as well as study designs that we have outlined will assist in further delineating the genetic

architecture of neurodevelopmental conditions and psychiatric disorders, dissect their phenotypic complexity, generate new discoveries of the underlying neurobiology, and ultimately achieve translational and clinical applications.

Acknowledgments

We thank Naomi Wray and Stephan Sanders for comments on earlier versions of this chapter. We also thank all the members of the Forum who contributed to the discussions in our group, in particular Jordan Smoller and Lea Davis. Views expressed herein do not necessarily represent the views of the NIH, HHS, or the United States Government.

Series Editor: J. R. Lupp
Editorial Assistance: C. Stephen, A. Ducey-Gessner,
Lektorat: BerlinScienceWorks

The book was set in TimesNewRoman and Arial.

Library of Congress Cataloging-in-Publication Data

Names: Gordon, Joshua A., editor. | Binder, Elisabeth (Elisabeth B.), editor.
Title: Exploring and exploiting genetic risk for psychiatric disorders / edited by Joshua A. Gordon and Elisabeth B. Binder.
Description: Cambridge, Massachusetts : The MIT Press, [2023] | Series: Strüngmann forum reports | Includes bibliographical references and index.
Identifiers: LCCN 2023021101 (print) | LCCN 2023021102 (ebook) | ISBN 9780262547383 (paperback) | ISBN 9780262377430 (epub) | ISBN 9780262377423 (pdf)
Subjects: LCSH: Mental illness--Genetic aspects.
Classification: LCC RC455.4.G4 E97 2023 (print) | LCC RC455.4.G4 (ebook)
| DDC 616.89/042--dc23/eng/20230527
LC record available at https://lccn.loc.gov/2023021101
LC ebook record available at https://lccn.loc.gov/2023021102

Ernst Strüngmann Forum (31st : 2022 : Frankfurt am Main, Germany)

10 9 8 7 6 5 4 3 2 1