

Intragenic deletions from whole genome sequencing of 1054 suicide deaths

DiBlasi E,¹ Shabalin AA,¹ Nicholas TJ,² Monson ET,¹ Ferris E,³ Yefimov L,⁴ Han S,¹ Baird LM,¹ Callor WB,⁴ Staley MJ,⁴ Li QS,⁵ Willour VL,⁶ Coon H^{1*}

¹ Department of Psychiatry & Huntsman Mental Health Institute, University of Utah, Salt Lake City, UT

² Department of Human Genetics, Utah Center for Genetic Discovery, University of Utah, Salt Lake City UT

³ Department of Neurobiology, University of Utah, Salt Lake City, UT

⁴ Utah Department of Health and Human Services, Salt Lake City, UT

⁵ Neuroscience Therapeutic Area, Janssen Research and Development, Titusville, NJ

⁶ Department of Psychiatry, University of Iowa, Iowa City, IA

* Corresponding author: Hilary Coon, Department of Psychiatry, University of Utah School of Medicine, Salt Lake City, UT, 84112, USA. Phone: 801-540-3817. E-mail: hilary.coon@utah.edu

Abstract

Suicide is an urgent public health crisis that claimed over 48,000 lives in the US in 2022. The importance of genetics in suicide risk has been established by classical twin and family studies, and confirmed with recent large genome-wide association studies (GWAS). While the GWAS are beginning to reveal genetic risk due to common variants each with small effect on liability, these results explain only a fraction of the genetic risk. As with other complex health conditions, some of this unexplained risk is likely due to rarer variants with larger effect on liability. Using whole genome sequencing (WGS) data from 1,054 population-ascertained Utah suicide deaths, we investigated intragenic deletions as a class of genomic variation highly likely to disrupt gene function. To minimize the chance of false positive results, studied deletions were limited to those found in three large publicly-available control datasets (1000 Genomes, GnomAD, and Centers for Common Disease Genomics). Additional internal replication also required deletions to occur at least twice in WGS from an initial cohort of 670 suicide deaths then again in a second cohort of 384 suicide deaths. All results meeting these filters were manually validated. There were 11 validated deletions with at least 2-fold increase in frequency over occurrence in controls (range 2.28 to 4.46). These results implicated genes associated with risk of mental health conditions (*MPST*, *IL4R*, *CDH13*), epilepsy (*CLCA4*), intellectual disability (*ZNF44*), neuronal function (*OSBPL2*), metabolic function (*FBOX36*), lipid metabolism (*TM9SF3*), immune related functions (*PIPOX*, *IL4R*), and transcriptional repression (*ZHX3*). SNPs in genes implicated by the deletions have also been associated with mental health conditions, neuronal function, immune response, and other critical biological pathways including neuroinflammation and cellular response to stress. Demographic and clinical associations of suicide deaths with specific genetic deletions, highlight the prevalence of mood, anxiety and bipolar disorders and variations in age at suicide death among affected individuals. This work is the largest genome-wide analyses of WGS variation in suicide deaths to date. Pending replication, results will guide future functional studies with the eventual goals of increased understanding of mechanisms leading to risk.

Introduction

In the U.S., the suicide death rate is the highest it has been in over 80 years (Curtin et al., 2023); suicide prediction and prevention are top public health priorities. Identification of risk factors leading to suicide death remains challenging due to the complexity of suicide risks; however, genetic factors comprise one important aspect of risk that may allow substantial progress in bridging this knowledge gap, with estimates of heritability of approximately 50% (McGuffin et al., 2001; N. L. Pedersen & Fiske, 2010; Statham et al., 1998).

Recent genome-wide results based on large samples have begun to reveal common genetic factors associated with suicide attempt and mortality (Docherty et al., 2023; Kimbrel et al., 2023; Q. S. Li et al., 2022; Mullins et al., 2022). Biological pathway analyses among GWAS with significant loci implicate nervous system function and development as well as immune function, neuroinflammation and cellular response to stress. While these findings represent significant progress, they also highlight the complex polygenic nature of suicidal behavior and the need to detect additional risk loci. GWAS studies provide estimates of SNP heritability ranging from 3.9 – 5.7%, considerably less than the heritability of 50% from the classical behavioral genetics designs involving aggregated family and twin studies. This “missing heritability” likely derives from effects of rare genomic variation and from the contribution of gene x gene effects (Génin, 2020).

While extensive research has focused on common genetic variants in psychiatric phenotypes, far less attention has been given to rarer genetic variation, particularly structural variation (Owen & Williams, 2021). Structural variants (SVs) are a broad class of genome variation that is extremely diverse in type and size (typically 50 bp to several Mb). SVs includes copy number variants (CNVs), such as deletions and duplications, simple re-arrangements (e.g., inversions), mobile element insertions and other complex rearrangements (multiple combinations of SVs) (Ho et al., 2020). SVs are present in all human genomes often have severe functional consequences due to their ability to disrupt genes, cause gene fusions, rearrange regulatory elements and alter gene dosage (Weischenfeldt et al., 2013). SVs have historically been harder to detect than common variation data due size and limitations in associated sequencing technologies and detection algorithms (Ho et al., 2020). Thus, like rarer SNVs, structural variants have largely been excluded from genotyping arrays used in large-scale genetic studies. This gap in knowledge is significant, as other types of genetic variation may play a

crucial role in understanding the genetic underpinnings of complex psychiatric conditions (Andreassen et al., 2023). Indeed, studies have begun to implicate rare variation associated with risk of suicide death (DiBlasi et al., 2021; DiBlasi et al., 2021). However, it is crucial to note that these results await replication in larger, independent cohorts. Further research is needed to validate and expand upon these initial discoveries, potentially opening new avenues for suicide prevention and targeted interventions.

The current study provides results from a genome-wide screen of whole genome sequence (WGS) data in a large (N=1,054), population-ascertained sample of suicide deaths. These aspects address previous gaps in studies of rare variation, including small sample sizes, reliance on genotyping rather than WGS data, and/or ascertainment within individuals with specific clinical diagnoses. Using WGS data, emerging approaches can be used to more reliably detect and call SVs (Ho et al., 2020). The current genome-wide investigation focuses on deletions (a class of SVs) overlapping coding sequence of genes; these represent genomic events most likely to affect gene function. We have chosen to focus on deletions as the category of SVs with the most robust ascertainment as they are more amenable to detection by available robust computational tools. We use an independent internal replication strategy for deletion results in our sample, and present cross-associations implicated by previous rare variant studies and by current large GWAS of a range of suicide outcomes.

Methods

Samples and data linking. Samples from Utah suicide deaths used in this study were obtained through a long-term collaboration with the Utah State Office of the Medical Examiner (OME). Suicide determination was made by the OME following detailed investigation of the scene and circumstances of death. High quality DNA was extracted from whole blood as previously described (Coon et al., 2020). Identifiers from suicide deaths were securely transferred by the OME to staff at the Utah Population Database (UPDB, <https://uofuhealth.utah.edu/huntsman/utah-population-database>). The UPDB is a statewide database with >27 million data records, including demographics, electronic health records data, and genealogical records. After linking suicide deaths to health data, identifiers were stripped by UPDB staff prior to secure data transfer to the research team. This study was approved by Institutional Review Boards at the University of Utah,

Intermountain Healthcare, and the Utah Department of Health and Human Services.

