

# Genome-Wide Association Study of Suicide Death and Polygenic Prediction of Clinical Antecedents

Anna R. Docherty, Ph.D., Andrey A. Shabalina, Ph.D., Emily DiBlasi, Ph.D., Eric Monson, M.D., Niamh Mullins, Ph.D., Daniel E. Adkins, Ph.D., Silviu-Alin Bacanu, Ph.D., Amanda V. Bakian, Ph.D., Sheila Crowell, Ph.D., Danli Chen, Ph.D., Todd M. Darlington, Ph.D., William B. Callor, M.S., Erik D. Christensen, M.D., Douglas Gray, M.D., Brooks Keeshin, M.D., Michael Klein, M.S., John S. Anderson, B.S., Leslie Jerominski, M.S., Caroline Hayward, Ph.D., David J. Porteous, Ph.D., Andrew McIntosh, M.D., Qingqin Li, Ph.D., Hilary Coon, Ph.D.

**Objective:** Death by suicide is a highly preventable yet growing worldwide health crisis. To date, there has been a lack of adequately powered genomic studies of suicide, with no sizable suicide death cohorts available for analysis. To address this limitation, the authors conducted the first comprehensive genomic analysis of suicide death using previously unpublished genotype data from a large population-ascertained cohort.

**Methods:** The analysis sample comprised 3,413 population-ascertained case subjects of European ancestry and 14,810 ancestrally matched control subjects. Analytical methods included principal component analysis for ancestral matching and adjusting for population stratification, linear mixed model genome-wide association testing (conditional on genetic-relatedness matrix), gene and gene set-enrichment testing, and polygenic score analyses, as well as single-nucleotide polymorphism (SNP) heritability and genetic correlation estimation using linkage disequilibrium score regression.

**Results:** Genome-wide association analysis identified two genome-wide significant loci (involving six SNPs: rs34399104,

rs35518298, rs34053895, rs66828456, rs35502061, and rs35256367). Gene-based analyses implicated 22 genes on chromosomes 13, 15, 16, 17, and 19 ( $q < 0.05$ ). Suicide death heritability was estimated at an  $h^2_{\text{SNP}}$  value of 0.25 (SE=0.04) and a value of 0.16 (SE=0.02) when converted to a liability scale. Notably, suicide polygenic scores were significantly predictive across training and test sets. Polygenic scores for several other psychiatric disorders and psychological traits were also predictive, particularly scores for behavioral disinhibition and major depressive disorder.

**Conclusions:** Multiple genome-wide significant loci and genes were identified and polygenic score prediction of suicide death case-control status was demonstrated, adjusting for ancestry, in independent training and test sets. Additionally, the suicide death sample was found to have increased genetic risk for behavioral disinhibition, major depressive disorder, depressive symptoms, autism spectrum disorder, psychosis, and alcohol use disorder compared with the control sample.

*Am J Psychiatry* 2020; 177:917–927; doi: 10.1176/appi.ajp.2020.19101025

Death by suicide is a behavioral event that reflects a complex, heritable phenotype with diverse clinical antecedents and environmental risk factors (1). The estimated annual age-adjusted prevalence rate of suicide death is approximately 0.014% in the United States (2), and this rate has been steadily increasing since 2000 (3). Suicide is now ranked the second leading cause of death among all persons 15–24 years old in the United States (4). As such, suicide is considered a major public health challenge (5), which has spurred research into its etiology and clinical prediction (6, 7).

While it is established that suicide is significantly heritable (8), well-powered genomic research on the topic has been largely limited to the study of suicide-related behaviors rather

than the ultimate phenotype of suicide death. And although results from well-powered genome-wide association studies (GWASs) of suicidal behavior have advanced our understanding of the process (9), the vast majority of suicidal behavior does not result in suicide death (1, 10). Thus, suicidal behavior represents a less severe and likely more heterogeneous phenotype than suicide death, characteristics that may adversely affect statistical power to detect genetic associations. Conversely, the unambiguous phenotype of suicide death avoids several confounders inherent in the study of suicidal behavior or ideation and also focuses study on one of the single most critical contemporary public health outcomes. Previous genomic studies of suicidal behavior have also been limited by

See related feature: **Editorial** by Dr. Goldman (p. 881)

narrow ascertainment, examining only individuals with specific diagnoses (e.g., mood disorders, psychotic disorders) in order to maximize severity and accommodate post hoc study designs. Here, we avoided this limitation, instead using population-based sampling with ascertainment wholly independent of any co-occurring diagnoses, thus improving the representativeness of cases to the corresponding population and, consequently, the generalizability of results.

Suicide death, like its associated clinical antecedents, including schizophrenia and depression, has a highly multifactorial, and likely highly polygenic, etiology (9, 11). Currently, the scientific literature lacks well-powered studies of suicide death in relation to molecular genetic risk for any medical or psychiatric diagnoses, and no robust polygenic scores have yet been developed for the critical outcome of suicide death. There are, however, a few modestly powered GWASs of suicide death in the literature (12, 13), the largest of which combined two independent Japanese cohorts totaling 746 deaths by suicide (13). Additionally, a recent UK Biobank GWAS of suicidal behavior conducted preliminary polygenic analyses of suicide death using a very modest number of deaths by suicide ( $N=127$ ) (14). Of note, these smaller studies reported evidence of polygenicity in both suicidal behavior and suicide death but with widely varying single-nucleotide polymorphism (SNP)-based heritability estimates, from an  $h^2_{\text{SNP}}$  of 4.6 (15) and 7.6% (14) for suicidal behavior not including death to an  $h^2_{\text{SNP}}$  of 35%–48% for suicide death (13). The present study marks a notable advance from earlier research, because we used the world's largest DNA data bank of suicide death, merged with a massive bank of electronic medical record and sociodemographic data (16, 17), to model common variant genetic, and clinical phenotypic, precursors of suicide death.

This study adds to the growing body of genomic research on suicide (13, 14), providing the first adequately powered GWAS of suicide death. In our analyses, we integrated data on modes of suicide death, medical and psychiatric diagnostic (ICD-10) codes (18), and medical and psychiatric polygenic scores to comprehensively model common variant genetic risk for suicide death. Secondary analyses of sex differences were conducted as a result of the substantial sex differences in suicide death rates and modes of suicide death (19, 20). Reliable and significant prediction of case-control status was achieved, adjusting for ancestry, using both a novel polygenic score for suicide death and polygenic scores for a range of comorbid psychiatric and medical risk factors, particularly behavioral disinhibition, major depressive disorder, and depressive symptoms. Additionally, gene set enrichment, SNP heritability, and genetic correlations were examined in this unique, unpublished data resource. We conclude with a brief discussion of the relevance of our findings to the broader field and promising future research directions.

## METHODS

### Sample Ascertainment

*Case samples.* In collaboration with the centralized statewide Utah Office of the Medical Examiner (OME), we obtained

DNA samples from approximately 6,000 persons who died by suicide. The centralized OME and conservative determination helped to maximize the accuracy of suicide case status (21). Suicide cause-of-death determination resulted from detailed investigation of the scene of the death and circumstances of death, determination of medical conditions by full autopsy, review of medical and other public records concerning the case, interviews with survivors, and standard toxicology workups. Suicide determination is traditionally made quite conservatively because of its effect on surviving relatives.

DNA from individuals who died by suicide was extracted from whole blood using the Qiagen Autopure Cell Lysis Solution automated DNA extractor ([www.qiagen.com](http://www.qiagen.com)). Genotyping is described below. Data collection and genotyping for the Utah Suicide Study is ongoing. To date, 4,381 samples have been genotyped in two waves (waves 1 and 2) as determined by the date of sample receipt and the availability of genotyping funds (for further details on the sample collection, see the Methods section in the online supplement). After quality-control procedures and ancestry analysis were performed, the data comprised 3,413 samples of individuals who died by suicide in Utah. The Utah population is primarily Northwestern European in ancestry, a relatively genetically homogeneous group with low inbreeding across generations, comparable to the rest of the United States (22).

