1 **Genome-wide association study of suicide death and polygenic prediction of clinical antecedents**
2
3
4
5
6 Word Count:
7 Summary: 142
8 Introduction: 414
9 Results: 992
10 Discussion: 303
11 Methods: 2472
12 Tables/Figures: 1/5
13 Supplemental Tables/Figures:38/18
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

Summary

Suicide death is a preventable yet growing worldwide health crisis. To date, genetic discovery efforts for the extreme phenotype of suicide death have been virtually nonexistent, as no sizeable cohorts have been available for study. We have conducted the first GWAS of suicide death, with 3,413 population-ascertained cases of European ancestry and 14,810 matched controls, implicating two loci and 10 genes on chromosomes 13, 15, 16, 17, and 19. In this report, we successfully validate prediction of case-control status, accounting for ancestry, across independent training and test sets using polygenic risk scores for suicide death. Furthermore, we report that suicide death cases carry significantly increased genetic risk for autism spectrum disorder, major depression, psychosis, and alcohol use disorder relative to controls. Results validate several known epidemiological risk factors and suggest that our genetic research can lead to reliable biomarkers of suicide death.

Introduction

Suicide death is a behavioral event which reflects a complex, heritable phenotype with diverse clinical antecedents and environmental contributing factors. The rate of suicide death has been steadily increasing,[1] and in the United States, suicide is now ranked the second leading cause of death for all persons 15-24 years old.[2] Despite a significant heritability of suicide death,[3] genetic research on suicide has been limited to the study of suicide-related behaviors rather than the extreme phenotype of suicide death.

Suicidal behaviors present in diverse ways and reflect varying levels of risk and severity[4,5] and this variation may be leveraged to isolate genetic precursors of suicide death. However, most suicidal behavior does not result in suicide death, and the unambiguous phenotype of a suicide death skirts several confounds inherent in the use of suicidal behavior phenotypes.

Previous genetic research on suicidal behavior phenotypes has also tended toward rigorous ascertainment, studying only individuals with specific diagnoses (e.g., mood disorders, psychotic disorders) in order to maximize severity. In a population-based sample with ascertainment independent of any co-occurring diagnoses, the distribution, prevalence, and interaction of variables can be assumed to exist in the population.

Individuals who die by suicide, like those who suffer from schizophrenia or depression, have a condition that is likely highly complex and polygenic.[6,7] Currently the scientific literature lacks any examination of 1) suicide death in relation to molecular genetic risk for any medical or psychiatric diagnoses, and 2) molecular genetic risk for suicide death in relation to clinical diagnostic precursors. This study leveraged the world's largest DNA databank of suicide death, and these data were merged with a massive bank of electronic medical record and demographic data for all cases, to comprehensively model common variant genetic and clinical phenotypic precursors of suicide death.

This study represents the first genome-wide association study of suicide death. Furthermore, analyses leveraged comprehensive data on five modes of suicide, medical and psychiatric diagnostic codes, and medical and psychiatric polygenic risks to predict common variant genetic risk for suicide death. This study additionally followed up with models of sex differences to address a typically high male-to-female prevalence ratio.

This study sought to reliably differentiate cases from controls, accounting for critical covariates such as ancestry and sex, on 1) polygenic risk for suicide death and 2) on polygenic risk for specific psychiatric and medical risk factors. We additionally examined whether clinical phenotypes are stronger predictors of mode of suicide than are polygenic scores. Finally, we review important considerations relating to ancestry in suicide genetics research.

113 **Results**
114 **Genome-Wide Association**
115 A total of six variants from two loci met genome-wide criteria for statistical association with suicide death
116 ($p < 5 \times 10^{-8}$). An additional 52 variants were nominally significant at $q < 0.05$ and mapped to 19 genes. (λ
117 = 1.015, **Figure 1** and **Table 1**).[8,9] All results on the full cohort are derived from analyses adjusting for
118 effects of ancestry and sex. Genes associated with top genomic regions are presented in
119 **Supplementary Table S1**. Chromosome 13 and 15 regions were supported by additional positive results
120 that were suggestive but below threshold. Ten additional genes were identified in gene-based tests
121 meeting threshold for nominal significance (**Figure 2**). The large number of signals in the SNP-based
122 tests prompted quality control analyses varying the degree of LD pruning prior to PCA for the purposes of
123 sensitivity analysis, and results and respective λ's were consistent across these analyses.
124 **Supplementary Figures S1-S8** present additional plots of the top signals in each of nine regions.
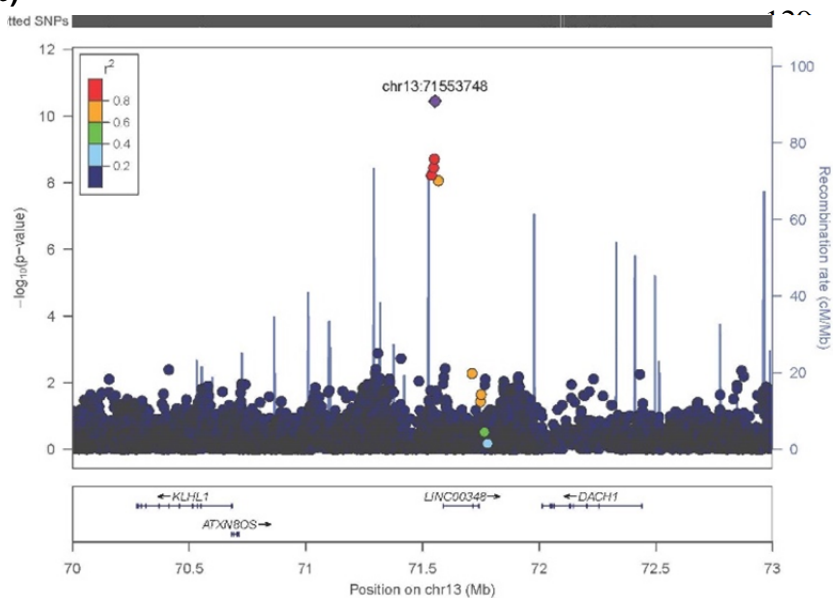125
126 **(a)**
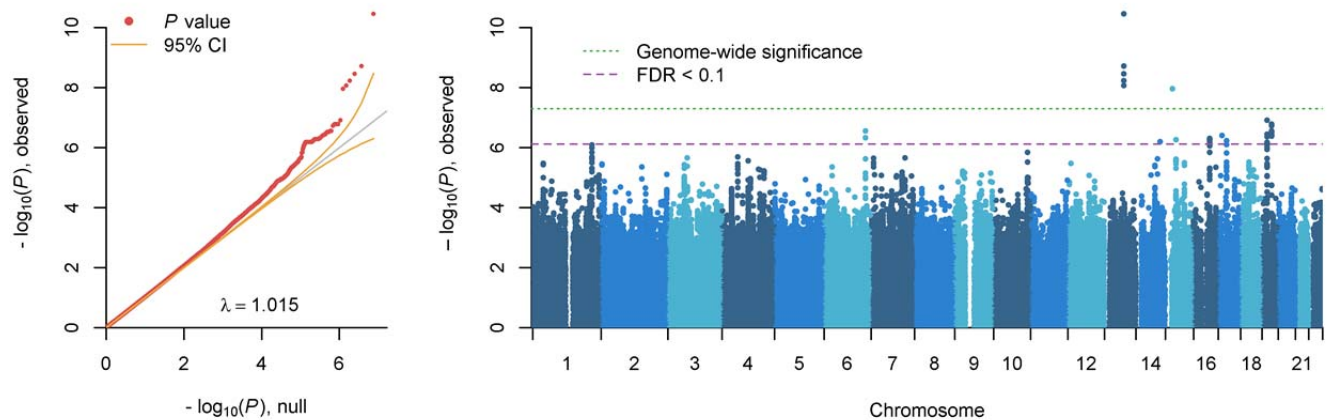
127
128 **(b)**

Figure 1. GWAS SNP-Based Results. (a) QQ and Manhattan plots from the 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 threshold for false discovery rate (FDR) corrected nominal statistical significance, the green dotted line representing the threshold for genome-wide significance after FDR correction. 57 SNPs met threshold for nominal significance and 6 met

149 genome-wide significance. **(b)** Locus zoom plot of genome-wide significant loci on chromosome 13.
150
151
152
153

154

Table 1: Top Signals from GWAS of Death by Suicide.

Chr	Position	SNP	A1	A2	AF	β	SE	p value	q value	Nearest Gene	Function	CADD
13	71,538,274	rs35502061	G	A	0.016	0.089	0.0153	5.97×10^{-9}	5.63×10^{-3}	*SOGA2P1*	Intergenic	2.69
13	71,547,393	rs34053895	A	C	0.016	0.091	0.0154	3.51×10^{-9}	3.78×10^{-3}	*LINC00348*	Intergenic	0.39
13	71,550,518	rs35518298	T	C	0.016	0.092	0.0154	1.92×10^{-9}	2.42×10^{-3}	*LINC00348*	Intergenic	0.33
13	71,553,748	rs34399104	T	C	0.017	0.098	0.0147	3.54×10^{-11}	6.67×10^{-5}	*LINC00348*	Intergenic	19.86
13	71,567,365	rs66828456	A	C	0.017	0.086	0.0150	8.63×10^{-9}	7.24×10^{-3}	*LINC00348*	Intergenic	4.24
15	25,962,209	rs35256367	G	A	0.016	0.088	0.0155	1.10×10^{-8}	8.33×10^{-3}	*ATP10A*	Intronic	4.41

Note: SNP = single nucleotide polymorphism, CHR = chromosome, A1, A2 = alleles 1 (minor) and 2, AF = allele 1 frequency, CADD[10,11] = combined annotation dependent depletion score.
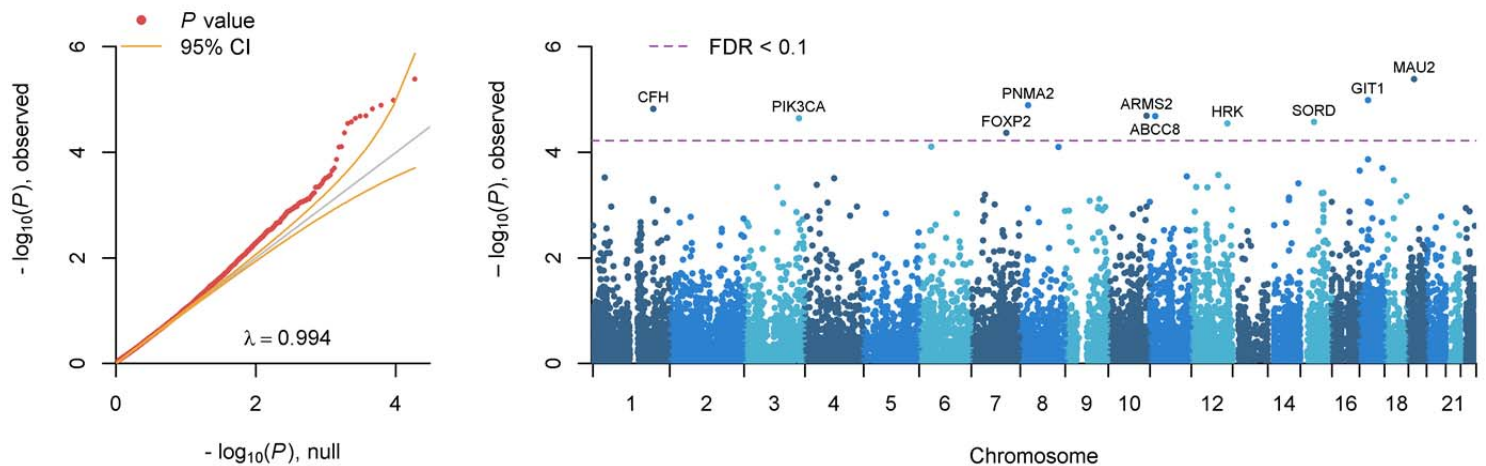

156
157
158 **Figure 2. GWAS Gene-Based Results.** Gene-based tests: qq plot and Manhattan plot of >18,000 genes.
159 Y-axes for both plots are identical and reflect observed p-values. The x-axis on the qq-plot is the number
160 of significant p-values expected under H_0, and the x-axis on the Manhattan plot maps each chromosome.
161 The purple dashed line indicates threshold for FDR-corrected nominal statistical significance; 10 genes
162 met this threshold for nominal significance. A polygenic signal reflected in a slow, gradual lift from the
163 diagonal, is observed in both SNP and gene QQ plots in Figures 2 and 3.
164
165
166 **Gene- and Pathway-Based Functional Enrichment Tests**
167 Gene-base analysis using MAGMA (FUMA[1]) identified 19 genes associated with suicide death.
168 Associations were observed between chr13 SNPs and Daschund family transcription factor 1 (*DACH1*),
169 Ubiquitin-protein ligase protein (*UBE3A*), and Kelch-like family member 1 (*KLHL1*). Eleven of the 19
170 associated genes carry prior evidence of association with suicidal behaviors **(Supplementary Table S2).**
171 GO pathway results included enrichment of histone modification sites *SETD6, COPR5, GATAD2A.*
172 Comprehensive gene and Gene Ontology (GO, http://www.geneontology.org/) pathway enrichment
173 results are presented in **Supplementary Tables S3-S4.** In addition to functional pathways, a significant
174 association with schizophrenia results in the GWAS Catalog was identified ($p=1 \times 10^{-11}$)
175 (**https://www.ebi.ac.uk/gwas/**). Psychiatric associated traits are in green. IW-scoring in SNP-Nexus[12]
176 suggested regulatory functional significance for one SNP (chr13:71553748:C/T). Ten of the implicated

