Emerging phenotyping strategies will advance our understanding of psychiatric genetics

Sandra Sanchez-Roige[1,*],

Abraham A Palmer[1,2,*]

[1]Department of Psychiatry, University of California San Diego, La Jolla, CA 92093

[2]Institute for Genomic Medicine, University of California San Diego, La Jolla, CA 92093

Abstract

Over the last decade, genome-wide association studies (**GWAS**) of psychiatric disorders have identified numerous significant loci. Whereas these studies initially depended on cohorts ascertained for specific disorders, there has been a gradual shift in the ascertainment strategy towards population-based cohorts (**PBCs**) for which both genotype and heterogeneous phenotypic information are available. One of the advantages of PBCs is that, in addition to clinical diagnoses and various proxies for diagnoses ("minimal phenotyping"), many of them also provide non-clinical phenotypes, including putative endophenotypes, that can be used to study domains of normal function in addition to, or instead of, clinical diagnoses. By studying endophenotypes it is possible to both dissect psychiatric disorders ("splitting") and to combine multiple phenotypes ("clumping"), which can either reinforce or challenge traditional diagnostic categories. Such endophenotypes may also permit a deeper exploration of the neurobiology of psychiatric disorders. A coordinated effort to fully exploit the potential of endophenotypes is overdue.

For decades, the field of psychiatric genetics largely failed to identify replicable associations. A watershed moment occurred several years ago, with the first successful genome-wide association study (**GWAS**) for schizophrenia[1]. Since then, numerous GWAS for other psychiatric disorders have enjoyed similar success[2]. It has now become clear that very large sample sizes are needed because common psychiatric diseases are highly polygenic, perhaps even omnigenic[3]. Whereas the initial focus was on identifying individual genes and loci[4], larger samples and polygenic methodologies have emphasized the importance and utility of sub-genomewide significant signals. In addition, recent evidence has confirmed an important role for rare and even de novo variants[5–8]. Well-powered case control studies have been essential to understand the genetics of psychiatric disorders, and ascertaining more cases and controls will certainly yield more genome-wide significant associations. *But is that all that we should be doing?*

In recent psychiatric genetics studies, the largest and most rapid growth in sample size has not come from ramping up ascertainment of cases, but rather from utilizing increasingly abundant population-based cohorts (**PBCs**), such as UK Biobank (**UKB**), Million Veterans

---

*****Co-corresponding Authors**: Sandra Sanchez-Roige, sanchezroige@health.ucsd.edu, Abraham A. Palmer, aap@ucsd.edu.

Project (**MVP**), and cohorts from genetics-focused companies such as deCode Genetics and 23andMe (Box 1). In some cases, PBCs have simply provided additional cases and controls. For example, the first successful GWAS for schizophrenia included cases and controls from deCode Genetics[1]. Much more recently, MVP has provided by far the largest cohorts for diseases such as alcohol use disorder (**AUD**)[9], opiate use disorder (**OUD**)[10], and post-traumatic stress disorder (**PTSD**)[11] by using electronic health records as a source of phenotypic information[12,13]. Moreover, self-reported clinical diagnosis collected by 23andMe (e.g. "Have you ever been diagnosed with clinical depression?") provided the majority of the data for a recent GWAS for MDD[14]. Self-reported case status was also used for replication in a recent GWAS of ADHD[15]. Finally, both UKB and 23andMe have been rich sources of non-disease phenotypes, such as neuroticism[16], insomnia[17] and risk tolerance[18], which are continuously distributed and can therefore be measured in the general population but are still relevant to multiple psychiatric disorders (Figure 1). In this perspective, we will explore how PBCs are changing ascertainment and phenotyping strategies in ways that create new challenges but may also provide opportunities for a deeper understanding of psychiatric disorders.

Advantages and caveats of population-based cohorts

Whereas genetic studies have traditionally ascertained cases for a particular disorder, PBCs may contain individuals who can serve as cases (and controls) for numerous different disorders. However, several limitations need to be considered. The ascertainment of PBCs, while not focused on a specific diagnosis, is never random and therefore does not represent the general population[19]. For example, 23andMe and UKB[20] research participants are more highly educated and have higher SES than the general population. In addition, similar to traditionally ascertained genetic cohorts, current PBCs are overwhelmingly made up of individuals of European ancestry; although MVP is a notable exception[21]. Another limitation of PBCs is that certain disorders are underrepresented; for example, in UKB, the frequency of schizophrenia (524 research participants with ICD codes for schizophrenia out of 410,293) is lower than the general population[22], perhaps reflecting the lower rate at which schizophrenia patients volunteered to participate in such a rigorous study. The age of subjects in PBCs is another potential limitation. For example, the use of diagnoses for childhood onset disorders like ADHD and autism have changed dramatically over the past few decades, meaning that older subjects will have a lower than expected prevalence of these diagnoses. In addition, the prevalence of environmental exposures (e.g. smoking), which modulate the prevalence of many traits and diseases, have changed over time, which may confound various genetic studies. Lastly, privacy and intellectual property concerns restrict the sharing of raw data and even the results obtained from some PBCs, these restrictions impede data sharing. Despite these limitations, PBCs are attractive because they are economical, offer the potential to dramatically increase sample size, provide a much greater diversity of phenotypes, and lend themselves to innovative study designs.

