



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## Genome-wide association study of suicide death and polygenic prediction of clinical antecedents.

**Citation for published version:**

McIntosh, A 2020, 'Genome-wide association study of suicide death and polygenic prediction of clinical antecedents.', *American Journal of Psychiatry*. <https://doi.org/10.1176/appi.ajp.2020.19101025>

**Digital Object Identifier (DOI):**

[10.1176/appi.ajp.2020.19101025](https://doi.org/10.1176/appi.ajp.2020.19101025)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

American Journal of Psychiatry

**Publisher Rights Statement:**

This is the authors' peer-reviewed manuscript as accepted for publication.

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



**GENOME-WIDE ASSOCIATION STUDY OF SUICIDE DEATH  
AND POLYGENIC PREDICTION OF CLINICAL ANTECEDENTS**

Journal:	<i>The American Journal of Psychiatry</i>
Manuscript ID	AJP-19-10-1025.R2
Manuscript Type:	Article
Date Submitted by the Author:	27-Apr-2020
Complete List of Authors:	<p>Docherty, Anna; University of Utah School of Medicine, Psychiatry; Virginia Commonwealth University, Psychiatry Shabalin, Andrey; University of Utah School of Medicine DiBlasi, Emily; University of Utah School of Medicine, Psychiatry Monson, Eric; University of Utah School of Medicine, Psychiatry Mullins, Niamh; Icahn School of Medicine at Mount Sinai, Genetics and Genomic Sciences Adkins, Daniel; University of Utah School of Medicine, Psychiatry; University of Utah, Sociology Bacanu, Silviu; Virginia Commonwealth University, Psychiatry Bakian, Amanda; University of Utah School of Medicine, Psychiatry Crowell, Sheila; University of Utah School of Medicine, Psychiatry Darlington, Todd; University of Utah School of Medicine, Psychiatry Callor, W; Utah Department of Health, Office of the Medical Examiner Christensen, Erik; Utah Department of Health, Office of the Medical Examiner Gray, Douglas; University of Utah School of Medicine, Psychiatry Keeshin, Brooks; University of Utah School of Medicine, Pediatrics Klein, Michael; University of Utah School of Medicine, Health Sciences Center Core Research Facility Anderson, John ; University of Utah School of Medicine, Psychiatry Jerominski, Leslie; University of Utah School of Medicine, Psychiatry Hayward, Caroline; University of Edinburgh School of Philosophy Psychology and Language Sciences, Department of Psychology Porteous, David; University of Edinburgh Western General Hospital, Institute of Genetics and Molecular Medicine McIntosh, Andrew; University of Edinburgh Western General Hospital, Centre for Genomic and Experimental Medicine Li, Qingqin; Janssen Research and Development LLC Coon, Hilary; University of Utah School of Medicine, Psychiatry</p>
Keywords:	Suicide and Self-Harm, Major Depressive Disorder < Depressive Disorders, Psychosis see Schizophrenia Spectrum and Other Psychotic Disorders, Genetics / Genomics

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



## Genome-wide association study of suicide death and polygenic prediction of clinical antecedents

Dr. Anna R. Docherty, Ph.D.<sup>1,2,3\*</sup>, Dr. Andrey A. Shabalina, Ph.D.<sup>1</sup>, Dr. Emily DiBlasi, Ph. D.<sup>1</sup>, Dr. Eric Monson, M.D.<sup>1</sup>, Dr. Niamh Mullins, Ph.D.<sup>4</sup>, Prof. Daniel E. Adkins, Ph.D.<sup>1</sup>, Prof. Silviu-Alin Bacanu, Ph.D.<sup>3</sup>, Dr. Amanda V. Bakian, Ph.D.<sup>1</sup>, Prof. Sheila Crowell, Ph.D.<sup>1,5</sup>, Dr. Todd M. Darlington, Ph.D.<sup>1</sup>, Mr. Brandon Callor, M.S.<sup>6</sup>, Dr. Erik D. Christensen, M.D.<sup>6</sup>, Prof. Douglas Gray, M.D.<sup>1,7</sup>, Dr. Brooks Keeshin, M.D.<sup>8</sup>, Mr. Michael Klein, M.S.<sup>9</sup>, Mr. John S. Anderson, B.S.<sup>1+</sup>, Ms. Leslie Jerominski, M.S.<sup>1</sup>, Prof. Caroline Hayward, Ph.D.<sup>10</sup>, Prof. David J. Porteous, Ph.D.<sup>10,11</sup>, Prof. Andrew McIntosh, M.D.<sup>10,11</sup>, Dr. Qingqin Li Ph.D.<sup>12</sup>, Prof. Hilary Coon Ph.D.<sup>1</sup>

1. Department of Psychiatry, University of Utah School of Medicine, Salt Lake City, UT USA
  2. Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT USA
  3. Virginia Institute for Psychiatric & Behavioral Genetics, Virginia Commonwealth University School of Medicine, Richmond, VA USA
  4. Pamela Sklar Division of Psychiatric Genomics, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA
  5. Department of Psychology, University of Utah, Salt Lake City, UT USA
  6. Utah State Office of the Medical Examiner, Utah Department of Health, Salt Lake City, UT USA
  7. Mental Illness Research, Education and Clinical Center (MIRECC), Veterans Integrated Service Network 19 (VISN 19), George E. Whalen Department of Veterans Affairs Medical Center, Salt Lake City, UT, USA
  8. Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, UT, USA
  9. Health Sciences Center Core Research Facility, University of Utah, Salt Lake City, UT USA
  10. MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, United Kingdom
  11. Centre for Genomic and Experimental Medicine, Institute of Genetics & Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, United Kingdom
  12. Janssen Research & Development, LLC, Neuroscience Therapeutic Area, Titusville, NJ USA
- + MD/PhD student, \* Corresponding author

**Previous presentation:** World Congress of Psychiatric Genetics, Oct 26-31, 2019, Anaheim, CA, USA.

**Address for reprint:** Corresponding author, Anna R. Docherty, Departments of Psychiatry & Human Genetics, University of Utah School of Medicine, 501 Chipeta Way, Salt Lake City, Utah 84110, USA, tel. +1 612 227 2754. Email: [anna.docherty@utah.edu](mailto:anna.docherty@utah.edu).

**Disclosures:** Dr. Qingqin Li, a co-author, is employed by Janssen Pharmaceuticals. All other authors have no financial relationships with commercial interests.

**Acknowledgements:** This work was supported by the National Institute of Mental Health (H.C., grant number R01MH099134; A.D., grant number K01MH093731), a research contract from Janssen Research, LLC (H.C. & Q.L.); the American Foundation for Suicide Prevention (A.D. & A.B.), the Brain & Behavior Research Foundation (A.D. & A.S., NARSAD Young Investigator Awards); the University of Utah EDGE Scholar Program (A.D.); the Clark Tanner Foundation (HC, TD, AVB), and a donation from the Sharon Kae Lehr Endowed Research Fund in Memory of James Raymond Crump (DG). Partial support for all datasets housed within the Utah Population Data Base is provided by the Huntsman Cancer Institute (HCI), <http://www.huntsmanccancer.org/>, and the HCI Cancer Center Support grant, P30-CA-42014 from the NIH. Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates (CZD/16/6) and the Scottish Funding Council (HR03006). Genotyping of the GS 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 UK and the Wellcome Trust (Wellcome Trust Strategic Award “Stratifying Resilience and Depression Longitudinally” (STRADL) Reference 104036/Z/14/Z). Genotyping was performed by U of Utah Genomic Core and by Illumina, Inc with support from Janssen Research & Development, LLC. This work was also supported by research funding from Janssen Research & Development, LLC to University of Utah.

## GENOME-WIDE ASSOCIATION STUDY OF SUICIDE DEATH AND POLYGENIC PREDICTION OF CLINICAL ANTECEDENTS

### ABSTRACT

**Objective:** Suicide death 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 study. To address this limitation, we conducted the first comprehensive genomic analysis of suicide death using previously unpublished genotype data from a large, population-ascertained cohort.

**Methods:** The analysis sample consisted of 3,413 population-ascertained cases of European ancestry and 14,810 ancestrally matched controls. Analytical methods included principal components 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, polygenic score analyses, as well as SNP heritability and genetic correlation estimation using LD score regression.

**Results:** GWAS identified two genome-wide significant loci (6 SNPs,  $p < 5 \times 10^{-8}$ , including 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  $h^2_{SNP} = .25$  (SE = .04), and .16 (.02) when converted to a liability scale. Notably, suicide polygenic scores were significantly predictive. Polygenic scores for several other psychiatric disorders and psychological traits were also predictive, particularly behavioral disinhibition and major depressive disorder.

**Conclusions:** In this report, we identify multiple genome-wide significant loci/genes and demonstrate polygenic score prediction of suicide death case-control status, adjusting for ancestry, in independent training and test sets. Additionally, we report that suicide death cases have increased genetic risk for behavioral disinhibition, major depression, depressive symptoms, autism spectrum disorder, psychosis, and alcohol use disorder relative to controls.