177 genes from positional or gene-based testing have evidenced genome-wide significant differential gene
178 expression in postmortem brain in either schizophrenia, autism, or bipolar disorder (FDR<0.05;
179 PsychENCODE Consortium, **Supplementary Table S5**).[13]
180
181 **Cross-Validation from Derived Suicide Death Polygenic Risk Scores**
182 In European ancestry training and test samples comprising independent case and control cohorts, and
183 accounting for five ancestry PCs and sex, suicide PRS robustly predicted suicide death case status.
184 Suicide waves 1 and 2 comprise approximately 1,321 and 2,092 suicide cases, respectively. These
185 predictions are plotted across 1000 p-value thresholds in **Figure 3**.
186
187
188
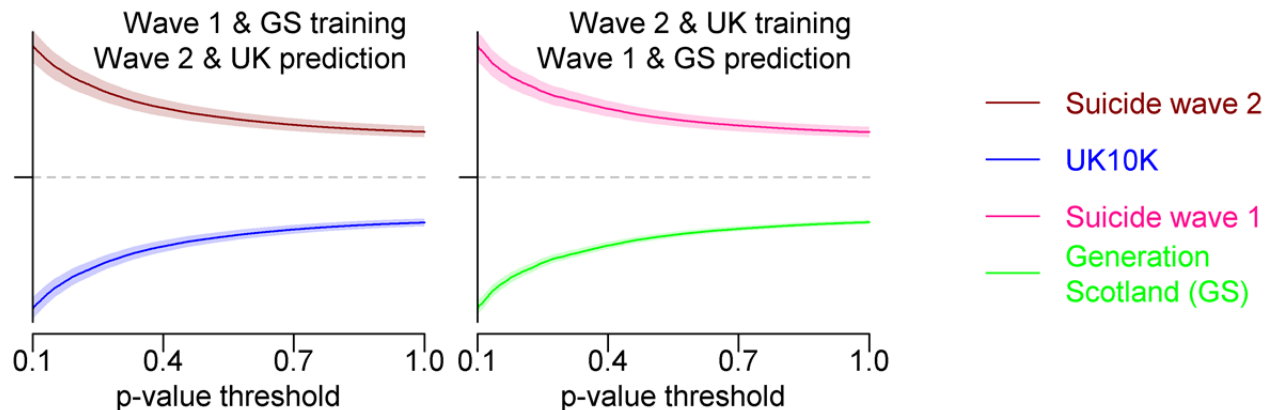
189
190
191 **Figure 3**. **Cross Validation of Suicide Polygenic Case-Control Prediction**. Polygenic prediction of
192 suicide death case status across two independent cohorts of cases and controls. Training GWAS
193 summary statistics are used to score the test set for suicide polygenic risk. P-value thresholds are plotted
194 on the x-axis from 0.1-1.0, reflecting the top 10% to 100% of the common variants from the training
195 GWAS. On the y-axis, all suicide PRS scores are centered at zero (dotted midline). 95% confidence
196 intervals around the scores are pictured for each cohort across all p-value thresholds.
197
198
199 **SNP-based h^2 and Polygenic Risk Score Association of Suicide with Multiple Complex Traits**
200 A Linkage Disequilibrium SCore regression (LDSC)[14,15] common variant h^2 estimate based on only the
201 summary statistics from a logistic GWAS, with five ancestry covariates and pruning to remove related
202 samples, was 0.2463, SE = 0.0356. Lambda in the latter model was inflated at 1.239. The suicide death
203 cases differed significantly from two UK control groups on PRS of phenotypes relevant to suicide death.
204 These differences were in the expected directions. Original discovery GWAS for all phenotypes were
205 filtered to exclude any using these control cohorts (**Supplementary Table S6**).
206
207 Consistent with hypotheses, significant PRS elevations included alcohol use, autism spectrum disorder,
208 child IQ, depressive symptoms, disinhibition, loneliness, and neuroticism (**Figure 4)**. Narrower error bars
209 correspond to increasing power, reflective of larger discovery GWAS. LD Hub[14] provided estimates of
210 SNP-based shared genetic covariance for several phenotypes (**Supplementary Table S7**).
211 PRS for suicide death, derived by cross-validation procedure, was regressed onto the other medical and
212 psychiatric PRS, including two ancestry PCs and sex as covariates **(Supplementary Figure S9).** PRS
213 association with autism spectrum disorder PRS was suggestive, though no associations remained
214 significant after testing correction. All summary statistics for PRS-PRS in cases are provided in
215 **Supplementary Tables S8-S10**.
216
217
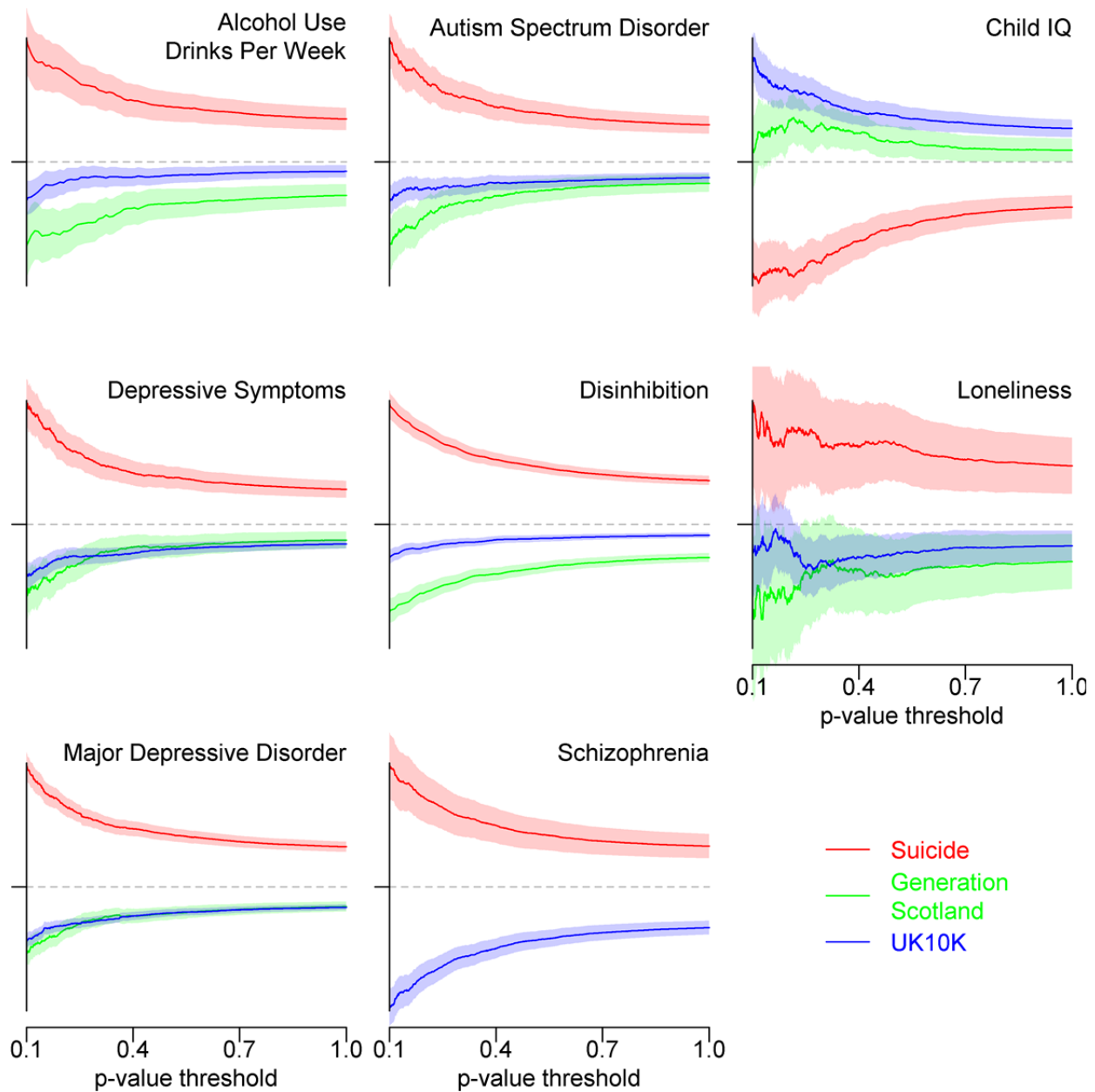
Figure 4. Notable Elevations of Psychiatric Polygenic Risk in Suicide Cases. Polygenic risk scores for eight phenotypes hypothesized to be relevant to suicide (centered, on the y-axis) plotted for suicide death case, and GS and UK10k control groups across a broad spectrum of PRS p-value thresholding. P-value thresholds are plotted on the x-axis from 0.1-1.0. 95% confidence intervals around the scores are pictured for each cohort across p-value thresholds. The schizophrenia GWAS meta-analysis included GS, thus GS were not scored for schizophrenia.

Clinical Diagnostic Associations with Mode of Death are presented in **Figure 5.** Suicide by gun was associated with the general absence of clinical diagnoses. Suicide by overdose was associated with many clinical diagnoses and most robustly with obesity and sleep disorders. Suicide by violent trauma was associated with a clinical diagnosis of schizophrenia. Corresponding βs and p-values for **Figure 5** are presented in **Supplementary Tables S11-S13.**
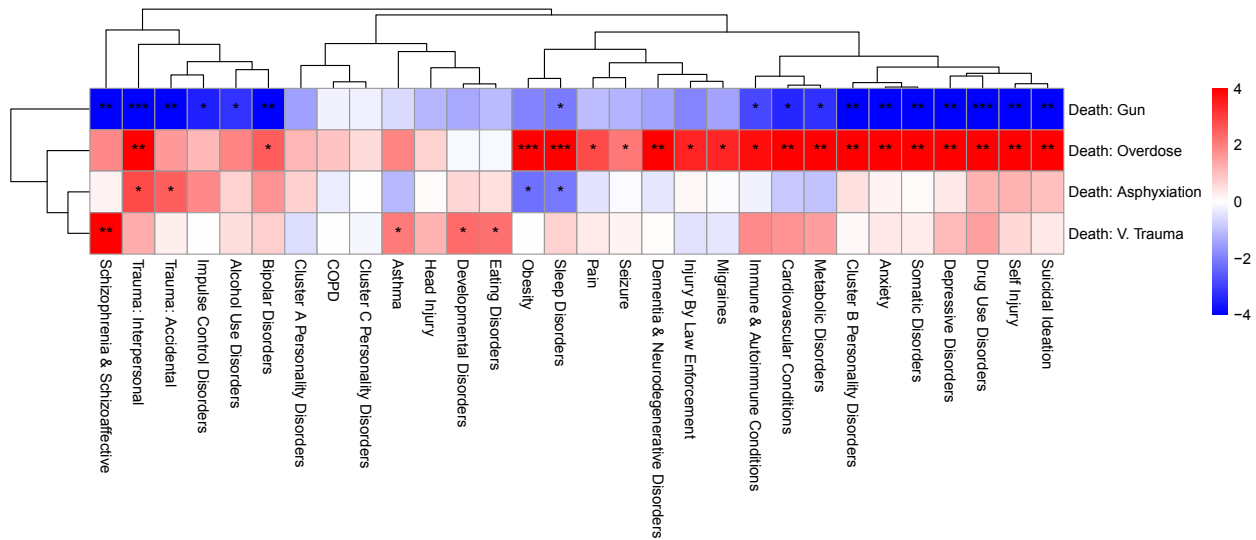
Figure 5. Diagnostic Antecedents of Specific Mode of Suicide. Medical record diagnoses (x-axis) mapped to mode of death (y-axis) in the suicide death cases. Dendrograms based on k-means clustering of Euclidean distances are featured on both axes. Shading reflects the test statistic value as positive (red) or negative (blue). $p<10^{-2}$ (*), $p<10^{-4}$ (**), and $p<10^{-10}$ (***).

PRS Associations with Mode of Death are presented in **Supplementary Figure S10**, including associations of all PRS (including suicide death PRS) with mode of death, covarying for ancestry and sex. Asthma and schizophrenia were negatively and positively associated with violent suicide, respectively, but did not reach significance after multiple testing correction. All corresponding βs and p-values for **Figure S10** are presented in **Supplementary Tables S14-S16**.

All Clinical Diagnostic Associations with PRS are presented in **Supplementary Figure S11**, including associations of all PRS (including suicide death PRS) with International Classification of Diseases (ICD) codes comprising 30 diagnostic categories, covarying for both ancestry and sex. Several informative patterns emerged but no significant associations were observed after FDR correction. βs and p-values for all associations are provided in **Supplementary Tables S17-S19**.

Sex Differences in Associations of Suicide Modality, Polygenic Risk, and Clinical Diagnosis
All associations of diagnoses and PRS with mode of death are presented for females and males separately in **Supplementary Figures S12-S17**. All PRS analyses include ancestry covariates. All corresponding statistics are reported in **Supplementary Tables S20-S37**. Suicide cases of both sexes evidenced clinical diagnostic clusters of 1) internalizing-trauma-cluster B psychiatric disorders and 2) metabolic-cardiovascular-obesity medical disorders. Female cases were observed to have a higher overall number of diagnoses relative to males, which could reflect increased severity in females, decreased severity in males, decreased likelihood of males receiving diagnosis, and/or decreased help-seeking in males. This is broadly consistent with observed higher relative prevalence rates of gun-related death in males and overdose death in females.

Power Analysis
With a suicide death prevalence rate of approximately .002%, this GWAS of European ancestry with 3,413 population-based suicide deaths and 14,810 ancestry-matched controls would be expected to have at least 80% power to detect common variants (MAF ≥ 0.15) with effect sizes ≥ 1.20 at $P<5\times10^{-8}$ and $P<1\times10^{-6}$ (**Supplementary Figure S18**). Power at $P < 1 \times 10^{-6}$ is relevant because 52 SNPs reach that threshold in the current analysis. Power is lower for less-common variants (MAF ≤ 0.05) even with odds ratios ≥ 1.20 at $P < 1 \times 10^{-6}$.

Discussion

273
274 This GWAS of suicide death resulted in genome-wide significance in a region on chromosome 13q21.33.
275 Eleven of 22 associated genes overlap with schizophrenia results from the GWAS Catalogue, and two of
276 these 11 genes have prior associations with risk of suicidal behavior (*HS3ST3B, NCAN*; for relevant
277 literature see **Supplementary Table S38**). Cross-validation procedures predicted case-control status with
278 summary statistics across training and test sets. Population-based suicide was associated with elevated
279 polygenic scores for multiple suicide risk factors. These included alcohol use, autism spectrum disorder,
280 child IQ, depressive symptoms, disinhibition, and loneliness. Suicide by violent trauma was associated
281 with both a clinical diagnosis of schizophrenia or schizoaffective disorder and with genome-wide
282 polygenic risk for schizophrenia.
283
284 Genetic overlap of actual suicide death with suicidal behaviors remains unclear to date. More common
285 suicidal behaviors are difficult to quantify and represent individuals with a range of risk for later suicide.[16]
286 To truly understand risk of suicide death and to implement highly effective interventions that provide
287 appropriate, targeted services to those most likely to die, we must understand the risks specifically
288 associated with suicide deaths. Moving closer to developing objective risk measures of suicide, future
289 modeling of the shared genetic covariance of suicide death and suicide behaviors may isolate important
290 genetic and environmental moderators of risk of death.
291
292 Importantly, we must study genetic risk in other ancestry groups to address the potential for increasing
293 health disparities stemming from polygenic risk research that relies only on European ancestry summary
294 statistics. To this end, the authors are working toward cross-ancestry replication of results in individuals of
295 Mexican American ancestry with ongoing collection of population-based cases. Future priorities also
296 include analyses of structural variation, methylation, and predicted gene expression in suicide,
297 investigation of the potential mediating role of substance/alcohol use in suicide, and ethical analysis of
298 genetic predictive models of suicide.
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323

Materials and Methods

1. Sample Ascertainment

Cases

In collaboration with the centralized, statewide Utah Office of Medical Examiner (OME), the authors obtained DNA samples from ~6000 persons who died by suicide. The centralized OME and conservative determination helped to maximize the accuracy of suicide case status.[17] Suicide cause-of-death determination results from a 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, in addition to standard toxicology workups. Suicide determination is traditionally made quite conservatively due to its impact on surviving relatives. DNA from suicide deaths extracted from whole blood using the Qiagen Autopure LS automated DNA extractor (www.qiagen.com). Genotyping was performed on 4,381 of these cases, as described below. After quality control procedures and ancestry analysis, data comprised 3,413 Utah suicide deaths. The Utah population is primarily Northwestern European in ancestry, a relatively genetically homogeneous group with very low inbreeding across generations, comparable to the rest of the United States.[18] Suicide determination results from a 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, in addition to standard toxicology workups. Suicide determination is traditionally made quite conservatively due to its impact on surviving relatives. The centralized OME and conservative determination helped to maximize the accuracy of suicide case status.[17]

Controls

Generation Scotland

Controls closely matching the Northern European ancestry of the cases were obtained from previously curated datasets in the UK. Wave 1 analysis included 3,623 founder controls from the population-based Generation Scotland Scottish Family Health Study.[19] The Generation Scotland Scottish Family Health Study (N > 24,000) constitutes an ancestrally comparable population-based cohort for comparison with the suicide decedents in Utah. To eliminate confounding arising from intra-dataset relatedness, only the 3,623 founders from the Generation Scotland dataset were used in analyses.

