

## Genome-wide Association Study of Dimensional Psychopathology Using Electronic Health Records

Thomas H. McCoy Jr., Victor M. Castro, Kamber L. Hart, Amelia M. Pellegrini, Sheng Yu, Tianxi Cai, and Roy H. Perlis

### ABSTRACT

**BACKGROUND:** Genetic studies of neuropsychiatric disease strongly suggest an overlap in liability. There are growing efforts to characterize these diseases dimensionally rather than categorically, but the extent to which such dimensional models correspond to biology is unknown.

**METHODS:** We applied a newly developed natural language processing method to extract five symptom dimensions based on the National Institute of Mental Health Research Domain Criteria definitions from narrative hospital discharge notes in a large biobank. We conducted a genome-wide association study to examine whether common variants were associated with each of these dimensions as quantitative traits.

**RESULTS:** Among 4687 individuals, loci in three of five domains exceeded a genome-wide threshold for statistical significance. These included a locus spanning the neocortical development genes *RFPL3* and *RFPL3S* for arousal ( $p = 2.29 \times 10^{-8}$ ) and one spanning the *FPR3* gene for cognition ( $p = 3.22 \times 10^{-8}$ ).

**CONCLUSIONS:** Natural language processing identifies dimensional phenotypes that may facilitate the discovery of common genetic variation that is relevant to psychopathology.

**Keywords:** Arousal, Genetic, Genomic, Social, Transdiagnostic, Valence

<https://doi.org/10.1016/j.biopsych.2017.12.004>

Family studies of psychiatric illnesses demonstrated decades ago the overlap in risk for these disorders, a finding that has now been confirmed by genome-wide association studies (1–3). Such overlap highlights the limitations of a nosologic system focused on categories of symptoms rather than dimensions. For this reason, recent initiatives emphasize the utility of identifying symptom domains that may better correspond to underlying neurobiology (4,5).

The rise of biobanks embedded in health care systems or national registries provides an opportunity to investigate the impact of genomic variation in a less biased fashion than traditional disease case-control designs. However, such biobanks typically capture primarily coded clinical data, i.e., categorical diagnoses. We have recently developed multiple methods to examine narrative clinical notes to extract symptom dimensions as a means of augmenting these coded data (6,7).

We hypothesized that symptom dimensions based on expert-curated terms capturing National Institute of Mental Health Research Domain Criteria (RDoC) domains would be associated with common genomic variation and could thereby implicate novel sets of genes related to psychopathology. As proof of concept, we therefore applied a newly described natural language processing (NLP) method for extracting dimensional phenotypes to hospital discharge

summaries drawn from the genomic biobank of an academic medical center (7) and used standard genome-wide association studies to investigate these novel phenotypes as quantitative traits.

### METHODS AND MATERIALS

#### Overview and Data Set Generation

We drew on three waves of participants in the Partners Biobank from the Brigham and Women's Hospital network and the Massachusetts General Hospital network, representing approximately the first 15,000 individuals genotyped as part of the Partners HealthCare Biobank initiative (8). Narrative discharge summaries were extracted from the longitudinal electronic health record of the Massachusetts General Hospital. We included any individuals 18 years of age or older who had at least one hospitalization between 2010 and 2015.

A datamart containing all clinical data was generated with i2b2 server software (version 1.6; i2b2, Boston, MA), a computational framework for managing human health data (9–11). The Partners HealthCare System Institutional Review Board approved both the study protocol and the release of biobank data, which were collected after acquiring written informed consent from participants and explicitly allowed identifiable data to be shared with qualified investigators.

## Study Design and Analysis

Primary analyses used a cohort design with all patients admitted for any reason during the time period noted above. Discharge documentation was used to estimate dimensional psychopathology scores for one encounter per individual; when an individual was hospitalized on multiple occasions during the study period, a single hospitalization was selected at random to minimize bias resulting from other means of ascertainment. The derivation of dimensional psychopathology has been described elsewhere (7); in brief, it began with a set of seed terms for each of the five National Institute of Mental Health RDoC definitions drawn from National Institute of Mental Health workgroup statements, then expanded these term lists to include synonyms (12). This second expansion step is important because it reduces potential bias introduced by a given specialty or set of providers who may use specific terminology to characterize symptoms, yielding a broader set of terms that should better generalize across providers and hospitals. Each note is assigned a score corresponding to a simple count of term appearance. We have developed simple code to facilitate dimension extraction in other data sets (7).