*Control samples.* In the Generation Scotland database, DNA from control subjects closely matching the ancestry of the case subjects was obtained from previously curated data sets in the United Kingdom. The wave 1 analysis included 3,623 founder control subjects from the population-based Generation Scotland Scottish Family Health Study (23). The Generation Scotland Scottish Family Health Study ( $N>24,000$ ) comprised an ancestrally comparable population-based cohort for comparison with the suicide decedents in Utah. To eliminate confounding arising from intra-data-set relatedness, only the 3,623 founders from the Generation Scotland data set were used in analyses.

A total of 11,049 control samples from the UK10K Rare Genetic Variants in Health and Disease Project (24) were analyzed in wave 2 and in GWAS analyses of both waves. This second control cohort comprised approximately 4,000 genomes along with 6,000 exomes from individuals in the United Kingdom with selected health phenotypes. We chose these data as a result of extensive phenotyping and characterization of any medical conditions present and to avoid choosing a cohort of entirely psychiatrically and medically healthy individuals. A total of 4,000 highly phenotyped “super control” samples were supplied from the TwinsUK Registry from King's College London and the Avon Longitudinal Study of Parents and Children. The UK10K Project was a collaborative project to examine obesity, autism, schizophrenia, familial hypercholesterolemia, thyroid disorders, learning disabilities, ciliopathies, congenital heart disease, coloboma, neuromuscular disorders, and rare disorders, including severe insulin resistance. Analyses of this control cohort included all

diagnostic groups. While this inclusion may have been conservative because of possible associations of chronic disease with suicide risk, determination of effects of specific disease states was beyond the scope of this study. Genotyping and sequencing procedures for the UK10K Project (<http://www.uk10k.org>) have been described elsewhere (24), and all molecular genetic data from the project were filtered to the hard-call variants present in our suicide death cohort before imputation of all cohorts simultaneously.

*1,000 Genomes Reference Panel.* The CEU population from the 1,000 Genomes Project (25), which includes only Utah residents carefully screened for Northwestern European ancestry, was utilized as a model for excluding ancestrally discordant suicide and control samples. These CEU data were downloaded from the 1,000 Genomes Project public repository. Data from unrelated individuals in the CEU provided a compelling, albeit small, ancestrally matched control resource (N=99). A variety of candidate control samples were assessed by principal component analysis for ancestral comparability to CEU and decedent data, with the UK10K Project and Generation Scotland founder data representing the closest match.

Control samples from the Utah cohort would have been an ideal match for the suicide case samples from this cohort, but as with many GWASs, local control samples were not readily available at the sample size required for GWAS. Samples from the Centre d'Etude du Polymorphisme Humain (CEPH) ancestry 1,000 Genomes were a useful comparison group to assess the likelihood that control samples from the United Kingdom were an appropriate match for our case samples. In addition (and described in more detail below), we performed a Generation Scotland control-UK10K Project control GWAS and subsequently eliminated any SNPs from the case-control analysis that evidenced signal between the control cohorts. This was performed to minimize the possibility of false positives in the case-control GWAS as a result of population and geographic stratification across cohorts.

### Genotyping and Quality Control

Suicide case subjects were genotyped using the Illumina Infinium PsychArray platform, measuring 593,260 SNPs. Generation Scotland samples were genotyped using the Illumina OmniExpress SNP GWAS and exome chip, measuring 700,000 and 250,000 SNPs, respectively (23). UK10K Project samples were whole-genome sequenced (24), and variants were extracted to match the available quality-controlled hard-called variants in the suicide case subjects. Genotypes were subsequently imputed in all case and control subjects. Both case and control data sets resulted from population-based ascertainment. Cryptic relatedness was modeled via the derivation of genomic-relatedness matrices. Genotyping quality control was performed using SNP clustering with Illumina GenomeStudio software. SNPs were retained if the GenTrain score was >0.5 and the cluster separation score was >0.4. SNPs were converted to an Hg19 plus strand. SNPs with >5%

missing genotypes were removed. Additionally, samples with a call rate <95% were removed.

Before case-control GWAS, a control-control GWAS was conducted (using the same methods described below) in order to detect differences between control groups (all variants, both pre- and postimputation, with GWAS control-control q values <0.10). For example, chromosome 4 toll-like receptor variants are often filtered from analyses involving Scottish control samples as a result of population stratification. A total of 167 variants in the control-control comparison were then filtered from subsequent case-control analyses. For the purposes of future meta-GWAS analyses and because the major histocompatibility complex is relevant to psychiatric risk, we included major histocompatibility complex-associated filtered variants in a second version of summary statistics.

### Data Analysis

*Principal component analysis.* The 1,000 Genomes super-populations and suicide case and control samples, both included and excluded, are presented in Figure S1 in the online supplement, plotted by the top two principal components. Approximately 20% of the population-based suicide case samples had a significant degree of non-Northwestern European ancestry (chiefly of admixed ancestry) and were excluded from analyses. The variation explained by the top four principal components was reduced 7.2-fold. The top four principal components explained 0.89% of variation before sample filtering and 0.12% of variation after filtering if calculated on pruned genotypes. For adequate statistical power, we examined only case samples from individuals of Northern European ancestry. However, the cohort comprised multiple ancestries, and research on suicide death in samples from individuals of non-European ancestries will reflect an important step beyond this first study.

Principal component analysis was performed in control, suicide, and 1,000 Genomes cohorts after linkage disequilibrium pruning at a threshold of 0.2. To maximize ancestral homogeneity, the top four principal components (defined as those components that accounted for >0.1% of the genotype variance) were used to establish principal component centroid limits centered around 1,000 Genomes CEU data such that 99% of the CEU data fell within the limits. Only suicide and control samples also falling within these limits were considered ancestrally homogeneous and were included in the association study. The ancestry principal component analysis was performed using RaMWAS, a Bioconductor package written by our analytical team, which comprises a complete tool set for high-dimensional genomic analyses (26, 27).

*Imputation.* Case and control subjects were well matched to 1,000 Genomes CEPH. The Haplotype Reference Consortium comprises in part control subjects from the United Kingdom used in this GWAS, and thus we elected to impute genotypes on the basis of the 1,000 Genomes reference panel

**TABLE 1. Genome-wide significant loci from a genome-wide association study of death by suicide<sup>a</sup>**

Chromosome	SNP	Imputation R <sup>2</sup>	Allele 1	Allele 2	Allele 1 Frequency, Case Subjects	Allele 1 Frequency, Control Subjects
13	rs34399104	0.75	T	C	0.027	0.015
13	rs35518298	0.81	T	C	0.024	0.014
13	rs34053895	0.81	A	C	0.024	0.014
13	rs35502061	0.72	G	A	0.024	0.014
13	rs66828456	0.74	A	C	0.027	0.015
15	rs35256367	0.50	G	A	0.023	0.014

<sup>a</sup> CADD=combined annotation-dependent depletion (61, 62); SNP=single-nucleotide polymorphism. For scale, a CADD score of 20 indicates that a variant is among the top 1% of deleterious variants in the human genome. All variants presented were imputed.

using Minimac3 (28) and Eagle (29). SNPs were imputed jointly from a common SNP list based on variant overlap across case and control data sets. SNPs with ambiguous strand orientation, >5% missing calls, or a Hardy-Weinberg equilibrium p value <0.001 were excluded. SNPs with a minor allele frequency <0.01 or an imputation R<sup>2</sup> <0.5 or average imputation call rate <0.9 were excluded after imputation. Genomic data were handled using PLINK (30, 31). Final GWAS analysis was performed on 7,519,308 variants that passed quality control.

