



King's Research Portal

DOI:

[10.1038/s41398-020-0738-5](https://doi.org/10.1038/s41398-020-0738-5)

Document Version

Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Fabbri, C., Kasper, S., Kautzky, A., Zohar, J., Souery, D., Montgomery, S., Albani, D., Forloni, G., Ferentinos, P., Rujescu, D., Mendlewicz, J., Uher, R., Lewis, C. M., & Serretti, A. (2020). A polygenic predictor of treatment-resistant depression using whole exome sequencing and genome-wide genotyping. *Translational psychiatry*, 10(1), Article 50. <https://doi.org/10.1038/s41398-020-0738-5>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

A polygenic predictor of treatment-resistant depression using whole exome sequencing and genome-wide genotyping

Chiara Fabbri ¹, Siegfried Kasper ², Alexander Kautzky ², Joseph Zohar ³, Daniel Souery ⁴, Stuart Montgomery ⁵, Diego Albani ⁶, Gianluigi Forloni ⁶, Panagiotis Ferentinos ⁷, Dan Rujescu ⁸, Julien Mendlewicz ⁹, Rudolf Uher ¹⁰, Cathryn M. Lewis ¹, Alessandro Serretti ¹¹

1: Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, United Kingdom

2: Department of Psychiatry and Psychotherapy, Medical University Vienna, Austria

3: Department of Psychiatry, Sheba Medical Center, Tel Hashomer, and Sackler School of Medicine, Tel Aviv University, Israel

4: Laboratoire de Psychologie Medicale, Université Libre de Bruxelles and Psy Pluriel, Centre Européen de Psychologie Medicale, Brussels

5: Imperial College School of Medicine, London, UK

6: Laboratory of Biology of Neurodegenerative Disorders, Neuroscience Department, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milan, Italy

7: Department of Psychiatry, Athens University Medical School, Athens, Greece

8: University Clinic for Psychiatry, Psychotherapy and Psychosomatic, Martin-Luther-University Halle-Wittenberg, Germany

9: Université Libre de Bruxelles

10: Department of Psychiatry, Dalhousie University, Halifax, Nova Scotia, Canada

11. Department of Biomedical and NeuroMotor Sciences, University of Bologna, Italy

Short title: exome sequencing in treatment-resistant depression

Corresponding author:

Alessandro Serretti, MD, PhD

Department of Biomedical and NeuroMotor Sciences, University of Bologna,

Viale Carlo Pepoli 5, 40123 Bologna, Italy

Tel +39 051 6584233

Fax +39 051 521030

Email: alessandro.serretti@unibo.it

Abstract

Treatment-resistant depression (TRD) occurs in ~30% of patients with major depressive disorder (MDD) but the genetics of TRD was previously poorly investigated.

Whole exome sequencing and genome-wide genotyping were available in 1209 MDD patients after quality control. Antidepressant response was compared to non-response to one treatment and non-response to two or more treatments (TRD). Differences in the risk of carrying damaging variants were tested. A score expressing the burden of variants in genes and pathways was calculated weighting each variant for its functional (Eigen) score and frequency. Gene- and pathway-based scores were used to develop predictive models of TRD and non-response using gradient boosting in 70% of the sample (training) which were tested in the remaining 30% (testing), evaluating also the addition of clinical predictors. Independent replication was tested in STAR*D and GENDEP using exome array-based data.

TRD and non-responders did not show higher risk to carry damaging variants compared to responders. Genes/pathways associated with TRD included those modulating cell survival and proliferation, neurodegeneration and immune response. Genetic models showed significant prediction of TRD vs. response and they were improved by the addition of clinical predictors, but they were not significantly better than clinical predictors alone. Replication results were driven by clinical factors, except for a model developed in subjects treated with serotonergic antidepressants, which showed a clear improvement in prediction at the extremes of the genetic score distribution in STAR*D.

These results suggested relevant biological mechanisms implicated in TRD and a new methodological approach to the prediction of TRD.

Keywords: treatment-resistant depression; exome sequencing; genome-wide association study; pathway; antidepressant; polygenic prediction; predictive modeling

1. Introduction

Major depressive disorder (MDD) is the second leading cause of disability in middle-aged adults on a global scale (1). Despite the availability of a number of different pharmacological treatments, treatment-resistant depression (TRD) is estimated to occur in ~30% of patients (2). TRD is usually defined as lack of response to at least two adequate treatments and it is associated with social and occupational impairment, suicidal thoughts, decline of physical health and increased health care utilization (3) (4). Annual costs for health care and lost productivity were estimated to be \$5,481 and \$4,048 higher, respectively, for a patient with TRD versus a patient with treatment-responsive depression (5).

In the future, biomarkers associated with TRD risk may contribute to improve the clinical management of MDD by providing an estimate of TRD genetic risk at baseline, by guiding the prescription of personalized treatments and the development of new drugs. Genetic variants are ideal biomarkers to predict treatment response and TRD: a genetic basis to treatment response has been demonstrated and genotyping can be performed in easily accessible samples with reasonable cost and time (6). The development of models able to predict the genetic risk of TRD at baseline would provide valuable information to personalize treatment prescription and hypothetically reduce the rate of TRD. Possible ways by which this could be achieved include: 1) identifying genetic predictors of non-response to specific antidepressant classes; 2) prescribing treatments with increased efficacy but limited availability because of costs constraints to patients having genetic risk for TRD. However, most existing pharmacogenomic studies were focused on measures of response to the last treatment without taking into account previous treatments, leaving the genetics of TRD largely unexplored (7). Another issue was the investigation of common variants only, while the possible role of rare variants was overlooked, despite they were suggested as one of the factors contributing to missing heritability of common traits (8). To the best of our knowledge, only a small pilot study (n=10) performed whole exome sequencing to the study of treatment response in MDD (but not TRD) and found that the bone morphogenetic protein (*BMP5*) gene may be associated with the therapeutic outcome (9).

The present study aimed to contribute in filling the existing gap in the knowledge of TRD genetics using whole exome sequencing and genome-wide genotyping to analyze the role of rare and common variants in the prediction of this phenotype and contribute to the development of predictive models potentially useful to personalize antidepressant prescription.

2. Patients and Methods

2.1. Sample

The Group for the Study of Resistant Depression (GSRD) sample was recruited within a multicenter, cross-sectional study including adult in- and outpatients with major depressive disorder (MDD) (DSM IV-TR criteria), as confirmed using the Mini International Neuropsychiatric Interview (MINI). Depressive symptom severity was assessed using the Montgomery and Åsberg Depression Rating Scale (MADRS) at study inclusion and at the onset of the current MDD episode. Information on previous and current antidepressant and other pharmacological treatments during the current MDD episode was collected as well as clinical-demographic characteristics. Antidepressant treatment was naturalistic according to best-clinical practice principles (Supplementary Table 1). The study protocol was approved by the local ethics committees and the participant signed the written informed consent. Further details can be found elsewhere (10).

2.2. Phenotype, training and testing samples

TRD (treatment-resistant depression) was defined according to the most common definition of lack of response to at least two adequate antidepressant treatments during the current depressive episode (11), while non-response was referred to one adequate treatment only. Adequate treatment was defined as an antidepressant treatment of minimum duration of four weeks at least at the minimum therapeutic dose according to drug labeling. Response was defined as a MADRS score < 22 and a score decrease of at least 50% compared to the onset of the current MDD episode. Responders could have had not more than one failed antidepressant treatment during the current depressive episode. After quality control, the sample was split in a training (70%) and testing set (30%) which were balanced in terms of phenotypic distribution (TRD, non-response and response) using the partition function of groupdata2 R package, and they did not differ for gender, age, baseline depression severity or centre of recruitment.