In some PBCs, clinical diagnoses are not available. However, self-reported clinical diagnoses may be available. For obvious reasons, these self-reported diagnoses must be interpreted with caution; however, the strength of the genetic correlation between gold-standard diagnoses and self-reported diagnoses helps to address this concern. For example,

self-reported MDD and clinician assigned MDD showed a robust genetic correlation (rg = 0.86)[14]. In other cases, self-reported diagnoses are unavailable, but screening tools can be used to approximate diagnoses. For example, scores from the Alcohol Use Disorder Identification Test (**AUDIT**[23]), which is as a screening tool for AUD, were available in research participants from 23andMe and UKB. Sanchez-Roige et al[24] found that when AUDIT scores were converted into a case control phenotype, they were highly genetically correlated with AUD (rg = 0.82)[25]. These examples demonstrate that, even when clinical diagnoses are not available, there is still significant value in using self-reported information from PBCs for genetic studies of psychiatric disorders.

Minimal phenotyping and endophenotypes for refining psychiatric genetics

In general, there is a tradeoff between phenotyping depth and sample size (Figure 2). The quest for larger sample sizes has led to the adoption of "minimal phenotyping" where a complex disease or trait may be reduced to a single yes or no question. Minimal phenotyping is sometimes criticized because it implicitly assumes that minimal phenotypes are merely noisy measurements of a true underlying phenotype[26]. Cai et al[26] sought to empirically examine this question by considering both self-reported diagnosis of MDD and clinician measurements of the cardinal symptoms of MDD and found that minimal phenotyping yielded a qualitatively different trait. Another empirical examination of minimal phenotyping used a multivariate framework (Genomic SEM[27]) to evaluate several psychiatric disorders and self-report measures of their cardinal symptoms[28]. That study identified large genetic correlations between some disorders and symptom pairs (e.g. MDD, depressive symptoms), but very modest genetic correlations between other pairs (e.g. bipolar disorder and manic symptoms; schizophrenia and psychotic symptoms). Despite these limitations, robust genetic signals -- of something -- can be obtained using minimal phenotyping; how useful these signals will be for understanding the pathophysiology of psychiatric disorders is a matter of ongoing debate, but when large, minimally phenotyped datasets exist, it seems natural that they should be analyzed.

Regardless of whether diagnoses are made by an expert clinician, a structured interview, or self-report, there is a broader question about whether or not the current diagnostic categories are optimal for genetic research, given that the DSM was never intended to be a research tool. A recent review summarized this issue with the memorable phrase "*our genes don't seem to have read the DSM*"[29]. Initiatives such as the National Institute of Mental Health (**NIMH**) Research Domain Criteria (**RDoC**)[30] and Hierarchical Taxonomy of Psychopathology (**HiTOP**[31]) provide new ways of classifying psychiatric disorders based on dimensions of observable behavioral and neurobiological measures, rather than diagnostic categories. These approaches have not been universally accepted[32]. Even before RDoC, there was widespread enthusiasm for genetic studies of endophenotypes (Box 2); however, studies of endophenotypes flourished in the era of candidate genes, when the necessity of large sample sizes was not generally understood. This may have fostered undue skepticism about the utility of endophenotypes for genetic research.

There are several recent examples of adequately powered *genome-wide* (rather than candidate gene) association studies of endophenotypes. For example, impulsivity, which has

been defined as *"actions which are poorly conceived, prematurely expressed, unduly risky or inappropriate to the situation, and that often result is undesirable consequences"*[33] appears to meet the criteria for an endophenotype for multiple psychiatric disorders, including attention-deficit/hyperactivity disorder (**ADHD**) and several substance use disorders (**SUD**). Numerous genetic studies have now shown that various measures of impulsivity[34–36] and sensation seeking[39] are heritable and that they are genetically correlated with both ADHD and various substance use related traits[34,35]. In addition, risk tolerance (*"would you describe yourself as someone who takes risks?"*), which has also been proposed as an endophenotype for both ADHD and substance use disorders, was recently measured in over one million individuals (primarily from UKB and 23andMe[18]). Although risk tolerance was measured using a minimal phenotype (a single vaguely worded question), risk tolerance was clearly heritable and the large sample size allowed identification of 124 genome-wide significant loci[18]. Some of these loci have also been implicated in clinically defined traits[25]. Furthermore, risk tolerance was positively genetically correlated ($rg > 0.3$) with numerous clinically relevant traits (e.g. ADHD, SUD). This study illustrates the power of minimal phenotyping to capture an endophenotype that informs complex disorders and also conforms to the RDoC framework. In a third example, Ibrahim-Verbaas et al[37] performed a GWAS for executive function, which can be considered an endophenotype for multiple psychiatric traits. Intriguingly, GWAS of sensation seeking[35], risk tolerance[18] and executive function[37] all identified a locus that included the gene *CAMD2*, which was subsequently associated with AUD[9]. Whether all of these associations are due to a single locus or multiple loci is far from clear[38], but the index SNPs for these studies are typically co-inherited (LD is ~0.9), consistent with a single causal locus.

Another example of an intriguing endophenotype is self-reported loneliness (e.g. "Do you often feel lonely?"), which is a strong predictor of mortality and life satisfaction and appears to precede the onset of MDD[39]. Several recent GWAS of loneliness[40–42] have identified several significant loci and shown that a genetic predisposition to loneliness is genetically correlated with psychiatric, cardiovascular, and metabolic disorders. By assigning polygenic risk scores to individuals for whom electronic medical records were also available, Dennis et al[43] showed that genetic liability for loneliness increased the risk to develop coronary artery disease more robustly than MDD. Thus, loneliness is an endophenotype that is relevant to both MDD and a variety of somatic disorders.

While some endophenotypes may be amenable to minimal phenotyping, others represent extremely deep and rich data types. For example, by passively collecting data from wearable devices and smartphones, certain endophenotypes relevant to psychiatric disorders can be measured[44]. In a recent GWAS of circadian rhythm, wearable devices were used to gather objective measures of sleep timing, duration and quality[45]. More recently, structural connectivity from fMRI was proposed as endophenotype for IQ[46]. Elliott et al[47] used 3,144 functional and structural brain imaging phenotypes from UKB to conduct GWAS that identified novel associations that included genes relevant to brain development, pathway signaling and plasticity.

