Pharmacogenomics-Driven Prediction of Antidepressant Treatment Outcomes: A Machine-Learning Approach With Multi-trial Replication

Arjun P. Athreya[1,2], Drew Neavin[2], Tania Carrillo-Roa[3], Michelle Skime[4], Joanna Biernacka[4], Mark A. Frye[4], A. John Rush[5,6,7], Liewei Wang[2], Elisabeth B. Binder[3,8], Ravishankar K. Iyer[1], Richard M. Weinshilboum[2] and William V. Bobo[9,*]

We set out to determine whether machine learning–based algorithms that included functionally validated pharmacogenomic biomarkers joined with clinical measures could predict selective serotonin reuptake inhibitor (SSRI) remission/response in patients with major depressive disorder (MDD). We studied 1,030 white outpatients with MDD treated with citalopram/escitalopram in the Mayo Clinic Pharmacogenomics Research Network Antidepressant Medication Pharmacogenomic Study (PGRN-AMPS; n = 398), Sequenced Treatment Alternatives to Relieve Depression (STAR*D; n = 467), and International SSRI Pharmacogenomics Consortium (ISPC; n = 165) trials. A genomewide association study for PGRN-AMPS plasma metabolites associated with SSRI response (serotonin) and baseline MDD severity (kynurenine) identified single nucleotide polymorphisms (SNPs) in *DEFB1*, *ERICH3*, *AHR*, and *TSPAN5* that we tested as predictors. Supervised machine-learning methods trained using SNPs and total baseline depression scores predicted remission and response at 8 weeks with area under the receiver operating curve (AUC) > 0.7 (P < 0.04) in PGRN-AMPS patients, with comparable prediction accuracies > 69% (P ≤ 0.07) in STAR*D and ISPC. These results demonstrate that machine learning can achieve accurate and, importantly, replicable prediction of SSRI therapy response using total baseline depression severity combined with pharmacogenomic biomarkers.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?
☑ In light of the phenotypic complexity of antidepressant response, social and demographic factors alone are insufficient to determine, prior to treatment initiation, whether selective serotonin reuptake inhibitors (SSRIs) will be effective in patients with major depressive disorder (MDD).

WHAT QUESTION DID THIS STUDY ADDRESS?
☑ This study tests the hypothesis that functionally validated single nucleotide polymorphisms (SNPs) associated with SSRI pharmacodynamics as predictor variables can enhance our ability to predict the response of patients with MDD to SSRI therapy.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?
☑ Pharmacogenomic SNPs' predictive capabilities potentially allow for guiding of clinical decisions relating to SSRI response.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?
☑ Alternative medication strategies can be considered by physicians if predictive models prior to SSRI initiation, using pharmacogenomic SNPs associated with MDD pathophysiology or SSRI response, forecast poor response.

Major depressive disorder (MDD) is the leading cause of disability related to chronic illnesses worldwide.[1] Selective serotonin reuptake inhibitors (SSRIs) are first-line pharmacotherapies for MDD, but only about half to two-thirds of patients respond to SSRI therapy, and several weeks of treatment must occur before an optimal therapeutic response is achieved.[2] Therefore, the ability

[1]Department of Electrical & Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA; [2]Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, Minnesota, USA; [3]Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany; [4]Department of Psychiatry & Psychology, Mayo Clinic, Rochester, Minnesota, USA; [5]Department of Psychiatry & Behavioral Sciences, Department of Medicine, Duke Institute of Brain Sciences, Duke University School of Medicine, Durham, North Carolina, USA; [6]Texas Tech University Health Sciences Center, Permian Basin, Texas, USA; [7]Duke-National University of Singapore Medical School, Singapore, Singapore; [8]Department of Psychiatry & Behavioral Sciences, Emory University School of Medicine, Atlanta, Georgia, USA; [9]Department of Psychiatry & Psychology, Mayo Clinic, Jacksonville, Florida, USA. *Correspondence: William Bobo (bobo.william@mayo.edu)

to identify patients with MDD who are most likely to respond to SSRI antidepressants before starting treatment (or soon after treatment initiation) would represent a significant therapeutic advance.

Statistical/machine-learning approaches have demonstrated that predictions obtained using clinical and sociodemographic factors can yield predictive performances for SSRI response (area under the receiver operating curve (AUC) of 0.54–0.67) that are significantly better than chance.[3–6] However, social and demographic factors alone have proven insufficient to individualize therapeutic decisions for depressed patients. The reason is that MDD is a heterogeneous disease (i.e., depression symptoms, such as sleep, mood, and appetite) and treatment outcomes vary greatly among patients, and social and demographic factors are often not consistently associated with either disease

severity or outcomes.[7] A limitation acknowledged by the authors of prior reports is the lack of inclusion of biological factors associated with depression severity or therapeutic response to antidepressants in the prediction models.[3–5,8] To address this limitation, recent machine-learning approaches have used genomics and/or metabolomics data to predict SSRI response, with AUC values of 0.68–0.78.[8–10] Although the studies demonstrated the feasibility of integrating biological factors with machine learning to achieve improved prediction performance, they were limited by lack of replication across multiple trials and multiple depression rating scales, in addition to lack of significance from a pharmacogenomics perspective.[8–11]

In the present study, we used a machine-learning workflow (depicted schematically in **Figure 1**) to study the capabilities of
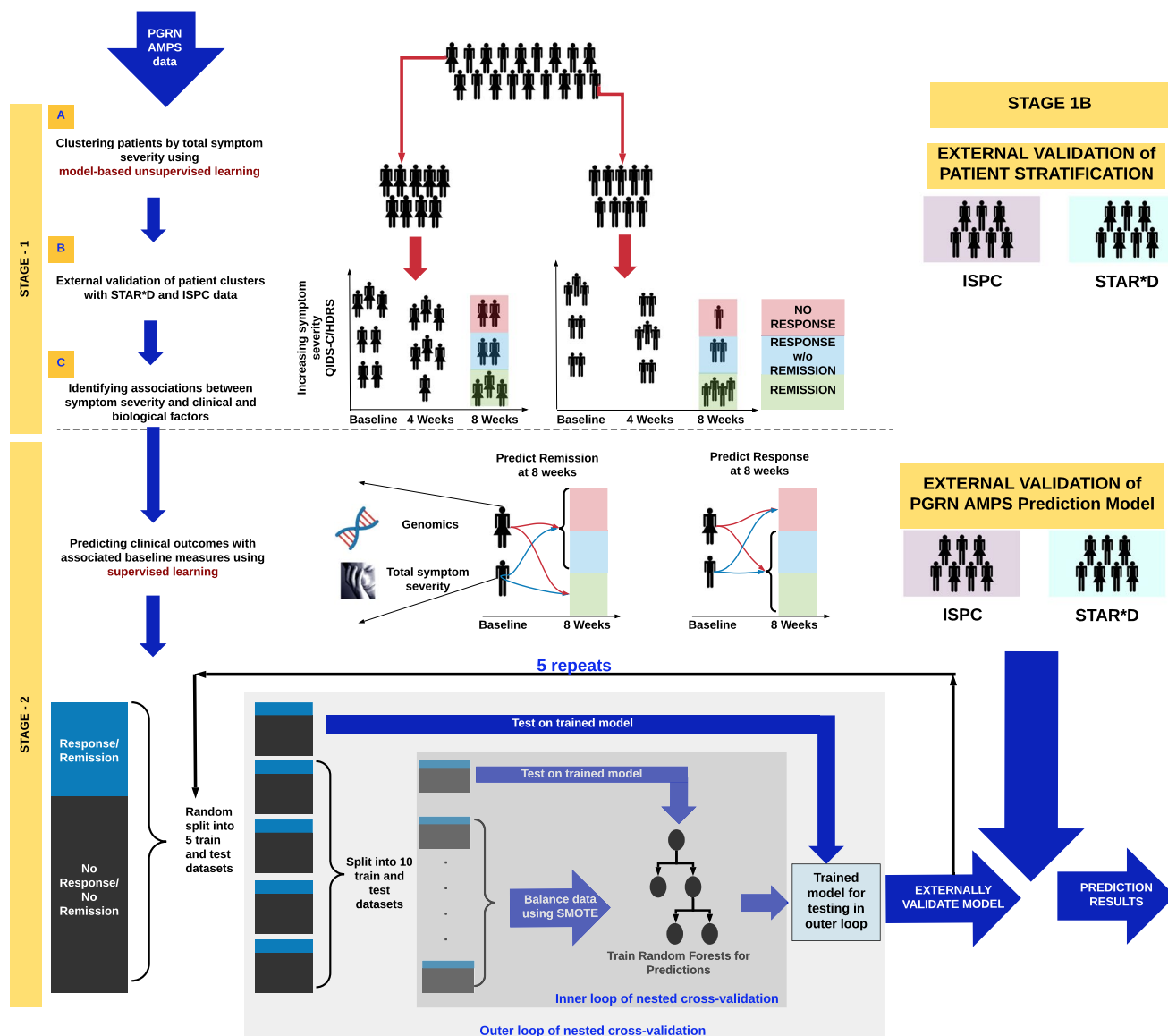