## Genotyping and Quality Control

DNA was extracted from buffy coat, and genotyping was done using three versions of the Illumina Multi-Ethnic Global (MEG) array (Illumina, Inc., San Diego, CA) (MEGA,  $n = 4927$ ; MEGA EX,  $n = 5353$ ; and MEG,  $n = 4784$ ; mappable variants available for each were 1,411,334, 1,710,339, and 1,747,639, respectively). These common variant arrays all incorporate content from the 1000 Genomes Project Phase 3. Single nucleotide polymorphism (SNP) coordinates were remapped based on the TopGenomicSeq provided by Illumina; all reference SNP cluster IDs correspond to build 142 of the Single Nucleotide Polymorphism Database. To determine the forward strand of the SNP, we aligned both SNP sequences (alleles A and B) to hg19 using the BLAST-like alignment tool (BLAT) with default parameters set by the University of California Santa Cruz Genome Browser (13).

Each cohort was cleaned, imputed, and analyzed separately to avoid batch effects. In each batch we included subjects with genotyping call rates exceeding 99%; no related individuals based on identity by descent were included (14). From these individuals, any genotyped SNP with a call rate of at least 95% and a Hardy-Weinberg equilibrium  $p$  value  $< 1 \times 10^{-6}$  was included. Imputation used the Michigan Imputation Server implementing Minimac3 (15–17). Imputation used all population subsets from 1000 Genomes Project Phase 3 version 5 as reference panel; haplotype phasing was performed using SHAPEIT (18).

For each batch, we applied principal components analysis of a linkage-disequilibrium-pruned set of genotyped SNPs to characterize population structure, based on EIGENSTRAT as implemented in PLINK software version 1.9 (19). We then plotted these components with superimposition of HapMap samples to confirm location of Northern European individuals. The present analysis included only individuals of Northern European genomic ancestry to minimize the risk for confounding by ancestry (i.e., population stratification) and

because the power to detect association in other ancestry groups would be limited (20–22).

## Analysis

We examined single-locus associations in each batch, then combined in inverse-variance-weighted fixed effects meta-analysis. In all analyses, only biallelic SNPs with minor allele frequencies of at least 1% in all batches were retained. Tests for association used linear regression assuming an additive allelic effect and examined each of the five dimensional measures as a quantitative trait, with adjustment for the first 10 principal components a priori. (In previous work, analyses incorporating five or 20 components did not yield meaningfully different results.) Association results are presented in terms of independent loci after pruning using the clump command in PLINK, with a 250-kb window and  $r^2 = .2$ . Locus plots were generated using LocusZoom software (19,23).

Reported  $p$  values are not adjusted for lambda or linkage disequilibrium scores; in previous work, adjustment for lambda-1000 or linkage disequilibrium score regression intercept did not meaningfully change relative results. Lambdas ranged from 0.998 to 1.003 (24).

