Utility of integrated pharmacogenomic testing to support the treatment of major depressive disorder in a psychiatric outpatient setting

Daniel K. Hall-Flavin[a], Joel G. Winner[e,f], Josiah D. Allen[f], Joseph M. Carhart[f], Brian Proctor[d], Karen A. Snyder[a], Maureen S. Drews[b], Linda L. Eisterhold[a], Jennifer Geske[c] and David A. Mrazek[a]

Objective The objective was to evaluate the potential benefit of an integrated, five-gene pharmacogenomic test and interpretive report (GeneSight) for the management of psychotropic medications used to treat major depression in an outpatient psychiatric practice.

Methods The open-label study was divided into two groups. In the first (unguided) group ($n=113$), pharmacogenomic information was not shared until all participants completed the study. In the second (guided) group ($n=114$), the pharmacogenomic report was provided to physicians for clinical use. Three depression ratings, the 17-item Hamilton Rating Scale for Depression (HAMD-17), the Quick Inventory of Depressive Symptomatology – Clinician Rated (QIDS-C16), and the Patient Health Questionnaire (PHQ-9), were collected at baseline, and at 2, 4, and 8 weeks.

Results The guided group experienced greater percent improvement in depression scores from baseline on all three depression instruments (HAMD-17, $P<0.0001$; QIDS-C16, $P<0.0001$; PHQ-9, $P<0.0001$) compared with the unguided group. Eight-week response rates were higher in the guided group than in the unguided group on all three measurements (HAMD-17, $P=0.03$; QIDS-C16, $P=0.005$; PHQ-9, $P=0.01$). Eight-week QIDS-C16 remission rates were higher in the guided group ($P=0.03$). Participants in the unguided group who at baseline were prescribed a medication that was most discordant with their genotype experienced the least improvement compared with other unguided participants (HAMD-17, $P=0.007$). Participants in the guided group and on a baseline medication most discordant with their genotype showed the greatest improvement compared with the unguided cohort participants (HAMD-17, $P=0.01$).

Conclusion These findings replicate previous studies and demonstrate significantly improved depression outcomes with use of GeneSight, an integrated, multigenetic pharmacogenomic testing platform. *Pharmacogenetics and Genomics* 00:000–000 © 2013 Wolters Kluwer Health | Lippincott Williams & Wilkins.

Introduction

Approximately 40% of patients treated with antidepressants experience a complete remission of their symptoms when initially treated with these agents [1]. In those patients who remain depressed, up to one-half are unlikely to experience substantial improvement with the use of a different antidepressant or adjunctive supplementation of another medication [2]. Adherence with taking antidepressant medications is also problematic and influenced by side effects, treatment ineffectiveness, or lack of appropriate follow-up care [3–6]. Substantial morbidity and interpatient variability is common with current antidepressant treatment strategies [7].

Interindividual variation in drug response depends on a number of factors, including diagnostic accuracy, drug–drug interactions, renal and hepatic function, medical and psychiatric comorbidity including substance use disorders, nutritional standing, and medication. In addition, genetically determined pharmacokinetic and pharmacodynamic variability can influence medication response [8]. For example, Kirchheiner *et al.* [9] suggested dosing changes based on the pharmacokinetic genotype. Mrazek *et al.* [10] reported on the association of variations in the *CYP2C19* gene and remission from depression with citalopram in individuals of European origin. Licinio and Wong recently summarized the influence of the 5HTTLPR polymorphism on response to selective serotonin reuptake inhibitors [11]. Several other authors have also reported on the impact of pharmacogenomic testing for the individualization of psychiatric care in both adult and pediatric populations [12–17].

The objective of personalized medicine is to improve outcomes of treatment, tolerability to medication, and adherence to medication regimens. Simon and Perlis [6] have discussed the need to move beyond a clinical evidence base composed of data that address the average effectiveness of treatments. Specific, measurable patient characteristics are needed to help guide treatment selection. Recent reviews and clinical reports have focused on the promise of individualized molecular psychiatry and the process of translational change [18–20].

Recently, a prospective, proof-of-concept study was conducted that demonstrated significant improvement in patient outcomes when pharmacogenomic testing was utilized in an outpatient psychiatric clinic that provided integrated treatment with a substantial emphasis on psychotherapy [21]. The present study, designed as a replication, utilized an identical study design but was conducted in an outpatient psychiatric clinic that primarily provides psychopharmacological treatment delivered by psychiatrists. Both studies describe the implementation of the GeneSight pharmacogenomic test and interpretive report designed to improve the safety and efficacy of prescribing antidepressant and antipsychotic medication in an outpatient psychiatric clinic. This algorithm is based on the genotyping of both copies of five pharmacokinetic and pharmacodynamic genes selected for their relevance to clinical response to antidepressants and antipsychotics (Fig. 1).

Methods
The present study was conducted at Franciscan Skemp Hospital in La Crosse, Wisconsin, a member of the Mayo Health System. Genotyping was performed by AssureRx Health, Inc. (Mason, Ohio, USA) (*http://www.assurerx-health.com*). The trial was conducted by Mayo Clinic personnel and was approved by the Mayo Clinic Institutional Review Board. All diagnoses were made clinically by board certified psychiatrists.

Genotyping procedure
Polymorphisms were measured among five genes that influence antidepressant and antipsychotic drug metabolism or response [8]. These included (a) the cytochrome P450 2D6 gene (*CYP2D6*); (b) the cytochrome P450 2C19 gene (*CYP2C19*); (c) the cytochrome P450 1A2 gene (*CYP1A2*); (d) the serotonin transporter gene (*SLC6A4*); and, (e) the serotonin 2A receptor gene (*HTR2A*).

CYP2D6, *CYP2C19*, and *CYP1A2* were genotyped using the Luminex xTAG system (Luminex Corporation, Austin, Texas, USA). Relevant regions were amplified using PCR and clarified using exonuclease I and shrimp alkaline phosphotase. Individual mutations were identified using allele-specific extension primers tagged for hybridization to Luminex xTAG beads. The following *CYP2D6* alleles were identified: *1, *2, *2A, *3, *4, *5,

*6, *7, *8, *9, *10, *11, *12, *14, *15, *17, *41. The presence of *CYP2D6* duplications was identified. The following *CYP2C19* alleles were identified: *1, *2, *3, *4, *5, *6, *7, *8. The following *CYP1A2* single nucleotide polymorphisms were identified: -3860G > A, -2467delT, -739T > G, -729C > T, -163C > A, 125C > G, 558C > A, 2116G > A, 2473G > A, 2499A > T, 3497G > A, 3533G > A, 5090C > T, 5166G > A, 5347C > T. Relevant regions of *SLC6A4* and *HTR2A* were amplified using PCR. *HTR2A* was then digested with the restriction enzyme *MSP*I (from *Moraxella* spp.). The *SLC6A4* PCR product and digested *HTR2A* PCR product were then run on a 2% gel to determine the genotype. The long and short forms of *SLC6A4* promoter due to a 44 bp indel were identified. The *HTR2A* single nucleotide polymorphism, T102C, was identified.

To improve the clinical relevance of the genotyping results for clinicians, genotype results were applied using a proprietary interpretive report (Fig. 2), referred to as GeneSight, which incorporates the genetic information with the known pharmacological profile for each of the 26 medications in the panel. The medications are then placed into advisory categories (bins) of 'use as directed' (hereafter referred to as 'green bin'), 'use with caution' (yellow bin), and 'use with caution and with more frequent monitoring' (red bin), which are meant to help categorize significant gene–drug interactions for each individual, with the red bin containing drugs whose outcome would be more likely to be impacted by the patient's individual genetics (see Fig. 2 for an example of a report for one patient). In addition, footnotes associated with each medication in the yellow or red cautionary bins provide the details of this interaction (e.g. serum levels may be too high, serum levels may be too low, genotype suggest less than optimal response), covering aspects of both reduced efficacy (e.g. *SLC6A4* S/S genotype, CYP450 UM phenotype, etc.) or increases in potential adverse events (e.g. CYP450 PM phenotype). Genotype results for the five genes were also provided.

