Rare Copy Number Variation in Treatment-Resistant Major Depressive Disorder

Colm O'Dushlaine, Stephan Ripke, Douglas M. Ruderfer, Steven P. Hamilton, Maurizio Fava, Dan V. Iosifescu, Isaac S. Kohane, Susanne E. Churchill, Victor M. Castro, Caitlin C. Clements, Sarah R. Blumenthal, Shawn N. Murphy, Jordan W. Smoller, and Roy H. Perlis

Background: While antidepressant treatment response appears to be partially heritable, no consistent genetic associations have been identified. Large, rare copy number variants (CNVs) play a role in other neuropsychiatric diseases, so we assessed their association with treatment-resistant depression (TRD).

Methods: We analyzed data from two genome-wide association studies comprising 1263 Caucasian patients with major depressive disorder. One was drawn from a large health system by applying natural language processing to electronic health records (i2b2 cohort). The second consisted of a multicenter study of sequential antidepressant treatments, Sequenced Treatment Alternatives to Relieve Depression. The Birdsuite package was used to identify rare deletions and duplications. Individuals without symptomatic remission, despite two antidepressant treatment trials, were contrasted with those who remitted with a first treatment trial.

Results: CNV data were derived for 778 subjects in the i2b2 cohort, including 300 subjects (37%) with TRD, and 485 subjects in Sequenced Treatment Alternatives to Relieve Depression cohort, including 152 (31%) with TRD. CNV burden analyses identified modest enrichment of duplications in cases (empirical $p = .04$ for duplications of 100–200 kilobase) and a particular deletion region spanning gene *PABPC4L* (empirical $p = .02$, 6 cases: 0 controls). Pathway analysis suggested enrichment of CNVs intersecting genes regulating actin cytoskeleton. However, none of these associations survived genome-wide correction.

Conclusions: Contribution of rare CNVs to TRD appears to be modest, individually or in aggregate. The electronic health record-based methodology demonstrated here should facilitate collection of larger TRD cohorts necessary to further characterize these effects.

Key Words: Antidepressant, copy number, deletion, duplication, pharmacogenetic, pharmacogenomic, rare genetic variation

A third or more of individuals treated for major depressive disorder (MDD) do not reach symptomatic remission despite multiple adequate antidepressant treatment trials (1). Treatment-resistant depression (TRD), defined as failure to remit despite two or more treatment trials, contributes substantially to the morbidity associated with MDD, increasing health care costs, as well as functional impairment (2), suicide liability, and increased risk of relapse even following remission (1). Despite its clinical importance, little is known of the underlying neurobiology, likely because identification of TRD cohorts requires multiple treatment trials so few such cohorts exist. Identifying genetic associations with TRD could facilitate risk stratification and development of novel interventions for this patient population (3).

From the Stanley Center for Psychiatric Research (CO, SR, JWS, RHP), Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge; and Department of Psychiatry (MF, CCC, SRB, JWS, RHP), Massachusetts General Hospital, Boston, Massachusetts; Department of Psychiatry (SPH), University of California, San Francisco, California; Icahn School of Medicine at Mount Sinai (DMR, DVI), New York, New York; and Information Systems (ISK, SEC, VMC, SNM), Partners HealthCare System; and Laboratory of Computer Science and Department of Neurology (VMC, SNM), Massachusetts General Hospital, Boston, Massachusetts.

Address correspondence to Roy H. Perlis, M.D., Massachusetts General Hospital, Center for Experimental Drugs and Diagnostics, Department of Psychiatry and Center for Human Genetic Research, 185 Cambridge St, Boston, MA 02114; E-mail: rperlis@partners.org.

Received Aug 1, 2013; revised Oct 3, 2013; accepted Oct 26, 2013.

Prior genetic studies of antidepressant response have focused on common variation in individuals receiving a single treatment (4–6), while rarer copy number variants (CNVs) (i.e., deletions and duplications) have not been examined, despite a burgeoning body of evidence implicating them in neuropsychiatric disorders (7–11), including major depressive disorder (12–14). These data suggest that common phenotypes may still be associated with rare variants. In particular, a recent in silico investigation of genes coding for known drug targets suggests the possibility that copy number variation is likely to have large effects on treatment response (15). An alternate hypothesis, also examined here, is that a small subset of individuals with treatment-resistant depression fail to respond to treatment because of phenotypic overlap with another neuropsychiatric disorder mediated by CNVs that may be less responsive to antidepressant treatment; in particular, we anticipated that we might observe an increased frequency of CNVs previously implicated in schizophrenia, autism, or related disorders.

To examine these hypotheses, data were identified from a novel treatment-response cohort drawn from electronic health records (EHR) (16,17), referred to as the i2b2 cohort, as well as from the largest prospective investigation of treatment resistance to date, the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study (1). The former cohort represents one of the first applications in psychiatry of EHR data for genetic investigation, an approach that may be particularly useful for studying rare or otherwise difficult to ascertain clinical phenotypes.