## INTRODUCTION

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

While it is established that suicide is significantly heritable,<sup>8</sup> 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 though results from well-powered GWAS of suicidal behavior have advanced our understanding of the process,<sup>9</sup> the vast majority of suicidal behavior does not result in suicide death.<sup>1,10</sup> Thus, suicidal behavior represents a less severe and likely more heterogeneous phenotype than suicide death, characteristics which may adversely affect statistical power to detect genetic associations. Conversely, the unambiguous phenotype of suicide death avoids several confounds 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 narrow ascertainment, studying only individuals with specific diagnoses (e.g., mood disorders, psychotic disorders) in order to maximize severity and accommodate post hoc study designs. We avoid this limitation here, 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.<sup>9,11</sup> 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 have been, however, a few modestly powered GWAS of suicide death published,<sup>12,13</sup> the largest of which combined two independent Japanese cohorts totaling 746 completed suicide cases.<sup>13</sup> Further, a recent UK Biobank GWAS of suicidal behavior has conducted preliminary polygenic analyses

1  
2  
3 of suicide death using a very modest number completed suicide cases ( $n = 127$ ).<sup>14</sup> Of note, these smaller  
4 studies reported evidence of polygenicity in both suicidal behavior and suicide death, but with widely  
5 varying SNP-based heritability estimates, from  $h^2_{SNP} = 4.6$ <sup>15</sup> and 7.6%<sup>14</sup> for suicidal behavior not including  
6 death, to  $h^2_{SNP} = 35$ –48% for suicide death.<sup>13</sup> The current study marks a notable advance from earlier  
7 research, as it uses the world's largest DNA databank of suicide death, merged with a massive bank of  
8 electronic medical record and sociodemographic data,<sup>16,17</sup> to comprehensively model common variant  
9 genetic, and clinical phenotypic, precursors of suicide death.  
10  
11  
12  
13  
14  
15

16 This study adds to the growing body of genomic research on suicide,<sup>13,14</sup> providing the first  
17 adequately powered genome-wide association study of suicide death. Analyses integrated data on modes  
18 of suicide death, medical and psychiatric diagnostic (ICD-10) codes,<sup>18</sup> and medical and psychiatric  
19 polygenic scores to comprehensively model common variant genetic risk for suicide death. Secondary  
20 analyses of sex differences were conducted due to the substantial sex differences in suicide death rates  
21 and modes of suicide death.<sup>19,20</sup> Reliable and significant prediction of case-control status was achieved,  
22 adjusting for ancestry, using both (1) a novel polygenic score for suicide death and (2) polygenic scores  
23 for a range of comorbid psychiatric and medical risk factors, particularly behavioral disinhibition, major  
24 depressive disorder, and depressive symptoms. Additionally, gene set enrichment, SNP heritability, and  
25 genetic correlations were examined in this unique, unpublished data resource. We conclude with a brief  
26 discussion of the relevance of our findings to the broader field and promising future research directions.  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

## 40 **METHODS**

### 41 **Sample Ascertainment**

#### 42 Cases

43  
44 In collaboration with the centralized, statewide Utah Office of Medical Examiner (OME), the authors  
45 obtained DNA samples from ~6000 persons who died by suicide. The centralized OME and conservative  
46 determination helped to maximize the accuracy of suicide case status.<sup>21</sup> Suicide cause-of-death  
47 determination results from a detailed investigation of the scene of the death and circumstances of death,  
48 determination of medical conditions by full autopsy, review of medical and other public records  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 concerning the case, interviews with survivors, and standard toxicology workups. Suicide determination is  
4 traditionally made quite conservatively due to its impact on surviving relatives.  
5

6  
7 DNA from suicide deaths were extracted from whole blood using the Qiagen Autopure LS  
8 automated DNA extractor (www.qiagen.com). Genotyping is described below. Data collection and  
9 genotyping for the Utah Suicide Study is ongoing. To date, 4,381 samples have been genotyped in two  
10 waves as determined by the date of sample receipt and the availability of genotyping funds (for detailed  
11 information on sample collection see Supplementary Methods S1. After quality control procedures and  
12 ancestry analysis, data are comprised of 3,413 Utah suicide deaths. The Utah population is primarily  
13 Northwestern European in ancestry, a relatively genetically homogeneous group with low inbreeding  
14 across generations, comparable to the rest of the United States.<sup>22</sup>  
15  
16  
17  
18  
19  
20  
21  
22  
23

#### 24 Controls

25  
26 *Generation Scotland.* Controls closely matching the Northern European ancestry of the cases were  
27 obtained from previously curated datasets in the UK. Wave 1 analysis included 3,623 founder controls  
28 from the population-based Generation Scotland Scottish Family Health Study.<sup>23</sup> The Generation Scotland  
29 Scottish Family Health Study (N > 24,000) constitutes an ancestrally comparable population-based cohort  
30 for comparison with the suicide decedents in Utah. To eliminate confounding arising from intra-dataset  
31 relatedness, only the 3,623 founders from the Generation Scotland dataset were used in analyses.  
32  
33  
34  
35  
36

37 *UK10K.* A total of 11,049 UK10K controls<sup>24</sup> were analyzed in wave 2 and GWAS analyses of  
38 both waves. This second control cohort is comprised of approximately 4000 genomes from the UK along  
39 with 6000 exomes from UK individuals with selected health phenotypes. We chose these data due to the  
40 extensive phenotyping and characterization of any medical conditions present, and to avoid choosing a  
41 cohort of entirely psychiatrically and medically healthy individuals. 4,000 highly phenotyped “super  
42 control” samples were supplied from the TwinsUK study from King’s College London and the Avon  
43 Longitudinal Study of Parents and Children. UK10K was a collaborative project to examine obesity,  
44 autism, schizophrenia, familial hypercholesterolemia, thyroid disorders, learning disabilities, ciliopathies,  
45 congenital heart disease, coloboma, neuromuscular disorders, and rare disorders including severe insulin  
46 resistance. Analyses of this control cohort included all diagnostic groups. While this inclusion may have  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 been conservative due to possible associations of chronic disease with suicide risk, determination of  
4 effects of specific disease states was beyond the scope of the current study. Genotyping and sequencing  
5 procedures for UK10K have been previously described<sup>24</sup> (<http://www.uk10k.org>) and all molecular genetic  
6 data from UK10K were filtered to the hard call variants present in our suicide death cohort prior to  
7 imputation of all cohorts simultaneously.  
8  
9  
10  
11

12 *1000 Genomes Reference Panel*. The CEU population from the 1000 Genomes Project<sup>25</sup> (1KG),  
13 which includes only Utah residents carefully screened for Northwestern European ancestry, was utilized  
14 as a model for excluding ancestrally discordant suicide and control samples. These CEU data were  
15 downloaded from the 1000 Genomes Project public repository. Unrelated individuals in the CEU provide a  
16 compelling, albeit small, ancestrally matched control resource (n = 99). A variety of candidate control  
17 samples were assessed via PCA for ancestral comparability to CEU and decedent data, with UK10K and  
18 Generation Scotland founder data representing the closest match.  
19  
20  
21  
22  
23  
24  
25

26 Utah controls would be an ideal match for the suicide cases, but as with many GWAS, local  
27 controls were not readily available at the sample size needed for GWAS. CEPH ancestry 1KG were a  
28 useful comparison group to assess the likelihood that UK controls were an appropriate match for the  
29 cases. In addition (and described in more detail below), we performed a Generation Scotland control-  
30 UK10K control GWAS and subsequently eliminated any SNPs from the case-control analysis that  
31 evidenced signal between the control cohorts. This was performed to minimize the possibility of false  
32 positives in the case control GWAS due to population and geographic stratification across cohorts.  
33  
34  
35  
36  
37  
38  
39  
40

### 41 **Genotyping and Quality Control**

42 Suicide cases were genotyped using Illumina Infinium PsychArray platform measuring 593,260 single  
43 nucleotide polymorphisms (SNPs). Generation Scotland samples were genotyped using Illumina  
44 OmniExpress SNP GWAS and exome chip, measuring 700,000 and 250,000 SNPs, respectively.<sup>23</sup>  
45 UK10K samples were whole genome sequenced<sup>24</sup> and variants were extracted to match the available  
46 QC'd hard-called variants in the suicide cases. Genotypes were subsequently imputed in all cases and  
47 controls (details of imputation are presented in Analytics, below). Both case and control datasets resulted  
48 from population-based ascertainment. Cryptic relatedness was modeled via the derivation of genomic  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 relatedness matrices. Genotyping quality control was performed using SNP clustering in  
4  
5 Illumina Genome Studio ([https://www.illumina.com/techniques/microarrays/array-data-analysis-](https://www.illumina.com/techniques/microarrays/array-data-analysis-experimental-design/genomestudio.html)  
6  
7 [experimental-design/genomestudio.html](https://www.illumina.com/techniques/microarrays/array-data-analysis-experimental-design/genomestudio.html)). SNPs were retained if the GenTrain score was  $> 0.5$  and the  
8  
9 Cluster separation score was  $> 0.4$ . SNPs were converted to HG19 plus strand. SNPs with  $>5\%$  missing  
10  
11 genotypes were removed. Samples with a call rate  $< 95\%$  were removed.