Study criteria and description
Two hundred and thirty-three male and female patients between the ages of 18 and 72 with a primary *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed. diagnosis of major depressive disorder or depressive disorder NOS were approached for consent. Three patients refused consent. Potential individuals with a diagnosis of bipolar type I, schizophrenia, or schizoaffective disorders were excluded from the study. A minimum score of 14 on the 17-item Hamilton Rating Scale for Depression (HAMD-17) was required for patient inclusion and three participants failed to meet these criteria [22,23]. Of the 227 remaining participants (113 in the unguided group and 114 in the guided group), 62 (20 in the unguided group and 42 in the guided group) failed to complete one or more of the follow-up calls (Fig. 3). No significant

Fig. 1

Antidepressants

Use as directed	Use with caution	Use with increased caution and with more frequent monitoring
Citalopram (Celexa) Desvenlafaxine (Pristiq) Escitalopram (Lexapro) Fluvoxamine (Luvox) Selegiline (Emsam) Sertraline (Zoloft)	Duloxetine (Cymbalta) [1] Miratazapine (Remeron) [1] Trazodone (Desyrel) [1]	Amitriptyline (Elavil) [6] Bupropion (Wellbutrin) [6] Clomipramine (Anafranil) [6] Desipramine (Norpramin) [6] Fluoxetine (Prozac) [6] Imipramine (Tofranil) [6] Nortriptyline (Pamelor) [6] Paroxetine (Paxil) [6] Venlafaxine (Effexor) [6]

Antipsychotics

Use as directed	Use with caution	Use with increased caution and with more frequent monitoring
Quetiapine (Seroquel) Ziprasidone (Geodon)	Clozapine (Clozaril) [1] Olanzapine (Zyprexa) [1] Risperidone (Risperdal) [1]	Aripiprazole (Abilify) [6] Haloperidol (Haldol) [6] Perphenazine (Trilafon) [6]

[1] Serum level may be too high, lower doses may be required.
[6] Use of this drug is associated with an increased risk of side effects.

Patient genotypes and phenotypes

Gene	Genotype	Predicted phenotype
CYP2D6	*4/*4	Poor metabolizer
CYP2C19	*1/*1	Extensive metabolizer
CYP1A2	−163C>A − C/A	Extensive metabolizer
SLC6A4	L/L	High activity
HTR2A	C/C	Reduced activity

An example of a pharmacogenomic interpretive report for one individual participant. Pharmacogenomic test results are used to categorize medications based on the individual pharmacokinetic and pharmacodynamic factors that are salient to each medication.

differences in age, sex, or baseline HAMD-17 score were observed between groups. These and other demographic variables are shown in Table 1. The frequencies of the *CYP2D6* phenotype differed between the unguided and guided groups and did not differ for the other four genes (Table 2).

Participants were enrolled into two consecutive groups. In the first (unguided) group, DNA was collected by buccal swab, and the report was created, but it was not shared with the physician until the completion of the trial. Thus, these participants received clinical treatment

as usual, without the use or knowledge of genotyping results by their physician. In the second (guided) group, for which enrollment immediately followed the unguided group, DNA was collected by buccal swab at baseline and, within 48 h of sample collection, the report was provided to the treating physicians.

No patient participated in both groups. Other than the pharmacogenomic testing and interpretive report, no additional pharmacogenomic education was provided to physicians or participants in the guided group. The pharmacogenomic testing was provided free of

Fig. 2
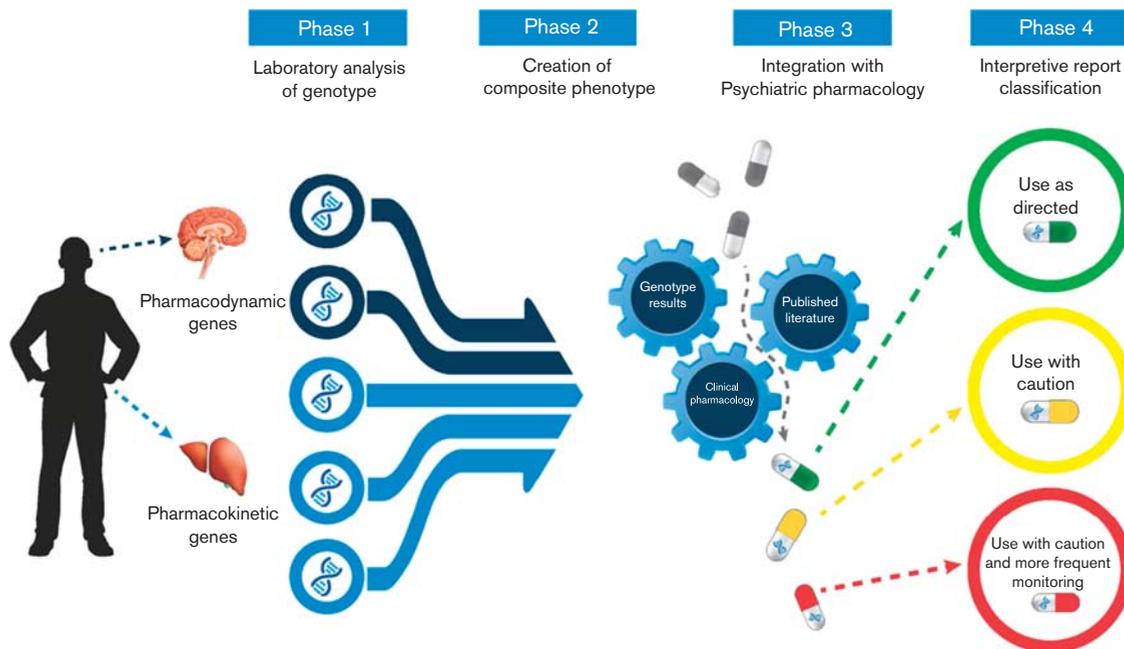
Medication binning methodology of the GeneSight interpretive report.

charge to study participants and no additional incentive was provided. Study physicians were not incentivized to participate. None of the study physicians had previously ordered pharmacogenomic testing and were free to implement the results at their clinical discretion.

Data collection
Data were collected at baseline and at 2, 4, and 8 weeks thereafter. Clinical rating instruments were selected to facilitate comparison with the Sequenced Treatment Alternatives to Relieve Depression Study (STAR*D) [24]. Baseline visits were conducted in person following informed consent. At baseline, participants were given the HAMD-17, the Quick Inventory of Depressive Symptomatology – Clinician Rated (QIDS-C16), and the Patient Health Questionnaire (PHQ-9) [25–28]. Demographic information and a psychotropic medication history were also collected at the baseline visit.

Subsequent patient study visits were conducted through the telephone to reduce participatory burden. The same indices measured at baseline were also assessed at the 2-, 4-, and 8-week visits. Participants who missed a visit were called every other day for 1 week after the absence before the visit was classified as 'missed.' Medication reconciliation was performed at each visit. At the 8-week visit, a patient survey regarding satisfaction with clinical care was also administered. Physicians were also given a satisfaction survey following the 8-week visit. The physician survey questions addressed turnaround time,

ease of use, overall satisfaction, confidence in the utility of pharmacogenomic testing as a tool to help guide treatment, and perception of patient acceptance.

Statistical analysis
A sample size of 86 participants per group was calculated as the number of patients needed to provide 90% power to detect on average a 15% reduction in symptom scores over 8 weeks with a common SD of 30% at an α level of 0.05. Repeated measures analysis was performed using a mixed model that treated participants as random effects to examine the effect of time and group (guided vs. unguided treatment) on the reduction of depression rating score. Treatment group, weeks (i.e. time), weeks-squared, and the interaction between treatment group and time were treated as fixed effects. Under mixed-model assumptions, F tests were declared significant if P was less than 0.05.

