RESEARCH ARTICLE

American Journal of medical genetics B Neuropsychiatric Genetics WILEY

Rare protein-coding variants implicate genes involved in risk of suicide death

Emily DiBlasi[1,2] | Andrey A. Shabalin[1,2] | Eric T. Monson[1,2] |
Brooks R. Keeshin[1,2,3,4] | Amanda V. Bakian[1,2] | Anne V. Kirby[5] | Elliott Ferris[6] |
Danli Chen[1,2] | Nancy William[1,2] | Eoin Gaj[1,2] | Michael Klein[7] |
Leslie Jerominski[1,2] | W. Brandon Callor[8] | Erik Christensen[8] | Ken R. Smith[9] |
Alison Fraser[9] | Zhe Yu[9] | Douglas Gray[1,2] | PsychChip Investigators of the
Psychiatric Genomics Consortium | Nicola J. Camp[10] | Eli A. Stahl[11,12] |
Qingqin S. Li[13] | Anna R. Docherty[1,14,2] | Hilary Coon[1,2]

[1]Department of Psychiatry, University of Utah School of Medicine, Salt Lake City, Utah

[2]University of Utah Health, Huntsman Mental Health Institute, Salt Lake City, Utah

[3]Department of Pediatrics, University of Utah, Salt Lake City, Utah

[4]Safe and Healthy Families, Primary Children's Hospital, Intermountain Healthcare, Salt Lake City, Utah

[5]Department of Occupational & Recreational Therapies, University of Utah, Salt Lake City, Utah

[6]Department of Neurobiology & Anatomy, University of Utah School of Medicine, Salt Lake City, Utah

[7]Health Sciences Center Core Research Facility, University of Utah, Salt Lake City, Utah

[8]Utah State Office of the Medical Examiner, Utah Department of Health, Salt Lake City, Utah

[9]Pedigree & Population Resource, Huntsman Cancer Institute, University of Utah, Salt Lake City, Utah

[10]Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, Utah

[11]Pamela Sklar Division of Psychiatric Genomics, Icahn School of Medicine at Mount Sinai, New York, New York

[12]Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts

[13]Neuroscience Data Science, Janssen Research & Development LLC, Titusville, New Jersey

[14]Virginia Institute for Psychiatric & Behavioral Genetics, Virginia Commonwealth School of Medicine, Richmond, Virginia

Correspondence
Emily DiBlasi, Department of Psychiatry, University of Utah, 383 Colorow Dr, Salt Lake City, Utah 84108.
Email: emily.diblasi@hsc.utah.edu

Identification of genetic factors leading to increased risk of suicide death is critical to combat rising suicide rates, however, only a fraction of the genetic variation influencing risk has been accounted for. To address this limitation, we conducted the first comprehensive analysis of rare genetic variation in suicide death leveraging the largest suicide death biobank, the Utah Suicide Genetic Risk Study (USGRS). We conducted a single-variant association analysis of rare (minor allele frequency <1%) putatively functional single-nucleotide polymorphisms (SNPs) present on the Illumina PsychArray genotyping array in 2,672 USGRS suicide deaths of non-Finnish European (NFE) ancestry and 51,583 NFE controls from the Genome Aggregation Database. Secondary analyses used an independent control sample of 21,324 NFE

controls from the Psychiatric Genomics Consortium. Five novel, high-impact, rare SNPs were identified with significant associations with suicide death (*SNAPC1*, rs75418419; *TNKS1BP1*, rs143883793; *ADGRF5*, rs149197213; *PER1*, rs145053802; and *ESS2*, rs62223875). 119 suicide decedents carried these high-impact SNPs. Both *PER1* and *SNAPC1* have other supporting gene-level evidence of suicide risk, and psychiatric associations exist for *PER1* (bipolar disorder, schizophrenia), and for *TNKS1BP1* and *ESS2* (schizophrenia). Three of the genes (*PER1*, *TNKS1BP1*, and *ADGRF5*), together with additional genes implicated by genome-wide association studies on suicidal behavior, showed significant enrichment in immune system, homeostatic and signal transduction processes. No specific diagnostic phenotypes were associated with the subset of suicide deaths with the identified rare variants. These findings suggest an important role for rare variants in suicide risk and implicate genes and gene pathways for targeted replication.

KEYWORDS
genetic risk, rare genetic variation, suicide

1 | INTRODUCTION

Nearly 800,000 preventable suicide deaths occur each year (WHO, 2019). While environmental factors are undeniably important, evidence suggests that genetic factors play a major role in suicide death, with estimated heritability from meta-analyses of close to 50% (McGuffin, Marusic, & Farmer, 2001; Pedersen & Fiske, 2010). The identification of suicide genetic risk factors is critical to understanding the biological basis of suicide risk and improving prevention strategies. However, to date, only a fraction of the genetic variation influencing suicide risk has been accounted for with most studies focusing on suicidal behaviors (e.g., ideation or attempt) as opposed to suicide death (Erlangsen et al., 2018; Mullins et al., 2019).

Most genetic studies of suicide focus on common genetic risk variants (variants carried by more than 1% of a given population), despite the high likelihood that rare variants (minor allele frequency; MAF <1%) also significantly contribute to suicide risk (Monson et al., 2017; Tombácz et al., 2017). Discoveries of rare genetic risk variants for other complex psychiatric and nonpsychiatric medical conditions [e.g., lipid disorders (Surakka et al., 2015), Alzheimer's disease (Marei et al., 2016), schizophrenia (Bergen et al., 2019), and obesity (Justice et al., 2019; Turcot et al., 2018)] have identified functional gene pathways and now explain a sizeable proportion of the genetic risk for these disorders. Functional variants that alter gene function or expression tend to be rare in populations because natural selection acts to reduce their prevalence (Kryukov, Pennacchio, & Sunyaev, 2007). Discoveries of risk genes and pathways affected by these rare but high-impact variants lead directly to hypotheses regarding risk mechanisms and are critical to the development of targeted interventions.

A primary reason why most research focuses on common variants is that detection and reliable estimates of frequencies of rare variants in both case and control groups requires large genetic studies (Auer & Lettre, 2015; Kosmicki, Churchhouse, Rivas, & Neale, 2016; Nicolae, 2016). Whole-genome sequencing (WGS) and whole-exome sequencing studies are becoming more common, yet costs involved are considerable and these studies often have more modest sample sizes compared with genome-wide association studies (GWAS) using more affordable genotyping arrays. Large genetic datasets from genotyping arrays are an overlooked source of important rare variation. Although most of the content on genotyping arrays is focused on common intergenic variants, carefully chosen, medically relevant, rare variants are often also included on these platforms.

The large, population-based genetic dataset of suicide deaths available in the USGRS provides unique opportunities to uncover both common and rare variants leading to risk. For this analysis, we studied 4,382 Utah suicide deaths genotyped with the Illumina PsychArray BeadChip, a high-density, genome-wide microarray developed specifically for large-scale genetic studies of psychiatric disorders. The PsychArray includes 265,000 common single-nucleotide polymorphisms (tag-SNPs), as well as 245,000 markers from Illumina's Exome BeadChip, developed to capture coding variants that alter gene function, and 50,000 additional markers associated with psychiatric conditions (https://www.illumina.com/products/by-type/microarray-kits/infinium-psycharray.html). These latter markers were included on the array specifically for their potential functional impact and relevance to psychiatric disease.