Figure 1 The two-stage analysis workflow. Our analysis workflow proceeded in two stages. In stage 1, we identified depressive symptom severity clusters in the Mayo Clinic Pharmacogenomics Research Network Antidepressant Medication Pharmacogenomic Study (PGRN-AMPS) dataset, separately for men and women, using a data-driven approach (stage 1A); we then validated those clusters using data from Sequenced Treatment Alternatives to Relieve Depression (STAR*D) and International SSRI Pharmacogenomics Consortium (ISPC; stage 1B). Factors that differentiated the validated depressive symptom clusters were identified in stage 1C. In stage 2, predictive models were developed using PGRN-AMPS data and were externally validated using STAR*D and ISPC data. HDRS, Hamilton Depression Rating Scale; QIDS-C, Quick Inventory of Depressive Symptomatology; SMOTE, Synthetic Minority Oversampling Technique.

functionally validated single nucleotide polymorphisms (SNPs) associated with SSRI pharmacodynamics, combined with clinical data, to predict SSRI response. For this study, we (i) algorithmically grouped (clustered) patients with MDD in the Mayo Clinic Pharmacogenomics Research Network Antidepressant Medication Pharmacogenomic Study (PGRN-AMPS) trial, using an unsupervised learning approach; (ii) predicted remission/response to citalopram/escitalopram treatment using supervised machine-learning methods that considered clinical and pharmacogenomic data from the PGRN-AMPS trial as predictor variables; and (iii) externally validated these patient clusters and statistical/machine-learning models using data from the Sequenced Treatment Alternatives to Relieve Depression (STAR*D[12]) and the International SSRI Pharmacogenomics Consortium (ISPC[13]) datasets. These analyses are sex-stratified, given the established evidence of sex differences in the prevalence of depression and increasing evidence of sex differences in antidepressant response. [11, 14–21]

The pharmacogenomic biomarkers included in the present study are six SNPs in or near the *TSPAN5* (rs10516436), *ERICH3* (rs696692), *DEFB1* (rs5743467, rs2741130, and rs2702877), and *AHR* (rs17137566) genes. Each of these SNPs were the "top" SNP in its respective genomewide association study (GWAS) SNP signal, except that for *DEFB1*, we included the "top" SNP as well as two others in different haplotype blocks. The GWAS had used as phenotypes plasma serotonin and kynurenine concentrations assayed in PGRN-AMPS samples.[22,23] Of the plasma metabolites assayed in those patients, serotonin and kynurenine concentrations were the most highly associated with SSRI outcomes at 8 weeks[22] or with baseline depressive symptom severity—one of the most important predictors of eventual antidepressant treatment response.[23] The application of a research strategy that involved the use of metabolomics to "guide" genomics represented a step toward the inclusion of biological data (i.e., metabolite concentrations associated with outcomes), in an effort to move beyond the traditional rating scales used in psychiatry (e.g., the Hamilton Depression Rating Scale (HDRS) and the Quick Inventory of Depressive Symptomatology (QIDS-C)). Subsequent functional genomic studies showed that knockdown of both TSPAN5 and ERICH3 in neuronally derived cell lines resulted in decreased serotonin in the cell culture media, and that alterations in the expression of both DEFB1 and AHR

could influence kynurenine biosynthesis as well as the effects of mediators of inflammation,[9,10] a process that has been shown to play an important role in MDD pathophysiology.[24–26]

RESULTS
Total depressive symptom severity clusters
We first observed that the distribution of total depression severity scores (**Figure 2a**) comprised multiple Gaussian distributions (**Figure 2b**). Our unsupervised learning approach inferred subgroups of patients (clusters) based on which Gaussian distribution their score belonged to, and algorithmically identified three distinct clusters for both men and women ($P < 1.3E-09$; the significance level for the test is 0.05/3, because we are comparing differences between 3 distributions at each time point) in PGRN-AMPS based on their total depression scores at each time point. The distribution of scores representing each of the three clusters at baseline (A1, A2, and A3), after 4 weeks (B1, B2, and B3), and 8 weeks (C1, C2, and C3) of SSRI treatment is illustrated in **Figure 3**. For the cluster assignments (shown in **Figure 3**), the letters (e.g., A, B, and C) represent the treatment time points, and the numeric suffix at each time point represents the level of depression severity, with "3" being the most severely depressed subjects, "1" being mild depression, and "2" being moderate levels of depression.

Because treatment outcomes are defined after 8 weeks of SSRI treatment (see Methods), we were interested in how the clusters might relate to standard definitions of treatment outcomes (e.g., remission or response). In both men and women, C1 included all patients who achieved remission (i.e., their HDRS or QIDS-C total scores were ≤7 or ≤5, respectively). Eighty-seven percent of patients in C2 achieved response (the remaining patients were nonresponders), defined as a decrease in either the HDRS or QIDS-C total score of at least 50% but without achieving remission. Multivariate clustering on individual depressive item scores for both scales did not yield three clusters at 8 weeks that conformed to accepted definitions of response or remission (**Figure S1**).