Given the distribution of the data, nonparametric analysis of variance models were used to compare continuous outcomes (e.g. mean change in depression rating scores from baseline) between groups using Wilcoxon's rank sum test. For all dichotomous or categorical outcomes (e.g. response, remission, physician satisfaction), χ^2 tests were used to test for differences between the two treatment groups at week 8. Thresholds for remission were determined using current psychiatric guidelines (HAMD-17 of < 8, QIDS-C16 of < 6, and PHQ-9 of < 5). The Cochran–Mantel–Haenszel test was used to calculate reported odds ratios (ORs). Fisher's
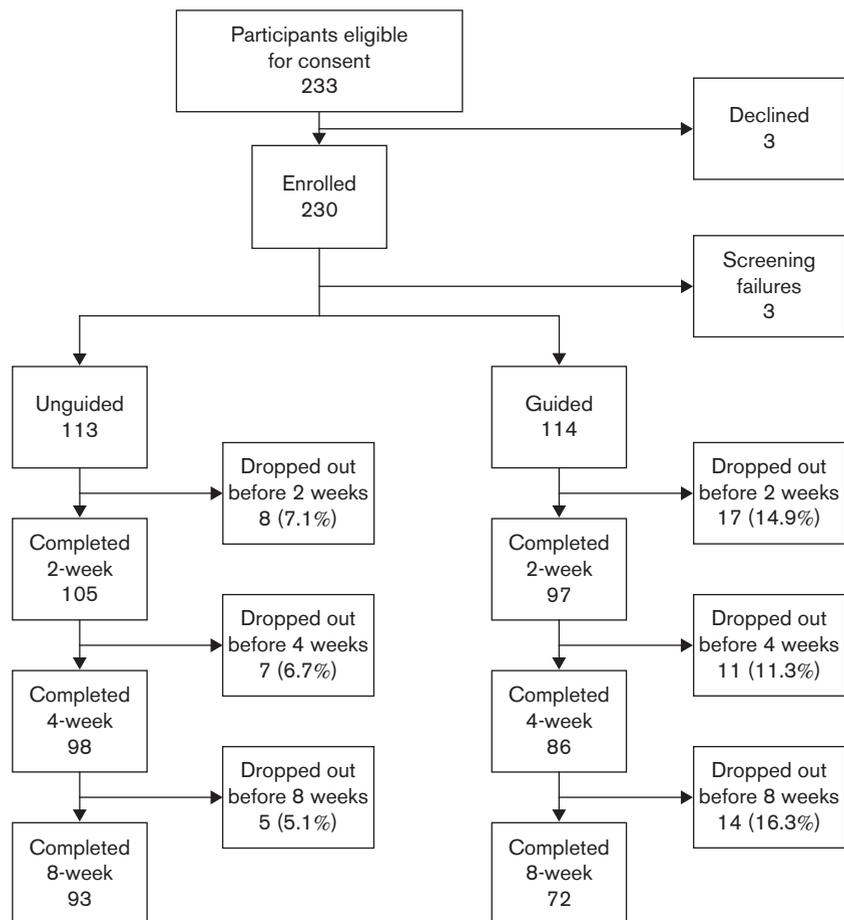
Fig. 3

Diagram of participant accrual and dropouts.

Table 1   Demographic characteristics of study population

	Unguided ($n=113$)	Guided ($n=114$)	P value[a]
Age	44.0 (12.1)	41.0 (12.8)	NS
Sex (female) (%)	77.0	69.3	NS
HAMD-17 score at baseline	22.5 (5.4)	23.0 (5.07)	NS
QIDS-C16 score at baseline	16.0 (3.8)	17.5 (3.7)	0.003
PHQ-9 score at baseline	16.9 (5.6)	17.5 (5.5)	NS
Previous psychiatric medication trials	4.7 (3.52)	3.6 (3.50)	0.021
Previous panel medication trials	3.9 (2.81)	3.0 (3.03)	0.026
Baseline psychiatric medications	2.6 (1.56)	2.2 (1.48)	NS
Baseline panel medications	1.5 (0.91)	1.3 (0.90)	NS

Values are represented as mean±SD.
HAMD-17, Hamilton Depression Rating-17 Item; NS, not significant; PHQ-9, Patient Health Questionnaire – 9 Item; QIDS-C16, Quick Inventory of Depressive Symptomatology – Clinician Rated 16 Item.
[a]P values calculated using the Wilcoxon test.

exact test was used to compare physician satisfaction items between the two groups.

We also evaluated clinical outcomes for patient subgroups that were defined by the bin status of the medications they were taking at baseline (i.e. prescribed before entering the study), and these subgroups were compared within and between the unguided and the guided treatment groups using Kruskal–Wallis one-way analysis of variance. Twelve participants in the unguided group and eight participants in the guided group were not taking medications at baseline among the 26 categorized by the interpretive report and thus these 20 participants were not included in this analysis.

Table 2 Distribution of phenotypes by cohort

Gene	Phenotype	Unguided (%)	Guided (%)	P
CYP2D6	Extensive	49.5	36.1	0.03
	Intermediate	33.3	36.1	
	Poor	9.7	25.0	
	Ultrarapid	7.5	2.8	
CYP2C19	Extensive	72.0	75.0	NS
	Intermediate	26.9	22.2	
	Poor	1.1	2.8	
CYP1A2	Extensive	41.5	45.7	NS
	Intermediate	0	1.43	
	Poor	0	0	
	Ultrarapid	58.5	52.9	
SLC6A4	High activity	26.9	41.7	NS
	Moderate activity	54.8	45.8	
	Low activity	18.3	12.5	
HTR2A	Normal activity	12.9	18.1	NS
	Intermediate activity	45.2	56.9	
	Reduced activity	41.9	25	

NS, not significant.
P values were derived using the χ^2-test.

Participants who were prescribed one or more of the medications represented in the GeneSight panel of drugs were classified according to the most severe bin status for any drug. For example, a participant on two medications, one classified in the yellow bin and the other in the red bin, would be assigned to the red bin for all subsequent analyses. Thus, three patient subgroups were defined: those who were taking only medication(s) in the green bin category, those prescribed at least one medication in the yellow bin category, and those with at least one medication in the red bin category. Participants were evaluated on the basis of their bin status using the percent change in QIDS-C16 and HAMD-17 score from baseline to 8 weeks.

Given the high rate of participant attrition, two different methods of data imputation, an expectation maximization (EM) algorithm and a last observation carried forward method (LOCF), were also utilized during post-hoc analyses to cross-validate the observed results. Using SAS software (version 9.3; SAS Institute Inc., Cary, North Carolina, USA), PROC MI was used to generate the first cross-validation data set. This procedure combines Markov-chain Monte Carlo simulation with the EM algorithm to arrive at a value for a given column vector (i.e. variable) whose probability represents the greatest maximum likelihood of occurrence, as modeled by the observed data. This value is imputed for all missing observations within the given column vector. The process is then repeated for all variables (i.e. columns) with missing values. The second method of cross-validation used LOCF to create a second data set, using the last observed value for each participant for a given measure and carrying the last observed value forward for each remaining time point that was missed because the participant dropped out of the study.

In all analyses, two-tailed tests were conducted and significance was declared when P was less than 0.05.