12  
13 Prior to case-control GWAS, a control-control GWAS was conducted (using the same methods  
14  
15 described in Data Analysis: Genome-Wide Association Testing, below) in order to detect differences  
16  
17 between control groups (all variants both pre- and post-imputation with GWAS control-control  $q$  values  
18  
19  $< .10$ ). For example, chromosome 4 TLR variants are often filtered from analyses involving Scottish  
20  
21 controls due to population stratification. 167 variants in the control-control comparison were then filtered  
22  
23 from subsequent case-control analyses. For the purposes of future meta-GWAS analyses and because  
24  
25 the MHC is relevant to psychiatric risk, we included MHC-associated filtered variants in a second version  
26  
27 of summary statistics.

## 30 **Data Analysis**

### 31 Principal Component Analysis (PCA)

32  
33 Supplementary Figure S1 shows 1000 Genomes superpopulations and suicide case/control samples,  
34  
35 both included and excluded, plotted by the top two principal components (PCs). Approximately 20% of the  
36  
37 population-based suicide cases had a significant degree of non-Northwestern European ancestry (chiefly  
38  
39 of admixed ancestry) and were excluded from analyses. The variation explained by top four PCs was  
40  
41 reduced 7.2-fold. The top four PCs explain 0.89% of variation before sample filtering and 0.12% of  
42  
43 variation after filtering if calculated on pruned genotypes. For adequate statistical power, we examined  
44  
45 only cases of Northern European ancestry. However, it is clear from Figure S1 (a) and (c) that the cohort  
46  
47 was comprised of multiple ancestries and that research on suicide death in non-European ancestries will  
48  
49 reflect an important step beyond this first study.

50  
51 PCA was performed on control, suicide, and 1000 Genomes cohorts after linkage disequilibrium  
52  
53 (LD) pruning at a 0.2 threshold. To exclude ancestrally heterogeneous samples, the top four principal  
54  
55 components (defined as those components which accounted for  $> 0.1\%$  of the genotype variance) were  
56  
57  
58  
59  
60

1  
2  
3 used to establish PC centroid limits centered around 1000 Genomes CEU data such that 99% of the CEU  
4 data fell within the limits. Only suicide and control samples also falling within these limits were considered  
5 ancestrally homogenous and were included in the association study. The ancestry PCA was performed  
6 using RaMWAS, a Bioconductor package written by our analytical team, which comprises a complete  
7 toolset for high dimensional genomic analyses.<sup>26,27</sup>  
8  
9  
10  
11  
12  
13

#### 14 Imputation

15  
16 European ancestry cases and controls were well-matched to 1000 Genomes CEPH. The Haplotype  
17 Reference Consortium is comprised in part by UK controls used here so we elected to impute genotypes  
18 based on the 1000 Genomes reference panel using minimac3<sup>28</sup> and Eagle.<sup>29</sup> SNPs were imputed jointly  
19 from a common SNP list based on variant overlap across case and control datasets. SNPs with  
20 ambiguous strand orientation, >5% missing calls, or Hardy-Weinberg equilibrium  $p < 0.001$  were  
21 excluded. SNPs with minor allele frequency below 0.01 or imputation  $R^2 < 0.5$  or average imputation call  
22 rate  $< 0.9$  were excluded after imputation. Genomic data were handled using PLINK.<sup>30,31</sup> Final GWAS  
23 analysis was performed on 7,519,308 variants passing quality control.  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33

#### 34 Genome-wide Association Testing

35  
36 A Linear Mixed Model (LMM) algorithm tested variant association with suicide death and provided follow-  
37 up examination of significant hits for linkage disequilibrium and gene set enrichment. GWAS were  
38 performed using GEMMA,<sup>32</sup> a computationally efficient and open-source LMM algorithm for GWAS that  
39 models population stratification remaining after PCA by use of genomic relatedness matrices. Sex was  
40 not included as a covariate in GWAS analyses due to the association of suicide with sex status at a ratio  
41 of approximately 4:1 males to females. GWAS with hard call-only and then with imputed data were  
42 examined separately to assess potential population stratification unique to our imputed GWAS. Prior to  
43 case-control GWAS, control-control GWAS was implemented to filter signal likely due to population  
44 stratification in the controls. To assess the likelihood that observed results may be due to chance, we  
45 assessed FDR at 5%.  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

### Gene and Gene Set Enrichment and Functional Mapping

SNP results from the GWAS were then mapped to genes within 1kb of the SNP. These genes were examined for gene set enrichment and LD using FUMA<sup>33</sup> and GREAT.<sup>34</sup> 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 non-coding 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 SNP-trait p-value <1.0x10<sup>-5</sup> in the overall (initial GWAS + replication) population. The most significant SNP from each independent locus is extracted.

### Polygenic Risk Scores, SNP Heritability ( $h^2$ ), and Genetic Correlations ( $r_G$ )

Discovery GWAS summary statistics for phenotypes were compiled to score each cohort for polygenic risk. A polygenic score (PS) for suicide death was derived using PRSice 2.0<sup>35</sup> and summary statistics from a 10-fold cross validation procedure to avoid overfitting. To elaborate, k-folds cross-validation is a well-established method<sup>36</sup> allowing a single dataset 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 current dataset. Of several thousand medical and psychological GWAS now available as training weights for calculating polygenic scores, only GWAS with N>10,000 individuals and >1,000 cases (or for population-based studies, adequate base rates) were selected for these analyses. These generally included the largest medical and psychiatric GWAS. When multiple versions of GWAS were available for the same phenotype (for example, neuroticism or depression), we selected the most comprehensive (see the online GWAS Atlas at <http://atlas.ctglab.nl/>). PRSice 2.0 was used to calculate individual PS for 59

1  
2  
3 phenotypes with estimated risk allele effect sizes for each discovery sample trait. A PS is traditionally  
4 calculated as a weighted sum score, where a score for an individual in the target sample is calculated by  
5 the summation of each SNP multiplied by the effect size of that SNP in the discovery GWAS. Based on  
6 the cross-disorder psychiatric genomics findings to date, we hypothesized significant positive prediction of  
7 suicide with PS for depressive symptoms, depressive disorders, behavioral disinhibition, schizophrenia,  
8 autism, loneliness, child IQ, alcohol use, and neuroticism.  
9  
10  
11  
12  
13

14 Linkage Disequilibrium Score regression (LDSC)<sup>37,38</sup> was used to calculate the observed scale  
15 common variant  $h^2$  using summary statistics from a logistic regression model with five ancestry covariates  
16 and pruning related samples at 0.05  $\hat{p}$  from IBD. LDSC was also used to calculate common variant  
17 molecular genetic correlations ( $r_G$ ) with psychiatric and medical phenotypes. Finally, we performed  
18 secondary analyses to characterize genetic predictors and clinical antecedents of suicide (Supplementary  
19 Methods S2) and mode of suicide death (Supplementary Methods S3). Specifically, in these secondary  
20 analyses, we conducted (1) epidemiological association tests between four sufficiently prevalent/powerful  
21 modes of death (i.e., gun, overdose, asphyxiation, and violent trauma) and 30 ICD-10 derived clinical  
22 antecedents, and (2) association tests the suicide polygenic score with modes of death, adjusting for five  
23 ancestry covariates in multivariate regressions.  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33

### 34 35 Sex Differences

36 As suicide rates and modes of suicide death are characterized by substantial sex differences,<sup>19,20</sup> we  
37 performed secondary epidemiological and genomic analyses to characterize sex differences in mode of  
38 death and clinical antecedents. These sex stratified analyses mirrored the full sample analyses described  
39 above, including (1) sex-stratified epidemiological association tests between four sufficiently  
40 prevalent/powerful modes of death (i.e., gun, overdose, asphyxiation, and violent trauma) and 30 ICD-10  
41 derived clinical antecedents, and (2) sex stratified association tests the suicide polygenic score with  
42 modes of death, adjusting for five ancestry covariates in multivariate regressions. We constrained these  
43 exploratory analyses to only those medical diagnoses with frequencies high enough in either females or  
44 males to provide decent power for testing and report false discovery rate (FDR) corrected p-values;  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 nonetheless, it is worth noting that power for these secondary analyses of sex differences was limited by  
4  
5 N, which was restricted to cases-only and stratified by sex and mode of death.  
6  
7