Modern GWAS focus on the detection of common genetic risk factors, and SNPs with MAF <1%, such as the rare variants included on the PsychArray that are the focus of this study, are traditionally removed as being less informative. However, the removal of content

specifically chosen for both functional and disease relevance represents a missed opportunity for the discovery of rare, more highly penetrant risk factors for complex psychiatric traits. Although the PsychArray is a popular genotyping platform, there has been no large-scale effort to evaluate its rare variant content for psychiatric or suicide risk. This analysis requires special methodological considerations, including rigorous attention to quality control issues followed by comparison with external control data as well as secondary molecular validation of variants of interest.

To investigate the role of rare variants in suicide death, we first identified significantly associated rare variant content on the PsychArray using a case–control design, comparing USGRS cases against publicly available control data from the large Genome Aggregation Database (gnomAD) resource. Next, we prioritized rare variants by comparison with an independent control cohort genotyped on the PsychArray, additionally including only variants with validation in the USGRS with sequencing data. A follow-up literature search and gene pathway analyses of these variants identified important mechanistic pathways and cross-trait associations. Finally, we characterized the phenotypic attributes of all suicide cases with these rare variants using demographic data, autopsy reports, and electronic health records.

2 | METHODS

2.1 | Utah Suicide Genetic Research Study cohort

In Utah, suicide rates have risen steadily for the past 20 years, and Utah is among 10 states with the highest suicide death rate in the nation (CDC/National Center for Health Statistics, 2019). The USGRS has a biospecimen resource of >6,500 population-ascertained DNA samples from suicide decedents. Cases are about 78% male. This resource grows by ~650–700 cases per year reflective of the rate of suicide in Utah. Samples are ascertained through a two-decade collaboration with the single centralized Utah Office of the Medical Examiner (OME), ensuring consistency of suicide determination, and facilitating near-complete ongoing state-wide collection. Collection of de-identified DNA samples from suicide deaths is done with Institutional Review Board (IRB) approvals from the University of Utah, the Utah Department of Health, and Intermountain Healthcare. DNA is extracted from blood using the Qiagen Autopure LS automated DNA extractor (www.qiagen.com). Case identifiers from the OME are then securely linked to the Utah Population Database (UPDB), which includes demographic, genealogical, geographic, and medical information for over 11 million individuals. Medical information includes determination of medical conditions by OME reports (e.g., autopsy, toxicology report, interview with family members) and electronic medical record (EMR) data, specifically International Classification of Diseases (ICD) diagnostic codes (ICD-9/ICD-10). EMR data come from three sources in the UPDB and include data from all statewide inpatient and ambulatory care encounters from the Utah State Health Department, and data from outpatient encounters from the two

largest clinical data providers, University of Utah Healthcare and Intermountain Healthcare. After linkage to the UPDB, data are stripped of identifying information before transfer to the research team. All cases are referenced by anonymous IDs and no contact is possible with living family members. The state-wide sample ascertainment is population-based, and not limited to cases within a research or clinical cohort with a specific psychiatric diagnosis.

2.2 | PsychArray genotyping and quality control

Genotypes from 4,379 suicide cases were generated with the Illumina PsychArray BeadChip. This is a high-density, genome-wide microarray developed specifically for large-scale genetic studies of psychiatric conditions. Genotyping was performed in six batches with duplicate samples across batches to allow for quantification and modeling of any batch effects. Quality control of all PsychArray genotypes was performed in GenomeStudio for each batch separately. SNPs were retained if the GenTrain score was >0.5 and the Cluster separation score was >0.4. The GenTrain score is computed from the GenTrain clustering algorithm in GenomeStudio. This statistical score indicates SNP calling quality and ranges from 0 to 1, with a higher value indicating better SNP calling quality. SNPs were converted to HG19 plus strand format then batches were combined. PLINK (Chang et al., 2015; Purcell et al., 2007) was used to remove duplicated SNPs or individuals, poorly genotyped cases, monomorphic SNPs, and non-autosomal SNPs. A total of 368,152 SNPs were retained after quality control.

2.3 | Ancestry estimation and sample relatedness

Ancestry estimates were computed using ADMIXTURE (Alexander, Novembre, & Lange, 2009). PsychArray genotypes were compared to 1,000 Genomes (https://www.internationalgenome.org/data/) reference panel. Ten percent of Utah suicide cases are not of predominately European ancestry (e.g., Latinx, Asian, African, or of mixed ancestries). Rare variants are expected to display stronger patterns of population stratification than common variants (Mathieson & McVean, 2012). To control for possible allele frequency differences due to ancestry, we therefore confined our analyses to the 2,735 cases that had estimates of at least 90% non-Finnish European (NFE) ancestry (Supplementary Figure 1). This threshold represents a conservative estimate of ancestry as most samples in the USGRS are predominately European; 85% of the USGRS suicide cases are >75% NFE ancestry.

Estimates of pairwise identity by descent (IBD, measured using pi-hat values) were calculated using PLINK. Pairs of related individuals (third degree or closer) were identified with pi-hat values greater than 0.12. Because relatedness can artificially inflate estimates of association, one member of each of the 63 identified related pairs was randomly removed from the NFE subset and 2,672 remaining individuals were assessed.

2.4 | Control data

Control data were downloaded from the Genome Aggregation Database (gnomAD v2.1) (Lek et al., 2016). gnomAD provides allele frequencies from aggregated genomic data from sequencing studies of over 141,000 unrelated individuals. The NFE gnomAD subset excluding cases with neurological phenotypes, referred to as non-neuro (NN), was used in this study. This NFE-NN subset provides a comparison group eliminating possibly confounding cases with neuropsychiatric conditions. Detailed phenotyping, including psychiatric diagnoses, was not available for this resource. Genome and exome VCFs from the gnomAD data were filtered using bcftools (Li, 2011) to keep only SNPs at genomic sites that passed all gnomAD quality filters (see https://macarthurlab.org/2018/10/17/gnomad-v2-1/). Genomic sites identified as multiallelic in gnomAD were also removed. Biallelic SNPs present in both genome and exome datasets and passing QC filters were combined. SNPs with MAF <0.01 (2,813,863 sites) in the combined NFE-NN dataset were matched to filtered PsychArray SNPs, resulting in a set of 36,215 SNPs in common between the two datasets.

2.5 | Variant annotation

Before proceeding with association tests, the 36,215 SNPs in both the PsychArray and gnomAD data were prioritized for functional relevance by annotating all SNPs using Combined Annotation Dependent Depletion (CADD) (Kircher et al., 2014). CADD prioritizes functional, deleterious, and pathogenic coding and noncoding variants across many functional annotations. If variants had at least one of the following CADD annotations they were retained in the analysis: (a) probably or possibly damaging in PolyPhen; (b) deleterious in SIFT; or (c) CADD scaled C-score (phred) ≥ 10 (a score reflecting a deleterious ranking of the variant compared to all possible substitutions in the genome). Higher CADD phred scores indicate an increased likelihood that a variant will have meaningful functional consequences (Tin et al., 2018). In total, 30,377 SNPs were retained after CADD filtering.

2.6 | Single variant analyses

Allele counts in suicide cases and gnomAD controls were compared using Fisher's exact test. Genomic sites having significant p-values after Bonferroni correction for 30,377 multiple comparisons were retained for Stage 2 analyses (see Figure 1). We chose this conservative correction method assuming independence of all comparisons. After this first stage, 27 variants survived multiple testing. Next, in a Stage 2 analysis, we performed an extension of the gnomAD comparisons using an independent control dataset genotyped on the PsychArray. Stage 2 considered only significant results from the gnomAD comparisons, comparing the frequencies of 27 variants in the USGRS suicides to an independent cohort of 21,324 NFE controls. These control samples have limited phenotype information and

were chosen because they do not have any known psychiatric conditions and were all genotyped on the PsychArray platform. Stage 2 control samples are healthy samples ascertained and genotyped by many groups within the Psychiatric Genomics Consortium (PGC) (for details see Supporting Information). SNP allele counts in suicide cases and PGC control individuals were compared using a Fisher's exact test using a replication significance threshold of $p < 2.17E-03$. Genomic sites having significant p-values after Bonferroni correction were retained for Stage 3 validation analysis.