**Genome-wide association testing.** A linear mixed model (LMM) algorithm tested variant association with suicide death and provided follow-up examination of significant hits for linkage disequilibrium and gene set enrichment. GWASs were performed using genome-wide efficient mixed-model analysis (32), a computationally efficient and open-source LMM algorithm for GWASs that models population stratification remaining after principal component analysis by use of genomic-relatedness matrices. Sex was not included as a covariate in GWAS analyses because of the association of suicide with sex status at a male:female ratio of approximately 4:1. GWASs with hard-call only and then with imputed data were examined separately to assess potential population stratification unique to our imputed GWAS. Before case-control GWAS, control-control GWAS was implemented to filter signal that was likely a result of population stratification in the control samples. To assess the likelihood that observed results may be due to chance, we assessed the false discovery rate at 5%.

**Gene and gene set enrichment and functional mapping.** SNP results from the GWAS were then mapped to genes within 1 kb of the SNP. These genes were examined for gene set enrichment and linkage disequilibrium using functional mapping and annotation of genetic associations (FUMA) (33) and the Genomic Regions Enrichment of Annotations Tool (GREAT) (34). FUMA annotates SNPs, uses MAGMA to identify associated genes (of approximately 18,612), and provides gene and gene pathway enrichment analysis (of approximately 10,649 pathways). GREAT analyzes the functional significance of sets of cis-regulatory regions by modeling the genome regulatory landscape using multiple information sources and can determine the functional domain of intergenic

variants. GREAT improves upon the identification of genes associated with noncoding genomic regions through statistically rigorous incorporation of more distal binding sites from 20 ontologies. The GWAS Catalog (<https://www.ebi.ac.uk/gwas>) includes studies and associations if they include a primary GWAS analysis from >100,000 SNPs with a SNP-trait p value <1.0×10<sup>-5</sup> in the overall (initial GWAS plus replication) population. The most significant SNP from each independent locus is extracted.

**Polygenic risk scores, SNP heritability (h<sup>2</sup>), and genetic correlations (r<sub>G</sub>).** Discovery GWAS summary statistics for phenotypes were compiled to score each cohort for polygenic risk. A polygenic score for suicide death was derived using PRSice, version 2.0 (35), and summary statistics were derived from a 10-fold cross-validation procedure to avoid overfitting. To elaborate, k-folds cross-validation is a well-established method (36) allowing a single data set to serve as both training and testing data for the purpose of suicide death polygenic score development and validation. We conservatively set the p-value threshold for predicting case status based on the data to 1.0. This eliminated overfitting arising from choosing the threshold based on the phenotype.

Using related methods, we calculated polygenic scores for several psychiatric and psychological traits in the present data set. Of hundreds of medical and psychological GWASs now available as training weights for calculating polygenic scores, only GWASs with >10,000 individuals and >1,000 case subjects (or for population-based studies, adequate base rates) were selected for these analyses. These generally included the largest medical and psychiatric GWASs. When multiple versions of GWASs were available for the same phenotype (for example, neuroticism or depression), we selected the most comprehensive version (for further details, see the GWAS Atlas at <http://atlas.ctglab.nl>). PRSice was used to calculate individual polygenic scores for 59 phenotypes with estimated risk-allele effect sizes for each discovery sample trait. A polygenic score is traditionally calculated as a weighted sum score, where a score for an individual in the target sample is calculated by the summation of each SNP multiplied by the effect size of that SNP in the discovery GWAS. Based on the cross-disorder psychiatric genomics findings to date, we hypothesized significant positive prediction of suicide with polygenic score for depressive

Odds Ratio	95% CI	p	False Discovery Rate	Nearest Gene	Function	CADD
1.769	1.48–2.11	$3.54 \times 10^{-11}$	$2.7 \times 10^{-4}$	<i>LINC00348</i>	Intergenic	19.86
1.726	1.43–2.07	$1.92 \times 10^{-9}$	$7.2 \times 10^{-3}$	<i>LINC00348</i>	Intergenic	0.33
1.718	1.42–2.06	$3.51 \times 10^{-9}$	$8.7 \times 10^{-3}$	<i>LINC00348</i>	Intergenic	0.39
1.703	1.41–2.05	$5.97 \times 10^{-9}$	0.011	<i>SOGA2P1</i>	Intergenic	2.69
1.707	1.42–2.04	$8.63 \times 10^{-9}$	0.013	<i>LINC00348</i>	Intergenic	4.24
1.605	1.43–2.07	$1.10 \times 10^{-8}$	0.014	<i>ATP10A</i>	Intronic	4.41

symptoms, major depressive disorder, behavioral disinhibition, schizophrenia, autism, loneliness, child IQ, alcohol use, and neuroticism.

Linkage disequilibrium score regression (37, 38) was used to calculate the observed scale common variant  $h^2$  using summary statistics from a logistic regression model with five ancestry covariates and pruning related samples at 0.05  $\hat{p}$  from identical by descent. Linkage disequilibrium score regression was also used to calculate common variant molecular genetic correlations ( $r_G$ ) with psychiatric and medical phenotypes. Finally, we performed secondary analyses to characterize genetic predictors and clinical antecedents of suicide and mode of suicide death (see the Methods section in the online supplement). Specifically, in these secondary analyses, we conducted epidemiological association tests between four sufficiently prevalent and powered modes of death (gunshot, drug overdose, asphyxiation, and violent trauma) and 30 ICD-9 and ICD-10-derived clinical antecedents, and we conducted association tests of the suicide polygenic score with modes of death, adjusting for five ancestry covariates in multivariate regressions.

**Sex differences.** Because suicide rates and modes of suicide death are characterized by substantial sex differences (19, 20), we performed secondary epidemiological and genomic analyses to characterize sex differences in mode of death and clinical antecedents. These sex-stratified analyses mirrored the full sample analyses described above, including sex-stratified epidemiological association tests between four sufficiently prevalent and powered modes of death and 30 ICD-9 and ICD-10-derived clinical antecedents, as well as sex-stratified association tests of the suicide polygenic score with modes of death, adjusting for five ancestry covariates in multivariate regressions. We constrained these exploratory analyses to only those medical diagnoses with frequencies high enough in either females or males to provide decent power for testing and reporting false-discovery-rate corrected p values. Nonetheless, it is worth noting that power for these secondary analyses of sex differences was limited by the number of subjects, which was restricted to case subjects only and stratified by sex and mode of death.

## RESULTS

### Genome-Wide Association: Positional Evidence

A total of six variants from two loci met genome-wide criteria for statistical association with suicide death ( $p < 5 \times 10^{-8}$ ). An