## 8 9 **RESULTS**

### 10 **Genome-wide Association: Positional Evidence**

11  
12 A total of six variants from two loci met genome-wide criteria for statistical association with suicide death  
13  
14 ( $p < 5 \times 10^{-8}$ ). An additional 52 variants were nominally significant at  $q < 0.1$  and mapped to 22 genes ( $\lambda =$   
15  
16 1.015, Figure 1 and Table 1).<sup>39,40</sup> All results on the full cohort are derived from analyses adjusting for  
17  
18 effects of ancestry. Genes associated with top genomic regions are presented in Supplementary Table  
19  
20 S1. The significant chromosome 13 and 15 regions were supported by additional positive results that  
21  
22 were suggestive but below threshold. The large number of signals in the SNP-based tests prompted  
23  
24 quality control analyses varying the degree of LD pruning prior to PCA for the purposes of sensitivity  
25  
26 analysis. Results and respective  $\lambda$ 's were consistent across these analyses. Supplementary Figures S2-  
27  
28 S8 present additional plots of the top signals in each of eight regions.  
29  
30

### 31 **Gene-based Results and Pathway Functional Enrichment Tests**

32  
33 Mapping top positional SNP hits to genes suggested 22 gene associations, including chromosome 13  
34  
35 genes, Daschund family transcription factor 1 (*DACH1*), Ubiquitin-protein ligase protein (*UBE3A*), and  
36  
37 Kelch-like family member 1 (*KLHL1*) on chromosome 15 (Supplementary Table S1). Gene-based analysis  
38  
39 using MAGMA (FUMA<sup>1</sup>) identified an additional ten genes significantly associated ( $q < 0.1$ ) with suicide  
40  
41 death (Figure 2; Supplementary Table S1). GO pathway results included enrichment of histone  
42  
43 modification sites *SETD6*, *COPR5*, *GATAD2A*. Full gene and Gene Ontology (GO,  
44  
45 <http://www.geneontology.org/>) pathway enrichment results are presented in Supplementary Tables S2-  
46  
47 S3. Genes showing psychiatric associations are in green (Supplementary Table S2).  
48

49  
50 In addition to functional pathways, FUMA analyses indicated significant enrichment for  
51  
52 schizophrenia ( $p = 1 \times 10^{-11}$ ) and bipolar disorder ( $p = 1 \times 10^{-17}$ ) in the GWAS Catalog  
53  
54 (<https://www.ebi.ac.uk/gwas/>; Table S1). IW-scoring in SNP-Nexus<sup>41</sup> suggested regulatory functional  
55  
56 significance for one SNP (chr13:71553748:C/T). Ten of the implicated genes from positional or gene-  
57  
58  
59  
60

1  
2  
3 based testing have evidenced genome-wide significant differential gene expression in postmortem brain  
4 analyses in schizophrenia, autism, or bipolar disorder (FDR<0.1; PsychENCODE Consortium,  
5 Supplementary Table S4).<sup>42</sup>  
6  
7  
8  
9

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

11  
12 In European ancestry training and test samples comprising independent case and control cohorts, and  
13 accounting for 20 ancestry covariates, suicide PS significantly predicted suicide death case status.  
14  
15 Suicide waves 1 and 2 comprise 1,321 and 2,092 suicide cases, respectively, and were pruned for  
16 relatedness within waves prior to training and testing analyses. These case-control predictions are plotted  
17 across  $p$ -value thresholds in Figure 3.  
18  
19  
20  
21

22 The LDSC common variant  $h^2$  estimate, based on the summary statistics from a logistic GWAS  
23 with five ancestry covariates and pruning to remove related samples, was 0.2463, SE = 0.0356. The  $h^2$   
24 estimate on the liability scale = 0.1567, SE = 0.0218 when considering the .2% reported population  
25 prevalence of completed suicide in Utah.  $\lambda$  in the model was 1.20,  $\lambda_{1000} = 1.04$ . These  $\lambda$ s remained stable  
26 with follow-up testing of 20 PCs and with increased pruning. The suicide death cases differed significantly  
27 from the two UK control groups on PS of phenotypes relevant to suicide death. These differences were in  
28 the expected directions. Original discovery GWAS for all PS phenotypes were filtered to exclude any that  
29 used these control cohorts (Supplementary Table S5).  
30  
31  
32  
33  
34  
35  
36

37 Consistent with hypotheses, significant PS elevations included alcohol use, autism spectrum  
38 disorder, child IQ, depressive symptoms, disinhibition, major depression, loneliness, and schizophrenia.  
39 The largest effect sizes were observed for PS for behavioral disinhibition, major depressive disorder, and  
40 schizophrenia (Figure 4). LD Hub<sup>37</sup> provided estimates of SNP-based shared genetic covariance for  
41 several phenotypes (Supplementary Table S6), though none of these SNP-based genetic correlations  
42 reached statistical significance. Additionally, we disaggregated suicide by mode of death into four  
43 categories (see Supplementary Methods S2): gun, overdose, asphyxiation, and violent trauma. We  
44 epidemiologically characterized these groups by testing associations with 30 ICD-10 derived clinical  
45 antecedents (Supplementary Figure S9; Supplementary Tables S7-S9). Additionally, we conducted PS  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 association testing with mode of death in all cases. No associations met multiple testing adjusted  
4  
5 significance criteria ( $q < 0.1$ ; Supplementary Figure S10; Supplementary Tables S10-S12).  
6  
7

### 8 9 **Sex Differences**

10 Epidemiological analyses of sex differences in electronic health records data indicated that suicide cases  
11 of both sexes evidenced clinical diagnostic clusters of 1) internalizing-trauma-cluster B psychiatric  
12 disorders and 2) metabolic-cardiovascular-obesity medical disorders (Supplementary Tables S13-  
13 S18). Female cases were observed to have a higher overall number of diagnoses relative to males, which  
14 could reflect increased severity in females, decreased severity in males, decreased likelihood of males  
15 receiving diagnosis, and/or decreased help-seeking in males. This is broadly consistent with observed  
16 higher relative prevalence rates of gun-related death in males and overdose death in females. All  
17 associations of ICD and PS with mode of death are presented for females and males separately in  
18 Supplementary Figures S11-S14. All PS analyses include ancestry covariates. All corresponding statistics  
19 are reported in Supplementary Tables S19-S24. No sex-stratified polygenic scores findings met multiple  
20 testing adjusted significance criteria ( $q < 0.1$ ).  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33

### 34 **DISCUSSION**

35 Results from this analysis, the first adequately powered comprehensive genomic study of suicide death,  
36 yield several insights and suggest important applications for future extension. The GWAS of suicide death  
37 identified two genome-wide significant loci on chromosomes 13 and 15. The significant SNP-based  
38 heritability estimate of suicide death in this GWAS (25%, or 16% on the liability scale) is greater than  
39 those previously reported for suicide ideation/attempt (5%)<sup>14,15</sup> and attempt within psychiatric diagnosis  
40 (3-10%)<sup>43</sup> and is closer to previous estimates reported for suicide death (35-48%).<sup>13</sup> Fully half of the  
41 genes implicated by our results overlap with schizophrenia results from the GWAS Catalog ( $p = 1 \times 10^{-11}$ ),  
42 and two of these 11 genes have prior associations with risk of suicidal behavior (*HS3ST3B*, *NCAN*; for  
43 relevant literature see Supplementary Table S25; allele frequencies and effect sizes by cohort in our  
44 study are presented in Table S26).  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 All but four of these schizophrenia-associated genes also drove a significant GWAS Catalog  
4 association with bipolar disorder ( $p=1 \times 10^{-17}$ ), supporting a hypothesis of genetic liability to more non-  
5 specific psychiatric illness. Of particular interest, the region on chromosome 19p13.11 in our study has  
6 evidence from the most recent, well-powered studies of both bipolar disorder<sup>44</sup> and schizophrenia.<sup>45</sup> In  
7 addition, our region on chromosome 16q21 has evidence not only from several large GWAS of  
8 schizophrenia, but also from a large meta-analysis of autism spectrum disorder.<sup>45-50</sup>

9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
Genome-wide significant SNPs on chromosome 13 (Table 1) are all intergenic and vary with  
respect to combined annotation dependent depletion (CADD) score, an index for scoring the  
deleteriousness of single nucleotide variants as well as insertion/deletions variants in the human genome.  
Notably the variant with the strongest GWAS effect size, SNV rs34399104, reaches a CADD score of  
almost 20, indicating that it is in the top 1% of deleterious variants on the genome.

Direct comparison of this GWAS with 110 effect sizes reported in previous literature<sup>12-15,43,51</sup>  
identified just one nominally significant variant in the largest GWAS of suicidal behavior,<sup>43</sup> rs4870888,  
nearest to *FER1L6* and *FER1I6-AS2*. Table S27 provides direct comparison of the previously reported  
effect sizes with those in this GWAS. SNV rs7989250 (chr 13) was reported in the UK Biobank GWAS for  
ordinal suicidality and reached a  $p < .05$  threshold for replication in the Otsuka et al. suicide GWAS, but  
did not reach threshold in this GWAS (Table S27).