A path forward for researching endophenotypes

Despite compelling examples like these, there has not been a coordinated effort to define and explore the endophenotype space. Whereas psychiatric disorders require ascertainment of cases and controls, endophenotypes are continuous and could therefore be measured at scale in PBCs (Figure 1). The Psychiatric Genomics Consortium (**PGC**) has subdivided psychiatric genetic studies into working groups for each major diagnostic category; in contrast, while individual groups have been formed around specific projects (e.g. the Social Science Genetic Association Consortium, https://www.thessgac.org; the Cognitive Genomics Consortium), *there is no coordinated effort to establish a similar set of working groups focused on GWAS of endophenotypes or RDoC traits; however, we feel such an effort is overdue.*

The approach we are proposing will be orthogonal to the efforts of the PGC because RDoC traits and endophenotypes "split" diagnostic categories into discrete units of analysis. The SUD field provides a good example of how a complex disorder can be split into smaller, more biologically meaningful units. SUD develop in accordance with an obligate longitudinal pattern: drug experimentation → regular use → harmful use → transition to compulsive use → quit attempts → relapse (Figure 3a). Approaching SUD with a case control framework merges the genetic liability for each of these stages into a single phenotype, obscuring the distinct biological factors relevant at each stage. In contrast, several recent projects have focused on individual stages of SUD, which can help to address this limitation. For example, GSCAN used data from almost 1 million individuals to examine a number of SUD-related traits, including smoking initiation[48]. In another example, the genetic relationship between alcohol consumption and AUD was explored using the AUDIT, a 10-item questionnaire that measures alcohol use and misuse[24]. By dissecting the genetic contribution for alcohol consumption (first 3 items) vs problematic use (final 7 items), Sanchez-Roige et al[24] and Kranzler et al[9] showed a surprisingly low correlation between alcohol consumption and AUD (rg=0.33 and 0.52, respectively); however, the correlation between problematic alcohol use and AUD was stronger (rg=0.63)[24].

Even when the temporal stages of a psychiatric disorder cannot be so clearly delineated, it can be helpful to split diagnoses into endophenotypes that are associated with the disease of interest. For example, a recent GWAS of insomnia[17], which is a core symptom of multiple psychiatric disorders and a DSM criterion for MDD, identified 202 loci and showed strong genetic correlations with MDD (rg=0.5) and several other psychiatric conditions. Similarly, neuroticism, which shares a common genetic basis with MDD but can be more easily measured, could serve as a clinical stratifying factor for antidepressant actions[16]. However, it can be difficult to determine what level of dissection is required; a recent study suggested that neuroticism reflected two genetic dimensions, one capturing depressed affect, and another capturing worry[49]. Another example comes from several GWAS of impulsive personality[35], which has been proposed as an endophenotype for several psychiatric disorders including ADHD. The UPPS-P is a self-reported questionnaire that measures 5 different aspects of impulsive personality. Only two of those five were significantly associated with ADHD; in contrast, all three subscales of BIS-11, which is another impulsive personality questionnaire, were significantly associated with ADHD[35].

These examples illustrate how disease phenotypes can be dissected into component parts. Nonetheless, despite the original claim that endophenotypes would have a simpler genetic architecture[50], all studies conducted to date have shown that both disease diagnoses and endophenotypes are highly polygenic.

Once the traits that reflect domains of normal function have been measured ("split") in genotyped cohorts, it becomes possible to explore their empirical relationships with one another ("clumping") beyond those that are already defined by traditional psychiatric nosology (see Figure 3b). Genomic SEM[27] and related techniques are now being used in a number of such efforts. Luningham et al[51] used genomic SEM to test multiple models of psychopathology among fourteen psychiatric disorders and related traits. They identified three factors (namely Externalizing, Internalizing, and Thought Problems), and an uncorrelated Neurodevelopmental Disorders factor. These factors showed distinct patterns of genetic correlations and accounted for substantial genetic variance. These empirically identified clusters may provide better targets for GWAS than individual disorders. In another example, Baselmans et al[52] showed that it was possible to increase power by using Genomic SEM to integrate multiple traits (life satisfaction, positive affect, neuroticism, depressive symptoms) into a measure of "well-being spectrum". By aggregating data from different sources of correlated traits, they reached a sample size of over 2.3 million individuals, which allowed them to identify 304 independent signals associated with well-being; a similar analysis suggested a two factor model that distinguishes "lower end" and "higher end" well-being factors[53]. In a third example, Thorp et al[54] used Genomic SEM to identify two factors, which they referred to as "psychological" and "somatic" from the 9-item Patient Health Questionnaire (**PHQ-9**). Recently, several related methods have been developed (e.g. reverse GWAS, **RGWAS**[55] and BUHMBOX[56]). Using RGWAS, Dahl et al[55] proposed a stress subtype in MDD, and identified three novel subtypes of metabolic traits. Using BUHMBOX (Breaking Up Heterogeneous Mixture Based On Cross-locus correlations), Han et al[56] found that seropositive and seronegative rheumatoid arthritis could be subdivided to form a new subgroup within seronegative-like cases. Conversely, they identified a genetic correlation between MDD and SCZ, but there was no evidence that this correlation was due to subgroup heterogeneity.