2.7 | Validation of rare variant content

A subset of 286 suicide cases with PsychArray data also has WGS data generated with Illumina technology. USGRS WGS data were jointly processed using a BWA/GATK-Haplotype caller sequence analysis pipeline (Depristo et al., 2011) with control WGS data, including unrelated cases from the 1,000 Genomes CEU cohort, 1,000 Genomes subjects from Great Britain (matching in ancestry to Utah cases), and a control group of Utah elderly (>age 90) healthy subjects. USGRS WGS data had mean coverage of 30×–60×. Data were assessed for duplicate rate, fraction of properly paired reads, absolute and relative number of calls for different variant classes, and the Ts/Tv ratio within coding sequences. These sequence data allowed for direct validation of rare variants where cases had both array and sequence data. If cases carrying alternate alleles surviving Stages 1 and 2 were not present in this WGS dataset but were significant in Stage 1 and 2 analyses, we directly validated the variant using Sanger sequencing (Sanger, Nicklen, & Coulson, 1977). Only variants where presence of the allele was validated in sequence data were considered.

2.8 | Phenotyping

All cases in the USGRS have demographic information and cause of death information from Vital Records data and Medical Examiner autopsy reports. The phenotypic characteristics of the 2,672 individuals used in the rare variant analyses were described by age, sex, and cause of death (Table 2). Additionally, approximately 80% of USGRS cases have EMR data, specifically International Classification of Diseases (ICD) diagnostic codes (ICD-9/ICD-10). Missing EMR data can occur for many reasons (e.g., individuals not seeking medical attention, care received out of the period of UPDB record coverage). We compared EMR data between suicide cases with and without the identified rare risk variants. Suicide cases were compared to each other in order to assess if there was a specific phenotype, or suicide subtype, that was associated with having a rare risk variant. EMR data were available for 2,117 suicide cases without identified risk SNPs and 97 suicide cases with identified risk SNPs. The diagnostic burden was compared across cases with and without rare variants by comparing the mean number of diagnostic codes per individual between the two groups with Welch's t test. Thirty relevant co-occurring psychiatric
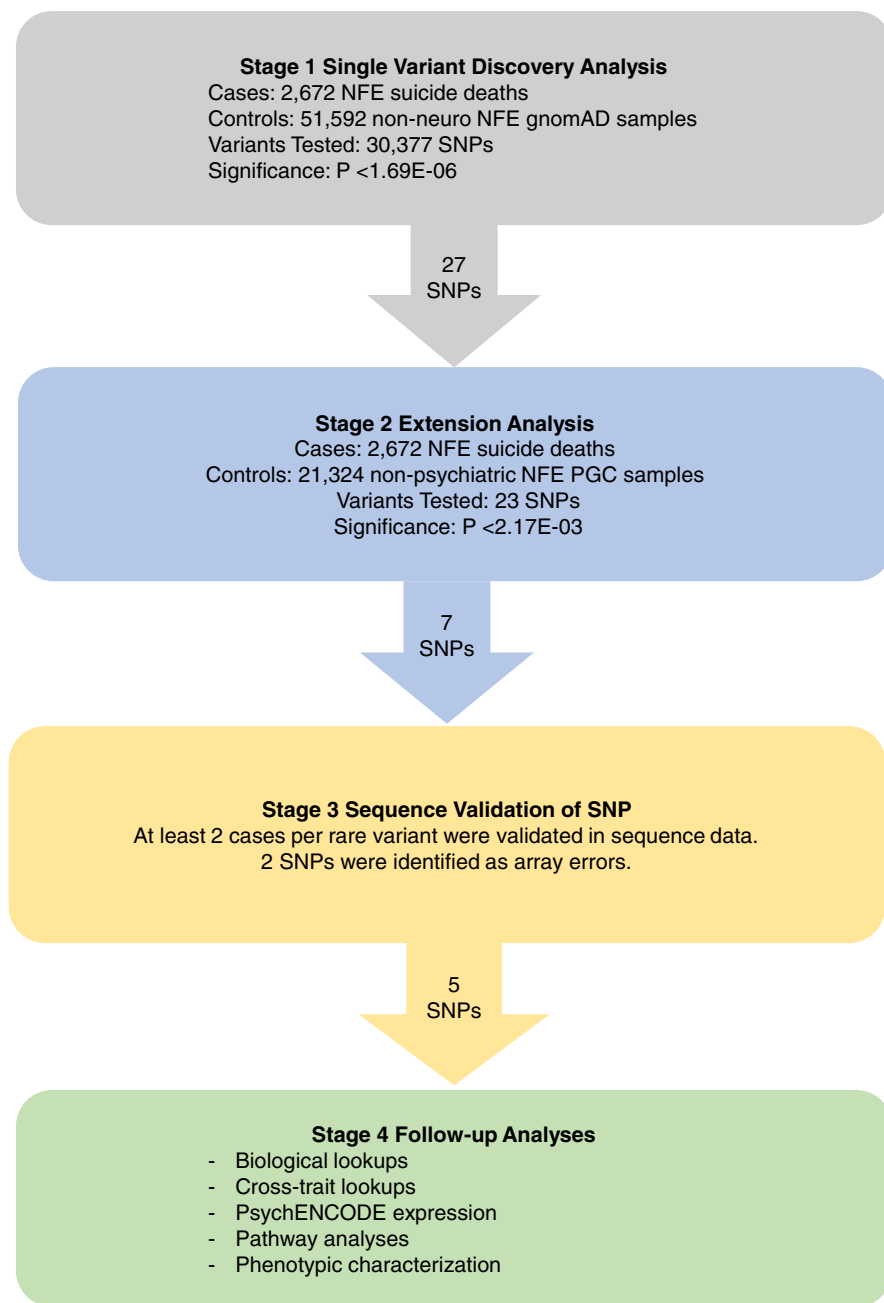
FIGURE 1 Study design and workflow diagram for rare variant analyses. gnomAD, Genome Aggregation Database v2.1; non-neuro, individuals who were not ascertained for having a neurological condition in a neurological case/control study; NFE, non-Finish European

and related medical phenotypes were established by hierarchically aggregating ICD codes in EMR data using the ICD system and with review by clinical experts on our team (A.D., B.K., and E.M.) (Supplementary Table S1). The diagnostic burden within the 30 phenotypes was compared across cases with and without rare variants by comparing the mean number of diagnostic codes per individual between the two groups with Welch's t test. Counts of occurrences of each of the 30 relevant EMR phenotypes were compared between cases with and without identified risk SNPs using Fisher's exact test with a Bonferroni significance threshold of $p < .0017$. We hypothesized that cases with rare variants may have an earlier age of onset of psychiatric and relevant medical conditions compared to cases without rare variants. We chose 10 diagnostic phenotypes that were

highly prevalent in cases with risk SNPs or most related to risk SNP gene function (depression, anxiety, bipolar, schizophrenia/schizoaffective, pain, drug abuse, sleep disorders, suicidal ideation, self-injury, and immune disorders). For these phenotypes, we compared the mean first age of diagnosis between cases with and without risk SNPs using Welch's t test with a Bonferroni significance threshold of $p < .005$.