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

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 (Figure S15 presents  
additional PRS plots). These results were consistent with expectations, as strong associations in cases

1  
2  
3 relative to both control cohorts were with behavioral disinhibition and major depressive disorder—  
4  
5 arguably the two most critical antecedents of suicide death.<sup>52</sup> More generally, these results are consistent  
6  
7 with the emerging consensus that genetic correlations across psychiatric diagnoses and traits are  
8  
9 substantial.<sup>53</sup> Future research with an even larger suicide death cohort will be needed to determine the  
10  
11 ultimate cause of these PS associations with suicide death, with leading hypotheses arguing for the  
12  
13 presence of a “p-factor” capturing general liability to psychopathology,<sup>54</sup> contrasting those arguing for  
14  
15 pleiotropic effects across a small set of psychiatric disorder clusters.<sup>55</sup>

### 16 *Limitations*

17  
18 Several limitations should be noted. First, it is possible that population stratification or platform  
19  
20 confounds stemming from convenience controls could lead to false positives. To address this risk, we  
21  
22 employed very conservative filters of both loci and subjects to minimize false positives. Values of  
23  
24  $\lambda_{GC} < 1.05$  are generally considered benign, although inflation in  $\lambda_{GC}$  is proportional to sample size.<sup>56</sup> Both  
25  
26 Generation Scotland and UK10k founder controls were extensively assessed for ancestral comparability  
27  
28 to both 1000 Genomes Utah CEU<sup>25</sup> and Utah European ancestry suicides, both matching extremely well.  
29  
30 Incorporating two separate control cohorts (and formal analytical comparison of these cohorts to the  
31  
32 corresponding 1KG populations) also allowed us to detect and filter out loci related to platform-specific  
33  
34 technical artifacts, and subtle variation in population structure, prior to case-control analyses. Table S28  
35  
36 presents results of tests for differences in loadings for up to 200 PCs by cohort and by case-control  
37  
38 status. Before and after imputation, we excluded any and all variants generating q-values  $< .1$  in a control-  
39  
40 control GWAS. Additionally, employing ancestral PC Mahalanobis (i.e., centroid) distance pruning, we  
41  
42 have leveraged the fact the CEU 1KG subjects are largely ancestrally homogenous with our case  
43  
44 samples to restrict our controls (and cases) to a highly ancestrally homogenous group.

45  
46 It is also possible that the use of mixed healthy controls and controls with common medical  
47  
48 conditions, as in UK10K, could exert a mild bias on effect estimates toward the null, which we do not view  
49  
50 as highly problematic, given that the direction of the bias effectively reduces the risk of false positive  
51  
52 findings. In the suicide case cohort, rates of common medical conditions are much higher than population  
53  
54 base rates. However, in stratified control analyses, effect sizes specific to cases vs. UK10K were elevated  
55  
56 relative to Generation Scotland, which is inconsistent with potential confounding of signal due to other  
57  
58  
59  
60

1  
2  
3 disorders. Furthermore, effect sizes for PS for other disorders, across 1000 p-value thresholds, were far  
4 greater between the cases and controls than between the control groups, wherein differences were  
5 negligible, and distributions of PS often overlapped. None of the implicated variants were significantly  
6 different in AF between control groups in the control-control GWAS. However, it should be noted that AF  
7 is relatively low for the top variants across all cohorts.  
8  
9  
10  
11

12 Another limitation relates to electronic medical record data, in which prevalence for disorders is  
13 always confounded with age, and in a number of ways. For example, missing diagnosis could be more or  
14 less likely in a given decade during which a case was assessed. Younger ages are more likely to have  
15 complete electronic medical record information yet be less likely to be diagnosed with co-morbid medical  
16 conditions, schizophrenia, or personality disorders.  
17  
18  
19  
20  
21

22 Importantly, in this study there was insufficient diversity to examine effect sizes for non-European  
23 ancestries. We expect a sufficient number of cases for a Mexican-American ancestry GWAS within five  
24 years, and we will prioritize diverse ancestry groups in order to minimize potential health disparities  
25 stemming from homogeneous summary statistics.<sup>57</sup>  
26  
27  
28  
29

### 30 *Future Directions*

31 A recent review of suicide prediction models has indicated that positive predictive values for  
32 suicide attempt are quite high (.9), but positive predictive values for suicide death continue to hover near  
33 zero.<sup>58,59</sup> In this study, we have discovered multiple genome-wide significant loci/genes, strong polygenic  
34 signal, and significantly increased genetic risks for behavioral disinhibition, major depression, depressive  
35 symptoms, autism spectrum disorder, psychosis, and alcohol use disorder in cases. Future modeling of  
36 multiple polygenic risks and phenotypic risk factors may help isolate important moderators of risk and  
37 improve objective risk measures of suicide death. Importantly, ethical challenges associated with  
38 predictive models of suicide death are significant, and must be addressed proactively across psychiatric  
39 and medical domains.<sup>6,60</sup>  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## REFERENCES

1. Turecki, G. & Brent, D.A. Suicide and suicidal behaviour. *Lancet (London, England)* **387**, 1227-1239 (2016).
2. Hedegaard H., C.S., Warner M. Suicide mortality in the United States, 1999–2017. in *National Center for Health Statistics Data Brief* Vol. 330 (CDC/NCHS, Hyattsville, MD, 2018).
3. Kochanek, K., Murphy, S., Xu, J. & Tejada-Vera, B. Deaths: Final Data for 2014. *National Vital Statistics Reports* **65**, 1-121 (2016).
4. Murphy, S.L., Mathews, T.J., Martin, J.A., Minkovitz, C.S. & Strobino, D.M. Annual Summary of Vital Statistics: 2013-2014. *Pediatrics* **139**(2017).
5. Case, A. & Deaton, A. Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century. *Proceedings of the National Academy of Sciences* **112**, 15078 (2015).
6. Barak-Corren, Y. *et al.* Predicting Suicidal Behavior From Longitudinal Electronic Health Records. *American Journal of Psychiatry* **174**, 154-162 (2017).
7. Franklin, J.C. *et al.* Risk Factors for Suicidal Thoughts and Behaviors: A Meta-Analysis of 50 Years of Research. *Psychological Bulletin* **143**, 187-U121 (2017).
8. McGuffin, P., Marusic, A. & Farmer, A. What can psychiatric genetics offer suicidology? *Crisis* **22**, 61-5 (2001).
9. Mullins, N. *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* **165b**, 428-37 (2014).
10. Turecki, G. Suicidal behavior: is there a genetic predisposition? *Bipolar Disord* **3**, 335-49 (2001).
11. Coon, H. *et al.* Genome-wide significant regions in 43 Utah high-risk families implicate multiple genes involved in risk for completed suicide. *Mol Psychiatry* (2018).
12. Galfalvy, H. *et al.* A genome-wide association study of suicidal behavior. *Am J Med Genet B Neuropsychiatr Genet* **168**, 557-63 (2015).
13. Otsuka, I. *et al.* Genome-wide association studies identify polygenic effects for completed suicide in the Japanese population. *Neuropsychopharmacology* **44**, 2119-2124 (2019).
14. Strawbridge, R.J. *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* **41**, 517-525 (2019).
15. Erlangsen, A. *et al.* Genetics of suicide attempts in individuals with and without mental disorders: a population-based genome-wide association study. *Mol Psychiatry* (2018).
16. O'Brien, E. *et al.* Familial mortality in the Utah population database: characterizing a human aging phenotype. **62**, 803-812 (2007).
17. Kerber, R., O'Brien, E., Smith, K.R. & Mineau, G.P. Familial aggregation of elderly cause-specific mortality: analysis of extended pedigrees in Utah, 1904–2002. in *Kinship and demographic behavior in the past* 243-258 (Springer, 2008).
18. World Health, O. ICD-10 : international statistical classification of diseases and related health problems : tenth revision. 2nd ed edn (World Health Organization, Geneva, 2004).
19. Schrijvers, D.L., Bollen, J. & Sabbe, B.G.J.J.o.a.d. The gender paradox in suicidal behavior and its impact on the suicidal process. **138**, 19-26 (2012).
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. *Social Psychiatry and Psychiatric Epidemiology* **39**, 113-120 (2004).
21. Utah Medical Examiner Act. in 26-4 ([https://le.utah.gov/xcode/Title26/Chapter4/C26-4\\_1800010118000101.pdf](https://le.utah.gov/xcode/Title26/Chapter4/C26-4_1800010118000101.pdf), United States, 2016).
22. Jorde, L.B. Inbreeding in the Utah Mormons: an evaluation of estimates based on pedigrees, isonymy, and migration matrices. *Ann Hum Genet* **53**, 339-55 (1989).
23. Smith, B.H. *et al.* Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. *BMC Medical Genetics* **7**, 74 (2006).
24. Walter, K. *et al.* The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82-90 (2015).