## RESULTS

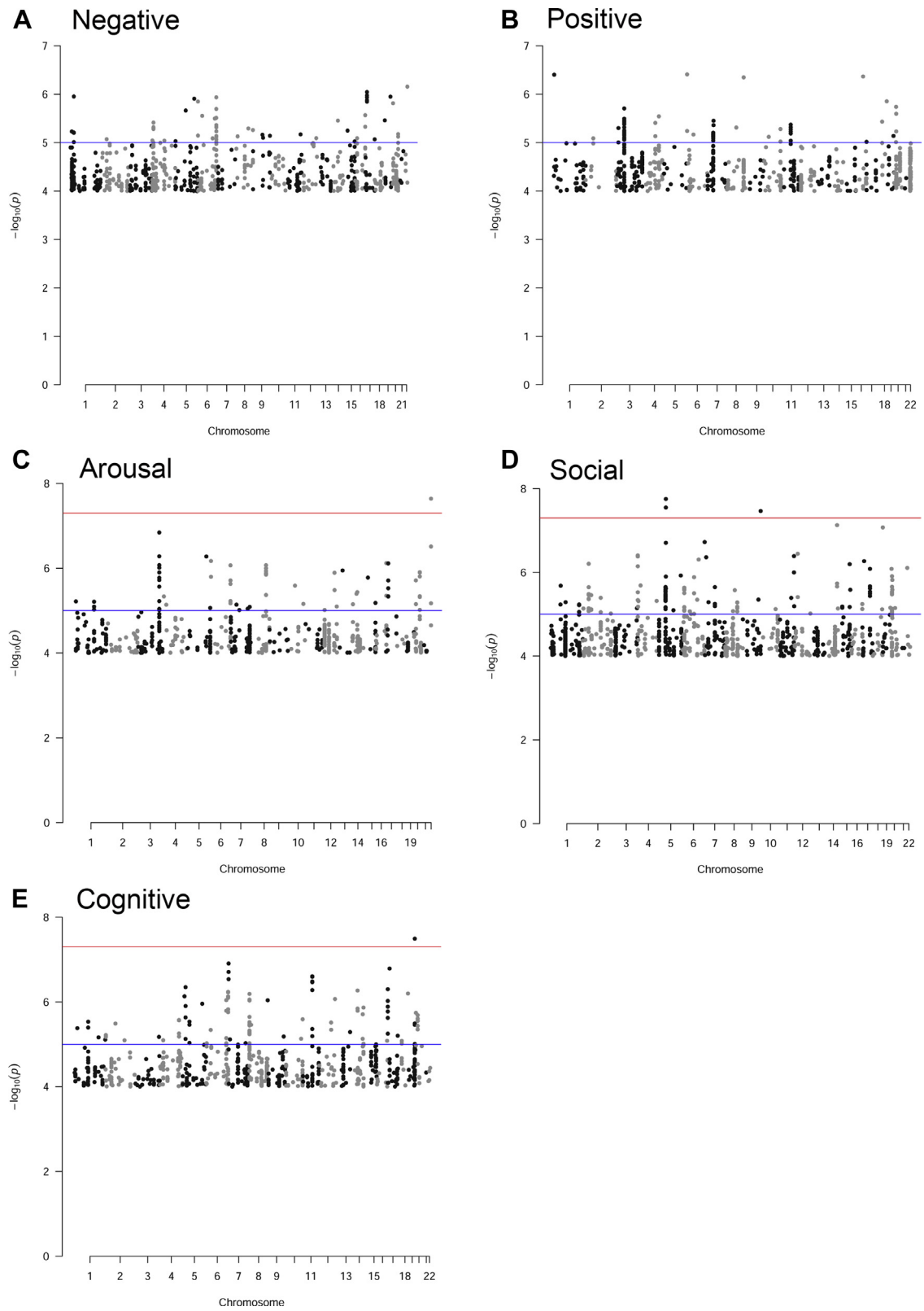
We examined 4687 individuals of Northern European ancestry across the three batches (wave 1, 1589; wave 2, 1547; wave 3, 1551), with meta-analysis of 893,900 SNPs with minor allele frequency of 0.01 or greater. The cohorts included 2363 females (50.4%), and the mean age was 64.3 years (SD, 14.9 years). Figure 1 shows Manhattan plots for each of the five dimensional phenotypes (Q-Q plots are shown in Supplemental Figure S1).

For each of the dimensions, the 10 independent loci with strongest evidence of association are described in Table 1. Overall, one locus was associated with arousal, two with social, and one with cognition at a standard genome-wide significance threshold ( $p < 5 \times 10^{-8}$ ); these four regions are depicted in Figure 2. Notably, for arousal, the associated locus spans *RFPL3* and *RFPL3S*; this family of proteins has been suggested to be important in primate neocortical evolution (25). For cognition, the associated locus spans *FPR3*, a chemoattractant (15623572) that has been suggested to be relevant in immune response in Alzheimer's disease (26).

## DISCUSSION

In this analysis of 4687 individuals drawn from a biobank spanning two academic medical centers, we identified four loci associated with dimensional psychopathology at a standard genome-wide threshold based on natural language processing of narrative hospital discharge notes. Two of these span genes are associated with neurodevelopment (*RFPL3*) or neurodegeneration (*PFR3*). While both are known to be brain expressed, neither has previously been strongly associated with neuropsychiatric disease, suggesting the potential utility of the approach we describe in understanding brain function in a manner that is unbiased by traditional nosology.

While not achieving a genome-wide threshold for significance, we also note the observed association between the calcium channel subunit *CACNA2D3* and positive valence.



