Appendix S1

Methods

Study subjects

Patients who met diagnostic criteria for MDD without psychosis or mania and who had a 17-item

Hamilton Depression Rating Scale (HAMD-17) score of at least 14 were eligible for inclusion in

the trial. They should not have been, or currently be on an antidepressant, antipsychotic, or

mood-stabilizing medication. The patients were treated initially with either 10 mg of S-CT or 20

mg of CT. Clinical response was determined using the 16-item Quick Inventory of Depressive

Symptomatology (QIDS-C16) scores 4 weeks and 8 weeks after the initiation of SSRI therapy.

The drug dose could be increased to 20 mg of S-CT or 40 mg of CT at 4 weeks for patients with

insufficient clinical response, and it could also be decreased or treatment could be discontinued if

the patient developed intolerable side effects. Blood samples were obtained for metabolomic

profiling at baseline as well as 4 and 8 weeks. The 4- and 8-week samples were also used to

assay plasma drug and drug metabolite concentrations. DNA for genetic analyses was extracted

from the baseline blood samples. All patients provided written informed consent and the study

was reviewed and approved by the Mayo Clinic Institutional Review Board.

Assay of blood drug and drug metabolite concentrations

Plasma concentrations of S-CT, CT and their metabolites were determined using liquid

chromatography tandem mass spectrometry (LC-MS/MS). A chiral LC column was used for CT

and CT metabolite assays in patients treated with that drug. In brief, 500 µl of plasma, an

internal standard (dextrorphan, 20 µL of 10 µg/mL in methanol), 500 µL of 1 M glycine/1M

NaOH buffer (adjusted to pH = 11.3), and 6 mL of hexane: isopropyl alcohol (95:5) was added.

After mixing in a shaker for 5 min, the samples were centrifuged at 3600 rpm for 10 min at

10°C. The organic phase was evaporated to dryness using a speed vacuum, and the residue was

reconstituted with 100 µL of mobile phase that consisted of acetonitrile:ammonium formate (45

mM in water, adjusted to pH = 4.0); 40:60 (v/v) from which 50 µL was injected onto the LC-

MS/MS system. The LC-MS/MS analysis was performed using an API 2000 MS/MS triple

quadruple system (Applied Biosystems, Foster City, CA) equipped with a turbo ion spray

coupled with a Shimadzu (Columbia, MD) HPLC system consisting of an LC-20AB pump and

an SIL-20A HT autosampler, all controlled by Analyst 1.4.2 software (Applied Biosystems/MDS

Sciex, Foster City, CA) in conjunction with Windows 2000 (Microsoft, Redmond, WA). S-CT

and S-CT metabolites were separated using a 3 µm Luna C18(2) column (100 x 2.00 mm)

(Phenomenex, Torrance, CA), a 4 µM Nova-Pak C18 guard column (Waters, Inc., MA, USA),

and a mobile phase that consisted of acetonitrile:ammonium formate (45 mM in water , adjusted

to pH = 4.0); 40:60 (v/v) delivered at a flow rate of 0.300 mL/min. Citalopram and metabolites

were separated using a 5 µm Chirobiotic V chiral column (250 x 4.6mm) (Sigma-Aldrich, MO,

USA) and a mobile phase that consisted of methanol: triethylamine: acetic acid (100: 0.04: 0.16)

(v/v) delivered at 1.000 mL/min. Mass spectrometry optimization was achieved by adjustment

of both compound-dependent and instrument-dependent parameters for the analytes.

Quantification of R- and S-CT, R- and S-DCT, R- and S-DDCT, and the internal standard

dextrorphan was performed by multiple reactions monitoring (MRM) in a positive mode using

the most abundant precursor/product ion transitions (m/z: 325.3/109.00; 297.2/109.00;

311.2/109.00; and 258.25/157.10 respectively). Drugs and metabolites were quantified using the

ratio of peak area of the analyte to the peak area of the internal standard, and calibration curves were constructed using known concentrations of the analytes added to blank plasma.