Clumping has been used to test the hypothesis, originally suggested by twin studies, that psychiatric disorders share a single common genetic factor (the "p-factor")[57]. One of the earliest studies to use GWAS data to test this hypothesis showed that SNPs associated with schizophrenia were also associated with bipolar disorder[58]. Specific genes have been identified that confer risk for multiple psychiatric disorders (e.g. *CACNA1C*[59–61]). Evidence that the risk for substance abuse is shared across multiple substances (e.g. alcohol, tobacco[48]) is also consistent with earlier results from twin studies showing both substance-specific and substance-independent genetic risk. An example of this genetic overlap is the gene *CADM2*, which has been associated several substances (alcohol use[24,62], tobacco and cannabis initiation[63]) and risky behavior[18,35]. Joint analysis of correlated traits may outperform that of single phenotypes and allows the possibility to disentangle genetic effects that are specific to each trait from those that capture a latent construct (Figure 3b).

Clumping can also lead to new splits. For example, Bansal et al[64] used GWAS results from two correlated traits: schizophrenia (a disorder) and educational attainment (a continuously distributed non-disorder trait) to propose two distinct etiologies of schizophrenia, one that resembled bipolar disorder and was characterized by high education, and another that reflected a cognitive disorder and was independent of education. Studies like this one provide greater flexibility to explore the phenotypic space, which can lead to novel insights and challenge established nosologies.

The utility of endophenotypes for translation to cellular and animal models

Throughout this perspective, we have alluded to GWAS producing novel biological insights; however GWAS have numerous limitations[65] and do not themselves produce actionable new knowledge. The influence of locus on a phenotype may be due to a coding difference or a regulatory difference. The former can be directly identified from sequence data (although the interpretation of sequence variants is still challenging), whereas regulatory polymorphisms are typically identified by using complementary data from GTEx[66], PsychENCODE (resource.psychencode.org), CommonMind (https://www.synapse.org/#!Synapse:syn2759792/tables/) and Brainspan (http://www.brainspan.org/); or protein QTLs (**pQTLs**; e.g. [67]). Once identified, the protein products of such genes can be intensively studies and evaluated as possible drug targets[68]. Another way to follow up on GWAS results is to use cellular and animal models. However, these approaches have been challenging because psychiatric diseases cannot be recapitulated in cells or non-human animals. On the contrary, certain endophenotypes can be more readily modeled in animals, which provides an opportunity to evaluate the role of genes identified by GWAS at the molecular, cellular and circuit level[56]. Individual genes can also be manipulated in animal models by using viral vectors, genetically engineered null alleles (knock outs), over expression alleles, conditional alleles or knockins of humanized alleles. These approaches provide a gene-by-gene approach to translation -- what is still lacking are robust methods for examining the polygenic nature of complex traits in animal models. In contrast, methods for using human-derived cellular models to examine the polygenic signals obtained from GWAS are better established. The ability to more directly model endophenotypes in experimental model systems will be critical in extracting biological insights from GWAS and thus realizing the full potential of psychiatric genetic studies.

Conclusion

Over the last 10 years, GWAS for psychiatric disorders have turned a corner and begun to identify numerous significant loci for all major psychiatric disorders. It is generally understood that larger samples will extend on these successes. In this perspective, we have considered additional study designs that go beyond disease diagnoses. Although they are not without limitations, PBC are quickly becoming the predominant ascertainment strategy. Direct-to-consumer genetic companies, which collectively account for millions of research participants (https://thednageek.com/dna-tests/), are the largest PBCs. Publicly funded PBC already account for millions of research participants and are linked to electronic health records and other rich data types (e.g. questionnaire data, imaging data, epigenetics). Unlike previous ascertainment strategies, PBC have provided adequate sample sizes for GWAS

of endophenotypes. This has allowed for increasingly sophisticated study designs. These resources are already leading to neurobiological insights about the molecular, cellular and circuit underpinnings associated with psychiatric disorders that will facilitate the translation of psychiatric genetic insights to other fields of neuroscience.

Acknowledgements and Disclosures

References

1. Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. Genome-wide association study identifies five new schizophrenia loci. Nat. Genet. 43, 969–976 (2011). [PubMed: 21926974]

2. Visscher PM et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am. J. Hum. Genet. 101, 5–22 (2017). [PubMed: 28686856]

3. Boyle EA, Li YI & Pritchard JK An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell 169, 1177–1186 (2017). [PubMed: 28622505]

4. Manolio TA et al. Finding the missing heritability of complex diseases. Nature 461, 747–753 (2009). [PubMed: 19812666]

5. Wainschtein P et al. Recovery of trait heritability from whole genome sequence data. 10.1101/588020 (2019) doi:10.1101/588020.

6. Sebat J et al. Strong association of de novo copy number mutations with autism. Science 316, 445–449 (2007). [PubMed: 17363630]

7. Iossifov I et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature 515, 216–221 (2014). [PubMed: 25363768]

8. Brazel DM et al. Exome Chip Meta-analysis Fine Maps Causal Variants and Elucidates the Genetic Architecture of Rare Coding Variants in Smoking and Alcohol Use. Biol. Psychiatry (2018) doi:10.1016/j.biopsych.2018.11.024.

9. Kranzler HR et al. Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. Nat. Commun. 10, 1499 (2019). [PubMed: 30940813]

10. Zhou H et al. GWAS including 82,707 subjects identifies functional coding variant in OPRM1 gene associated with opioid use disorder. 10.1101/19007039 (2019) doi:10.1101/19007039.

11. Gelernter J et al. Genome-wide association study of post-traumatic stress disorder reexperiencing symptoms in >165,000 US veterans. Nat. Neurosci. 22, 1394–1401 (2019). [PubMed: 31358989]

12. Sanchez-Roige S & Palmer AA Electronic Health Records Are the Next Frontier for the Genetics of Substance Use Disorders. Trends Genet. 35, 317–318 (2019). [PubMed: 30797598]

13. Smoller JW The use of electronic health records for psychiatric phenotyping and genomics. Am. J. Med. Genet. Part B Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet. 177, 601–612 (2018).