2.9 | Follow-up analyses

Evidence for supporting clinically relevant trait associations were assessed at both the SNP and gene level. Identified SNPs that

were significant in all three stages were checked against the ClinVar (Landrum et al., 2018), dbSNP (Sherry, 2001), and DisGeNET (Piñero et al., 2017) databases. At the gene level, supporting trait associations were assessed by a search of the literature, using the NHGRI-EBI GWAS Catalog (Buniello et al., 2019), and PsychENCODE differential gene expression (Gandal et al., 2018). Pathway analyses were also conducted at the gene level using FUMA GENE2FUNC software (Watanabe, Taskesen, van Bochoven, & Posthuma, 2017) to determine pathway enrichments in the gene ontology (GO) gene sets compared to all genes in the genome. We assume that, like other psychiatric conditions, both rare and common genetic variants will contribute to the risk of suicide. We wished to determine if the significant rare variants from our analysis added to gene enrichment found from common variant GWAS discovery. Therefore, to understand where these genes fit into the current understanding of pathways potentially contributing to risk, we conducted pathway analyses using genes with rare variants meeting significance and validation in this study and including genes implicated in independent GWAS studies of suicidal behavior. 102 genes from GWAS studies focusing on the discovery of common variants associated with suicide attempt and death that were prioritized by González-Castro et al. (2019) were used in these combined pathway analyses. p-Values for enrichment analyses were adjusted for multiple testing using the Benjamini–Hochberg false discovery rate (FDR) method (Ferreira, 2007).

3 | RESULTS

3.1 | Single-variant analyses

We conducted a single-variant association analysis of 30,377 rare putatively functional SNPs present on the PsychArray genotyping array in 2,672 USGRS suicides of NFE ancestry and 51,583 publicly available NFE controls from gnomAD. The average CADD PHRED score for the 30,377 variants tested was 21.54 (5.96 SD). 2,528 SNPs had nominally significantly elevated allele frequencies in suicide cases compared with gnomAD NFE-NN controls. The average CADD PHRED score for the 2,528 nominally significant variants was 22.05 (6.25 SD). In our primary (stage 1) analysis, we identified 27 functionally annotated PsychArray SNPs with significantly elevated allele frequencies after Bonferroni correction in suicide cases compared with gnomAD NFE-NN controls (Supplementary Table S2). We chose this conservative correction method due to possible array heterogeneity and estimations of increased error rates in genotyping rare variants (Li, 2016).

To control for differences in molecular platforms we performed Stage 2 extension analyses with an independent PsychArray control set. Stage 2 considered only significant results from the Stage 1 comparisons. 23 of the 27 significant SNPs from Stage 1 had available comparable genotypes in the Stage 2 PGC control set comprised of 21,324 NFE controls without psychiatric conditions. Four of the 27 SNPs were not comparable as they did not pass QC thresholds in the PGC control sample. Seven of the 23 tested variants had

significantly elevated allele frequencies in suicide cases compared to Stage 2 PGC controls after Bonferroni correction (Supplementary Table S3).

In Stage 3 analyses, we validated the genotype calls from the PsychArray genotyping platform using sequence data. Two of the seven variants did not validate. Five novel, high-impact, rare SNPs (MAF <1%) survived all stages of our study (Table 1; Figure 2). All five variants had CADD PHRED scores greater than 22 (Table 2) indicative of the top 1% of deleterious substitutions in the human genome (Kircher et al., 2014). These risk variants are protein-coding missense variants with predicted high deleterious impact in *SNAPC1* (rs75418419), *TNKS1BP1* (rs143883793), *ADGRF5* (rs149197213), *PER1* (rs145053802), and *ESS2* (rs62223875). Together, these five rare variants were present in 119 of the 2,672 suicide cases of NFE ancestry (4.45% of cases). Supplementary Figure S1 shows odds ratios and confidence intervals for each variant.

3.2 | Trait associations, gene function, expression, and pathway analysis

No prior trait associations were found in the literature or in searched databases for the five specific SNPs identified in this study. At the gene level, two of the identified genes with rare risk variants had prior associations with suicidal behavior [*PER1* (Levey et al., 2016; Sequeira et al., 2012) and *SNAPC1* (Coon et al., 2018)]. Three genes had prior associations with severe psychopathology (*PER1*, *TNKS1BP1*, and *ESS2*) (see Section 4 for details).

FUMA pathway enrichment analyses using only the five final genes remaining after Stage 3 analysis did not implicate any biological pathways. When the five genes were combined with 102 prioritized genes from genome-wide studies of suicidal behavior (Supplementary Table S4), *PER1*, *TNK1S1BP1*, and *ADGRF5* showed significant enrichments in GO biological process gene sets, as follows. *TNKS1BP1* and *ADGRF5* were significant in the Homeostatic Process gene set (M11223). *PER1* was a member of the Regulation of Intracellular Signal Transduction (M11395) pathway. *ADGRF5* has also implicated the Regulation of Immune System Process (M13496) pathway.

3.3 | Phenotypic associations of Utah suicide deaths with significant rare variants

Demographic data (Table 2) and diagnostic data (Table 3) were compared between Utah NFE suicide deaths without rare risk variants and Utah suicide deaths with identified rare risk variants. Overall, no significant differences were found for demographic variables (Table 2). The use of firearms was the most common method of death for suicide cases both with and without rare variants. Occurrences of the 30 EMR diagnostic categories did not significantly differ (Bonferroni significant threshold $p < .0017$) between suicide cases with and without identified risk SNPs (Table 3).

TABLE 1 Rare missense variants associated with suicide death

SNP location (GRCh37)	Reference allele	Alternate allele	Amino acid substitution	Rs#	Gene name	CADD PHRED score	gnomAD MAF[a]	Suicide MAF[b]	gnomAD reference allele count	gnomAD alternate allele count	Suicide reference allele count	Suicide alternate allele count	p-value[c]	# suicide cases
17:8045136	G	A	Ala ->Val, missense, coding	rs145053802	PER1	23.1	4.48E-04	3.56E-03	102,634	46	5,325	19	2.06E-10	19
6:46851334	C	T	Ala ->Thr, missense, coding	rs149197213	ADGRF5 (GPR116)	24.1	8.82E-04	4.68E-03	101,994	90	5,317	25	3.94E-10	25
11:57070069	C	T	Ser ->Asn, missense, coding	rs143883793	TNKS1BP1	23.7	1.58E-03	5.43E-03	102,618	162	5,313	29	8.14E-08	28[d]
14:62244854	C	T	Thr ->met, missense, coding	rs75418419	SNAPC1	22.3	2.05E-03	6.18E-03	101,432	208	5,311	33	1.74E-07	33
22:19127403	C	T	Ala ->Thr, missense, coding	rs62223875	ESS2 (DGCR14)	22.8	4.75E-04	2.81E-03	103,117	49	5,327	15	3.80E-07	15

[a]Minor allele frequency in gnomAD v2.1 non-Finnish-European, non-neuro subset.
[b]Minor allele frequency in non-Finnish-European suicide cases.
[c]p-value <1.7E-6 (Bonferroni correction for 30,377 tests).
[d]One individual was homozygous for the alternate allele.