After we applied the same unsupervised learning approach to the STAR*D (QIDS-C–measured severity) and ISPC (HDRS-measured severity) datasets, three clusters of men and women were again identified at all time points. These clusters did not differ statistically ($P > 0.1$) from those inferred in PGRN-AMPS,
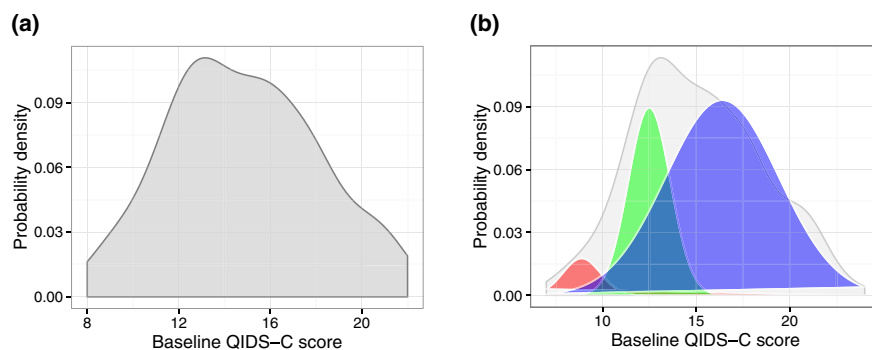
Figure 2 Probability density functions (PDFs) of depression severity scores. Baseline Quick Inventory of Depressive Symptomatology (QIDS-C) symptom severity scores in men (**a**), and the estimated components of the PDF using an expectation-maximization algorithm (**b**). [Colour figure can be viewed at wileyonlinelibrary.com]
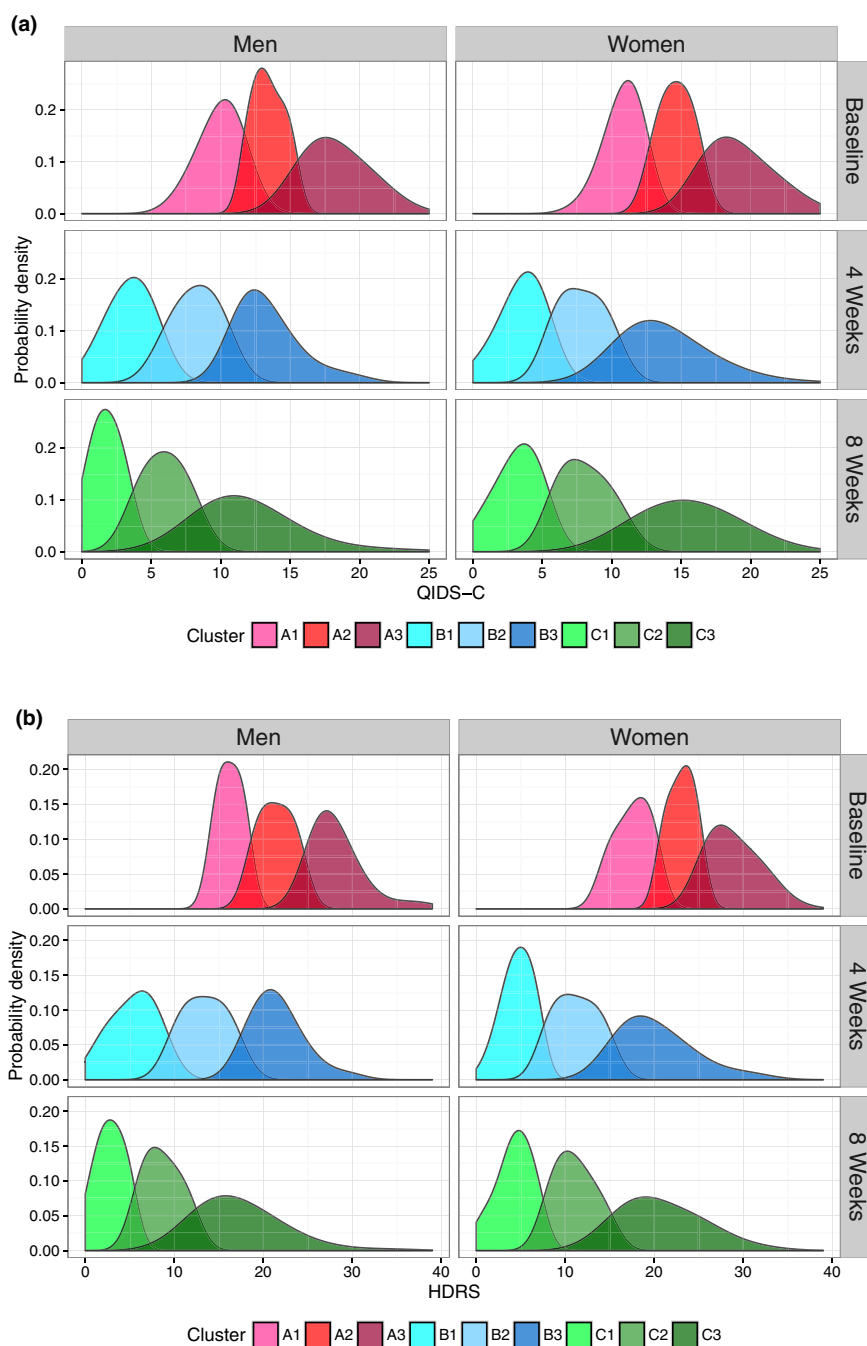
Figure 3 Depressive symptom–based clusters identified by data-driven unsupervised learning using Gaussian mixture models. Probability densities of symptom severity in clusters at baseline, 4 weeks, and 8 weeks of the Mayo Clinic Pharmacogenomics Research Network Antidepressant Medication Pharmacogenomic Study trial for both the Quick Inventory of Depressive Symptomatology (QIDS-C) (**a**) and Hamilton Depression Rating Scale (HDRS) (**b**) scales. Probability densities are proportional to the fraction of patients with the associated symptom severity scores.

providing external validation. As observed in PGRN-AMPS, the 8-week clusters (C1, C2, and C3) identified in STAR*D and ISPC conformed to accepted clinical definitions of remission, response (without remission), and nonresponse, respectively, for both depression rating scales. These externally validated clusters allowed us to identify associations of depression severity with the clinical and demographic factors listed in **Table S1**.

Association of clinical and demographic factors, cytochrome P450 *2C19* metabolizer phenotypes, and plasma drug levels with severity-based clusters

For citalopram-treated or escitalopram-treated PGRN-AMPS patients across different drug dosages after 4 and 8 weeks of treatment, and across all three clusters for both men and women at any time point (*P* > 0.1, **Figures S2–S4**), there were no significant

differences in the distributions of any of the clinical or demographic factors listed in **Table S1**, in drug dosing, or in plasma drug levels.

Given the lack of associations of clinical/demographic factors or cytochrome P450 (*CYP*)*2C19* metabolizer phenotypes with depression severity clusters at baseline or at 8 weeks, we focused on testing the capability of pharmacogenomic SNP biomarkers combined with baseline depression severity to predict remission (i.e., patients found in cluster C1 at 8 weeks) or response, regardless of the baseline cluster in which they began treatment. We trained prediction models stratified by sex for each rating scale.

Response/remission prediction performance

Prediction performance using only sociodemographic factors. In our prior work,[9] the accuracy (percent of correctly predicted outcomes) and AUC when only depression severity (QIDS-C or HDRS) scores, together with social and demographic factors, were used as predictors and were 48–55% and 0.54–0.67%, respectively. We later compared those results with the prediction performances of classifiers that used both baseline depression severity and pharmacogenomic SNP data.

Training performance using PGRN-AMPS data. In PGRN-AMPS (for which we used nested cross-validation to train the prediction models), baseline depression severity combined with pharmacogenomic biomarkers predicted sex-specific response and remission status with accuracies of 73–88% ($P \leq 0.01$; AUC 0.7–0.9) and 71–86% ($P \leq 0.04$; AUC 0.75–0.9), respectively (**Table 1**; the ranges represent results for both sexes). When the *CYP2C19* metabolizer phenotype was included as a predictor variable, the prediction accuracies were reduced by $\geq 4\%$ for remission and response in both sexes and both scales ($P > 0.3$).

Top predictor variables during training. We next evaluated the contribution of each of the SNP biomarkers and baseline HDRS and QIDS-C scores to the prediction accuracy of the algorithm. As shown in **Figure 4**, for outcomes defined using HDRS, the top predictor for remission for both women and men was baseline depression severity, followed by the *DEFB1*_2 (rs2741130) and *DEFB1*_1 (rs5743467) SNPs—biomarkers identified during our GWAS for plasma kynurenine concentrations. The top SNPs for response for men were the *TSPAN5* SNPs, which was the top hit in our GWAS for plasma serotonin concentration, followed by the *DEFB1*_1 and *DEFB1*_2 SNPs. For response in women, the top predictor was the *DEFB1*_1 SNP, followed by baseline depression severity and the *DEFB1*_2 SNP. **Figure S5** shows comparable contributions of predictors when QIDS-C rather than HDRS was used to define remission and response. When QIDS-C was used, DEFB1_1, DEFB1_2, and total depression severity were consistently among the top predictors of either outcome, just as with HDRS.