14. Howard DM et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat. Neurosci. 22, 343–352 (2019). [PubMed: 30718901]

15. Demontis D et al. Discovery of the first genome-wide significant risk loci for attention deficit/ hyperactivity disorder. Nat. Genet. 51, 63–75 (2019). [PubMed: 30478444]

16. Luciano M et al. Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. Nat. Genet. 50, 6–11 (2018). [PubMed: 29255261]

17. Jansen PR et al. Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. Nat. Genet. 51, 394–403 (2019). [PubMed: 30804565]

18. Karlsson Linnér R et al. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. Nat. Genet. 51, 245–257 (2019). [PubMed: 30643258]

19. Keyes KM & Westreich D UK Biobank, big data, and the consequences of non-representativeness. Lancet Lond. Engl. 393, 1297 (2019).

20. Adams M et al. Factors associated with sharing email information and mental health survey participation in two large population cohorts. 10.1101/471433 (2018) doi:10.1101/471433.

21. Gaziano JM et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. J. Clin. Epidemiol. 70, 214–223 (2016). [PubMed: 26441289]

22. Moreno-Küstner B, Martín C & Pastor L Prevalence of psychotic disorders and its association with methodological issues. A systematic review and meta-analyses. PloS One 13, e0195687 (2018).

23. Saunders JB, Aasland OG, Babor TF, de la Fuente JR & Grant M Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II. Addict. Abingdon Engl. 88, 791–804 (1993).

24. Sanchez-Roige S et al. Genome-Wide Association Study Meta-Analysis of the Alcohol Use Disorders Identification Test (AUDIT) in Two Population-Based Cohorts. Am. J. Psychiatry 176, 107–118 (2019). [PubMed: 30336701]

25. Walters RK et al. Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. Nat. Neurosci. 21, 1656–1669 (2018). [PubMed: 30482948]

26. Cai N et al. Minimal phenotyping yields GWAS hits of low specificity for major depression. 10.1101/440735 (2018) doi:10.1101/440735.

27. Grotzinger AD et al. Genomic SEM Provides Insights into the Multivariate Genetic Architecture of Complex Traits. 10.1101/305029 (2018) doi:10.1101/305029.

28. Mallard TT et al. Not just one p: Multivariate GWAS of psychiatric disorders and their cardinal symptoms reveal two dimensions of cross-cutting genetic liabilities. bioRxiv (2019).

29. Smoller JW et al. Psychiatric genetics and the structure of psychopathology. Mol. Psychiatry 24, 409–420 (2019). [PubMed: 29317742]

30. Insel T et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. Am. J. Psychiatry 167, 748–751 (2010). [PubMed: 20595427]

31. Kotov R, Krueger RF & Watson D A paradigm shift in psychiatric classification: the Hierarchical Taxonomy Of Psychopathology (HiTOP). World Psychiatry Off. J. World Psychiatr. Assoc. WPA 17, 24–25 (2018).

32. Weinberger DR, Glick ID & Klein DF Whither Research Domain Criteria (RDoC)?: The Good, the Bad, and the Ugly. JAMA Psychiatry 72, 1161–1162 (2015). [PubMed: 26558844]

33. Daruna JH & Barnes PA A neurodevelopmental view of impulsivity. in The impulsive client: Theory, research, and treatment. (eds. McCown WG, Johnson JL & Shure MB) 23–37 (American Psychological Association, 1993). doi:10.1037/10500-002.

34. Sanchez-Roige S et al. Genome-wide association study of delay discounting in 23,217 adult research participants of European ancestry. Nat. Neurosci. 21, 16–18 (2018). [PubMed: 29230059]

35. Sanchez-Roige S et al. Genome-Wide Association Studies of Impulsive Personality Traits (BIS-11 and UPPS-P) and Drug Experimentation in up to 22,861 Adult Research Participants Identify Loci in the CACNA1I and CADM2 genes. J. Neurosci. Off. J. Soc. Neurosci. 39, 2562–2572 (2019).

36. Anokhin AP, Grant JD, Mulligan RC & Heath AC The genetics of impulsivity: evidence for the heritability of delay discounting. Biol. Psychiatry 77, 887–894 (2015). [PubMed: 25555481]

37. Ibrahim-Verbaas CA et al. GWAS for executive function and processing speed suggests involvement of the CADM2 gene. Mol. Psychiatry 21, 189–197 (2016). [PubMed: 25869804]

38. Morris J et al. Genetic variation in CADM2 as a link between psychological traits and obesity. Sci. Rep. 9, 7339 (2019). [PubMed: 31089183]

39. Holt-Lunstad J, Smith TB, Baker M, Harris T & Stephenson D Loneliness and social isolation as risk factors for mortality: a meta-analytic review. Perspect. Psychol. Sci. J. Assoc. Psychol. Sci. 10, 227–237 (2015).

40. Gao J et al. Genome-Wide Association Study of Loneliness Demonstrates a Role for Common Variation. Neuropsychopharmacol. Off. Publ. Am. Coll. Neuropsychopharmacol. 42, 811–821 (2017).

41. Day FR, Ong KK & Perry JRB Elucidating the genetic basis of social interaction and isolation. Nat. Commun. 9, 2457 (2018). [PubMed: 29970889]

42. Abdellaoui A et al. Phenome-wide investigation of health outcomes associated with genetic predisposition to loneliness. Hum. Mol. Genet. (2019) doi:10.1093/hmg/ddz219.

43. Dennis J et al. Genetic risk for major depressive disorder and loneliness in gender-specific associations with coronary artery disease: supplementary. 10.1101/512541 (2019) doi:10.1101/512541.