Genotyping quality control

Genotyping quality control assessments included concordance rates for duplicate samples and Mendelian inheritance checks performed using a CEPH parent-child trio. For each SNP, the minor allele frequency, call rate and departure from Hardy–Weinberg equilibrium were evaluated. Observed call rates, total heterozygosity and inbreeding coefficients were assessed for each subject using PLINK [1]. Sex-checks based on X-chromosome heterozygosity were performed, and measures of identity-by-descent were used to identify potentially related subjects. After this quality control process was completed, 545,115 genotyped SNPs were available for inclusion in the analysis. Genome-wide imputation was then performed using the BEAGLE software [2] and the November 23, 2010 release of the 1000 Genomes Project as the reference data.

To evaluate the possibility of population stratification, the measured SNPs were used to determine eigenvectors of the SNP correlation matrix, using the EigenStrat software [3] to determine eigenvalues for the SNP correlation matrix. Based on Tracy–Widom test p-values, 4 eigenvalues differed statistically from zero. However, none of the corresponding eigenvectors were associated with plasma drug/metabolite concentrations, so none of the eigenvectors were included in the LMM as adjusting covariates.

Genome-wide imputation

Genome-wide imputation was performed using BEAGLE software [2] and the November 23,

2010 release of the 1000 Genomes Project as reference data. This phased reference data was

oriented on the forward strand with build 37 annotation. There were 1094 subjects included in

the reference data, with a total of 37,138,905 SNPs. For quality purposes, reference SNPs with

minor allele frequency (MAF) values less than 0.005 were removed. This resulted in 12,067,002

un-typed SNPs for imputation. The 545,115 measured SNPs were aligned on the forward "+"

strand using a combination of reference files from genotyping and the BEAGLE utility check

strands.py (from Brian Browning University of Washington). Chromosomes were divided into

20 MB regions, and BEAGLE v3.3.1 was run on these 20 MB regions, plus a 1 MB buffer region

to the right and left of the main region because ends are generally poorly imputed. The reference

data was phased, and the observed genotype data was un-phased. SNPs with a dose $R^2 < 0.3$

were removed from the analyses [4], as well SNPs with a MAF < 0.01, resulting in 7,537,437

SNPs (observed or imputed) that were used in the GWA analyses.

Database mining and informatics analyses

Functional annotation of SNPs was performed with the Ensembl Genome Browser

(http://useast.ensembl.org) and SNPnexus, a web-based annotation tool (http://snp-nexus.org) [5]

as well as GeneMANIA (http://www.genemania.org) for genomic and proteomic data. These

tools were applied to identify function and relationship of genes observed from the top GWAS

associations [6]. Regional association results and LD values were plotted using LocusZoom

(http://csg.sph.umich.edu/locuszoom) [7] and SNAP (SNP Annotation and Proxy Search from

the Broad Institute, http://www.broadinstitute.org/mpg/snap), respectively. The Mouse Phenome

Database at the Jackson Laboratory (http://phenome.jax.org) was used to investigate the function

of mouse homologue genes, including *TRIML1* that were identified during our GWA analyses.

References

1. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81: 559-75.
2. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am J Hum Genet 2009; 84: 210-23.
3. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006; 38: 904-9.
4. de Bakker PI, Ferreira MA, Jia X, Neale BM, Raychaudhuri S, Voight BF. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum Mol Genet 2008; 17: R122-8.
5. Dayem Ullah AZ, Lemoine NR, Chelala C. SNPnexus: a web server for functional annotation of novel and publicly known genetic variants (2012 update). Nucleic Acids Res 2012; 40: W65-70.
6. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, Maitland A, Mostafavi S, Montojo J, Shao Q, Wright G, Bader GD, Morris Q. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res 2010; 38: W214-20.
7. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 2010; 26: 2336-7.