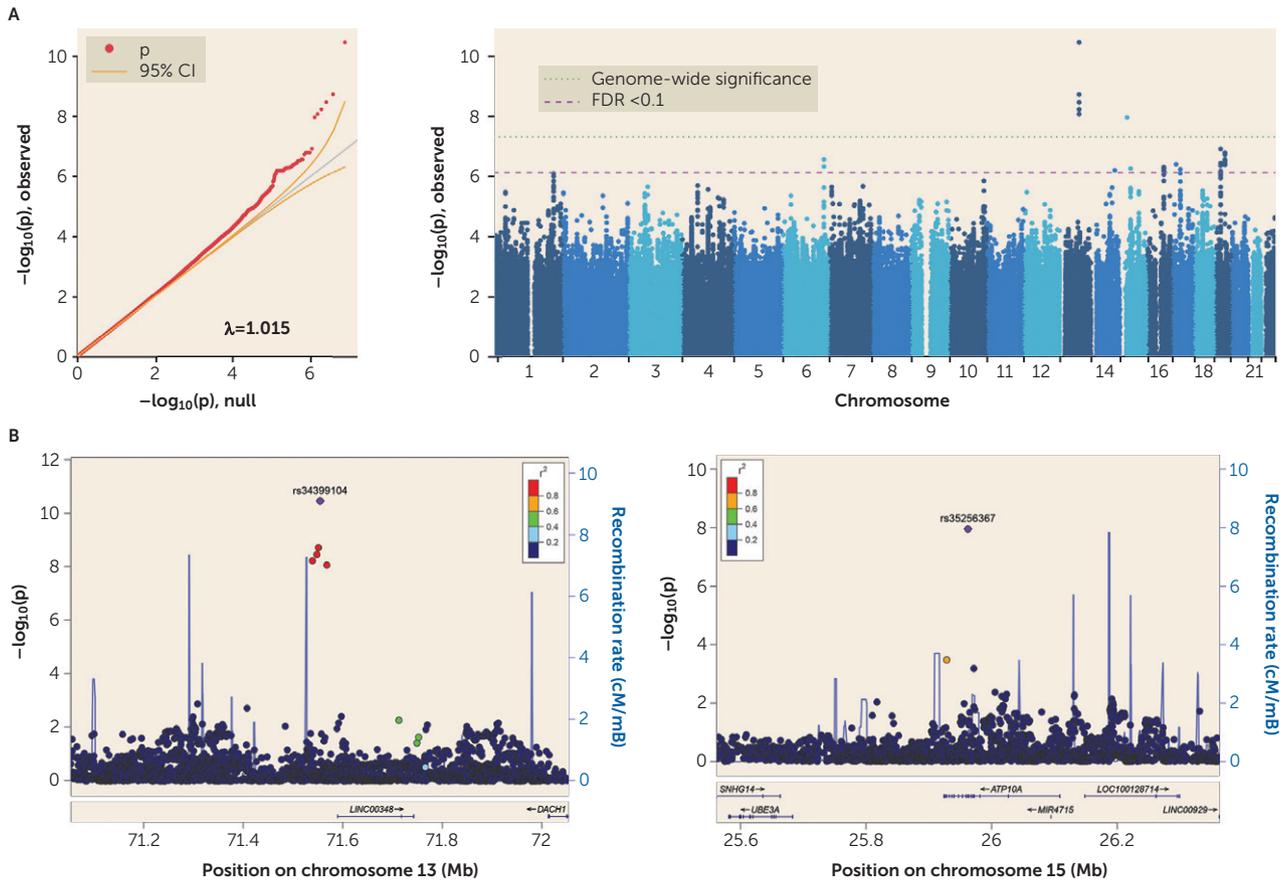
additional 52 variants were nominally significant at a q value  $< 0.1$  and mapped to 22 genes ( $\lambda = 1.015$ ) (Table 1, Figure 1) (39, 40). All results on the full cohort were derived from analyses that were adjusted for the effects of ancestry. Genes associated with top genomic regions are presented in Table S1 in the online supplement. The significant associations found in the chromosome 13 and chromosome 15 regions were supported by additional positive results that were suggestive of association but below the threshold. The large number of signals in the SNP-based tests prompted quality-control analyses varying the degree of linkage disequilibrium pruning before principal component analysis for the purposes of sensitivity analysis. Results and respective lambda values were consistent across these analyses. Additional plots of the top signals in each of eight regions are presented in Figures S2–S8 in the online supplement.

### Gene-Based Results and Pathway Functional Enrichment Tests

Mapping top positional SNP hits to genes suggested 22 gene associations, including chromosome 13 genes, Dachshund family transcription factor 1 (*DACHI*), Ubiquitin-protein ligase (*UBE3A*), and Kelch-like family member 1 (*KLHL1*) on chromosome 15 (see Table S1 in the online supplement). Gene-based analysis using MAGMA (FUMA [1]) identified an additional 10 genes significantly associated ( $q < 0.1$ ) with suicide death (Figure 2; see also Table S1 in the online supplement). Gene Ontology (GO; <http://www.geneontology.org>) pathway results included enrichment of histone modification sites *SETD6*, *COPR5*, and *GATAD2A*. Full gene and GO pathway enrichment results are presented in Tables S2–S3 in the online supplement. Genes showing psychiatric associations are also presented (see Table S2 in the online supplement).

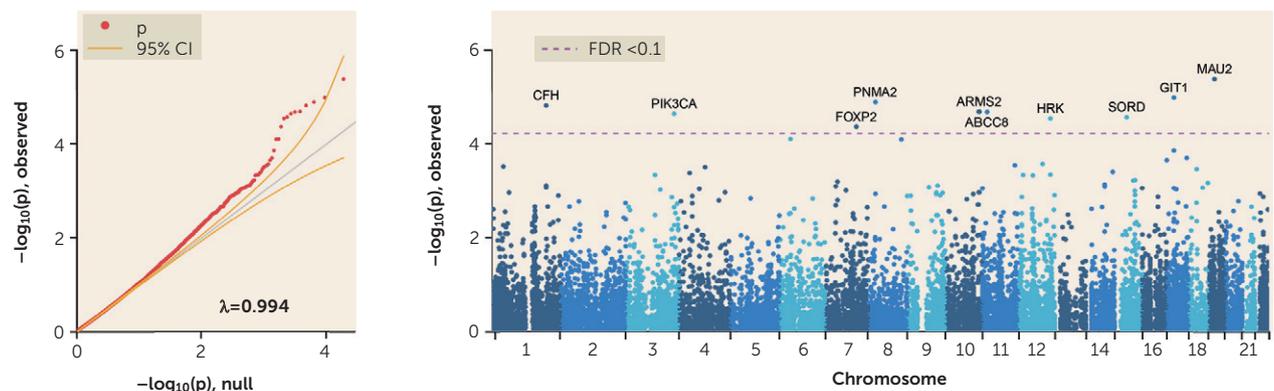
In addition to functional pathways, FUMA analyses indicated significant enrichment for schizophrenia ( $p = 1 \times 10^{-11}$ ) and bipolar disorder ( $p = 1 \times 10^{-17}$ ) in the GWAS Catalog (<https://www.ebi.ac.uk/gwas>; see also Table S1 in the online supplement). Integrative-weighted-scoring in SNP-Nexus (41) suggested regulatory functional significance for one SNP (chr13:71553748:C/T). Ten of the implicated genes from positional or gene-based testing have evidenced genome-wide significant differential gene expression in post-mortem brain analyses of schizophrenia, autism, and bipolar disorder (false discovery rate  $< 0.1$ ; PsychENCODE Consortium [42]; see also Table S4 in the online supplement).

**FIGURE 1. Genome-wide significance for suicide death on chromosomes 13 and 15 in a genome-wide association study<sup>a</sup>**



<sup>a</sup> Panel A shows QQ and Manhattan plots from a genome-wide association study of suicide death. Y-axes for both plots reflect observed p values. The x-axis on the QQ plot indicates the number of significant p values expected under  $H_0$ , and the x-axis on the Manhattan plot maps each chromosome. The purple dashed line indicates false-discovery-rate (FDR)-corrected nominal statistical significance, and the green dotted line represents the threshold for genome-wide significance with Bonferroni correction. Fifty-seven single-nucleotide polymorphisms met threshold for nominal significance, and six met genome-wide significance. Panel B shows regional plots of genome-wide significant loci on chromosomes 13 and 15.

