BJPsych | The British Journal of Psychiatry (2022)
220, 219–228. doi: 10.1192/bjp.2022.28

Using polygenic scores and clinical data for bipolar disorder patient stratification and lithium response prediction: machine learning approach

Micah Cearns, Azmeraw T. Amare, Klaus Oliver Schubert, Anbupalam Thalamuthu, Joseph Frank, Fabian Streit, Mazda Adli, Nirmala Akula, Kazufumi Akiyama, Raffaella Ardau, Bárbara Arias, Jean-Michel Aubry, Lena Backlund, Abesh Kumar Bhattacharjee, Frank Bellivier, Antonio Benabarre, Susanne Bengesser, Joanna M. Biernacka, Armin Birner, Clara Brichant-Petitjean, Pablo Cervantes, Hsi-Chung Chen, Caterina Chillotti, Sven Cichon, Cristiana Cruceanu, Piotr M. Czerski, Nina Dalkner, Alexandre Dayer, Franziska Degenhardt, Maria Del Zompo, J. Raymond DePaulo, Bruno Étain, Peter Falkai, Andreas J. Forstner, Louise Frisen, Mark A. Frye, Janice M. Fullerton, Sébastien Gard, Julie S. Garnham, Fernando S. Goes, Maria Grigoroiu-Serbanescu, Paul Grof, Ryota Hashimoto, Joanna Hauser, Urs Heilbronner, Stefan Herms, Per Hoffmann, Andrea Hofmann, Liping Hou, Yi-Hsiang Hsu, Stephane Jamain, Esther Jiménez, Jean-Pierre Kahn, Layla Kassem, Po-Hsiu Kuo, Tadafumi Kato, John Kelsoe, Sarah Kittel-Schneider, Sebastian Kliwicki, Barbara König, Ichiro Kusumi, Gonzalo Laje, Mikael Landén, Catharina Lavebratt, Marion Leboyer, Susan G. Leckband, Mario Maj, the Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, Mirko Manchia, Lina Martinsson, Michael J. McCarthy, Susan McElroy, Francesc Colom, Marina Mitjans, Francis M. Mondimore, Palmiero Monteleone, Caroline M. Nievergelt, Markus M. Nöthen, Tomas Novák, Claire O'Donovan, Norio Ozaki, Vincent Millischer, Sergi Papiol, Andrea Pfennig, Claudia Pisanu, James B. Potash, Andreas Reif, Eva Reininghaus, Guy A. Rouleau, Janusz K. Rybakowski, Martin Schalling, Peter R. Schofield, Barbara W. Schweizer, Giovanni Severino, Tatyana Shekhtman, Paul D. Shilling, Katzutaka Shimoda, Christian Simhandl, Claire M. Slaney, Alessio Squassina, Thomas Stamm, Pavla Stopkova, Fasil Tekola-Ayele, Alfonso Tortorella, Gustavo Turecki, Julia Veeh, Eduard Vieta, Stephanie H. Witt, Gloria Roberts, Peter P. Zandi, Martin Alda, Michael Bauer, Francis J. McMahon, Philip B. Mitchell, Thomas G. Schulze, Marcella Rietschel, Scott R. Clark and Bernhard T. Baune

Background

Response to lithium in patients with bipolar disorder is associated with clinical and transdiagnostic genetic factors. The predictive combination of these variables might help clinicians better predict which patients will respond to lithium treatment.

Aims

To use a combination of transdiagnostic genetic and clinical factors to predict lithium response in patients with bipolar disorder.

Method

This study utilised genetic and clinical data ($n = 1034$) collected as part of the International Consortium on Lithium Genetics (ConLi^{+}Gen) project. Polygenic risk scores (PRS) were computed for schizophrenia and major depressive disorder, and then combined with clinical variables using a cross-validated machine-learning regression approach. Unimodal, multimodal and genetically stratified models were trained and validated using ridge, elastic net and random forest regression on 692 patients with bipolar disorder from ten study sites using leave-site-out cross-validation. All models were then tested on an independent test set of 342 patients. The best performing models were then tested in a classification framework.

Results

The best performing linear model explained 5.1% ($P = 0.0001$) of variance in lithium response and was composed of clinical variables, PRS variables and interaction terms between them. The best performing non-linear model used only clinical variables and explained 8.1% ($P = 0.0001$) of variance in lithium response. *A priori* genomic stratification improved non-linear model performance to 13.7% ($P = 0.0001$) and improved the binary classification of lithium response. This model stratified patients based on their meta-polygenic loadings for major depressive disorder and schizophrenia and was then trained using clinical data.

Conclusions

Using PRS to first stratify patients genetically and then train machine-learning models with clinical predictors led to large improvements in lithium response prediction. When used with other PRS and biological markers in the future this approach may help inform which patients are most likely to respond to lithium treatment.

Keywords

Mood stabilisers; bipolar affective disorders; genetics; outcome studies; depressive disorders.

[†] The online version of this article has been updated since original publication. A notice detailing the changes has also been published at https://doi.org/10.1192/bjp.2022.55.

Background

Lithium remains a commonly used first-line treatment for bipolar disorder,[1–3] highly effective for both acute manic episodes[4,5] and maintenance treatment.[3] However, in around 65% of acute manic presentations, response is incomplete and 35% of patients do not respond to treatment at all.[6,7] In maintenance treatment, approximately 30% of patients report an excellent long-term response, around 30% report an intermediate response and 30% respond poorly.[8] In addition, lithium has a range of serious acute and chronic side-effects, including increased risk of renal failure and suppression of thyroid and parathyroid function.[2] Moreover, lithium can be toxic at high doses so plasma levels often need to be monitored.[9] These varying response rates and side-effect profiles suggest the need to better tailor lithium treatment for individual patients, ensuring timely prescription of the right drug for the right patient at the right time. Better understanding of the link between genetic and clinical factors may assist in achieving this personalised approach.

Regarding associated clinical factors, variables reflecting an episodic pattern of mania–depression intervals, a later age at onset and fewer hospital admissions preceding treatment have shown significant associations with lithium response.[10,11] Using a large range of clinical factors, Nunes et al[12] demonstrated the ability of machine-learning models to classify lithium responders from non-responders. Beyond these clinical factors, a genetic basis to lithium response has also been found. In genome-wide association studies (GWASs), multiple genetic variants have been associated with lithium response. However, the effects of these variants have been too small to facilitate lithium response prediction.[8] Combining these variants into polygenic risk scores (PRS) has improved their performance, however, these scores still only explain ~1% of variance in lithium response.[13]

Given this biopsychosocial basis to lithium response, one solution may be to combine clinical factors with PRS to improve lithium response prediction. This genotype–phenotype approach was recently used by Antonucci et al[14] to classify patients with schizophrenia (SCZ) from healthy controls. Using environmental and genetic data, they *a priori* stratified patients into tertiles (thirds) based on decision scores from two support vector machines. Making predictions with patients in the lower and upper tertiles led to an increase in balanced accuracy from 77.7% to 89.4%.

In the context of lithium response prediction, Amare et al found evidence for a non-linear stratified relationship between SCZ PRS and lithium response.[13] In particular, patients in the lowest decile of the SCZ PRS distribution were 3.46 times more likely to be lithium responders when compared with patients in the tenth decile. In addition, Amare et al also found that patients with bipolar disorder with a low polygenic load for major depressive disorder (MDD) were more likely to respond to lithium treatment, with the largest differences observed between the quartiles of the MDD PRS distribution.[15] This transdiagnostic and polygenic basis to lithium response is not without precedent. For example, recent studies have shown significant genetic overlap and shared biological pathways between SCZ, MDD and bipolar disorder.[16,17] Exploiting this genetic overlap, Maier et al[17] used multitrait models to utilise correlations between, as well as a participant's individual risk for SCZ, MDD and bipolar disorder. This multivariate approach led to an equivalent increase in sample size of 34% for SCZ, 68% for bipolar disorder and 76% for MDD when compared with single trait models.

Aims

Given these findings, we have conducted a range of analyses to test the predictive ability of combined transdiagnostic genetic and clinical data for lithium response prediction. First, to measure the predictive contribution of PRS alongside clinical variables, we trained both uni- and multimodal prediction models of lithium response in patients with bipolar disorder. Second, we trained uni- and multimodal models containing interaction terms within and across variables from each data modality to measure non-linear and biopsychosocial effects between each modality. Third, to measure the effects of a patient's PRS loadings on clinical model accuracy,[14] we used MDD and SCZ PRS, as well as a combined MDD and SCZ meta-PRS to *a priori* stratify patients according to their polygenic loadings prior to the supervised prediction of lithium response with clinical data. This approach was then directly compared with the traditional method of including PRS and clinical data modalities together in a single predictive model.[18] Finally, to test the effects of model linearity on prediction, we compared the use of linear and non-linear machine-learning models for regression analyses and validated all findings on a geographically stratified test set. We then assessed the best performing models in a classification framework.

Method

The International Consortium on Lithium Genetics

The International Consortium on Lithium Genetics (ConLi$^+$Gen, www.ConLiGen.org) is an initiative by the National Institute of Mental Health and the International Group for the Study of Lithium-Treated Patients (www.IGSLI.org) and was established with the aim of studying the genetic basis of lithium treatment response in patients with bipolar disorder.[19] The ConLi$^+$Gen study involved patients with bipolar disorder from Europe, South America, USA, Asia and Australia[20] who have been treated with lithium. A series of quality control procedures were implemented on the genotype data before and after imputation as described below. Sample characteristics have been published in previous works.[8,13] This study used consortium data through an international collaboration.

The University of Heidelberg Ethics Committee provided central ethics approval for the consortium. Written consent was obtained from each patient according to the study protocols of the participating cohorts.