**Figure 1.** Manhattan plots from genome-wide association studies for each of the five dimensions of psychopathology.

**Table 1. Independent Loci With Strongest Evidence of Association for Each Dimension of Psychopathology**

CHR	SNP	$p$ Value	SNPs, $n^a$	Locus Span	Locus Size, kb	Genes in Locus	A1	A2	MAF
Negative									
22	22:32750463	$7.00 \times 10^{-7}$	8	chr22:32738156..32800705	62.55	LOC339666, <i>RFPL3</i> , <i>RFPL3S</i> , <i>RTCB</i>	A	C	0.013
17	17:14495883	$9.03 \times 10^{-7}$	12	chr17:14382314..14497857	115.544		A	C	0.012
1	1:27963259	$1.12 \times 10^{-6}$	23	chr1:27963259..27981314	18.056		T	A	0.011
19	19:51371390	$1.13 \times 10^{-6}$	1	chr19:51371390..51371390	0.001		T	C	0.033
6	6:155563548	$1.16 \times 10^{-6}$	18	chr6:155562206..155733894	171.689	<i>CLDN20</i> , <i>NOX3</i> , <i>TFB1M</i> , <i>TIAM2</i>	G	A	0.014
5	5:155373160	$1.24 \times 10^{-6}$	4	chr5:155221281..155397446	176.166		T	C	0.015
6	6:5821150	$1.42 \times 10^{-6}$	3	chr6:5815134..5851189	36.056		G	T	0.014
20	20:15042429	$1.54 \times 10^{-6}$	1	chr20:15042429..15042429	0.001	<i>MACROD2</i>	C	T	0.013
5	5:86039654	$2.18 \times 10^{-6}$	4	chr5:86037590..86040114	2.525		G	T	0.010
16	16:83664928	$2.71 \times 10^{-6}$	38	chr16:83656037..83753512	97.476	<i>CDH13</i>	A	T	0.084
Positive									
6	6:5821150	$3.91 \times 10^{-7}$	3	chr6:5815134..5851189	36.056		G	T	0.014
1	1:6039258	$3.97 \times 10^{-7}$	3	chr1:5953811..6039453	85.643	<i>NPHP4</i>	T	C	0.012
16	16:56251428	$4.32 \times 10^{-7}$	2	chr16:56095547..56251428	155.882	<i>DKFZP434H168</i> , <i>GNAO1</i> , <i>LOC283856</i>	G	A	0.013
8	8:132532229	$4.51 \times 10^{-7}$	121	chr8:132404887..132532936	128.05		A	G	0.157
18	18:77374268	$1.40 \times 10^{-6}$	1	chr18:77374268..77374268	0.001		G	C	0.011
20	20:15696084	$1.83 \times 10^{-6}$	3	chr20:15690854..15696084	5.231	<i>MACROD2</i>	C	A	0.093
3	3:54508115	$1.98 \times 10^{-6}$	71	chr3:54488508..54575770	87.263	<i>CACNA2D3</i>	T	C	0.338
20	20:16560345	$2.54 \times 10^{-6}$	51	chr20:16509803..16605627	95.825	<i>KIF16B</i>	C	T	0.167
4	4:127370341	$2.88 \times 10^{-6}$	102	chr4:127360862..127402924	42.063		T	C	0.100
7	7:47324136	$3.55 \times 10^{-6}$	3	chr7:47324136..47328060	3.925	<i>TNS3</i>	T	C	0.081
Arousal									
22	22:32750463	$2.29 \times 10^{-8}$	8	chr22:32738156..32800705	62.55	LOC339666, <i>RFPL3</i> , <i>RFPL3S</i> , <i>RTCB</i>	A	C	0.013