TABLE 2 Descriptive data for 2,672 Utah suicide deaths with and without any of the five identified risk SNPs

	Cases without risk SNPs	Cases with risk SNPs
# Individuals	2,553	119
Mean age at death (*SD*)	41.42 (17.63)	40.71 (17.74)
Male	0.79	0.82
Female	0.21	0.18
Firearm death	0.54	0.53
Asphyxiation death	0.30	0.30
Overdose/poison death	0.12	0.13
Violent trauma death	0.03	0.03
Other methods of death	0.01	0.01

No significant differences were found between groups.
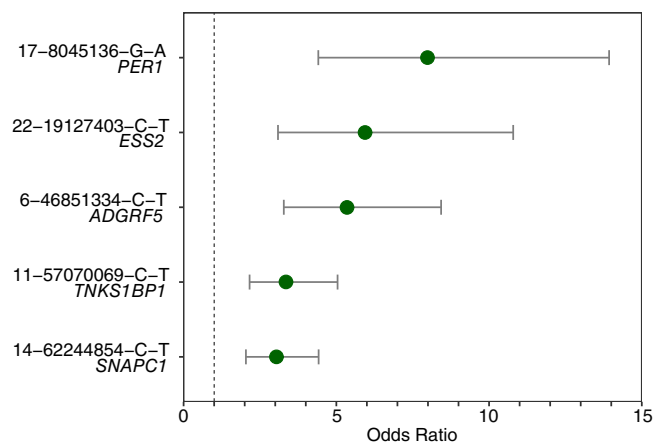
FIGURE 2 Estimated odds ratios and 95% confidence intervals of significant variants in suicide cases compared to PGC controls. Dashed line is placed at 1. Variant labels are in the format chromosome, position, reference allele, alternate allele

The mean number of all EMR diagnoses did not differ between suicide cases without risk SNPs and cases with risk SNPs [$t(104.4) = -0.31, p = .76$]. The mean number of EMR diagnoses for the 30 relevant phenotypes did not differ between suicide cases without risk SNPs and cases with risk SNPs [$t(93.1) = -0.52, p = .60$]. Average age of onset measured by the earliest EMR diagnosis for 10 phenotypes did not differ significantly between groups after correction for multiple comparisons [depression; $t(46.3) = 0.35, p = .73$, anxiety; $t(15.6) = -2.15, p = .05$, bipolar $t(7.7) = 2.48, p = .04$, schizophrenia/schizoaffective $t(7.1) = 2.16, p = .07$, pain $t(32.0) = -0.60, p = .55$, drug abuse; $t(13.4) = -0.24, p = .82$, sleep disorders; $t(13.7) = 0.38, p = .71$, suicidal ideation; $t(5.2) = 0.13, p = .90$, self-injury; $t(8.5) = 0.62, p = .55$, immune disorders; $t(14.1) = -2.4, p = .02$].

Eight percent of cases carrying rare risk SNPs had bipolar diagnoses and 5% had Schizophrenia or Schizoaffective diagnoses. One individual who died by suicide had rare missense variants in both *SNAPC1*

and *ADGRF5*. This case was male and 16 years old at the time of death. Only one ICD code was observed for this individual (ICD9 787.03 "vomiting alone") and other information was not available.

Of the 15 suicide cases with the rare missense variant in *ESS2*, 13 had EMR data. 12 of 13 of these cases had ICD diagnosis codes indicative of accidental trauma (Supplementary Table S5). For cases carrying a rare risk variant at any one of the five SNPs who also had linked EMR data, at least 50% of individuals had chronic/acute pain diagnoses (Supplementary Table S5; Supplementary Table S6). For cases with identified risk variants in *TNKS1BP1*, *ESS2*, *ADGRF5* or *PER1* and EMR data, >45% of cases had depression diagnoses. For cases with identified risk variants in *SNAPC1*, *TNKS1BP1*, *ESS2* or *PER1* and EMR data, >55% of cases had diagnoses related to accidental trauma. However, these frequencies are similar when compared against suicide cases without the risk variants.

4 | DISCUSSION

This study identified five novel variants likely to impact gene function that are significantly associated with suicide death. These variants implicate the following genes: *SNAPC1*, *TNKS1BP1*, *ADGRF5*, *PER1*, and *ESS2*. Our study included rigorous thresholds for quality control and ancestry, and all surviving variants were validated in sequence data, addressing concerns of potential error in PsychArray rare variant content. Of note, the significant variants in our study are closer to 1% in frequency; variants that were <0.1% did not validate in sequence data.

At the gene level, our analyses revealed two genes with prior evidence linking them to suicide risk: *PER1* and *SNAPC1*, though specific functional risk variants were not identified in those previous studies. *PER1* is a circadian clock gene that plays an essential role in generating circadian rhythms (Gery et al., 2006). *PER1* has been implicated in long-term memory formation (Kwapis et al., 2018), response to insufficient sleep (Moller-Levet et al., 2013), and regulation of neuroinflammation (Fonken et al., 2016). Altered expression of *PER1* has been found in individuals with schizophrenia (Aston, Jiang, & Sokolov, 2004; Sun et al., 2016) and *PER1* knock-out mice exhibit ADHD-like symptoms and reduced levels of dopamine (Huang et al., 2015). *PER1* was significantly downregulated in the prefrontal cortex of individuals who died by suicide (Sequeira et al., 2012) and was differentially expressed in suicide attempters compared with controls (Levey et al., 2016). Postmortem brain tissue expression results from the psychENCODE sample showed *PER1* was differentially expressed in schizophrenia (FDR $p = .0006$) and bipolar cases (FDR $p = .0016$) compared to controls.

SNAPC1 encodes a subunit of the Small Nuclear RNA Activating Complex (SNAPc), which activates RNA polymerase II and III and is required for translation of small nuclear RNA (Baillat, Gardini, Cesaroni, & Shiekhattar, 2012; Henry, Sadowski, Kobayashi, & Hernandez, 1995). *SNAPC1* was also identified in a genome-wide significant region in extended Utah families enriched for suicide deaths (Coon et al., 2018). The prior *SNAPC1* association was discovered

TABLE 3 Prevalence of broad EMR diagnostic categories (ICD codes defining each category are listed in Supplementary Table S1) for suicide cases with and without five identified risk SNPs

EMR category	EMR phenotype	Cases without risk SNPs ($N = 2{,}117$)	Cases with risk SNPs ($N = 97$)
Psychiatric	Suicidal ideation	0.163	0.125
	Self-injury	0.212	0.198
	Depression	0.497	0.469
	Anxiety (non-trauma)	0.366	0.344
	Bipolar	0.128	0.083
	Schizophrenia/schizoaffective	0.038	0.052
	PD cluster A	0.003	0
	PD cluster B	0.093	0.063
	PD cluster C	0.011	0.000
	Interpersonal trauma	0.345	0.365
	Impulse control disorders	0.093	0.042
	Eating disorders	0.011	0
	Alcohol abuse	0.246	0.250
	Drug abuse	0.275	0.292
Developmental	Dementia/neurodegenerative	0.117	0.125
	Developmental disorders	0.016	0.031
	Seizure	0.040	0.031
Pain/ injury	Migraine	0.090	0.125
	Pain	0.681	0.656
	Head injury	0.066	0.083
	Accidental trauma	0.634	0.604
	Injury by law enforcement	0.004	0
Other relevant conditions	Immune/autoimmune disorders	0.193	0.271
	Somatic disorders	0.093	0.104
	Asthma	0.112	0.125
	COPD	0.073	0.052
	Cardiovascular disorders	0.458	0.406
	Nutrition/metabolism	0.415	0.365
	Obesity	0.208	0.250
	Sleep disorders	0.258	0.302

Adjusting for multiple comparisons, no significant differences between groups were found.

using an extended familial design with a small subset of 216 Utah suicides from our current study cohort. There was no overlap in cases with rare risk variants in *SNAPC1* in this study and those cases that implicated *SNAPC1* in the previous study.