Computing PRS

PRS were computed for people with SCZ and MDD, two traits previously associated with lithium response.[13,15] Each PRS was calculated using individual genetic data from ConLi+Gen[8] and summary statistics from the previous largest GWASs available for MDD[21] and SCZ.[22] PRS were calculated at different GWAS *P*-value thresholds, however, the best predictive score was selected for each trait based on previous analysis. More details on PRS calculation, genotyping, imputation and quality control steps can be found in previous publications[13,15,23] and in the Supplementary Methods available at https://doi.org/10.1192/bjp.2022.28.

Study participants

As our aim was to assess both uni- and multimodal regression models of lithium response, a requirement for inclusion was complete PRS data and no more than 20% missingness on clinical predictors. Any clinical predictors above this threshold were removed (Missing data table in Supplementary Table 1). Only ConLi$^+$Gen GWAS 1 was used for analysis ($n = 1163$) as it contained both clinical and genetic data, whereas GWAS 2 only contained genetic data. Fifteen demographic, clinical, substance use and comorbid psychiatric illness predictors were used in analyses. To ensure geographic homogeneity, only samples of European descent were used (Halifax $n = 240$, University California San Diego $n = 216$, Cagliari $n = 196$, Poznan $n = 97$, Wuerzburg $n = 91$, Geneva $n = 46$, Prague $n = 45$, Dresden $n = 43$,

National Institute for Mental Health = 36, John Hopkins University n = 24), leaving a final sample of n = 1034 patients.

To ensure the unbiased approximation of the model's generalisability to new patients, we partitioned 33% (n = 342) of our sample into a holdout set for model testing, leaving 692 observations for training and validation. This partitioning was stratified by data-collection site to ensure the same distribution of sites across partitions.

Regression and classification target

Lithium treatment response was assessed using the validated 'Retrospective Criteria of Long-Term Treatment Response in Research Subjects with Bipolar Disorder' scale, also known as the Alda scale.[24] To arrive at a total Alda score, this scale measures symptom improvement over the course of treatment (A score, range 0–10), which is then weighted against five criteria (B score) that assesses the quality of evidence for the response score.[8] For the predictive regression analyses, the total Alda score was used.[25] For a subset of models assessed in a classification framework, patients with a score ≥7 were coded as responders, whereas patients scoring <7 were coded as non-responders.[25]

Unimodal and multimodal machine-learning pipelines

On our training sample of 692 patients, we fit pipelines that conducted imputation, polynomial feature engineering (interaction terms only), standardisation, feature selection, hyperparameter optimisation, and the fitting of linear regression (regularised with ridge and the elastic net) and random forest models. For the unimodal linear regression models, we imputed predictors using multivariate imputation by chained equations (MICE) with the ten nearest predictors used in the imputation process.[26] As regularised linear regression models may perform poorly if variables are on different scales, we standardised all predictors to have a mean of 0 and a s.d. of 1. As the number of predictors were low in the unimodal and multimodal PRS and clinical models (PRS 2, clinical 15), all predictors were included in analyses. Finally, we fit a linear regression model with ridge regularisation to the training sample. For the interaction term models, interaction terms between all predictors were engineered in the pipeline prior to feature selection and hyperparameter tuning. As the number of predictor variables grew exponentially in these analyses, we conducted feature selection with the elastic net, a form of penalised regression that removes highly correlated predictors from the model while retaining the most predictive subset for model fitting.[27]

This same procedure was then repeated using an random forest model, however, as random forest models are scale invariant, we did not scale the predictors prior to model training.[28] Instead of regularised linear predictor selection, we conducted non-linear predictor selection according to the mean decrease in variance provided by each predictor in the random forest model.[28] For the subset of classification models, equivalent classifiers replaced the regression models in each pipeline.

For all pipelines, we used a random search of 60 iterations to tune model hyperparameters. When less than 60 hyperparameter combinations were present we used an exhaustive grid search. See Supplementary Table 2 for the tuned hyperparameter values. This process and all steps inside the pipelines were completed using leave-site-out cross-validation. This method trains on all data-collection sites minus one. The excluded site is then used to assess the selected features and hyperparameters, with the combination that minimises the root mean squared error on the held-out site selected.[29] As there were ten collection sites, this equates to tenfold cross-validation for model selection. This site-based stratification protects against the optimisation of hyperparameters and selection of features that may proxy for disparities in feature and outcome distributions across sites and result in ungeneralisable estimates.[30] All tuned and selected models from training and validation were then further tested in the *a priori* held-out set of 342 patients.

We then re-ran all analyses with clinical variables on patients who were in the lower and upper quartiles of the PRS distributions for MDD, SCZ, and MDD and SCZ combined. See Supplementary Methods. In addition, we ran supplementary analyses to control for sample size effects and changes in the number of predictor variables across analyses (Supplementary Methods).

Results

Cohort characteristics

The final analysis cohort contained 1034 patients with an average age of 47.7 years (s.d. = 14) years and an average age at onset of bipolar disorder of 24.9 years (s.d. = 11). Of these patients, 627 (60.6%) were male and 803 (77.7%) were classified as having bipolar I disorder. The average Alda score for lithium response was 4.3 (s.d. = 3.3) out of 10. See Supplementary Fig. 1 for the full distribution of Alda scores. See Table 1 for more information on participant characteristics.

Unimodal and multimodal models

According to the coefficient of determination (R^2), the unimodal linear regression PRS and clinical models explained 1.2% and 1.8% of variance in lithium response, respectively, and the combined multimodal model explained 4.7% of variance in lithium response. Re-running the three models including interaction terms between all variables resulted in 1.4%, 4.5% and 5.1% explained variance. For the non-linear random forest models, the unimodal PRS and clinical models explained 2% and 8.1% of variance in lithium response, and the combined multimodal model explained 7.4% of variance in lithium response. Re-running the three models and including interaction terms between all variables resulted in −0.9%, 6.7% and 5.2% explained variance.

Stratified PRS analyses

For the stratified analysis using patients in the upper and lower quartiles of the MDD PRS distribution, the clinical linear and clinical linear interaction models explained −2.8% and 2.7% of variance in lithium response, whereas the non-linear random forest and random forest interaction models explained 3.5% and 1.8% of variance. For the stratified SCZ PRS analyses, the clinical linear and clinical linear interaction models explained 7.1% and 9% of variance in lithium response, and the non-linear random forest and random forest interaction models explained 7.2% and 9.3% of variance. Finally, for the stratified meta-PRS analyses, the clinical linear and clinical linear interaction models explained 12.1% and 9.2% of variance in lithium response, and the non-linear random forest and random forest interaction models explained 13.7% and 4.5% of variance. All models were statistically significant after false discovery rate (FDR) corrections. See Fig. 1 and Table 2 for all model results.

Completing 1000 runs of the Monte-Carlo sampling procedure to control for decreases in sample size on the stratified meta-PRS model, we attained an average R^2 = 2.7% (s.d. = 5, P = 0.002). Therefore, the superior performance of our meta-PRS stratified model was not explainable by increased performance variability resulting from decreased sample size.[31] In addition, increases in R^2 were not explained by changes in the number of predictor variables across models (Pearson's r = 0.17, P = 0.44) (Supplementary Fig. 1). See Table 2 for results and Supplementary Tables 3–6 for all model metrics. After controlling for these confounds, the best performing meta-PRS stratified model explained 69% more

Table 1 Characteristics of bipolar disorder patients included in analyses, stratified by their study site[a]