External validation using STAR*D and ISPC data. The classifier trained using PGRN-AMPS baseline depression severity and SNP data predicted response and remission, as defined by the QIDS-C

scores, in STAR*D patients with accuracies for men of: 66%, women: 66% ($P \leq 0.06$) and men: 75%, women: 65% ($P \leq 0.07$), respectively (**Table 1**). The classifier trained using PGRN-AMPS baseline depression severity and SNP data predicted response and remission, as defined by HDRS scores, in ISPC patients with accuracies for men of: 77%, women: 75% ($P \leq 0.07$) and men: 77%, women: 74% ($P \leq 0.07$), respectively (**Table 1**).

DISCUSSION
Improved predictions and mechanistic significance
We have shown that robust prediction of citalopram/escitalopram treatment outcomes can be achieved in depressed patients by using machine-learning approaches that integrate baseline depression severity with functionally validated pharmacogenomic SNP biomarkers. The AUC of 0.70 or higher achieved in this work represents an advance over our prior work, in which we used sociodemographic and clinical factors as predictor variables in a machine-learning algorithm applied to PGRN-AMPS data that resulted in an AUC of 0.54.[27] The prediction of antidepressive response must account for multiple interactions among biological, psychological, and environmental factors. Because of the phenotypic complexity of antidepressive response, others have defined an AUC of 0.70 or higher as being clinically meaningful—that is, sufficiently accurate to guide clinical decision making.[5] Crucially, we demonstrated cross-trial replication of prediction performance across rating scales in both STAR*D (QIDS-C scale) and ISPC (HDRS scale) trials with precision similar to that observed in training with PGRN-AMPS data. This work also represents an advance over traditional pharmacogenetic candidate gene approaches that identify plausible genes and SNPs associated with outcomes.[28–34] We achieved that advance by asking whether the application of machine-learning approaches that combine clinical assessments with a group of functionally validated pharmacogenomic SNPs as predictor variables might make it possible to predict SSRI treatment outcomes. Taken as a whole, our findings represent an important step toward the goal of algorithmically determining whether SSRIs are likely to be effective in patients with MDD prior to treatment initiation.

The pharmacogenomic biomarkers used in this study, namely SNPs in the *DEFB1*, *AHR*, *TSPAN5*, and *ERICH3* genes, were chosen based on the important roles of these genes in serotonin or kynurenine biosynthesis or in inflammation—mechanisms that are known to be associated with MDD disease risk and/or antidepressant response.[9,10] As noted earlier, prior experimental work showed that knockdown of the expression of both TSPAN5 and ERICH3 in neuronally derived cell lines resulted in decreased serotonin release into the culture media.[9] The *DEFB1* gene encodes a protein expressed in gastrointestinal mucosa that can inactivate lipopolysaccharides and, in turn, inhibit both inflammation and the biosynthesis of kynurenine, which is enhanced by inflammatory mediators.[10] The facts that the *DEFB1* SNPs figured so prominently and that this gene encodes a gut mucosal protein that can inactivate both lipopolysaccharides and gut bacteria highlight the potential importance of the rapidly evolving concept of a gut–brain axis.[25,35]

Table 1 Prediction performance of random forests using baseline depression severity and functionally validated SNPs of *TSPAN5*, *AHR*, *DEFB1*, and *ERICH3*

Rating scale	Trial	Training data	Gender	Accuracy (%)	95% CI in training cross-validation	NIR	P value	Sensitivity	Specificity	PPV	NPV	AUC in training cross-validation	Top 3 predictors in cross-validation
Response													
QIDS-C	PGRN-AMPS	10-fold cross-validation	Men	73	(63, 82)	0.66	0.01	0.7	0.78	0.86	0.57	0.85	DEFB1_2, Baseline severity, DEFB1_1
			Women	74	(65, 83)	0.63	0.0003	0.71	0.8	0.85	0.62	0.7	TSPAN5, DEFB1_1, Baseline severity
	STAR*D	PGRN-AMPS	Men	69	NA	0.68	0.06	0.67	0.72	0.82	0.51	NA	NA
			Women	66	NA	0.68	0.0007	0.68	0.63	0.78	0.52	NA	NA
HDRS	PGRN-AMPS	10-fold cross-validation	Men	86	(81, 94)	0.68	5.70E-04	0.9	0.85	0.93	0.79	0.88	TSPAN5, DEFB1_1, DEFB1_2
			Women	88	(78, 93)	0.7	7.30E-09	0.9	0.82	0.91	0.7	0.9	DEFB1_1, Baseline severity, DEFB1_2
	ISPC	PGRN-AMPS	Men	77	NA	0.69	0.05	0.8	0.71	0.85	0.62	NA	NA
			Women	75	NA	0.65	0.01	0.78	0.68	0.82	0.63	NA	NA
Remission													
QIDS-C	PGRN-AMPS	10-fold cross-validation	Men	78	(69, 86)	0.63	5.40E-08	0.81	0.75	0.84	0.69	0.86	Baseline severity, DEFB1_1, DEFB1_2
			Women	69	(60, 80)	0.62	0.0001	0.6	0.83	0.84	0.59	0.75	Baseline severity, DEFB1_2, DEFB1_1
	STAR*D	PGRN-AMPS	Men	75	NA	0.55	0.008	0.79	0.69	0.75	0.72	NA	NA
			Women	66	NA	0.5	0.001	0.59	0.72	0.67	0.63	NA	NA
HDRS	PGRN-AMPS	10-fold cross-validation	Men	86	(75, 90)	0.55	0.0001	0.9	0.84	0.87	0.87	0.84	Baseline severity, DEFB1_2, DEFB1_1
			Women	83	(75, 90)	0.51	0.03	0.87	0.8	0.82	0.85	0.9	Baseline severity, DEFB1_2, DEFB1_1
	ISPC	PGRN-AMPS	Men	76	NA	0.56	0.04	0.8	0.71	0.77	0.73	NA	NA
			Women	74	NA	0.52	0.07	0.76	0.72	0.74	0.73	NA	NA

Results of 10-fold internal cross-validation using PGRN-AMPS patients' data are reported in the light-blue blocks. Results from external validation in STAR*D (QIDS-C) and ISPC (HDRS) of the prediction model trained using PGRN-AMPS patients' data are reported in the light-orange blocks.