**FIGURE 2. Functional mapping and annotation of genetic association gene-based test results in a genome-wide association study<sup>a</sup>**

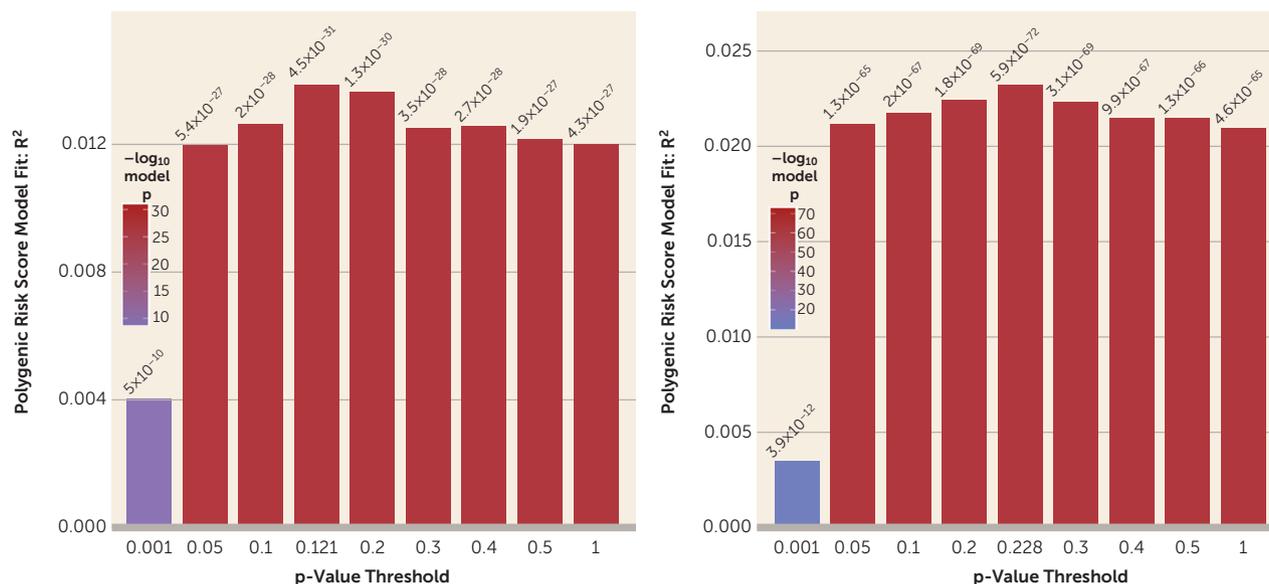


<sup>a</sup> QQ and Manhattan plots of  $>18,000$  genes are presented. Y-axes for both plots are identical and reflect observed p values. The x-axis on the QQ plot indicates the number of significant p values expected under  $H_0$ , and the x-axis on the Manhattan plot maps each chromosome. The purple dashed line indicates the threshold for false-discovery-rate (FDR)-corrected nominal statistical significance; 10 genes met this threshold for nominal significance.

**Polygenic Scores, SNP Heritability, and Genetic Correlations**

In European ancestry training and test samples comprising independent case and control cohorts and accounting for

20 ancestry covariates, suicide polygenic scores significantly predicted suicide death case status. Suicide waves 1 and 2 comprised 1,321 and 2,092 suicide cases, respectively, and were pruned for relatedness within waves before training and

**FIGURE 3. Cross-validation of suicide polygenic case-control prediction in a genome-wide association study<sup>a</sup>**

<sup>a</sup> Polygenic prediction of suicide death case status across two independent cohorts of case and control samples is shown. Training genome-wide association study (GWAS) summary statistics were used to score the test set for suicide polygenic risk. The p-value thresholds are plotted on the x-axis from 0.1 to 1.0, reflecting the top 10%–100% of the common variants from the training GWAS. Plots present model fit when predicting case-control status with wave 1 suicide compared with summary statistics from the Generation Scotland summary database and when predicting with wave 2 suicide compared with summary statistics from the UK10K Project database. Scores are adjusted for 20 principal components. Summary data for the best-fitting and 1.0 p-value threshold models are presented in Table S29 in the online supplement.

testing analyses. These case-control predictions are plotted across p-value thresholds in Figure 3.

The linkage disequilibrium score regression common variant  $h^2$  estimate, based on the summary statistics from a logistic GWAS with five ancestry covariates and pruning to remove related samples, was 0.2463 (SE=0.0356). The  $h^2$  estimate on the liability scale was 0.1567 (SE=0.0218) when considering the 0.2% reported population prevalence of death by suicide in Utah. The lambda statistic in the model was 1.20 ( $\lambda_{1000}=1.04$ ). This lambda statistic remained stable with follow-up testing of 20 principal components and with increased pruning. The suicide death cases differed significantly from the two U.K. control groups on polygenic scores of phenotypes relevant to suicide death. These differences were observed in the expected directions. Original discovery GWASs for all polygenic score phenotypes were filtered to exclude any studies that used these control cohorts (see Table S5 in the online supplement).

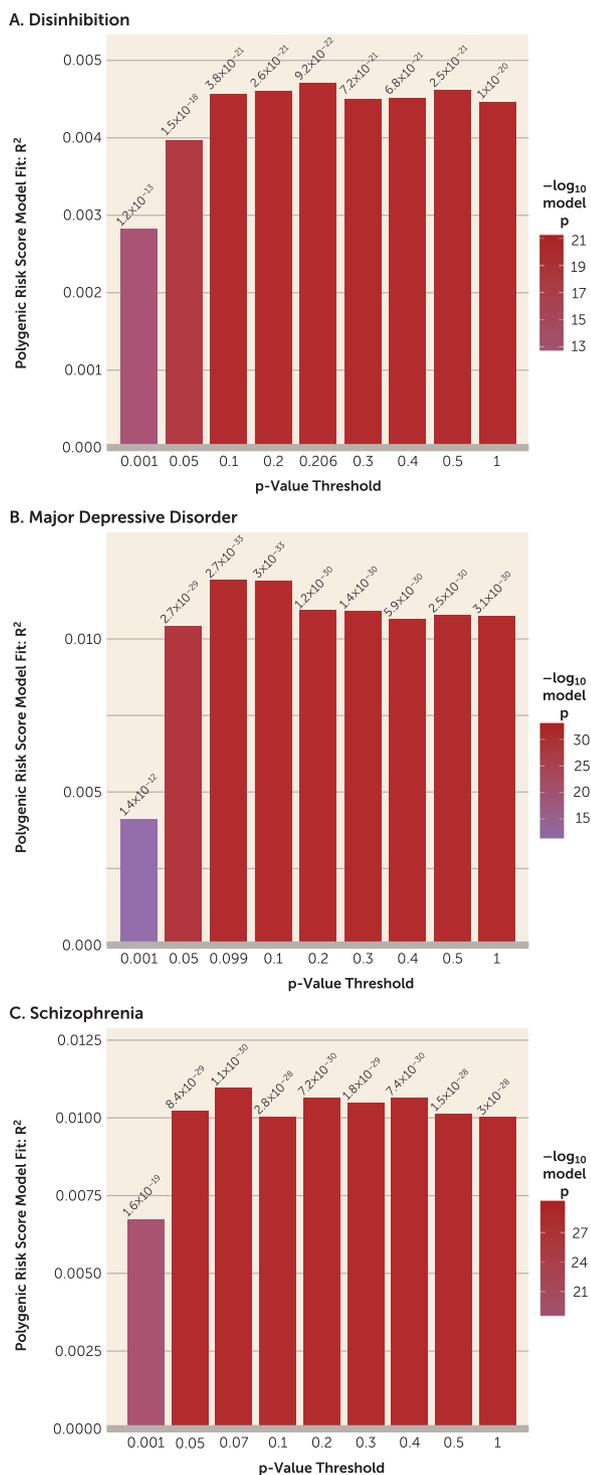
Consistent with our hypotheses, significant polygenic score elevations included alcohol use, autism spectrum disorder, child IQ, depressive symptoms, disinhibition, major depressive disorder, loneliness, and schizophrenia. The largest effect sizes were observed for polygenic scores for behavioral disinhibition, major depressive disorder, and schizophrenia (Figure 4). The LD Hub database (37) provided estimates of SNP-based shared genetic covariance for several phenotypes (see Table S6 in the online supplement), although none of these SNP-based genetic correlations reached statistical significance. Additionally, we disaggregated suicide by mode of death into four categories: gunshot, drug overdose,

asphyxiation, and violent trauma (see the Methods section in the online supplement). We epidemiologically characterized these groups by testing associations with 30 ICD-10-derived clinical antecedents (see Figure S9 and Tables S7–S9 in the online supplement). In addition, we conducted polygenic score association testing with mode of death in all cases. No associations met multiple-testing-adjusted significance criteria ( $q < 0.1$ ; see also Figure S10 and Tables S10–S12 in the online supplement).