	Total	Cagliari/ Sardinian	Dresden	Geneva	Halifax	JHU	NIMH	Poznan	Prague/ Czech	San Diego/ UCSD	Wuerzburg	Statistic	*P*
n	1034	196	43	46	240	24	36	97	45	216	91	–	–
Age at interview, mean (s.d.)	47.7 (14.0)	46.0 (14.7)	43.5 (14.4)	46.9 (9.2)	48.5 (13.4)	44.0 (10.0)	45.6 (12.8)	61.9 (10.4)	44.8 (14.6)	45.0 (12.9)	–	15.4	4.5×10^{-21}
Age at onset of bipolar disorder, years: mean (s.d.)	24.9 (11.0)	27.1 (11.0)	25.4 (9.1)	21.4 (9.0)	26.1 (9.8)	21.7 (9.6)	19.9 (6.8)	31.5 (10.2)	28.8 (10.7)	17.6 (9.7)	28.8 (11.0)	21.9	1.3×10^{-33}
Age at onset MDD, mean (s.d.)	26.2 (12.2)	27.1 (14.3)	26.9 (11.5)	26.6 (10.1)	28.6 (11.0)	25.6 (12.0)	22.2 (10.7)	32.4 (10.2)	29.4 (10.9)	19.4 (10.1)	30.0 (10.4)	13.5	3.7×10^{-20}
Alda scale total score, mean (s.d.)	4.3 (3.3)	4.1 (3.1)	3.7 (3.2)	3.5 (3.0)	5.9 (3.6)	4.9 (3.1)	4.2 (2.7)	6.2 (2.8)	4.2 (3.0)	2.9 (2.6)	2.8 (3.0)	19.4	9.1×10^{-30}
Lithium responder, yes: *n* (%)	330 (31.9)	55 (28.1)	11 (25.6)	10 (21.7)	138 (57.5)	8 (33.3)	10 (27.8)	47 (48.5)	13 (28.9)	23 (10.6)	15 (16.5)	144.3	3.1×10^{-26}
Bipolar I versus rest, *n* (%)	803 (77.7)	147 (75.0)	31 (72.1)	36 (78.3)	173 (72.1)	24 (100.0)	33 (91.7)	80 (82.5)	39 (86.7)	187 (86.6)	53 (58.2)	49.9	1.4×10^{-07}
Bipolar I schizoaffective bipolar disorder versus rest, *n* (%)	821 (79.4)	147 (75.0)	31 (72.1)	36 (78.3)	177 (73.8)	24 (100.0)	35 (97.2)	80 (82.5)	39 (86.7)	199 (92.1)	53 (58.2)	70.0	2.3×10^{-11}
Alcohol dependence, yes: *n* (%)	159 (17.5)	19 (9.7)	3 (7.3)	14 (30.4)	7 (2.9)	13 (54.2)	6 (16.7)	10 (12.5)	9 (20.0)	78 (37.0)	0 (0.0)	170.1	3.7×10^{-32}
Substance dependence, yes: *n* (%)	152 (16.8)	17 (8.7)	1 (2.4)	13 (28.3)	1 (0.4)	11 (45.8)	2 (5.6)	42 (52.5)	1 (2.2)	64 (30.5)	0 (0.0)	231.4	1.0×10^{-44}
OCD, yes: *n* (%)	46 (5.1)	0 (0.0)	3 (7.3)	4 (8.7)	6 (2.5)	2 (8.3)	5 (13.9)	6 (7.5)	0 (0.0)	20 (9.4)	0 (0.0)	69.8	8.6×10^{-12}
PTSD, yes: *n* (%)	62 (7.2)	0 (0.0)	0 (0.0)	2 (4.4)	0 (0.0)	0 (0.0)	0 (0.0)	16 (20.0)	1 (2.2)	43 (20.3)	0 (0.0)	112.9	1.1×10^{-21}
Panic disorder, yes: *n* (%)	124 (13.8)	0 (0.0)	5 (12.2)	6 (13.0)	15 (6.3)	5 (20.8)	9 (25.0)	43 (53.8)	4 (8.9)	37 (17.5)	0 (0.0)	189.1	4.9×10^{-36}
Gender, male: *n* (%)	627 (60.6)	133 (67.9)	23 (53.5)	25 (54.3)	139 (57.9)	17 (70.8)	28 (77.8)	60 (61.9)	30 (66.7)	109 (50.5)	63 (69.2)	25.1	2.8×10^{-03}
Polarity 1st episode (depression onset), *n* (%)	655 (63.3)	130 (66.3)	25 (58.1)	22 (47.8)	150 (62.5)	13 (54.2)	23 (63.9)	51 (52.6)	38 (84.4)	133 (61.6)	70 (76.9)	28.0	1.0×10^{-03}
Polarity 1st episode (mania onset), *n* (%)	185 (17.9)	61 (31.1)	7 (16.3)	3 (6.5)	33 (13.8)	11 (45.8)	11 (30.6)	12 (12.4)	4 (8.9)	33 (15.3)	10 (11.0)	55.4	1.4×10^{-08}
Polarity 1st episode (hypomania onset), *n* (%)	67 (6.5)	5 (2.6)	3 (7.0)	2 (4.3)	8 (3.3)	0 (0.0)	0 (0.0)	1 (1.0)	2 (4.4)	46 (21.3)	0 (0.0)	103.0	6.4×10^{-18}
Any suicidal features, no: *n* (%)	433 (41.9)	125 (63.8)	0 (0.0)	9 (19.6)	104 (43.3)	15 (62.5)	21 (58.3)	61 (62.9)	14 (31.1)	41 (19.0)	43 (47.3)	154.7	2.2×10^{-28}
Any suicidal features, yes: *n* (%)	449 (43.4)	70 (35.7)	18 (41.9)	34 (73.9)	14 (17.1)	9 (37.5)	15 (41.7)	19 (19.6)	31 (68.9)	164 (75.9)	48 (52.7)	220.8	7.8×10^{-42}
Any suicidal features, unknown: *n* (%)	152 (14.7)	1 (0.50)	25 (58.1)	3 (6.5)	95 (39.6)	0 (0.0)	0 (0.0)	17 (17.5)	0 (0.0)	11 (5.1)	0 (0.0)	267.4	1.9×10^{-51}
DSM diagnosis (bipolar I disorder), *n* (%)	803 (77.7)	147 (75.0)	31 (72.1)	36 (78.3)	173 (72.1)	24 (100.0)	33 (91.7)	80 (82.5)	39 (86.7)	187 (86.6)	53 (58.2)	49.9	1.4×10^{-07}
DSM diagnosis (bipolar II disorder), *n* (%)	203 (19.6)	49 (25.0)	12 (27.9)	10 (21.7)	63 (26.3)	0 (0.0)	1 (2.8)	17 (17.5)	6 (13.3)	15 (6.9)	30 (33.0)	58.3	4.1×10^{-09}
DSM diagnosis (schizoaffective bipolar disorder), *n* (%)	18 (1.7)	0 (0.0)	0 (0.0)	0 (0.0)	4 (1.7)	0 (0.0)	2 (5.6)	0 (0.0)	0 (0.0)	12 (5.6)	0 (0.0)	31.0	3.1×10^{-04}
DSM diagnosis (bipolar III disorder), *n* (%)	3 (0.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	3 (3.3)	31.2	3.1×10^{-04}
DSM diagnosis (bipolar disorder not otherwise specified), *n* (%)	7 (0.7)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	2 (0.9)	5 (5.5)	36.6	3.7×10^{-05}
Bipolar disorder family history, yes: *n* (%)	351 (33.9)	83 (42.3)	31 (72.1)	0 (0.0)	44 (18.3)	0 (0.0)	12 (33.3)	42 (43.3)	10 (22.2)	129 (59.7)	0 (0.0)	213.5	2.23×10^{-40}
Bipolar disorder family history, no: *n* (%)	386 (37.3)	113 (57.7)	9 (20.9)	0 (0.0)	154 (64.2)	0 (0.0)	0 (0.0)	38 (39.2)	35 (77.8)	37 (17.1)	0 (0.0)	300.1	3.42×10^{-58}
Bipolar disorder family history, unknown: *n* (%)	297 (28.7)	0 (0.0)	3 (7.0)	46 (100.0)	42 (17.5)	24 (100.0)	24 (66.7)	17 (17.5)	0 (0.0)	50 (23.1)	91 (100.0)	555.9	1.65×10^{-112}

JHU, John Hopkins University; NIMH,National Institute of Mental Health; UCSD, University California San Diego; MDD, major depressive disorder; OCD, obsessive–compulsive disorder; PTSD, post-traumatic stress disorder.
a. Statistics calculated using one-way ANOVA for continuous variables and Fischer exact tests for categorical variables. All *P*-values were false discovery rate-corrected using the Benjamini and Hochberg method.

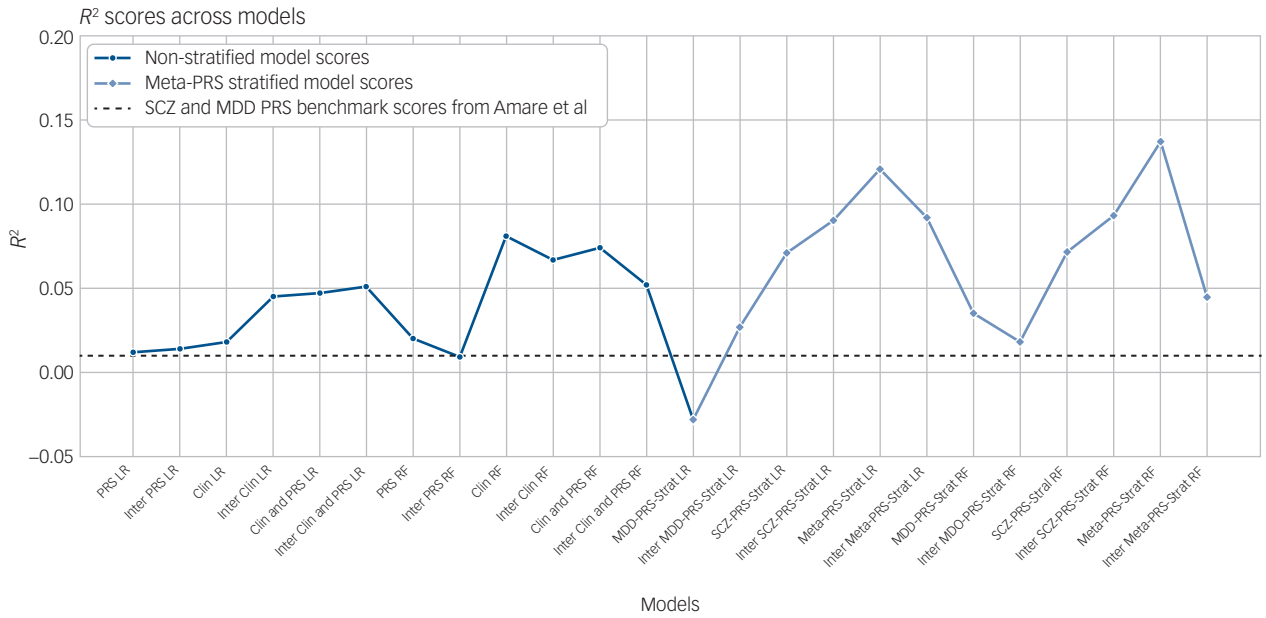

Fig. 1 Line plot for all model R^2 scores. Black dashed line: R^2 benchmarks for schizophrenia (SCZ) and major depressive disorder (MDD) polygenic risk scores (PRS) and their relationship with lithium response in previous works from Amare et al.[17,18] Dark blue line: R^2 scores for the non-PRS stratified models. Light blue line: R^2 scores for the PRS stratified models.
Clin, clinical; Inter, interactions terms; LR, linear regression; RF, random forest; strat, stratified.

variance ($R^2 = 13.7\%$, $P = 0.0001$) than the equivalent model containing no *a priori* meta-PRS stratification ($R^2 = 8.1\%$, $P = 0.0001$). In this model, all clinical variables were retained in model selection (See Supplementary Table 7). Re-running these two best performing models in a classification framework led to balanced accuracies of 58.95% and 63.65%, respectively. See Supplementary Table 8 for all classification metrics.