AUC, area under the receiver operating curve; CI, confidence interval; HDRS, Hamilton Depression Rating Scale; ISPC, International SSRI Pharmacogenomics Consortium; NA, not applicable; NIR, null information rate; NPV, negative predictive value; PGRN-AMPS, Mayo Clinic Pharmacogenomics Research Network Antidepressant Medication Pharmacogenomic Study; PPV, positive predictive value; QIDS-C, Quick Inventory of Depressive Symptomatology; SNP, single nucleotide polymorphism; STAR*D, Sequenced Treatment Alternatives to Relieve Depression.
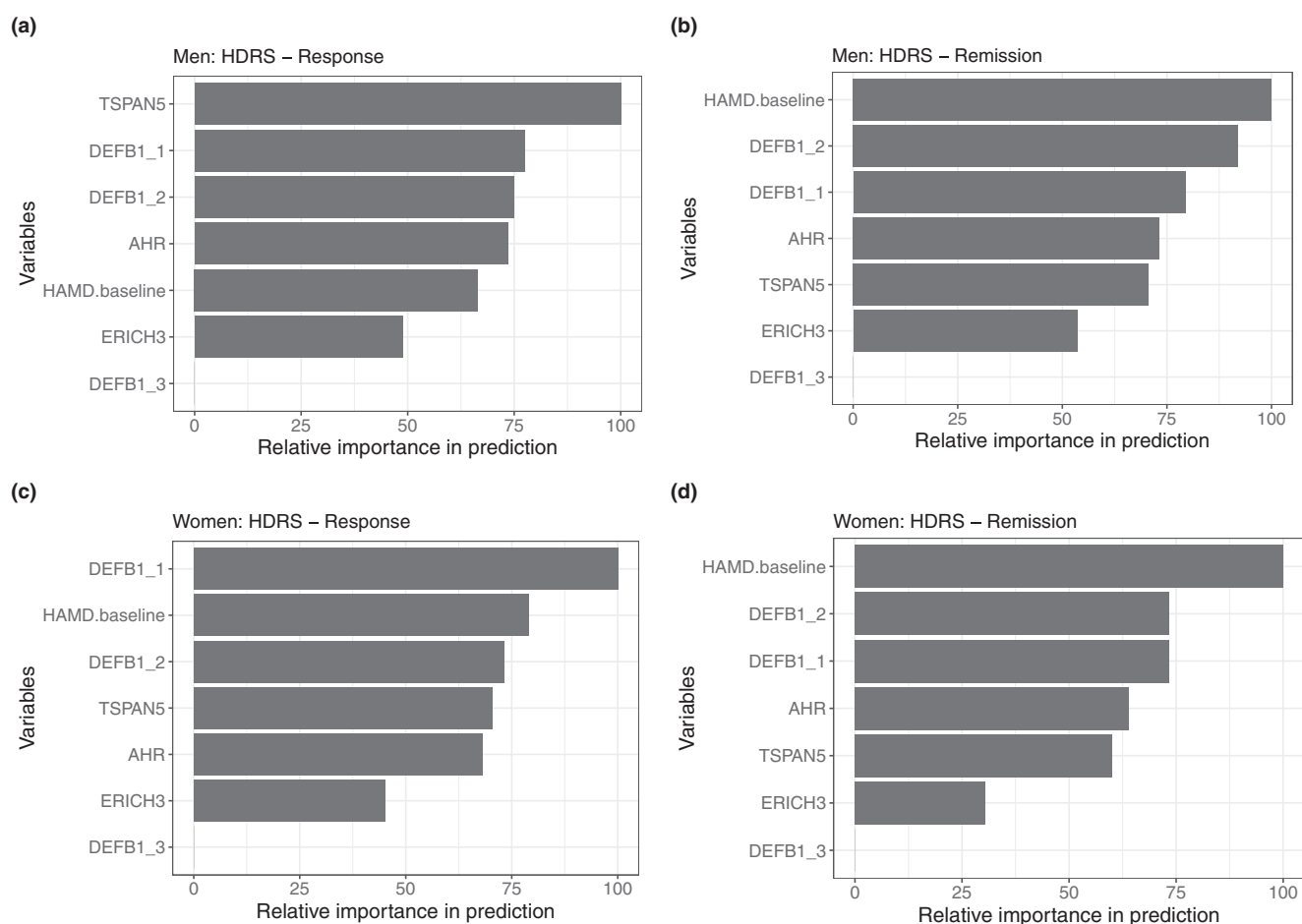
Figure 4 Importance of variables for predicting clinical outcomes measured using Hamilton Depression Rating Scale (HDRS). SNPs, single nucleotide polymorphisms.

The identification of these "top hit" SNPs during GWAS was performed for quantitative biological traits (i.e., metabolite concentrations), rather than measures of MDD clinical symptom severity (i.e., HDRS or QIDS-C), as our use of phenotypes represented a conscious attempt to move our analyses toward the biological underpinning of SSRI response. Because another of our goals involved cross-trial replication, we focused on pharmacogenomic SNP biomarkers in our predictive model because DNA data were more widely available across datasets than were other "omics" data. Furthermore, unlike metabolomics data, DNA sequences are stable and are less susceptible to variation related to environmental exposures or specimen handling and processing.

We acknowledge that the SNPs included in our study are not the only SNPs that might contribute to the predictability of antidepressant outcomes with this type of computational approach. Future investigation with methodological innovations will make it possible to screen a large number of SNPs across the human genome that may be more highly predictive of SSRI treatment outcomes than those used in this initial effort. Our results (as described in this work) from using pharmacodynamic biomarkers are promising because they suggest that, if similar approaches to derivation of biomarkers to study clinical responses are used with other antidepressants (such as serotonin-norepinephrine reuptake

inhibitors or esketamine), subsequent studies using machine-learning approaches like ours may lead to the development of drug-specific or of drug-agnostic (regardless of antidepressant subtype) predictive models that could guide treatment selection.

Clinical implications of patient clustering
The following are the clinical implications of the patient clusters inferred in this work.

Toward clinically actionable modeling of longitudinal effects of antidepressants. In practice, clinicians' ability to forecast eventual antidepressant treatment outcomes rests on their ability to factor baseline depression severity and subsequent changes in symptoms at intermediate time points, before a therapeutic trial is complete. To study the longitudinal effects of antidepressants, the patient clusters inferred in this work served as nodes of a probabilistic graph that made it possible to capture the longitudinal variation of depression symptoms over time, conditioned on baseline characteristics and changes in those characteristics at intermediate time points—a process that we have referred to as "symptom dynamics."[27] Therefore, replication of the cluster patterns at baseline and at 4 weeks is just as important as the replication of clinically valid clusters at 8 weeks. Understanding the symptom dynamics within

clusters of patients defined by depression severity and biological characteristics at baseline (predictive outcome markers) and at intermediate time points (change markers) may lead to further improvement in our understanding of antidepressant response. Specifically, in clinical settings where genomic biomarkers are not assayable, a symptom-based model that analyzes improvements in the severity of depressive symptoms at 4 weeks could still be used to provide prognoses of treatment outcomes at 8 weeks. A detailed understanding of the symptom dynamics across multiple time points may enable clinicians to change treatments if the predicted chances for response/remission are low.

Biological associations with depression severity. Our clustering approach can also be used to iteratively investigate the effects of multiple biological measures (e.g., metabolomics and genomics), individually and in groups, for predicting antidepressant response. Systematic studies using a variety of biological measures and other antidepressants may lead to improved understanding of the underlying neurobiology of antidepressant response and an enhanced ability to match individual patients with MDD with specific antidepressants based on their biological profiles.

Sex differences

When antidepressants are being chosen, potential sex differences in the underlying biology of antidepressant response are often overlooked. It is clear that sex represents an important risk factor for MDD, with virtually all studies reporting twice as many affected women as men.[15] Although sex has been reported to influence response to antidepressants in some studies,[11,17,21,36,37] prior machine-learning approaches using sociodemographic factors as predictors did not identify sex as a robust predictor of remission.[4] The sex-specific differences in some top predictors of treatment outcomes in our study (see **Figures 4 and S5**), and in recent targeted metabolomics-based antidepressant prediction studies,[9,11] suggest that sex-specific biological mechanisms may play an important role in antidepressant response.

CYP2C19 metabolizer phenotype and depression severity clusters

Our observation that the *CYP2C19* metabolizer phenotype was not significantly associated with citalopram/escitalopram treatment outcomes or depression severity clusters is similar to findings from previous research. Although functional *CYP2C19* allele variants are associated with citalopram/escitalopram metabolism and some drug side effects,[38] the impact of CYP P450 genotypes, including *CYP2C19*, on therapeutic outcomes has been less clear.[39] Some studies in depressed patients have found a significant association between the *CYP2C19* genotype and treatment response to citalopram or escitalopram,[40,41] whereas other studies have failed to demonstrate such an association.[42,43] There are similar inconsistencies in the results of studies attempting to link serum concentrations of antidepressants, including citalopram, with antidepressant response.[44,45] The lack of improved predictability of treatment outcomes through use of the *CYP2C19* genotype does not mean that

pharmacokinetic mechanisms and the *CYP2C19* genotype are not clinically relevant. That is especially true with respect to adverse responses, such as dose-dependent risk for corrected QT interval prolongation with citalopram.[39]