The other genes impacted by rare variants in our study have been implicated in psychiatric disorders or immune-related phenotypes. *TNKS1BP1* plays a role in DNA damage repair, genome stability, and regulation of the actin cytoskeleton (Ohishi et al., 2017; Tan et al., 2017; Zou et al., 2015). Expression results from psychENCODE showed increased expression differences for *TNKS1BP1* in schizophrenia (FDR $p = .02$). *ADGRF5*, formally *GPR116*, encodes a G protein-coupled receptor that plays a critical role in regulating pulmonary immune response (Brown et al., 2017; Yang et al., 2013). The GWAS catalog identified *ADGRF5* as having associations with four

blood-related traits (blood protein levels, white blood cell count, mean platelet volume, and monocyte count) and adolescent idiopathic scoliosis. *ESS2* (formally known as *DGCR14*) is a nuclear protein that enhances the transcriptional activity of the retinoic acid receptor and is a component of the spliceosome C complex which removes introns from pre-mRNA (Takada et al., 2018). Mutations in *ESS2* are associated with risk for schizophrenia (Funato et al., 2010; Wang et al., 2006). In addition, deletions in *ESS2* are associated with DiGeorge syndrome, which causes immunodeficiency, heart defects, skeletal abnormalities, psychiatric disturbance, and alterations in REM sleep (Funato et al., 2010).

The associations of *AGDRF5* with pulmonary immune response, blood phenotypes, and scoliosis and the association of *PER1* with neuroinflammation may reflect recent data suggesting the importance of

immune response and inflammation in risk generally of psychopathology (Pouget et al., 2019) and specifically of suicide (Sudol & Mann, 2017). Supporting this association, *ADGRF5*, with genes implicated by GWAS studies of suicidal behavior showed an enrichment in immune pathways (González-Castro et al., 2019).

The 119 suicide cases with these variants represent 4.45% of the European genotyped study sample that was the focus of this study. 97 of the 119 cases (81.5%) had EMR data and these cases showed evidence of chronic pain, depression, and trauma phenotypes in EMRs data. No clear differences with regards to medical diagnoses from EMR were noted between suicide cases with and without the five identified risk SNPs, although as expected for rare variants, we were underpowered to detect differences if they did exist. While previously published associations suggested risk of schizophrenia for *PER1*, *ESS2*, and *TNKS1BP1*, only 10% of cases with risk alleles in these genes had schizophrenia or schizoaffective diagnoses, versus 4% of cases without these rare variants. *PER1* has also previously been associated with bipolar disorder, yet only one of the 16 cases with the risk SNP and EMR data had a bipolar diagnosis.

Our analysis presents an efficient, cost-effective study of putative functional variants in the USGRS, a research resource that includes a large sample of molecular data from suicide deaths linked to extensive demographic and medical record data. Importantly, most large studies of the genetics of suicide focus on suicidal behaviors, which are relatively common [4.3% per year (Hedegaard, Curtin, & Warner, 2018)]. However, suicidal behaviors imperfectly predict the more extreme outcome of suicide death [0.01–0.02% per year (Hedegaard et al., 2018)]. The USGRS provides a resource for genetic discovery of risks specifically associated with the outcome of suicide death. Every discovery of specific genetic risk factor associated with suicide death will bring us closer to the goal of targeted prevention for those at risk. This study complements our recent genome-wide association study (Docherty et al., 2020), which focuses on the discovery of common variation associated with suicide death.

Here we demonstrate an important role for rare variants in the polygenic architecture of suicide risk. Importantly, rare variants are more functionally tractable than common variants and may reveal unique insights into the mechanisms of risk. We have made careful use of rare, functional content on the genome-wide array to reach the discoveries reported in this study.

This study has important limitations. While taking a conservative approach and refining our data to preselected, rare array content with previous evidence of relevance to psychiatry, these data represent only a tiny proportion of the potential rare, functional risk variation in the genome. A full examination of rare variants in WGS data in suicide cases will be necessary to uncover additional variants as well as other classes of variants such as structural variants. These efforts are currently underway with WGS data in the USGRS cohort.

Rare variants are particularly susceptible to ancestry effects (Kosmicki et al., 2016). Therefore, we confined our analysis to individuals of NFE ancestry, the group making up the largest proportion of our data resource. Allele frequencies for the five validated variants are not elevated in other ancestries in gnomAD and these variants seem to be more specific to NFE populations. This likely limits the generalizability of these findings to other ancestry groups.

It is also possible that population stratification or potential confounding could stem from our use of convenience control samples in these analyses. Cases and controls were collected for different purposes and are from different geographic regions. We used publicly available controls from gnomAD in the Stage 1 analysis. Detailed demographic data and individual-level genotype data are not available for these controls, therefore we were not able to use ancestry, sex, or other variables as covariates in the analysis. Detailed demographic data were also not available for PGC controls from the Stage 2 analyses.

In spite of these limitations, our results represent important novel risk variants and genes associated with risk for suicide death. Our success with this approach may in part reflect the design of the PsychArray, providing highly relevant content. Success may also be driven by our analysis of suicide death, an extreme phenotypic outcome. Overall, the research community has made great progress with genetic discovery of both common and rare variants leading to risk of complex phenotypes. The aggregated effects of common and rare variants are starting to make a difference for individuals with these conditions. While the genetic study of suicide death has far to go, as discoveries accumulate, aggregation of results into larger risk pathways will begin to have an impact on personalized risk prediction and prevention.

CONFLICT OF INTEREST

Author Qingqin Li is an employee of Janssen Research & Development. No other co-authors have conflicts of interest relevant to the content of this manuscript, including no financial interest, relationships, or affiliations.

AUTHOR CONTRIBUTIONS

Emily DiBlasi and Hilary Coon: Conceived and designed the method. Emily DiBlasi conducted analyses and wrote the first draft and revisions of the manuscript. **Andrey Shabalin, Elliott Ferris, Danli Chen, and Eric Monson**: Contributed to data quality control and analysis

pipelines. **Nancy William, Eoin Gaj, Michael Klein, Leslie Jerominski, Brandon Callor, Erik Christensen, Ken Smith, Alison Fraser, and Zhe Yu**: Contributed to data preparation. **Anna Docherty, Qingqin Li, Eli Stahl, Nicola Camp, Brooks Keeshin, Douglas Gray, Amanda Bakian, and Anne Kirby**: Contributed to the overall interpretation of the method and results. All authors revised and approved the manuscript.

ORCID

Emily DiBlasi https://orcid.org/0000-0002-1266-7297
Anna R. Docherty https://orcid.org/0000-0001-7139-7007
Hilary Coon https://orcid.org/0000-0002-8877-5446

REFERENCES

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Aston, C., Jiang, L., & Sokolov, B. P. (2004). Microarray analysis of postmortem temporal cortex from patients with schizophrenia. *Journal of Neuroscience Research*, *77*(6), 858–866. https://doi.org/10.1002/jnr.20208

Auer, P. L., & Lettre, G. (2015). Rare variant association studies: Considerations, challenges and opportunities. *Genome Medicine*, *7*(1), 16. https://doi.org/10.1186/s13073-015-0138-2

Baillat, D., Gardini, A., Cesaroni, M., & Shiekhattar, R. (2012). Requirement for SNAPC1 in transcriptional responsiveness to diverse extracellular signals. *Molecular and Cellular Biology*, *32*(22), 4642–4650. https://doi.org/10.1128/MCB.00906-12

Bergen, S. E., Ploner, A., Howrigan, D., O'Donovan, M. C., Smoller, J. W., Sullivan, P. F., ... Kendler, K. S. (2019). Joint contributions of rare copy number variants and common SNPs to risk for schizophrenia. *American Journal of Psychiatry*, *176*(1), 29–35. https://doi.org/10.1176/appi.ajp.2018.17040467

Brown, K., Filuta, A., Ludwig, M.-G., Seuwen, K., Jaros, J., Vidal, S., ... Bridges, J. P. (2017). Epithelial Gpr116 regulates pulmonary alveolar homeostasis via Gq/11 signaling. *JCI Insight*, *2*(11), e93700. https://doi.org/10.1172/jci.insight.93700

Buniello, A., Macarthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., ... Parkinson, H. (2019). The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research*, *47*(D1), D1005–D1012. https://doi.org/10.1093/nar/gky1120

CDC/National Center for Health Statistics. (2019). Retrieved from https://www.cdc.gov/nchs/pressroom/sosmap/suicide-mortality/suicide.htm.