The Sidak correction was applied to all planned contrasts to account for multiple testing. Analyses were performed using the JMP (SAS Institute Inc.) and SAS software packages.
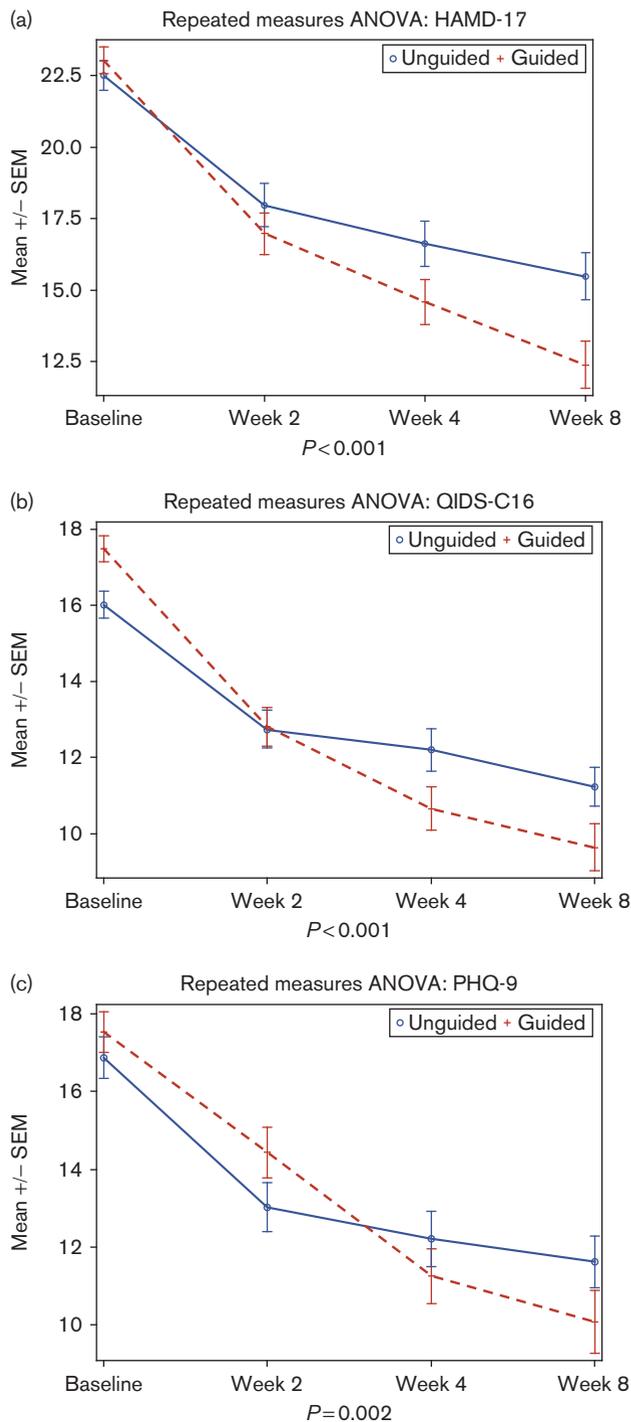
Results
Depression outcomes
Using repeated measures analysis for all participants who completed the 8-week study, a greater reduction of symptoms was observed for guided group participants compared with unguided group participants for HAMD-17 ($F = 22.40$, $P < 0.0001$), QIDS-C16 ($F = 29.70$, $P < 0.0001$), and PHQ-9 ($F = 7.07$, $P = 0.002$). Although the scores for each of the three scales were similar at baseline and at week 2, all three clinical scores declined more rapidly by week 4 in the guided group than in the unguided treatment group. The separation was largest at week 8 (Fig. 4), indicating more rapid improvement by participants whose pharmacogenomic report was available to help guide their physician regarding medication decisions.

The repeated measures analysis was also carried out using data derived from EM and LOCF imputation. The trends and results obtained from the data set using EM remained highly significant and ran parallel to those obtained from the analysis of completing patients. Significantly greater reductions in depression ratings were observed across the duration of the study for the guided group using HAMD-17 ($F = 15.43$, $P = 0.0004$), QIDS-C16 ($F = 25.98$, $P < 0.0001$), and PHQ-9 ($F = 6.14$, $P = 0.004$). Results using LOCF were also statistically significant. Greater reductions in depression ratings were observed across the duration of the study for the guided group using the HAMD-17 ($F = 17.62$, $P < 0.0001$), QIDS-C16 ($F = 23.66$, $P < 0.0001$), and PHQ-9 ($F = 4.02$, $P = 0.03$).

Endpoint analysis compared the unguided and guided treatment groups at 2, 4, and 8 weeks as a function of the mean percent reduction in clinical scores from baseline at each time point. Significant differences were not observed at week 2 for any time point. However, the guided group showed significantly greater percent reductions at week 4 for the HAMD-17 ($z = 2.82$, $P = 0.0002$) and QIDS-C16 ($z = 3.69$, $P = 0.0002$) scales. The guided group showed significant improvement with all three measures at the 8-week time point. The 46.9% reduction in HAMD-17 score in the guided group at week 8 exceeded the 29.9% reduction among the unguided group ($z = 3.14$, $P < 0.0001$). The 44.8% reduction in QIDS-C16 score from baseline to the 8-week visit in the guided group exceeded the 26.4% reduction in the unguided group ($z = 3.24$, $P < 0.0001$). Using the PHQ-9 scale, a 40.1% reduction in PHQ-9 score was observed in the guided group compared with a 19.5% reduction in the unguided group ($z = 3.26$, $P < 0.0001$; Fig. 5).

Fig. 4

(a)

Repeated measures ANOVA: HAMD-17

P < 0.001

(b)

Repeated measures ANOVA: QIDS-C16

P < 0.001

(c)

Repeated measures ANOVA: PHQ-9

P = 0.002

Repeated measures analysis by depression scale. Results from mixed-model ANOVA. Points represent mean for each group at each time point. Bars represent SEM. ANOVA, analysis of variance; HAMD-17, Hamilton Depression Rating-17 Item; PHQ-9, Patient Health Questionnaire – 9 Item; QIDS-C16, Quick Inventory of Depressive Symptomatology – Clinician Rated 16 Item.

Endpoint analysis was repeated using data derived from the two imputation methods. The endpoint analysis using EM data showed nearly identical results to the initial analysis. Using the LOCF data, the guided group showed significantly greater improvement at weeks 2 ($P = 0.001$), 4 ($P < 0.0001$), and 8 ($P < 0.0001$) for HAMD-17. QIDS-C16 and PHQ-9 showed significantly greater improvements ($P < 0.05$) at week 2 and 8, but not week 4.

Significantly greater response rates (i.e. a $\geq 50\%$ reduction in score from baseline) were found in the guided group versus the unguided group at week 8 (Fig. 6), as determined with QIDS-C16; 44.4% of participants in the guided group responded, compared with 23.7% of participants in the unguided group [OR = 2.58; 95% confidence interval (CI) 1.33–5.03; $P = 0.005$]. HAMD-17 (OR = 2.06; 95% CI 1.07–3.95; $P = 0.03$) and PHQ-9 (OR = 2.27; 95% CI 1.20–4.30; $P = 0.01$) showed nearly identical results. This analysis was repeated with the data sets including imputed values. Using EM, response rates were significant for HAMD-17 (OR = 2.69; 95% CI 1.57–4.60; $P = 0.0003$) and PHQ-9 (OR = 2.95; 95% CI 1.71–5.08; $P < 0.0001$), but not QIDS-C16. The LOCF method produced significant response rates at week 8 with QIDS-C16 (OR = 3.06; 95% CI 1.70–5.51; $P = 0.0002$), HAMD-17 (OR = 3.59; 95% CI 2.03–6.36; $P < 0.0001$), and PHQ-9 (OR = 3.93; 95% CI 2.23–6.95; $P < 0.0001$).

Significantly greater remission rates at week 8 were also obtained (Fig. 6). Guided participants had a higher rate of remission (26.4%) than unguided participants (12.9%) as measured by QIDS-C16 (OR = 2.42; 95% CI 1.09–5.39; $P = 0.03$). Remission rates as measured by the HAMD-17 and PHQ-9 followed a similar trend, but did not reach significance (Fig. 6). However, imputation analysis of remission was significant ($P < 0.05$) across most measures (Table 3). Using EM, remission was significant for HAMD-17 (OR = 2.34; 95% CI 1.37–3.99; $P = 0.002$) and PHQ-9 (OR = 2.57; 95% CI 1.49–4.41; $P = 0.0006$), but not QIDS-C16. Using LOCF, the rate of remission at week 8 in guided participants using HAMD-17 was 38.6%, compared with 17% in unguided participants (OR = 2.92; 95% CI 1.58–5.39; $P = 0.0005$). Significance was also obtained for remission with QIDS-C16 (OR = 3.27; 95% CI 1.62–6.61; $P = 0.0007$) and PHQ-9 (OR = 3.85; 95% CI 1.92–7.72; $P < 0.0001$).