UK10k

A total of 11,049 UK10K controls[20] were analyzed in wave 2 and GWAS analyses of both waves. This second control cohort is comprised of approximately 4000 genomes from the UK along with 6000 exomes from UK individuals with selected health phenotypes. We chose these data due to the extensive phenotyping and characterization of any medical conditions present, and to avoid choosing a cohort of entirely psychiatrically and medically healthy individuals. 4,000 highly phenotyped "super control" samples were supplied from the King's College London registry and the Avon Longitudinal Study of Parents and Children. UK10K 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. Genotyping and sequencing procedures for UK10k have been previously described[20] (http://www.uk10k.org) and all molecular genetic data from UK10k were filtered to the hard call variants present in our suicide death cohort prior to imputation of all cohorts simultaneously.

1000 Genomes Reference Panel

The CEU population from the 1000 Genomes Project,[21] 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 1000 Genomes Project public repository. Unrelated individuals in the CEU provide a compelling, albeit small, ancestrally matched control resource (n = 99). A variety of candidate control samples were assessed via PCA for ancestral comparability to CEU and decedent data, with UK10k and Generation Scotland founder data representing the closest match.

Utah controls would be an ideal match for the suicide cases, but as with most GWAS, local controls were not readily available at the sample size needed for GWAS. CEPH ancestry 1KG were a useful comparison group to assess the likelihood that UK controls were an appropriate match for the cases. In addition (and described in more detail below) we performed a GS control-UK10K control GWAS and subsequently eliminated any SNPs from the case-control analysis that evidenced signal between the

379   control cohorts. This was performed to minimize the possibility of false positives in the case control
380   GWAS due to population/geographic stratification across cohorts.
381   **2. Genotyping and Quality Control**
382   Suicide cases were genotyped using Illumina Infinium PsychArray platform measuring 593,260 single
383   nucleotide polymorphisms (SNPs). Generation Scotland samples were genotyped using Illumina
384   OmniExpress SNP GWAS and exome chip, measuring 700,000 and 250,000 SNPs, respectively.[19]
385   UK10k samples were whole genome sequenced[20] and variants were extracted to match the available
386   QC'd hard-called 7,519,308 variants in the suicide cases. Genotypes were subsequently imputed in all
387   cases and controls (details of imputation are presented in Analytics, below). Both case and control
388   datasets resulted from population-based ascertainment, and cryptic relatedness was modeled via the
389   derivation of genomic relatedness matrices. Genotyping quality control was performed using SNP
390   clustering in the Illumina Genome Studio software
391   (https://www.illumina.com/techniques/microarrays/array-data-analysis-
392   experimentaldesign/genomestudio.html). SNPs were retained if the GenTrain score was > 0.5 and the
393   Cluster separation score was > 0.4. SNPs were converted to HG19 plus strand, and SNPs with >5%
394   missing genotypes were removed. Samples with a call rate < 95% were removed. Prior to case-control
395   GWAS, a control-control GWAS was run (using the same methods described in Analytics: GWAS, below)
396   to detect signal between control groups to filter out of the case-control GWAS (control-control q value
397   >.10). For example, chromosome 4 variants within the MHC are often filtered from analyses involving
398   Scottish controls, due to prevalent population stratification in this region. We performed a stringent screen
399   for population-specific signal in the controls. While this method is somewhat conservative, it was deemed
400   necessary to address potential geographic stratification confounds. Signal detected from this control-
401   control GS vs. UK10k comparison was then filtered from subsequent case-control analyses. For the
402   purposes of future meta-GWAS analyses, and because the MHC is relevant to psychiatric risk, we
403   included these filtered variants in a second version of our summary statistics, available upon request.
404   **3. Analytics**
405   **Principal Components Analysis (PCA)**
406   **Figure S18** shows 1000 Genomes (1KG) superpopulations and suicide case/control samples, both
407   included and excluded, plotted by the top 2 principal components (PC). Approximately 20% of the
408   population-based suicide cases had a significant degree of non-Northwestern European ancestry (chiefly
409   of admixed ancestry) and were excluded from analyses. The variation explained by top 4 PCs was
410   reduced 7.2-fold. The top 4 PCs explain 0.89% of variation before sample filtering and 0.12% of variation
411   after filtering, if calculated on pruned genotypes. For adequate statistical power, we examined only cases
412   of Northern European ancestry. However, it is clear from (a) and (c) that the cohort was comprised of
413   multiple ancestries and that research on suicide death in non-European ancestries will reflect an
414   important step beyond this first study.
415   PCA was performed on control, suicide, and 1000 Genomes cohorts after LD pruning at a 0.2 threshold.
416   To exclude ancestrally heterogeneous samples, the top principal components (defined as those
417   components which accounted for > 0.1% of the genotype variance, n_{pc} = 4) were used to establish PC
418   centroid limits centered around 1000 Genomes CEU data, such that 99% of the CEU data fell within the
419   limits. Only suicide and control samples also falling within these limits were considered ancestrally
420   homogenous and thus were included in the association study. The ancestry PCA was performed using
421   RaMWAS,[22] a Bioconductor package which comprises a complete toolset for GWAS and methylome-wide
422   association studies. RaMWAS includes functions for PCA for capturing batch effects and detection of
423   outliers, association analysis while correcting for top PCs and covariates, creation of QQ-plots and
424   Manhattan plots, and annotation of significant results.
425   **Imputation**
426   European ancestry cases and controls were well-matched to 1000 Genomes CEPH. The Haplotype
427   Reference Consortium is comprised in part by UK controls used here, so we elected to impute genotypes
428   based on the 1000 Genomes reference panel using minimac3[23] and Eagle.[24] SNPs with ambiguous
429   strand orientation, >5% missing calls, or Hardy-Weinberg equilibrium p < 0.001 were excluded. SNPs with
430   minor allele frequency below 0.01 or imputation R^2 < 0.5 were also excluded. Genomic data were handled
431   using PLINK. [25,26] Final GWAS analysis was performed on 7,519,308 variants passing quality control.
432   **GWAS**
433   A Linear Mixed Model (LMM) algorithm tested variant association with suicide death, with follow-up
434   examination of significant hits for linkage disequilibrium and gene set enrichment. GWAS were performed

435 using GEMMA,[27] a computationally efficient and open-source LMM algorithm for GWAS that models
436 population stratification remaining after PCA by use of genomic relatedness matrices. Sex was not
437 included as a covariate in GWAS analyses due to the association of suicide with sex status at a ratio of
438 approximately 3:1 males: females. GWAS with hard call-only and then with imputed data were examined
439 separately to assess potential population stratification unique to our imputed GWAS. Prior to case-control
440 GWAS, control-control GWAS was implemented to filter signal likely due to population stratification in the
441 controls.

442 **SNP-Based Heritability** (h^2) **and Shared Polygenic Covariance** (r_G)
443 LDSC was used to calculate common variant h^2 using summary statistics from a logistic regression model
444 with five ancestry covariates and pruning related samples at 0.2 $\hat{p}$ from IBD. LDSC was also used to
445 calculate common variant molecular genetic correlations (r_G) with psychiatric and medical phenotypes.

446 **Gene Set Enrichment and Functional Mapping**
447 SNP results from the GWAS were then mapped to genes within 1kb of the SNP and these genes were
448 examined for gene set enrichment and LD using FUMA[28] and GREAT.[29] FUMA annotates SNPs, uses
449 MAGMA to identify associated genes (of approximately 18,612) and provide gene and gene pathway
450 enrichment analysis (of approximately 10,649 pathways). GREAT analyzes the functional significance of
451 sets of *cis*-regulatory regions, by modeling the genome regulatory landscape using multiple
452 information sources and can determine the functional domain of intergenic variants. GREAT improves
453 upon the identification of genes associated with non-coding genomic regions through statistically rigorous
454 incorporation of more distal binding sites from 20 ontologies. The GWAS catalog
455 (https://www.ebi.ac.uk/gwas/) includes studies and associations if they: include a primary GWAS analysis
456 from >100,000 SNPs, SNP-trait p-value <1.0 x 10-5 in the overall (initial GWAS + replication) population.
457 The most significant SNP from each independent locus is extracted.

458 **Polygenic Risk Scoring**
459 Discovery GWAS summary statistics for phenotypes were compiled to score each cohort for polygenic
460 risk. PRS for suicide death was derived using PRSice 2.0[30] and summary statistics from a 10-fold cross
461 validation procedure to avoid overfitting. We did not allow PRSice to select the p-value threshold for
462 predicting case status based on the data, and set it to 1.0. This eliminated the overfitting arising from
463 choosing the threshold based on the phenotype. Of several thousand medical and psychiatric GWAS now
464 available, only GWAS with N>10,000 individuals and >1,000 cases (or for population-based studies,
465 adequate base rates) were selected for these analyses. These generally included the largest medical and
466 psychiatric GWAS (**Supplementary Table 1**), and when several versions of GWAS were available for the
467 same phenotype (for example, neuroticism or depression) we selected the most comprehensive. For a
468 helpful reference to GWAS available, see Watanabe et al.'s online GWAS Atlas (http://atlas.ctglab.nl/).
469 PRSice 2.0 was used to calculate individual PRS for 59 phenotypes with estimated risk allele effect sizes
470 for each discovery sample trait. A PRS is traditionally calculated as a weighted sum score, where a score
471 for an individual in the target sample is calculated by the summation of each SNP multiplied by the effect
472 size of that SNP in the discovery GWAS.

473 **Cross-Disorder PRS Association**
474 Betas, p-values, and q-values (after correcting the p-values for the FDR of 5%) were derived for PRS-on-
475 suicide PRS regressions using R and adjusting for five ancestry principle components. Based on the
476 cross-disorder psychiatric genomics findings to date, we hypothesized significant positive prediction of
477 suicide with PRS for depressive symptoms, schizophrenia, autism, loneliness, child IQ, alcohol use,
478 disinhibition, and neuroticism.

479 **International Classification of Diseases (ICD-10) Code Ascertainment**
480 Decedent data are linked within the Utah Population Database (UPDB), a unique resource that houses
481 data on > 9 million individuals and contains vital statistic and demographic records and electronic health
482 records from the two major hospital systems in the Utah.[31] ICD codes for suicide cases were recovered
483 from the electronic data warehouses of Utah's two largest health care providers, Intermountain
484 Healthcare and the University of Utah. Ambulatory and inpatient ICD codes were obtained directly from
485 UPDB. EMR data, while comprehensive and population-based, are also subject to missingness (random
486 and non-random), and severity of health disparity may correspond with the presence and absence of ICD
487 codes in an EMR. For this reason, we checked ICD code associations with polygenic risk for the same
488 disorders and assessed relationships of ICD with PRS across age group and date of code. The rationale
489 for the latter analyses was that younger individuals are less likely to have specific ICD codes due to either
490 lack of contact with the system or to lack of maturation of overt psychopathology (e.g., psychosis,

491   personality disorders). In addition, ICD codes from early EMR development in Utah (pre-year 2000) are
492   notably sparser. To manage the wide diversity of codes, we limited our analyses to 30 categories of
493   codes reflecting 30 relevant diagnostic domains. Data were subsequently completely de-identified prior to
494   analysis. The study was approved by the Institutional Review Boards of the University of Utah,
495   Intermountain Healthcare, and the Utah Department of Health.
496   **Association of Psychiatric and Medical PRSs, ICD Diagnoses, and Mode of Death**
497   Statistical associations of 59 PRS (Listed in **Table S6**), 30 ICD diagnostic categories (Y/N; x-axis of
498   **Figure 5**), and four modes of death (each Y/N; gunshot, asphyxiation, overdose, violent trauma) were
499   examined in exploratory analyses within all suicide cases. Each regression was run using R to compare
500   full (e.g., PRS, five ancestry principle components, age predicting mode of death) and restricted models
501   where PRS was removed. Multiple testing correction False Discovery Rate (FDR) was 5%. Dendrograms
502   according to k-means clustering using Euclidean distances were configured in R. Due to decreased
503   likelihood of many of the ICD codes in individuals <25 years of age, we included for those analyses
504   individuals >25 who had reached the age of diagnosis for many common psychiatric disorders (e.g.,
505   schizophrenia, personality disorders). Examination of PRS-phenotype associations included cluster
506   analysis of 59 polygenic risks, 30 categories of ICD diagnoses, and five modes of suicide death (gun,
507   poison, asphyxiation, violent trauma, and other).
508   **Sex Differences**
509   After modeling PRS and ICD associations with suicide PRS in the cases, we followed up with
510   examination of sex differences. We examined associations of suicide PRS and mode of death with clinical
511   diagnostic codes in females and males separately, including five ancestry covariates in multivariate
512   regressions. Model fit was compared across models predicting suicide PRS or mode of death
513   (categorical) using ancestry PCs, with and without clinical code (Y/N) included as an IV. We constrained
514   these exploratory analyses to only those medical diagnoses with frequencies high enough in either
515   females or males to provide adequate power for testing and report false discovery rate (FDR) corrected p-
516   values.
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546

