bioRxiv preprint doi: https://doi.org/10.1101/195644; this version posted September 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Identification of genome-wide significant shared genomic segments in large extended Utah families at high risk for completed suicide

Hilary Coon PhD,[1,£] Todd M. Darlington PhD,[1] W. Brandon Callor MS,[2] Elliott Ferris MS,[3] Alison Fraser MS,[4] Zhe Yu MS,[4] Nancy William MS,[1] Sujan C. Das PhD,[1] Sheila E. Crowell PhD,[5] Megan Puzia MS,[5] Michael Klein MS,[6] Anna Docherty PhD,[1] Leslie Jerominski MS,[1] Dale Cannon PhD,[1] Ken R. Smith PhD,[4] Brooks Keeshin MD,[7] Amanda V. Bakian PhD,[1] Erik Christensen MD,[2] Nicola J. Camp PhD,[8] Douglas Gray MD[1]

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

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

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

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

[5] Department of Psychology, University of Utah, Salt Lake City, UT

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

[7] Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, UT

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

[£] Corresponding author; address for correspondence: Hilary Coon, University of Utah Department of Psychiatry, 650 Komas Drive, Suite 206, Salt Lake City, UT 84108.

bioRxiv preprint doi: https://doi.org/10.1101/195644; this version posted September 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Abstract

Suicide is the 10th leading cause of death in the US. While environment has undeniable impact, evidence suggests that genetic factors play a major role in completed suicide. We have >4,500 DNA samples from completed suicides through a collaboration with the Utah Medical Examiner. We have linked the records from these cases to the Utah Population Database which includes multi-generation genealogies, demographic data, and medical information on over 8 million individuals. This linking has resulted in extended families (7-9 generations) with significant familial risk of completed suicide. Familial aggregation across distant relatives minimizes effects of shared environment, provides more genetically homogeneous risk groups, and magnifies genetic risks through familial repetition. We analyzed DNA from 215 suicide cases in 43 of our largest high-risk families and identified 16 regions with genome-wide significance in 10 families. Of the 163 genes in these regions, 25% were associated with psychiatric risk. We also found 13 regions with genome-wide suggestive evidence where the region overlaps in >1 family (p-values from 4.63E-09 to <1E-16). Of the 101 genes in these overlapping regions, seven have been previously associated with suicide risk (*RGS18, BRINP3, RHEB, CDK5, CTNNA3, STAT1*, and *HTR2A)*; only one gene with specific suicide risk would have been expected by chance. Our most significant region on chromosome 5q23.3 was shared by 21 cases across three families. This region contains several genes associated with both psychiatric conditions and inflammation: *HINT1*, *RAPGEF6*, *ACSL6*, *IL3*, S*LC22A4*, *CSF2*, and *IRF1*. Our study provides 249 genes in the significant regions in our study represent important new candidate genes for suicide. The genome-wide significant family-specific regions may reveal more rare risk variants, while risk variants in regions that overlap across families may be more common.

bioRxiv preprint doi: https://doi.org/10.1101/195644; this version posted September 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Author Summary

Suicide is the 10th leading cause of death in the US. While environment has undeniable impact, evidence suggests that genetic factors play a major role in completed suicide. We have used DNA from suicide cases related to each other in very large extended high-risk families (7-9 generations) to discover regions of the genome likely to contain genetic changes leading to increased suicide risk. Studying these distantly related cases minimizes effects of shared environment and allows us to detect genetic risks more easily through their familial repetition. The genomic regions discovered in this study of our 43 largest high-risk families contained seven genes with corroborating evidence of association with suicide from previous studies, in addition to many genes with known psychiatric associations. We also discovered genes with links to other processes, such as inflammation or immune response, suggesting possible new ways to study suicide risk. The genes identified by this study will stimulate further study in our data resources, and in studies of suicide risk in the worldwide scientific community.

bioRxiv preprint doi: https://doi.org/10.1101/195644; this version posted September 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Introduction