2.3. Whole exome sequencing and genome-wide genotyping

Whole exome sequencing was performed using the Illumina HiSeq platform with 100 bp read length. Genome-wide genotyping was performed using the Illumina Infinium PsychArray 24 BeadChip (Illumina, Inc., San Diego) and these data were imputed as described in Supplementary Methods. Rare variants were extracted from exome sequence data and were defined as those having minor allele frequency (MAF) $< 1/\sqrt{(2n)}$, where n is the sample size (12), which corresponded to 0.02 in GSRD. Information about DNA extraction, quality control of exome sequence data and genome-wide data are reported as Supplementary Methods. We compared the concordance of genotypes of SNPs

available in both exome sequence and array data, splitting them in genotyped and imputed and by minor allele frequency (MAF). These comparisons were also relevant to determine the putative reliability of rare imputed variants in the replication samples. Subjects with discrepancies between genome-wide and exome sequence data were excluded (non-major homozygote genotype concordance $\leq 90\%$ for rare variants and $\leq 95\%$ for common variants).

2.4. Statistical analysis

2.4.1. Variant annotation and distribution of functional variants

We tested if predicted detrimental/damaging variants obtained through exome sequencing were differently distributed between TRD patients, non-responders and responders. Variant annotation was performed using Variant Effect Predictor (Vep) release 90, using the `–pick` flag that chooses one block of annotation per variant, based on an ordered set of criteria (13). Annotations from SIFT, PolyPhen and functional consequence scores from the Sequence Ontology (SO) project were used to estimate the relative pathogenicity of variants (14) (15) (17). The use of scores which combine different variant annotations was also pursued and it is described in the next paragraph. The risk of carrying SIFT deleterious variants (scores <0.05), PolyPhen damaging or probably damaging variants (scores >0.45) and variants with SO functional score ≥ 0.90 and ≥ 0.70 in specific genes was compared across TRD patients, non-responders and responders using regression models adjusted for three population principal components and center of recruitment. Bonferroni correction was applied to account for multiple testing (the number of included genes was between 14,353 and 18,600 depending from the considered annotation). Additional details are reported as Supplementary Methods.

2.4.2. Exome risk scores

These analyses aimed to estimate a weighted measure reflecting the burden of rare genetic variants exome-wide and in a gene- and pathway-based way. Secondly, we combined these measures with analogous estimations for common variants.

For rare variants, a score was calculated for each individual as

$$\sum_{i=1}^n v_{all} \times w_s \times w_f$$

where n is the number of genetic variants within the considered unit (whole exome, gene or pathway), v_{all} is the number of alternative alleles, w_s is the corresponding functional score and w_f is the frequency weight for that variant. In this way, the score is not dependent from the presence of individual variants which could not be observed in some of the tested samples. A similar approach was previously applied to the study of schizophrenia risk using exome sequence data (18), but it was modified in this study by using different functional weighting (composite scores reflecting multiple annotations) and

different frequency weighting (to allow the inclusion of rare but also common variants). Different sources for determining w_s were tested and compared (Eigen scores (19), CADD scores (16) and SO functional scores (15), see Supplementary Methods). The frequency weight was determined using a beta distribution based on the frequency of the alternative allele alt_all ($w_f=dbeta(alt_all,1,25)$), according to the previous literature (12), see the corresponding curve in Supplementary Figure 1). Rare variants were extracted from exome sequence data as those with minor allele frequency (MAF $< 1/\sqrt{(2n)}$, where n is the sample size (12), which corresponded to 0.02 in GSRD). Common intragenic variants were extracted from genome-wide genotyping data and clumped based on their functional scores w_s and linkage disequilibrium (LD) using Plink v.1.9 (Supplementary Methods). A smoother beta distribution was used to weight these variants based on frequency ($w_f=dbeta(alt_all,0.5,0.5)$) (12), see curve in Supplementary Figure 1).

The obtained scores were tested for different distribution among the phenotypic groups considering rare variants only and the sum of the scores for rare and common variants. These tests were performed using regression models adjusted for three population principal components and centre of recruitment.

2.4.3. Predictive modeling

Gene- and pathway-based scores (adjusted for the described confounders, more details in Supplementary Methods) were entered into a predictor selection process in the training sample using a five-fold cross-validation repeated 100 times for pathways and 20 times for genes, 500 and 100 rounds in total, respectively. In each round, one fifth of the training dataset was left out, and in the remaining four-fifths of the training dataset a Correlation-Adjusted T (CAT) score was estimated (i.e. a multivariate generalization of the standard univariate T-test statistic that takes the correlation among variables explicitly into account (20) (21)) and the Local False Discovery Rate (LFDR) (i.e. the probability of a variable to be non-informative with regard to phenotype prediction given its CAT score) for each potential predictor. We selected predictors that had a LFDR smaller than 0.8 in $> 50\%$ of the rounds (22). This process reduces dimensionality and select variables with higher probability of being informative, reducing the risk of overfitting. These predictors were used to develop predictive models in the training sample using a gradient boosting machine (GBM) algorithm with a five-fold cross-validation repeated 100 and 20 times when predictors were pathway and gene scores, respectively. Cross-validation in this phase was used to provide better estimates of predictor contribution and empirically estimate model parameters (number of trees and interaction depth; shrinkage was set to 0.1 and minimum number of observations in each terminal node was set to 10). GBM produces a prediction model in the form of an ensemble of weak prediction models based on decision trees and it was demonstrated to be a suitable algorithm to learn from weak predictors, when

there is not a large amount of available data for training and predictors may interact among each other (23) (24). Models using gene-based scores as predictors included both rare and common variants, because the inclusion of rare variants only would have created scores very skewed towards zero which could not be realistically adjusted for confounders, while models using gene-set scores were tested for rare variants only and rare combined with common variants.

The performance of the developed models in predicting TRD or non-response in the testing sample was estimated using the area under the curve (AUC) of ROC (receiver operating characteristic) curves. Predictive models were developed in the whole training sample and in the subsamples treated with serotonergic antidepressants (5-HT ADs) and noradrenergic or noradrenergic-serotonergic antidepressants (NA ADs) according to the pharmacology domain reported in the NbN classification (Neuroscience based Nomenclature) (25). Different genetic profiles were indeed previously found for antidepressants belonging to these pharmacology domains (22). Only the current treatment was considered and subjects treated with combinations of 5-HT ADs and NA ADs were not included in this analysis (Supplementary Table 1). The addition of a clinical risk score to the genetic predictors was evaluated. The clinical risk score was calculated as a weighted sum of the variables independently associated with TRD or non-response in the training sample in a regression model after Bonferroni correction (Supplementary Table 2). Each variable included in the clinical score was weighted for its effect size (z score) and divided by the number of variables available in each subject ($\sum_{predictor=1}^n predictor \times z/n$) in order to avoid the exclusion of subjects with one or two missing values. We compared the ROC curves including genetic predictors with those including clinical or clinical-genetic predictors using the DeLong's method.