The 1,054 suicide deaths in this study were selected for whole genome sequencing (WGS) based on factors associated with increased likelihood of greater genetic risk, including younger age at death and evidence for significant extended familial risk of suicide death using Utah genealogical data (Coon et al., 2022), and also diagnoses of bipolar disorder, which is associated with high risk of suicide death (Miller & Black, 2020). Processing of WGS data was done in two waves, the first including 674 Utah suicide deaths and the second including 384 additional Utah suicide deaths.

Control data. Joint processing of WGS data from Utah suicide deaths was done with individual WGS data from 1,241 controls chosen to most closely match the predominantly Northern European ancestry of the Utah population from three sources: 1) 622 were from the 1000 Genomes Project (1000G) (Byrska-Bishop et al., 2022) 2) 512 were unrelated individuals from the Centre d'Etude du Polymorphisme Humain (CEPH) genetic reference families (Dausset et al., 1990); and 3) 96 were from a Utah study of longevity (Tschanz et al., 2004).

Relatedness and ancestry. Pairwise relatedness was tested using PLINK (Purcell et al., 2007). Among suicide deaths, sample pairs with third-degree relatedness or closer were retained to allow for correction of initial results and prioritization for potential future analyses if familial transmitted variants were observed. Among controls, only one randomly selected case for each pair with third degree or closer relatedness was retained for analysis. Any controls related to suicide deaths at the level of third degree or closer were eliminated.

Ancestry was determined through comparison of genotypes to reference population data available in the 1000G. Small numbers of non-European samples were expected in the suicide death cohort, as samples were ascertained from the Utah population which is predominantly White. In addition, selection of suicide deaths for WGS using increased extended familial risk was expected to further elevate the percentage of suicides with European ancestry, as familial risk assessments required the use of UPDB genealogical data which derives primarily from family founders of European descent. In this study, all suicide deaths were retained regardless of ancestry. Checks were then implemented to determine the extent to which any result was driven by non-

European carriers of the deletions.

Whole genome sequencing pipeline. WGS data was generated on the 1,054 Utah suicide deaths and 1,241 jointly called control samples by using Illumina NGS technology with an average read depth of ~30x. Alignment and variant calling and joint genotyping of suicide deaths and control WGS datasets was performed at the Utah Center for Genetic Discovery (UCGD) Core Facility, part of the Health Sciences Center Cores at University of Utah. The UCGD pipeline (v 2.13.77) called variants using the Sentieon software package (Freed et al., 2017), which incorporates GATK best practices (Depristo et al., 2011; Van der Auwera et al., 2013) with other innovative elements. Use of this UCGD pipeline has been described in detail elsewhere (Nicholas et al., 2022). Briefly, sequence reads were aligned to GRCh38 (Genome Reference Consortium Human Build 38) using BWA-MEM (H. Li & Durbin, 2009). The Haplotyper algorithm in Sentieon (Freed et al., 2017) was used to produce genomic Variant Call Format (gVCF) files. Suicide death gVCF files were then combined and jointly processed with data from the 1,241 individual control samples. Processing and analysis of data was done in two waves, one with 670 suicide deaths and the second with 384 suicide deaths. However, the final VCF file with suicide deaths and controls was recalled across both waves of suicide data together with controls to limit false positive calls.

Structural variant (SV) calling. Detection, attribution, and filtering of structural variants (SVs) was done using Smoove (version 0.2.6; <https://github.com/brentp/smoove>) and Manta (version 1.6.0) (Chen et al., 2016), again using the GRCh38 reference genome. Smoove produces a VCF file that contains deletions, duplications, inversions, and unclassified breakends (BNDs). Overlapping genes (Ensembl GRCh38.p12) and Duphold (B. S. Pedersen & Quinlan, 2019) quality metrics were added to the Smoove-generated VCF files. SVAfotate (<https://github.com/fakedrtom/SVAfotate>) was also used to add population SV allele frequencies from control datasets for comparison. SV frequencies from WGS datasets were available from the large publicly available cohorts described above (CCDG, GnomAD, 1000G, and CEPH, N=603). The reciprocal overlap fraction requirement of 0.8 was used to match control datasets in SVAfotate. This study focuses on the deletions, as these are the most robustly detected class of SV.

Additional filtering criteria for deletions. Additional filtering was performed using Slivar (B. S.

Pedersen et al., 2021) to minimize artifacts and maximize potential functional deletions. Criteria included: 1) removing deletions <100 base pairs (bp) and >1 megabase (Mb) as likely artifacts; 2) requiring deletions to overlap an exon using Smoove_gene INFO values (<https://github.com/brentp/smoove>); 3) requiring deletions to be present in at least two suicide deaths; and 4) requiring deletions to pass duphold (B. S. Pedersen & Quinlan, 2019) metrics (Duphold_pass flag and fold change DHFFC <0.7). Estimation of population frequencies of deletions were made using four large public datasets: 1) the Genome Aggregation Database comprising a subset of individuals from studies of adult-onset common diseases (Collins et al., 2020); 2) Centers for Common Disease Genomics (CCDG) comprising 10,738 individuals from studies of cardiovascular, neuropsychiatric, and immune-mediated diseases (Abel et al., 2020); 3) 1000G comprising 3,202 individuals (Byrska-Bishop et al., 2022); and 4) CEPH comprising 603 unrelated individuals in 34 multigenerational families (Dausset et al., 1990). Deletions were required to have frequency of <0.2 in each of these large control datasets. These criteria were required to be independently met in the first set of WGS data from 670 Utah suicides, and then also in the second set of 384 Utah suicide WGS data. This internal replication was followed to increase the robustness of our design. Frequencies of deletions were compared to observed frequencies in the jointly processed controls using exact binomial tests.

Manual validation of deletions. Smaller deletions (less than 5,000 bp) were verified by PCR and subsequent electrophoresis of the PCR product on agarose gels. First, primers flanking the putative deletions were designed using the Primer3 program (v. 0.4.0). 50 ng of template DNA was then amplified with these primers in a PCR reaction using Q5®High-Fidelity DNA Polymerase (New England Biolabs®) per manufacturer's recommended conditions. PCR product was subsequently electrophoresed on 1XTBE 2% agarose gels in the presence of Ethidium Bromide and then visualized on a Fotodyne™ FOTO/Analyst™ Workstation. The deletion was verified when product bands of two different sizes were present after PCR, one band being the larger wild type allele and the other band being a smaller deleted allele. Larger deletions (greater than 5,000 bp) were verified using pre-designed Taqman™ Copy Number Assays purchased from Thermo Fisher Scientific. Quadruplicate reactions were carried out per manufacturer's recommendations, and the data was analyzed with CopyCaller® Software.

Additional validation was done using SAMPLOT (Belyeu et al., 2021) to visualize the deletions *in silico* for comparison with manual validation results and to obtain indications of read depth sequence alignments in the region of the deletion. SAMPLOT images were reviewed for call and confidence intervals. Specifically, each deletion was inspected in the called region to have a loss in the depth of coverage and split or discordant read signatures.

Characterization of SNPs within deletions and of implicated genes. Potential functional impact of SNPs within each significant deletion was characterized using public data from the GTEx consortium (<https://gtexportal.org/>). Additional characterizations of the genes implicated by the deletions included publicly-available genetic associations from the GWAS Catalog (Sollis et al., 2023), and drug interaction data (Cannon et al., 2024; Gaulton et al., 2012).