Suicide is the 10th leading cause of death in the United States; over 44,000 individuals die by suicide in the US every year.[1] While environmental variables have undeniable impact, evidence suggests that genetic factors play a role in completed suicide, with heritability estimated at 45%.[2,3] Discovery of specific genetic risk variants could result in better insight into biological mechanisms of risk. Recent growth in the number of suicide genetic studies has resulted in promising genetic findings from candidate gene and genome-wide association studies,[4] though not surprisingly given the high degree of predicted genetic heterogeneity for suicide risks, replicated findings have been rare. Replication is also hampered by sample differences across studies, including differences in demographics, primary diagnoses, or outcomes, which range from suicidal ideation and behaviors to completed suicide; evidence suggests these etiologies may be distinct.[5] Statistical issues remain challenging. While candidate gene studies are reasonably well powered, they assume the correct choice of candidate gene(s), a premise not well supported by non-replication in psychiatric genetics research.[6] Genome-wide association (GWAS) studies avoid gene choices, but heterogeneity and complexity dictate the use of very large sample sizes to achieve acceptable statistical power.

The Utah Suicide Research Group has a unique opportunity to study genetic risk for familial suicide while addressing multiple study design limitations. We have a resource of >4,500 DNA samples from completed suicide cases, and permission to link the records from these samples to the Utah Population Database (UPDB, https://healthcare.utah.edu/huntsmancancerinstitute/research/updb), a computerized database including multi-generational genealogies, as well as death certificates, demographic data, and current medical information on over 8 million individuals. Through this linking we have ascertained very large families with significantly elevated suicide risk and at least three familial cases with DNA. Familial aggregation across distant relatives in these families minimizes the impact of shared environment. The families also may provide more genetically homogeneous risk groups, and they increase statistical power to detect familial variants. Using the Shared Genomic Segments method (SGS),[7] we can identify genomic regions that have high likelihood of harboring variants leading to risk within a family.

In this report, we give details of analyses of 43 large Utah families at high risk for suicide. We have identified 16 chromosomal regions with genome-wide significance for familial sharing in a single family, and an additional 13 regions where the shared familial region overlaps in at least two families. There are 249 genes in these regions that we are

bioRxiv preprint doi: https://doi.org/10.1101/195644; this version posted September 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

currently following up in our data, and that can be followed up in other data resources in the research community.

Results

Descriptive results. The familial risk statistics (Familial Standardized Incidence Ratio, FSIR)[8] for the 43 families ranged from 2.04-4.41 times the expected risk (p=0.003 to <0.0001; Table 1). The average number of cases per family with genotyping was 6.2 (range 3-13), and the average number of meioses between analyzed cases was 29.6 (range 15-70). Theoretically derived family-specific genome-wide significant and suggestive thresholds for SGS depended upon family size, structure, dispersion of cases, and number of cases analyzed. SGS analyses were done on all 43 families. Most of the families (35/43=81.4%) gave at least one genome-wide suggestive region, and 10 families (23.3%) gave at least one genome-wide significant region.

The genealogies are complex. Although the total number of cases across all families listed in Table 1 is 267, 52 of these cases occurred in multiple families. See Figure 1 for an example of this complexity. Because analyses to identify genomic shared regions are done within family, we included cases each time they occurred under each founder in our analyses, as we don't know *a priori* where true sharing may occur. It is possible that, for example, cases 38882 and 44679 shown in Fig. 1 share risk variant(s) from the founders of 209487 with other cases in that family, but then also share other risk variant(s) from founder 233769 with other cases in this second family. The inclusion of cases in more than one family is therefore done to ensure we capture all possible evidence of sharing.