Supplementary Tables
S1. Genes Associated with Top Genomic Regions.
S2-S4. FUMA and GREAT Results, GWAS Catalogue, SNP Nexus
S5. Genome-wide significant differential gene expression in postmortem brain in schizophrenia, autism, or bipolar disorder
S6. PRS Discovery GWAS Summary Statistic Citation Table
S7. LD Hub Genetic Correlation Results
S8-S10. All Case PRS-PRS Summary Statistics: Betas, p-values, result summaries
S11-13. All Case ICD-MD Summary Statistics: Betas, p-values, result summaries
S14-16. All Case PRS-MD Summary Statistics: Betas, p-values, result summaries
S17-19. All Case PRS-ICD Summary Statistics: Betas, p-values, result summaries
S20-25. Separate Female and Male ICD-MD Summary Statistics: Betas, p-values, result summaries
S26-S31. Separate Female and Male PRS-MD Summary Statistics: Betas, p-values, result summaries
S32-37. Separate Female and Male PRS-ICD Summary Statistics: Betas, p-values, result summaries
S38. Gene Literature (brief)

Supplementary Figures
S1. LocusZoom Plot of Chromosome 6 Region
S2. LocusZoom Plot of Chromosome 14 Region
S3. LocusZoom Plot of Chromosome 15 Region
S4. LocusZoom Plot of Chromosome 15 Region
S5. LocusZoom Plot of Chromosome 16 Region
S5. LocusZoom Plot of Chromosome 17 Region
S7. LocusZoom Plot of Chromosome 17 Region
S8. LocusZoom Plot of Chromosome 19 Region
S9. Suicide PRS Associations with Multiple Medical and Psychiatric PRS
S9. PRS x MD: All suicide deaths
S10. ICD x PRS plots: All suicide deaths > 25 years of age
S11. ICD x MD plots: Female suicide deaths > 25 years of age
S12. ICD x MD plots: Male suicide deaths > 25 years of age
**S13. PRS x MD: Female suicide deaths**
S14. PRS x MD: Male suicide deaths
S15. ICD x PRS plots: Female suicide deaths > 25 years of age
**S16. ICD x PRS plots: Male suicide deaths > 25 years of age**
S17. Effect sizes across frequencies required to detect both genome-wide and nominal levels of significance
**S18. PCA of 1KG super-populations, cases, controls, and excluded samples**

References

1.   Kochanek, K., Murphy, S., Xu, J. & Tejada-Vera, B. Deaths: Final Data for 2014. *National Vital Statistics Reports* **65**, 1-121 (2016).

2.   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).

3.   McGuffin, P., Marusic, A. & Farmer, A. What can psychiatric genetics offer suicidology? *Crisis* **22**, 61-5 (2001).

4.   Turecki, G. Suicidal behavior: is there a genetic predisposition? *Bipolar Disord* **3**, 335-49 (2001).

5.   Turecki, G. & Brent, D.A. Suicide and suicidal behaviour. *Lancet (London, England)* **387**, 1227-1239 (2016).

6.   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).

7.   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).

8.   Narum, S.R. Beyond Bonferroni: less conservative analyses for conservation genetics. *Conservation genetics* **7**, 783-787 (2006).

9.   Hochberg, Y. & Benjamini, Y. More powerful procedures for multiple significance testing. *Stat Med* **9**, 811-8 (1990).

10.   Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nature genetics* **46**, 310 (2014).

11.   Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic acids research* **47**, D886-D894 (2018).

12.   Dayem Ullah, A.Z. *et al.* SNPnexus: assessing the functional relevance of genetic variation to facilitate the promise of precision medicine. *Nucleic acids research* **46**, W109-W113 (2018).

13.   Gandal, M.J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**(2018).

14.   Zheng, J. *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* **33**, 272-279 (2017).

15.   Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics* **47**, 291 (2015).

16.   Younes, N. *et al.* Attempted and completed suicide in primary care: not what we expected? *Journal of affective disorders* **170**, 150-154 (2015).

17.   Utah Medical Examiner Act. in *26-4* (https://le.utah.gov/xcode/Title26/Chapter4/C26-4_1800010118000101.pdf, United States, 2016).

18.   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).

19.   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).

20.   Walter, K. *et al.* The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82-90 (2015).

21.   The Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68 (2015).

22.   Shabalin, A., Clark, S., Hattab, M., Aberg, K. & van den Oord, E. *RaMWAS: Fast Methylome-Wide Association Study Pipeline for Enrichment Platforms* . R package version 1.2.0 edn (2017).

23.   Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-1287 (2016).

24.   Loh, P.R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet* **48**, 1443-1448 (2016).

25.   Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

658   26.   *Purcell, S.*, *Chang, C.*, *Sinai, N.-N.L.o.B.M.P.L.a.M.* & *School of Medicine*. *PLINK 1.9 beta.*
659   (2017).
660   27.   Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies.
661   *Nat Genet* **44**, 821-4 (2012).
662   28.   Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and
663   annotation of genetic associations with FUMA. *Nature Communications* **8**, 1826 (2017).
664   29.   McLean, C.Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat*
665   *Biotechnol* **28**, 495-501 (2010).
666   30.   Euesden, J., Lewis, C.M. & O'Reilly, P.F. PRSice: polygenic risk score software. *Bioinformatics*
667   **31**, 1466-1468 (2014).
668   31.   University of Utah Health. Utah Population Database Overview. (2017).
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689

