Using genetics to increase specificity of outcome prediction in psychiatric disorders: prospects for progression.

Liability to psychiatric disorders such as schizophrenia, bipolar disorder and unipolar depression is known to have a significant genetic component, which is partially shared between disorders (1). Polygenic risk scores (PRS) are a convenient way of summarising the contribution of disease-associated variants across multiple loci to give individual-specific estimates of risk (2). In this issue of the Journal, Musliner and colleagues (3) show that PRS for psychiatric disorders are associated with progression from unipolar depression to bipolar disorder or psychotic disorder in 16,949 people from the iPSYCH2012 Danish population cohort. Specifically: bipolar disorder PRS was associated with progression to bipolar disorder, and schizophrenia PRS was associated with progression to psychotic disorder. Association of PRS with longitudinal progression is novel and has potential clinical utility in early identification of patients likely to need treatments.

Association of polygenic risk scores with disease outcomes has three main aims: to predict individuals who will develop the disease, to refine the definition of disease phenotypes, and to highlight genetic mechanisms relevant to disease. The accuracy of a predictor of disease risk is defined by the true positive rate (the probability that it correctly identifies individuals with the disease as being affected) and false positive rate (the probability that it incorrectly identifies individuals without the disease as being affected). These quantities can be calculated for varying values of the threshold used to identify affected individuals and plotted against each other to give the receiver operating characteristic (ROC) curve. The probability that a predictor correctly classifies disease status is measured by the area under the ROC curve, the AUC. The higher the AUC, the more accurate the predictor, with a random predictor having an AUC of 0.5. Typically, a predictor requires an AUC of 0.8 to be regarded as clinically useful (4). While this criterion has been reached in Alzheimer's disease (5), in general prediction is much less accurate for psychiatric disorders – for example, PRS typically achieves an AUC of 0.6-0.7 in schizophrenia (6), and is weaker still in other psychiatric disorders(7). The accuracy of genetic predictors is known to be limited by the heritability and prevalence of the phenotype being predicted (8). Therefore, attention has focused increasingly on using genetics to refine phenotype definition, to reduce the clinical heterogeneity typically observed in traditional psychiatric diagnoses and define disease subgroups that map more closely onto the underlying biological mechanisms. For example, schizophrenia PRS can distinguish between schizophrenia, bipolar disorder with mood-incongruent psychotic features, bipolar disorder with mood-congruent psychotic features and bipolar disorder without psychosis (9). SCZ PRS can also distinguish between SCZ and other forms of psychosis (10).

Disease progression can yield genetically informative phenotypes even in small samples (11) and is therefore a promising avenue for further study. There have been several studies associating PRS with disease progression in neurodegenerative disorders, for example: predicting cognitive decline in PD (12), and progression of mild cognitive impairment to AD (13). There have been fewer genetic studies of disease progression in psychiatric disorders. Significant associations have been shown between psychiatric PRS and psychiatric, cognitive, and behavioural phenotypes in childhood and adolescence (14, 15, 16). These studies suggest that PRS may predict progression from adolescent traits to psychiatric disorders in adulthood. However, none of them performed the longitudinal follow-up necessary for confirmation of this hypothesis. Jonas et al. (17) present the results of a 20-year study of a group of first admission patients with psychosis, in which schizophrenia PRS was found to predict increased illness severity, along with worse cognition and also which individuals will progress from mood disorder with psychosis to a schizophrenia spectrum disorder. This study is interesting because it shows how genetic risk can be related to progression of psychosis *longitudinally within patients*, and can be used to postulate a hypothesis for genetic risk initially predicting cognitive deficits and negative symptoms prior to an eventual diagnosis of non-affective

psychosis. However, given the small sample size (n=249) of this study, the results need to be replicated. The study of Musliner et al (3) presented in this issue, by contrast, uses a large sample to examine the development of progression to bipolar or psychotic disorders within individuals with unipolar depression, thereby widening the range of psychiatric phenotypes beyond psychosis. Interestingly, PRS for BD was associated with progression to bipolar disorder, with PRS for SCZ associated with progression to psychotic disorder, but the combination of high PRS-BD and high PRS-SCZ was associated with progression to affective psychosis. This suggests the potential both for refinement of the phenotype and genotype to increase the specificity of the association. However, the PRS account for a relatively small proportion of phenotypic variance, thus limiting their clinical utility to predict disease progression.