44. Freimer NB & Mohr DC Integrating behavioural health tracking in human genetics research. Nat. Rev. Genet. 20, 129–130 (2019). [PubMed: 30523274]

45. Jones SE et al. Genome-wide association analyses of chronotype in 697,828 individuals provides insights into circadian rhythms. Nat. Commun. 10, 343 (2019). [PubMed: 30696823]

46. Hatoum AS, Reineberg AE, Kragel PA, Wager TD & Friedman NP Inferring the Genetic Influences on Psychological Traits Using MRI Connectivity Predictive Models: Demonstration with Cognition. 10.1101/777821 (2019) doi:10.1101/777821.

47. Elliott LT et al. Genome-wide association studies of brain imaging phenotypes in UK Biobank. Nature 562, 210–216 (2018). [PubMed: 30305740]

48. Liu M et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nat. Genet. 51, 237–244 (2019). [PubMed: 30643251]

49. Nagel M et al. Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. Nat. Genet. 50, 920–927 (2018). [PubMed: 29942085]

50. Gottesman II & Gould TD The endophenotype concept in psychiatry: etymology and strategic intentions. Am. J. Psychiatry 160, 636–645 (2003). [PubMed: 12668349]

51. Luningham JM, Poore HE, Yang J & Waldman ID Testing Structural Models of Psychopathology at the Genomic Level. (2018) doi:10.1101/502039.

52. Baselmans BML et al. Multivariate genome-wide analyses of the well-being spectrum. Nat. Genet. 51, 445–451 (2019). [PubMed: 30643256]

53. Baselmans BML et al. A Genetic Investigation of the Well-Being Spectrum. Behav. Genet. 49, 286–297 (2019). [PubMed: 30810878]

54. Thorp JG et al. Genetic heterogeneity in self-reported depressive symptoms identified through genetic analyses of the PHQ-9. Psychol. Med. 1–12 (2019) doi:10.1017/S0033291719002526.

55. Dahl A et al. Reverse GWAS: Using genetics to identify and model phenotypic subtypes. PLoS Genet. 15, e1008009 (2019).

56. Han B et al. A method to decipher pleiotropy by detecting underlying heterogeneity driven by hidden subgroups applied to autoimmune and neuropsychiatric diseases. Nat. Genet. 48, 803 (2016). [PubMed: 27182969]

57. Selzam S, Coleman JRI, Caspi A, Moffitt TE & Plomin R A polygenic p factor for major psychiatric disorders. Transl. Psychiatry 8, 205 (2018). [PubMed: 30279410]

58. International Schizophrenia Consortium et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460, 748–752 (2009). [PubMed: 19571811]

59. Ferreira MAR et al. Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. Nat. Genet. 40, 1056–1058 (2008). [PubMed: 18711365]

60. Liu Y et al. Meta-analysis of genome-wide association data of bipolar disorder and major depressive disorder. Mol. Psychiatry 16, 2–4 (2011). [PubMed: 20351715]

61. Ripke S et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat. Genet. 45, 1150–1159 (2013). [PubMed: 23974872]

62. Clarke T-K et al. Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N=112 117). Mol. Psychiatry 22, 1376–1384 (2017). [PubMed: 28937693]

63. Pasman JA et al. GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. Nat. Neurosci. 21, 1161–1170 (2018). [PubMed: 30150663]

64. Bansal V et al. Genome-wide association study results for educational attainment aid in identifying genetic heterogeneity of schizophrenia. Nat. Commun. 9, 3078 (2018). [PubMed: 30082721]

65. Tam V et al. Benefits and limitations of genome-wide association studies. Nat. Rev. Genet. 20, 467–484 (2019). [PubMed: 31068683]

66. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348, 648–660 (2015). [PubMed: 25954001]

67. Yao C et al. Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease. Nat. Commun. 9, 3268 (2018). [PubMed: 30111768]

68. Zheng J et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. 10.1101/627398 (2019) doi:10.1101/627398.

69. Bycroft C et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 562, 203–209 (2018). [PubMed: 30305743]

70. Sudlow C et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLOS Med. 12, e1001779 (2015).

71. Khan R & Mittelman D Consumer genomics will change your life, whether you get tested or not. Genome Biol. 19, 120 (2018). [PubMed: 30124172]

Box 1.

**Some of the major population-based cohorts in psychiatric genetics**

- **UK Biobank (UKB)**: prospective cohort with rich phenotypic information including biological, lifestyle, biomarkers, and imaging data and genetic information from approximately 500,000 volunteer research participants across the United Kingdom. Participants are predominantly of European ancestry, middle age and older[69], and with higher socioeconomic backgrounds than the general population[70].

- **Million Veterans Project (MVP)**: observational cohort study in the Department of Veterans Affairs (**VA**) health care system containing deep phenotyping, including the VA electronic health records, and genotypes[21]. At the time of this writing more than 450,000 individuals have already been genotyped. Most participants are male, with a mean age of 64. While the majority of the participants are of European ancestry, this cohort also contains significant numbers of individuals from other ancestral groups.

- **deCODE Genetics, Inc.**: biopharmaceutical company based in Reykjavík, Iceland, funded in 1996 to study genetic risk factors for several diseases (https://www.decode.com). deCODE contains genotypic and medical data from >160,000 volunteer participants, comprising about half of the adult population in Iceland.

- **23andMe, Inc**: direct-to-consumer genetic company with over 5 million genotyped individuals[71]; research participants tend to have higher education levels and socioeconomic status than the general population (e.g. [34]).