Table 2 gives descriptive information for the 43 selected families. We removed all duplicates from the 43 families, leaving the 215 independent cases, and compared descriptive data with the larger sample of cases with DNA. Within the 43 families, 172 cases were male (80.0%), similar to the 79.2% rate in the larger sample. Average age at death in the family sample was 34.28 years (standard deviation=16.28), significantly lower than the average age of 40.01 years (standard deviation=17.39) in the larger sample (t=4.74; p<0.0001). Method of suicide in the family sample was predominantly gun-related (110/215=51.1%), followed by other violent methods (78/215=36.3%), then overdose (27/215=12.6%), similar to rates in the larger sample of 52.6%, 32.0%, and 15.3%, respectively. Death certificate data identified 212/215 cases as White-Non-Hispanic, and three as Black-Non-Hispanic (one case each in families 41469, 233769, and 587072), similar to the larger sample, where death certificate data identified 96.89% as White-Non-

bioRxiv preprint doi: https://doi.org/10.1101/195644; this version posted September 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Hispanic. A principal component race/ethnicity analysis of the SNP data from our 215 cases and 1000 Genomes population data confirmed the three Black cases (though two were also of Hispanic ethnicity), and identified 11 other cases with evidence of Asian and/or Hispanic ancestry, giving overall rates of 3.3% non-White and 5.6% Hispanic cases.

Shared Genomic Segments (SGS)[7] results: genome-wide significant regions. SGS analyses revealed 16 regions with p-values below our derived thresholds for genome-wide significance. These regions occurred in 10 of the 43 families (Table 3). Several families showed evidence for more than one region; these were the larger families, where there was more opportunity for multiple subsets to show sharing evidence. Diagrams of each of these families are presented in Supplemental Figure S1.

Genes of interest in significant family-specific regions. Supplemental Table S1 shows all genes (N=163) with coding or regulatory sequence in the 16 genome-wide significant regions. Our search of the literature revealed 41/163=25% of these genes had prior psychiatric associations (highlighted in Table S1). Variants associated with other non-psychiatric genes in these regions may also be important; comprehensive follow-up analyses of the significant regions are in progress.

SGS regions overlapping in more than one family. In addition to the genome-wide significant regions, there were 145 regions in 35 of the 43 families with p-values below derived genome-wide suggestive thresholds (Supplemental Table S2). Because these suggestive regions are more likely to reflect false positive results, current follow-up efforts are focused on the 13 suggestive regions where evidence also overlapped across more than one family (Table 4). The p-values of these overlapping regions were approximated using Fisher's combined probability test.[11] All overlapping regions were supported by independent cases, with one exception. For the region on chromosome 5, person 112304 is a descendant of both 553615 and 603471, and person 95765 is a descendant in both 553615 and 176860. To satisfy the independence requirement for computing the Fisher's combined p-value, we omitted these two shared cases from family 553615 and recomputed the familial significance for the Fisher's p-value estimate.

Supplementary table S3 shows the subset of genes with coding or regulatory sequence in the regions overlapping in more than one family (N=101). Seven genes in these regions have been previously implicated in suicide risk in the literature: regulator of G-protein signaling 18 (*RGS18*), retinoic acid inducible neural specific 3 (*BRINP3*), signal transducer and activator of transcription 1 (*STAT1*), Ras homolog enriched in brain (*RHEB*), cyclin-dependent kinase 5

bioRxiv preprint doi: https://doi.org/10.1101/195644; this version posted September 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

(*CDK5*), catenin (cadherin-associated protein) alpha 3 (*CTNNA3*), and 5-hydroxytryptamine (serotonin) receptor 2A (*HTR2A*). Other genes in these regions are of high interest due to neuronal and/or psychiatric involvement.

Our most significant region shared by 21 cases across three families was on chromosome 5q23.3-q.31.1. This region contains several genes associated with both psychiatric conditions and inflammation: histidine triad nucleotide binding protein gene (*HINT1*), Rap guanine nucleotide exchange factor (*RAPGEF6*), Acyl-CoA synthetase long-chain family member 6 (*ACSL6*), interleukin 3 (*IL3*), solute carrier family 22, member 4 (S*LC22A4*), colony stimulating factor 2 gene (*CSF2*), and interferon regulatory factor 1 (*IRF1*). Supplemental Figure S2 shows families that contribute to these overlapping regions.