Supplementary Figures

S1. LocusZoom Plot of Chromosome 6 Region
S2. LocusZoom Plot of Chromosome 14 Region
S3. LocusZoom Plot of Chromosome 15 Region
S4. LocusZoom Plot of Chromosome 15 Region
S5. LocusZoom Plot of Chromosome 16 Region
S5. LocusZoom Plot of Chromosome 17 Region
S7. LocusZoom Plot of Chromosome 17 Region
S8. LocusZoom Plot of Chromosome 19 Region
S9. Suicide PRS Associations with Multiple Medical and Psychiatric PRS.
S9. PRS x MD: All suicide deaths.
S10. ICD x PRS plots: All suicide deaths > 25 years of age.
S11. ICD x MD plots: Female suicide deaths > 25 years of age.
S12. ICD x MD plots: Male suicide deaths > 25 years of age.
S13. PRS x MD: Female suicide deaths.
S14. PRS x MD: Male suicide deaths.
S15. ICD x PRS plots: Female suicide deaths > 25 years of age.
S16. ICD x PRS plots: Male suicide deaths > 25 years of age.
S17. Effect sizes across minor allele frequencies required to detect both genome-wide and nominal levels of significance.
S18. PCA of 1KG super-populations, cases, controls, and excluded samples
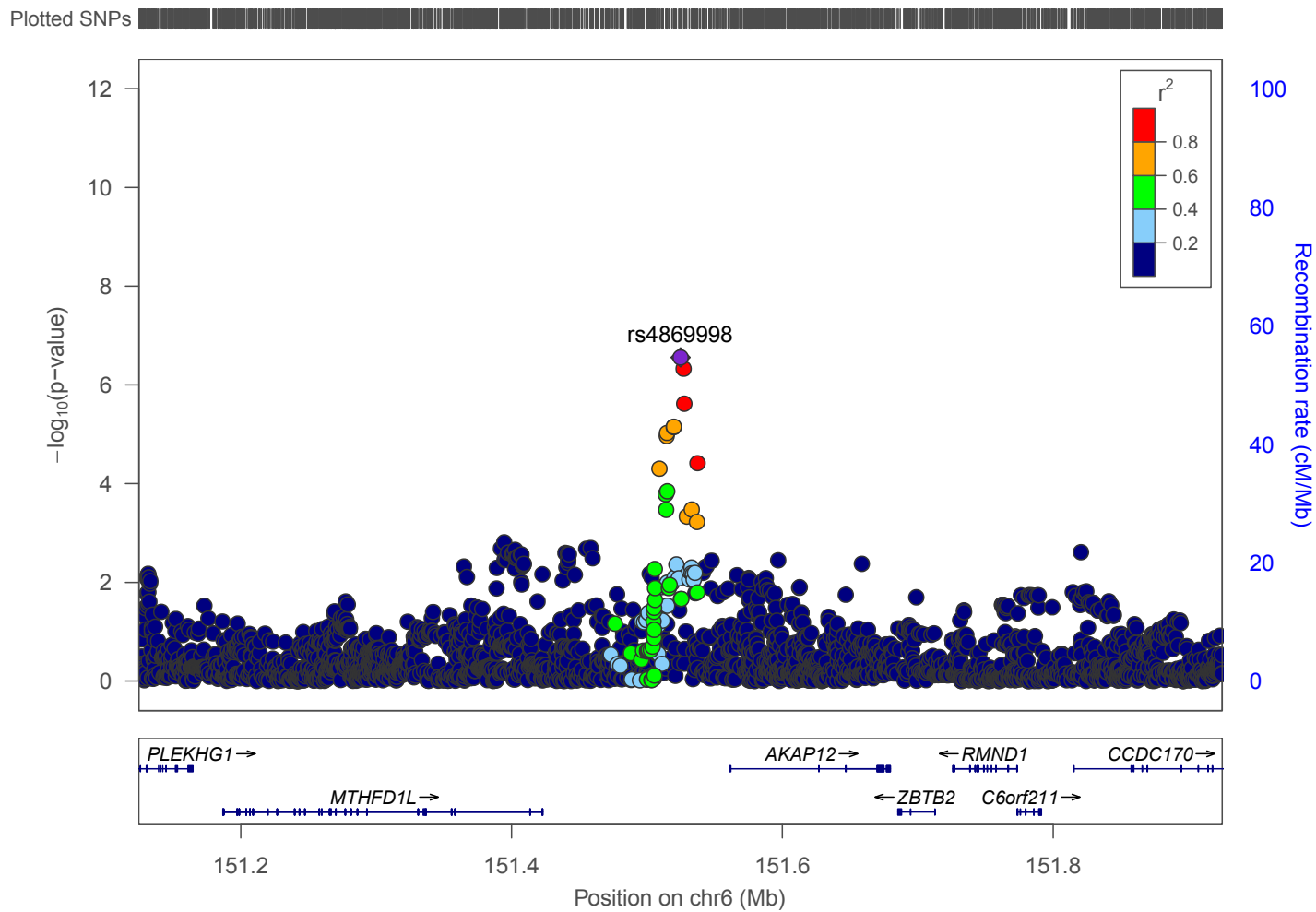
S1. Chromosome 6
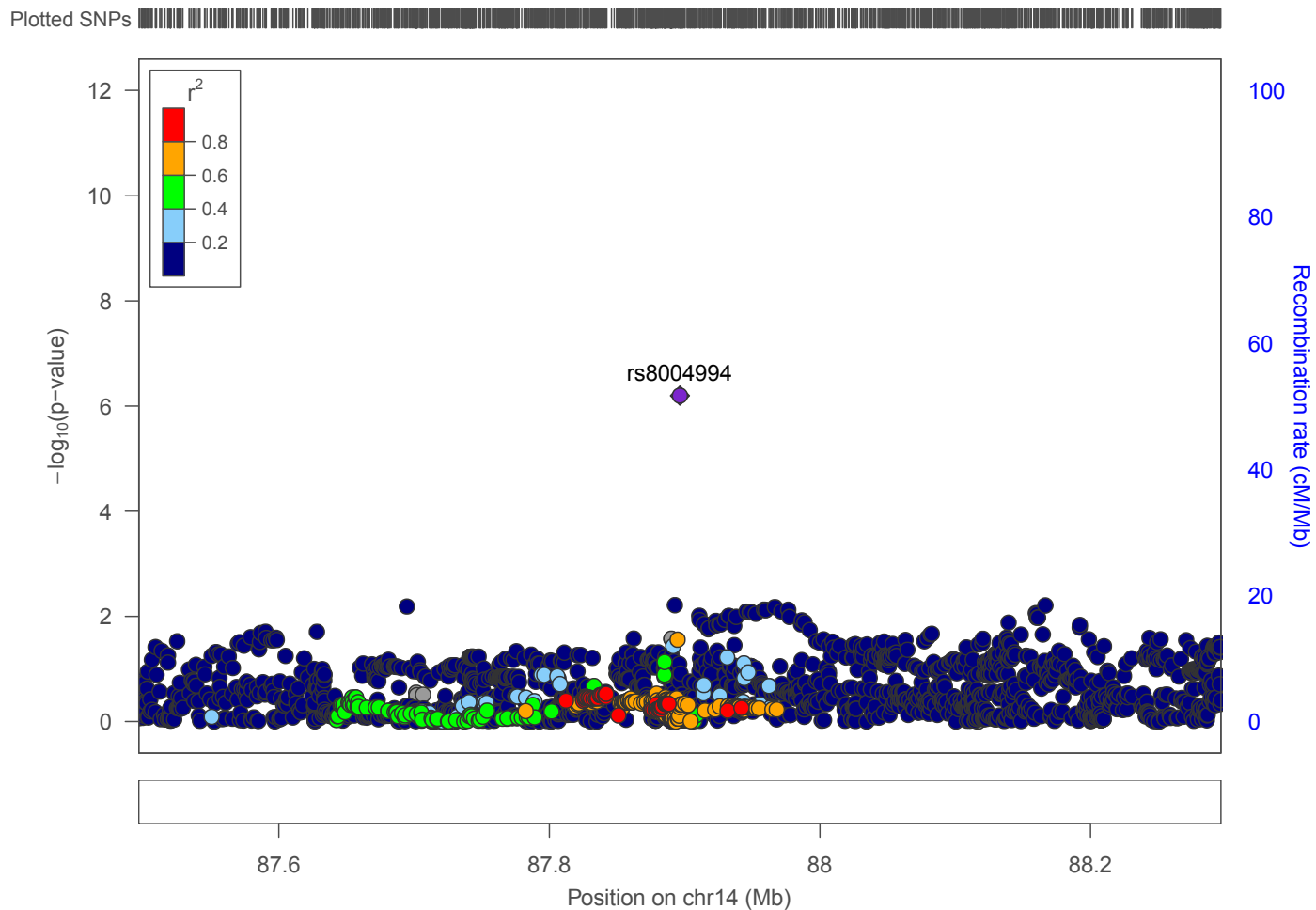
S2. Chromosome 14
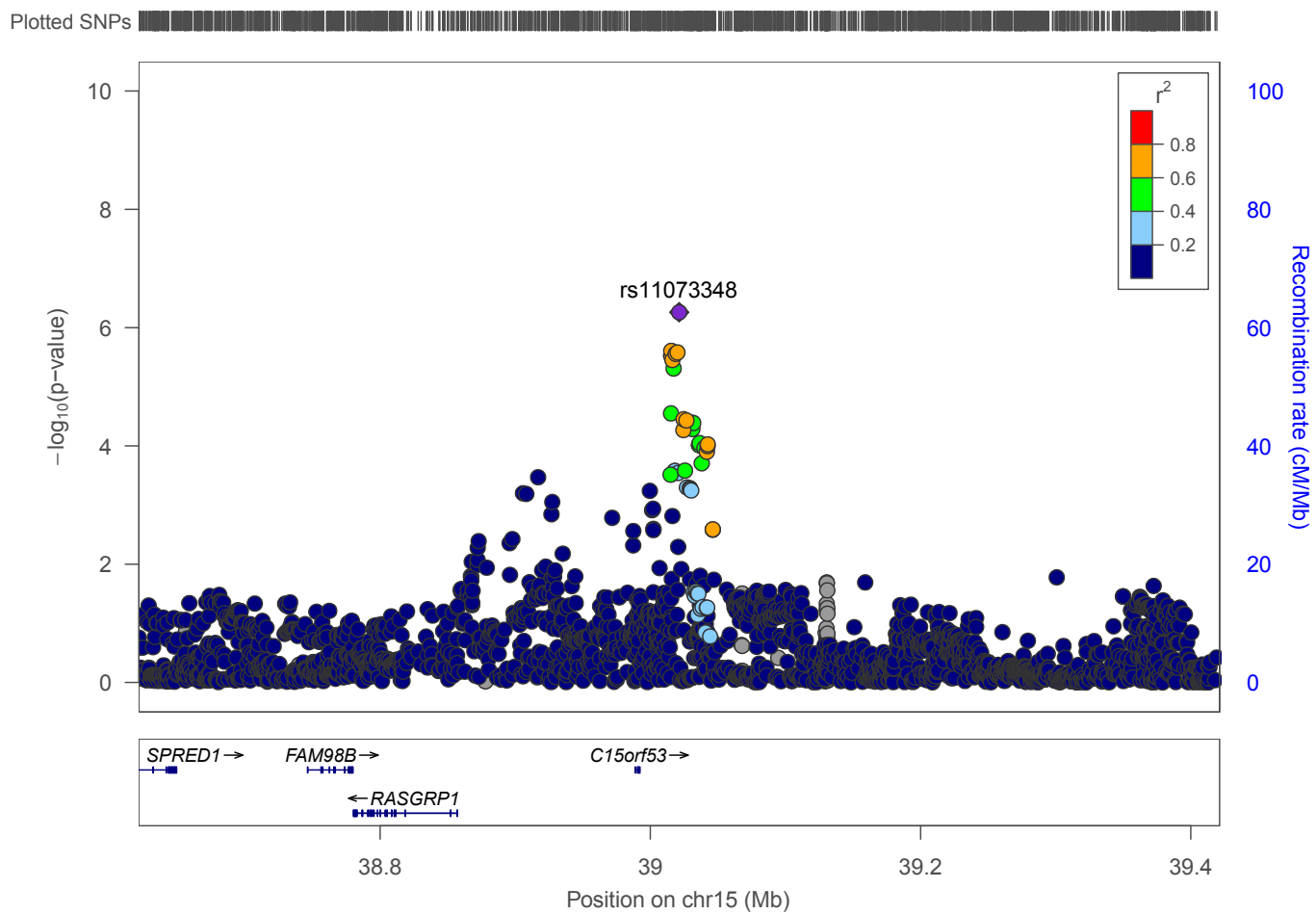
S3. Chromosome 15
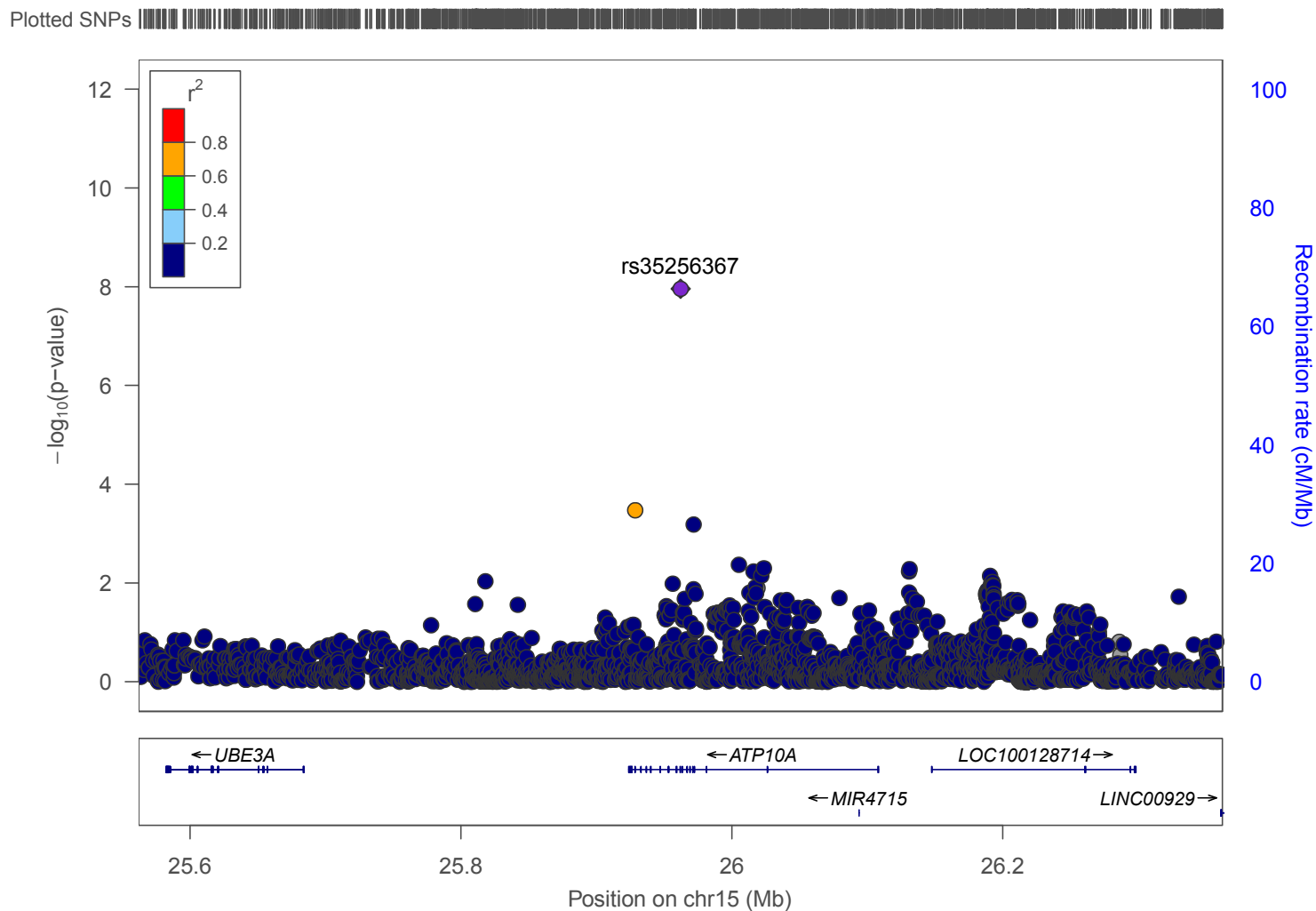
S4. Chromosome 15
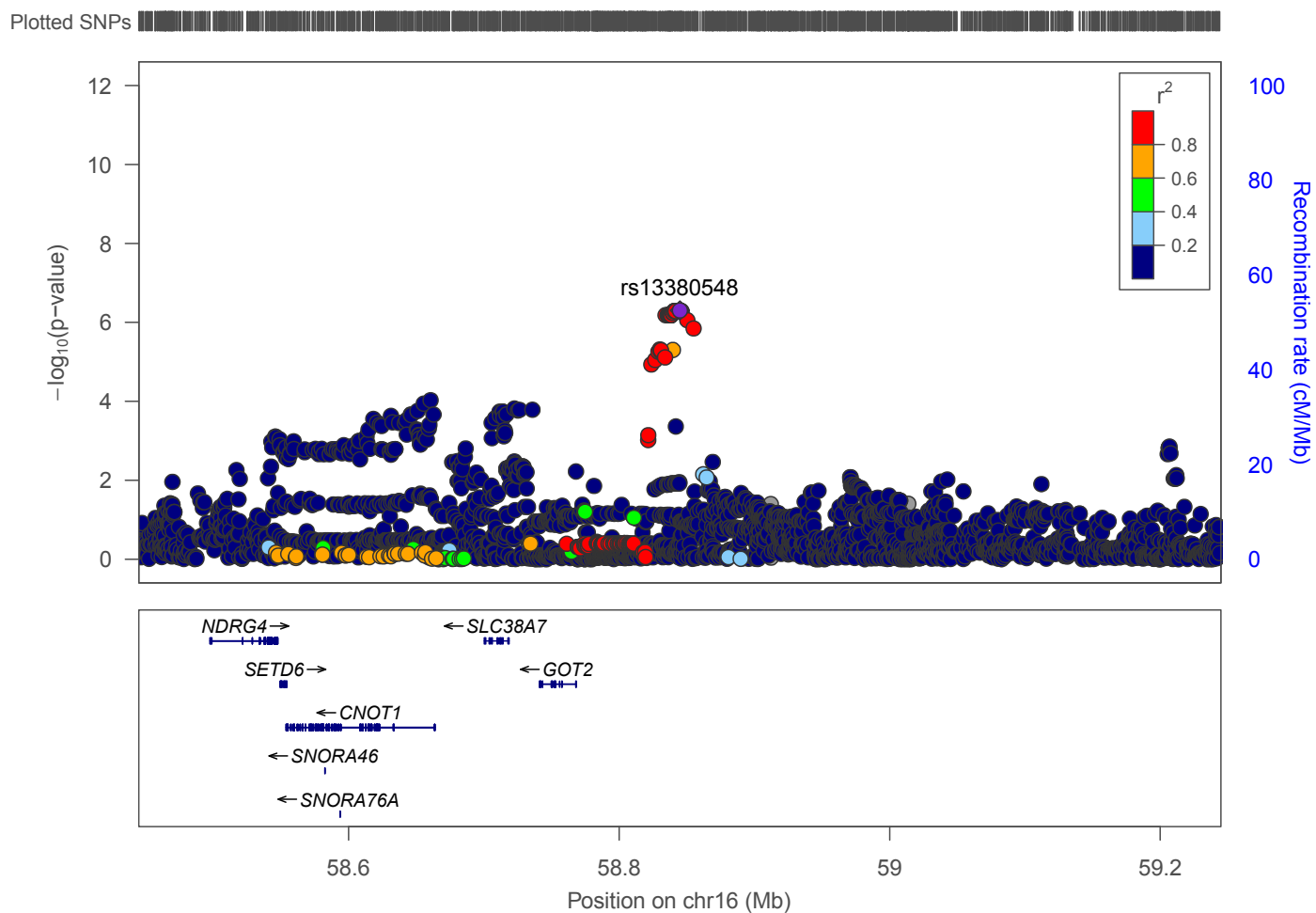
S5. Chromosome 16
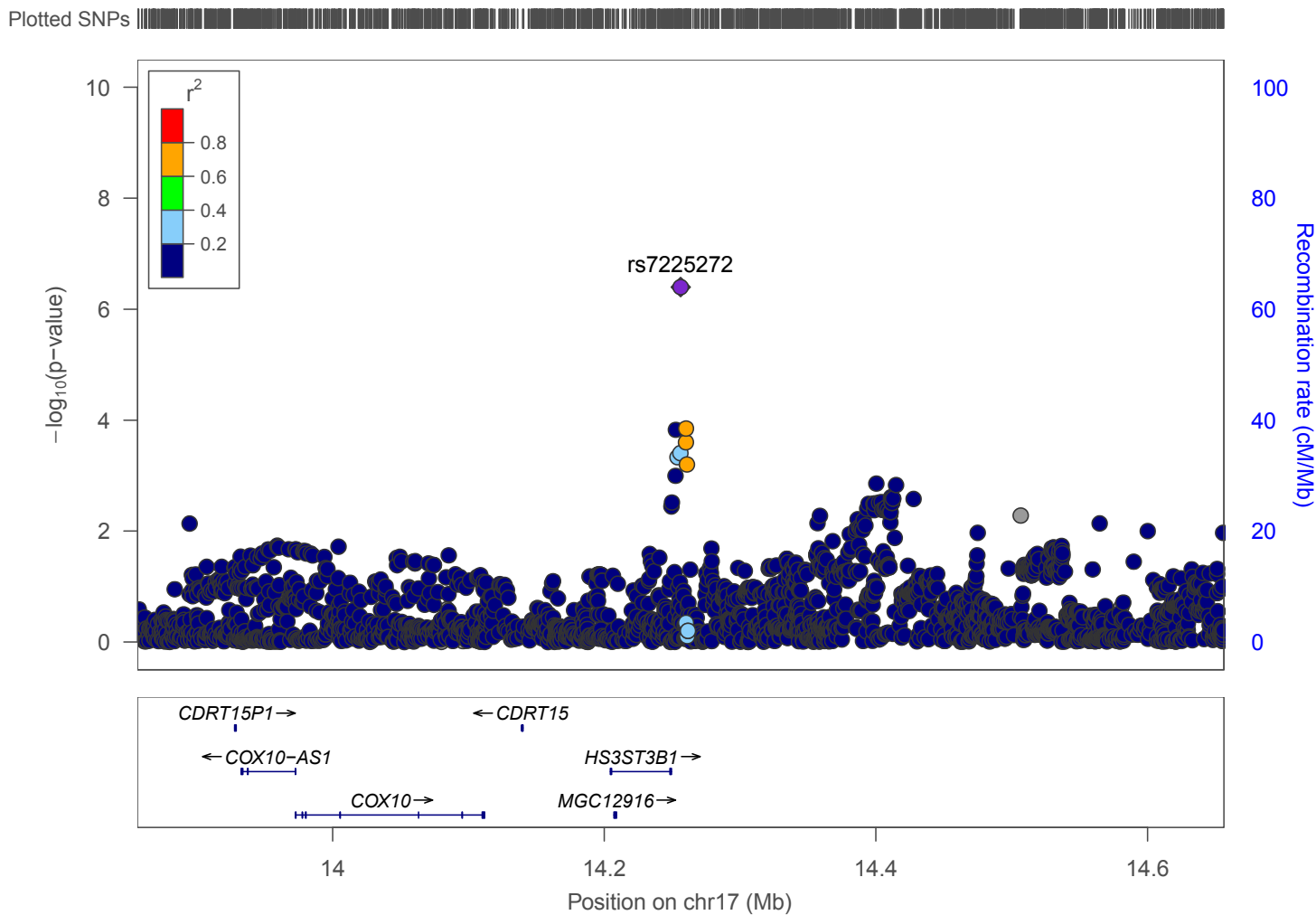
S6. Chromosome 17
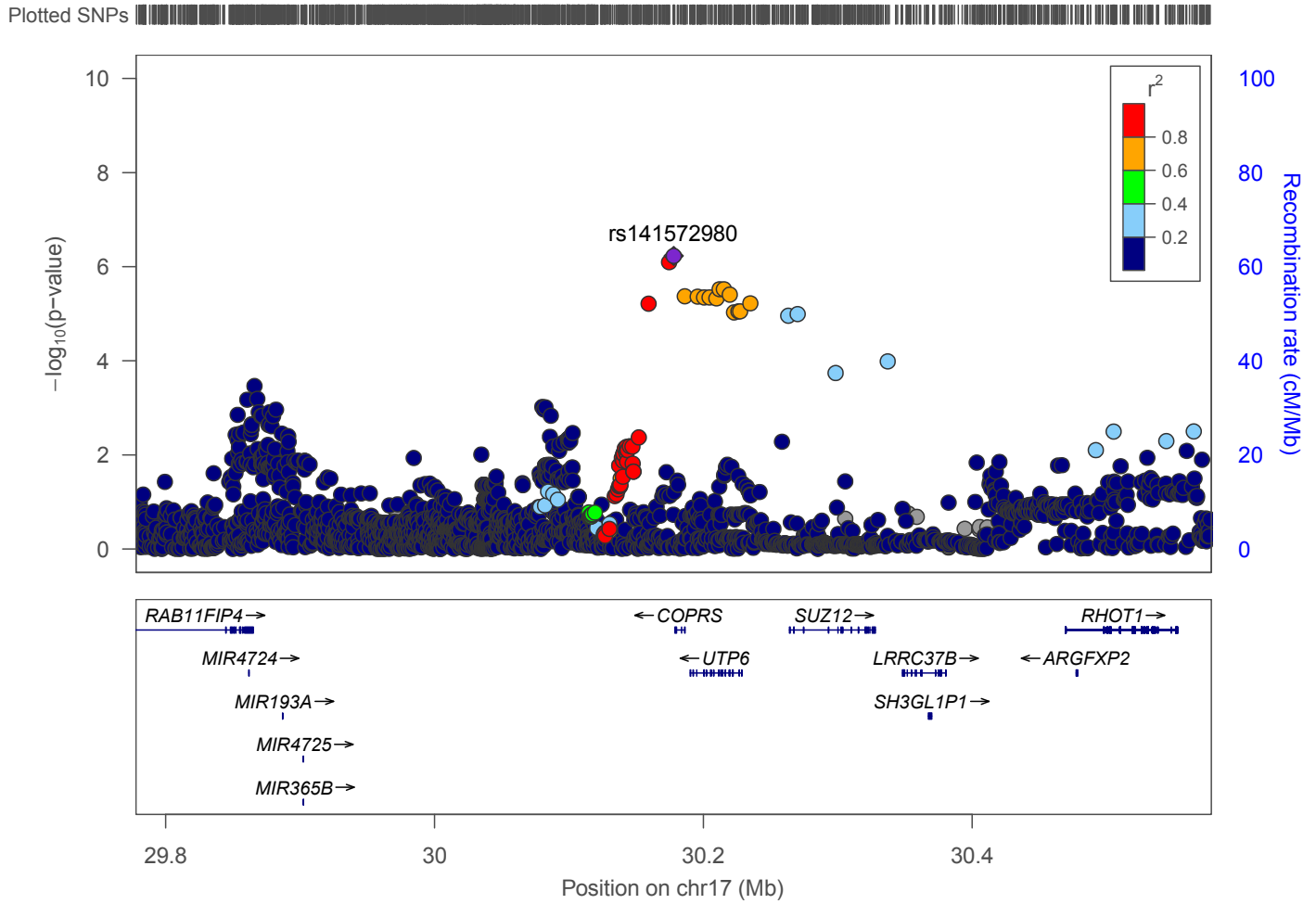
S7. Chromosome 17
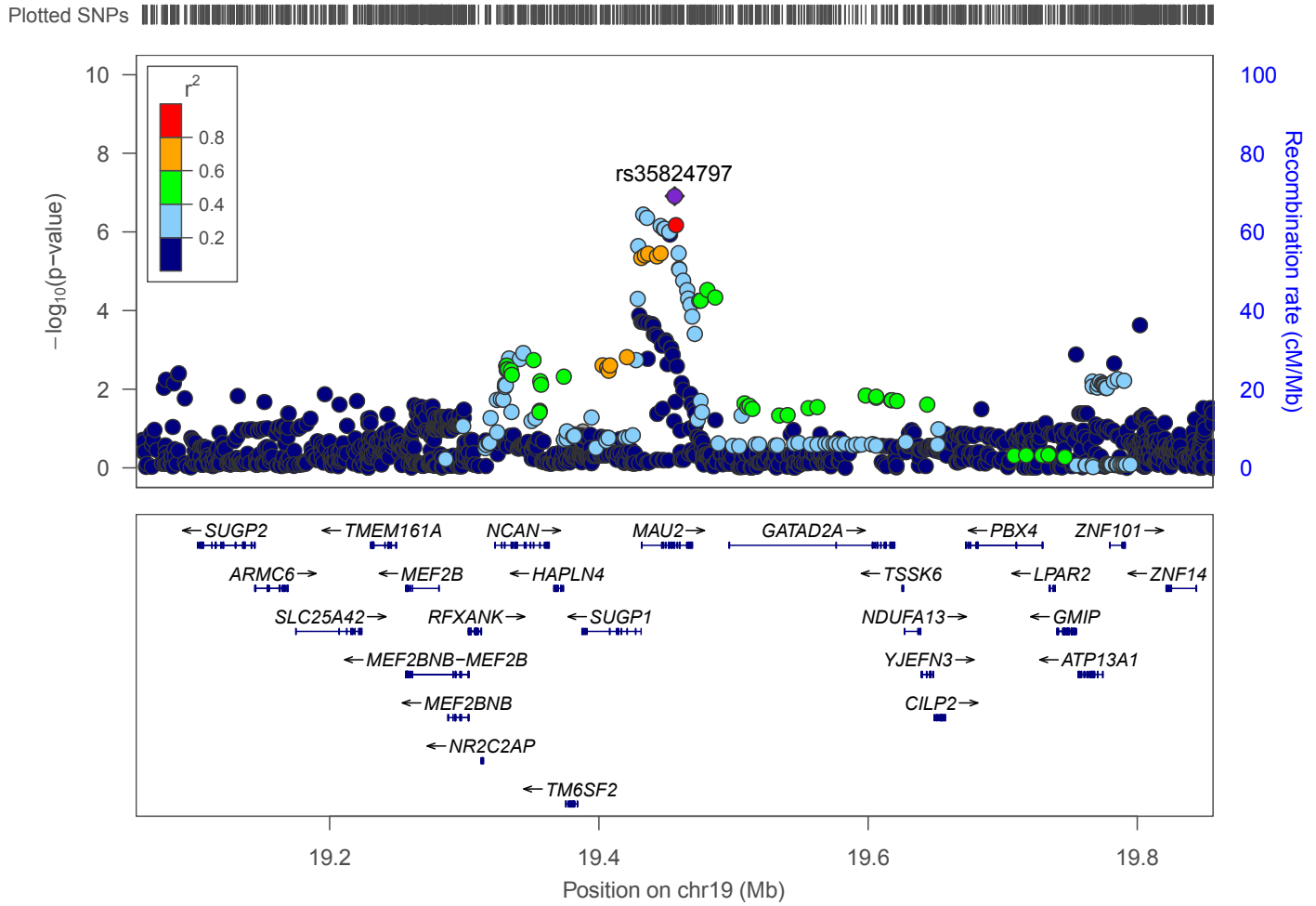
S8. Chromosome 19
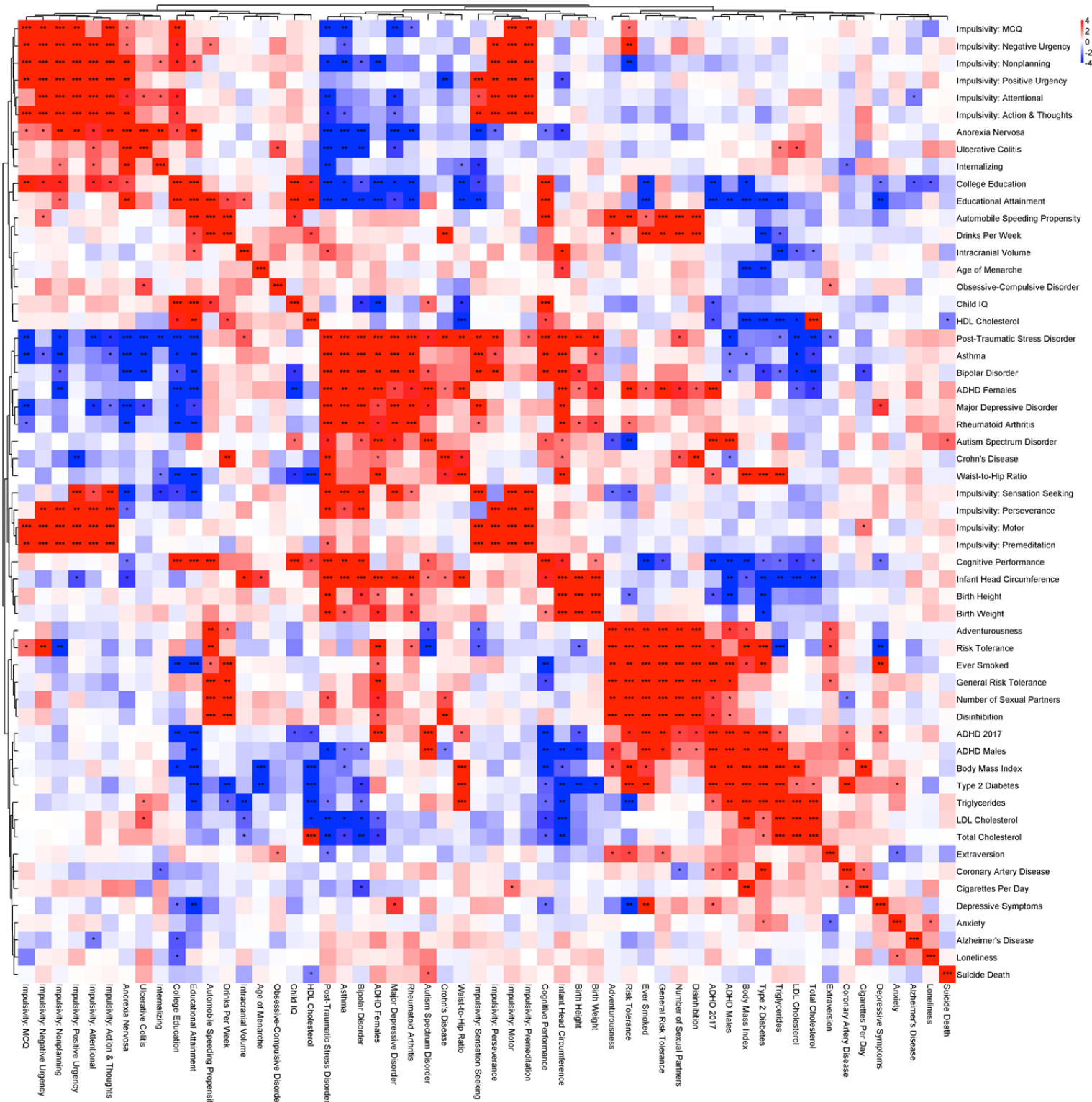
**S9. Suicide PRS Associations with Multiple Medical and Psychiatric PRS.**
Suicide death PRS in cases, based on 10-fold cross validation, regressed onto PRS for medical and psychiatric phenotypes. Shading reflects the test statistic value as positive (red) or negative (blue). Dendrograms based on