The risk of TRD or non-response may increase particularly at the extremes of the genetic score distribution. Thus, we also tested the significant models including only subjects with a genetic score ≤ 30 or ≥ 70 percentiles; we used this threshold to balance the risk of instability of findings due to the limited sample size, particularly in the subsamples treated with specific drug classes. The total genetic score was calculated in each subject as a sum of the gene/pathway scores included in the model of interest, each of them weighted for its importance in the predictive model. This approach is a simplification, since it does not reflect the non-linearity of the developed models and possible interactions.

We did not perform multiple-testing correction for these analyses because: 1) these tests were correlated among each other and not independent (for example, patients in the tails of the genetic score are a subset); 2) we looked at the consistency of results of correlated analyses (i.e. we analysed patients in the tails or added the clinical score for further testing models which showed non-random prediction in the basic test).

The following R cran packages were used for the described analyses: caret, nnet, sda, crossval, pROC.

2.5. Replication

Replication of the significant predictive models was tested in STAR*D and GENDEP (26) (27), using the same approach described for creating gene- and pathway-based risk scores (including rare and common variants according to the definition reported in paragraph 2.4.2, more details are in Supplementary Methods). In replication samples we used a genetic score ≤ 20 or ≥ 80 percentiles to identify subjects with extreme genetic scores since the larger sample size. In both these samples genome-wide genotyping was available, including standard genome-wide arrays and an exome array (Illumina Infinium Exome-24 v1.0 BeadChip) (28), but not exome sequence data. Further information on genotyping methods and quality control was previously reported (29) and it is described also in the Supplementary Methods. Imputation was carried out using the Michigan imputation server and the Haplotype Reference Consortium (HRC, version r1.1 2016) as reference panel (30). Different imputation quality thresholds were used to prune rare and common variants according to the previous literature ($R^2 > 0.30$ and $R^2 > 0.60$ for common and rare variants, respectively (31)(32)). The comparability between the available rare variants in GENDEP/STAR*D and GSRD was tested in terms of number and functional annotation. Phenotypes were defined in a way comparable to the GSRD sample, only TRD and response were considered because of their univocal phenotypic definition (part of non-responders are expected to become TRD) and these analyses aimed to replicate significant results in GSRD (which were concentrated to the comparison TRD vs. response). Further details on phenotype definition are reported in Supplementary Methods (paragraph “Replication samples: STAR*D and GENDEP”).

2.6. Power estimation

GSRD sample size after quality control ($n=1209$) provides adequate power (≥ 0.80) in 865 out of 1000 simulations when testing a set of 45 simulated rare variants ($MAF < 0.02$) and 100 simulated common variants (which reflects the median number of variants in the analysed genes), having effect sizes (β) randomly distributed between -0.25 and 0.25, at $\alpha=2.69e-06$ (Bonferroni corrected p-value for number of genes). R cran libraries KATSP, minqa and CompQuadForm were used for power estimation (33).

3. Results

The number of subjects available after quality control was 1209 (details on number of excluded subjects are in Supplementary Figure 2). A comprehensive description of the clinical-demographic

characteristics of the samples is reported in Supplementary Table 1, while a condensed overview is shown in Table 1. The number of included variants split by variant type and MAF is reported in Supplementary Table 3 (exome sequence data). Five subjects showed low concordance between genotypes available in both exome and genome-wide data and they were excluded from the analyses including both rare and common variants, since exome sequencing repeated on one of these subjects demonstrated genotype concordance >99% with the initial sequencing results. The comparison between sequenced rare variants and rare variants imputed from genome-wide data showed a mean concordance of 75% (SD=5%) considering only non-major homozygote genotypes. The mean concordance considering the same comparison but for genotyped rare variants (array data) was 93% (SD=2%) (Supplementary Figure 3), suggesting that the use of rare variants obtained from an array may be feasible even though not optimal. From the genome-wide data, 476,319 intragenic common variants in low LD and 1180 subjects were included after quality control.

The variables included in the clinical risk score were suicidal risk, number of previous depressive episodes, chronic depression and two MADRS factors (pessimism and interest-activity) (Supplementary Table 2).

3.1. Distribution of damaging variants

Patients with TRD and non-responders did not show an increased risk to carry SIFT/PolyPhen damaging variants compared to responders or variants with SO functional score ≥ 0.90 or ≥ 0.70 (Supplementary Table 4 and Figure 1). When considering individual genes (Supplementary Tables 5-6), we did not identify any difference among phenotypic groups after Bonferroni correction. The top gene was *WDR90* (WD Repeat Domain 90) which showed variants with SO functional score ≥ 0.90 in 21 patients with TRD but only in 4 non-responders and 2 responders ($p=3.44e-05$).

3.2. Exome-wide, gene and pathway scores

The distribution of the of exome-wide scores for the three tested functional weights were substantially overlapping. Six patients were excluded from the subsequent analyses as they scored outside five standard deviations from the sample mean (Supplementary Figure 4). Patients with TRD and non-responders did not show higher exome-wide scores compared to responders ($p>0.05$ for all three tested functional weights). The correlations between gene scores calculated using the three tested functional weights were high (mean correlation coefficient between 0.89 and 0.95 with SD from 0.04 to 0.06 in pair-wise comparisons, Supplementary Figure 5). In consideration of these high correlation coefficients, the demonstration that Eigen scores have better discriminatory ability using disease-associated and putatively benign variants from published studies compared to CADD scores (19), and

the lower functional precision of SO functional scores, only Eigen-based functional weighting was used in subsequent analysis.

Gene- and pathway-based scores were not associated with phenotypic groups after Bonferroni correction (Supplementary Tables 7-8). The top genes were *NBN* and *ZNF418* ($p=4.34e-05$ and $5.18e-05$, respectively, whole sample, Supplementary Table 4) and the top pathways were PID (protein interaction database) CD40 pathway in the subsample treated with serotonergic drugs and GO (gene ontology) response to cocaine in the subsample treated noradrenergic drugs ($p=5.28e-05$ and $5.61e-05$, respectively, Supplementary Table 8).

3.3. Predictive modeling

Pathway-based models for TRD vs. response in the whole sample including only rare genetic variants showed non-random prediction in the testing sample ($n=237$, AUC 0.61 [95% CI 0.54-0.69], Table 2 and Figure 2) and in patients treated with 5-HT ADs ($n=272$ and $n=118$ in the training and testing samples respectively, AUC 0.62 [95% CI 0.52-0.73], Table 2 and Figure 2). The list of pathways used as predictors is in Supplementary Table 9 and their relative contribution to the models in Supplementary Figure 6. No significant prediction of TRD vs. response was observed in patients treated with NA ADs or when comparing non-responders vs. responders or TRD plus non-responders vs. responders (Supplementary Table 10). Prediction was improved by adding the clinical risk score to genetic predictors in both the whole sample and patients treated with 5-HT ADs (AUC 0.73 [0.66-0.79] and AUC 0.65 [0.55-0.76], respectively, Table 2 and Figure 2), and this effect was more evident in subjects having extreme genetic scores for the included pathways ($n=142$, AUC 0.75 [0.67-0.83] and $n=71$, AUC 0.68 [0.55-0.82], respectively; Table 2 and Figure 2). However, there was no significant difference between the AUC obtained using the clinical risk score and that of the models including genetic and clinical predictors ($p=0.89$ and $p=0.68$ for the whole testing sample and for 5-HT ADs, respectively). The clinical risk score showed similar or better AUC compared to the models including genetic predictors alone ($p=0.03$ and $p=0.45$ for the whole testing sample and for 5-HT ADs, respectively). A possible interpretation of this finding can be found in the observation that patients in the 5-HT ADs group had a lower clinical risk score compared to the others ($p=9.73e-09$). Pathway-based models including rare and common genetic variants did not show predictive effect in the testing sample in almost all scenarios (Supplementary Table 10).