- 1
- 2
- 3 25. The Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**,
- 4 68 (2015).
- 5 26. Shabalin, A.A. *et al.* RaMWAS: fast methylome-wide association study pipeline for enrichment
- 6 platforms. *Bioinformatics* **34**, 2283-2285 (2018).
- 7 27. Shabalin, A., Clark, S., Hattab, M., Aberg, K. & van den Oord, E. *RaMWAS: Fast Methylome-*
- 8 *Wide Association Study Pipeline for Enrichment Platforms* . R package version 1.2.0 edn (2017).
- 9 28. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-
- 10 1287 (2016).
- 11 29. Loh, P.R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat*
- 12 *Genet* **48**, 1443-1448 (2016).
- 13 30. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets.
- 14 *Gigascience* **4**, 7 (2015).
- 15 31. Purcell, S., Chang, C., Sinai, N.-N.L.o.B.M.P.L.a.M. & School of Medicine. *PLINK 1.9 beta.*
- 16 (2017).
- 17 32. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies.
- 18 *Nat Genet* **44**, 821-4 (2012).
- 19 33. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and
- 20 annotation of genetic associations with FUMA. *Nature Communications* **8**, 1826 (2017).
- 21 34. McLean, C.Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat*
- 22 *Biotechnol* **28**, 495-501 (2010).
- 23 35. Euesden, J., Lewis, C.M. & O'Reilly, P.F. PRSice: polygenic risk score software. *Bioinformatics*
- 24 **31**, 1466-1468 (2014).
- 25 36. Hastie, T., Tibshirani, R. & Friedman, J.H. *The Elements of Statistical Learning: Data Mining,*
- 26 *Inference, and Prediction*, (Springer, 2009).
- 27 37. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression
- 28 that maximizes the potential of summary level GWAS data for SNP heritability and genetic
- 29 correlation analysis. *Bioinformatics* **33**, 272-279 (2017).
- 30 38. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in
- 31 genome-wide association studies. *Nature genetics* **47**, 291 (2015).
- 32 39. Narum, S.R. Beyond Bonferroni: less conservative analyses for conservation genetics.
- 33 *Conservation genetics* **7**, 783-787 (2006).
- 34 40. Hochberg, Y. & Benjamini, Y. More powerful procedures for multiple significance testing. *Stat*
- 35 *Med* **9**, 811-8 (1990).
- 36 41. Dayem Ullah, A.Z. *et al.* SNPnexus: assessing the functional relevance of genetic variation to
- 37 facilitate the promise of precision medicine. *Nucleic acids research* **46**, W109-W113 (2018).
- 38 42. Gandal, M.J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and
- 39 bipolar disorder. *Science* **362**(2018).
- 40 43. Mullins, N. *et al.* GWAS of Suicide Attempt in Psychiatric Disorders and Association With Major
- 41 Depression Polygenic Risk Scores. *Am J Psychiatry* **176**, 651-660 (2019).
- 42 44. Stahl, E.A. *et al.* Genome-wide association study identifies 30 loci associated with bipolar
- 43 disorder. *Nat Genet* **51**, 793-803 (2019).
- 44 45. Pardinias, A.F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and
- 45 in regions under strong background selection. *Nat Genet* **50**, 381-389 (2018).
- 46 46. Autism Spectrum Disorders Working Group of The Psychiatric Genomics, C. Meta-analysis of
- 47 GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at
- 48 10q24.32 and a significant overlap with schizophrenia. *Mol Autism* **8**, 21 (2017).
- 49 47. Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108
- 50 schizophrenia-associated genetic loci. *Nature* **511**, 421-427 (2014).
- 51 48. Ikeda, M. *et al.* Genome-Wide Association Study Detected Novel Susceptibility Genes for
- 52 Schizophrenia and Shared Trans-Populations/Diseases Genetic Effect. *Schizophr Bull* **45**, 824-
- 53 834 (2019).
- 54 49. Periyasamy, S. *et al.* Association of Schizophrenia Risk With Disordered Niacin Metabolism in an
- 55 Indian Genome-wide Association Study. *JAMA Psychiatry* (2019).
- 56 50. Goes, F.S. *et al.* Genome-wide association study of schizophrenia in Ashkenazi Jews. *Am J Med*
- 57 *Genet B Neuropsychiatr Genet* **168**, 649-59 (2015).
- 58
- 59
- 60

- 1  
2  
3 51. Ruderfer, D.M. *et al.* Significant shared heritability underlies suicide attempt and clinically  
4 predicted probability of attempting suicide. *Mol Psychiatry* (2019).
- 5 52. Nock, M.K. *et al.* Prevalence, correlates, and treatment of lifetime suicidal behavior among  
6 adolescents: results from the National Comorbidity Survey Replication Adolescent Supplement.  
7 *JAMA Psychiatry* **70**, 300-10 (2013).
- 8 53. Anttila, V. *et al.* Analysis of shared heritability in common disorders of the brain. **360**, eaap8757  
9 (2018).
- 10 54. Caspi, A. *et al.* The p factor: one general psychopathology factor in the structure of psychiatric  
11 disorders? **2**, 119-137 (2014).
- 12 55. Lee, P.H. *et al.* Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight  
13 Psychiatric Disorders. *Cell* **179**, 1469-1482.e11 (2019).
- 14 56. Price, A.L., Zaitlen, N.A., Reich, D. & Patterson, N. New approaches to population stratification in  
15 genome-wide association studies. *Nat Rev Genet* **11**, 459-63 (2010).
- 16 57. Martin, A.R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities.  
17 *Nature Genetics* **51**, 584-591 (2019).
- 18 58. Belsher, B.E., Smolenski, D.J. & Pruitt, L.D. Positive Predictive Values and Potential Success of  
19 Suicide Prediction Models-Reply. *JAMA Psychiatry* (2019).
- 20 59. Belsher, B.E. *et al.* Prediction Models for Suicide Attempts and Deaths: A Systematic Review and  
21 Simulation. *JAMA Psychiatry* **76**, 642-651 (2019).
- 22 60. Adkins, D.E. Machine Learning and Electronic Health Records: A Paradigm Shift. *American*  
23 *Journal of Psychiatry* **174**, 93-94 (2017).
- 24 61. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic  
25 variants. *Nature genetics* **46**, 310 (2014).
- 26 62. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J. & Kircher, M. CADD: predicting the  
27 deleteriousness of variants throughout the human genome. *Nucleic acids research* **47**, D886-  
28 D894 (2018).
- 29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