Difference in outcome by bin status and treatment group

We examined the participants' medication bin classifications within their treatment groups (guided and unguided) to evaluate clinical outcomes as a function of participants' bin status. No differences in the distribution of bin status at baseline were observed. A significant association between bin status and outcome was observed within the unguided group with HAMD-17 ($z = 2.22$, $P = 0.02$) and QIDS-C16 ($z = 3.96$, $P = 0.02$). Participants classified in the red bin category had less improvement (16.6%) from baseline to week 8 according

Fig. 5
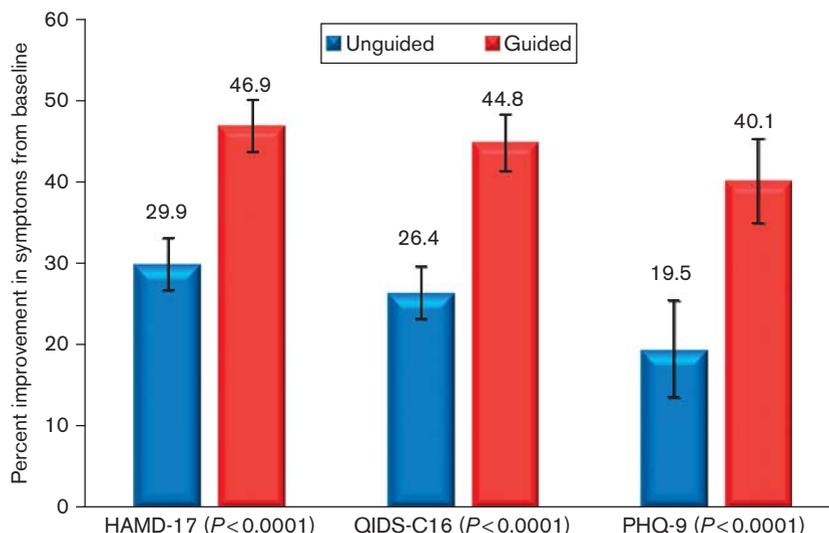
Percent improvement in depression ratings. Values above histogram bars represent group mean. Bars represent SEM. *P* values are derived using the Wilcoxon rank sum test. HAMD-17, Hamilton Depression Rating-17 Item; PHQ-9, Patient Health Questionnaire – 9 Item; QIDS-C16, Quick Inventory of Depressive Symptomatology – Clinician Rated 16 Item.

to HAMD-17 than those not classified in this category (36.1%; $z = -2.76$, $P = 0.007$; Fig. 7). Participants classified in the red bin category had less improvement (11%) from baseline to week 8 according to QIDS-C16 than those not classified in this category (32.3%; $z = -2.76$, $P = 0.007$). In the guided group, neither HAMD-17 nor QIDS-C16 demonstrated significant differences in improvement between participants classified in the red bin category compared with participants classified in other bin categories because of medication changes guided by the report (Fig. 7).

Differential improvement by bin status was suggested in the guided group in the imputed data sets. Using EM, participants in the red bin outperformed participants in the green bin (HAMD-17, $P = 0.02$; QIDS-C16, $P = 0.03$) and yellow bin (HAMD-17, $P = 0.06$; QIDS-C16, $P = 0.07$) because of medication changes guided by the report. Using the LOCF, data set produced similar improvement for participants in the red bin relative to participants in the green bin (HAMD-17, $P = 0.03$; QIDS-C16, $P = 0.09$). However, no significant difference in percent improvement as a function of bin status was observed in the unguided group using either imputation method.

Participants were then compared within their advisory categories across treatment groups. Guided participants classified in the red bin category had a greater percent improvement (42.5%) according to the HAMD-17 than unguided participants (16.6%) in the same advisory category ($P = 0.01$) (Fig. 7). Similarly, guided participants classified in the red bin category had a greater percent improvement (41.9%) according to the QIDS-C16 than
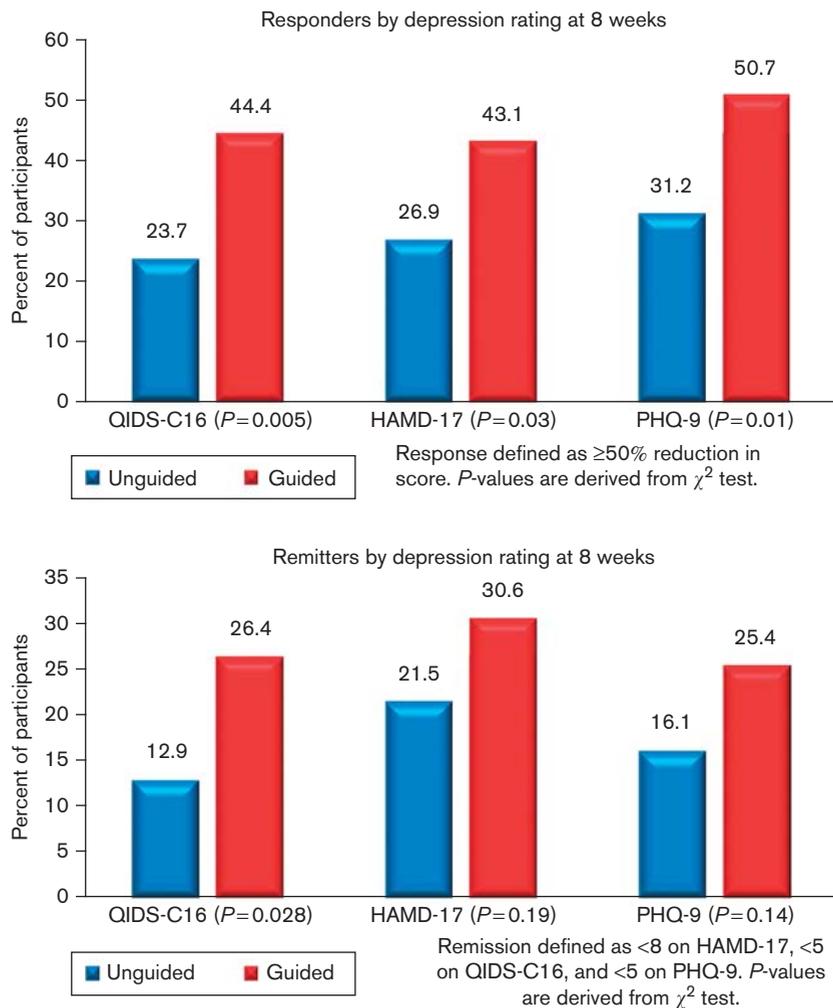
unguided participants (11%) in the same advisory category ($P = 0.004$). Similar comparisons of guided and unguided groups classified in the yellow bin category was significant with QIDS-C16 ($P = 0.02$), but not HAMD-17. Comparisons between guided and unguided participants classified in the green bin category were significant in the same direction with HAMD-17 ($P = 0.05$), but not QIDS-C16 (Fig. 7).

Using the EM data set and QIDS-C16, guided participants in the red bin category had a greater percent improvement (57.5%) than unguided participants (26.5%) in the same category ($F = 11.4$, $P = 0.0009$). Analysis of percent improvement between guided (41.9%) and unguided (23.1%) participants in the yellow bin category was also significant ($F = 10.4$, $P = 0.002$). No significant differences were observed between guided (38.6%) and unguided (35.6%) participants in the green bin category ($F = 0.17$, $P = 0.68$). Results with LOCF applied to QIDS-C16 values also showed significant differences between guided and unguided participants in the red ($F = 4.3$, $P = 0.04$) and yellow bin categories ($F = 4.0$, $P = 0.049$), but not the green bin category ($F = 0.7$, $P = 0.4$).

Pharmacogenomic report utilization and physician satisfaction

Physicians changed participants' medication regimens (i.e. medications were switched, augmented, or dose was adjusted) more often for participants in the guided group (76.8%) than the unguided group (44.1%) ($\chi^2 = 17.33$; $P < 0.0001$; Fig. 8). Of the 18 unguided participants classified in the red bin category at baseline, only 10

Fig. 6

Responders by depression rating at 8 weeks

Response defined as ≥50% reduction in score. *P*-values are derived from χ^2 test.

Remitters by depression rating at 8 weeks

Remission defined as <8 on HAMD-17, <5 on QIDS-C16, and <5 on PHQ-9. *P*-values are derived from χ^2 test.

Response and remission percentages. Values above histogram bars represent group mean. *P* values are derived using χ^2 tests. HAMD-17, Hamilton Depression Rating-17 Item; PHQ-9, Patient Health Questionnaire – 9 Item; QIDS-C16, Quick Inventory of Depressive Symptomatology – Clinician Rated 16 Item.