Gene-based models including rare and common variants predicted TRD vs. response in the whole testing sample and in subjects treated with 5-HT ADs ($n=230$, AUC 0.61, [0.53-0.69]; $n=113$, AUC 0.65 [0.55-0.76], respectively; Table 2). The lists of genes used as predictors is shown in Supplementary Table 9 and their relative contribution to the models in Supplementary Figure 6. The

addition of the clinical risk score improved the prediction while the subgroups having scores in the extreme percentiles did not show different results (Table 2). There was no significant difference between the AUC of the model including only clinical predictors and that of the models including genetic and clinical predictors ($p=0.74$ and $p=0.70$ for the whole testing sample and for 5-HT ADs, respectively). The clinical risk score showed similar or better AUC compared to the models including genetic predictors alone ($p=0.02$ and $p=0.50$ for the whole testing sample and for 5-HT ADs, respectively).

Predictive models of non-response vs. response showed marginal significance in the whole sample ($n=211$, AUC 0.59 [0.51-0.67]) but better values in the sample treated with 5-HT ADs ($n=121$, AUC 0.64, [0.53-0.74]; Supplementary Table 10). However, given that models including non-responders were significant in a smaller number of scenarios compared to those focused on TRD, we did not further investigate them in the replication samples.

3.5. Replication in STAR*D and GENDEP

Despite the availability of genotypes from an exome array, a low covering of coding regions was obtained compared to exome sequence data, limiting the comparability of these data with those available in GSRD (Supplementary Figure 7). In GENDEP, *LCE1B* gene was not covered and we had to re-train the corresponding predictive model (gene scores in patients treated with 5-HT ADs) without this gene, with no major change in predictive performance in the GSRD testing sample (not shown). The number of included subjects and their main clinical-demographic characteristics are reported in Supplementary Table 11.

None of the models including only genetic variables predicted TRD, apart from the rare variant pathway-based model developed in patients treated with 5-HT ADs. In STAR*D, this genetic model showed significant prediction of TRD risk in subjects with scores ≤ 10 or ≥ 90 percentiles (we looked at more extreme percentiles because of the larger sample size; $n=134$, AUC 0.73, 95% CI 0.61-0.86, Table 3). The AUC of this model was not different from that of the model including clinical and genetic variables ($p=0.63$), but it was better compared to the model including clinical variables only (AUC of the clinical predictor: 0.55 [0.49-0.62], $p=0.01$). The other models showing replication (all including genetic and clinical predictors) are reported in Table 3; the ROC AUC of these models were not significantly different from those of the models based on the clinical risk score (all $p > 0.05$). An overview of all replication results is provided in Supplementary Table 12.

4. Discussion

This study found no overall difference in the distribution of functional and deleterious/damaging variants between TRD patients, non-responders and responders within the whole exome or within individual genes. The closest gene to the significance threshold was *WDR90* (WD Repeat Domain 90), which product function is still poorly known but it is thought to participate in microtubule organization within the presynaptic axon terminal (34). The tested risk scores were not associated with TRD at gene or gene set level, with *NBN* (nibrin) and *ZNF418* (Zinc Finger Protein 418) genes, PID CD40 and GO response to cocaine pathways as top results. *NBN* is thought to be involved in DNA double-strand break repair, DNA damage-induced checkpoint activation and telomere integrity (35). It may be involved in neurodegenerative disorders (36). Variants in the *ZNF418* region had a non-significant trend of association with MDD in a previous PGC (Psychiatric Genetic Consortium) mega-analysis (37) and in an exome sequence study (38). The PID CD40 gene set includes 31 genes, it is involved in the modulation of inflammation and CD40 ligand has been previously associated with MDD (39).

The lack of strong signals coming from individual genes or pathways was expected as it is in line with a previous genome-wide association study of copy number variants (CNVs) that reported no significant enrichment of CNVs in TRD (40). Thus, it is reasonable to hypothesize that if genetics contributes to TRD, multiple genes/pathways must be involved with complex interactions. This mirrors the highly polygenic liability to MDD that is emerging from other studies (41). On the basis of this hypothesis, we applied predictive modeling to assess TRD risk using gene- and pathway genetic as well as clinical scores as predictors. Predictive modeling combining genetic and clinical predictors has been used by only two previous studies to predict antidepressant response to the best of our knowledge (22) (42), both these studies used SNPs from genome-wide genotyping as genetic predictors. In contrast to the present study, they did not perform any independent replication and the second study did not distinguish between training and testing sets (42).

The present study applied an innovative approach which combined gene and pathway polymorphisms in genetic scores weighted by their functional relevance, using exome sequence and genome-wide data. The predictive models comparing TRD vs. response showed significant prediction in a higher number of scenarios compared to models including non-responders, confirming the biological relevance of TRD as a distinct phenotype. In this regard, it should be noted that non-responders are a more heterogeneous group than TRD patients, because part of them is expected to develop TRD. In the GSRD testing sample, both gene- and pathway-based models showed significant prediction of TRD vs. response (Table 2). The genes/pathways included in these models (Supplementary Table 9) are mostly involved in cell survival, cell growth and replication, cell migration, neurodegenerative

processes, neuroplasticity, immune system, hormonal regulation (sex and thyroid hormones) and second messenger cascades. Predictive performance was often improved by adding clinical risk factors and in the extreme percentiles of the score distribution. However, none of the genetic or genetic-clinical models showed a significantly better ROC AUC compared to the model including the clinical risk score only. We hypothesized two possible scenarios which could make the genetic predictors useful: 1) in patients with no clinical risk factors; 2) in patients having genetic scores at the extremes of the distribution. We preliminary tested the first hypothesis in GSRD whole testing sample: the pathway-based model showed AUC of 0.67 (0.54-0.81) in patients with no clinical risk factors (n=64) vs. AUC=0.61 (0.54-0.69) in the whole testing sample. The number of patients was limited (for this reason we did not explore this hypothesis in other subsamples), but the result supports the hypothesis that our genetic predictors perform slightly worse in patients with clinical risk factors, presumably because they are largely independent from them (i.e. genetic factors are not able to predict TRD cases caused by clinical variables having a distinct genetic or environmental basis). In line with this, there was no correlation between the cumulative genetic score (for any model) and the clinical risk score and genetic models were not able to predict TRD classification according to the clinical risk score. We hypothesized that the high impact of clinical risk factors in GSRD (most patients were complex cases of MDD, recruited in tertiary health care centres) may have led to a relative down-weighting of genetic predictors in the clinical-genetic models (Supplementary Figure 6), explaining the fact that they did not show better performance in predicting TRD compared to the models including only the clinical risk score. We could not explore the contribution of the individual risk variables included in the risk score, because we used a cumulative score aimed to avoid the exclusion of subjects with partially missing data.

The fact that genetic models developed in patients treated with 5-HT AD had better AUC point estimates (Table 2) may be explained by the fact that these patients had significantly lower clinical risk factors compared to the others ($p=9.73e-09$), since treatment prescription was naturalistic in GSRD. This means that the different gene/pathways selected in the whole sample compared to those selected in patients treated with 5-HT ADs may reflect their different clinical characteristics rather than differences due to distinctive biological mechanisms implicated in response to different drug classes. None of the analyses performed in the group treated with NA ADs was significant, a probable consequence of the small size of this group. In this regard, we also underline that polypharmacy was frequent in this sample, including combination and augmentation strategies (10), thus our classification according to the antidepressants pharmacology represented a simplified approach.