TABLES

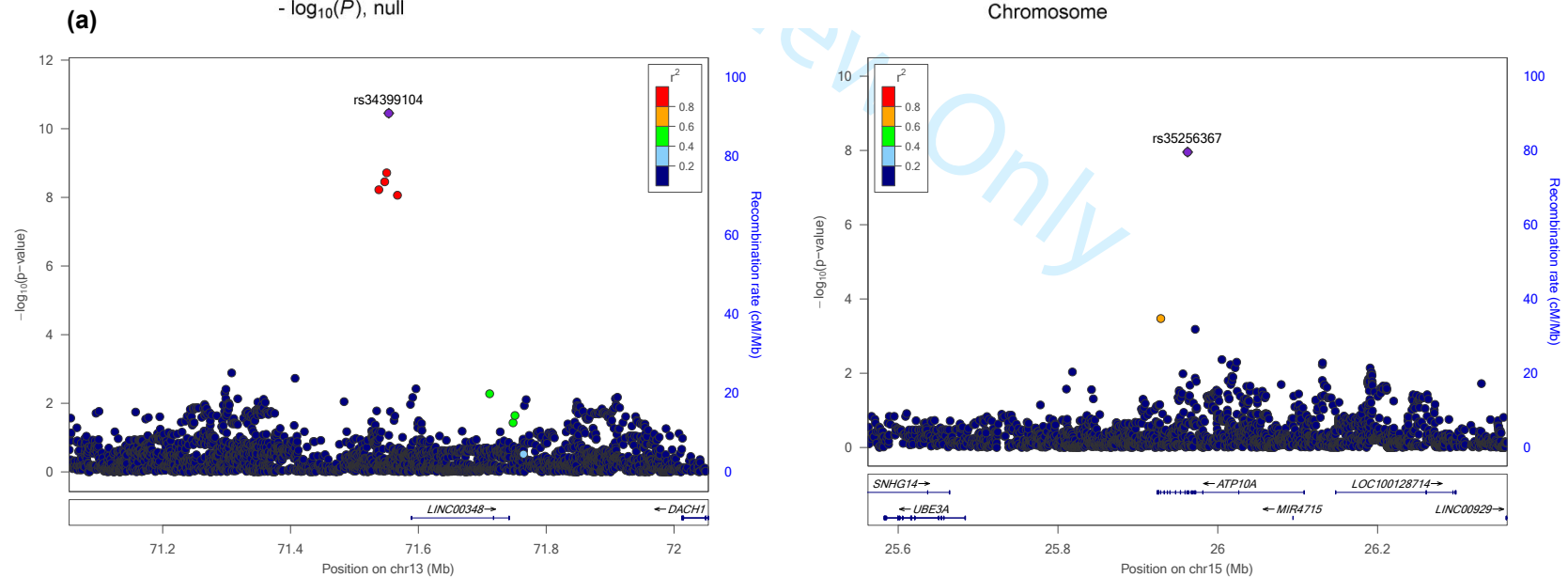
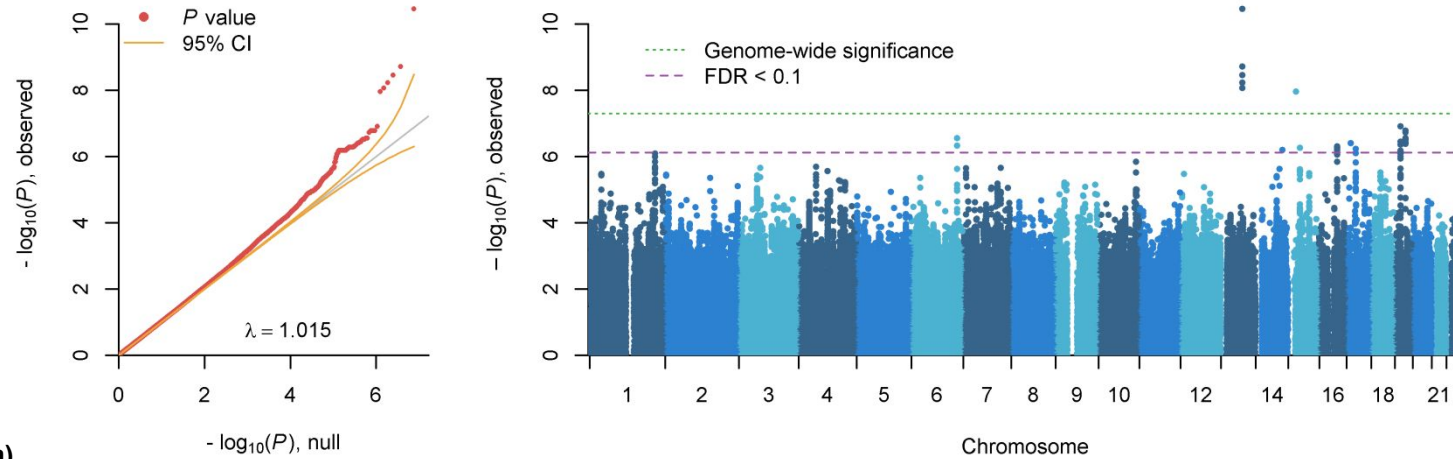
Table 1. Genome-wide significant loci from GWAS of death by suicide.

Chr	SNP	Imputation R <sup>2</sup>	A1	A2	AF Cases	AF Controls	OR	95% CI	p value	FDR	Nearest Gene	Function	CADD
13	rs34399104	0.75	T	C	0.027	0.015	1.769	1.48–2.11	3.54x10 <sup>-11</sup>	2.7x10 <sup>-4</sup>	LINC00348	Intergenic	19.86
13	rs35518298	0.81	T	C	0.024	0.014	1.726	1.43–2.07	1.92x10 <sup>-9</sup>	7.2x10 <sup>-3</sup>	LINC00348	Intergenic	0.33
13	rs34053895	0.81	A	C	0.024	0.014	1.718	1.42–2.06	3.51x10 <sup>-9</sup>	8.7x10 <sup>-3</sup>	LINC00348	Intergenic	0.39
13	rs35502061	0.72	G	A	0.024	0.014	1.703	1.41–2.05	5.97x10 <sup>-9</sup>	.011	SOGA2P1	Intergenic	2.69
13	rs66828456	0.74	A	C	0.027	0.015	1.707	1.42–2.04	8.63x10 <sup>-9</sup>	.013	LINC00348	Intergenic	4.24
15	rs35256367	0.50	G	A	0.023	0.014	1.605	1.43–2.07	1.10x10 <sup>-8</sup>	.014	ATP10A	Intronic	4.41

Note: Chr = chromosome, SNP = single nucleotide polymorphism, A1, A2 = alleles 1 (minor) and 2, AF = allele 1 frequency, OR = odds ratio, SE = standard error, CADD<sup>61,62</sup> = combined annotation dependent depletion score. For scale, note that a CADD score of 20 means that a variant is among the top 1% of deleterious variants in the human genome. All variants presented in the table were imputed.

## FIGURES

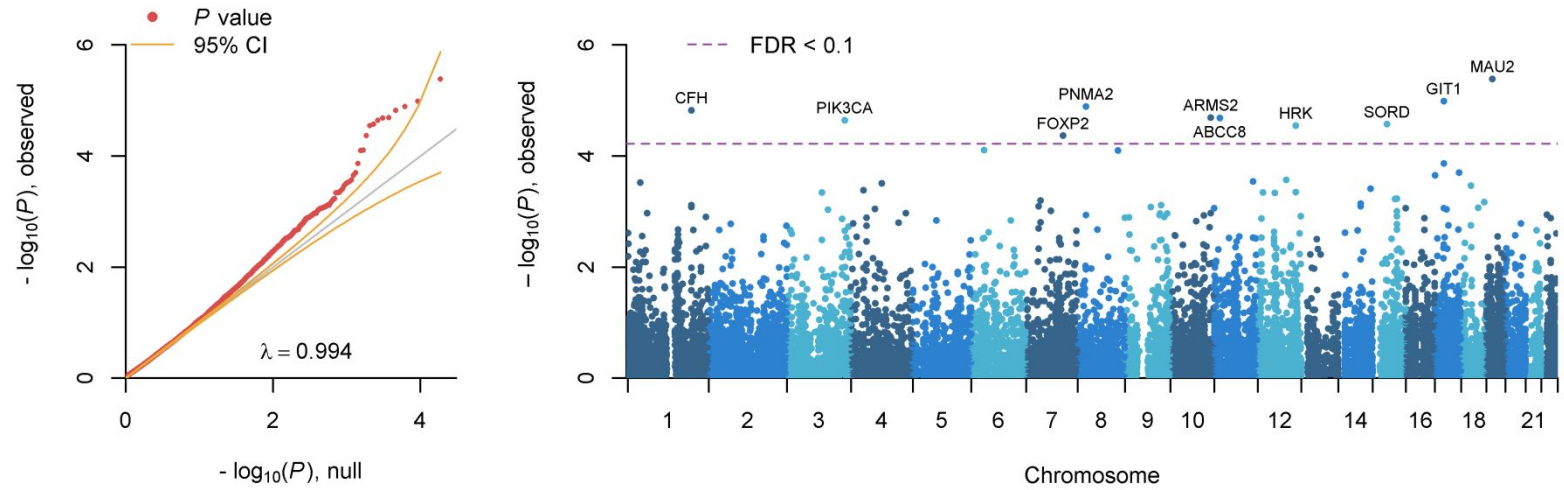
**Figure 1. GWAS Results.** (a) QQ and Manhattan plots from GWAS of suicide death. Y-axes for both plots reflect observed p-values. The x-axis on the qq-plot is the number of significant p-values expected under  $H_0$ , and the x-axis on the Manhattan plot maps each chromosome. The purple dashed line indicates FDR corrected nominal statistical significance, the green dotted line representing the threshold for genome-wide significance with Bonferroni correction. 57 SNPs met threshold for nominal significance and 6 met genome-wide significance. (b) Regional plots of genome-wide significant loci on chromosomes 13 and 15.



(b)