Patient characteristics in the genetically stratified cohort

After FDR corrections, significant differences in clinical characteristics were found between those in quartiles 1 (low meta-PRS load) and 4 (high meta-PRS load) of the combined meta-PRS distribution for binary lithium response (ALDA ≥ 7) ($\chi^2 = 12.214$, $P = 0.005$), bipolar I disorder versus rest (bipolar II disorder and schizoaffective disorder) ($\chi^2 = 12.755$, $P = 0.005$) and DSM diagnosis (bipolar I disorder, bipolar II disorder and schizoaffective disorder) ($\chi^2 = 13.33$, $P = 0.027$).

In quartile 1 of the meta-PRS distribution, 70% had bipolar I disorder, 26% had bipolar II disorder and 4% had schizoaffective disorder. In total, 39% of these patients were lithium responders. In quartile 4, 86% had bipolar I disorder, 12% had bipolar II disorder and 3% had schizoaffective disorder. In total, 22% of these patients were lithium responders.

Overall, those in quartile 1 were 67.7% more likely to be lithium responders compared with those in quartile 4 (odds ratio 1.677, 95% CI 1.14–2.47, $P = 0.009$). For all other clinical characteristics, including variables that attained nominal significance, see Table 3.

Discussion

Main findings

This is the first study to provide evidence for the combined predictive ability of routine clinical data and PRS for lithium response. Specifically, we show that first using PRS to stratify patients according to their polygenic loadings, followed by training with clinical data explains more variance in lithium response and improves model accuracy in a classification setting.[13,14] Interestingly, the combination of PRS with clinical data performed best in the linear models, but not in the non-linear models. Outside of the best performing stratified meta-PRS model, neither of these multimodal models performed best overall. Moreover, unimodal clinical models outperformed their PRS equivalents.

Interpretation of our findings

This observation of clinical variables outperforming their biological counterparts has been repeatedly demonstrated across a range of multimodal machine-learning studies.[18,32] The most intuitive explanation is that the small effect sizes yielded by biological variables, when compared with clinical variables, leads to overfitting and/or their lack of selection in cross-validation, resulting in underperformance for biomarker models when tested out of sample.

The next consideration is why effects are smaller for biological variables. To answer this, we need to consider how psychiatric traits are constructed and the implications this has for studies attempting to elucidate a biological basis for psychiatric phenomenon. In comparison with other disorders, psychiatric phenotypes are defined by deviations from normative behaviours and emotional–cognitive experiences, rather than from well-defined physiological processes.[30] Therefore, it is plausible that they bias towards larger effect sizes for clinical variables that correlate with clinical data already used in the construction of phenotypes and illness trajectories. Consequently, this tautology in the formation of diagnostic and prognostic constructs may limit the predictive contribution of biological data. In theory, this problem could be circumvented by first parsing patient heterogeneity at the biological level, and then using clinical variables in secondary analyses.

This rationale informed our stratified analyses where we *a priori* partitioned patients based on their polygenic loadings for MDD, SCZ and their combination in the form of a standardised meta-PRS. Interestingly, this method was most predictive of lithium

Table 2 Train/validation ($n = 692$) and test ($n = 342$) results across all models[a]

	Train mean (s.d.)	Validate mean (s.d.)	Test	*P*
Linear (ridge and elastic net regression)				
PRS	0.01 (0.0)	0.20 (0.18)	0.012	0.0131
Interaction PRS	0.01 (0.0)	0.20 (0.18)	0.014	0.01
Clinical	0.12 (0.02)	0.32 (0.17)	0.018	0.0004
Interaction clinical	0.12 (0.02)	0.26 (0.19)	0.045	0.0001
Clinical and PRS	0.09 (0.02)	0.20 (0.17)	0.047	0.0001
Interaction clinical and PRS	0.13 (0.02)	0.27 (0.19)	0.051	0.0001
Random forest regression				
PRS	0.09 (0.01)	0.22 (0.18)	0.02	0.0001
Interaction PRS	0.09 (0.01)	0.22 (0.19)	0.009	0.0001
Clinical	0.21 (0.02)	0.21 (0.11)	0.081	0.0001
Interaction clinical	0.24 (0.02)	0.23 (0.14)	0.067	0.0001
Clinical and PRS	0.22 (0.02)	0.21 (0.12)	0.074	0.0001
Interaction clinical and PRS	0.25 (0.02)	0.23 (0.16)	0.052	0.0001
A priori PRS stratified regression (trained with clinical predictors)				
Linear MDD stratified	0.12 (0.01)	0.12 (0.31)	−0.028	0.0171
Linear interaction MDD PRS stratified	0.13 (0.01)	0.21 (0.36)	0.027	0.0013
Linear SCZ PRS stratified	0.12 (0.01)	0.41 (0.42)	0.071	0.0004
Linear interaction SCZ PRS stratified	0.12 (0.01)	0.31 (0.21)	0.09	0.0001
Linear Meta-PRS stratified	0.09 (0.01)	0.67 (0.39)	0.121	0.0001
Linear interaction meta-PRS stratified	0.10 (0.01)	0.51 (0.34)	0.092	0.0001
Random forest MDD PRS stratified	0.18 (0.01)	0.16 (0.22)	0.035	0.0001
Random forest interaction MDD PRS stratified	0.23 (0.01)	0.20 (0.29)	0.018	0.0001
Random forest SCZ PRS stratified	0.24 (0.01)	0.24 (0.14)	0.072	0.0001
Random forest interaction SCZ PRS stratified	0.27 (0.02)	0.26 (0.14)	0.093	0.0001
Random forest meta-PRS stratified	0.23 (0.02)	0.40 (0.29)	0.137	0.0001
Random forest interaction meta-PRS stratified	0.70 (0.01)	0.39 (0.35)	0.045	0.0001

PRS, polygenic risk scores; MDD, major depressive disorder; SCZ, schizophrenia.
a. Unimodal, multimodal and interaction term predictors spaces were measured using both linear regression (ridge and the elastic net) and random forest regression models. In addition, PRS stratified models composed of MDD PRS, SCZ PRS, and their standardised combinations in the form of a meta-PRS were assessed across model types and feature interaction combinations. Mean and (s.d.) represent the mean (s.d.) from the leave-site-out train and validation procedures. All *P*-values were false discovery rate-corrected with the Benjamini and Hochberg method.

response overall, explaining 69% more variance than the equivalent model with no *a priori* meta-PRS stratification. These results support the view that first parsing biological heterogeneity may improve the prediction of bipolar disorder lithium response with clinical data.

When assessing the best performing non-stratified model, the clinical random forest model, and the best performing overall model, the meta-PRS stratified model, we also observed an increase in model performance in a classification framework, albeit a smaller percentage change. This observation warranted an inspection and interpretation of the lithium response distributions between patients in the low and high meta-PRS quartiles (Supplementary Fig. 1). Here, we observe disproportionate densities of very low (0) and moderate response (5–7) scores for patients with high meta-PRS loadings. Conversely, for patients with low meta-PRS loadings, we observe disproportionate densities of very low (0) and very high response (7–10) scores. Both quartiles of meta-PRS loadings demonstrated high densities of very low response, yet differences between moderate and high response scores were evident. More specifically, patients with low meta-PRS loadings belonged to a continuous bimodal response distribution, whereas those with high meta-PRS loadings appeared to be mixed across the distribution and skewed towards lower response. When dichotomising lithium response, this nuanced understanding between a patient's genetic loadings and lithium response was lost.

This observation is interesting in light of recent work that quantified the asymmetrical reliability of the Alda scale, finding higher interrater reliability in the upper tail of the response distribution.[33] Therefore, a dichotomous representation of lithium response was generally argued for, even after considering the resultant loss in statistical power. However, rather than deciding *a priori* to discretise this distribution, an alternative approach would be to tune and select models in a leave-site-out cross-validation framework, as was done in the current work. This is because we would expect to see the highest amount of interrater disagreement between data-collection sites, as purported by Nunes et al.[33] If it was high enough to warrant *a priori* discretisation, these across site models would not generalise because of their disagreement in lithium response. However, in the current work nearly all models generalised across sites to the out-of-sample-test sets that were excluded from model construction, demonstrating that the use of leave-site-out cross-validation ensured that each model was tuned to learn parameters and relationships that generalised regardless of any disagreement between raters across sites. In addition, this established that there was enough agreement between raters to learn meaningful, informative and generalisable patterns in the continuous lithium response distribution.

In future works, an alternative to dichotomising the Alda scale would be to use the full scale and run analyses using spline regression.[34] With this technique, we would not build one model for the entire data-set, but instead, divide the data-set into multiple bins and fit each bin with its own model. Some of these models may be linear, whereas others may be polynomial. This approach would allow us to fit PRS to the lithium response distribution and account for the linear and non-linear relationships between different strata of the PRS and lithium response distributions.[34]

Regarding the clinical characteristics of patients in each meta-PRS quartile, we observed significant differences between the types of psychiatric diagnosis. Quartile 1 (low meta-PRS load) had lower proportions of bipolar I disorder diagnoses and higher proportions of bipolar II disorder and schizoaffective disorder, whereas the opposite was true for those in quartile 4 (high meta-PRS load). Given that higher meta-PRS loadings are associated with poorer lithium response, and that people with 'purer' forms of bipolar I disorder are considered better responders to lithium,[2,35,36] this is an unexpected finding: one might have hypothesised that there would be a higher proportion of people with bipolar I disorder in the low, and therefore less 'contaminated', meta-PRS group. Our finding suggests that the relationship between meta-polygenic disposition for SCZ and MDD and actual phenotypical expression of bipolar spectrum disorders is more complex,[37] and that people with seemingly unfavourable genetic constellations may still benefit from lithium once other clinical and environmental parameters come into play. Similarly, patients with seemingly less favourable diagnoses for lithium response (i.e. bipolar II disorder and schizoaffective disorder) may still benefit if their polygenic disposition points towards better responsiveness.