(55.6%) experienced a medication change or dose adjustment during the 8-week study period, compared with 15 of 16 participants in the guided group (93.8%) in the red bin category ($\chi^2 = 6.35$; $P = 0.01$; Fig. 8). Although the differences were not significant, at the end of the 8 weeks, 40% of the guided group was on green bin medications (up from 26.6% at baseline) compared with 27.6% of the unguided group (up from 25.9% at baseline).

Physicians were directed to complete a survey for each participant detailing their experiences during the study period. Physicians completed surveys for 88 participants (78%) from the unguided group and for 37 participants (32%) from the guided group. Physicians also reported on their perception of each patient's satisfaction with their care. Physician perception of patient satisfaction increased in the guided group with 40.5% reporting very

high satisfaction compared with 14.8% in the unguided group ($\chi^2 = 11.71$; $P = 0.008$). Physician satisfaction with care also increased, with the guided group reporting 94.6% satisfaction rate, compared with 61.8% in the unguided group ($\chi^2 = 14.4$; $P = 0.0007$). There was a significant association between physician confidence in choice of medication and treatment group. The proportion of physicians reporting 'confident' or 'very confident' in medication selection was substantially higher in the guided group (91.9%), compared with those in the unguided group (61.8%). This result was significant at P equal to 0.003 ($\chi^2 = 11.41$).

Discussion

For pharmacogenomics to change the paradigm of current psychiatric practice, the following proposed criteria should be met: (a) pharmacogenomic information must

Table 3 Percentage of participants who achieved remission at 8 weeks by depression score with imputed data

	EM	LOCF
QIDS-C16		
Unguided	38.9	11
Guided	42.8	29.8
OR (95% CI)	1.18 (0.70–2.01)	3.27 (1.62–6.07)
P value	0.54	0.0007
HAMD-17		
Unguided	35.4	17.7
Guided	56.1	38.6
OR (95% CI)	2.34 (1.37–3.99)	2.92 (1.58–5.39)
P value	0.002	0.0005
PHQ-9		
Unguided	31.0	11.5
Guided	53.5	33.3
OR (95% CI)	2.57 (1.49–4.41)	3.85 (1.92–7.72)
P value	0.0006	<0.0001

Odds ratios for guided versus unguided were derived using the Cochran–Mantel–Haenszel test. P values were derived using the χ^2-test.
CI, confidence interval; EM, expectation maximization; HAMD-17, Hamilton Depression Rating-17 Item; LOCF, last observation carried forward method; OR, odds ratio; PHQ-9, Patient Health Questionnaire – 9 Item; QIDS-C16, Quick Inventory of Depressive Symptomatology – Clinician Rated 16 Item.

be predictive of those individuals whose specific treatments are likely to be intolerable or nonefficacious; (b) pharmacogenomic information must be easily integrated into the clinical workflow; and (c) it must effectively guide treatment decisions, resulting in improved clinical outcomes. The present study replicated a smaller prospective pilot, which also examined these characteristics. In the previous pharmacogenomic study ($n = 44$), the reduction in depression scores from the baseline to the 8-week visit was greater in the guided than in the unguided group (e.g. a 31.2% reduction in depression scores for guided participants compared with the 7.2% reduction in QIDS-C16 depression scores for unguided participants was significant at $P = 0.002$) [21]. The findings of the present study replicate the magnitude of effect previously observed and expand upon these earlier results. In addition, the results of the present study demonstrate the ability of the GeneSight test and interpretive report to meet the criteria for predictive value, workflow integration, and improved patient outcomes.

A major goal of psychiatric pharmacogenomics is to prospectively predict individuals who are more or less likely to have a favorable outcome with specific pharmacotherapies. In this study, unguided participants classified in the red bin category exhibited the poorest clinical response when compared with the unguided yellow bin and green bin categories and when compared with all three advisory categories in the guided group. Given that unguided participants were unaware of their pharmacogenomic test results, this affirms the clinical validity of the test and its ability to identify problematic gene–drug interactions that lead to poor outcomes.

In order for pharmacogenomic testing to be of value, it must be appropriately integrated into clinical practice. Jürgens et al. [29], found little clinical impact of CYP450 genotyping when test results were 'absent from tools used for adjustment of pharmacological treatment' and when test results arrived 'too late' to affect treatment decisions. Compared with the 3-week turnaround time observed by Jürgens and colleagues, the genotyping and accompanying interpretive report used in the present study had a turnaround time of 2 business days. In addition, Jürgens and colleagues observed that clinical action was taken only 52% of the time for participants with a 'deviant' gene–drug interaction. Our results, however, indicate that physicians took clinical action for 93.8% of the guided participants on a red bin medication by switching them to medications in the green bin category or tailoring the dose of the medication to reflect their metabolic capacity (Fig. 8).

As an example of the diverse ways in which physicians utilize pharmacogenomic information, one patient in the guided group on a medication in the red bin who did not undergo dose or medication change was originally placed on citalopram 20 mg daily and the dose was not increased throughout the study. This participant was a poor metabolizer (*2/*2) for CYP2C19 polymorphism, which contributed to citalopram being moved to the red bin for this individual. The FDA has created a recommendation regarding citalopram and CYP2C19 metabolizer status stating, '20 mg/day is the maximum recommended dose for patients who are CYP2C19 poor metabolizers because [this factor leads] to increased blood levels of citalopram, increasing the risk of QT interval prolongation and Torsades de Pointes' [30]. Ostensibly, by not increasing the suggested initial dose, the clinician was following current FDA recommendations regarding the gene–drug interaction for citalopram.

When pharmacogenomic testing is utilized to support medication treatment decisions, patient outcomes should improve, particularly for those individuals who may have a significant gene–drug interaction. In the present study, robust results were achieved with pharmacogenomic-guided treatment and were superior to the improvements seen in the unguided group utilizing repeated measures analysis, endpoint analysis, response analysis, and remission analysis. Improvements were evident by week 4 of the study and continued until the end of the study. Overall, participants in the guided group experienced a 1.5–2-fold improvement in outcomes associated with symptom improvement and were 2.4 times more likely to achieve remission than participants in the unguided group.

Guided participants who were classified in the red bin category at baseline performed equally as well by the end of the study as did participants initially on medications classified in the green bin or yellow bin categories because of changes in treatment supported by the report. Further, patients on red bin medications in the guided group who were switched to green bin medications showed a consistent two- to three-fold improvement.