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*, 7. https://doi.org/10.1186/s13742-015-0047-8

Coon, H., Darlington, T. M., DiBlasi, E., Callor, W. B., Ferris, E., Fraser, A., ... Gray, D. (2018). Genome-wide significant regions in 43 Utah high-risk families implicate multiple genes involved in risk for completed suicide. *Molecular Psychiatry*, *25*(11), 3077–3090. https://doi.org/10.1038/s41380-018-0282-3

Depristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., ... Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*, 491–498. https://doi.org/10.1038/ng.806

Docherty, A. R., Shabalin, A. A., DiBlasi, E., Monson, E., Mullins, N., Adkins, D. E., ... Coon, H. (2020). Genome-wide association study of suicide death and polygenic prediction of clinical antecedents. *American Journal of Psychiatry*, *177*(10), 917–927. https://doi.org/10.1176/appi.ajp.2020.19101025

Erlangsen, A., Appadurai, V., Wang, Y., Turecki, G., Mors, O., Werge, T., ... Agerbo, E. (2018). Genetics of suicide attempts in individuals with and without mental disorders: A population-based genome-wide association study. *Molecular Psychiatry*, *25*(10), 2410–2421. https://doi.org/10.1038/s41380-018-0218-y

Ferreira, J. A. (2007). The Benjamini-Hochberg method in the case of discrete test statistics. *International Journal of Biostatistics*, *3*(1), 11. https://doi.org/10.2202/1557-4679.1065

Fonken, L. K., Kitt, M. M., Gaudet, A. D., Barrientos, R. M., Watkins, L. R., & Maier, S. F. (2016). Diminished circadian rhythms in hippocampal microglia may contribute to age-related neuro-inflammatory sensitization. *Neurobiology of Aging*, *47*, 102–112. https://doi.org/10.1016/j.neurobiolaging.2016.07.019

Funato, H., Sato, M., Sinton, C. M., Gautron, L., Williams, S. C., Skach, A., ... Yanagisawa, M. (2010). Loss of Goosecoid-like and DiGeorge syndrome critical region 14 in interpeduncular nucleus results in altered regulation of rapid eye movement sleep. *Proceedings of the National Academy of Sciences*, *107*(42), 18155–18160. https://doi.org/10.1073/pnas.1012764107

Gandal, M. J., Zhang, P., Hadjimichael, E., Walker, R. L., Chen, C., Liu, S., ... Geschwind, D. H. (2018). Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science*, *362*(6420), eaat8127. https://doi.org/10.1126/science.aat8127

Gery, S., Komatsu, N., Baldjyan, L., Yu, A., Koo, D., & Koeffler, H. P. (2006). The circadian gene Per1 plays an important role in cell growth and DNA damage control in human cancer cells. *Molecular Cell*, *20*(7), 2195–2201. https://doi.org/10.1016/j.molcel.2006.03.038

González-Castro, T. B., Tovilla-Zárate, C. A., Genis-Mendoza, A. D., Juárez-Rojop, I. E., Nicolini, H., López-Narváez, M. L., & Martínez-Magaña, J. J. (2019). Identification of gene ontology and pathways implicated in suicide behavior: Systematic review and enrichment analysis of GWAS studies. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, *180*(5), 320–329. https://doi.org/10.1002/ajmg.b.32731

Hedegaard, H., Curtin, S. C., & Warner, M. (2018). Suicide mortality in the United States, 1999–2017. In *NCHS data brief*. Hyattsville, MD: National Center for Health Statistics.

Henry, R. W., Sadowski, C. L., Kobayashi, R., & Hernandez, N. (1995). A TBP–TAF complex required for transcription of human snRNA genes by RNA polymerases II and III. *Nature*, *374*, 653–656. https://doi.org/10.1038/374653a0

Huang, J., Zhong, Z., Wang, M., Chen, X., Tan, Y., Zhang, S., ... Wang, H. (2015). Circadian modulation of dopamine levels and dopaminergic neuron development contributes to attention deficiency and hyperactive behavior. *Journal of Neuroscience*, *35*(6), 2572–2587. https://doi.org/10.1523/JNEUROSCI.2551-14.2015

Justice, A. E., Karaderi, T., Highland, H. M., Young, K. L., Graff, M., Lu, Y., ... Lindgren, C. M. (2019). Protein-coding variants implicate novel genes related to lipid homeostasis contributing to body-fat distribution. *Nature Genetics*, *51*(3), 452–469. https://doi.org/10.1038/s41588-018-0334-2

Kircher, M., Witten, D. M., Jain, P., O'roak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*, *46*, 310–315. https://doi.org/10.1038/ng.2892

Kosmicki, J. A., Churchhouse, C. L., Rivas, M. A., & Neale, B. M. (2016). Discovery of rare variants for complex phenotypes. *Human Genetics*, *135*, 625–634. https://doi.org/10.1007/s00439-016-1679-1

Kryukov, G. V., Pennacchio, L. A., & Sunyaev, S. R. (2007). Most rare missense alleles are deleterious in humans: Implications for complex

disease and association studies. *The American Journal of Human Genetics*, 80(4), 727–739. https://doi.org/10.1086/513473

Kwapis, J. L., Alaghband, Y., Kramár, E. A., López, A. J., Vogel Ciernia, A., White, A. O., ... Wood, M. A. (2018). Epigenetic regulation of the circadian gene Per1 contributes to age-related changes in hippocampal memory. *Nature Communications*, 9(1), 3323. https://doi.org/10.1038/s41467-018-05868-0

Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., ... Maglott, D. R. (2018). ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Research*, 46(D1), D1062–D1067. https://doi.org/10.1093/nar/gkx1153

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., ... Williams, A. L. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 536(7616), 285–291. https://doi.org/10.1038/nature19057

Levey, D. F., Niculescu, E. M., Le-Niculescu, H., Dainton, H. L., Phalen, P. L., Ladd, T. B., ... Niculescu, A. B. (2016). Towards understanding and predicting suicidality in women: Biomarkers and clinical risk assessment. *Molecular Psychiatry*, 21(6), 768–785. https://doi.org/10.1038/mp.2016.31

Li, G. (2016). A new model calling procedure for Illumina BeadArray data. *BMC Genetics*, 17(1), 90. https://doi.org/10.1186/s12863-016-0398-x

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Marei, H. E., Althani, A., Suhonen, J., El Zowalaty, M. E., Albanna, M. A., Cenciarelli, C., ... Caceci, T. (2016). Common and rare genetic variants associated with Alzheimer's disease. *Journal of Cellular Physiology*, 231(7), 1432–1437. https://doi.org/10.1002/jcp.25225

Mathieson, I., & McVean, G. (2012). Differential confounding of rare and common variants in spatially structured populations. *Nature Genetics*, 44(3), 243–246. https://doi.org/10.1038/ng.1074

McGuffin, P., Marusic, A., & Farmer, A. (2001). What can psychiatric genetics offer suicidology? *Crisis*, 22, 61–65. https://doi.org/10.1027//0227-5910.22.2.61