Considerations for clinical use

This leads to two main considerations for clinical use. The overall increase in variance from combining clinical and genetic data may

Table 3 Descriptive statistics for the clinical profiles of patients with bipolar disorder in the lower and upper quartiles of the meta-polygenic risk scores (PRS) distribution[a,b]

	Quartile 1 (low meta-PRS)	Quartile 4 (high meta- PRS)	Statistic	*P*	FDR *P*
n	259	259	–	–	–
Alda scale total score, mean (s.d.)	4.852 (3.412)	4.078 (3.171)	2.412	**0.016**	0.065
Lithium response (Alda ≥7), no/yes: %	61/39	78/22	12.214	**0.001**	**0.005**
Age at onset of bipolar, mean (s.d.)	25.723 (11.606)	23.548 (10.784)	1.993	**0.047**	0.104
Age at onset of depression, mean (s.d.)	27.520 (12.587)	25.790 (12.010)	1.427	0.154	0.237
Age at interview, mean (s.d.)	48.277 (14.052)	48.434 (14.010)	−0.114	0.909	0.909
Bipolar I disorder versus rest, no/yes: %	30/70	14/86	12.755	**<0.001**	**0.005**
Polarity first episode, depression/mania/hypomania onset: %	74/17/10	67/21/12	1.961	0.375	0.491
Alcohol dependence, no/yes: %	87/13	78/22	4.567	**0.033**	0.093
Substance dependence, no/yes: %	87/13	82/18	1.916	0.166	0.238
Obsessive–compulsive disorder, no/yes: %	95/5	94/06	0.270	0.603	0.670
Post-traumatic stress disorder, no/yes: %	96/4	91/09	3.726	0.054	0.107
Panic disorder, no/yes: %	92/8	86/14	2.980	0.084	0.153
Any suicidal features, no/yes/unknown: %	46/39/15	39/52/09	6.986	**0.030**	0.093
Gender, female/male: %	42/58	46/54	0.552	0.458	0.539
DSM diagnosis, bipolar I disorder/bipolar II disorder/schizoaffective disorder: %	70/26/4	86/12/3	13.330	**0.004**	**0.027**
Bipolar family history, no/yes/unknown: %	42/46/12	42/41/16	1.869	0.393	0.491

FDR, false discovery rate.
a. Statistic: calculated using independent samples *t*-tests for continuous variables and χ^2 tests for categorical variables. All *P*-values were FDR-corrected using the Benjamini and Hochberg method.
b. Nominal and FDR-corrected *P*-values in bold.

be of use for clinicians to improve the accuracy of their clinical decision-making overall, especially when combined with other PRS and biomarkers in future works and then incorporated into classification models. Further and more immediate benefit could be derived from using this added genetic data to reconsider patients who would be traditionally ruled out as favourable responders to lithium based solely on their clinical presentation if their meta-polygenic loadings suggest otherwise.

Limitations

A number of limitations exist in the current work. First, there was a limited amount of clinical data available for analysis in this cohort. Future studies should aim to collect a wider range of clinical data (for example symptom scales) to elucidate the relationship between PRS and clinical characteristics, as well as their combined predictive ability. Ideally, prospective studies of lithium response will be required in the future to quantify the predictive ability of machine-learning models in an environment that is analogous to clinical practice.

Second, correctly operationalising bipolar I disorder, bipolar II disorder and schizoaffective disorder DSM phenotypes is difficult in real-world practice. Relying on patient's retrospective reporting of symptoms and past episodes to form these diagnoses, as was done in the current study, can lead to misdiagnosis of bipolar subtypes.[2,3]

Another consideration concerns the selection of quantile-based PRS stratification over tertile[12] and decile[13] stratification used in previous works. Choosing the number of PRS strata involves considering the trade-offs between a higher number of bins (i.e. decile stratification) that would likely contain larger differences in clinical characteristics, lithium response and polygenic risk, but result in a smaller sample (only 20% of the original sample would be retained when taking the extreme deciles). Alternatively, a lower number of bins (i.e. tertiles) would result in the opposite being true. To balance this trade-off, we chose quartile-based stratification. When taking the two extreme quartiles, we retained 50% of the original sample, while removing the middle of the PRS distributions that shows the smallest genetic differences in lithium response.[13,15] However, future studies could attempt to find the optimal number of strata through the use of cross-validation with spline regression, where the optimal number of strata could be tuned and selected according to the minimisation of a loss function. If completed in a leave-site-out framework, between site rater disagreement would be controlled for and the full lithium response distribution could be modelled.

The next limitation pertains to the patients that do not fall in the tails of the PRS distributions and who would therefore be excluded from prediction with this model. However, such a stratified model that confers superior predictive ability could first be used for patients that fall within these strata, and for patients that do not, models without stratification could be used[12] or other stratifying biomarkers could be incorporated.[38] Through this lens, we envision a stepwise process in clinical deployment where the choice of model itself would be tailored to individual patients depending on their unique clinical and biological characteristics (Supplementary Fig. 2). An alternative approach to parse biological heterogeneity would be to use unsupervised machine-learning models.[39] However, the disproportionately small effect sizes afforded by PRS,[40] the large risk of overfitting on unlabelled data,[41] the high level of polygenic collinearity across psychiatric traits[16,42] and the resultant demands these considerations impose on statistical power,[43] led us to take a simpler approach informed by previous findings.[13,14]

Implications

In conclusion, using PRS to stratify patients genetically and then train machine-learning models with clinical predictors led to large improvements in lithium response prediction over other forms of unimodal and multimodal modelling. Clinical data explained the most variance and both clinical and PRS data showed non-linear relationships with lithium response. To adequately model the linear and non-linear relationships between these PRS and lithium response across different genetic strata, future works should consider modelling these relationships using spline regression. Moreover, engineering a direct lithium response PRS and using this to parse heterogeneity may further improve model performance. In addition, parsing heterogeneity with biomarkers from neuroimaging and omics domains should also be considered. Finally, data-sets with a larger range of clinical variables will likely improve prediction following genetic stratification.