Fig. 7
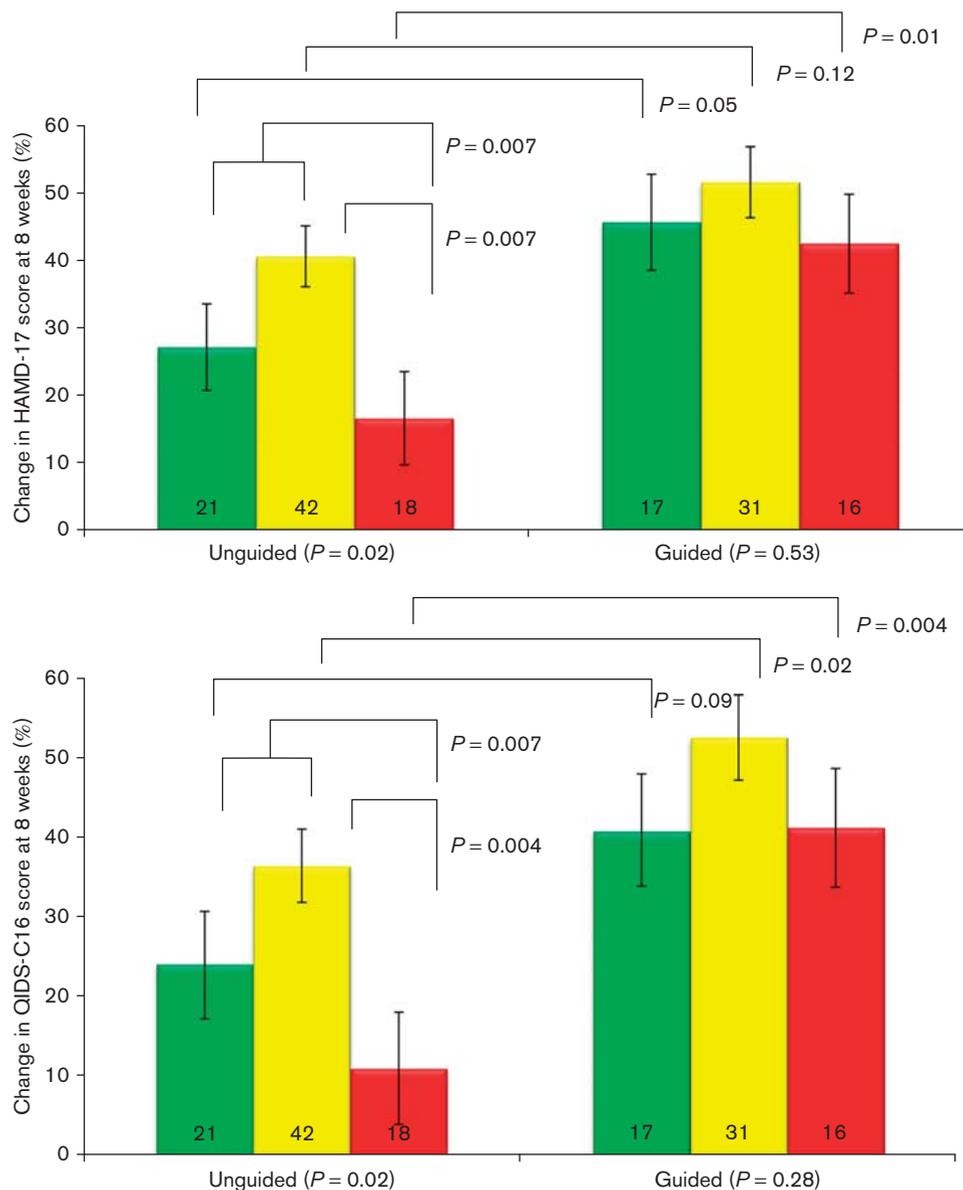
Differences in outcome by bin status. Green, yellow, and red histogram bars represent the 'Use as directed,' 'Use with caution,' and 'Use with increased caution and with more frequent monitoring' bins, respectively. Values in histogram bars represent group sample size. Bars represent SEM. P values are derived using factorial ANOVA. ANOVA, analysis of variance; HAMD-17, Hamilton Depression Rating-17 Item; QIDS-C16, Quick Inventory of Depressive Symptomatology – Clinician Rated 16 Item.

Thus, GeneSight pharmacogenomic-informed interventions (e.g. implementing alternative pharmacologic therapies, adjusting dose of medications) in the guided group enhanced the improvement in those specific individuals with the most potential for gene–drug interactions (Fig. 7).

Beyond statistical significance, pharmacogenomic testing must also produce a clinically meaningful difference in patient outcomes relative to empirical treatment as usual prescribing. Historically, antidepressant therapies have

been found to have a consistent, but relatively narrow improvement over placebo response [31–33]. For example, a meta-analysis of early multicenter controlled studies of the use of fluoxetine versus placebo showed a mean difference of only 2.8 in HAMD-17 score change from baseline [34]. In our study, at 8 weeks the guided group had achieved a mean 10.9-point drop from baseline with the HAMD-17, compared with a 6.5-point drop in the unguided group ($P < 0.0001$). This 4.4-point difference in HAMD-17 score between treatment as usual (unguided) and pharmacogenomic-informed (guided)

Fig. 8
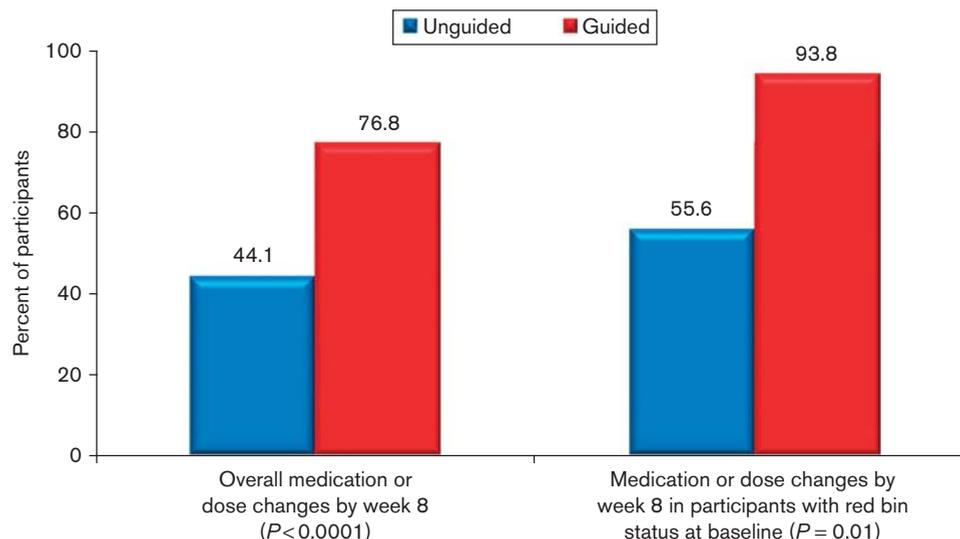
Medication adjustments based on pharmacogenomic report utilization. Values above histogram bars represent group mean. P values are derived using χ^2 tests.

treatment represents a 57% increase over the 2.8-point differential improvement seen for fluoxetine monotherapy for major depressive disorder. Further, the 4.4-point difference exceeds the 3-point standard for clinical significance established by the National Institute for Clinical Excellence [35].

One limitation of the study was the higher dropout rate in the guided group relative to the unguided group. A possible explanation could be that poorly performing participants in the guided group could have dropped out of the trial, thereby inflating the mean improvement of this group. However, there were no significant differences in depression scores between completers and dropouts in either group at any time point. In addition, conservative data imputation methods (e.g. LOCF) produced results that were commensurate with the findings for those who completed the trial. Hypothetically the increased dropout rate in the guided group may have been a result of differential timing of incentives. Participants in the unguided group received their test results only after they completed the trial, which may have increased the incentive to continue participation through the full 8 weeks of the study. Conversely, the guided group received their test results within 2 days of beginning the trial. This differential timing of access to pharmacogenomic information may have resulted in a greater retention of participants in the unguided group, but does not appear to have influenced the differential outcome.

Because of the naturalistic study design, one must also take into account the possible role of the placebo effect. Patients in the guided group had knowledge of their genotyping results and this could have influenced treatment response. However, one would expect a placebo response to affect treatment earlier, rather than at later time points (where differences were greater), as test results were available to physicians within 2 days. The fact that the two groups did not differ in their depression rating scores in the first 4 weeks of the study also argues against an explanation of our results based on a placebo effect. Moreover, the fully blinded unguided group showed the least improvement in those for whom test results predicted the poorest response (i.e. those entering the study on red bin medications). Although it remains a possible confounding factor, placebo effect alone appears insufficient to explain the greater improvement in the guided group. Neither does it explain the ability of the pharmacogenomic testing to identify the unguided patients with the poorest prognosis.

It was noted that the guided group had more *CYP2D6*-poor metabolizer phenotypes than did the treatment-as-usual group. However, adding *CYP2D6* phenotype status as a covariate during statistical modeling did not change any of the original parameter estimates by more than 10%. Therefore, the differential distribution of *CYP2D6* phenotypes between groups at baseline does not appear to have imparted a statistically meaningful effect and was not considered to be a confounder. All other comparisons between treatment group and phenotype status were nonsignificant. As depressive symptoms were higher in the guided group as measured by the QIDS-C16, one would hypothesize that this group would have had a greater difficulty in reaching remission or reaching an average symptom score less than the unguided group. However, the guided group's improvement was

significant over treatment-as-usual on both of these measures. Other limitations of this study include the almost exclusively European ancestry of the participants, and the slightly higher number of previous medication trials in the unguided group relative to the guided group. Replication in additional studies of patients of diverse ancestral origins would further increase confidence in the broader utility of these findings.