Methodological considerations in clustering patients

We focused on the use of total depression scale scores rather than individual depression scale items for three reasons. First, total depression scores at baseline were the most robust predictor of clinical outcomes in prior machine-learning studies.[4] Second, total depression scores have been widely used to define nonresponse, response, and remission in clinical trials.[46] Finally, we showed in this work that multivariate clustering approaches that use individual depression item scores did not yield clustering patterns at 8 weeks that conformed to accepted definitions of response or remission (**Figure S1**). The lack of associations between social/demographic factors and any of the depressive symptom severity clusters also agrees with prior work demonstrating that social/demographic factors individually or in aggregate cannot accurately predict antidepressant treatment outcomes.[4,6,47,48]

Predictive pharmacogenomic biomarkers for stratified randomization of clinical trials

Our results have potential implications for the design of future antidepressant trials. The predictive biomarkers in this study were pharmacodynamic in nature and are linked to important mechanisms underlying MDD risk and/or antidepressant response. If our results are replicated and extended to other antidepressants, these biomarkers may serve as genetic factors that may be used to screen out (exclude) patients on the basis of a high predicted likelihood of treatment failure. Alternatively, this information could be used to stratify clinical trial participants into categories with higher and lower risk of treatment failure prior to randomization. If so, randomization to treatment condition would be conducted within each risk group (i.e., there would be stratified randomization),[49] thus ensuring an optimum balance in outcome prognoses between treatment groups.

LIMITATIONS

The patient samples studied comprised white subjects, which reduced confounding by race but limits the generalizability of the predictions. We had no direct measures of socioeconomic status and comorbid anxiety, factors associated with poorer response to antidepressants.[48,50] Because we included complete cases, we cannot exclude the possibility of confounding by patients who dropped out. Although the improvement in outcome predictions was replicated across clinical trials of citalopram/escitalopram, this work has not been replicated for other antidepressants. Finally, patients were not excluded on the basis of body mass index or comorbid general medical conditions that might have influenced the interaction between drug treatment and genomic profile.

CONCLUSIONS

In summary, this study demonstrates that statistical/machine-learning approaches that integrate baseline depression severity with functionally validated pharmacogenomic SNP biomarkers

can be used to enhance our ability to predict antidepressant drug response phenotypes during short-term treatment. The patient clusters inferred and replicated across trials and depression rating scales could lead to better understanding of the underlying pathophysiology of MDD. Extension of this work to additional antidepressants may have the potential to increase the precision of antidepressant drug selection for individual patients and may serve as a platform for the use of antidepressant drugs as molecular probes to identify underlying mechanisms of disease and molecular subsets of MDD, a disease that is currently defined by symptoms rather than biological mechanisms.

METHODS

Data sources

PGRN-AMPS (NCT 00613470) was an 8-week, single-arm, open trial that assessed clinical outcomes in adults with MDD in response to citalopram/escitalopram and examined metabolomic and genomic factors associated with those outcomes.[51] Subjects were recruited from primary-care and specialty-care settings from March 2005 to May 2013. Psychiatric diagnoses were confirmed using modules A, B (screen-only version), and D of the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (SCID).[52] Clinical and demographic variables from the PGRN-AMPS dataset used in the analyses (**Table S1**) were assessed at baseline using standardized questionnaires.

Data from the initial phase of the STAR*D trial (NCT 00021528)[12] and ISPC[13] were used to externally validate the depressive symptom response subgroups inferred in the PGRN-AMPS subjects, and the prediction models trained using PGRN-AMPS's data. The initial phase of STAR*D was a 12-week clinical trial of citalopram for adults with MDD conducted in the United States from June 2001 to April 2004. Subjects were recruited from primary-care and specialty-care settings. ISPC comprised seven member sites that contributed data from seven clinical trials of SSRIs for depression carried out in North America, Europe, and Asia to examine genetic factors driving variation in clinical response to SSRIs.[13] Details of STAR*D's study procedures and a description of each contributing study in ISPC have been published previously,[12,13,53,54] and are described in **Section S1**.

The PGRN-AMPS study protocol was approved by the Institutional Review Board at Mayo Clinic. All study sites included in the ISPC analyses were approved for participation in the ISPC consortium by their local institutional review boards. The University of Texas Southwestern Medical Center (Dallas, TX), the institutional review boards at each clinical site, and the Data Coordinating Center and the Data Safety and Monitoring Board of the National Institute of Mental Health approved and monitored the study protocol. Finally, all PGRN-AMPS, STAR*D, and ISPC participants provided written informed consent before study entry.

For the present analyses, we utilized data from 398 (men: 144, women: 254) white citalopram-treated PGRN-AMPS subjects, 467 (men: 182, women: 285) white citalopram-treated STAR*D subjects, and 165 (men: 62, women: 103) white citalopram/escitalopram–treated ISPC subjects who had genotype and complete clinical data (no missing values) at baseline and at 4 and 8 weeks. All PGRN-AMPS and ISPC subjects also had *CYP2C19* metabolizer genotype data at baseline, and plasma drug levels at 4 and 8 weeks. Details of genotyping and GWAS of the PGRN-AMPS, STAR*D, and ISPC subjects have been previously published.[22,23,51]

Clinical outcomes

In all three trials, treatment outcomes were established using the clinician-rated version of the 16-item QIDS-C[55] or the 17-item HDRS.[56]

Remission was defined as a QIDS-C score ≤ 5[55] (HDRS score ≤ 7[56]) at 4 or 8 weeks. Response was defined as a ≥50% reduction in QIDS-C or HDRS total score from baseline to either 4 or 8 weeks. Across the three datasets, 60–66% of subjects were classified as responders, and 37–50% as remitters, at 8 weeks.

Analysis workflow

Sex-stratified analyses. Given several prior research efforts that identified sex differences in MDD prevalence and biological factors related to antidepressant treatment outcomes,[14–21,57] all analyses in this work were stratified by sex. This allowed us to study sex-specific contributions of pharmacogenomic SNPs to prediction of SSRI response.

Analysis overview. A sequence of the application of an unsupervised machine-learning approach (because clustering is an inferential task) followed by supervised learning (to predict SSRI treatment outcome) comprised a machine-learning workflow (illustrated in **Figure 1**) that is described next; additional details of the implementation are in **Section S2**.

Stage 1

Aim. Identify depressive symptom severity clusters in PGRN-AMPS (stage 1A), replicate the cluster patterns using STAR*D and ISPC data (stage 1B), and identify sociodemographic factors associated with clusters (stage 1C).

Approach. Unsupervised learning was used to identify clusters of patients based on total QIDS-C and HDRS scores at baseline, 4 weeks, and 8 weeks. The overall distribution of QIDS-C and HDRS total scores comprised multiple normal distributions (**Figure 2**). Mixture-model–based unsupervised learning[27] with Gaussian mixture models was used to algorithmically identify the minimum number of Gaussians that best approximated the actual distribution of depressive symptom severity in PGRN-AMPS patients at each time point.[9,27] The use of the Gaussian mixture model clustering approach was further justified by the unsuitability of longitudinal clustering/trajectory techniques,[58] given the eventual goal of associating biological measures with depression severity during discrete treatment time points.

To validate the clustering approach developed in stage 1A, we used STAR*D (for QIDS-C) and ISPC (for HDRS) datasets in stage 1B to investigate, using Kolmogorov–Smirnov tests, whether the distributions of depression severity were the same in the three independent datasets.

In stage 1C, Kolmogorov–Smirnov (continuous data) and two-way χ^2 (categorical data) tests were used to identify clinical and sociodemographic factors (listed in **Table S1**) associated with the depression severity clusters at all time points in all three datasets. Any associated clinical/sociodemographic factors were then combined with pharmacogenomic SNPs to predict treatment outcomes in stage 2.

Stage 2

Aim. Predict antidepressant remission/response using pharmacogenomic biomarkers and baseline depression severity.