Clinical data. Once SVs were identified and validated, additional electronic health records (EHR) data linked to suicide deaths were then used to assess for clinical associations of deletions with co-occurring psychiatric and physical health conditions among carriers of prioritized variants. Diagnostic data from the International Classification of Diseases (ICD-9 and ICD-10) system in the EHR was tabulated for suicides with deletions. Diagnoses were collapsed into phecodes (Denny et al., 2010), curated groups of ICD codes created to capture clinically meaningful phenotypes, using the hierarchical classification of diagnoses in the PheWAS catalog (PheCode Map X) (Shuey et al., 2023). Diagnoses that occurred within one week prior to death were excluded to eliminate diagnoses associated with the final fatal suicide event rather than those reflecting prior risk. Diagnoses relating to “Mood [affective] disorders” (phecode MB_286); “Bipolar disorder” (phecode MB_286.1); “Psychoactive substance related disorders” (phecode MB_280); “Pain” (phecode SS_809); “Psychotic disorder” (phecode MB_287); “Anxiety and anxiety disorders” (phecode MB_288) and “Sleep disorders” (NS_333) were studied.

Results

Sequencing quality control. Overall, WGS in the Utah suicide deaths was of high quality, with coverage from 25x to 48x across both batches, averaging 30x. In the first set of 674 suicides, four were omitted due to

low quality, leaving 670 for analysis. In addition, one sample in this batch had relatively low alignment (77.6%), four had alignment between 80-90%, and four had alignment between 90-95%. All of these were retained but scrutinized at the manual validation stages. In the second set of 384 suicides, one suicide selected for sequencing failed at the library prep phase, and was replaced with another suicide death sample. One sample in this second set had relatively low alignment (74.2%), two others had alignment between 80-90%, and three others had alignment between 90-95%. All of these were again retained for analysis and scrutinized at the manual validation stages.

For two of the suicides in this first WGS set and three in the second set, sex in the data records did not match genetic sex. Because all of the samples collected in our protocol are suicide deaths, the five samples with sex mismatches still represent suicides, so they were retained for tests of genetic ancestry and deletion frequency analyses. However, they were omitted for other phenotypic analyses. The suicide death cohort size was therefore 1,054 for frequency tests, but 1,049 for other phenotypic analyses.

Selection characteristics of suicides for WGS. Descriptive data of the suicides with WGS are shown in Table 1. A study design decision to further enhance our selection for bipolar disorder between cohorts 1 and 2 resulted in inclusion in the second cohort of more individuals with older age at death and also more females. Tests between the initial cohort of 670 and the second cohort of 384 suicides resulted in significant differences in these demographic characteristics (age test: $t=10.93$, $p<0.0001$; sex test: $\chi^2=11.75$, $p=0.0006$). Our overall WGS sample selection strategy also resulted in the combined cohort of Utah WGS suicides being significantly younger at death than in 5,386 other Utah suicides with genotyping but without WGS (32.13 years vs. 42.26 years; $p<0.0001$). The full sample of suicides with WGS were also significantly more female than in the larger genotyped cohort (27.84% female vs. 19.97% female; $p<0.0001$).

Utah suicides with WGS were of predominantly European ancestry (98.10%). There were 15 WGS suicide deaths with $<90\%$ EUR ancestry (between 51% and 84%; see Supplemental Table 1). These were retained in our initial analyses, with sensitivity tests to determine the extent to which any results were driven by occurrence of variants in non-European case samples. Control samples were chosen to match expected primarily European ancestry of population-ascertained cases.

Relatedness tests among the suicide deaths indicated nine third-degree relative pairs (cousins), five second-degree pairs (grandparent-grandchild or avuncular or half-siblings), and six first-degree relative pairs (siblings or parent-child). All related suicides were retained. Our analysis design dictated that if both relatives of a relative pair were found to share a variant or deletion, statistics for the variant/deletion would be re-adjusted, and the variant/deletion considered for potential future familial analyses.

Control data was also checked for relatedness. Among controls, there were two third-degree relative pairs, one second-degree pair, and two first-degree pairs. One random related control from each relative pair was deleted for the analysis. Finally, there were five suicide deaths where controls were found with a third degree relationships, and two suicides where controls were found with second degree relationships. These related controls were not considered in the analyses.

SV discovery and filtering. In the initial set of 670 suicide deaths, used as the discovery set, 58,012 deletions were discovered. While additional structural variants were apparent (e.g., insertions, inversions), deletions are the most common event and also most robustly detected, so further analyses focused on this class of structural variant. Of the 58,012 deletions, 7,616 crossed an exon using the Smoove_gene info prediction. Our next filters requiring the Duphold pass flag, fold change DHFFC<0.7, and presence in at least two suicide deaths resulted in 831 deletions. Using the SVAfotate tool, each deletion was compared to the public data cohorts (GnomAD, CCDG, 1000G, and CEPH). Deletions were deemed to be present in controls if there was at least 0.8 overlap in deletion boundaries. We only retained those with a frequency of <20% across these control cohorts (to eliminate common events), resulting in 391 deletions. Of these 391, 29 were >1Mb, and 42 were <100bp; these were omitted as likely artifacts.

After all of these filtering steps, there were 320 deletions, 113 of which were not found in any of the control cohorts; most of these unique deletions (N=96) were called as imprecise. We retained only the 207 found in the large control datasets. Our next step was to match these deletions to the second set of 384 suicide deaths. We required precise base-pair matching of the start and stop of the deletion in two suicides in the second set of 384 suicides for this replication step. Of the 25 deletions that met this requirement, there were 14 deletions with at least two occurrences in this second set, and with at least a two-fold increase in frequency of

each deletion over and above its frequency in jointly processed controls. Final determination of significance was made only after manual validation of each deletion in each case, as detailed below. Eleven deletions were significant after manual validation (Table 1).

Manual validation of deletions. Validation for seven of the smaller deletions was done using PCR (*PIPOX*, *TM9SF3*, *ZNF44*, *OSBPL2*, *ZHX3*, *APOOL*, *MPST*; see Supplemental Tables S1-S2 for primers and details). Primers for the *LILRAI* deletion (chr19:54601316-54601631) were not possible to define given extensive homology across the region of the deletion with three other regions on chromosome 19. For the *APOOL* deletion (chrX:85087992-85088221), only seven of the 16 suicides clearly showed the deletion. These two deletions were therefore not studied further in this analysis. Validation of *MPST* showed evidence of varying copy numbers of this SV in both suicides and controls. Though results still indicate case-control differences (see Supplemental Figure S1), this variation indicates interpretation of this deletion should be approached with caution. For the other deletions validated by PCR, all suicides determined to have the deletion by silico methods (Samplot; Belyeu et al., 2021) showed validation with PCR with the exception of one sample for the *TM9SF3* deletion, leaving 10 suicides with that deletion (case frequency = 0.00476, fold increase over controls = 3.33).

For the larger deletions requiring Taqman validation, *FBXO36*, *LMNTD1*, *CLCA4*, *FBX036*, and *IL4R* were manually validated for all suicides with these deletions. For *CDH13*, two of the 20 suicides initially identified as having the deletion failed to validate, leaving 18 with the deletion (case frequency = 0.000857, fold increase over controls = 2.06).

Table 2 gives details of each validated deletion. These 11 deletions were significantly elevated over the frequency in jointly-processed WGS data from controls at a false discovery rate (FDR) of 0.05. Frequencies from the publicly-available GnomAD “non-neuro” subset (Gudmundsson et al., 2021), which omit genomes from individuals with neuropsychiatric conditions, are also shown for comparison.