Box 2:

criteria for a trait to be considered an endophenotype, as defined by[50]

1. The endophenotype is associated with illness in the population [genetic correlation].

2. The endophenotype is heritable.

3. The endophenotype is primarily state-independent (manifests in an individual whether or not illness is active).

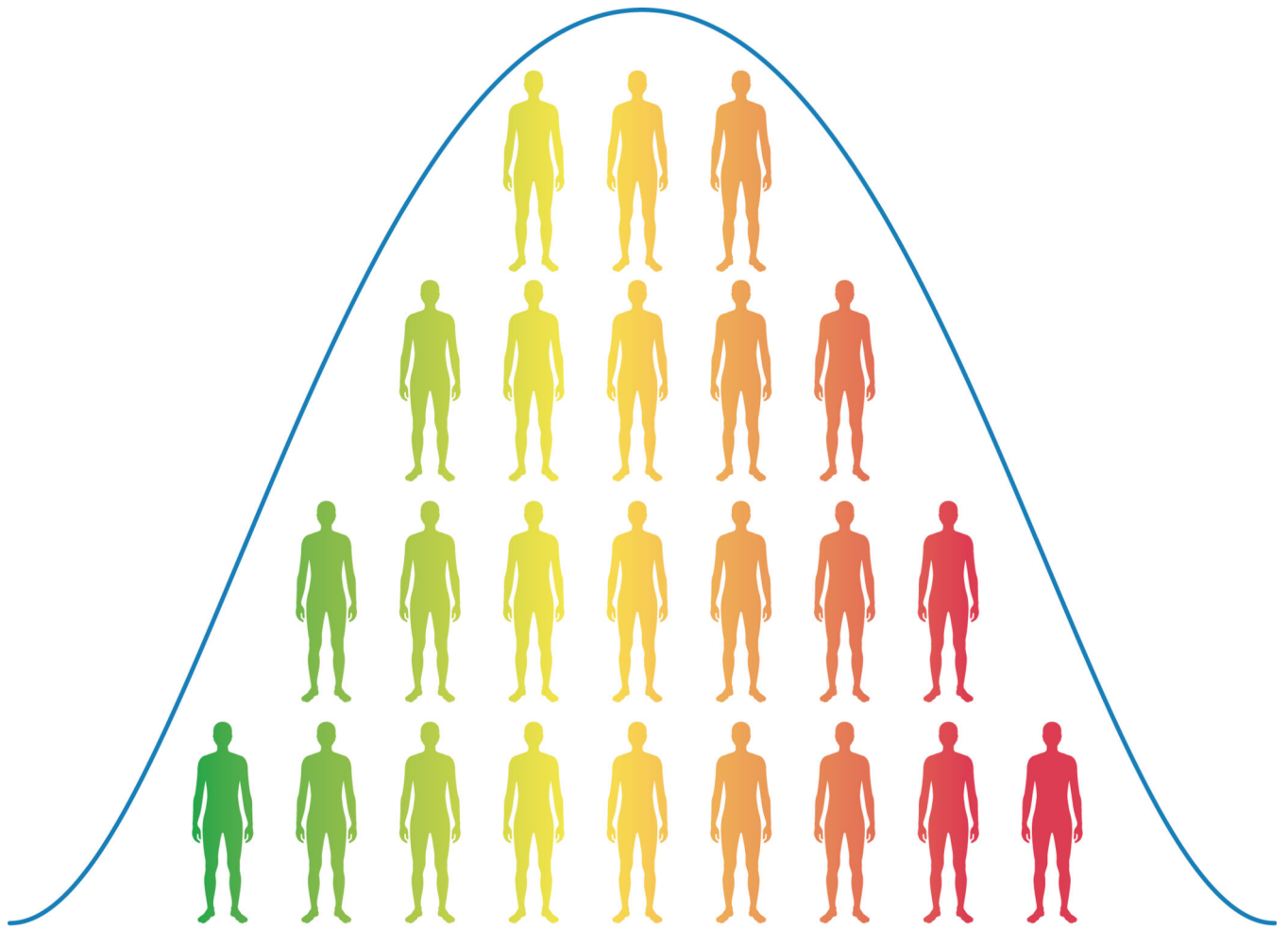4. Within families, endophenotype and illness co-segregate.
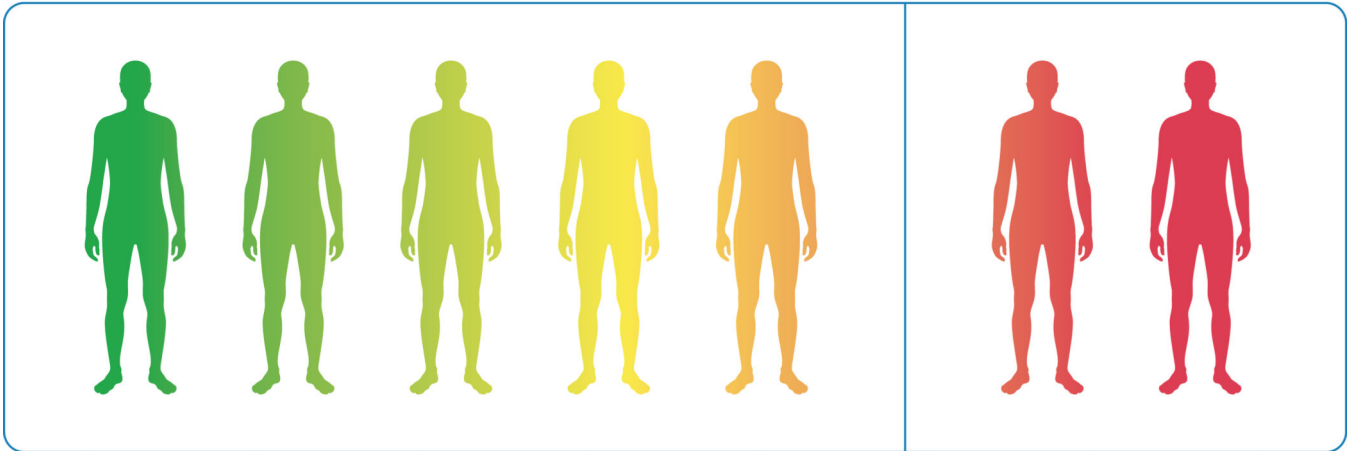
Figure 1. Case control vs continuous phenotypes.
Psychiatric disorders require ascertainment of cases and controls and because of their binary nature they do capture variability within the two classes, a problem that is addressed by a continuous phenotype.
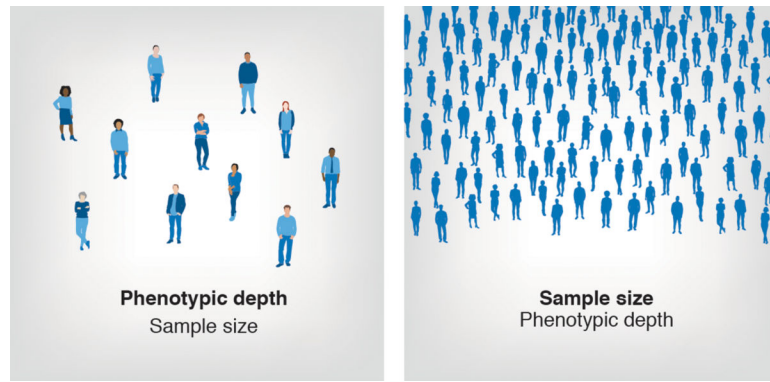
Figure 2. The trade-off between phenotyping depth and sample size.
Deep phenotyping is more expensive and time consuming; therefore, when the available budget is fixed, greater phenotyping depth comes at the expense of sample size. In contrast, scalable phenotyping strategies, which are more commonly used in PBCs, allow for larger sample sizes.