The second scenario in which genetic predictors may be more relevant is in subjects with genetic risk scores at the extremes of the distribution. This case is exemplified by the clear improvement of

prediction in subjects with genetic scores ≤ 10 or ≥ 90 percentiles in STAR*D (Table 3), the largest available sample in our study, which allowed to test more extreme percentiles compared to GSRD and GENDEP (at least the top 5% of the distribution was suggested to be meaningful for increased risk when using polygenic risk scores (43), but we had no power for this). The corresponding model was the only one showing replication of genetic predictors only and superiority over the clinical risk score, while prediction in other models showing replication in STAR*D or GENDEP was driven by the clinical score. Unfortunately, the genetic data available in the replication samples were poorly comparable with those available in GSRD (only arrays, with low coverage of coding regions) and there were also clinical differences between STAR*D, GENDEP and GSRD. For example, patients in STAR*D had very long depressive episodes of relatively mild severity, while in GENDEP there were no patients with chronic MDD according to the standard definition (≥ 2 years) and they had on average a lower number of previous episodes (Supplementary Table 11). Unlike the other samples, MADRS was not available in STAR*D and equivalent scores were calculated using the QIDS-C16 scale (Supplementary Methods). The definition of the phenotype was performed slightly differently in each sample, because of the differences in study design.

Bearing in mind the discussed limitations, our results contributed to clarify the genetic factors involved in TRD and it was the first study to assess the contribution of rare genetic variants through whole exome sequencing, if we exclude a very small pilot study performed on 10 subjects (9). No individual gene or pathway probably plays a major role in TRD, thus models including multiple genes/pathways and able to account for their interactions are probably the best strategy. Theoretically, pathway-based models are more suitable to take into account the complex genetic component of antidepressant response compared to gene-based models and they are expected to be more replicable, as confirmed by our top replication results. Our study represents a new approach to the prediction of treatment resistance in MDD and future improvements in larger samples may lead to clinical applications, at least in patients with extreme genetic scores or those with no clinical risk factors. In patients having genetic risk for TRD, treatment strategies with demonstrated higher efficacy (e.g. pharmacotherapy combined with psychotherapy (44)) but limited availability for cost constraints could be implemented as first line treatment, when these patients first seek treatment and there are still no clinical signs of severe MDD and no clinical risk factors for TRD, reducing the proportion of patients at risk who progresses towards resistance.

Acknowledgements

We thank the NIMH for having had the possibility of analyzing their data on the STAR*D sample. We also thank the authors of previous publications in this dataset, and foremost, we thank the patients and their families who accepted to be enrolled in the study. Data and biomaterials were obtained from the limited access datasets distributed from the NIH-supported “Sequenced Treatment Alternatives to Relieve Depression” (STAR*D). The study was supported by NIMH Contract No. N01MH90003 to the University of Texas Southwestern Medical Center. The ClinicalTrials.gov identifier is NCT00021528.

The GENDEP project was supported by a European Commission Framework 6 grant (contract reference: LSHB-CT-2003-503428). The Medical Research Council, United Kingdom, and GlaxoSmithKline (G0701420) provided support for genotyping. This paper represents independent research part-funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King’s College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. High performance computing facilities were funded with capital equipment grants from the GSTT Charity (TR130505) and Maudsley Charity (980).

We thank Intomics (Copenhagen, Denmark) for genotype calling and contribution to quality control of exome sequence data in the GSRD sample.

Conflict of interest

Dr. Souery D. has received grant/research support from GlaxoSmithKline and Lundbeck; has served as a consultant or on advisory boards for AstraZeneca, Bristol-Myers Squibb, Eli Lilly, Janssen and Lundbeck. Prof. Montgomery S. has been a consultant or served on Advisory boards: AstraZeneca, Bristol Myers Squibb, Forest, Johnson & Johnson, Leo, Lundbeck, Medelink, Neurim, Pierre Fabre, Richter. Prof. Kasper S. received grants/research support, consulting fees and/or honoraria within the last three years from Angelini, AOP Orphan Pharmaceuticals AG, Celegne GmbH, Eli Lilly, Janssen-Cilag Pharma GmbH, KRKA-Pharma, Lundbeck A/S, Mundipharma, Neuraxpharm, Pfizer, Sanofi, Schwabe, Servier, Shire, Sumitomo Dainippon Pharma Co. Ltd. and Takeda. Prof. Zohar J. has received grant/research support from Lundbeck, Servier, Brainsway and Pfizer, has served as a consultant or on advisory boards for Servier, Pfizer, Abbott, Lilly, Actelion, AstraZeneca and Roche, and has served on speakers’ bureaus for Lundbeck, Roch, Lilly, Servier, Pfizer and Abbott. Prof. Mendlewicz J. is a member of the Board of the Lundbeck International Neuroscience Foundation and of Advisory Board of Servier. Prof. Serretti A. is or has been consultant/speaker for: Abbott, Abbvie, Angelini, Astra Zeneca, Clinical Data, Boheringer, Bristol Myers Squibb, Eli Lilly, GlaxoSmithKline,

Innovapharma, Italfarmaco, Janssen, Lundbeck, Naurex, Pfizer, Polifarma, Sanofi, Servier. The other authors declare no conflict of interest. Cathryn Lewis is a member of the Scientific Advisory Board of Myriad Neurosciences

Funding

Chiara Fabbri is supported by a Marie Skłodowska-Curie Actions Individual Fellowship funded by the European Community (EC Grant agreement number: 793526; project title: Exome Sequencing in stages of Treatment Resistance to Antidepressants - ESTREA).

Cathryn M. Lewis is part-funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. Dr Rudolf Uher is supported by the Canada Research Chairs Program.

This study was supported by an unrestricted grant from Lundbeck for the Group for the Study of Resistant Depression (GSRD). Lundbeck had no further role in the study design, in the collection, analysis, and interpretation of data, in the writing of the report, and in the decision to submit the paper for publication. All authors were actively involved in the design of the study, the analytical method of the study, the selection and review of all scientific content. All authors had full editorial control during the writing of the manuscript and approved it.

Data availability

Genetic and clinical data from STAR*D study can be obtained by submitting an application at: <https://www.nimhgenetics.org>. Genetic and clinical data from GENDEP can be obtained by submitting an enquiry to cathryn.lewis@kcl.ac.uk. Genetic and clinical data from GSRD is available to analysts of the Psychiatric Genomics Consortium (<https://www.med.unc.edu/pgc/>), for enquires contact the corresponding author (alessandro.serretti@unibo.it).