SNP-level and gene-level associations of significant deletions. For the 11 significant deletions, tissue associations are summarized in Table 3 for SNPs within each significant deletion with overlapping in GTEx data (<https://gtexportal.org/home/>). Detail of the GTEx SNP-level data is given in Supplemental Table S3. At

the gene level, Table 3 gives a summary of associations for each gene implicated by the 11 deletions listed in the GWAS catalog (<https://www.ebi.ac.uk/gwas/home>), and gene-drug interactions from data in the Drug-Gene Interaction Database (DGIdb; <https://www.dgidb.org/>).

Phenotypic associations. There were no related pairs among the suicide deaths sharing validated deletions in Table 2, nor were any of these cases among those with known sex mismatching. None of the validated deletions occurred in samples from individuals with <95% European ancestry with the exception of the deletion in *TM9SF3* on chromosome 10. This deletion was carried in two suicides with <95% European ancestry, one with 84.1% European ancestry, 15.0% Admixed American ancestry, and 0.9% South Asian ancestry, and the other with 81.3% European ancestry, 17.9% East Asian ancestry and 0.8% Admixed American ancestry. Removing these two occurrences from the data lowered the observed frequency for this frequency to 0.00381, leaving a fold increase of 2.67 and a binomial test significance of 1.19E-02 which remained significant at FDR (0.05).

Table 4 presents demographic and clinical associations of suicides carrying each deletion compared to overall frequencies in all suicide deaths with WGS data. Results are aggregated to protect confidentiality. 1,030/1,053 samples with WGS data had ICD diagnoses and could be assessed for phecode phenotypes. Of the samples with WGS and ICD data, mood disorders diagnoses were prevalent in 69.1% of suicide deaths, anxiety disorder (50.5%), bipolar disorder (42%), and substance related disorders (48.2%) were also common diagnoses. All suicide deaths with identified deletions in *TM9SF3* and *ZHX3* had mood disorder diagnoses. Suicide deaths with a deletion in *MPST* had the youngest average age at death at 28.3 years.

Discussion

This study represents the most extensive genome-wide investigation of structural variation, specifically intragenic deletions, in suicide decedents to date. Our analysis of whole genome sequencing (WGS) data from 1,054 suicide cases and 1,241 controls identified 11 validated deletions with significant enrichment in suicide deaths compared to controls. The use of WGS data allowed for a comprehensive analysis of deletions, overcoming limitations of previous studies that relied on genotyping arrays which do not contain this genetic

information. We also employed an internal replication strategy and manual validation of deletions. These deletions are predicted to impact genes associated with mental health conditions, neuronal function, immune response, and other critical biological pathways, offering new insights into the genetic underpinnings of suicide risk.

Results highlight the potential role of rare intragenic deletions in suicide risk, a dimension less explored compared to common single nucleotide polymorphisms (SNPs). The identified deletions overlap genes involved in diverse biological functions, such as mental health (*MPST*, *IL4R*, *PIPOX*, *CDH13*) (Bonk et al., 2024; Ruggeri et al., 2015; Salatino-Oliveira et al., 2015; Schoormans et al., 2012), neuronal function (*OSBPL2*), and immune related functions (*PIPOX*, *IL4R*) (Paulovičová et al., 2021; Vogelaar et al., 2018). Notably, several of these genes, such as *PIPOX*, *ZHX3*, *CDH13* (Børglum et al., 2014; Coleman et al., 2020; Fabbri et al., 2019; Mullins et al., 2021) have been implicated in previous GWAS related to psychiatric and neurodevelopmental disorders.

The association of deletions with *CLCA4* (He et al., 2021), linked to epilepsy, and *ZNF44* (Bassuk et al., 2013), associated with intellectual disabilities, also highlights the potential overlap between suicide risk and other neurodevelopmental and neurological conditions. Additionally, metabolic function (*FBXO36*) (Riveros-Mckay et al., 2020), lipid metabolism (*TM9SF3*) (Zhu et al., 2022), immune related functions (*PIPOX*, *IL4R*), and transcriptional repression (*ZHX3*) (Igata et al., 2022). *LMNTD1* is involved in structural molecule and may be associated with cancer risk and motor memory (Nguyen et al., 2022).

Study Limitations and Future Directions

Most of our suicide death cohort is of European ancestry, which may limit the generalizability of the results in other populations. Since the observed deletions were found predominantly in individuals of European ancestry, further studies in diverse populations are necessary to validate these findings across different genetic backgrounds. Replication in larger and independent cohorts will be needed to confirm these findings.

Additionally, while our study provides support for the role of intragenic deletions in suicide risk, characterizing the biological impact of these deletions on gene expression and function, and exploring potential interactions

with other genetic and environmental risk factors.

The next steps in this work will include a broader investigation that includes non-coding deletions, other structural variant types (e.g., duplications, inversions), and a more comprehensive assessment of additional putatively functional single nucleotide variants (SNVs). Future work will also include interactions of common polygenic backgrounds and/or environmental exposures with the occurrence of rare variants. The study of rare genetic variation leading to suicide risk will increase understanding of biological mechanisms and serve to identify potential drug targets with the ultimate goal of improved prevention and personalized interventions.

It is important to emphasize that having a deletion associated with suicide death does not deterministically confer suicide risk but rather represents just one component of multifaceted genetic and environmental interactions. Complex traits such as suicide are influenced by multiple genetic variants, each often exerting small, additive, or interactive effects, along with epigenetic modifications and environmental factors that modulate gene expression and phenotype manifestation. A single deletion may increase risk, but its impact is contingent on broader genomic context, gene-gene interactions, and external influences. Thus, while discovering deletions provides valuable insight into disease mechanisms, it should be interpreted as one piece of a larger, intricate puzzle rather than a definitive predictor of the trait

Conclusion

This work advances our understanding of the genetic factors contributing to suicide risk by examining the role of rare intragenic deletions. Our findings highlight the importance of integrating structural variant analysis with existing knowledge from GWAS and other genetic studies. As the field progresses, further research and replication in independent cohorts will be essential to confirm these results and understand the underlying biological mechanisms, ultimately contributing to improved strategies for suicide prevention.