Appendix Table S1. Linear mixed model covariate p-values. A p-value of 0.05 was considered significant.

pheno	dose	age	genderM	Week 8
log(S.Cital)	1.60E-14	1.35E-14	7.38E-04	2.34E-01
log(S.Desmethyl.Cital)	0.00E+00	9.52E-04	4.29E-06	1.70E-01
log(S.Didesmethyl.Cital+.01)	6.16E-04	1.02E-01	5.98E-01	9.40E-01
log10(ratio.S.Des.Cit)	1.49E-01	3.22E-08	7.16E-01	9.56E-01
log10(ratio.S.Dides.Des)	3.73E-01	7.02E-01	2.74E-01	4.69E-01

Appendix Table S2. Top 10 associations for the GWAS of S-desmethylcitalopram (S-DCT) concentrations.

S-Desmethylcitalopram (S-DCT) concentration

SNP	p-value	genotype	MAF	CHR	BP	gene	gene location/function	distance to gene
rs9747992*	1.55E-07	A/G	0.086	17	77804840	CBX4	3'-downstream	8388
rs2059865	2.58E-07	A/T	0.235	5	31766433	PDZD2	intronic	0
rs34884728	4.33E-07	A/G	0.272	5	31781647	PDZD2	intronic	0
rs7707966*	5.10E-07	A/G	0.276	5	31789679	PDZD2	intronic	0
rs10461903	5.99E-07	C/T	0.325	5	31766703	PDZD2	intronic	0
rs9747031	6.13E-07	C/G	0.105	17	77804936	CBX4	3'-downstream	8292
rs117020818	6.80E-07	A/G	0.046	11	13129163	CTC-497E21.5	intronic	0
rs79847705	7.07E-07	A/G	0.047	11	13134780	CTC-497E21.5	intronic	0
rs13182171	8.02E-07	C/A	0.196	5	31812179	PDZD2	intronic	0
rs34301697	8.42E-07	G/A	0.193	5	31813639	PDZD2	intronic	0

CYP2C19-adjusted S-desmethylcitalopram (S-DCT) concentration

SNP	p.snp	genotype	MAF	CHR	BP	gene	gene location/function	distance to gene
rs9747992*	1.16E-07	A/G	0.086	17	77804840	CBX4	3'-downstream	8388
rs9747031	4.32E-07	C/G	0.105	17	77804936	CBX4	3'-downstream	8292
rs2059865	5.09E-07	A/T	0.235	5	31766433	PDZD2	intronic	0
rs77724872	6.84E-07	T/C	0.047	12	104754500	EID3	3'-downstream intragenic	55520
rs117020818	7.15E-07	A/G	0.046	11	13129163	CTC-497E21.5	5'-downstream intragenic	96516
rs8102150	7.18E-07	G/C	0.028	19	50230327	CTB-33G10.1	non-coding /CpG island 43	0
rs3810263	7.21E-07	G/A	0.027	19	50223147	CTB-33G10.1	non-coding /CpG island 43	0
rs79847705	7.44E-07	A/G	0.047	11	13134780	CTC-497E21.5	5'-downstream intragenic	102133
rs3810264	8.22E-07	G/A	0.028	19	50223230	CTB-33G10.1	non-coding /CpG island 43	0
rs3810265	8.39E-07	A/G	0.028	19	50223266	CTB-33G10.1	non-coding /CpG island 43	0

* indicates genotyped SNPs.