Prospects for the use of genetic risk to predict psychiatric phenotypes centre on two main avenues of research: improved genetic measures and improved phenotypic measures (Figure 1).
Since the predictive ability of PRS depends on the power of the GWAS used as the training sample (18), one important way of improving the genetic measures is the collection and analysis of large GWAS of psychiatric disorders. These are currently being co-ordinated by the Psychiatric Genomics Consortium, and biobanks will also become increasingly important. Methods are also being developed for deriving PRS that are more powerful than those obtained from the standard p-value thresholding approach (2), for example: PRS-CS (19) and SBayesR (20). Psychiatric disorders are genetically correlated (1), such that PRS from multiple disorders are often associated with clinical phenotypes. Therefore, methods such as genomic structural equation modelling (gSEM) which partition genetic variation from multiple GWAS into portions corresponding to that shared between disorders and disorder-specific components can increase power and specificity of PRS associations (21). Ultimately, the predictive power of common-variant PRS is limited by the genetic architecture of the disease (8). In that case, consideration of rare variants such as CNVs may be useful in improving prediction, since the penetrance of these is often high (22). Furthermore, CNVs associated with neurodevelopmental and psychiatric disorders are associated with psychiatric, cognitive, and behavioural phenotypes in children (23), suggesting that they may be useful for modelling longitudinal trait progression. There is evidence that an increased number of deleterious rare variants in functionally intolerant genes is associated with reduced IQ in autism cases carrying CNVs associated with neurodevelopmental phenotypes (24), and that schizophrenia PRS acts additively with the 22q11.2 deletion to increase schizophrenia liability (25). These studies motivate the combination of CNVs with other types of variation to model phenotypic outcomes. The specificity of genetic risk measures for phenotypic prediction may be further improved by incorporating information on biological pathways. This approach has been applied to PRS in Alzheimer's disease (26) and may also be informative in psychiatric disorders. The predictive ability of genetic risk measures may also be improved by integrating expression data (27) and functional annotation (28).

It is also crucial to refine definitions of phenotypes to improve the correlation with genetic measures. This can be done simultaneously, for example by a combination of gSEM and factor analysis to provide gene-phenotype associations that cross traditional disease diagnoses (29) and thus provide novel biological and clinical insights. As noted earlier, consideration of longitudinal progression is a promising avenue for deriving informative phenotypes with a genetic basis. Intensively phenotyped cohorts are useful in this regard as they allow multivariate modelling of disease trajectories (11). Biobanks are another promising source of novel phenotypes. Zemedikun et al. (30) showed that certain mental and physical conditions clustered together (multimorbidity) in the UK Biobank, although using a cross-sectional, rather than longitudinal, analysis. Recent evidence from the Danish National Registry population cohort indicates that diagnosis of mental disorders influences the future risk of other medical conditions (31) under a survival analysis (Cox proportional hazards). These studies motivate the construction of clusters of multimorbidity that cut across

traditional diagnostic boundaries, and the next step is to discover the genetic (and other) factors that underlie them, ideally in a longitudinal framework.

To conclude: the use of genetics to predict traditional psychiatric outcomes has produced some useful insights into disease mechanisms but has been limited in terms of risk prediction. To address this issue, it will be necessary to derive novel phenotypes that cut across traditional diagnostic boundaries. This will require the analysis of multivariate phenotypic data of various types, including longitudinal progression measures, alongside environmental risk factors. Likewise, genetic predictors should integrate multiple types of variant (common SNP, rare SNV, CNV) with functional and biological information. Analyses of these multidimensional datasets will be challenging and require the development of novel methodology, with machine learning methods being a promising approach (32). Criteria for assessing risk prediction in multivariate outcomes (33) have been proposed to measure the performance of these methods. Thus, while there is still considerable work do be done, the prospects for genomics to accurately predict psychiatric outcomes, and thus target treatments to patients more precisely (34), are bright.