References

1. GBD 2015 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* 2016 08;388(10053):1545–602.
2. Souery D, et al. Treatment resistant depression: methodological overview and operational criteria. *Eur Neuropsychopharmacol J Eur Coll Neuropsychopharmacol*. 1999;9(1–2):83–91.
3. Trivedi MH, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry*. 2006;163(1):28–40.
4. Souery D, et al. Switching antidepressant class does not improve response or remission in treatment-resistant depression. *J Clin Psychopharmacol*. 2011;31(4):512–6.
5. Mrazek DA, Hornberger JC, Altar CA, Degtiar I. A review of the clinical, economic, and

- societal burden of treatment-resistant depression: 1996-2013. *Psychiatr Serv Wash DC*. 2014;65(8):977–87.
6. Tansey KE, et al. Contribution of Common Genetic Variants to Antidepressant Response. *Biol Psychiatry*. 2013;73(7):679–82.
 7. Fabbri C, et al. The Genetics of Treatment-Resistant Depression: A Critical Review and Future Perspectives. *Int J Neuropsychopharmacol*. 2018; 22(2):93-104.
 8. Gibson G. Rare and common variants: twenty arguments. *Nat Rev Genet*. 2012;13(2):135–45.
 9. Tammiste A, et al. Whole-exome sequencing identifies a polymorphism in the BMP5 gene associated with SSRI treatment response in major depression. *J Psychopharmacol*. 2013;27(10):915–20.
 10. Dold M, et al. Clinical correlates of augmentation/combination treatment strategies in major depressive disorder. *Acta Psychiatr Scand*. 2018;137(5):401–12.
 11. Souery D, et al. Clinical factors associated with treatment resistance in major depressive disorder: results from a European multicenter study. *J Clin Psychiatry*. 2007;68(7):1062–70.
 12. Ionita-Laza I, Lee S, Makarov V, Buxbaum JD, Lin X. Sequence kernel association tests for the combined effect of rare and common variants. *Am J Hum Genet*. 2013;92(6):841–53.
 13. McLaren W, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016;17(1):122.
 14. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4(7):1073–81.
 15. Adzhubei IA, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7(4):248–9.
 16. Kircher M, et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46(3):310–5.
 17. Koscielny G, et al. Open Targets: a platform for therapeutic target identification and validation. *Nucleic Acids Res*. 2017;45(D1):D985–94.
 18. Curtis D. Construction of an Exome-Wide Risk Score for Schizophrenia Based on a Weighted Burden Test. *Ann Hum Genet*. 2018;82(1):11–22.
 19. Ionita-Laza I, McCallum K, Xu B, Buxbaum JD. A spectral approach integrating functional genomic annotations for coding and noncoding variants. *Nat Genet*. 2016;48(2):214–20.
 20. Zuber V, Strimmer K. Gene ranking and biomarker discovery under correlation. *Bioinforma Oxf Engl*. 2009;25(20):2700–7.
 21. Zuber V, Strimmer K. High-dimensional regression and variable selection using CAR scores. Vol. *Statistical Applications in Genetics and Molecular Biology*. 2011.
 22. Iniesta R, et al. Antidepressant drug-specific prediction of depression treatment outcomes from genetic and clinical variables. *Sci Rep*. 2018;8(1):5530.
 23. Zhang Z, Zhao Y, Canes A, Steinberg D, Lyashevskaya O, written on behalf of AME Big-Data Clinical Trial Collaborative Group. Predictive analytics with gradient boosting in clinical medicine. *Ann Transl Med*. 2019;7(7):152.
 24. Ogutu JO, Piepho H-P, Schulz-Streeck T. A comparison of random forests, boosting and support vector machines for genomic selection. *BMC Proc*. 2011;5 Suppl 3:S11.
 25. Wilson S. Naming the drugs we use: neuroscience-based nomenclature, a helpful innovation. *Ther Adv Psychopharmacol*. 2018;8(6):171–2.
 26. Rush AJ, et al. Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. *Control Clin Trials*. 2004;25(1):119–42.
 27. Uher R, et al. Genome-wide pharmacogenetics of antidepressant response in the GENDEP project. *Am J Psychiatry*. 2010;167(5):555–64.
 28. Fabbri C, et al. New insights into the pharmacogenomics of antidepressant response from the GENDEP and STAR*D studies: rare variant analysis and high-density imputation. *Pharmacogenomics J*. 2018;18(3):413–21.
 29. Fabbri C, et al. Genome-wide association study of treatment-resistance in depression and

- meta-analysis of three independent samples. *Br J Psychiatry*. 2019;214(1):36–41.
30. Das S, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48(10):1284–7.
 31. Sidore C, et al. Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers. *Nat Genet*. 2015;47(11):1272–81.
 32. Pistis G, et al. Rare variant genotype imputation with thousands of study-specific whole-genome sequences: implications for cost-effective study designs. *Eur J Hum Genet*. 2015;23(7):975–83.
 33. Wu B, Pankow JS. On Sample Size and Power Calculation for Variant Set-Based Association Tests. *Ann Hum Genet*. 2016;80(2):136–43.
 34. Drew K, et al. Integration of over 9,000 mass spectrometry experiments builds a global map of human protein complexes. *Mol Syst Biol*. 2017;13(6):932.
 35. Cilli D, et al. Identification of the interactors of human nibrin (NBN) and of its 26 kDa and 70 kDa fragments arising from the NBN 657del5 founder mutation. *PloS One*. 2014;9(12):e114651.
 36. Katsel P, Tan W, Fam P, Purohit DP, Haroutunian V. Cell cycle checkpoint abnormalities during dementia: A plausible association with the loss of protection against oxidative stress in Alzheimer's disease [corrected]. *PloS One*. 2013;8(7):e68361.
 37. Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium, et al. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry*. 2013;18(4):497–511.
 38. Subaran RL, Odgerel Z, Swaminathan R, Glatt CE, Weissman MM. Novel variants in ZNF34 and other brain-expressed transcription factors are shared among early-onset MDD relatives. *Am J Med Genet Part B Neuropsychiatr Genet*. 2016;171B(3):333–41.
 39. Leo R, et al. Association between enhanced soluble CD40 ligand and proinflammatory and prothrombotic states in major depressive disorder: pilot observations on the effects of selective serotonin reuptake inhibitor therapy. *J Clin Psychiatry*. 2006;67(11):1760–6.
 40. O'Dushlaine C, et al. Rare copy number variation in treatment-resistant major depressive disorder. *Biol Psychiatry*. 2014;76(7):536–41.
 41. Howard DM, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci*. 2019;22(3):343–52.
 42. Lin E, et al. A Deep Learning Approach for Predicting Antidepressant Response in Major Depression Using Clinical and Genetic Biomarkers. *Front Psychiatry*. 2018;9:290.
 43. Lewis CM, Vassos E. Prospects for using risk scores in polygenic medicine. *Genome Med*. 2017;9(1):96.
 44. Cuijpers P, et al. Adding psychotherapy to antidepressant medication in depression and anxiety disorders: a meta-analysis. *World Psychiatry*. 2014;13(1):56–67.

Figure 1: representation of exome-wide distribution of variants with Sequence Ontology (SO) functional score ≥ 0.90 , SIFT deleterious variants, PolyPhen damaging/probably damaging variants. The examined phenotypic groups (x axis) were treatment-resistant depression (TRD), non-response and response. The number of variants in each phenotypic group is reported on the y axis.