### Sex Differences

Epidemiological analyses of sex differences in electronic medical record data indicated that suicide cases among both sexes revealed clinical diagnostic clusters of internalizing-trauma-cluster B psychiatric disorders and metabolic cardiovascular-obesity medical disorders (see Tables S13–S18 in the online supplement). Females were observed to have a higher overall number of diagnoses compared with males, which may reflect greater severity among females, greater severity among males, lower likelihood of males receiving a diagnosis, or lower help-seeking among males. This is broadly consistent with the observed higher relative prevalence rates of gun-related deaths among males and overdose deaths among females. All associations of ICD diagnoses and polygenic scores with mode of death are presented for females and males separately in Figures S11–S14 in the online supplement. All polygenic score analyses include ancestry covariates. All corresponding statistics are presented in Tables S19–S24 in the online supplement. No sex-stratified polygenic score findings met multiple-testing-adjusted significance criteria ( $q < 0.1$ ).

**FIGURE 4. Notable elevations of psychiatric polygenic risk in suicide cases in a genome-wide association study<sup>a</sup>**



<sup>a</sup> The p-value thresholds are plotted on the x-axis. Scores are adjusted for 20 principal components. The largest effect sizes and significance levels were observed for major depressive disorder, schizophrenia, and behavioral disinhibition. Summary data for the best-fitting and 1.0 p-value threshold models for all eight phenotypes—also including alcohol use, autism spectrum disorder, child IQ, depressive symptoms, and loneliness—are presented in Table S29 in the online supplement.

**DISCUSSION**

These results yield several insights and suggest important applications for future extension. To our knowledge, this is the first adequately powered comprehensive genomic study of suicide death. This GWAS of suicide death identified two genome-wide significant loci on chromosomes 13 and 15. The significant SNP-based heritability estimate of suicide death (25%, or 16% on the liability scale) is greater than that previously reported for suicide ideation and suicide attempt (5%) (14, 15) and for suicide attempt within psychiatric diagnosis (3%–10%) (43) and is closer to previous estimates reported for suicide death (35%–48%) (13). Fully half of the genes implicated by our results overlap with schizophrenia results from the GWAS Catalog ( $p=1\times 10^{-11}$ ), and two of these 11 genes (*HS3ST3B*, *NCAN*) have previous associations with risk of suicidal behavior (for details on the relevant literature, see Table S25 in the online supplement; for allele frequencies and effect sizes by cohort, see Table S26 in the online supplement).

All but four of these schizophrenia-associated genes also drove a significant GWAS Catalog association with bipolar disorder ( $p=1\times 10^{-17}$ ), supporting the hypothesis of genetic liability to more nonspecific psychiatric illness. Of particular interest, the region on chromosome 19p13.11 in our study is supported by evidence from the most recent, well-powered studies of both bipolar disorder (44) and schizophrenia (45). In addition, our region on chromosome 16q21 is supported by evidence not only from several large GWASs of schizophrenia but also from a large meta-analysis of autism spectrum disorder (45–50).

Genome-wide significant SNPs on chromosome 13 (Table 1) are all intergenic and vary with respect to a combined annotation-dependent depletion score, an index for scoring the deleteriousness of single-nucleotide variants as well as insertion and deletion variants in the human genome. Notably, the variant with the strongest GWAS effect size, single-nucleotide variant rs34399104, reaches a combined annotation-dependent depletion score of almost 20, indicating that it is in the top 1% of deleterious variants in the genome.

Direct comparison of the effect sizes in our GWAS with 110 effect sizes reported in previous studies (12–15, 43, 51) identified only one nominally significant variant in the largest GWAS of suicidal behavior (43), rs4870888, nearest to *FER1L6* and *FER1L6-AS2*. (See Table S27 in the online supplement for a direct comparison of previously reported effect sizes with those in our study.) The single-nucleotide variant rs7989250 (chromosome 13) was reported for ordinal suicidality in the UK Biobank GWAS and reached a p-value threshold  $<0.05$  for replication in the suicide GWAS conducted by Otsuka et al. (13) but did not reach the threshold in our study (see Table S27 in the online supplement).

We were also able to cross-validate prediction of suicide death case-control status with polygenic scores for suicide death, accounting for population stratification.

While within-case polygenic score validation relies on stronger assumptions than other polygenic score case-control comparisons, the considerable overlapping polygenic signal across p-value thresholds in this study (suicide deaths,  $N=3,413$ ) is consistent with the two existing smaller polygenic analyses predicting suicidality (13, 14). In one of those studies, suicidal behavior polygenic risk score predicted increased odds of suicide death, with 127 deaths (14), and in the other, a significant SNP-based genetic correlation of suicide death was observed between two independent cohorts (suicide deaths,  $N=746$ ) (13).

Perhaps most compellingly, suicide death was strongly associated with polygenic scores for multiple psychiatric and psychological traits. As noted above, these included alcohol use, autism spectrum disorder, child IQ, depression, disinhibition, schizophrenia, and loneliness (for additional polygenic risk score plots, see Figure S15 in the online supplement). These results were consistent with expectations, because strong associations in case samples compared with both control cohorts were observed with behavioral disinhibition and major depressive disorder—arguably the two most critical antecedents of suicide death (52). More generally, these results are consistent with the emerging consensus that genetic correlations across psychiatric diagnoses and traits are substantial (53). Further research with an even larger suicide death cohort will be needed to determine the ultimate cause of these polygenic score associations with suicide death, with leading hypotheses arguing for the presence of a p-factor capturing general liability to psychopathology (54), contrasting with those arguing for pleiotropic effects across a small set of psychiatric disorder clusters (55).

### Limitations

Several limitations to this study should be noted. First, it is possible that population stratification or platform confounders stemming from convenience control samples could lead to false positives. To address this risk, we employed very conservative filters of both loci and study subjects to minimize false positives. Values of  $\lambda_{GC} < 1.05$  are generally considered benign, although inflation in  $\lambda_{GC}$  is proportional to the sample size (56). Control samples in both the Generation Scotland database and the UK10K Project cohort were extensively assessed for ancestral comparability to samples comprising individuals who died by suicide in both the 1,000 Genomes Utah CEU (25) and Utah European ancestry databases, and both matched extremely well. Incorporating two separate control cohorts (and formal analytical comparison of these cohorts to the corresponding 1,000 Genomes populations) also allowed us to detect and filter out loci related to platform-specific technical artifacts and subtle variation in population structure before case-control analyses. Results of tests for differences in loadings for up to 200 principal components by cohort and by case-control status are presented in Table S28 in the online supplement. Before and after imputation, we excluded any and all variants generating q values  $< 0.1$  in a control-control GWAS. Additionally, employing

ancestral Mahalanobis (i.e., centroid) distance pruning, we have leveraged the fact that the CEU 1,000 Genomes Project subjects are largely ancestrally homogeneous with our case samples to restrict our control subjects (and case subjects) to a highly ancestrally homogeneous group.