References

- Abel, H. J., Larson, D. E., Regier, A. A., Chiang, C., Das, I., Kanchi, K. L., Layer, R. M., Neale, B. M., Salerno, W. J., Reeves, C., Buyske, S., Abecasis, G. R., Appelbaum, E., Baker, J., Banks, E., Bernier, R. A., Bloom, T., Boehnke, M., Boerwinkle, E., ... Hall, I. M. (2020). Mapping and characterization of structural variation in 17,795 human genomes. *Nature*, 583(7814). <https://doi.org/10.1038/s41586-020-2371-0>
- Andreassen, O. A., Hindley, G. F. L., Frei, O., & Smeland, O. B. (2023). New insights from the last decade of research in psychiatric genetics: discoveries, challenges and clinical implications. *World Psychiatry*, 22(1). <https://doi.org/10.1002/wps.21034>
- Bassuk, A. G., Geraghty, E., Wu, S., Mullen, S. A., Berkovic, S. F., Scheffer, I. E., & Mefford, H. C. (2013). Deletions of 16p11.2 and 19p13.2 in a family with intellectual disability and generalized epilepsy. *American Journal of Medical Genetics, Part A*, 161(7). <https://doi.org/10.1002/ajmg.a.35946>
- Belyeu, J. R., Chowdhury, M., Brown, J., Pedersen, B. S., Cormier, M. J., Quinlan, A. R., & Layer, R. M. (2021). Samplot: a platform for structural variant visual validation and automated filtering. *Genome Biology*, 22(1). <https://doi.org/10.1186/s13059-021-02380-5>
- Bonk, S., Eszlari, N., Kirchner, K., Gezsi, A., Garvert, L., Koukkanen, M., Cano, I., Grabe, H. J., Antal, P., Juhasz, G., & Van der Auwera, S. (2024). Impact of gene-by-trauma interaction in MDD-related multimorbidity clusters. *Journal of Affective Disorders*, 359, 382–391. <https://doi.org/10.1016/j.jad.2024.05.126>
- Børglum, A. D., Demontis, D., Grove, J., Pallesen, J., Hollegaard, M. V., Pedersen, C. B., Hedemand, A., Mattheisen, M., Uitterlinden, A., Nyegaard, M., Ørntoft, T., Wiuf, C., Didriksen, M., Nordentoft, M., Nöthen, M. M., Rietschel, M., Ophoff, R. A., Cichon, S., Yolken, R. H., ... Mors, O. (2014). Genome-wide study of association and interaction with maternal cytomegalovirus infection suggests new schizophrenia loci. *Molecular Psychiatry*, 19(3). <https://doi.org/10.1038/mp.2013.2>
- Byrska-Bishop, M., Evani, U. S., Zhao, X., Basile, A. O., Abel, H. J., Regier, A. A., Corvelo, A., Clarke, W. E., Musunuri, R., Nagulapalli, K., Fairley, S., Runnels, A., Winterkorn, L., Lowy, E., Eichler, E. E., Korbel, J. O., Lee, C., Marschall, T., Devine, S. E., ... Zody, M. C. (2022). High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *Cell*, 185(18). <https://doi.org/10.1016/j.cell.2022.08.004>
- Cannon, M., Stevenson, J., Stahl, K., Basu, R., Coffman, A., Kiwala, S., McMichael, J. F., Kuzma, K., Morrissey, D., Cotto, K., Mardis, E. R., Griffith, O. L., Griffith, M., & Wagner, A. H. (2024). DGIdb 5.0: rebuilding the drug-gene interaction database for precision medicine and drug discovery platforms. *Nucleic Acids Research*, 52(D1). <https://doi.org/10.1093/nar/gkad1040>
- Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., Cox, A. J., Kruglyak, S., & Saunders, C. T. (2016). Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*, 32(8). <https://doi.org/10.1093/bioinformatics/btv710>
- Coleman, J. R. I., Gaspar, H. A., Bryois, J., Byrne, E. M., Forstner, A. J., Holmans, P. A., de Leeuw, C. A., Mattheisen, M., McQuillin, A., Whitehead Pavlides, J. M., Pers, T. H., Ripke, S., Stahl, E. A., Steinberg, S., Trubetskoy, V., Trzaskowski, M., Wang, Y., Abbott, L., Abdellaoui, A., ... Breen, G. (2020). The Genetics of the Mood Disorder Spectrum: Genome-wide Association Analyses of More Than 185,000 Cases and 439,000 Controls. *Biological Psychiatry*, 88(2). <https://doi.org/10.1016/j.biopsych.2019.10.015>
- Collins, R. L., Brand, H., Karczewski, K. J., Zhao, X., Alföldi, J., Francioli, L. C., Khera, A. V., Lowther, C., Gauthier, L. D., Wang, H., Watts, N. A., Solomonson, M., O'Donnell-Luria, A., Baumann, A., Munshi, R., Walker, M., Whelan, C. W., Huang, Y., Brookings, T., ... Talkowski, M. E. (2020). A structural variation reference for medical and population genetics. *Nature*, 581(7809). <https://doi.org/10.1038/s41586-020-2287-8>
- Coon, H., Darlington, T. M., DiBlasi, E., Callor, W. B., Ferris, E., Fraser, A., Yu, Z., William, N., Das, S. C., Crowell, S. E., Chen, D., Anderson, J. S., Klein, M., Jerominski, L., Cannon, D., Shabalin, A., Docherty, A., Williams, M., Smith, K. R., ... Gray, D. (2020). Genome-wide significant regions in 43 Utah high-risk

- families implicate multiple genes involved in risk for completed suicide. *Molecular Psychiatry*, 25(11). <https://doi.org/10.1038/s41380-018-0282-3>
- Coon, H., Shabalin, A., Bakian, A. V., DiBlasi, E., Monson, E. T., Kirby, A., Chen, D., Fraser, A., Yu, Z., Staley, M., Callor, W. B., Christensen, E. D., Crowell, S. E., Gray, D., Crockett, D. K., Li, Q. S., Keeshin, B., & Docherty, A. R. (2022). Extended familial risk of suicide death is associated with younger age at death and elevated polygenic risk of suicide. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, 189(3–4). <https://doi.org/10.1002/ajmg.b.32890>
- Curtin, S. C., Garnett, M. F., & Ahmad, F. B. (2023). Provisional Estimates of Suicide by Demographic Characteristics: United States, 2022. *Vital Statistics Rapid Release*, 34(34).
- Dausset, J., Cann, H., Cohen, D., Lathrop, M., Lalouel, J. M., & White, R. (1990). Centre d'Etude du polymorphisme humain (CEPH): Collaborative genetic mapping of the human genome. *Genomics*, 6(3). [https://doi.org/10.1016/0888-7543\(90\)90491-C](https://doi.org/10.1016/0888-7543(90)90491-C)
- Denny, J. C., Ritchie, M. D., Basford, M. A., Pulley, J. M., Bastarache, L., Brown-Gentry, K., Wang, D., Masys, D. R., Roden, D. M., & Crawford, D. C. (2010). PheWAS: Demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics*, 26(9). <https://doi.org/10.1093/bioinformatics/btq126>
- Depristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., Del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernysky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43(5). <https://doi.org/10.1038/ng.806>
- DiBlasi, E., Kang, J., & Docherty, A. R. (2021). Genetic contributions to suicidal thoughts and behaviors. *Psychological Medicine*, 51(13). <https://doi.org/10.1017/S0033291721001720>
- DiBlasi, E., Shabalin, A. A., Monson, E. T., Keeshin, B. R., Bakian, A. V., Kirby, A. V., Ferris, E., Chen, D., William, N., Gaj, E., Klein, M., Jerominski, L., Callor, W. B., Christensen, E., Smith, K. R., Fraser, A., Yu, Z., Gray, D., Camp, N. J., ... 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(8). <https://doi.org/10.1002/ajmg.b.32861>
- Docherty, A. R., Mullins, N., Ashley-Koch, A. E., Qin, X., Coleman, J. R. I., Shabalin, A., Kang, J. E., Murnyak, B., Wendt, F., Adams, M., Campos, A. I., DiBlasi, E., Fullerton, J. M., Kranzler, H. R., Bakian, A. V., Monson, E. T., Rentería, M. E., Walss-Bass, C., Andreassen, O. A., ... Ruderfer, D. M. (2023). GWAS Meta-Analysis of Suicide Attempt: Identification of 12 Genome-Wide Significant Loci and Implication of Genetic Risks for Specific Health Factors. *American Journal of Psychiatry*, 180(10). <https://doi.org/10.1176/appi.ajp.21121266>
- Fabbri, C., Kasper, S., Kautzky, A., Bartova, L., Dold, M., Zohar, J., Souery, D., Montgomery, S., Albani, D., Raimondi, I., Dikeos, D., Rujescu, D., Uher, R., Lewis, C. M., Mendlewicz, J., & Serretti, A. (2019). Genome-wide association study of treatment-resistance in depression and meta-analysis of three independent samples. *British Journal of Psychiatry*, 214(1). <https://doi.org/10.1192/bjp.2018.256>
- Freed, D., Aldana, R., Weber, J. A., & Edwards, J. S. (2017). The Sentieon Genomics Tools - A fast and accurate solution to variant calling from next-generation sequence data. *BioRxiv*.
- Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B., & Overington, J. P. (2012). ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Research*, 40(D1). <https://doi.org/10.1093/nar/gkr777>
- Génin, E. (2020). Missing heritability of complex diseases: case solved? In *Human Genetics* (Vol. 139, Issue 1). <https://doi.org/10.1007/s00439-019-02034-4>
- He, H., Guzman, R. E., Cao, D., Sierra-Marquez, J., Yin, F., Fahlke, C., Peng, J., & Stauber, T. (2021). The molecular and phenotypic spectrum of CLCN4-related epilepsy. *Epilepsia*, 62(6). <https://doi.org/10.1111/epi.16906>
- Ho, S. S., Urban, A. E., & Mills, R. E. (2020). Structural variation in the sequencing era. In *Nature Reviews Genetics* (Vol. 21, Issue 3). <https://doi.org/10.1038/s41576-019-0180-9>