Micah Cearns, BSc (Hons), PhD, Discipline of Psychiatry, School of Medicine, University of Adelaide, Australia; **Azmeraw T. Amare**, MPH, MSc, PhD, Discipline of Psychiatry, School of Medicine, University of Adelaide; **Klaus Oliver Schubert**, MD, PhD, Discipline of Psychiatry, School of Medicine, University of Adelaide, Australia and Northern Adelaide Local Health Network, Mental Health Services, Australia; **Anbupalam Thalamuthu**, PhD, Centre for Healthy Brain Ageing (CHeBA), School of Psychiatry, University of New South Wales, Australia; **Joseph Frank**, PhD, Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Germany; **Fabian Streit**, PhD, Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Germany; **Mazda Adli**, MD, Department of Psychiatry and Psychotherapy, Charité - Universitätsmedizin Berlin, Campus Charité Mitte, Germany; **Nirmala Akula**, PhD, Intramural Research Program, National Institute of Mental Health, National Institutes of Health, US Department of Health & Human Services, USA; **Kazufumi Akiyama**, MD, Department of Biological Psychiatry and Neuroscience, Dokkyo Medical University School of Medicine, Japan; **Raffaella Ardau**, MD, Unit of Clinical Pharmacology, Hospital University Agency of Cagliari, Italy; **Bárbara Arias**, PhD, Unitat de Zoologia i Antropologia Biològica (Dpt. Biologia Evolutiva, Ecologia i Ciències Ambientals), Facultat de Biologia and Institut de Biomedicina (IBUB), University of Barcelona, CIBERSAM, Spain; **Jean-Michel Aubry**, MD, Department of Psychiatry, Mood Disorders Unit, HUG - Geneva University Hospitals, Switzerland; **Lena Backlund**, MD, PhD,, Department of Molecular Medicine and Surgery, Karolinska Institute, Sweden and Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden; **Abesh Kumar Bhattacharjee**, MD,, Department of Psychiatry, University of California San Diego, USA; **Frank Bellivier**, MD, PhD, INSERM UMR-S 1144, Université Paris Diderot, Département de Psychiatrie et de Médecine Addictologique, AP-HP, Groupe Hospitalier Saint-Louis-Lariboisière-F.Widal, France; **Antonio Benabarre**, MD, PhD, Bipolar Disorder Program, Institute of Neuroscience, Hospital Clinic, University of Barcelona, IDIBAPS, CIBERSAM, Spain; **Susanne Bengesser**, MD, Department of Psychiatry and Psychotherapeutic Medicine, Research Unit for bipolar affective disorder, Medical University of Graz, Austria; **Joanna M. Biernacka**, PhD, Department of Health Sciences Research, Mayo Clinic, USA and Department of Psychiatry and Psychology, Mayo Clinic, USA; **Armin Birner**, MD, Department of Psychiatry and Psychotherapeutic Medicine, Research Unit for Bipolar Affective Disorder, Medical University of Graz, Austria; **Clara Brichant-Petitjean**, MD, INSERM UMR-S 1144, Université Paris Diderot, Département de Psychiatrie et de Médecine Addictologique, AP-HP, Groupe Hospitalier Saint-Louis-Lariboisière-F.Widal, France; **Pablo Cervantes**, MD, The Neuromodulation Unit, McGill University Health Centre, Canada; **Hsi-Chung Chen**, MD, PhD, Department of Psychiatry & Center of Sleep Disorders, National Taiwan University Hospital, Taiwan; **Caterina Chillotti**, MD, Unit of Clinical Pharmacology, Hospital University Agency of Cagliari, Italy; **Sven Cichon**, PhD, Institute of Human Genetics, University of Bonn and Department of Genomics, Life & Brain Center, Germany; and Human Genomics Research Group, Department of Biomedicine, University Hospital Basel, Switzerland; **Cristiana Cruceanu**, PhD, Douglas Mental Health University Institute, McGill University, Canada; **Piotr M. Czerski**, PhD, Psychiatric Genetic Unit, Poznan University of Medical Sciences, Poland; **Nina Dalkner**, MSc, Department of Psychiatry and Psychotherapeutic Medicine, Research Unit for Bipolar Affective Disorder, Medical University of Graz, Austria; **Alexandre Dayer**, MD, Department of Psychiatry, Mood Disorders Unit, HUG - Geneva University Hospitals, Switzerland; **Franziska Degenhardt**, MD, Institute of Human Genetics, University of Bonn and Department of Genomics, Life & Brain Center, Germany; **Maria Del Zompo**, MD, Department of Biomedical Sciences, University of Cagliari, Italy; **J. Raymond DePaulo**, MD, Department of Psychiatry and Behavioral Sciences, Johns Hopkins University, USA; **Bruno Étain**, MD, PhD, INSERM UMR-S 1144, Université Paris Diderot, Département de Psychiatrie et de Médecine Addictologique, AP-HP, Groupe Hospitalier Saint-Louis-Lariboisière-F.Widal, France; **Peter Falkai**, MD, Department of Psychiatry and Psychotherapy, Ludwig-Maximilian-University Munich, Germany; **Andreas J. Forstner**, MD, Institute of Human Genetics, University of Bonn and Department of Genomics, Life & Brain Center, Germany; Human Genomics Research Group, Department of Biomedicine, University Hospital Basel, Switzerland and Department of Psychiatry (UPK), University of Basel, Switzerland; **Louise Frisen**, MD, Department of Molecular Medicine and Surgery, Karolinska Institute, Sweden and Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden; **Mark A. Frye**, MD,, Department of Psychiatry and Psychology, Mayo Clinic, USA; **Janice M. Fullerton**, PhD, Neuroscience Research Australia, Australia and School of Medical Sciences, University of New South Wales, Australia; **Sébastien Gard**, MD, Service de Psychiatrie, Hôpital Charles Perrens, France; **Julie S. Garnham**, BN, Department of Psychiatry, Dalhousie University, Canada; **Fernando S. Goes**, MD,, Department of Psychiatry and Behavioral Sciences, Johns Hopkins University, USA; **Maria Grigoroiu-Serbanescu**, PhD, Biometric Psychiatric Genetics Research Unit, Alexandru Obregia Clinical Psychiatric Hospital, Romania; **Paul Grof**, MD, PhD, Mood Disorders Center of Ottawa, Canada; **Ryota Hashimoto**, MD, PhD, Molecular Research Center for Children's Mental Development, United Graduate School of Child Development, Osaka University, Japan and Department of Psychiatry, Osaka University Graduate School of Medicine, Japan; **Joanna Hauser**, MD, Psychiatric Genetic Unit, Poznan University of Medical Sciences, Poland; **Urs Heilbronner**, PhD, Institute of Psychiatric Phenomics and Genomics (IPPG), University Hospital, LMU Munich, Germany and Department of Psychiatry and Psychotherapy, University Medical Center (UMG), Georg-August University Göttingen, Germany; **Stefan Herms**, Dipl Biol, Institute of Human Genetics, University of Bonn and Department of Genomics, Life & Brain Center, Germany and Human Genomics Research Group, Department of Biomedicine, University Hospital Basel, Switzerland; **Per Hoffmann**, PhD, Institute of Human Genetics, University of Bonn and Department of Genomics, Life & Brain Center, Germany and Human Genomics Research Group, Department of Biomedicine, University Hospital Basel, Switzerland; **Andrea Hofmann**, PhD, Institute of Human Genetics, University of Bonn and Department of Genomics, Life & Brain Center, Germany; **Liping Hou**, PhD, Intramural Research Program, National Institute of Mental Health, National Institutes of Health, US Department of Health & Human Services, USA; **Yi-Hsiang Hsu**, PhD, Program for Quantitative Genomics, Harvard School of Public Health, USA and HSL Institute for Aging Research, Harvard Medical School, USA; **Stephane Jamain**, PhD, Univ Paris Est Créteil, INSERM, IMRB, Translational Neuropsychiatry, Fondation FondaMental, France; **Esther Jiménez**, PhD, Bipolar Disorder Program, Institute of Neuroscience, Hospital Clinic, University of Barcelona, IDIBAPS, CIBERSAM, Spain; **Jean-Pierre Kahn**, MD, PhD, Service de Psychiatrie et Psychologie Clinique, Centre Psychothérapique de Nancy - Université de Lorraine, France; **Layla Kassem**, PhD, Intramural Research Program, National Institute of Mental Health, National Institutes of Health, US Department of Health & Human Services, USA; **Po-Hsiu Kuo**, PhD, Department of Public Health & Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taiwan; **Tadafumi Kato**, MD, PhD, Department of Psychiatry & Behavioral Science, Juntendo University, Graduate School of Medicine, Japan; **John Kelsoe**, MD, Department of Psychiatry, University of California San Diego, USA; **Sarah Kittel-Schneider**, MD, Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, University Hospital Frankfurt, Germany; **Sebastian Kliwicki**, MD, Department of Adult Psychiatry, Poznan University of Medical Sciences, Poland; **Barbara König**, MSc, Department of Psychiatry and Psychotherapeutic Medicine, Landesklinikum Neunkirchen, Austria; **Ichiro Kusumi**, MD, Department of Psychiatry, Hokkaido University Graduate School of Medicine, Japan; **Gonzalo Laje**, MD, Intramural Research Program, National Institute of Mental Health, National Institutes of Health, US Department of Health & Human Services, USA; **Mikael Landén**, MD, Institute of Neuroscience and Physiology, the Sahlgrenska Academy at the Gothenburg University, Sweden and Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Sweden; **Catharina Lavebratt**, PhD, Department of Molecular Medicine and Surgery, Karolinska Institute, Sweden and Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden; **Marion Leboyer**, MD, PhD, Univ Paris Est Créteil, INSERM, IMRB, Translational Neuropsychiatry, AP-HP, Mondor University Hospital, DMU Impact, Fondation FondaMental, France; **Susan G. Leckband**, BSc, Office of Mental Health, VA San Diego Healthcare System, USA; **Mario Maj**, MD, Department of Psychiatry, University of Campania 'Luigi Vanvitelli', Italy; ,for a full list of Major Depressive Disorder Working Group of the PGC Investigators, see the Supplementary Material; **the Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium**; **Mirko Manchia**, MD, PhD, Section of Psychiatry, Department of Medical Sciences and Public Health, University of Cagliari, Italy and Department of Pharmacology, Dalhousie University, Canada; **Lina Martinsson**, MD, Department of Clinical Neurosciences, Karolinska Institutet, Sweden; **Michael J. McCarthy**, MD, PhD, Department of Psychiatry, University of California San Diego, USA and Department of Psychiatry, VA San Diego Healthcare System, USA; **Susan McElroy**, MD, Department of Psychiatry, Lindner Center of Hope / University of Cincinnati, USA; **Francesc Colom**, PhD, Mental Health Research Group, IMIM-Hospital del Mar, Spain and Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Instituto de Salud Carlos III, Spain; **Marina Mitjans**, PhD, Mental Health Research Group, IMIM-Hospital del Mar, Spain and Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Instituto de Salud Carlos III, Spain; **Francis M. Mondimore**, MD, Department of Psychiatry and Behavioral Sciences, Johns Hopkins University, USA; **Palmiero Monteleone**, MD, Department of Medicine, Surgery and Dentistry 'Scuola Medica Salernitana', University of Salerno, Italy; **Caroline M. Nievergelt**, PhD,, Department of Psychiatry, University of California San Diego, USA; **Markus M. Nöthen**, MD, Institute of Human Genetics, University of Bonn and Department of Genomics, Life & Brain Center, Germany; **Tomas Novák**, MD, National Institute of Mental Health, Czech Republic; **Claire O'Donovan**, MB, Department of Psychiatry, Dalhousie University, Canada; **Norio Ozaki**, MD, Department of Psychiatry & Department of Child and Adolescent Psychiatry, Nagoya University Graduate School of Medicine, Japan; **Vincent Millischer**, PhD, Department of Molecular Medicine and Surgery, Karolinska Institute, Sweden and Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden and Department of Psychiatry and Psychotherapy, Medical University of Vienna, Austria; **Sergi Papiol**, PhD, Institute of Psychiatric Phenomics and Genomics (IPPG), University Hospital, LMU Munich, Germany and Department of Psychiatry and Psychotherapy, Ludwig-Maximilian-University Munich, Germany; **Andrea Pfennig**, MD, Department of Psychiatry and Psychotherapy, University Hospital Carl Gustav Carus, Medical Faculty, Technische Universität Dresden, Germany; **Claudia Pisanu**, MD, Department of Biomedical Sciences, University of Cagliari, Italy; **James B. Potash**, MD, MPH, Department of Psychiatry and Behavioral Sciences, Johns Hopkins University, USA; **Andreas Reif**, MD, Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, University Hospital Frankfurt, Germany; **Eva Reininghaus**, MD, Department of Psychiatry and Psychotherapeutic Medicine, Research Unit for bipolar affective disorder, Medical University of Graz, Austria; **Guy A. Rouleau**, MD, Montreal Neurological Institute and Hospital, McGill University, Canada; **Janusz K. Rybakowski**, MD, Department of Adult Psychiatry, Poznan University of Medical Sciences, Poland; **Martin Schalling**, MD, Department of Molecular Medicine and Surgery, Karolinska Institute, Sweden and Center for Molecular Medicine, Karolinska University Hospital, Stockholm, Sweden; **Peter R. Schofield**, PhD, DSc, Neuroscience Research Australia, Australia and School of Medical Sciences, University of New South Wales, Australia; **Barbara W. Schweizer**, RN, Department of Psychiatry and Behavioral Sciences, Johns Hopkins University, USA; **Giovanni Severino**, MD, Department of Biomedical Sciences, University of Cagliari, Italy; **Tatyana Shekhtman**, MSc, Department of Psychiatry, University of California San Diego, USA; **Paul D. Shilling**, PhD, Department of Psychiatry, University of California San Diego, USA; **Katzutaka Shimoda**, MD, Department of Psychiatry, Dokkyo Medical University School of Medicine, Japan; **Christian Simhandl**, MD, Bipolar Center Wiener Neustadt, Sigmund Freud University, Medical Faculty, Austria; **Claire M. Slaney**, RN, Department of Psychiatry, Dalhousie University, Canada; **Alessio Squassina**, PhD, Department of Biomedical Sciences, University of Cagliari, Italy; **Thomas Stamm**, MD, Department of Psychiatry and Psychotherapy, Charité - Universitätsmedizin Berlin, Campus Charité Mitte, Germany; **Pavla Stopkova**, MD, National Institute of Mental Health, Czech Republic; **Fasil Tekola-Ayele**, PhD, Epidemiology Branch, Division of Intramural Population Health Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, USA; **Alfonso Tortorella**, MD, Department of Psychiatry, University of Perugia, Italy; **Gustavo Turecki**, MD, Douglas Mental Health University Institute, McGill University, Canada; **Julia Veeh**, PhD, Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, University Hospital Frankfurt, Germany; **Eduard Vieta**, MD, PhD, Bipolar Disorder Program, Institute of Neuroscience, Hospital Clinic, University of Barcelona, IDIBAPS, CIBERSAM, Spain; **Stephanie H. Witt**, PhD, Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Germany; **Gloria Roberts**, PhD, School of Psychiatry, University of New South Wales, Australia; **Peter P. Zandi**, PhD, Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, USA; **Martin Alda**, MD, Department of Psychiatry, Dalhousie University, Canada; **Michael Bauer**, MD, PhD, Department of Psychiatry and Psychotherapy, University Hospital Carl Gustav Carus, Medical Faculty, Technische Universität Dresden, Germany;