3	3:167741670	$1.44 \times 10^{-7}$	81	chr3:167544555..167741670	197.116	<i>GOLIM4</i> , <i>LOC646168</i>	A	G	0.016
5	5:150327474	$5.28 \times 10^{-7}$	4	chr5:150115979..150327474	211.496	<i>DCTN4</i> , <i>IRGM</i> , <i>SMIM3</i> , <i>ZNF300</i> , <i>ZNF300P1</i>	T	G	0.057
6	6:5821150	$6.75 \times 10^{-7}$	2	chr6:5821150..5851189	30.04		G	T	0.014
16	16:83664928	$7.59 \times 10^{-7}$	49	chr16:83656037..83753512	97.476	<i>CDH13</i>	A	T	0.084
17	17:14496077	$7.69 \times 10^{-7}$	11	chr17:14495883..14497857	1.975		C	T	0.012
8	8:118469770	$8.49 \times 10^{-7}$	86	chr8:118379461..118588575	209.115	<i>MED30</i>	C	T	0.371
6	6:155563548	$8.55 \times 10^{-7}$	20	chr6:155562206..155733894	171.689	<i>CLDN20</i> , <i>NOX3</i> , <i>TFB1M</i> , <i>TIAM2</i>	G	A	0.014
13	13:43496853	$1.13 \times 10^{-6}$	1	chr13:43496853..43496853	0.001	<i>EPSTI1</i>	C	T	0.014
20	20:45375674	$1.24 \times 10^{-6}$	17	chr20:45315786..45385268	69.483	<i>SLC2A10</i> , <i>TP53RK</i>	T	A	0.063
Social									
5	5:52629643	$1.77 \times 10^{-8}$	22	chr5:52564100..52661007	96.908		A	G	0.012
9	9:137405964	$3.42 \times 10^{-8}$	3	chr9:137341500..137405964	64.465		T	C	0.021
14	14:97095154	$7.45 \times 10^{-8}$	4	chr14:97087772..97117785	30.014		A	G	0.043
18	18:77374268	$8.48 \times 10^{-8}$	5	chr18:77365764..77396240	30.477		G	C	0.011
7	7:2472517	$1.89 \times 10^{-7}$	2	chr7:2230076..2472517	242.442	<i>CHST12</i> , <i>EIF3B</i> , <i>FTSJ2</i> , <i>MAD1L1</i> , <i>MIR6836</i> , <i>NUDT1</i> , <i>SNX8</i>	T	C	0.023
12	12:27167220	$3.60 \times 10^{-7}$	8	chr12:27145587..27333632	188.046	<i>C12orf71</i> , <i>MED21</i> , <i>TM7SF3</i>	A	G	0.014
4	4:2483900	$3.94 \times 10^{-7}$	7	chr4:2483900..2732557	248.658	<i>FAM193A</i> , <i>RNF4</i>	A	C	0.013
11	11:125064877	$4.10 \times 10^{-7}$	65	chr11:125052718..125110079	57.362	<i>PKNOX2</i>	T	A	0.127
7	7:14309510	$4.37 \times 10^{-7}$	3	chr7:14294006..14309510	15.505	<i>DGKB</i>	T	C	0.014
6	6:125802803	$4.93 \times 10^{-7}$	1	chr6:125802803..125802803	0.001		A	G	0.013
Cognitive									
19	19:52351965	$3.22 \times 10^{-8}$	94	chr19:52306547..52377699	71.153	<i>FPR3</i> , <i>ZNF577</i>	T	C	0.321
7	7:3627391	$1.24 \times 10^{-7}$	3	chr7:3610381..3662960	52.58	<i>SDK1</i>	G	A	0.019
17	17:13683929	$1.63 \times 10^{-7}$	3	chr17:13680505..13806459	125.955		A	G	0.014
11	11:73586112	$2.49 \times 10^{-7}$	211	chr11:73340835..73672187	331.353	<i>COA4</i> , <i>DNAJB13</i> , <i>MRPL48</i> , <i>PAAF1</i> , <i>PLEKHB1</i> , <i>RAB6A</i>	G	C	0.303