It is also possible that the use of a mixed control group—both healthy subjects and those with common medical conditions, as in the UK10K Project—could have exerted a mild bias on effect estimates toward the null, which we do not view as highly problematic, given that the direction of the bias effectively reduces the risk of false positive findings. In the suicide case cohort, rates of common medical conditions are much higher than population base rates. However, in stratified control analyses, effect sizes specific to our case sample compared with those in the UK10K Project were elevated relative to the case sample from the Generation Scotland database, which is inconsistent with potential confounding of signal as a result of other disorders. Additionally, effect sizes for polygenic scores for other disorders, across 1,000 p-value thresholds, were far greater between case and control subjects than between control groups, where differences were negligible and distributions of polygenic scores often overlapped. None of the implicated variants were significantly different in allele frequency between control groups in the control-control GWAS. However, it should be noted that allele frequency was relatively low for the top variants across all cohorts.

Another limitation of this study lies in the use of electronic medical record data, in which prevalences for disorders are always confounded with age, and in a number of ways. For example, a missing diagnosis could be more or less likely in a given decade during which a case was assessed. Younger ages are more likely to have complete electronic medical record information yet are less likely to be diagnosed with comorbid medical conditions, schizophrenia, or personality disorders.

Importantly, there was insufficient diversity in the case cohort to examine effect sizes for any non-European ancestries. We expect a sufficient number of case samples for a Mexican-American ancestry GWAS within 5 years, and we will prioritize diverse ancestry groups in order to minimize potential health disparities stemming from homogeneous summary statistics (57).

### Future Directions

A recent review of suicide prediction models has indicated that positive predictive values for suicide attempt are quite high (0.9), but positive predictive values for suicide death continue to hover near zero (58, 59). In this study, we have discovered multiple genome-wide significant loci and genes, strong polygenic signal, and significantly increased genetic risks for behavioral disinhibition, major depression, depressive symptoms, autism spectrum disorder, psychosis, and alcohol use disorder among case samples. Future modeling of multiple polygenic risks and phenotypic risk factors may help isolate important moderators of risk and improve objective risk measures of suicide death. Importantly, ethical

challenges associated with predictive models of suicide death are significant and must be addressed proactively across psychiatric and medical domains (6, 60).

#### AUTHOR AND ARTICLE INFORMATION

Department of Psychiatry, University of Utah School of Medicine, Salt Lake City (Docherty, Shabalin, DiBlasi, Monson, Adkins, Bakian, Crowell, Chen, Darlington, Gray, Anderson, Jerominski, Coon); Huntsman Mental Health Institute, University of Utah School of Medicine, Salt Lake City (Docherty); Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University School of Medicine, Richmond (Docherty, Bacanu); Department of Genetics and Genomic Sciences, Pamela Sklar Division of Psychiatric Genomics, Icahn School of Medicine at Mount Sinai, New York (Mullins); Department of Psychology, University of Utah, Salt Lake City (Crowell); Utah State Office of the Medical Examiner, Utah Department of Health, Salt Lake City (Callor, Christensen); Mental Illness Research, Education, and Clinical Center, Veterans Integrated Service Network 19, George E. Whalen Department of Veterans Affairs Medical Center, Salt Lake City (Gray); Department of Pediatrics, University of Utah School of Medicine, Salt Lake City (Keeshin); Health Sciences Center Core Research Facility, University of Utah, Salt Lake City (Klein); Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, U.K. (Hayward, Porteous, McIntosh); Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, U.K. (Porteous, McIntosh); and Janssen Research and Development, Neuroscience Therapeutic Area, Titusville, N.J. (Li).

Send correspondence to Dr. Docherty (anna.docherty@utah.edu).

Presented at the World Congress of Psychiatric Genetics, October 26–31, 2019, Anaheim, Calif.

Supported by the American Foundation for Suicide Prevention (Drs. Docherty and Bakian), the Brain and Behavior Research Foundation (Drs. Docherty and Shabalin [NARSAD Young Investigator Award]), the Clark Tanner Foundation (Drs. Coon, Darlington, and Bakian), Janssen Research (Drs. Coon and Li), NIMH (grants R01MH123619 and K01MH109765 to Dr. Docherty and R01MH099134 to Dr. Coon), and the University of Utah EDGE Scholar Program (Dr. Shabalin), as well as a donation from the Sharon Kae Lehr Endowed Research Fund in Memory of James Raymond Crump (Dr. Gray). Partial support for all data sets housed within the Utah Population Database is provided by the Huntsman Cancer Institute (HCI), and the HCI Cancer Center (grant P30-CA-42014 from NIH). Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates (grant CZD/16/6) and the Scottish Funding Council (grant HR03006). Genotyping of the Generation Scotland samples was carried out by the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility, Edinburgh, Scotland, and was funded by the Medical Research Council United Kingdom and Wellcome Trust (Wellcome Trust Strategic Award Stratifying Resilience and Depression Longitudinally). Genotyping was performed by the University of Utah Genomics Core and by Illumina, with support from Janssen Research and Development. This work was also supported by research funding from Janssen Research and Development to the University of Utah.

Dr. Qingqin Li is employed by Janssen Pharmaceuticals. The other authors report no financial relationships with commercial interests.

Received October 5, 2019; revisions received February 22 and April 27, 2020; accepted May 27, 2020.

#### REFERENCES

1. Turecki G, Brent DA: Suicide and suicidal behaviour. *Lancet* 2016; 387:1227–1239
2. Hedegaard HCS, Warner M: Suicide mortality in the United States, 1999–2017, in National Center for Health Statistics Data Brief, vol 330. Hyattsville, Md, National Center for Health Statistics, 2018
3. Kochanek KD, Murphy SL, Xu J, et al: Deaths: final data for 2014. *Natl Vital Stat Rep* 2016; 65:1–122
4. Murphy SL, Mathews TJ, Martin JA, et al: Annual summary of vital statistics: 2013–2014. *Pediatrics* 2017; 139(6):e20163239 PubMed
5. Case A, Deaton A: Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century. *Proc Natl Acad Sci USA* 2015; 112:15078–15083
6. Barak-Corren Y, Castro VM, Javitt S, et al: Predicting suicidal behavior from longitudinal electronic health records. *Am J Psychiatry* 2017; 174:154–162
7. Franklin JC, Ribeiro JD, Fox KR, et al: Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. *Psychol Bull* 2017; 143:187–232
8. McGuffin P, Marusic A, Farmer A: What can psychiatric genetics offer suicidology? *Crisis* 2001; 22:61–65
9. Mullins N, Perroud N, Uher R, et al: Genetic relationships between suicide attempts, suicidal ideation and major psychiatric disorders: a genome-wide association and polygenic scoring study. *Am J Med Genet B Neuropsychiatr Genet* 2014; 165B:428–437
10. Turecki G: Suicidal behavior: is there a genetic predisposition? *Bipolar Disord* 2001; 3:335–349
11. Coon H, Darlington TM, DiBlasi E, et al: Genome-wide significant regions in 43 Utah high-risk families implicate multiple genes involved in risk for completed suicide. *Mol Psychiatry* (Epub ahead of print October 23, 2018) doi: 10.1038/s41380-018-0282-3
12. Galfalvy H, Haghghi F, Hodgkinson C, et al: A genome-wide association study of suicidal behavior. *Am J Med Genet B Neuropsychiatr Genet* 2015; 168:557–563
13. Otsuka I, Akiyama M, Shirakawa O, et al: Genome-wide association studies identify polygenic effects for completed suicide in the Japanese population. *Neuropsychopharmacology* 2019; 44:2119–2124
14. Strawbridge RJ, Ward J, Ferguson A, et al: Identification of novel genome-wide associations for suicidality in UK Biobank, genetic correlation with psychiatric disorders and polygenic association with completed suicide. *EBioMedicine* 2019; 41:517–525
15. Erlangsen A, Appadurai V, Wang Y, et al: Genetics of suicide attempts in individuals with and without mental disorders: a population-based genome-wide association study. *Mol Psychiatry* (Epub ahead of print August 16, 2018) doi: 10.1038/s41380-018-0218-y
16. O'Brien E, Kerber R, Smith K, et al: Familial mortality in the Utah population database: characterizing a human aging phenotype. *J Gerontol A Biol Sci Med Sci* 2007; 62:803–812
17. Kerber R, O'Brien E, Smith KR, et al: Familial aggregation of elderly cause-specific mortality: analysis of extended pedigrees in Utah, 1904–2002, in *Kinship and Demographic Behavior in the Past*. New York, Springer, 2008, pp 243–258
18. World Health Organization: *International Statistical Classification of Diseases and Related Health Problems—Tenth Revision*, 2nd ed. Geneva, World Health Organization, 2004
19. Schrijvers DL, Bollen J, Sabbe BGC: The gender paradox in suicidal behavior and its impact on the suicidal process. 2012; 138:19–26
20. Skogman K, Alsen M, Ojehagen A: Sex differences in risk factors for suicide after attempted suicide—a follow-up study of 1052 suicide attempters. *Soc Psychiatry Psychiatr Epidemiol* 2004; 39:113–120
21. Utah Medical Examiner. [https://le.utah.gov/xcode/Title26/Chapter4/C26-4\\_1800010118000101.pdf](https://le.utah.gov/xcode/Title26/Chapter4/C26-4_1800010118000101.pdf)
22. Jorde LB: Inbreeding in the Utah Mormons: an evaluation of estimates based on pedigrees, isonymy, and migration matrices. *Ann Hum Genet* 1989; 53:339–355
23. Smith BH, Campbell H, Blackwood D, et al: Generation Scotland: the Scottish Family Health Study—a new resource for researching genes and heritability. *BMC Med Genet* 2006; 7:74
24. Walter K, Min JL, Huang J, et al: The UK10K project identifies rare variants in health and disease. *Nature* 2015; 526:82–90