- Igata, T., Tanaka, H., Etoh, K., Hong, S., Tani, N., Koga, T., & Nakao, M. (2022). Loss of the transcription repressor ZHX3 induces senescence-associated gene expression and mitochondrial-nucleolar activation. *PLoS ONE*, *17*(1 January). <https://doi.org/10.1371/journal.pone.0262488>
- Kimbrel, N. A., Ashley-Koch, A. E., Qin, X. J., Lindquist, J. H., Garrett, M. E., Dennis, M. F., Hair, L. P., Huffman, J. E., Jacobson, D. A., Madduri, R. K., Trafton, J. A., Coon, H., Docherty, A. R., Mullins, N., Ruderfer, D. M., Harvey, P. D., McMahon, B. H., Oslin, D. W., Beckham, J. C., ... Hauser, M. A. (2023). Identification of Novel, Replicable Genetic Risk Loci for Suicidal Thoughts and Behaviors Among US Military Veterans. *JAMA Psychiatry*, *80*(2). <https://doi.org/10.1001/jamapsychiatry.2022.3896>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14). <https://doi.org/10.1093/bioinformatics/btp324>
- Li, Q. S., Shabalín, A. A., DiBlasi, E., Gopal, S., Canuso, C. M., Palotie, A., Drevets, W. C., Docherty, A. R., & Coon, H. (2022). Genome-wide association study meta-analysis of suicide death and suicidal behavior. *Molecular Psychiatry*. <https://doi.org/10.1038/s41380-022-01828-9>
- McGuffin, P., Marusic, A., & Farmer, A. (2001). What can psychiatric genetics offer suicidology? In *Crisis* (Vol. 22, Issue 2). <https://doi.org/10.1027//0227-5910.22.2.61>
- Miller, J. N., & Black, D. W. (2020). Bipolar Disorder and Suicide: a Review. In *Current Psychiatry Reports* (Vol. 22, Issue 2). <https://doi.org/10.1007/s11920-020-1130-0>
- Mullins, N., Forstner, A. J., O'Connell, K. S., Coombes, B., Coleman, J. R. I., Qiao, Z., Als, T. D., Bigdeli, T. B., Børte, S., Bryois, J., Charney, A. W., Drange, O. K., Gandal, M. J., Hagenaars, S. P., Ikeda, M., Kamitaki, N., Kim, M., Krebs, K., Panagiotaropoulou, G., ... Andreassen, O. A. (2021). Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nature Genetics*, *53*(6). <https://doi.org/10.1038/s41588-021-00857-4>
- Mullins, N., Kang, J. E., Campos, A. I., Coleman, J. R. I., Edwards, A. C., Galfalvy, H., Levey, D. F., Lori, A., Shabalín, A., Starnawska, A., Su, M. H., Watson, H. J., Adams, M., Awasthi, S., Gandal, M., Hafferty, J. D., Hishimoto, A., Kim, M., Okazaki, S., ... Willour, V. (2022). Dissecting the Shared Genetic Architecture of Suicide Attempt, Psychiatric Disorders, and Known Risk Factors. *Biological Psychiatry*, *91*(3). <https://doi.org/10.1016/j.biopsych.2021.05.029>
- Nguyen, H. D., Jo, W. H., Hoang, N. H. M., & Kim, M. S. (2022). In silico identification of the potential molecular mechanisms involved in protective effects of prolactin on motor and memory deficits induced by 1,2-Diacetylbenzene in young and old rats. *NeuroToxicology*, *93*. <https://doi.org/10.1016/j.neuro.2022.09.002>
- Nicholas, T. J., Al-Sweel, N., Farrell, A., Mao, R., Bayrak-Toydemir, P., Miller, C. E., Bentley, D., Palmquist, R., Moore, B., Hernandez, E. J., Cormier, M. J., Fredrickson, E., Noble, K., Rynearson, S., Holt, C., Karren, M. A., Bonkowsky, J. L., Tristani-Firouzi, M., Yandell, M., ... Malone Jenkins, S. (2022). Comprehensive variant calling from whole-genome sequencing identifies a complex inversion that disrupts ZFPM2 in familial congenital diaphragmatic hernia. *Molecular Genetics and Genomic Medicine*, *10*(4). <https://doi.org/10.1002/mgg3.1888>
- Owen, M. J., & Williams, N. M. (2021). Explaining the missing heritability of psychiatric disorders. In *World Psychiatry* (Vol. 20, Issue 2). <https://doi.org/10.1002/wps.20870>
- Paulovičová, E., Kroneková, Z., Paulovičová, L., Majerčíková, M., & Kronek, J. (2021). Cell-mediated immunoreactivity of poly(2-isopropenyl-2-oxazoline) as promising formulation for immunomodulation. *Materials*, *14*(6). <https://doi.org/10.3390/ma14061371>
- Pedersen, B. S., Brown, J. M., Dashnow, H., Wallace, A. D., Velinder, M., Tristani-Firouzi, M., Schiffman, J. D., Tvrdik, T., Mao, R., Best, D. H., Bayrak-Toydemir, P., & Quinlan, A. R. (2021). Effective variant filtering and expected candidate variant yield in studies of rare human disease. *Npj Genomic Medicine*, *6*(1). <https://doi.org/10.1038/s41525-021-00227-3>
- Pedersen, B. S., & Quinlan, A. R. (2019). Duphold: Scalable, depth-based annotation and curation of high-confidence structural variant calls. *GigaScience*, *8*(4). <https://doi.org/10.1093/gigascience/giz040>
- Pedersen, N. L., & Fiske, A. (2010). Genetic influences on suicide and nonfatal suicidal behavior: Twin study findings. *European Psychiatry*, *25*(5). <https://doi.org/10.1016/j.eurpsy.2009.12.008>

- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., De Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3). <https://doi.org/10.1086/519795>
- Riveros-Mckay, F., Oliver-Williams, C., Karthikeyan, S., Walter, K., Kundu, K., Ouwehand, W. H., Roberts, D., Angelantonio, E. Di, Soranzo, N., Danesh, J., Wheeler, E., Zeggini, E., Butterworth, A. S., & Barroso, I. (2020). The influence of rare variants in circulating metabolic biomarkers. *PLoS Genetics*, *16*(3). <https://doi.org/10.1371/journal.pgen.1008605>
- Ruggeri, B., Nymberg, C., Vuoksima, E., Lourdasamy, A., Wong, C. P., Carvalho, F. M., Jia, T., Cattrell, A., Macare, C., Banaschewski, T., Barker, G. J., Bokde, A. L. W., Bromberg, U., Büchel, C., Conrod, P. J., Fauth-Bühler, M., Flor, H., Frouin, V., Gallinat, J., ... Schumann, G. (2015). Association of protein phosphatase PPM1G with alcohol use disorder and brain activity during behavioral control in a genome-wide methylation analysis. *American Journal of Psychiatry*, *172*(6). <https://doi.org/10.1176/appi.ajp.2014.14030382>
- Salatino-Oliveira, A., Genro, J. P., Polanczyk, G., Zeni, C., Schmitz, M., Kieling, C., Anselmi, L., Menezes, A. M. B., Barros, F. C., Polina, E. R., Mota, N. R., Grevet, E. H., Bau, C. H. D., Rohde, L. A., & Hutz, M. H. (2015). Cadherin-13 gene is associated with hyperactive/impulsive symptoms in attention/deficit hyperactivity disorder. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, *168*(3). <https://doi.org/10.1002/ajmg.b.32293>
- Schoormans, D., Radonic, T., de Witte, P., Groenink, M., Azim, D., Lutter, R., Mulder, B. J. M., Sprangers, M. A. G., & Zwinderman, A. H. (2012). Mental Quality of Life Is Related to a Cytokine Genetic Pathway. *PLoS ONE*, *7*(9). <https://doi.org/10.1371/journal.pone.0045126>
- Shuey, M. M., Stead, W. W., Aka, I., Barnado, A. L., Bastarache, J. A., Brokamp, E., Campbell, M., Carroll, R. J., Goldstein, J. A., Lewis, A., Malow, B. A., Mosley, J. D., Osterman, T., Padovani-Claudio, D. A., Ramirez, A., Roden, D. M., Schuler, B. A., Siew, E., Sucre, J., ... Bastarache, L. (2023). Next-generation phenotyping: introducing phecodeX for enhanced discovery research in medical phenomics. *Bioinformatics*, *39*(11). <https://doi.org/10.1093/bioinformatics/btad655>
- Sollis, E., Mosaku, A., Abid, A., Buniello, A., Cerezo, M., Gil, L., Groza, T., Güneş, O., Hall, P., Hayhurst, J., Ibrahim, A., Ji, Y., John, S., Lewis, E., Macarthur, J. A. L., McMahon, A., Osumi-Sutherland, D., Panoutsopoulou, K., Pendlington, Z., ... Harris, L. W. (2023). The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Research*, *51*(1 D). <https://doi.org/10.1093/nar/gkac1010>
- Statham, D. J., Heath, A. C., Madden, P. A. F., Bucholz, K. K., Bierut, L., Dinwiddie, S. H., Slutske, W. S., Dunne, M. P., & Martin, N. G. (1998). Suicidal behaviour: An epidemiological and genetic study. *Psychological Medicine*, *28*(4). <https://doi.org/10.1017/S0033291798006916>
- Tschanz, J. T., Corcoran, C., Skoog, I., Khachaturian, A. S., Herrick, J., Hayden, K. M., Welsh-Bohmer, K. A., Calvert, T., Norton, M. C., Zandi, P., & Breitner, J. C. S. (2004). Dementia: The leading predictor of death in a defined elderly population: The Cache County Study. *Neurology*, *62*(7). <https://doi.org/10.1212/01.WNL.0000118210.12660.C2>
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, *SUPL.43*. <https://doi.org/10.1002/0471250953.bi1110s43>
- Vogelaar, C. F., Mandal, S., Lerch, S., Birkner, K., Birkenstock, J., Bühler, U., Schnatz, A., Raine, C. S., Bittner, S., Vogt, J., Kipnis, J., Nitsch, R., & Zipp, F. (2018). Fast direct neuronal signaling via the IL-4 receptor as therapeutic target in neuroinflammation. *Science Translational Medicine*, *10*(430). <https://doi.org/10.1126/scitranslmed.aao2304>
- Weischenfeldt, J., Symmons, O., Spitz, F., & Korbel, J. O. (2013). Phenotypic impact of genomic structural variation: Insights from and for human disease. In *Nature Reviews Genetics* (Vol. 14, Issue 2). <https://doi.org/10.1038/nrg3373>

Zhu, L., Liao, R., Huang, J., Yan, H., Xiao, C., Yang, Y., Wang, H., & Yang, C. (2022). The miR-216/miR-217 Cluster Regulates Lipid Metabolism in Laying Hens With Fatty Liver Syndrome via PPAR/SREBP Signaling Pathway. *Frontiers in Veterinary Science*, 9. <https://doi.org/10.3389/fvets.2022.913841>

Acknowledgements

This work was supported by the National Institute of Health (HC, grant number R01MH122412, R01MH123489; AD, grant number R01MH123619; AVB, grant number R01ES032028); the American Foundation for Suicide Prevention (ED), the Brain & Behavior Research Foundation--NARSAD (ED, grant number 28132; AS grant number 28686); and the Clark Tanner Foundation (HC, AS, EM, AVB). Partial support for all datasets housed within the Utah Population Data Base is provided by the Huntsman Cancer Institute (HCI), <http://www.huntsmancancer.org/>, and the HCI Cancer Center Support grant, P30CA42014 from the National Cancer Institute. Research was supported by NCCR grant “Sharing statewide health data for genetic research” R01RR021746 with additional support from the Utah Department of Health and Human Services and the University of Utah. We thank University of Utah Health Data Science Services for data and analytics support, and the University of Utah Pedigree and Population Resource and the University of Utah Health Enterprise Data Warehouse for establishing the Master Subject Index between the Utah Population Database and the University of Utah Health Sciences Center. DNA extraction was performed by the University of Utah Center for Clinical and Translational Science supported by the National Center for Advancing Translational Sciences of the NIH (grant number UL1TR002538). Genotyping was performed by the University of Utah Genomic Core (UL1TR002538) and by Illumina, Inc. with support from Janssen Research & Development, LLC. The support and resources from the Center for High Performance Computing at the University of Utah are gratefully acknowledged.

Table 1. Demographic characteristics of Utah suicide deaths with WGS data.

Cohort	N with WGS	% female	Avg. age at death in years (sd)	% European ancestry	N with bipolar disorder
Cohort 1	670	24.25%	28.87 (sd=13.08)	97.72%	145
Cohort 2	384	34.12%	37.84 (sd=12.21)	98.58%	180
All suicides with WGS	1,054	27.84%	32.13 (sd=13.47)	98.10%	325

Note. Demographic age and sex data and clinical data omits 5 suicides known to have mismatched sex between records and genotyping. As all collected samples were suicide deaths, these 5 cases were included for ancestry tests.

Table 2. Manually validated deletions present in at least two suicides in each of the two suicide death cohorts and showing > two-fold increase in frequency over jointly processed control sequence data.