Appendix Table S3. Top 10 associations for the GWAS of S-didesmethylcitalopram (S-DDCT) concentration and S-didesmethylcitalopram/S-desmethylcitalopram (S-DCT/S-CT) ratio. Two *CYP2D6* nonsynonymous cSNPs are boxed.
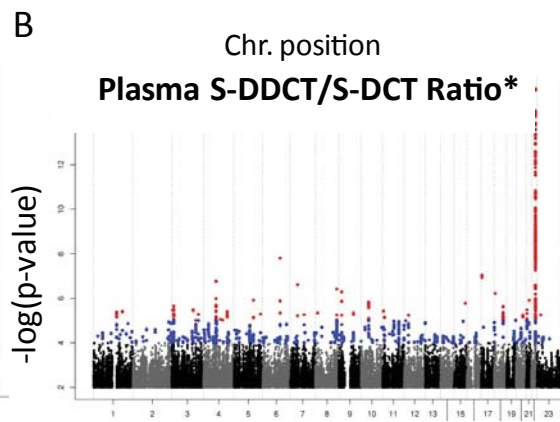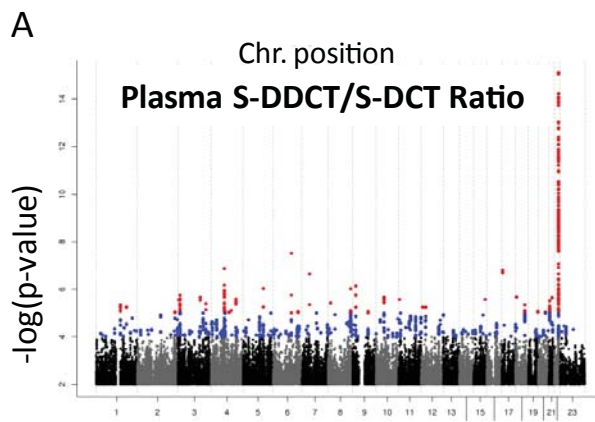
S-Didesmethylcitalopram (S-DDCT) concentration

SNP	p-value for S-CT**	for S-DDCT/ S-DCT**	genotype	MAF	CHR	BP	gene	gene location/function	Distance to *CYP2D6*, bp
c22.42519657.b37p0	2.00E-16	7.99E-16	A/C	0.198	22	42519657	*LOC100132273/CYP2D6*	non-coding intronic/3'-downstream	2844
c22.42533645.b37p0	2.00E-16	7.99E-16	T/C	0.332	22	42533645	*CYP2D6*	5'-upstream	6737
rs1001586*	2.00E-16	7.99E-16	A/C	0.198	22	42670293	*LOC388906/CYP2D6*	non-coding/5'-upstream intragenic	143385
rs1001587*	2.00E-16	7.99E-16	A/G	0.201	22	42670111	*LOC388906/CYP2D6*	non-coding/5'-upstream intragenic	143203
rs1033459	2.00E-16	7.99E-16	G/A	0.198	22	42619067	*CYP2D6*	5'-upstream intragenic	92159
rs1033460	2.00E-16	7.99E-16	G/A	0.198	22	42619308	*CYP2D6*	5'-upstream intragenic	92400
rs1058172	2.00E-16	7.99E-16	T/C	0.155	22	42523528	*CYP2D6*	coding, nonsyn, R365/314H	0
rs1058174	2.00E-16	7.99E-16	A/G	0.151	22	42536263	*CYP2D6*	5'-upstream	9355
rs1062753*	2.00E-16	7.99E-16	A/G	0.307	22	42392811	*SEPT/CYP2D6*	Intronic/5'-upstream intragenic	12960
rs1065852	2.00E-16	7.99E-16	A/G	0.215	22	42526694	*CYP2D6*4,*10 and*14*	coding, nonsyn, P34S	0

* indicates genotyped SNPs.

* p-values remain the same for CYP2C19-adjusted S-DDCT level

Appendix Figure Legends

Figure S1. GWAS analysis results after 1000 Genomes imputation for ratios of S-didesmethylcitalopram (S-DDCT) plasma concentration over its demethylated precursor, S-desmethylcitalopram (S-DCT) (A) Manhattan plot for S-DDCT/S-DCT plasma concentration ratio. (B) Manhattan plot for S-DDCT/S-DCT plasma concentration ratio, adjusted for the effect of *CYP2C19* functional alleles. Red dots represent SNPs with p-values $\leq 10^{-5}$ and blue dots represent SNPs with p-values $\leq 10^{-4}$.

A

Chr. position
Plasma S-DDCT/S-DCT Ratio

B

Chr. position
Plasma S-DDCT/S-DCT Ratio*

Figure S1

* adjusted for the effect of
CYP2C19 functional alleles