25. Auton A, Brooks LD, Durbin RM, et al: A global reference for human genetic variation. *Nature* 2015; 526:68–74
26. Shabalín AA, Hattab MW, Clark SL, et al: RaMWAS: fast methylome-wide association study pipeline for enrichment platforms. *Bioinformatics* 2018; 34:2283–2285
27. Shabalín A, Clark S, Hattab M, et al: Fast Methylome-Wide Association Study Pipeline for Enrichment Platforms, R package version 1.2.0, 2017
28. Das S, Forer L, Schönherr S, et al: Next-generation genotype imputation service and methods. *Nat Genet* 2016; 48:1284–1287
29. Loh PR, Danecek P, Palamara PF, et al: Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet* 2016; 48:1443–1448
30. Chang CC, Chow CC, Tellier LC, et al: Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015; 4:7
31. Purcell S, Chang C, Sinai N: PLINK 1.9 beta, 2017
32. Zhou X, Stephens M: Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* 2012; 44:821–824
33. Watanabe K, Taskesen E, van Bochoven A, et al: Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* 2017; 8:1826
34. McLean CY, Bristor D, Hiller M, et al: GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 2010; 28:495–501
35. Euesden J, Lewis CM, O'Reilly PF: PRSice: polygenic risk score software. *Bioinformatics* 2015; 31:1466–1468
36. Hastie T, Tibshirani R, Friedman JH: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, Springer, 2009
37. Zheng J, Erzurumluoglu AM, Elsworth BL, et al: LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* 2017; 33:272–279
38. Bulik-Sullivan BK, Loh PR, Finucane HK, et al: LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 2015; 47:291–295
39. Narum SR: Beyond Bonferroni: less conservative analyses for conservation genetics. *Conserv Genet* 2006; 7:783–787
40. Hochberg Y, Benjamini Y: More powerful procedures for multiple significance testing. *Stat Med* 1990; 9:811–818
41. Dayem Ullah AZ, Oscanoa J, Wang J, et al: SNPnexus: assessing the functional relevance of genetic variation to facilitate the promise of precision medicine. *Nucleic Acids Res* 2018; 46(W1):W109–W113
42. Gandal MJ, Zhang P, Hadjimichael E, et al: Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* 2018; 362:(6420):eaat8127 doi: 10.1126/science.aat8127 PubMed
43. Mullins N, Bigdeli TB, Børglum AD, et al: GWAS of suicide attempt in psychiatric disorders and association with major depression polygenic risk scores. *Am J Psychiatry* 2019; 176:651–660
44. Stahl EA, Breen G, Forstner AJ, et al: Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat Genet* 2019; 51:793–803
45. Pardiñas AF, Holmans P, Pocklington AJ, et al: Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet* 2018; 50:381–389
46. Autism Spectrum Disorders Working Group of the Psychiatric Genomics Consortium: Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol Autism* 2017; 8:21
47. Schizophrenia Working Group of the Psychiatric Genomics Consortium: Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; 511:421–427
48. Ikeda M, Takahashi A, Kamatani Y, et al: Genome-wide association study detected novel susceptibility genes for schizophrenia and shared trans-populations/diseases genetic effect. *Schizophr Bull* 2019; 45:824–834
49. Periyasamy S, John S, Padmavati R, et al: Association of schizophrenia risk with disordered niacin metabolism in an Indian genome-wide association study. *JAMA Psychiatry* 2019; 76:1026–1034
50. Goes FS, McGrath J, Avramopoulos D, et al: Genome-wide association study of schizophrenia in Ashkenazi Jews. *Am J Med Genet B Neuropsychiatr Genet* 2015; 168:649–659
51. Ruderfer DM, Walsh CG, Aguirre MW, et al: Significant shared heritability underlies suicide attempt and clinically predicted probability of attempting suicide. *Mol Psychiatry* (Epub ahead of print, January 4, 2019) doi: 10.1038/s41380-018-0326-8
52. Nock MK, Green JG, Hwang I, et al: Prevalence, correlates, and treatment of lifetime suicidal behavior among adolescents: results from the National Comorbidity Survey Replication Adolescent Supplement. *JAMA Psychiatry* 2013; 70:300–310
53. Brainstorm Consortium; Anttila V, Bulik-Sullivan B, Finucane HK, et al: Analysis of shared heritability in common disorders of the brain. *Science* 2018; 360(6395):eaap8757
54. Caspi A, Houts RM, Belsky DW, et al: The p factor: one general psychopathology factor in the structure of psychiatric disorders? *Clin Psychol Sci* 2014; 119–137
55. Cross-Disorder Group of the Psychiatric Genomics Consortium: Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. *Cell* 2019; 179:1469–1482.e11
56. Price AL, Zaitlen NA, Reich D, et al: New approaches to population stratification in genome-wide association studies. *Nat Rev Genet* 2010; 11:459–463
57. Martin AR, Kanai M, Kamatani Y, et al: Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet* 2019; 51:584–591
58. Belsher BE, Smolenski DJ, Pruitt LD: Positive predictive values and potential success of suicide prediction models-reply. *JAMA Psychiatry* (Epub ahead of print June 26, 2019) doi: 10.1001/jamapsychiatry.2019.1510
59. Belsher BE, Smolenski DJ, Pruitt LD, et al: Prediction models for suicide attempts and deaths: a systematic review and simulation. *JAMA Psychiatry* 2019; 76:642–651
60. Adkins DE: Machine learning and electronic health records: a paradigm shift. *Am J Psychiatry* 2017; 174:93–94
61. Kircher M, Witten DM, Jain P, et al: A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014; 46:310–315
62. Rentzsch P, Witten D, Cooper GM, et al: CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* 2019; 47(D1):D886–D894