References

1. The Brainstorm Consortium: Analysis of shared heritability in common disorders of the brain. Science. 2018; 360(6395)
2. International Schizophrenia Consortium: Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009; 460(7256):748-752
3. Musliner KL, Krebs MD, Albinana C et al: Polygenic risk and progression to bipolar or psychotic disorders among individuals diagnosed with unipolar depression in early life. Am J Psychiatry in Advance (doi: 10.1176/appi.ajp.2020.19111195)
4. Schummers L, Himes KP, Bodnar LM et al: Predictor characteristics necessary for building a clinically useful risk prediction model: a simulation study. BMC Med Res Methodol. 2016; 16: 123
5. Escott-Price V, Myers AJ, Huentelman M et al: Polygenic risk score analysis of pathologically confirmed Alzheimer disease. Ann Neurol. 2017; 82(2):311-314
6. Ripke S, Neale BM, Corvin A et al: Biological insights from 108 schizophrenia-associated genetic loci. Nature 2014; 511: 421–427
7. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. Genome Medicine 2020; 12:44
8. Wray NR, Yang J, Goddard ME et al: The genetic interpretation of area under the ROC curve in genomic profiling. PLoS Genet. 2010; 6:e1000864
9. Allardyce J, Leonenko G, Hamshere ML et al: Association Between Schizophrenia-Related Polygenic Liability and the Occurrence and Level of Mood-Incongruent Psychotic Symptoms in Bipolar Disorder. JAMA Psychiatry 2018;75(1):28-35
10. Vassos E, di Forti M, Coleman J et al: An Examination of Polygenic Score Risk Prediction in Individuals With First-Episode Psychosis. Biol Psychiatry. 2017; 81(6):470-477
11. Moss DJH, Pardinas A, Langbehn D et al: Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study. Lancet Neurol. 2017; 16(9):701-711
12. Paul KC, Schulz J, Bronstein JM et al: Association of Polygenic Risk Score With Cognitive Decline and Motor Progression in Parkinson Disease. JAMA Neurol. 2018;75(3):360-366
13. Chaudhury S, Brookes KJ, Patel T et al: Alzheimer's disease polygenic risk score as a predictor of conversion from mild-cognitive impairment. Transl Psychiatry. 2019; 9(1):154
14. Pain O, Dudbridge F, Cardno AG et al: Genome-wide analysis of adolescent psychotic-like experiences shows genetic overlap with psychiatric disorders. Am J Med Genet B Neuropsychiatr Genet 2018; 177(4):416-425

15. Mistry S, Escott-Price V, Di Florio A et al: Investigating associations between genetic risk for bipolar disorder and cognitive functioning in childhood. J Affect Disord 2019;259:112-120

16. Jones HJ, Stergiakouli E, Tansey KE et al: Phenotypic Manifestation of Genetic Risk for Schizophrenia During Adolescence in the General Population. JAMA Psychiatry. 2016;73(3):221-228

17. Jonas KG, Lencz T, Li K et al: Schizophrenia polygenic risk score and 20-year course of illness in psychotic disorders. Transl Psychiatry. 2019; 9(1):300

18. Dudbridge F. Power and predictive accuracy of polygenic risk scores. PLoS Genet. 2013; 9(3):e1003348.

19. Ge T, Chen C-Y, Ni Y et al: Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat Commun. 2019; 10(1):1776

20. Lloyd-Jones LR, Zeng J, Sidorenko J et al: Improved polygenic prediction by Bayesian multiple regression on summary statistics. Nat Commun. 2019; 10(1):5086.

21. Grotzinger AD, Rhemtulla M, de Vlaming R et al: Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. Nature Human Behaviour 2019; 3:513–525

22. Kirov G, Rees E, Walters JTR et al: The penetrance of copy number variations for schizophrenia and developmental delay. Biol Psychiatry 2014; 75(5):378-85

23. Chawner SJRA, Owen MJ, Holmans P et al: Genotype-phenotype associations in children with copy number variants associated with high neuropsychiatric risk in the UK (IMAGINE-ID): a case-control cohort study. Lancet Psychiatry. 2019; 6(6):493-505

24. Pizzo L, Jensen M, Polyak A et al: Rare variants in the genetic background modulate cognitive and developmental phenotypes in individuals carrying disease-associated variants. Genet Med. 2019; 21(4):816-825

25. Cleynen I, Engchuan W, Hestand MS et al. Genetic contributors to risk of schizophrenia in the presence of a 22q11.2 deletion.. Mol Psychiatry. 2020; doi: 10.1038/s41380-020-0654-3. Online ahead of print

26. Ahmad S, Bannister C, van der Lee SJ et al: Disentangling the biological pathways involved in early features of Alzheimer's disease in the Rotterdam Study. Alzheimers Dement. 2018; 14(7):848-857

27. Marigorta UM, Denson LA, Hyams JS et al: Transcriptional risk scores link GWAS to eQTLs and predict complications in Crohn's disease. Nat Genet 2017;49(10):1517-1521

28. Hu Y, Lu Q, Powles R et al: Leveraging functional annotations in genetic risk prediction for human complex diseases. PLoS Comput Biol. 2017; 13(6): e1005589

29. Cross-Disorder Group of the Psychiatric Genomics Consortium: Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. Cell 2019; 179(7):1469-1482