**Table 1. Continued**

CHR	SNP	<i>p</i> Value	SNPs, <i>n</i> <sup>a</sup>	Locus Span	Locus Size, kb	Genes in Locus	A1	A2	MAF
5	5:22528391	4.49 <sup>e-7</sup>	5	chr5:22365713..22706775	341.063	<i>CDH12</i>	T	G	0.018
17	17:167312	5.02 <sup>e-7</sup>	11	chr17:149460..172591	23.132	<i>RPH3AL</i>	T	C	0.023
14	14:57183567	5.40 <sup>e-7</sup>	6	chr14:57182182..57194970	12.789		C	A	0.018
6	6:169616423	5.79 <sup>e-7</sup>	31	chr6:169596595..169622263	25.669	<i>THBS2</i>	T	C	0.257
18	18:77374268	6.33 <sup>e-7</sup>	1	chr18:77374268..77374268	0.001		G	C	0.011
8	8:6034653	6.48 <sup>e-7</sup>	43	chr8:6021491..6061234	39.744		A	G	0.363

A1, allele 1; A2, allele 2; CHR, chromosome; MAF, minor allele frequency; SNP, single nucleotide polymorphism.

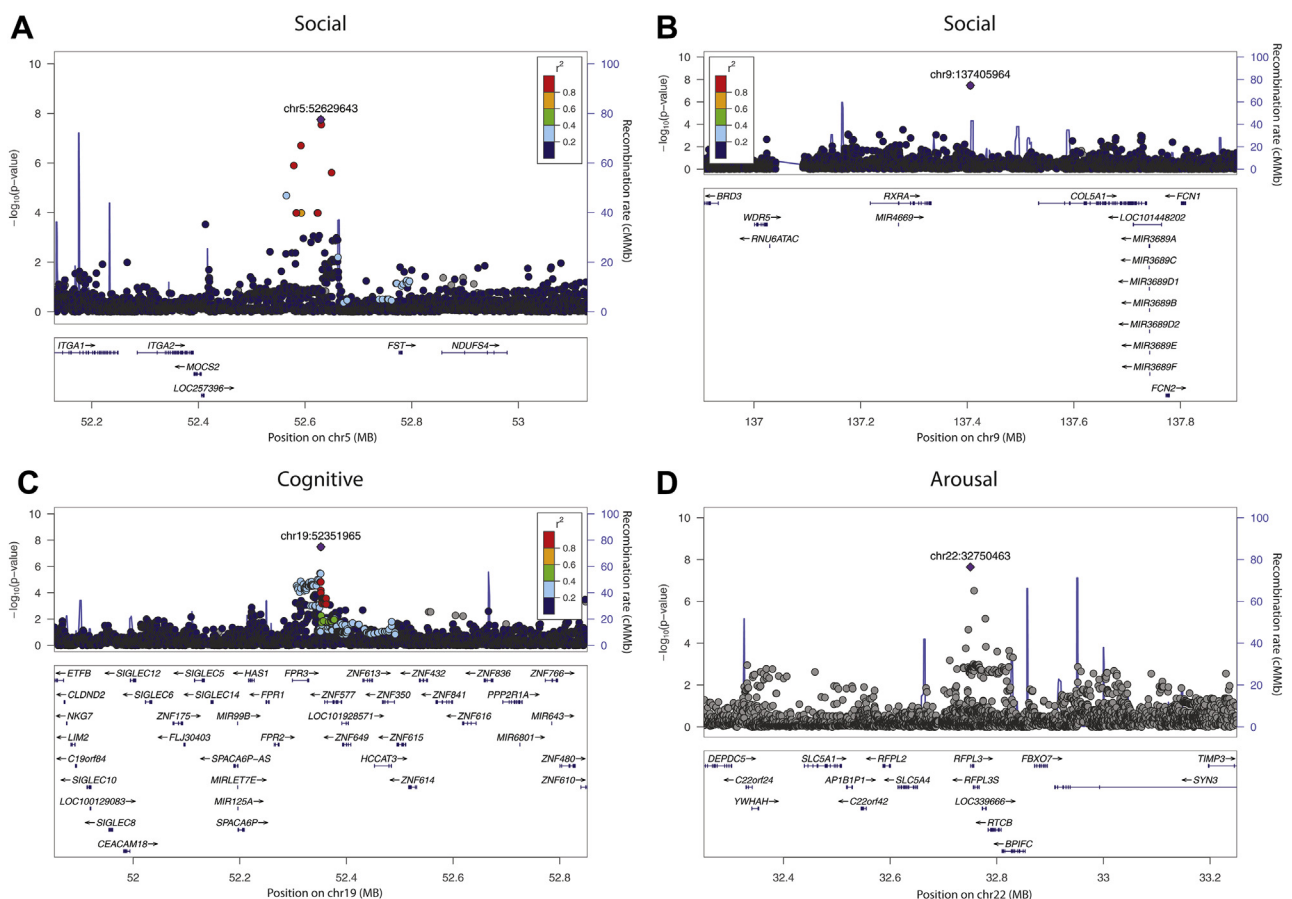
<sup>a</sup>Number of SNPs in the linkage disequilibrium block with nominal *p* < .01. See text for details.

This locus has previously been associated with pain sensitivity, which may impact reward responsiveness, suggesting convergent validity (i.e., assay sensitivity) (27). This family of subunits represents the target for multiple anticonvulsants used to treat neuropathic pain and has recently been shown to regulate accumulation of voltage-gated calcium channels and exocytosis at the synapse (28).

While these loci are promising as candidates for follow-up study, multiple limitations in this proof-of-concept study should be considered. First, while we exceed a standard threshold for genome-wide studies, replication will increase confidence in these results. (At a more stringent experiment-wide threshold, based upon correlation between these

domains, one could also argue that a threshold of  $2 \times 10^{-8}$  would be appropriate.) We elected to meta-analyze all data available to us, rather than holding out a replication set, and present these results in the hope that they will encourage other hospital-linked biobanks to consider our approach. Second, as with any common variant study, none of these variants can be considered causal, and biological studies will be required to characterize their effect.