Ch r	Start (hg38)	End	Length	Gene	Freq, jointly called controls ¹	Freq, suicid e deaths	Fold increase in suicides ²	Binomial test p- value ³	Compariso nfreq, in non-neuro GnomAD ⁴
22	37019286	37024652	-5366	<i>MPST</i>	0.000784	0.00381	4.86	3.12E-04	0.00089
17	29056851	29058075	-1224	<i>PIPOX</i>	0.000784	0.00381	4.86	3.12E-04	0.00102
2	229931901	229954572	-22671	<i>FBXO 36</i>	0.000876	0.00333	3.81	2.88E-03	0.00102
12	25494303	25501878	-7575	<i>LMNT D1</i>	0.002393	0.00905	3.78	1.44E-06	Not reported
10	96521588	96521862	-274	<i>TM9S F3</i>	0.001429	0.00476	3.33	1.09E-03	0.00168
19	12292303	12294048	-1745	<i>ZNF44</i>	0.002443	0.00762	3.12	8.99E-05	0.00294
20	62294459	62295274	-815	<i>OSBP L2</i>	0.001405	0.00429	3.05	3.39E-03	0.00102
20	41184282	41184388	-106	<i>ZHX3</i>	0.001106	0.00333	3.01	9.79E-03	0.00132
1	86562531	86573603	-11072	<i>CLCA 4</i>	0.002291	0.00571	2.49	4.02E-03	0.00240
16	27325117	27339881	-14764	<i>ILAR</i>	0.003522	0.00810	2.30	1.66E-03	Not reported
16	83162542	83176121	-13579	<i>CDHI 3</i>	0.004146	0.00857	2.07	3.75E-03	0.00450

¹ Jointly processed control data was chosen to match for ancestry from three sources: N=622 from the 1000 Genomes Project; N=512 unrelated individuals from Utah CEPH genetic reference families; N=96 from Utah study of longevity.

² Fold increase in suicides indicates the ratio of the frequency in suicides to the frequency in controls.

Notes. Listed frequencies in suicides are of deletions that were manually validated. A total of 25 deletions in suicides were prioritized with requirements of: meeting quality, length, and frequency criteria to minimize false positive occurrences (see text); and occurrence in at least two suicides in the initial cohort of 670 suicides and at least two suicides in the second set of 384 suicides (to increase robustness).

³ Significance tests compared suicide deaths to jointly called controls. The p-values listed were all significant at a false discovery rate (FDR) < 0.05.

⁴ These frequencies are given for comparison only, as data were not jointly processed. GnomAD is the Genome Aggregation Database; the “non-neuro” subset in GnomAD v2.1 (N=16,684) omits data from individuals with neuropsychiatric conditions.

Table 3. Additional characterizations of implicated genes.

Gene implicated by deletion	GTE _x expression of SNPs within deletion regions ¹	GWAS Catalog associations	DGI _{db} gene-drug interactions; DRUGBANK drug targets
<i>MPST</i>	No del SNPs overlap with GTE _x	Hemoglobin, gut microbiome, metabolite levels, putamen	DGI _{db} : Thyrozine (thyroid hormone levels)
<i>PIPOX</i>	4 SNPs, adrenal gland expression	MDD, Depression, AD, BMI, accelerated aging, language, smoking heaviness, smoking initiation	DRUGBANK DB00145, Glycine (nutraceutical; used for stroke, MI; fast inhibitory neurotransmitter, possible antipsychotic activity)
<i>FBXO36</i>	19 SNPs, 4 with brain expression	Triglycerides, creatinine, fat mass, BMI	None
<i>LMNTD1</i>	4 SNPs, testis expression	Gut microbiome, monocyte count	DGI _{db} : Capecitabine (cytotoxic agent, cancer treatment)
<i>TM9SF3</i>	No del SNPs overlap with GTE _x	Bone density	None
<i>ZNF44</i>	4 SNPs, 3 expressed in brain	Alcoholic liver disease	None
<i>OSBPL2</i>	1 SNP expressed in testis	Bipolar, platelet volume, nicotine dependence, reticulocyte fraction, cortical folding, educational attainment	Non
<i>ZHX3</i>	No del SNPs overlap with GTE _x	Unipolar depression, daytime napping, sleep duration, MDD, depressive symptoms, drinks per week, migraine, triglycerides, cholesterol, CAD, BMI, waist-hip ratio	None
<i>CLCA4</i>	11 SNPs, all associated with brain expression	IgG glycosylation, macrophage inflammatory protein 1b levels	None
<i>IL4R</i>	27 SNPs, 17 with brain expression	Asthma, allergy, C-reactive protein, respiratory diseases, IgE levels, eosinophil counts	DGI _{db} : Sunitinib (kinase inhibitor); DGI _{db} and DRUGBANK (DB12159): Dupilumab (monoclonal antibody, asthma, COPD, dermatitis); DRUGBANK: DB05078m AER001 (asthma treatment)
<i>CDH13</i>	No del SNPs overlap with GTE _x	ADHD, smoking initiation, clozapine response, amphetamine response, SZ, AD onset age, educational attainment, cognitive ability, CAD, blood pressure, adiponectin levels, COPD treatment response, BMI, antibody response to COVID-19 vaccine	Deutetrabenain (tardive dyskinesia/chorea in HD, NOTE: may increase suicidality); Dextro-amphetamine sulfate (ADHD, cognitive dysfunction) Clopidogrel (plavix; anti-thrombotic)

¹ Further detail of SNP names, p-values, and GTE_x-associated tissues appears in Supplemental Table S3. See Supplemental information for GWAS Catalog references.

Table 4. Demographic and clinical characteristics of suicide deaths with validated deletions.

All samples and by gene implicated by deletion	Avg. Age at death (SD)	% Female	% Anxiety and anxiety disorders	% Bipolar disorder	% Mood [affective] disorders	% Pain	% Psychoactive substance related disorders	% Psychotic disorder	% Sleep disorders
<i>MPST</i>	28.3 (11.8)	37.5	44.4	22.2	66.7	11.1	33.3	11.1	11.1
<i>PIPOX</i>	36.1 (16.7)	12.5	62.5	37.5	62.5	50.0	37.5	12.5	50.0
<i>FBXO36</i>	28.3 (15.0)	14.2	28.6	14.3	42.9	57.1	28.6	0	28.6
<i>LMNTD1</i>	35.8 (13.1)	36.8	57.9	47.4	73.7	36.8	42.1	21.1	21.1
<i>TM9SF3</i>	32.7 (11.5)	40.0	55.6	55.6	100	55.6	55.6	22.2	22.2
<i>ZNF44</i>	37.1 (13.6)	43.8	68.8	31.3	87.5	31.3	56.3	37.5	50.0
<i>OSBPL2</i>	36.3 (17.4)	33.3	44.4	66.7	66.7	33.3	44.4	33.3	55.6
<i>ZHX3</i>	39.00 (10.4)	42.9	71.4	71.4	100	42.9	71.4	28.6	14.3
<i>CLCA4</i>	37.7 (13.0)	16.7	66.7	85.7	83.3	50.0	50.0	41.7	66.7
<i>IL4R</i>	38.1 (13.5)	23.5	53.0	41.2	76.5	41.2	58.8	29.4	29.4
<i>CDH13</i>	32.8 (13.1)	22.2	44.4	55.6	66.7	61.1	44.4	16.7	27.8
All suicides with WGS	32.1 (13.5)	27.8	50.6	42.0	69.1	36.02	48.2	21.3	30.1