Francis J. McMahon, MD, Intramural Research Program, National Institute of Mental Health, National Institutes of Health, US Department of Health & Human Services, USA; **Philip B. Mitchell**, MD, School of Psychiatry, University of New South Wales, Australia; **Thomas G. Schulze**, MD, Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Germany; Intramural Research Program, National Institute of Mental Health, National Institutes of Health, US Department of Health & Human Services, USA, Department of Psychiatry and Behavioral Sciences, Johns Hopkins University, USA; Institute of Psychiatric Phenomics and Genomics (IPPG), University Hospital, LMU Munich, Germany and Department of Psychiatry and Psychotherapy, University Medical Center (UMG), Georg-August University Göttingen, Germany; **Marcella Rietschel**, MD, Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Germany; **Scott R. Clark**, Discipline of Psychiatry, School of Medicine, University of Adelaide, Australia; **Bernhard T. Baune**, MD, PhD,, Department of Psychiatry and Psychotherapy, University of Münster, Germany; Department of Psychiatry, Melbourne Medical School, University of Melbourne, Australia and The Florey Institute of Neuroscience and Mental Health, The University of Melbourne, Australia

Correspondence: Bernhard Baune. Email: Bernhard.Baune@ukmuenster.de

First received 9 Jul 2021, final revision 12 Jan 2022, accepted 17 Jan 2022

Supplementary material

Supplementary material is available online at https://doi.org/10.1192/bjp.2022.28.

Data availability

All data used in analysis is available to ConLi+Gen members. See http://www.conligen.org/ for more information.

Author contributions

Micah Cearns analysed the data, trained all models and drafted the manuscript. Azmeraw Amare calculated the PRS variables. Bernhard Baune, Oliver Schubert, and Scott Clark acted as senior authors providing supervision and overall guidance in the drafting of the manuscript. All ConLi+Gen members contributed clinical and genetic data and provided overall feedback on the manuscript.

Funding

The primary sources of funding were grants RI 908/7-1, FOR2107 and RI 908/11-1 from the Deutsche Forschungsgemeinschaft (Marcella Rietschel) and grant NO 246/10-1 (Markus M. Nöthen) and grant ZIA-MH00284311 from the Intramural Research Program of the National Institute of Mental Health (ClinicalTrials.gov identifier: NCT00001174). The genotyping was funded in part by the German Federal Ministry of Education and Research through the Integrated Network IntegraMent (Integrated Understanding of Causes and Mechanisms in Mental Disorders), under the auspices of the e:Med Programme (Thomas G. Schulze, Marcella Rietschel and Markus M. Nöthen). This study was supported by National Institutes of Health grants P50CA89392 from the National Cancer Institute and 5K02DA021237 from the National Institute of Drug Abuse. The Canadian part of the study was supported by grant 64410 from the Canadian Institutes of Health Research (Martin Alda). Collection and phenotyping of the Australian University of New South Wales sample was funded by program grant 1037196 from the Australian National Health and Medical Research Council (Philip B. Mitchell, Peter R. Schofield, Janice M. Fullerton). The collection of the Barcelona sample was supported by grants PI080247, PI1200906, PI12/00018, 2014SGR1636, 2014SGR398, and MSII14/00030 from the Centro de Investigación en Red de Salud Mental, Institut d'Investigacions Biomèdiques August Pi i Sunyer, the Centres de Recerca de Catalunya Programme/Generalitat de Catalunya, and the Miguel Servet II and Instituto de Salud Carlos III. The Swedish Research Council, the Stockholm County Council, Karolinska Institutet and the Söderström-Königska Foundation supported this research through grants awarded to Lena Backlund, Louise Frisen, Catharina Lavebratt and Martin Schalling. The collection of the Geneva sample was supported by grants Synapsy–The Synaptic Basis of Mental Diseases 51NF40-158776 and 32003B-125469 from the Swiss National Foundation. The work by the French group was supported by INSERM (Institut National de la Santé et de la Recherche Médicale), AP-HP (Assistance Publique des Hôpitaux de Paris), the Fondation FondaMental (RTRS Santé Mentale), and the labex Bio-PSY (Investissements d'Avenir program managed by the ANR under reference ANR-11-IDEX-0004-02). The collection of the Romanian sample was supported by a grant from Unitatea Executiva pentru Finantarea Invatamantului Superior, a Cercetarii, Dezvoltarii si Inovarii (Maria Grigoroiu-Serbanescu). The collection of the Czech sample was supported by the project Nr. LO1611 with a financial support from the MEYS under the NPU I program and by the Czech Science Foundation, grant Nr. 17-07070S. Azmeraw T. Amare is supported by 2019–2021 National Alliance for Research on Schizophrenia and Depression (NARSAD) Young Investigator Grant from the Brain & Behaviour Research Foundation (BBRF) and National Health and Medical Research Council (NHMRC) Emerging Leadership Investigator Grant 2021 – 2008000

Declaration of interest

Eduard Vieta has received grants and served as consultant, advisor or CME speaker for the following entities: AB-Biotics, Allergan, Angelini, AstraZeneca, Bristol-Myers Squibb, Dainippon Sumitomo Pharma, Farmindustria, Ferrer, Forest Research Institute, Gedeon Richter, Glaxo-Smith-Kline, Janssen, Lundbeck, Otsuka, Pfizer, Roche, Sanofi-Aventis, Servier, Shire, Sunovion, Takeda, the Brain and Behaviour Foundation, the Spanish Ministry of Science and Innovation (CIBERSAM), and the Stanley Medical Research Institute. Michael Bauer has received grants from the Deutsche Forschungsgemeinschaft (DFG), and Bundesministeriums für Bildung und Forschung (BMBF), and served as consultant, advisor or CME speaker for the following entities: Allergan, Aristo, Janssen, Lilly, Lundbeck, neuraxpharm, Otsuka, Sandoz, Servier and Sunovion outside the submitted work. Sarah Kittel-Schneider has received grants and served as consultant, advisor or speaker for the following entities: Medice Arzneimittel Pütter GmbH and Shire. Bernhard Baune has received grants and served as consultant, advisor or CME speaker for the following entities: AstraZeneca, Bristol-Myers Squibb, Janssen, Lundbeck, Otsuka, Servier, the National Health and Medical Research Council, the Fay Fuller Foundation, the James and Diana Ramsay Foundation. Tadafumi Kato received honoraria for lectures, manuscripts, and/or consultancy, from Kyowa Hakko Kirin Co, Ltd, Eli Lilly Japan K. K., Otsuka Pharmaceutical Co, Ltd, GlaxoSmithKline K.K., Taisho Toyama Pharmaceutical Co, Ltd, Dainippon Sumitomo Pharma Co, Ltd, Meiji Seika Pharma Co, Ltd, Pfizer Japan Inc., Mochida Pharmaceutical Co, Ltd, Shionogi & Co, Ltd, Janssen Pharmaceutical K.K., Janssen Asia Pacific, Yoshitomiyakuhin, Astellas Pharma Inc, Wako Pure Chemical Industries, Ltd, Wiley Publishing Japan, Nippon Boehringer Ingelheim Co Ltd, Kanae Foundation for the Promotion of Medical Science, MSD K.K., Kyowa Pharmaceutical Industry Co, Ltd and Takeda Pharmaceutical Co, Ltd. Tadafumi Kato also received a research grant from Takeda Pharmaceutical Co, Ltd. Peter Falkai has received grants and served as consultant, advisor or CME speaker for the following entities Abbott, GlaxoSmithKline, Janssen, Essex, Lundbeck, Otsuka, Gedeon Richter, Servier and Takeda as well as the German Ministry of Science and the German Ministry of Health. Eva Reininghaus has received grants and served as consultant, advisor or CME speaker for the following entities: Janssen and Institut Allergosan. Mikael Landén declares that, over the past 36 months, he has received lecture honoraria from Lundbeck and served as scientific consultant for EPID Research Oy; no other equity ownership, profit-sharing agreements, royalties or patent. Kazufumi Akiyama has received consulting honoraria from Taisho Toyama Pharmaceutical Co, Ltd The other authors have no other conflict of interest to disclose.