Study participants averaged approximately four prior medication trials for their illness, and were therefore more likely to have had a more complex treatment history. This suggests that most participants had treatment-resistant depression. Thus, the positive response of the patients in the guided group is encouraging for this difficult-to-treat population.

An important factor in the introduction of any new technology into clinical practice is the degree of physician confidence in its use and its impact on its ability to improve patient care. Physicians reported increased confidence in medication choices in the guided group compared with the unguided group (91.9 and 61.8% respectively; $P < 0.0001$). This may be expected to strengthen the physician–patient relationship as patient confidence in their caregivers is enhanced and as the decision-making process is clarified. The individualized guidance afforded by psychiatric pharmacogenomic testing and improvement in physician attitudes and confidence in their treatment decisions represent critical steps to more efficiently incorporate evidence-based technology into routine clinical practice.

Conclusion
This study demonstrates that 'real-world' clinical application of pharmacogenomic testing can facilitate individualized treatment strategies for patients prone to treatment resistance or suboptimal medication efficacy. The GeneSight multigenetic test and interpretive report represents an advance in translational medicine and is being increasingly used in clinical practice. The current findings replicate a previously published GeneSight clinical study and demonstrate how depression outcomes can be improved when pharmacogenomic information is translated into clinically relevant guidance at the point-of-care.

Acknowledgements

Conflicts of interest
There are no conflicts of interest.

References
1 Kemp AH, Gordon E, Rush AJ, Williams LM. Improving the prediction of treatment response in depression: integration of clinical, cognitive, psychophysiological, neuroimaging, and genetic measures. *CNS Spectr* 2008; **13**:1066–1086, quiz 1087–1088.
2 Rush AJ, Trivedi MH, Wisniewski SR, Stewart JW, Nierenberg AA, Thase ME, et al. Bupropion-SR, sertraline, or venlafaxine-XR after failure of SSRIs for depression. *N Engl J Med* 2006; **354**:1231–1242.
3 Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry* 2006; **163**:1905–1917.
4 Lieberman JA, Stroup TS, McEvoy JP, Swartz MS, Rosenheck RA, Perkins DO, et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *N Engl J Med* 2005; **353**:1209–1223.
5 Thase ME. STEP-BD and bipolar depression: what have we learned? *Curr Psychiatry Rep* 2007; **9**:497–503.
6 Simon GE, Perlis RH. Personalized medicine for depression: can we match patients with treatments? *Am J Psychiatry* 2010; **167**:1445–1455.
7 Perlis RH. Pharmacogenetic studies of antidepressant response: how far from the clinic? *Psychiatr Clin North Am* 2007; **30**:125–138.
8 Mrazek DA. *Psychiatric Pharmacogenomics*. New York, NY: Oxford University Press; 2010.
9 Kirchheiner J, Nickchen K, Bauer M, Wong ML, Licinio J, Roots I, Brockmoller J. Pharmacogenetics of antidepressants and antipsychotics: the contribution of allelic variations to the phenotype of drug response. *Mol Psychiatry* 2004; **9**:442–473.
10 Mrazek DA, Biernacka JM, O'Kane DJ, Black JL, Cunningham JM, Drews MS, et al. CYP2C19 variation and citalopram response. *Pharmacogenet Genomics* 2011; **21**:1–9.
11 Licinio J, Wong ML. Pharmacogenomics of antidepressant treatment effects. *Dialogues Clin Neurosci* 2011; **13**:63–71.
12 Rundell JR, Shinozaki G. Pharmacogenomic considerations in patients with both comorbid medical and psychiatric illness. *Prim Psychiatry* 2010; **17**:33–38.
13 Hall-Flavin DK, Schneekloth TD, Allen JA. Translational psychiatry: bringing pharmacogenomic testing into clinical practice. *Prim Psychiatry* 2010; **17**:39–44.
14 Kung S, Li X. The clinical use of pharmacogenomic testing in treatment-resistant depression. *Prim Psychiatry* 2010; **17**:46–51.
15 Wall CA, Oldenkamp C, Swintak C. Safety and efficacy pharmacogenomics in pediatric psychopharmacology. *Prim Psychiatry* 2010; **17**:53–58.
16 Rundell JR, Staab JP, Shinozaki G, Saad-Pendergrass D, Moore K, McAlpine D, Mrazek D. Pharmacogenomic testing in a tertiary care outpatient psychosomatic medicine practice. *Psychosomatics* 2011; **52**:141–146.
17 Rundell JR, Staab JP, Shinozaki G, McAlpine D. Serotonin transporter gene promotor polymorphism (5-HTTLPR) associations with number of psychotropic medication trials in a tertiary care outpatient psychiatric consultation practice. *Psychosomatics* 2011; **52**:147–153.
18 Mrazek DA, Lerman C. Facilitating clinical implementation of pharmacogenomics. *JAMA* 2011; **306**:304–305.

19 Collins FS. Reengineering translational science: the time is right. *Sci Transl Med* 2011; **3**:90cm17.
20 Gaughran F, Kapur S. How change comes: translating biological research into care. *The Psychiatrist* 2011; **35**:321–324.
21 Hall-Flavin DK, Winner JG, Allen JD, Jordan JJ, Nesheim RS, Snyder KA, et al. Using a pharmacogenomic algorithm to guide the treatment of depression. *Transl Psychiatry* 2012; **2**:e172. doi:10.1038/tp.2012.99.
22 Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960; **23**:56–62.
23 Hamilton M. Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol* 1967; **6**:278–296.
24 Rush AJ. Sequenced treatment alternatives to relieve depression (STAR*D) protocol. *Control Clin Trials* 2001; **25**:119–142.
25 Trivedi MH, Rush AJ, Ibrahim HM, Carmody TJ, Biggs MM, Suppes T, et al. The Inventory of Depressive Symptomatology, Clinician Rating (IDS-C) and Self-Report (IDS-SR), and the Quick Inventory of Depressive Symptomatology, Clinician Rating (QIDS-C) and Self-Report (QIDS-SR) in public sector patients with mood disorders: a psychometric evaluation. *Psychol Med* 2004; **34**:73–82.
26 Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001; **16**:606–613.
27 Rush AJ, Bernstein IH, Trivedi MH, Carmody TJ, Wisniewski S, Mundt JC, et al. An evaluation of the quick inventory of depressive symptomatology and the Hamilton rating scale for depression: a sequenced treatment alternatives to relieve depression trial report. *Biol Psychiatry* 2006; **59**:493–501.
28 Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, 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* 2003; **54**:573–583.
29 Jürgens G, Jacobsen CB, Rasmussen HB, Werge T, Nordentoft M, Andersen SE. Utility and adoption of CYP2D6 and CYP2C19 genotyping and its translation into psychiatric clinical practice. *Acta Psychiatr Scand* 2012; **125**:228–237.
30 FDA Drug Safety Communication: revised recommendations for Celexa (citalopram hydrobromide) related to a potential risk of abnormal heart rhythms with high doses. 28 March 2012. Available at: *http://www.fda.gov/Drugs/DrugSafety/ucm297391.htm* [Accessed 11 December 2012].
31 Perahia DG, Wang F, Mallinckrodt CH, Walker DJ, Detke MJ. Duloxetine in the treatment of major depressive disorder: a placebo- and paroxetine-controlled trial. *Eur Psychiatry* 2006; **21**:367–378.
32 Rossi A, Barraco A, Donda P. Fluoxetine: a review on evidence based medicine. *Ann Gen Hosp Psychiatry* 2004; **3**:1–8.
33 Apler A. Citalopram for major depressive disorder in adults: a systematic review and meta-analysis of published placebo-controlled trials. *BMJ Open* 2011; **1**:e000106.
34 Stark P, Harrison CD. A review of multicenter controlled studies of fluoxetine vs. imipramine and placebo in outpatients with major depressive disorder. *J Clin Psychiatry* 1985; **46**:53–58.
35 National Institute for Clinical Excellence. *Depression: management of depression in primary and secondary care. Clinical practice guideline No. 23.* London: NICE; 2004.