Methods and Materials

Subjects

For the i2b2 cohort, TRD and selective serotonin reuptake inhibitor responsive phenotypes were defined using a previously validated natural language processing tool (16), which classifies

0006-3223/$36.00
http://dx.doi.org/10.1016/j.biopsych.2013.10.028

clinical status cross-sectionally using the adaptive lasso approach to regression, then determines longitudinal outcome with a rules-based classifier. Individuals were defined as treatment-resistant if they had received two or more antidepressants during a period of depression or received electroconvulsive therapy following at least one documented antidepressant treatment trial, who would have been referred because of prior documented treatment failures. Individuals were defined as antidepressant-responsive if they achieved remission with the initial documented antidepressant treatment trial. Notably, we have previously demonstrated that similar treatment effects can be observed in analyses of clinical data from both i2b2 and STAR*D (17), suggesting the relative comparability of these two data sets despite the different means of ascertainment. Because the investigators did not interact with any individuals for the ascertainment of data or samples and samples were de-identified before receipt by investigators, the Massachusetts General Hospital Institutional Review Board elected to waive the requirement of seeking informed consent as detailed by Code of Federal Regulations, Title 45, Part 46, Section 116 (46.116). The sample collection utilizes a one-way hash to ensure that, once matched with phenotypic data, all identifiers are stripped.

For replication, subjects were drawn from the STAR*D cohort (18). Assessment of outcomes has been previously described (19). Treatment resistance was defined for primary analyses as Quick Inventory of Depressive Symptomatology-Self-Report of 10 or greater after two antidepressant trials as defined in the STAR*D protocol (i.e., guideline-based antidepressant treatment according to dosing parameters at levels 1 and 2). Selective serotonin reuptake inhibitor responsiveness was defined as Quick Inventory of Depressive Symptomatology-Self-Report of 5 or less after one or two antidepressant treatment trials. As the present analysis targeted pharmacologic treatment response, subjects who received cognitive therapy at level 2 of STAR*D were excluded from the analysis. The STAR*D clinical and genetics protocols were approved by institutional review boards at participating sites.

Genotyping and Quality Control

For the i2b2 discovery cohort, DNA was extracted from discarded blood samples. Genotyping for the two waves of this cohort utilized the Illumina Omni 1 MM ($n = 453$) or Omni Express ($n = 488$) array (Illumina Inc., San Diego, California) at the Broad Institute of Massachusetts Institute of Technology and Harvard University; all analyses were therefore stratified by array type. We included only samples with genotyping call rates ≥95%, non-outliers on multidimensional scaling measures of ancestry, and no evidence of substantial relatedness by pi-hat; resulting BeadStudio call rates exceeded 99%. Copy number variants were detected using a hidden Markov model as previously described, using the Birdsuite package (20), which performs well in comparisons with other CNV-calling tools (21). Subjects who failed to pass standard single nucleotide polymorphism (SNP) quality control and those with >20 total CNVs or >10 Mb of total CNV area were excluded. These thresholds were selected based on manual inspection of distributions within each cohort and genotyping platform. Consistent with prior reports (11), CNVs with frequency greater than 1% in any individual data set, those spanning centromeres or other genomic gaps, those overlapping with common CNVs in HapMap, those overlapping events of frequency >1% in the database of genomic variants, those with less than 10 probes/SNPs spanning the event, and those with size <100 kilobase (kb) were excluded.

Details of genotyping for STAR*D are presented elsewhere (4). The STAR*D cohort was originally genotyped on the Affymetrix 500k and 5.0 arrays (Affymetrix, Santa Clara, California); only the latter contains copy number variation probes, but SNP probe intensity may also be applied to identify CNVs, albeit more indirectly and with less precision. We obtained raw intensity data from both platforms from the investigator (S.P.H.) and utilized this data to call CNVs using the Birdsuite package (20). As with the i2b2 cohort, analyses were stratified by array type. The same quality control thresholds and methodology were applied as for the i2b2 cohort.

Analysis

Using an approach consistent with prior CNV analyses (11), we evaluated overall CNV burden for deletions and duplications considered separately, then for tranches of CNV frequencies (occurring once in the data set, or between two and six times) as well as tranches of CNV sizes (100–200 kb, 200–500 kb, and >500 kb). To compare CNV burden between cases and control subjects, one-sided tests were utilized, with 10,000 permutations used to evaluate statistical significance (22). The same approach, in which burden was examined for all duplications or deletions considered together, then for individual tranches, was used to compare proportion of genes intersected by CNVs in cases and control subjects. We also used permutation to identify individual loci where the proportion of CNVs observed in cases versus control subjects exceeded that expected by chance. Loci with specific CNVs that have previously been associated with schizophrenia or autism in a recent meta-analysis (10), as well as those associated with MDD (12–14), were examined to determine whether any were present in the TRD cases versus control subjects.

Finally, we examined curated pathways in Kyoto Encyclopedia of Genes and Genomes (http://www.genome.jp/kegg/) to examine whether individual pathways were enriched for duplications or deletions, using a test for gene-set enrichment described in Raychaudhuri *et al.* (23) implemented in PLINK (24). Such analyses may point to relevant biology even when individual variants fail to meet standard thresholds for statistical significance.

Table 1. Burden of Deletions and Duplications in Individuals with Treatment-Resistant Major Depressive Disorder and Control Subjects

		CNV Burden (Number)		CNV Burden (Gene Count)	
	Total	*p*	Case/Control Ratio	*p*	Case/Control Ratio
Deletions					
All	547	.984	.768	1.000	.475
Frequency					
1	203	.844	.789	.951	.496
2–6	243	.927	.786	.997	.429
Size (kb)					
100–200	359	.994	.732	.996	.512
200–500	150	.793	.871	.997	.382
500+	38	.849	.731	.831	.523
Duplications					
All	780	.235	1.033	.074	1.197
Frequency					
1	286	.260	1.073	.550	.934
2–6	345	.483	.969	.149	1.208
Size (kb)					
100–200	390	.692	.950	.036	1.433
200–500	265	.410	1.036	.615	.943
500+	125	.080	1.321	.205	1.295

CNV, copy number variant; kb, kilobase.