References

1 Miura T, Noma H, Furukawa TA, Mitsuyasu H, Tanaka S, Stockton S, et al. Comparative efficacy and tolerability of pharmacological treatments in the maintenance treatment of bipolar disorder: a systematic review and network meta-analysis. *Lancet Psychiatry* 2014; **1**: 351–9.

2 Malhi GS, Tanious M, Das P, Berk M. The science and practice of lithium therapy. *Aust N Z J Psychiatry* 2012; **46**: 192–211.

3 Malhi GS, Adams D, Berk M. Is lithium in a class of its own? A brief profile of its clinical use. *Aust N Z J Psychiatry* 2009; **43**: 1096–104.

4 Cipriani A, Barbui C, Salanti G, Rendell J, Brown R, Stockton S, et al. Comparative efficacy and acceptability of antimanic drugs in acute mania: a multiple-treatments meta-analysis. *Lancet* 2011; **378**: 1306–15.

5 Yildiz A, Vieta E, Leucht S, Baldessarini RJ. Efficacy of antimanic treatments: meta-analysis of randomized, controlled trials. *Neuropsychopharmacology* 2011; **36**: 375–89.

6 Miller F, Tanenbaum JH, Griffin A, Ritvo E. Prediction of treatment response in bipolar, manic disorder. *J Affect Disord* 1991; **21**: 75–7.

7 Machado-Vieira R, Luckenbaugh DA, Soeiro-de-Souza MG, Marca G, Henter ID, Busnello JV, et al. Early improvement with lithium in classic mania and its association with later response. *J Affect Disord* 2013; **144**: 160–4.

8 Hou L, Heilbronner U, Degenhardt F, Adli M, Akiyama K, Akula N, et al. Genetic variants associated with response to lithium treatment in bipolar disorder: a genome-wide association study. *Lancet* 2016; **387**: 1085–93.

9 Schubert KO, Wisdom A. Should the Australian Therapeutic Goods Administration recommend rapid dosing of lithium carbonate in acute mania? *Aust N Z J Psychiatry* 2018; **52**: 387.

10 Kleindienst N, Engel R, Greil W. Which clinical factors predict response to prophylactic lithium? A systematic review for bipolar disorders. *Bipolar Disord* 2005; **7**: 404–17.

11 Tighe SK, Mahon PB, Potash JB. Predictors of lithium response in bipolar disorder. *Ther Adv Chronic Dis* 2011; **2**: 209–26.

12 Nunes A, Ardau R, Berghöfer A, Chillotti C, Deiana V, Garnham J, et al. Prediction of lithium response using clinical data. *Acta Psychiatr Scand* 2020; **141**: 131–41.

13 Amare AT, Schubert KO, Hou L, Clark SR, Papiol S, Heilbronner U, et al. Association of polygenic score for schizophrenia and HLA antigen and inflammation genes with response to lithium in bipolar affective disorder: a genome-wide association study. *JAMA Psychiatry* 2018; **75**: 65–74.

14 Antonucci LA, Pergola G, Pigoni A, Dwyer D, Kambeitz-Ilankovic L, Penzel N, et al. A Pattern of cognitive deficits stratified for genetic and environmental risk reliably classifies patients with schizophrenia from healthy control subjects. *Biol Psychiatry* 2020; **87**: 697–707.

15 Amare AT, Schubert KO, Hou L, Clark SR, Papiol S, Cearns M, et al. Association of polygenic score for major depression with response to lithium in patients with bipolar disorder. *Mol Psychiatry* 2020; **26**(6): 2457–70.

16 Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, Perlis RH, et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* 2013; **45**: 984.

17 Maier R, Moser G, Chen G-B, Ripke S, Absher D, Agartz I, et al. Joint analysis of psychiatric disorders increases accuracy of risk prediction for schizophrenia, bipolar disorder, and major depressive disorder. *Am J Hum Genet* 2015; **96**: 283–94.

18 Cearns M, Opel N, Clark S, Kaehler C, Thalamuthu A, Heindel W, et al. Predicting rehospitalization within 2 years of initial patient admission for a major depressive episode: a multimodal machine learning approach. *Transl Psychiatry* 2019; **9**: 285.

19 Schulze TG, Alda M, Adli M, Akula N, Ardau R, Bui ET, et al. The International Consortium on Lithium Genetics (ConLiGen): an initiative by the NIMH and IGSLI to study the genetic basis of response to lithium treatment. *Neuropsychobiology* 2010; **62**: 72–8.

20 Malhi GS, Bassett D, Boyce P, Bryant R, Fitzgerald PB, Fritz K, et al. Royal Australian and New Zealand College of Psychiatrists clinical practice guidelines for mood disorders. *Aust N Z J Psychiatry* 2015; **49**: 1087–206.

21 Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet* 2018; **50**: 668–81.

22 Consortium SWGotPG. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511**: 421–7.

23 Amare AT, Schubert KO, Klingler-Hoffmann M, Cohen-Woods S, Baune BT. The genetic overlap between mood disorders and cardiometabolic diseases: a systematic review of genome wide and candidate gene studies. *Transl Psychiatry* 2017; **7**: e1007.

24 Manchia M, Adli M, Akula N, Ardau R, Aubry JM, Backlund L, et al. Assessment of response to lithium maintenance treatment in bipolar disorder: a Consortium on Lithium Genetics (ConLiGen) report. *PLoS One* 2013; **8**: e65636.

25 Scott J, Etain B, Manchia M, Brichant-Petitjean C, Geoffroy PA, Schulze T, et al. An examination of the quality and performance of the Alda scale for classifying lithium response phenotypes. *Bipolar Disord* 2020; **22**: 255–65.

26 Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *Int J Meth Psych Res* 2011; **20**: 40–9.

27 Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Series B Stat Methodol* 2005; **67**: 301–20.

28 Biau G, Devroye L, Lugosi G. Consistency of random forests and other averaging classifiers. *J Mach Learn Res* 2008; **9**: 2015–33.

29 Cearns M, Hahn T, Clark S, Baune BT. Machine learning probability calibration for high-risk clinical decision-making. *Aust N Z J Psychiatry* 2020; **54**: 123–6.

30 Cearns M, Hahn T, Baune BT. Recommendations and future directions for supervised machine learning in psychiatry. *Transl Psychiatry* 2019; **9**: 1–12.

31 Flint C, Cearns M, Opel N, Redlich R, Mehler DMA, Emden D, et al. Systematic misestimation of machine learning performance in neuroimaging studies of depression. *Neuropsychopharmacol* 2021; **46**: 1510–7.

32 Dinga R, Marquand AF, Veltman DJ, Beekman AT, Schoevers RA, van Hemert AM, et al. Predicting the naturalistic course of depression from a wide range of clinical, psychological, and biological data: a machine learning approach. *Transl Psychiatry* 2018; **8**: 241.

33 Nunes A, Trappenberg T, Alda M, Genetics i. Asymmetrical reliability of the Alda Score favours a dichotomous representation of lithium responsiveness. *PloS One* 2020; **15**: e0225353.

34 Durrleman S, Simon R. Flexible regression models with cubic splines. *Stat Med* 1989; **8**: 551–61.

35 Sportiche S, Geoffroy PA, Brichant-Petitjean C, Gard S, Khan JP, Azorin JM, et al. Clinical factors associated with lithium response in bipolar disorders. *Aust N Z J Psychiatry* 2017; **51**: 524–30.

36 Nunes A, Stone W, Ardau R, Berghöfer A, Bocchetta A, Chillotti C, et al. Exemplar scoring identifies genetically separable phenotypes of lithium responsive bipolar disorder. *Transl Psychiatry* 2021; **11**: 36.

37 Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. *Cell* 2018; **173**: 1705–15.

38 Clark S, Baune B, Schubert K, Lavoie S, Smesny S, Rice SM, et al. Prediction of transition from ultra-high risk to first-episode psychosis using a probabilistic model combining history, clinical assessment and fatty-acid biomarkers. *Transl Psychiatry* 2016; **6**: e897.

39 Dwyer DB, Kalman JL, Budde M, Kambeitz J, Ruef A, Antonucci LA, et al. An investigation of psychosis subgroups with prognostic validation and exploration of genetic underpinnings: the PsyCourse Study. *JAMA Psychiatry* 2020; **77**: 523–33.

40 Ivleva EI, Clementz BA, Dutcher AM, et al.Arnold SJ, Jeon-Slaughter H, Aslan S, et al. Brain Structure biomarkers in the psychosis biotypes: findings from the bipolar-schizophrenia network for intermediate phenotypes. *Biol Psychiatry* 2017; **82**: 26–39.

41 Buhmann JM, Held M. Unsupervised learning without overfitting: Empirical risk approximation as an induction principle for reliable clustering. *International Conference on Advances in Pattern Recognition (Plymouth, UK 23–25 Nov 1999).* Springer, 1999 (https://link.springer.com/chapter/10.1007/978-1-4471-0833-7_17).

42 Martin J, Taylor MJ, Lichtenstein P. Assessing the evidence for shared genetic risks across psychiatric disorders and traits. *Psychol Med* 2018; **48**: 1759–74.

43 Kasiulevičius V, Šapoka V, Filipavičiūtė R. Sample size calculation in epidemiological studies. *Gerontologija* 2006; **7**: 225–31.