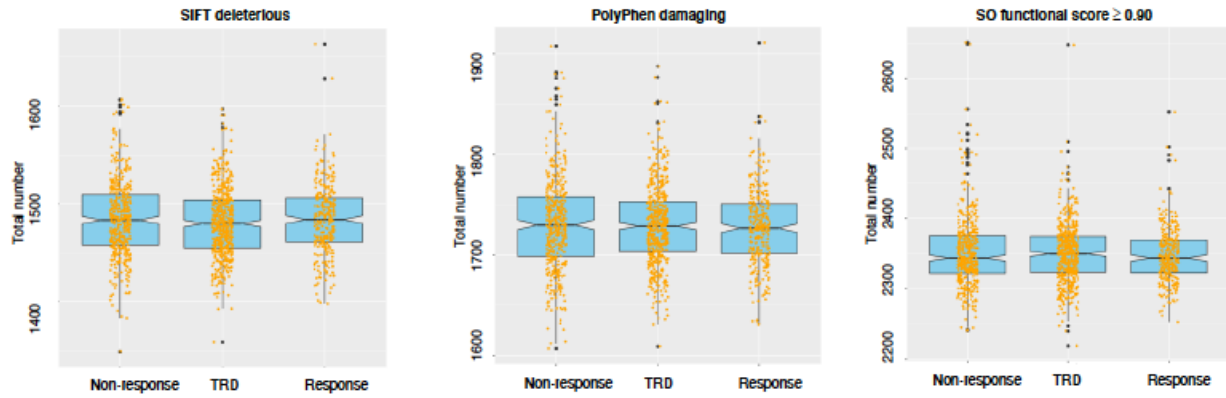
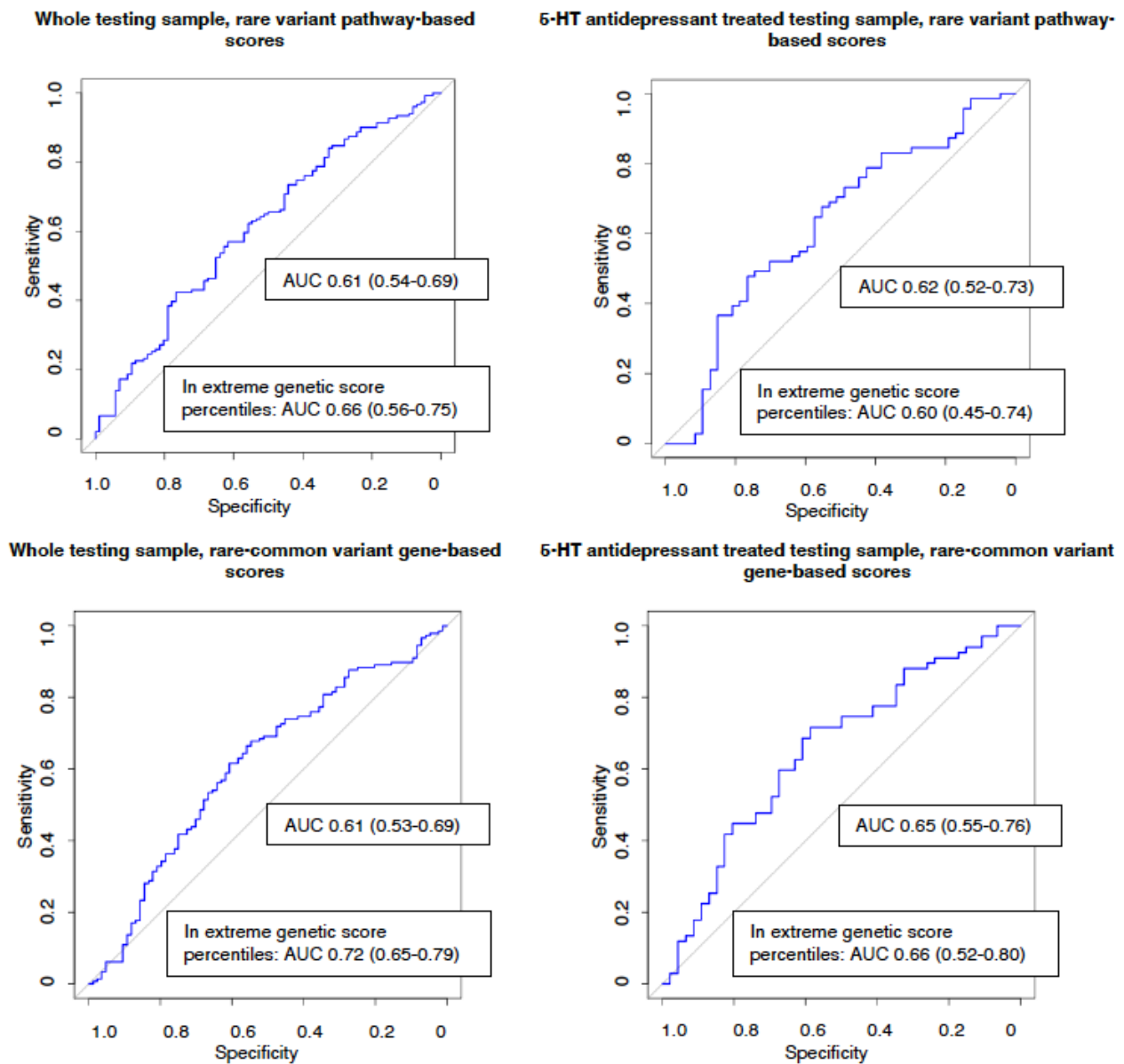


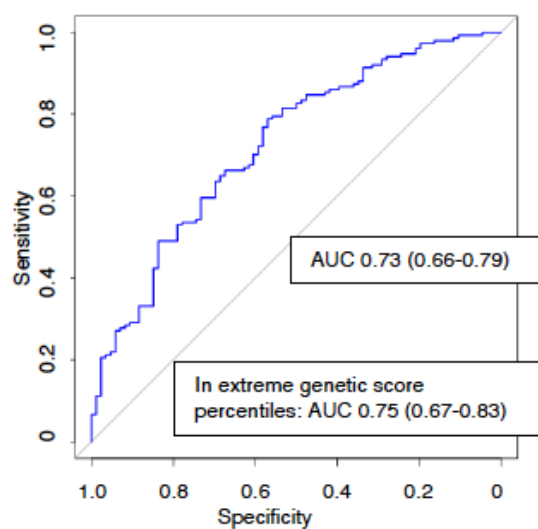
Figure 2: ROC curves of the non-random predictive models in GSRD testing sample and relative importance of the genetic predictors included in the models. When more than 20 predictors were included, only the first 20 are shown. 5-HT=serotonergic. The AUC values reached including only subjects with genetic scores ≤ 30 or ≥ 70 percentiles. **A.** Genetic predictors only. **B.** Genetic and clinical predictors.

A. Genetic predictors only

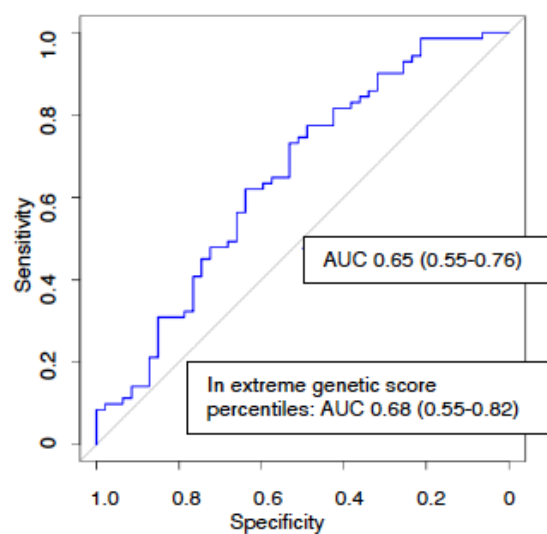


B. Genetic and clinical predictors

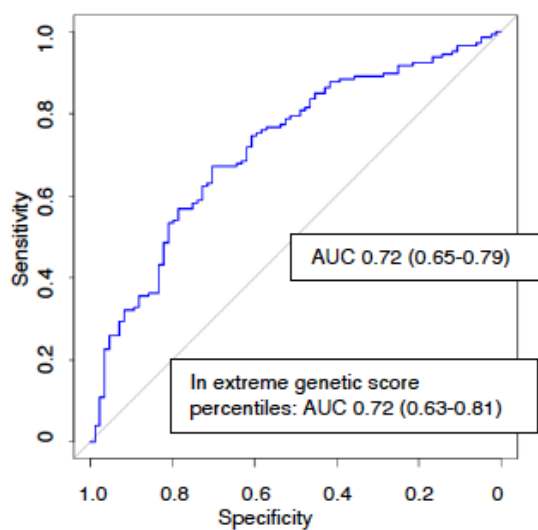
Whole testing sample, rare variant pathway-based scores