30. Zemedikun DT, Gray LJ, Khunti K et al: Patterns of Multimorbidity in Middle-Aged and Older Adults: An Analysis of the UK Biobank Data. Mayo Clin Proc. 2018 Jul;93(7):857-866

31. Momen NC, Plana-Ripoll O, Agerbo E et al: Association between Mental Disorders and Subsequent Medical Conditions. N Engl J Med. 2020; 382(18):1721-1731

32. Bracher-Smith M, Crawford K, Escott-Price V. Machine learning for genetic prediction of psychiatric disorders: a systematic review. Mol Psychiatry 2020; doi: 10.1038/s41380-020-0825-2. Online ahead of print.

33. Dudbridge F. Criteria for evaluating risk prediction of multiple outcomes. Stat Methods Med Res. 2020; 962280220929039. doi: 10.1177/0962280220929039. Online ahead of print.

34. Rees E, Owen MJ. Translating insights from neuropsychiatric genetics and genomics for precision psychiatry. Genome Med. 2020; 12(1):43
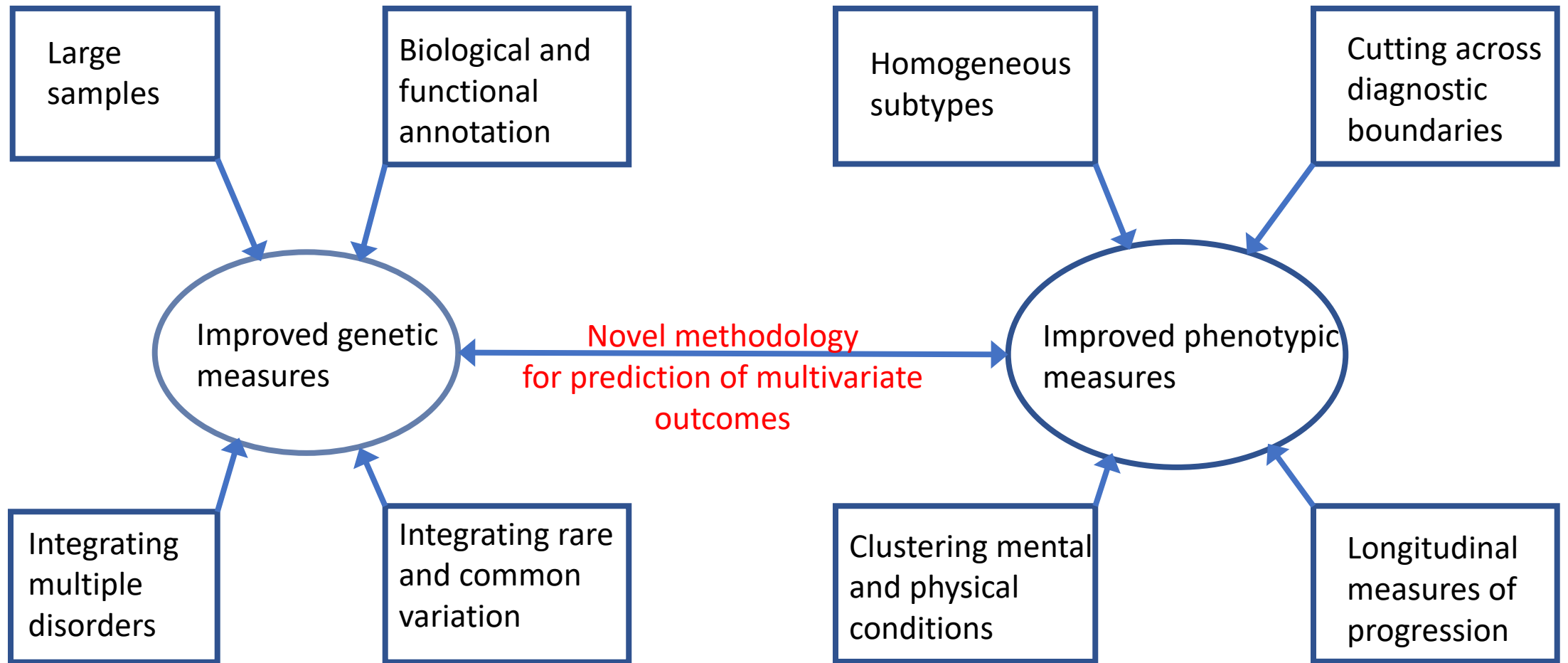
Schematic for improving outcome prediction from genetic measures