

LETTER TO THE EDITOR

Challenges in conducting genetic analyses based on data-driven classification of major depressive disorder

Molecular Psychiatry (2018) **23**, 494; doi:10.1038/mp.2016.205; published online 8 November 2016

Major depressive disorder (MDD) is characterized by the psychiatric symptoms of chronically low mood, low self-esteem and loss of interest or pleasure in life, and additional features such as sleep disturbances and changes in appetite and weight. It is not uncommon for different individuals to have contradicting manifestations on some of these traits, which makes it challenging to find suitable study designs and subclassification strategies for large-scale genetic studies.

Milaneschi and coauthors published a recent article in the *Journal*¹ that used innovative mathematical tools to dissect the associations between genome-wide common polymorphisms and MDD subtypes. This is an appealing strategy since it aims to enrich a specific phenotypic manifestation of MDD in an otherwise unremarkable subgroup, thus greatly enhancing the signal-to-noise ratio. Indeed, genome-wide signals for intermediate traits, which are functionally 'closer' to the DNA within a biological system than a clinical phenotype, can reveal novel genetic signals that could not be detected by the aggregate outcome observable in the clinics.² However, it is not immediately obvious if a data-driven MDD subtype, as defined in the paper by Milaneschi and coauthors, can be used similarly to an intermediate trait in order to reveal new signals.

In the paper, the authors used an unsupervised algorithm (latent class analysis, abbreviated as LCA) to identify three subtypes of MDD: two severe forms that were different with respect to appetite and weight and a third that represented less severe psychiatric symptoms. Furthermore, the authors also used a second classification based on appetite and weight change alone, which showed excellent consistency with the two severe earlier subtypes, but did not segregate for MDD with lifetime anxiety. This implies that the appetite-weight dimension was not correlated with an important psychiatric aspect of MDD, at least for the severe end of the spectrum. Therefore, one must ask if the broader metabolic changes captured via appetite and weight are, in fact, at all associated with the psychiatric or neurobiological profiles within MDD. In particular, the LCA results can be explained by a model where the severity of MDD acts as one of the many possible exposures that reveals the genetic susceptibility to weight gain and appetite (as a generic stress response), without the weight gain genes being part of MDD etiology *per se*. Essentially, the authors may have detected residual genetic correlations between appetite, weight gain and body mass index (BMI) (that arose from the subtype definition) rather than genetic aspects of specific psychiatric traits.

Milaneschi and coauthors previously investigated the complicated interplay between the *FTO* gene variants and demonstrated how the definition of MDD subtypes changed the interpretation of genetic associations; they found that an *FTO* variant increased 'atypical' MDD risk (defined via increased weight and appetite) and the signal remained significant after adjusting for baseline BMI,³ whereas a multi-cohort study of the conventional MDD definition found a protective signal.⁴ In the current study,¹ the authors showed that there was a statistical signal that linked the atypical MDD class with genetic scores for BMI and triglycerides, even when the genetic scores were adjusted for baseline BMI. Do these findings arise from genuine neurobiological subtypes that have clinical relevance, or do the differences mostly derive from the segregation of increased vs

decreased appetite response under non-specific stress that then leads to the secondary effects on weight and metabolism?

A more comprehensive phenotypic dissection of the dataset could provide additional confidence to the genetic signals. For instance, the LCA should be carefully controlled and the data patterns carefully investigated to see how and which of the inputs are associated pair-wise, and how these association patterns are likely to influence the final classification models. In a hypothetical model with appetite, weight and waist circumference; age and duration of MDD; and a depression score, the classes will most likely form based on the appetite-weight-waist dimension and the age-duration dimension simply because they are sets of correlated traits against a single weakly correlated depressive score. Furthermore, variables that are accurate are more likely to show mutual associations than noisy inputs (thus potentially skewing the classes), which is another motivation to carefully examine the association patterns between the inputs within the classification process. The authors did not provide details on how their LCA method managed this phenomenon, but these type of additional analyses can help rule out artifacts and help investigators to gain more confidence on the robustness of their models.

Ideally, replication across multiple cohorts should be sought to confirm any genetic associations with depression subtypes. Milaneschi and coauthors used an indirect route via genetic risk scores since the objective was not to discover new signals and compatible subtypes are difficult to define across cohorts. Nevertheless, stratification techniques in other studies⁵ can benefit from the data-driven subtypes. With a careful description of how the subtypes emerged from the data, and which where the driving factors for the subtyping and why, researchers can get a better sense of the subtype features and incorporate them in their analyses to meet some of the challenges in integrating the multiple depression datasets.

CONFLICT OF INTEREST

The author declares no conflict of interest.

V-P Mäkinen^{1,2,3}

¹Department of Heart Health, South Australian Health and Medical Research Institute, Adelaide, SA, Australia;

²School of Biological Sciences, Faculty of Sciences, University of Adelaide, Adelaide, SA, Australia and

³Computational Medicine, Faculty of Medicine, University of Oulu and Biocenter Oulu, Oulu, Finland

E-mail: ville-petteri.makinen@sahmri.com

REFERENCES

- Milaneschi Y, Lamers F, Peyrot WJ, Abdellaoui A, Willemsen G, Hottenga JJ *et al.* *Mol Psychiatry* 2016; **21**: 516–522.
- Kettunen J, Tukiainen T, Sarin A-P, Ortega-Alonso A, Tikkanen E, Lyttikäinen L-P *et al.* *Nat Genet* 2012; **44**: 269–276.
- Milaneschi Y, Lamers F, Mbarek H, Hottenga J-J, Boosma DI, Penninx BW. *Mol Psychiatry* 2014; **19**: 960–962.
- Samaan Z, Anand S, Zhang X, Desai D, Rivera M, Pare G *et al.* *Mol Psychiatry* 2013; **18**: 1281–1286.
- Power RA, Tansey KE, Buttenshøen H, Cohen-Woods S, Bigdeli T, Hall LS *et al.* *Biol Psychiatry* 2016 e-pub ahead of print 24 May 2016; doi: 10.1016/j.biopsych.2016.05.010.