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 to be done, the prospects for genomics to accurately predict psychiatric outcomes, and thus target treatments to patients more precisely (34), are bright.

References

7. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. Genome Medicine 2020; 12:44
Improved genetic measures

- Large samples
- Biological and functional annotation
- Integrating multiple disorders
- Integrating rare and common variation

Novel methodology for prediction of multivariate outcomes

Improved phenotypic measures

- Homogeneous subtypes
- Clustering mental and physical conditions
- Longitudinal measures of progression
- Cutting across diagnostic boundaries

Schematic for improving outcome prediction from genetic measures