k-means clustering of Euclidean distances map a general structure of PRS associations. Association of suicide PRS with autism (positive) and HDL (negative) did not survive correction for multiple testing. $p<10^{-2}$ (*), $p<10^{-4}$ (**), and $p<10^{-10}$ (***).
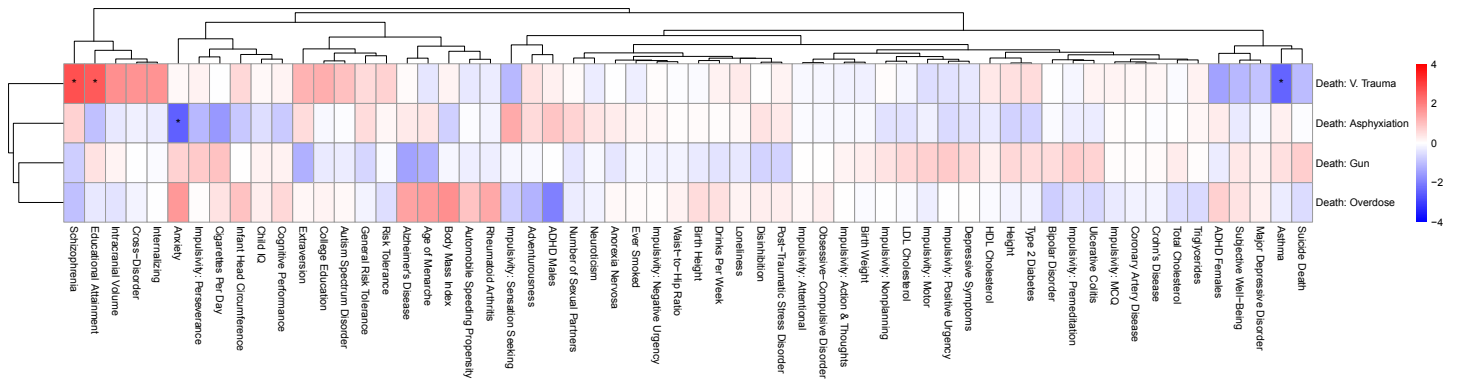
S10. PRS X MD: All Suicide Deaths.
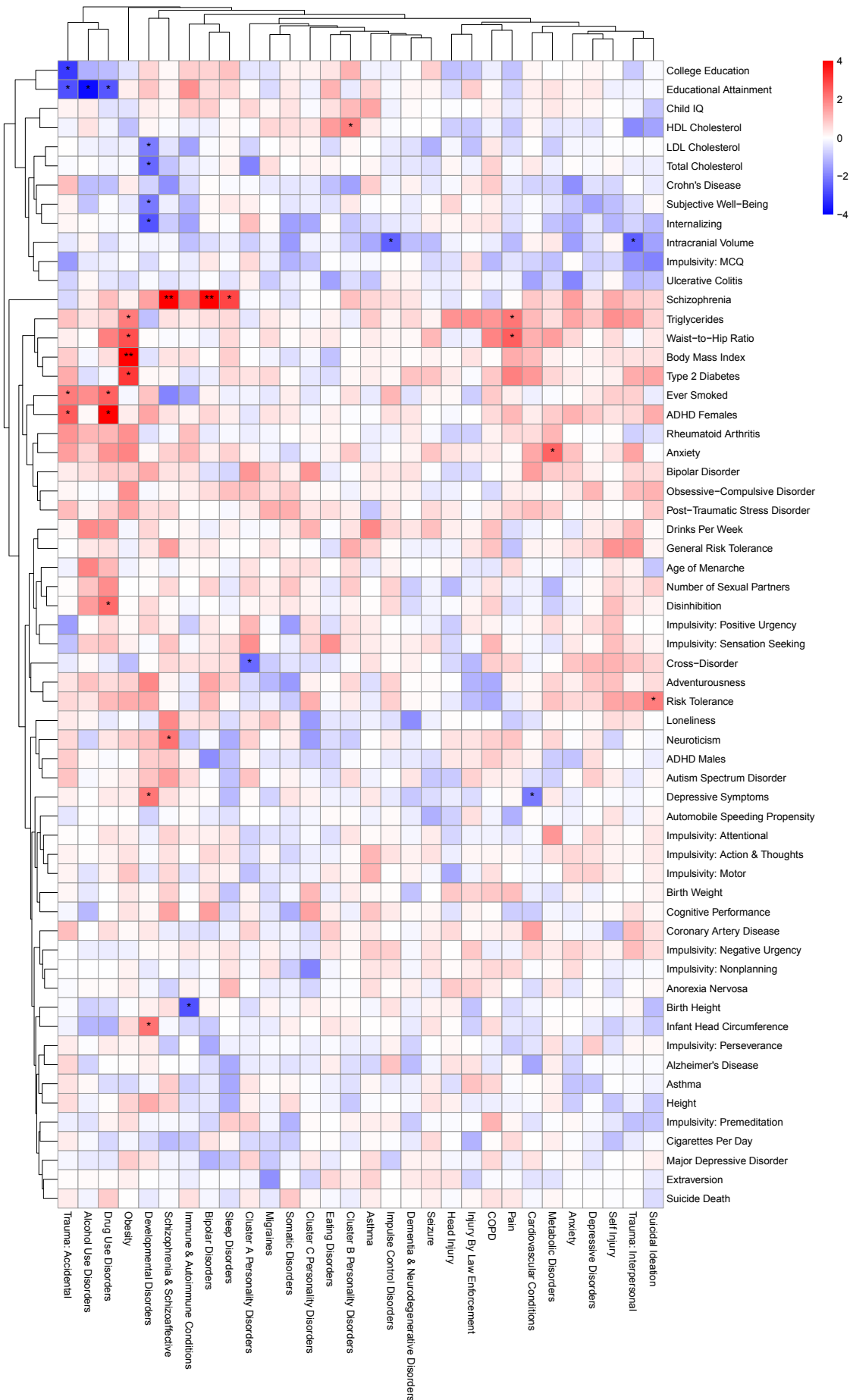
S11. ICD X PRS Plots: All Suicide Deaths > 25 Years of Age.
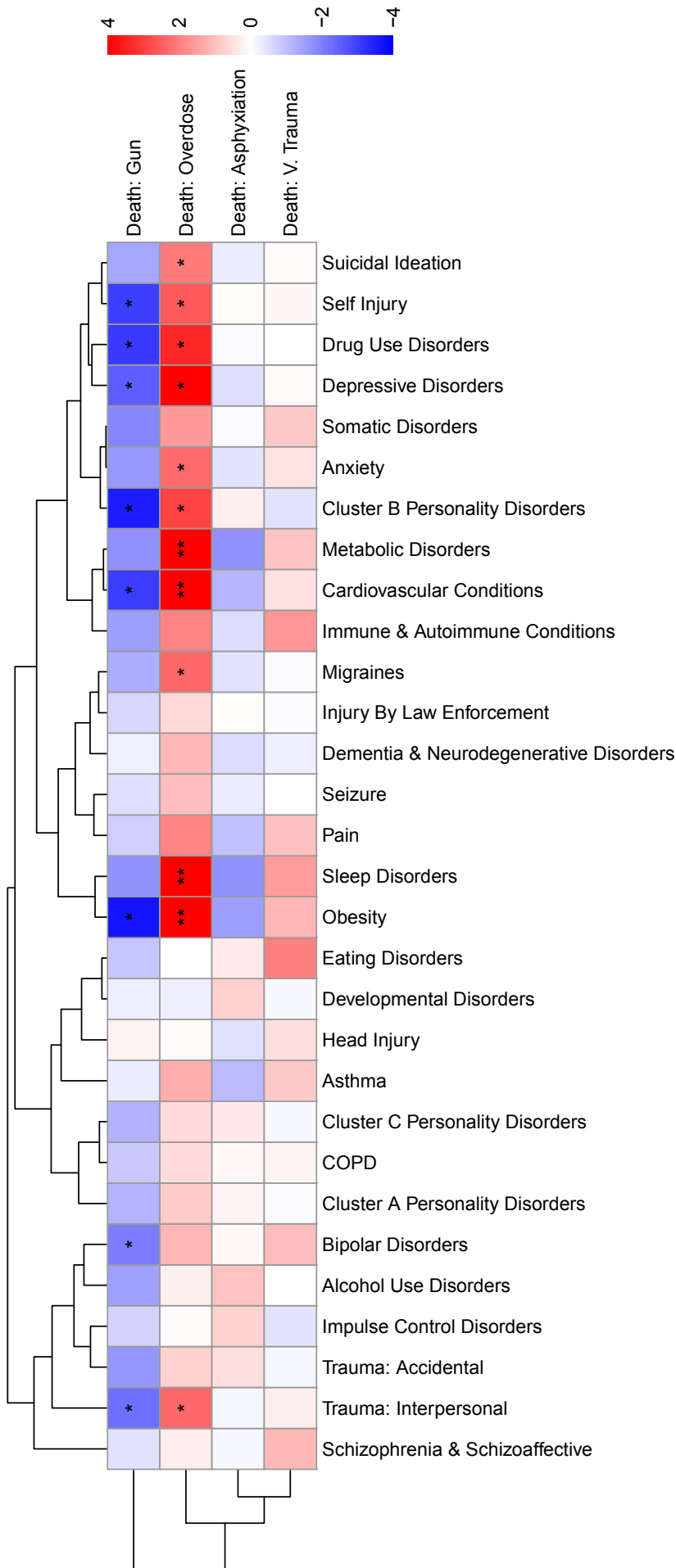
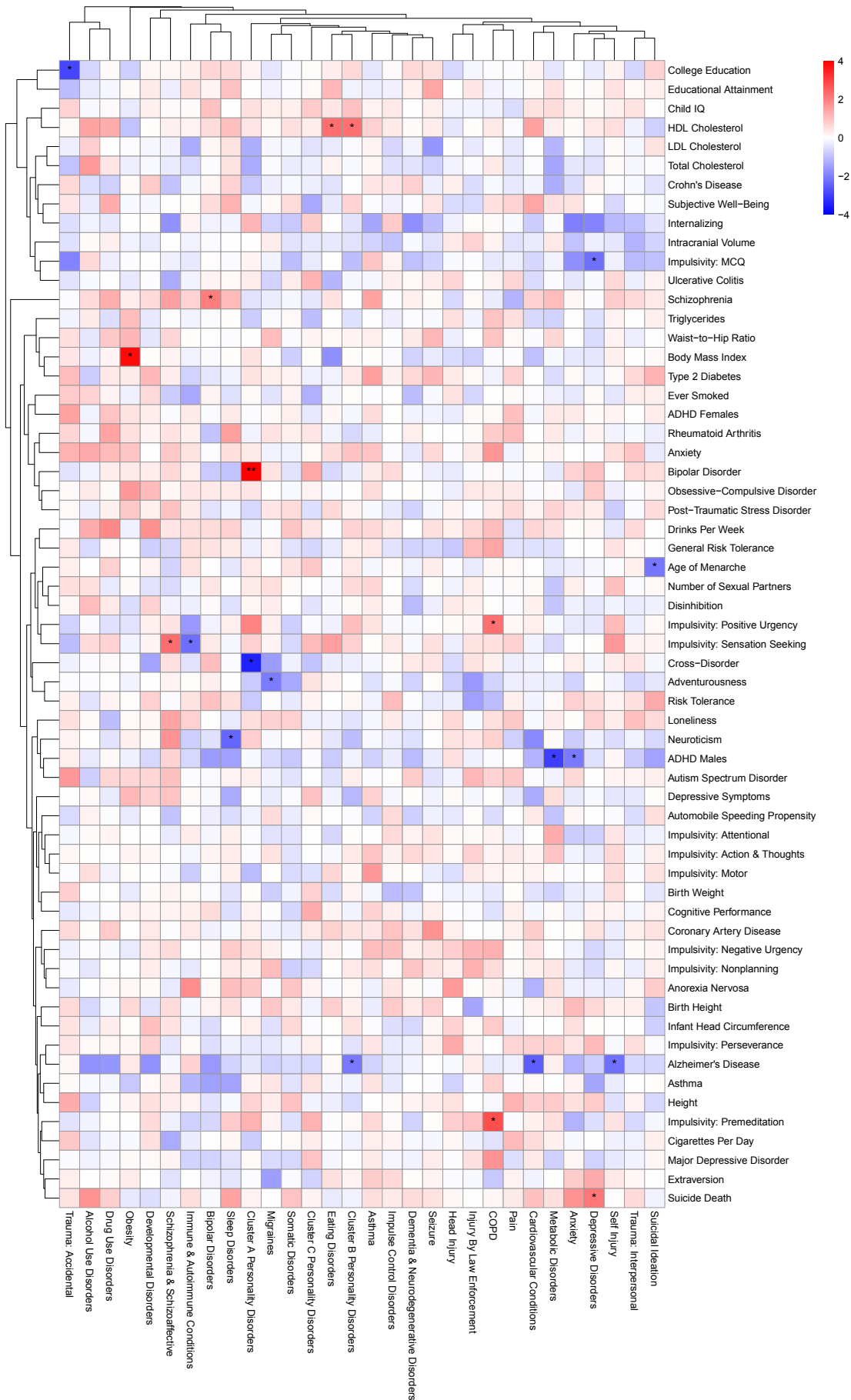S12. ICD X MD Plots: Female Suicide Deaths > 25 Years Of Age.
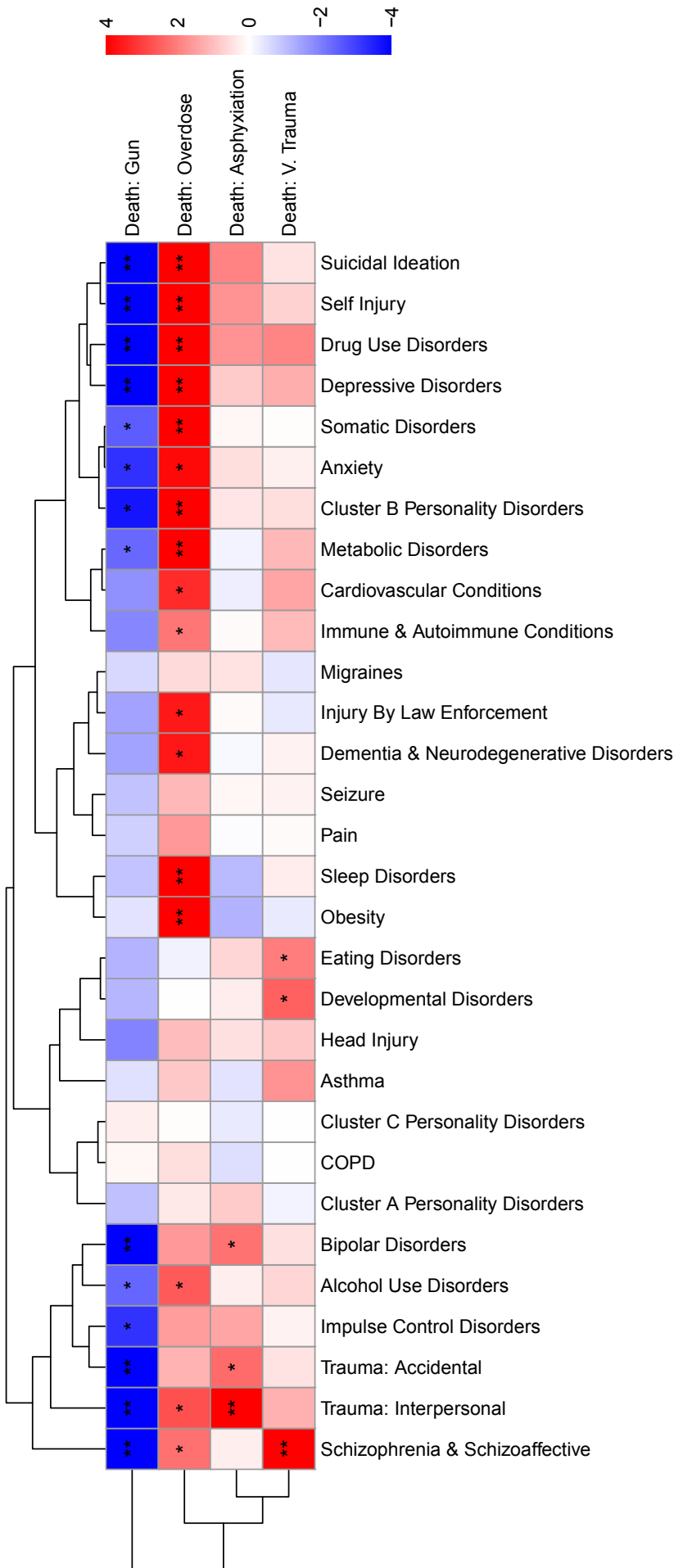
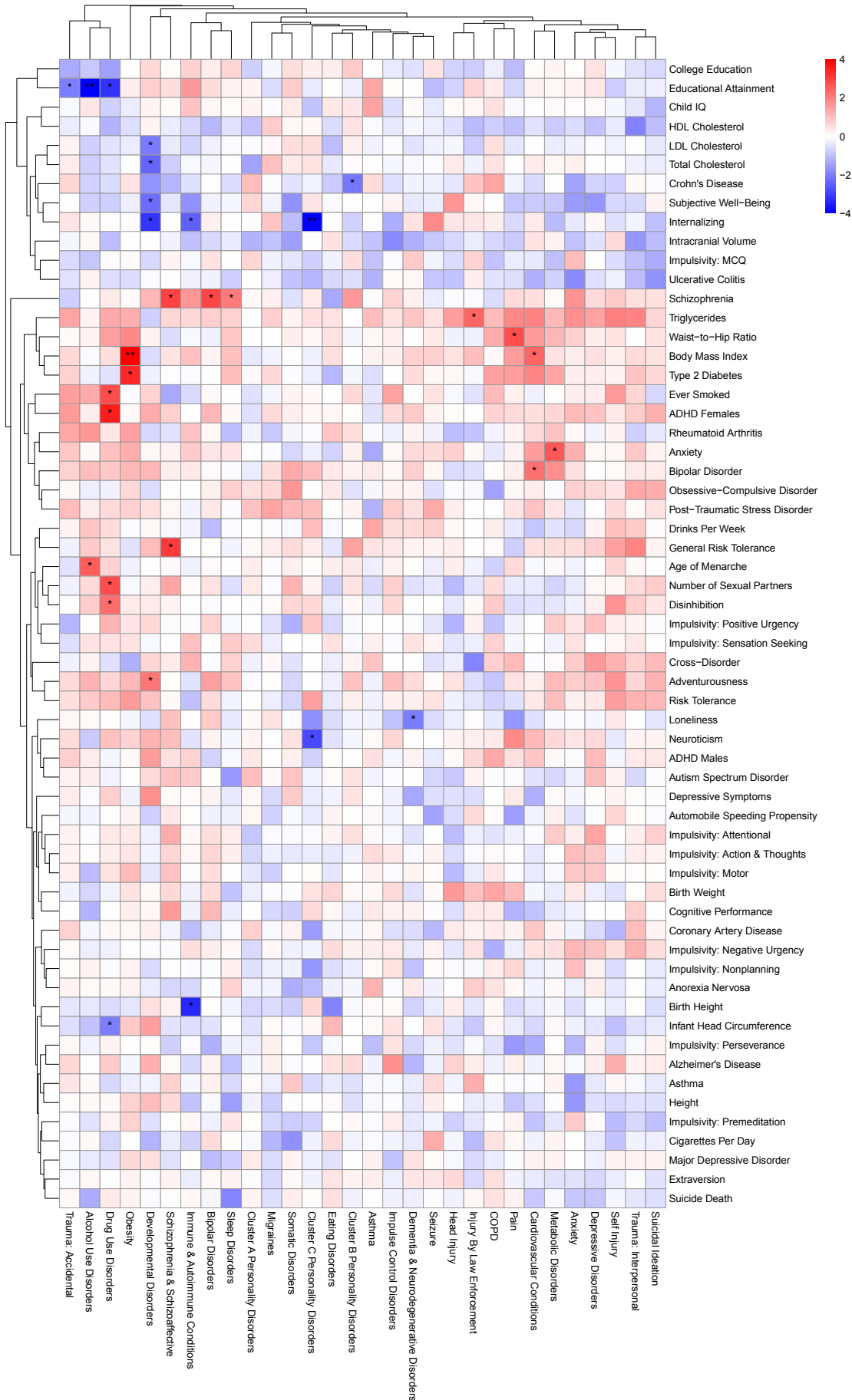S13. ICD X MD Plots: Male Suicide Deaths > 25 Years Of Age.
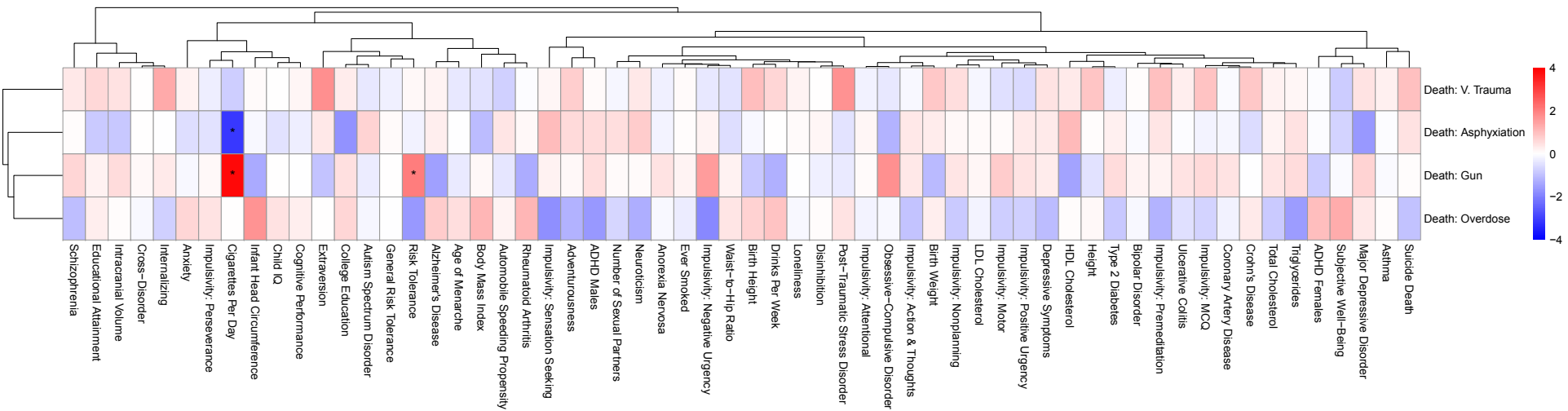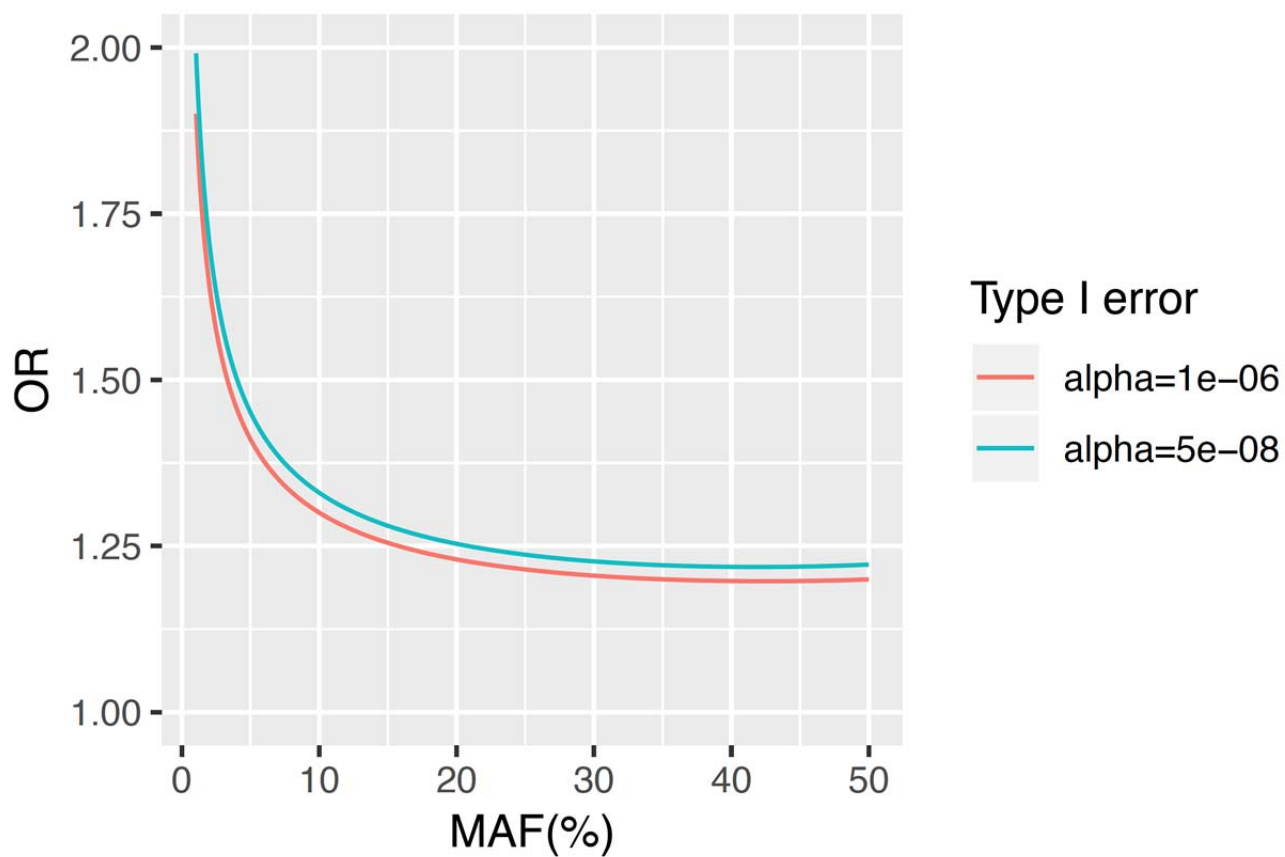
S14. PRS X MD: Female Suicide Deaths.
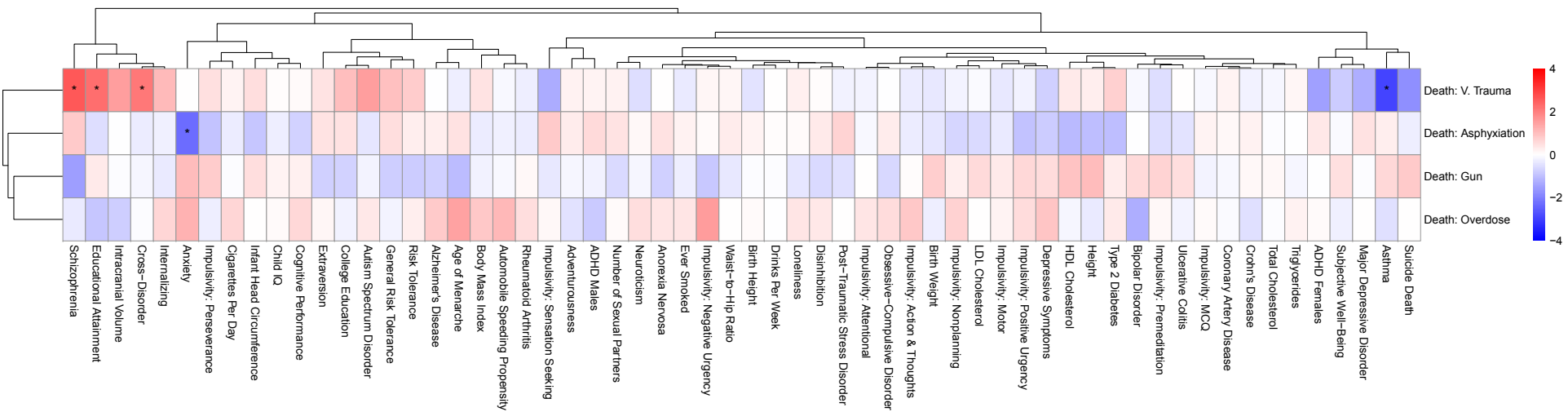
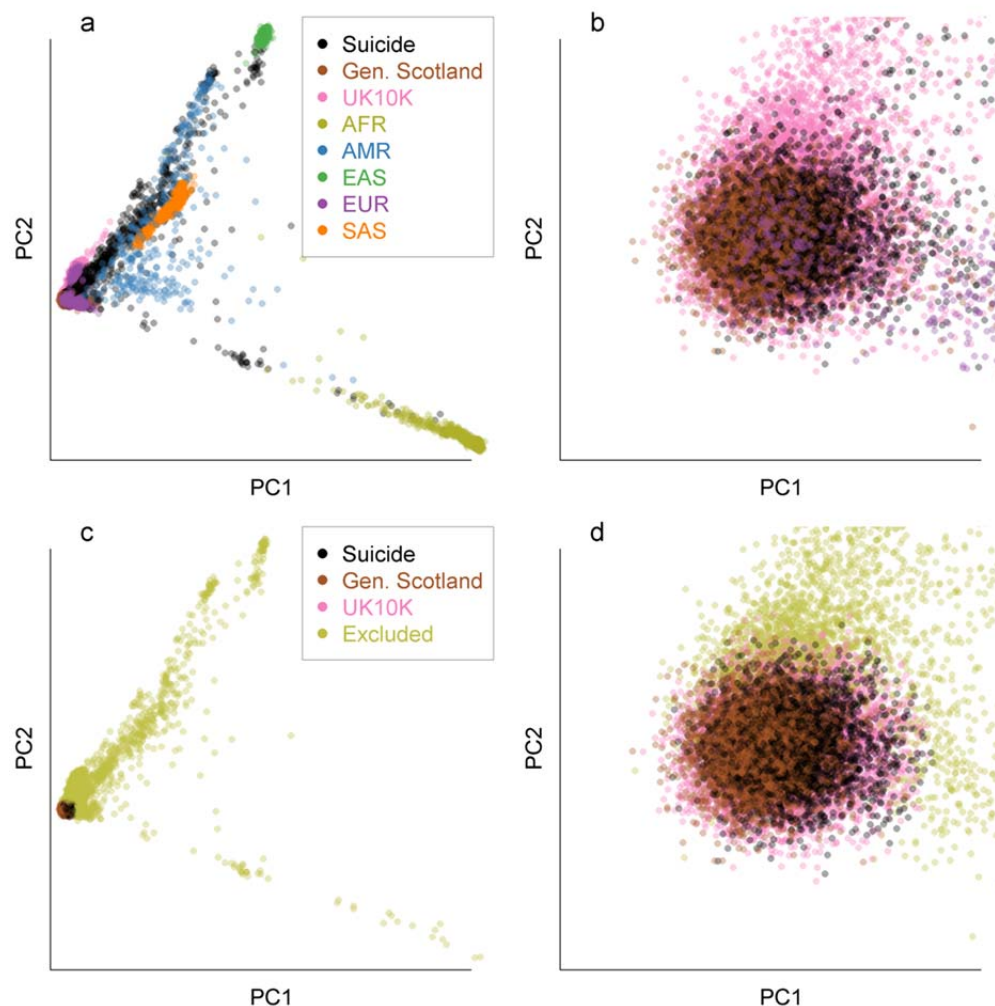**S15. PRS X MD: Male Suicide Deaths.**

S16. ICD X PRS Plots: Female Suicide Deaths > 25 Years Of Age.

S17. ICD X PRS Plots: Male Suicide Deaths > 25 Years Of Age.

S18. Power Plot: Power to detect both genome-wide and nominal levels of significance across MAF.

S19. PCA of 1KG super-populations, cases, controls, and excluded samples. Suicide death samples plotted by the principal component explaining the most variance (PC1) versus the principal component explaining the second most variance (PC2) in all 1KG super-populations (a) and focusing on only the Northern European cases and controls (b). (c) and (d) highlight excluded cases. For adequate statistical power, we examined only cases of Northern European ancestry. However, it is clear from (a) and (c) that the cohort was comprised of multiple ancestries and that research on suicide death in non-European ancestries will reflect an important step beyond this first study.