Approach. We trained random forests (i.e., the randomForest R library) using PGRN-AMPS's baseline depression severity and pharmacogenomics data (represented as numerical genotypes[22,23,51]) to predict remission/response, and we then externally validated the trained prediction model using STAR*D and ISPC data (see **Figure 1**). Because clinical/sociodemographic factors, the *CY2C19* phenotype, and plasma drug levels were not associated with the baseline or 4-week clusters, we assessed the predictive capability of the pharmacogenomic biomarkers when augmented only with baseline depression severity, not with stratification by baseline depression severity clusters.

Random forests were used because of their mathematical ability to handle discrete (e.g., with numerical genotypes), correlated predictor variables, which has demonstrated robust predictive capabilities in several clinical applications,[59] including psychiatric disorders.[60] Details of the 10-fold cross-validation with five repeats to minimize the effects of overfit and information leak, along with prediction performance statistics, are provided in **Section S2**.

SUPPORTING INFORMATION

Supplementary information accompanies this paper on the *Clinical Pharmacology & Therapeutics* website (www.cpt-journal.com).

Supplementary Material S1. Trial description.
Supplementary Material S2. Analyses.
Table S1. Clinical and demographic factors from AMPS analyzed in this work.
Figure S1. Comparison of depressive symptom clustering behavior, using various approaches.
Figure S2. Comparison of mean ages for men and women in clusters with comparable symptom severity at baseline, 4 weeks, and 8 weeks.
Figure S3. Comparison of mean body mass indices (BMIs; kg/m^2) for men and women in clusters with comparable symptom severity at baseline, 4 weeks, and 8 weeks.
Figure S4. Comparison of citalopram and escitalopram plasma drug concentrations between men and women with each depressive symptom severity cluster at 4 weeks (**a**) and 8 weeks (**b**).
Figure S5. Importance of variables for predicting clinical outcomes measured using QIDS-C.

CONFLICTS OF INTEREST
M.A.F. has grant support from AssureRx, the Mayo Foundation, Myriad, the National Institute of Alcohol Abuse and Alcoholism (NIAAA), the National Institute of Mental Health (NIMH), and Pfizer, and consults for Janssen, Mitsubishi Tanabe Pharma Corporation, Myriad, Neuralstem Inc., Otsuka America Pharmaceutical, Sunovion, and Teva Pharmaceuticals. L.W. and R.M.W. are cofounders and stockholders in OneOme LLC. W.V.B.'s research has been supported by the National Institute of Mental Health, the Agency for Healthcare Research & Quality, and the Mayo Foundation for Medical Education and Research. He has contributed chapters to UpToDate concerning the use of antidepressants and atypical antipsychotic drugs for treating adults with bipolar major depression. In the last 3 years, A.J.R. has received consulting fees from Akili Inc., Brain Resource Ltd, Compass Inc., Curbstone Consultant LLC, Emmes Corp., Holmusk, LivaNova, Santium Inc., Sunovion, Taj Medical, and Takeda USA; speaking fees from LivaNova; and royalties from Guilford Publications and the University of Texas Southwestern Medical Center. All others declared no competing interests for this work.

AUTHOR CONTRIBUTIONS
A.P.A., W.V.B., and R.M.W. wrote the manuscript. M.A.F., E.B., M.S., L.W., and R.M.W. designed the research. A.P.A., W.V.B., and R.M.W. performed the research. A.P.A., D.N., J.M.B., R.K.I., T.R., and E.B. analyzed the data.

1. Friedrich, M.J. Depression is the leading cause of disability around the world. *JAMA* **317**, 1517 (2017).
2. Trivedi, M.H. *et al.* Medication augmentation after the failure of SSRIs for depression. *N. Engl. J. Med.* **354**, 1243–1252 (2006).
3. Chekroud, A.M., Gueorguieva, R., Krumholz, H.M., Trivedi, M.H., Krystal, J.H. & McCarthy, G. Reevaluating the efficacy and predictability of antidepressant treatments: a symptom clustering approach. *JAMA Psychiatry* **74**, 370–378 (2017).
4. Chekroud, A.M. *et al.* Cross-trial prediction of treatment outcome in depression: a machine learning approach. *Lancet Psychiatry* **3**, 243–250 (2016).
5. Iniesta, R. *et al.* Combining clinical variables to optimize prediction of antidepressant treatment outcomes. *J. Psychiatr. Res.* **78**, 94–102 (2016).
6. Bagby, R.M., Ryder, A.G. & Cristi, C. Psychosocial and clinical predictors of response to pharmacotherapy for depression. *J. Psychiatry Neurosci.* **27**, 250–257 (2002).
7. Musliner, K.L., Munk-Olsen, T., Eaton, W.W. & Zandi, P.P. Heterogeneity in long-term trajectories of depressive symptoms: patterns, predictors and outcomes. *J. Affect. Disord.* **192**, 199–211 (2016).
8. Iniesta, R. *et al.* Antidepressant drug-specific prediction of depression treatment outcomes from genetic and clinical variables. *Sci. Rep.* **8**, 5530 (2018).
9. Athreya, A.P. *et al.* Augmentation of physician assessments with multi-omics enhances predictability of drug response: a case study of major depressive disorder. *IEEE Comput. Intell. Mag.* **13**, 20–31 (2018).
10. Lin, E., Kuo, P.H., Liu, Y.L., Yu, Y.W., Yang, A.C. & Tsai, S.J. A deep learning approach for predicting antidepressant response in major depression using clinical and genetic biomarkers. *Front. Psychiatry* **9**, 290 (2018).
11. Czysz, A.H. *et al.* Can targeted metabolomics predict depression recovery? Results from the CO-MED trial. *Transl. Psychiatry* **9**, 11 (2019).
12. Trivedi, M.H. *et al.* Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am. J. Psychiatry* **163**, 28–40 (2006).
13. Biernacka, J.M. *et al.* The international SSRI Pharmacogenomics Consortium (ISPC): a genome-wide association study of antidepressant treatment response. *Transl. Psychiatry* **5**, e553 (2015).
14. Berlanga, C. & Flores-Ramos, M. Different gender response to serotonergic and noradrenergic antidepressants. A comparative study of the efficacy of citalopram and reboxetine. *J. Affect. Disord.* **95**, 119–123 (2006).
15. Keers, R. & Aitchison, K.J. Gender differences in antidepressant drug response. *Int. Rev. Psychiatry* **22**, 485–500 (2010).
16. Khan, A., Brodhead, A.E., Schwartz, K.A., Kolts, R.L. & Brown, W.A. Sex differences in antidepressant response in recent antidepressant clinical trials. *J. Clin. Psychopharmacol.* **25**, 318–324 (2005).
17. Kornstein, S.G. *et al.* Gender differences in treatment response to sertraline versus imipramine in chronic depression. *Am. J. Psychiatry* **157**, 1445–1452 (2000).
18. Lawton, K.A. *et al.* Analysis of the adult human plasma metabolome. *Pharmacogenomics* **9**, 383–397 (2008).
19. Sramek, J.J., Murphy, M.F. & Cutler, N.R. Sex differences in the psychopharmacological treatment of depression. *Dialogues Clin. Neurosci.* **18**, 447–457 (2016).
20. Townsend, M.K. *et al.* Reproducibility of metabolomic profiles among men and women in 2 large cohort studies. *Clin. Chem.* **59**, 1657–1667 (2013).