More broadly, it is entirely possible—indeed, likely—that other dimensional features or extraction methods, as well as incorporation of other data types, would lead to identification of other loci. We adopted a new method for identifying dimensional psychopathology from narrative clinical notes



**Figure 2.** Region plots for four loci with genome-wide significance. chr, chromosome; cM/Mb, recombination rate.

based on seed terms extracted from RDoC workgroup statements, which we have recently described in more detail along with initial validation (7). These scores do not yet address subdomains; sensitivity likely varies by domain, and indeed, as with RDoC itself, the presence of terms loading on a given domain does not necessarily represent psychopathology and may instead capture normal or subsyndromal variation. We note that the present study represents an example of transfer learning: a model trained in one type of cohort (psychiatric hospitalizations) is applied to distinguish features of another (all-cause hospitalizations), but further investigations of portability will be important. In particular, this approach complements rather than replaces analysis of more traditional curated phenotypes (29,30). Beyond investigating other strategies for concept extraction, it will be valuable to understand the extent to which incorporating other types of notes or integrating these data with coded clinical data improve the identification of dimensions of psychopathology [for further discussion of general methodologic considerations, please see (7)].

With these caveats in mind, our results suggest an approach to identifying genes associated with psychopathology beyond traditional diagnostic categories, and they demonstrate the feasibility and potential utility of this broad class of approaches, aiming to be both transparent and portable. Narrative clinical notes may contain a wealth of clinical detail relevant to developing dimensional representations of brain diseases. With increasing availability of biobanks and registries as a resource for genomic discovery and translation, natural language processing represents a way to amplify their utility for investigating complex phenotypes that avoids the constraint of traditional psychiatric nosology.

### ACKNOWLEDGMENTS AND DISCLOSURES

This work was supported by National Human Genome Research Institute (NHGRI) Grant No. 1P50MH106933-04 and National Institute of Mental Health (NIMH) Grant No. 1R01MH106577-01A1 (to RHP) and the Broad Institute Stanley Center Fellowship and Brain and Behavior Foundation Grant No. 26489 (to THM). The sponsors had no role in study design, writing of the report, or data collection, analysis, or interpretation. The corresponding and senior authors had full access to all data and made the decision to submit for publication.

We thank the participants and administrators of the Partners HealthCare Biobank for their contribution to this work.

RHP serves on the scientific advisory board for Perfect Health, Genomind, and Psy Therapeutics and is a consultant for RID Ventures. The other authors report no biomedical financial interests or potential conflicts of interest.

### ARTICLE INFORMATION

From the Center for Quantitative Health and Department of Psychiatry (THM, VMC, KLH, AMP, RHP), Massachusetts General Hospital and Harvard Medical School, and the Harvard School of Public Health (TC), Boston, Massachusetts; and Tsinghua University (SY), Beijing, China.

Address correspondence to Thomas H. McCoy Jr., M.D., Massachusetts General Hospital, Simches Research Building, 6th floor, Boston, MA 02114; E-mail: [thmccoy@partners.org](mailto:thmccoy@partners.org).

Received Aug 10, 2017; revised Nov 17, 2017; accepted Dec 7, 2017.

Supplementary material cited in this article is available online at <https://doi.org/10.1016/j.biopsych.2017.12.004>.