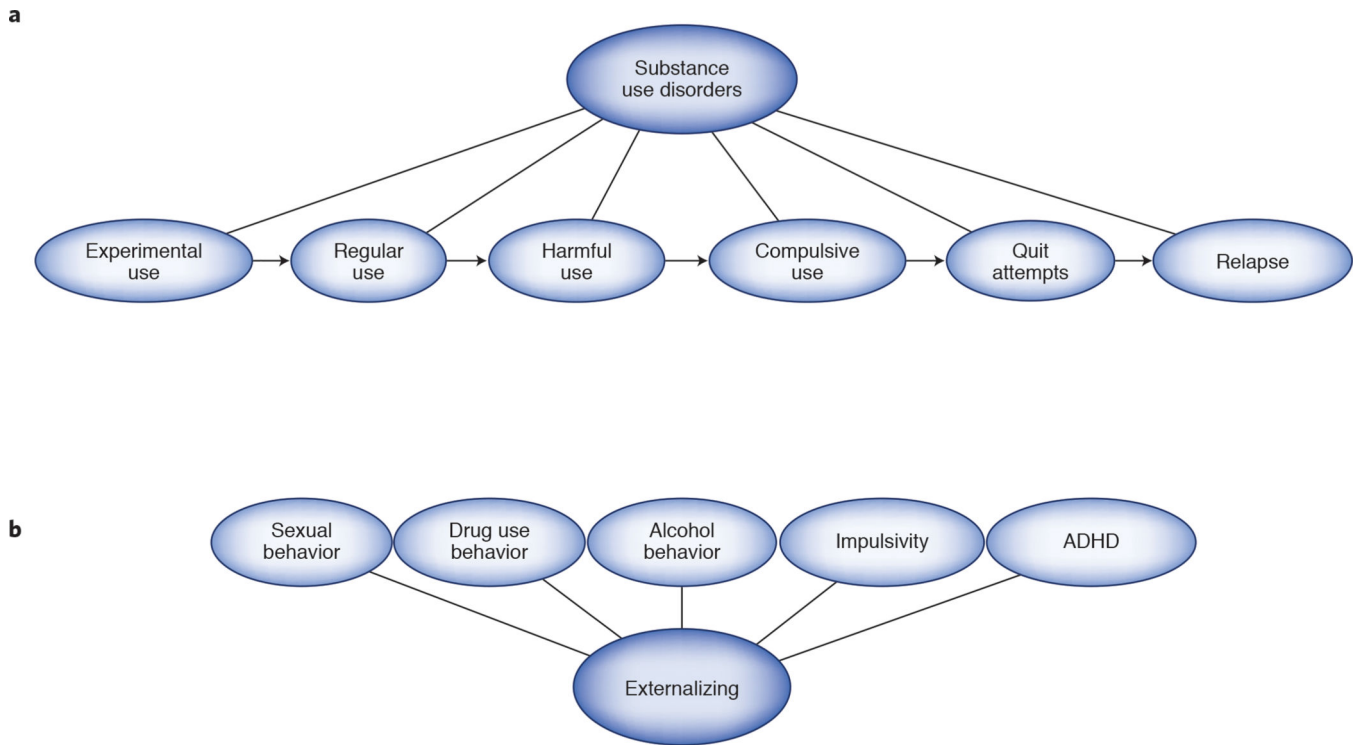
a

b

Figure 3. Splitting vs clumping.
Psychiatric disorders can be further dissected or "split" into discrete units of analysis; for example, SUDs can be split into smaller, more biologically meaningful traits, that manifest into an obligate longitudinal pattern: drug experimentation → regular use → harmful use → transition to compulsive use → quit attempts → relapse. The empirical correlations among these traits can be examined ("clumping") beyond those that are already defined by traditional psychiatric nosology. Recently, The Externalizing Consortium has sought to analyze the genetic correlations between different traits from the externalizing spectrum (https://osf.io/xkv36/), including sexual and drug-related phenotypes, impulsivity, and attention-deficit/hyperactivity disorder, with the goal of identifying loci involved in a shared underlying liability to externalizing versus genes unique to specific phenotypes.