5-HT antidepressant treated testing sample, rare variant pathway-based scores



Whole testing sample, rare-common variant gene-based scores



5-HT antidepressant treated testing sample, rare-common variant gene-based scores

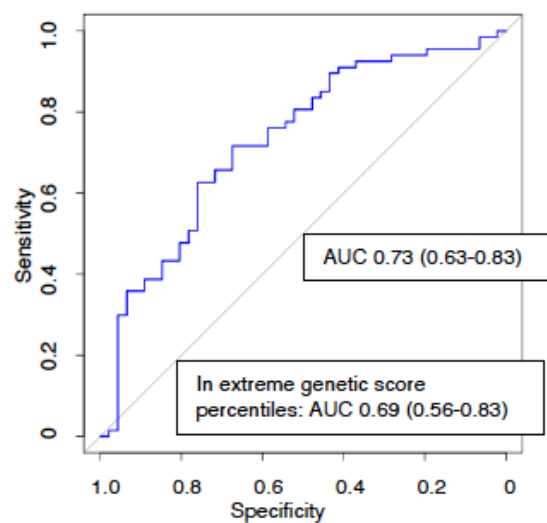


Table 1: main clinical-demographic characteristics of the training sample (n=847) and testing sample (n=362). The baseline MADRS score is referred to the beginning of the current depressive episode. MADRS=Montgomery and Åsberg Depression Rating Scale. TRD=treatment-resistant depression. Mean \pm standard deviation is reported for continuous variables and distribution for dichotomous ones. For a more comprehensive overview of patients' characteristics and results of comparisons between the characteristics of the two subsamples see Supplementary Table 1.

Variable	Training sample (n=847)	Testing sample (n=362)
Age	51.44 \pm 13.94	51.87 \pm 14.16
Gender (F/M)	566/281	235/127
Phenotype of interest	TRD n=353 Non-Responders n=291 Responders=203	TRD=151 Non-responders n=125 Responders=86
Baseline MADRS score	34.56 \pm 7.36	33.85 \pm 7.69
Current MADRS score	24.73 \pm 11.13	24.78 \pm 11.60
Treatment	Serotonergic n=421 Noradrenergic n=271 Serotonergic-noradrenergic n=128 Other n=27	Serotonergic n=192 Noradrenergic n=93 Serotonergic-noradrenergic n=59 Other n=18

Table 2: characteristics of the models showing significant prediction in the testing sample for the phenotype TRD vs. response. The results of the other tested models are shown in Supplementary Table 9. **A.** Pathway-based scores including rare variants; **B.** Gene-based scores including common and rare variants. Sens=sensitivity; spec=specificity; PPV=positive predictive value; NPV=negative predictive value. 5-HT=serotonergic drugs. AUC 95% confidence intervals are reported within parenthesis. * including only subjects with a genetic score ≤ 30 percentile or ≥ 70 percentile see paragraph 2.4.3 of the main manuscript for more details.

A

Subsample	Genetic predictors only	Genetic predictors + clinical risk score	Extreme genetic percentiles*, genetic predictors only	Extreme genetic percentiles*, genetic predictors + clinical score
Whole testing set	AUC 0.61 (0.54-0.69) Sens=0.42; spec=0.77; PPV=0.76; NPV=0.43	AUC 0.73 (0.66-0.79) Sens=0.79; spec=0.57; PPV=0.76; NPV=0.60	AUC 0.66 (0.56-0.75) Sens=0.53; spec=0.76; PPV=0.78; NPV=0.51	AUC 0.75 (0.67-0.83) Sens=0.79; spec=0.60; PPV=0.76; NPV=0.65
5-HT drugs only	AUC 0.62 (0.52-0.73) Sens=0.48; spec=0.77; PPV=0.76; NPV=0.49	AUC 0.65 (0.55-0.76) Sens=0.73; spec=0.53; PPV=0.70; NPV=0.57	AUC 0.60 (0.45-0.74)	AUC 0.68 (0.55-0.82) Sens=0.75; spec=0.61; PPV=0.75; NPV=0.61

B

Subsample	Genetic predictors only	Genetic predictors + clinical risk score	Extreme genetic percentiles*, genetic predictors only	Extreme genetic percentiles*, genetic predictors + clinical score
Whole testing set	AUC 0.61 (0.53-0.69) Sens=0.68; spec=0.55; PPV=0.72; NPV=0.49	AUC 0.72 (0.65-0.79) Sens=0.67; spec=0.70; PPV=0.80; NPV=0.55	AUC 0.61 (0.51-0.71) Sens=0.54; spec=0.70; PPV=0.72; NPV=0.52	AUC=0.72 (0.63-0.81) Sens=0.57; spec=0.82; PPV=0.82; NPV=0.57
5-HT drugs only	AUC 0.65 (0.55-0.76) Sens=0.72; spec=0.59; PPV=0.72; NPV=0.59	AUC 0.73 (0.63-0.83) Sens=0.72; spec=0.67; PPV=0.76; NPV=0.62	AUC 0.66 (0.52-0.80) Sens=0.76; spec=0.59; PPV=0.74; NPV=0.62	AUC 0.69(0.56-0.83) Sens=0.70; spec=0.67; PPV=0.76; NPV=0.60

Table 3: best predictive models of treatment-resistant depression (TRD) vs. response in the replication samples. For a detailed overview of all results in the replication samples see Supplementary Table 12. In STAR*D, more extreme percentiles of the genetic predictors were considered compared to other samples because of the larger sample size (for details see paragraph 3.5). 5-HT antidepressants=serotonergic antidepressants. CI=confidence intervals.

Sample	Genetic predictors	Genetic predictors AUC (95% CI)	Clinical and genetic predictors AUC (95% CI)
GENDEP, whole sample (n=321)	Pathways, rare variants	0.54 (0.47-0.60)	0.60 (0.54-0.65)
GENDEP, 5-HT antidepressants (n=188)	Genes, rare and common variants	0.58 (0.49-0.68)	0.62 (0.53-0.72)
STAR*D, whole sample (n=807)	Genes, rare and common variants	0.51 (0.46-0.55)	0.55 (0.51-0.59)
STAR*D, 5-HT antidepressants, ≤ 20 or ≥ 80 percentiles (n=266)	Pathways, rare variants	0.59 (0.48-0.69)	0.61 (0.51-0.71)
STAR*D, 5-HT antidepressants, ≤ 10 or ≥ 90 percentiles (n=134)	Pathways, rare variants	0.73 (0.61-0.86)	0.72 (0.58-0.86)