### REFERENCES

1. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, *et al.* (2015): An atlas of genetic correlations across human diseases and traits. *Nat Genet* 47:1236–1241.
2. Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, *et al.* (2013): Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* 45:984–994.
3. Gilman SE, Ni MY, Dunn EC, Breslau J, McLaughlin KA, Smoller JW, *et al.* (2015): Contributions of the social environment to first-onset and recurrent mania. *Mol Psychiatry* 20:329–336.
4. Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, *et al.* (2010): Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *Am J Psychiatry* 167:748–751.
5. Sanislow CA, Pine DS, Quinn KJ, Kozak MJ, Garvey MA, Heinssen RK, *et al.* (2010): Developing constructs for psychopathology research: Research Domain Criteria. *J Abnorm Psychol* 119:631–639.
6. McCoy TH, Castro VM, Rosenfield HR, Cagan A, Kohane IS, Perlis RH (2015): A clinical perspective on the relevance of research domain criteria in electronic health records. *Am J Psychiatry* 172:316–320.
7. McCoy TH Jr, Yu S, Hart KL, Castro VM, Brown HE, Rosenquist JN, *et al.* (2018): High throughput phenotyping for dimensional psychopathology in electronic health records. *Biol Psychiatry* 83:997–1004.
8. Gainer VS, Cagan A, Castro VM, Duey S, Ghosh B, Goodson AP, *et al.* (2016): The Biobank Portal for Partners Personalized Medicine: A query tool for working with consented biobank samples, genotypes, and phenotypes using i2b2. *J Pers Med* 6:11.
9. Murphy SN, Mendis M, Hackett K, Kuttan R, Pan W, Phillips LC, *et al.* (2007): Architecture of the open-source clinical research chart from Informatics for Integrating Biology and the Bedside. *AMIA Annu Symp Proc* 548–552.
10. Murphy S, Churchill S, Bry L, Chueh H, Weiss S, Lazarus R, *et al.* (2009): Instrumenting the health care enterprise for discovery research in the genomic era. *Genome Res* 19:1675–1681.
11. Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, *et al.* (2010): Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc* 17:124–130.
12. National Institute of Mental Health (2017): RDoC matrix. Available at: <https://www.nimh.nih.gov/research-priorities/rdoc/constructs/rdoc-matrix.shtml>. Accessed November 17, 2017.
13. Wang C, Ward ME, Chen R, Liu K, Tracy TE, Chen X, *et al.* (2017): Scalable production of iPSC-derived human neurons to identify tau-lowering compounds by high-content screening. *Stem Cell Rep* 9:1221–1233.
14. Henn BM, Hon L, Macpherson JM, Eriksson N, Saxonov S, Pe'er I, *et al.* (2012): Cryptic distant relatives are common in both isolated and cosmopolitan genetic samples. *PLoS One* 7:e34267.
15. Fuchsberger C, Abecasis GR, Hinds DA (2015): minimac2: Faster genotype imputation. *Bioinformatics* 31:782–784.
16. Sala M, Lazzaretti M, De Vidovich G, Caverzasi E, Barale F, d'Allio G, *et al.* (2009): Electrophysiological changes of cardiac function during antidepressant treatment. *Ther Adv Cardiovasc Dis* 3:29–43.
17. Wenger TL, Cohn JB, Bustrack J (1983): Comparison of the effects of bupropion and amitriptyline on cardiac conduction in depressed patients. *J Clin Psychiatry* 44:174–175.
18. Delaneau O, Marchini J, Zagury J-F (2012): A linear complexity phasing method for thousands of genomes. *Nat Methods* 9:179–181.
19. Purcell S, Chang C (2013): PLINK 1.90 beta. Available at: <https://www.cog-genomics.org/plink/1.9/>. Accessed November 17, 2017.
20. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006): Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38:904–909.

21. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015): Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* 4:7.
22. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, *et al.* (2007): PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575.
23. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Glied TP, *et al.* (2010): LocusZoom: Regional visualization of genome-wide association scan results. *Bioinformatics* 26:2336–2337.
24. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Patterson N, *et al.* (2015): LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 47:291–295.
25. Bonnefont J, Nikolaev SI, Perrier AL, Guo S, Cartier L, Sorce S, *et al.* (2008): Evolutionary forces shape the human RFPL1,2,3 genes toward a role in neocortex development. *Am J Hum Genet* 83:208–218.
26. Iribarren P, Zhou Y, Hu J, Le Y, Wang JM (2005): Role of formyl peptide receptor-like 1 (FPRL1/FPR2) in mononuclear phagocyte responses in Alzheimer disease. *Immunol Res* 31:165–176.
27. Neely GG, Hess A, Costigan M, Keene AC, Goulas S, Langeslag M, *et al.* (2010): A genome-wide *Drosophila* screen for heat nociception identifies alpha2delta3 as an evolutionarily conserved pain gene. *Cell* 143:628–638.
28. Hoppa MB, Lana B, Margas W, Dolphin AC, Ryan TA (2012): alpha2-delta expression sets presynaptic calcium channel abundance and release probability. *Nature* 486:122–125.
29. O’Dushlaine C, Ripke S, Ruderfer DM, Hamilton SP, Fava M, Iosifescu DV, *et al.* (2014): Rare copy number variation in treatment-resistant major depressive disorder. *Biol Psychiatry* 76:536–541.
30. Castro VM, Minnier J, Murphy SN, Kohane I, Churchill SE, Gainer V, *et al.* (2015): Validation of electronic health record phenotyping of bipolar disorder cases and controls. *Am J Psychiatry* 172:363–372.