21. Young, E.A. *et al.* Sex differences in response to citalopram: a STAR*D report. *J. Psychiatr. Res.* **43**, 503–511 (2009).
22. Gupta, M. *et al.* TSPAN5, ERICH3 and selective serotonin reuptake inhibitors in major depressive disorder: pharmacometabolomics-informed pharmacogenomics. *Mol. Psychiatry* **21**, 1717–1725 (2016).
23. Liu, D. *et al.* Beta-defensin 1, aryl hydrocarbon receptor and plasma Kynurenine in major depressive disorder: metabolomics-informed genomics. *Transl. Psychiatry* **8**, 10 (2018).
24. Bauer, M.E. & Teixeira, A.L. Inflammation in psychiatric disorders: what comes first? *Ann. NY Acad. Sci.* **1437**, 57–67 (2019).
25. Haase, S., Haghikia, A., Wilck, N., Muller, D.N. & Linker, R.A. Impacts of microbiome metabolites on immune regulation and autoimmunity. *Immunology* **154**, 230–238 (2018).
26. Moulton, C.D., Hopkins, C.W.P., Ismail, K. & Stahl, D. Repositioning of diabetes treatments for depressive symptoms: a systematic review and meta-analysis of clinical trials. *Psychoneuroendocrinology* **94**, 91–103 (2018).
27. Athreya, A.P. *et al.*, eds. Data-Driven Longitudinal Modeling and Prediction of Symptom Dynamics in Major Depressive Disorder: Integrating Factor Graphs and Learning Methods. IEEE International Conference on Computational Intelligence in Bioinformatics and Computational Biology; 2017: IEEE Computational Intelligence Society.
28. Ellsworth, K.A. *et al.* Contribution of FKBP5 genetic variation to gemcitabine treatment and survival in pancreatic adenocarcinoma. *PLoS One* **8**, e70216 (2013).
29. Gasso, P. *et al.* Association of regulatory TPH2 polymorphisms with higher reduction in depressive symptoms in children and adolescents treated with fluoxetine. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **77**, 236–240 (2017).
30. Ji, Y. *et al.* Glycine and a glycine dehydrogenase (GLDC) SNP as citalopram/escitalopram response biomarkers in depression: pharmacometabolomics-informed pharmacogenomics. *Clin. Pharmacol. Ther.* **89**, 97–104 (2011).
31. Kato, M. *et al.* Effect of 5-HT1A gene polymorphisms on antidepressant response in major depressive disorder. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **150B**, 115–123 (2009).
32. Kishi, T. *et al.* GTP cyclohydrolase 1 gene haplotypes as predictors of SSRI response in Japanese patients with major depressive disorder. *J. Affect. Disord.* **142**, 315–322 (2012).
33. Paroni, G. *et al.* Klotho gene and selective serotonin reuptake inhibitors: response to treatment in late-life major depressive disorder. *Mol. Neurobiol.* **54**, 1340–1351 (2017).
34. Qesseveur, G. *et al.* Genetic dysfunction of serotonin 2A receptor hampers response to antidepressant drugs: a translational approach. *Neuropharmacology* **105**, 142–153 (2016).
35. Johnson, K.V. & Foster, K.R. Why does the microbiome affect behaviour? *Nat. Rev. Microbiol.* **16**, 647–655 (2018).
36. Krumsiek, J. *et al.* Gender-specific pathway differences in the human serum metabolome. *Metabolomics* **11**, 1815–1833 (2015).
37. Martin, L.A., Neighbors, H.W. & Griffith, D.M. The experience of symptoms of depression in men vs women: analysis of the National Comorbidity Survey Replication. *JAMA Psychiatry* **70**, 1100–1106 (2013).
38. Chang, M., Tybring, G., Dahl, M.L. & Lindh, J.D. Impact of cytochrome P450 2C19 polymorphisms on citalopram/escitalopram exposure: a systematic review and meta-analysis. *Clin. Pharmacokinet.* **53**, 801–811 (2014).
39. Nassan, M., Nicholson, W.T., Elliott, M.A., Rohrer Vitek, C.R., Black, J.L. & Frye, M.A. Pharmacokinetic pharmacogenetic prescribing guidelines for antidepressants: a template for psychiatric precision medicine. *Mayo Clin. Proc.* **91**, 897–907 (2016).
40. Mrazek, D.A. *et al.* CYP2C19 variation and citalopram response. *Pharmacogenet. Genomics* **21**, 1–9 (2011).

41. Tsai, M.H. *et al.* Genetic polymorphisms of cytochrome P450 enzymes influence metabolism of the antidepressant escitalopram and treatment response. *Pharmacogenomics* **11**, 537–546 (2010).
42. Hodgson, K. *et al.* Genetic differences in cytochrome P450 enzymes and antidepressant treatment response. *J. Psychopharmacol.* **28**, 133–141 (2014).
43. Peters, E.J. *et al.* Pharmacokinetic genes do not influence response or tolerance to citalopram in the STAR*D sample. *PLoS One* **3**, e1872 (2008).
44. Florio, V., Porcelli, S., Saria, A., Serretti, A. & Conca, A. Escitalopram plasma levels and antidepressant response. *Eur. Neuropsychopharmacol.* **27**, 940–944 (2017).
45. Olfson, M. & Marcus, S.C. National patterns in antidepressant medication treatment. *Arch. Gen. Psychiatry* **66**, 848–856 (2009).
46. Rush, A.J. *et al.* Report by the ACNP Task Force on response and remission in major depressive disorder. *Neuropsychopharmacology* **31**, 1841–1853 (2006).
47. Mulder, R.T., Joyce, P.R., Frampton, C.M., Luty, S.E. & Sullivan, P.F. Six months of treatment for depression: outcome and predictors of the course of illness. *Am. J. Psychiatry* **163**, 95–100 (2006).
48. Jain, F.A., Hunter, A.M., Brooks, J.O. 3rd & Leuchter, A.F. Predictive socioeconomic and clinical profiles of antidepressant response and remission. *Depress. Anxiety* **30**, 624–630 (2013).
49. Kernan, W.N., Viscoli, C.M., Makuch, R.W., Brass, L.M. & Horwitz, R.I. Stratified randomization for clinical trials. *J. Clin. Epidemiol.* **52**, 19–26 (1999).
50. Fava, M. *et al.* Difference in treatment outcome in outpatients with anxious versus nonanxious depression: a STAR*D report. *Am. J. Psychiatry* **165**, 342–351 (2008).
51. Ji, Y. *et al.* Pharmacogenomics of selective serotonin reuptake inhibitor treatment for major depressive disorder: genome-wide associations and functional genomics. *Pharmacogenomics J.* **13**, 456–463 (2013).
52. First, M.B., Spitzer, R.L., Miriam, G. & Williams Janet, B.W. *Structured Clinical Interview for DSM-IV Axis I Disorders, Clinician Version (SCID-CV)* (American Psychiatric Press Inc., Washington, D.C., 1996).
53. Rush, A.J. *et al.* Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. *Control. Clin. Trials* **25**, 119–142 (2004).
54. Rush, A.J. *et al.* Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am. J. Psychiatry* **163**, 1905–1917 (2006).
55. Rush, A.J. *et al.* The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biol. Psychiatry* **54**, 573–583 (2003).
56. Hamilton, M. A rating scale for depression. *J. Neurol. Neurosurg. Psychiatry* **23**, 56–62 (1960).
57. Rahikainen, A.L., Palo, J.U., Haukka, J. & Sajantila, A. Post-mortem analysis of suicide victims shows ABCB1 haplotype 1236T-2677T-3435T as a candidate predisposing factor behind adverse drug reactions in females. *Pharmacogenet. Genomics* **28**, 99–106 (2018).
58. Genolini, C. & Falissard, B. KmL: k-means for longitudinal data. *Comput. Statistics* **25**, 317–328 (2010).
59. Fernandez-Delgado, M., Cernadas, E., Barro, S. & Amorim, D. Do we need hundreds of classifiers to solve real world classification problems? *J. Machine Learn. Res.* **15**, 3133–3181 (2014).
60. Ball, T.M., Stein, M.B., Ramsawh, H.J., Campbell-Sills, L. & Paulus, M.P. Single-subject anxiety treatment outcome prediction using functional neuroimaging. *Neuropsychopharmacology* **43**, 926 (2018).