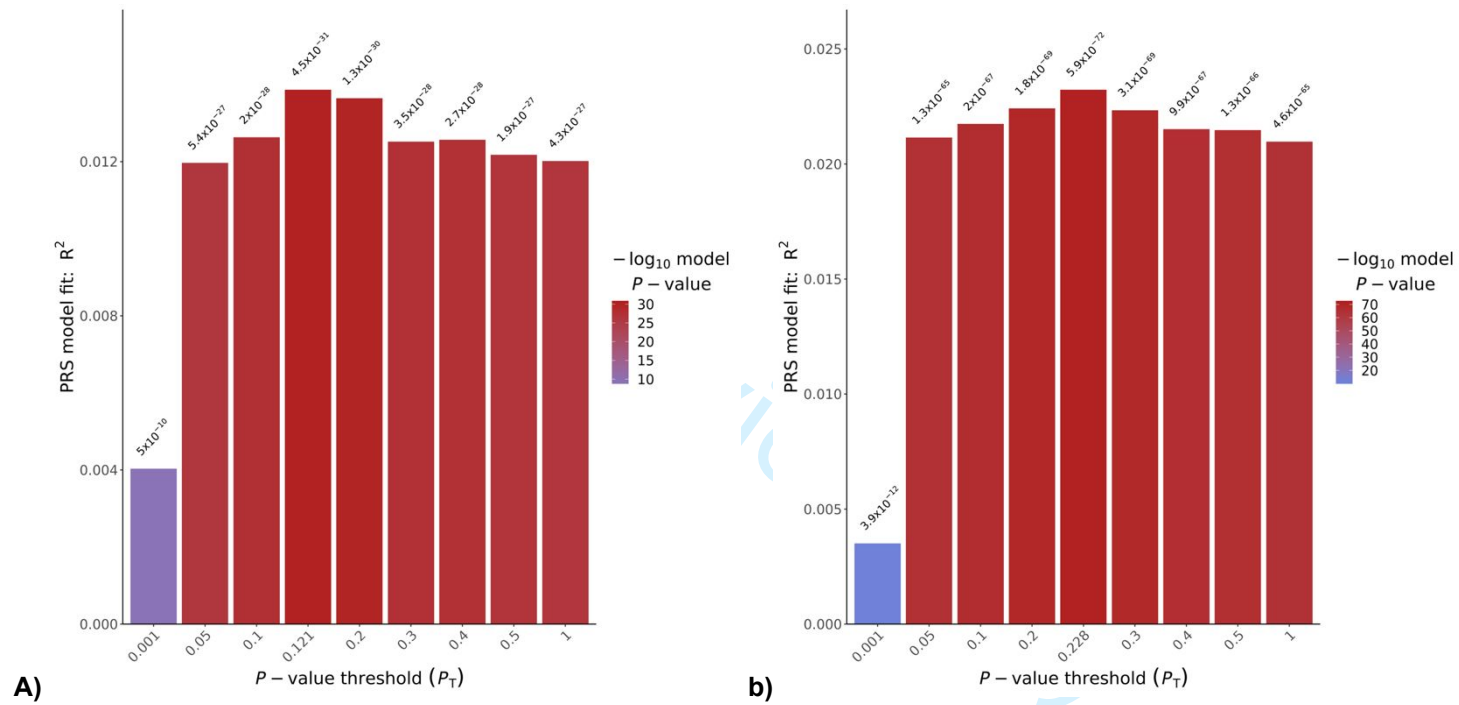


**Figure 2. GWAS Gene-Based Results.** FUMA gene-based test results: qq plot and Manhattan plot of >18,000 genes. Y-axes for both plots are identical and reflect observed p-values. The x-axis on the qq-plot is the number of significant p-values expected under  $H_0$ , and the x-axis on the Manhattan plot maps each chromosome. The purple dashed line indicates threshold for FDR-corrected nominal statistical significance; 10 genes met this threshold for nominal significance.

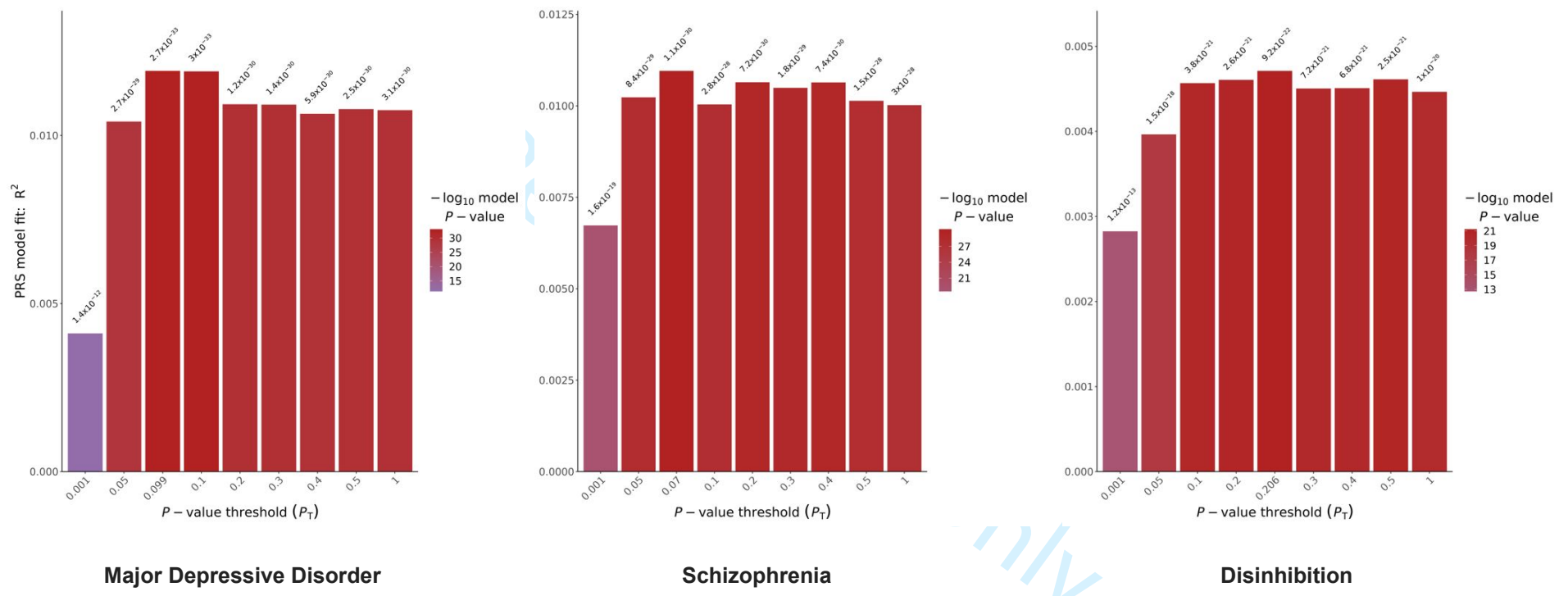


View Only

**Figure 3. Cross Validation of Suicide Polygenic Case-Control Prediction.** Polygenic prediction of suicide death case status across two independent cohorts of cases and controls. Training GWAS summary statistics are used to score the test set for suicide polygenic risk. P-value thresholds are plotted on the x-axis from 0.1-1.0, reflecting the top 10% to 100% of the common variants from the training GWAS. Plots present model fit when **a)** predicting case-control status with wave 1 suicide vs. Generation Scotland summary statistics and **b)** predicting with wave 2 suicide vs. UK10K summary statistics. Scores are adjusted for 20 PCs. Summary data for the best fitting and 1.0 p-value threshold models are presented in Table S29.



**Figure 4. Notable Elevations of Psychiatric Polygenic Risk in Suicide Cases.** P-value thresholds are plotted on the x-axis. Scores are adjusted for 20 PCs. 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.



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47

**SUPPLEMENTARY MATERIAL****Supplementary Methods**

- S1. International Classification of Diseases (ICD-10) Code Ascertainment
- S2. Clinical Diagnostic Associations with Mode of Death

**Supplementary Tables**

- S1. Genes Associated with Top Genomic Regions, FUMA Gene-Based Analysis
- S2-S3. GREAT, GWAS Catalog, SNP Nexus Results
- S4. Genome-Wide Significant Differential Gene Expression in PsychENCODE Postmortem Brain in Schizophrenia, Autism, or Bipolar Disorder
- S5. Polygenic Score (PS) Discovery GWAS Summary Statistic Citation Table
- S6. LD Hub Genetic Correlation Results
- S7-9. All Case Diagnostic Antecedents (ICD) by Mode of Suicide (MD) Association Summary Statistics
- S10-12. All Case PS-MD Summary Statistics
- S13-18. Separate Female and Male ICD-MD Summary Statistics
- S19-S24. Separate Female and Male PS-MD Summary Statistics
- S25. Gene Literature (brief)
- S26. Allele Frequencies and Summary Statistics for Top SNPs across Cohorts
- S27. Cross-Study Comparison of Effect Sizes
- S28. Tests for Differences in up to 200 PCs by Cohort and by Case-Control Status
- S29. Summary Statistics for the Best Fitting and 1.0 *P*-Value Threshold Models from Figures 3 & 4.

**Supplementary Figures**

- S1. PCA of 1KG super-populations, cases, controls, and excluded samples
- S2. Regional Plot of Chromosome 6 Region
- S3. Regional Plot of Chromosome 14 Region
- S4. Regional Plot of Chromosome 15 Region
- S5. Regional Plot of Chromosome 16 Region
- S6. Regional Plot of Chromosome 17 Region
- S7. Regional Plot of Chromosome 17 Region
- S8. Regional Plot of Chromosome 19 Region
- S9. Diagnostic Antecedents (ICD) of Specific Mode of Suicide (MD) Plot: All Suicide Deaths
- S10. PS x MD Plot: All Suicide Deaths
- S11. ICD x MD Plot: Female Suicide Deaths > 25 Years of Age
- S12. ICD x MD Plot: Male Suicide Deaths > 25 Years of Age
- S13. PS X MD Plot: Female Suicide Deaths Plot
- S14. PS X MD Plot: Male Suicide Deaths Plot
- S15. Plots of Additional Case-Control Differences in Polygenic Risks