Moller-Levet, C. S., Archer, S. N., Bucca, G., Laing, E. E., Slak, A., Kabiljo, R., ... Dijk, D.-J. (2013). Effects of insufficient sleep on circadian rhythmicity and expression amplitude of the human blood transcriptome. *Proceedings of the National Academy of Sciences*, 110(12), E1132–E1141. https://doi.org/10.1073/pnas.1217154110

Monson, E. T., Pirooznia, M., Parla, J., Kramer, M., Goes, F. S., Gaine, M. E., ... Willour, V. L. (2017). Assessment of whole-exome sequence data in attempted suicide within a bipolar disorder cohort. *Molecular Neuropsychiatry*, 3(1), 1–11. https://doi.org/10.1159/000454773

Mullins, N., Bigdeli, T. B., Børglum, A. D., Coleman, J. R. I., Demontis, D., Mehta, D., ... Lewis, C. M. (2019). GWAS of suicide attempt in psychiatric disorders and association with major depression polygenic risk scores. *American Journal of Psychiatry*, 176(8), 651–660. https://doi.org/10.1176/appi.ajp.2019.18080957

Nicolae, D. L. (2016). Association tests for rare variants. *Annual Review of Genomics and Human Genetics*, 17, 117–130. https://doi.org/10.1146/annurev-genom-083115-022609

Ohishi, T., Yoshida, H., Katori, M., Migita, T., Muramatsu, Y., Miyake, M., ... Seimiya, H. (2017). Tankyrase-binding protein TNKS1BP1 regulates Actin cytoskeleton rearrangement and cancer cell invasion. *Cancer Research*, 77(9), 2328–2338. https://doi.org/10.1158/0008-5472.CAN-16-1846

Pedersen, N. L., & Fiske, A. (2010). Genetic influences on suicide and nonfatal suicidal behavior: Twin study findings. *European Psychiatry*, 25(5), 264–267. https://doi.org/10.1016/j.eurpsy.2009.12.008

Piñero, J., Bravo, Á., Queralt-Rosinach, N., Gutiérrez-Sacristán, A., Deu-Pons, J., Centeno, E., ... Furlong, L. I. (2017). DisGeNET: A comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Research*, 45(D1), D833–D839. https://doi.org/10.1093/nar/gkw943

Pouget, J. G., Han, B., Wu, Y., Mignot, E., Ollila, H. M., Barker, J., ... Knight, J. (2019). Cross-disorder analysis of schizophrenia and 19 immune-mediated diseases identifies shared genetic risk. *Human Molecular Genetics*, 28(20), 3498–3513. https://doi.org/10.1093/hmg/ddz145

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., ... Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81(3), 559–575. https://doi.org/10.1086/519795

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), 5463–5467. https://doi.org/10.1073/pnas.74.12.5463

Sequeira, A., Morgan, L., Walsh, D. M., Cartagena, P. M., Choudary, P., Li, J., ... Vawter, M. P. (2012). Gene expression changes in the prefrontal cortex, anterior cingulate cortex and nucleus accumbens of mood disorders subjects that committed suicide. *PLoS One*, 7(4), e35367. https://doi.org/10.1371/journal.pone.0035367

Sherry, S. T. (2001). dbSNP: The NCBI database of genetic variation. *Nucleic Acids Research*, 29(1), 308–311. https://doi.org/10.1093/nar/29.1.308

Sudol, K., & Mann, J. J. (2017). Biomarkers of suicide attempt behavior: Towards a biological model of risk. *Current Psychiatry Reports*, 19(6), 31. https://doi.org/10.1007/s11920-017-0781-y

Sun, H. Q., Li, S. X., Chen, F. B., Zhang, Y., Li, P., Jin, M., ... Lu, L. (2016). Diurnal neurobiological alterations after exposure to clozapine in first-episode schizophrenia patients. *Psychoneuroendocrinology*, 64, 108–116. https://doi.org/10.1016/j.psyneuen.2015.11.013

Surakka, I., Horikoshi, M., Mägi, R., Sarin, A. P., Mahajan, A., Lagou, V., ... Ripatti, S. (2015). The impact of low-frequency and rare variants on lipid levels. *Nature Genetics*, 47(6), 589–597. https://doi.org/10.1038/ng.3300

Takada, I., Tsuchiya, M., Yanaka, K., Hidano, S., Takahashi, S., Kobayashi, T., ... Makishima, M. (2018). Ess2 bridges transcriptional regulators and spliceosomal complexes via distinct interacting domains. *Biochemical and Biophysical Research Communications*, 497(2), 597–604. https://doi.org/10.1016/j.bbrc.2018.02.110

Tan, W., Guan, H., Zou, L. H., Wang, Y., Liu, X. D., Rang, W. Q., ... Zhong, C. G. (2017). Overexpression of TNKS1BP1 in lung cancers and its involvement in homologous recombination pathway of DNA double-strand breaks. *Cancer Medicine*, 6(2), 483–493. https://doi.org/10.1002/cam4.995

Tin, A., Li, Y., Brody, J. A., Nutile, T., Chu, A. Y., Huffman, J. E., ... Köttgen, A. (2018). Large-scale whole-exome sequencing association studies identify rare functional variants influencing serum urate levels. *Nature Communications*, 9(1), 4228. https://doi.org/10.1038/s41467-018-06620-4

Tombácz, D., Maróti, Z., Kalmár, T., Csabai, Z., Balázs, Z., Takahashi, S., ... Boldogkoi, Z. (2017). High-coverage whole-exome sequencing identifies candidate genes for suicide in victims with major depressive disorder. *Scientific Reports*, 7, 7106. https://doi.org/10.1038/s41598-017-06522-3

Turcot, V., Lu, Y., Highland, H. M., Schurmann, C., Justice, A. E., Fine, R. S., ... Loos, R. J. F. (2018). Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. *Nature Genetics*, 50(1), 26–35. https://doi.org/10.1038/s41588-017-0011-x

Wang, H., Duan, S., Du, J., Li, X., Xu, Y., Zhang, Z., ... He, L. (2006). Transmission disequilibrium test provides evidence of association between promoter polymorphisms in 22q11 gene DGCR14 and schizophrenia. *Journal of Neural Transmission*, 113(10), 1551–1561. https://doi.org/10.1007/s00702-005-0420-3

Watanabe, K., Taskesen, E., van Bochoven, A., & Posthuma, D. (2017). Functional mapping and annotation of genetic associations with

FUMA. *Nature Communications*, *8*(1), 1826. https://doi.org/10.1038/s41467-017-01261-5

WHO. (2019). *Suicide in the world: Global health estimates*. Retrieved from https://apps.who.int/iris/bitstream/handle/10665/326948/WHO-MSD-MER-19.3-eng.pdf.

Yang, M. Y., Hilton, M. B., Seaman, S., Haines, D. C., Nagashima, K., Burks, C. M., … St.Croix, B. (2013). Essential regulation of lung surfactant homeostasis by the orphan G protein-coupled receptor GPR116. *Cell Reports*, *3*(5), 1457–1464. https://doi.org/10.1016/j.celrep.2013.04.019

Zou, L.-H., Shang, Z.-F., Tan, W., Liu, X.-D., Xu, Q.-Z., Song, M., … Zhou, P.-K. (2015). TNKS1BP1 functions in DNA double-strand break repair though facilitating DNA-PKcs autophosphorylation dependent on PARP-1. *Oncotarget*, *6*(9), 7011–7022. https://doi.org/10.18632/oncotarget.3137

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: DiBlasi, E., Shabalin, A. A., Monson, E. T., Keeshin, B. R., Bakian, A. V., Kirby, A. V….Coon, H. (2021). Rare protein-coding variants implicate genes involved in risk of suicide death. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *186*B:508–520. https://doi.org/10.1002/ajmg.b.32861