A summary of the phenotypic data available for cases driving the evidence of sharing in the most significant overlapping region (chromosome 5q) and in the regions with genes previously implicated in suicide risk appears in Supplementary Table S4.

Discussion

This study applies a powerful analysis method, SGS,[7] to a unique resource of 43 extended high-risk families. The method allows for the identification of genomic regions likely to harbor risk variants, accounting for likely within-family heterogeneity. The families in our analysis have suicide cases where the average age at death was significantly lower than the age at death in our larger cohort with DNA ($p<0.0001$), a result that may reflect a greater genetic liability in younger cases. Families are predominantly White, as verified using the genotyping data. Our analyses focus on genomic segments inherited from the ancestors within families, so non-White race/ethnicity is of genetic importance only for other novel genetic risk that may interact with familial risk. Such interaction is of interest, but is beyond the scope of our current study.

A quarter of the 163 genes in the genome-wide significant regions had prior psychiatric associations (Table S1). Follow-up studies of these and other genes in these significant regions are ongoing. Though risk variants in these significant regions may be unique to each family, it is possible that they still may shed light on mechanisms of risk. Gene pathway analyses of the regions are premature in the absence of more definitive knowledge of specific risk variants within the regions. Each region contains multiple genes, and we do not yet have sufficient data to determine which

bioRxiv preprint doi: https://doi.org/10.1101/195644; this version posted September 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

gene(s) may drive the shared genomic segment result. A gene pathway analysis may spuriously reflect the known genomic co-localization of genes with similar function,[12] rather than pathways that truly indicate suicide risk.

A current focus of our work is on significant genomic regions shared across more than one high-risk family. We are now generating sequence to analyze variants in these regions, with initial attention to seven genes with previous evidence for suicide risk in the literature. Compared to the estimated number of ~19,000 genes in the human genome,[13] there are a relatively small number of genes in the literature specifically associated with suicide: about 100 genes from candidate gene, expression, and epigenetic studies, and approximately another 100 from GWAS studies[4] (200/19000=~1%). From our overlapping regions, 7/101=6.93% have prior suicide associations, significantly greater than the expected rate (Z=5.69, p<0.0001). Follow up studies in our data resource focus on these genes with lines of evidence from the published literature linking them to suicide risk.

<u>Genes in SGS regions with prior suicide evicdence</u>. *RGS18* and *BRINP3*, have regulatory sequence within the region on chromosome 1q31.1–31.2. *RGS18* is significantly associated with serious suicide attempts from a study of major depression,[14] and is also significantly associated with neuroticism and a mouse model of emotionality.[15] *BRINP3* (*FAM5C*) is associated with suicidal ideation,[16] and also heart disease and vascular inflammation.[17] Of the 10 cases driving this shared region (Table S4), the seven with available diagnostic data had depression (5/7), anxiety (3/7), and alcohol dependence (3/7). Two cases had multiple psychiatric diagnoses, and four cases had previous documented attempts. The three oldest cases had cardiovascular diagnoses and/or obesity.

STAT1 has coding sequence in the chromosome 2q32.2-q32.3 shared region. This gene is a transcription activator regulated by interferon-alpha (IFN-alpha); treatment with IFN-alpha is known to result in serious neuropsychiatric complications.[18] *STAT1* gene expression is significantly increased in individuals with severe depression, and in postmortem brain tissue of individuals who died by suicide.[19] Of the six cases with phenotype data sharing this region, four had depression, three had anxiety, two had personality disorders, three had evidence of obesity, and two had asthma.

The *RHEB* and *CDK5* genes have coding sequence in the chromosome 7q36.1 shared region. *RHEB* is a involved in rapamycin (mTOR) signaling and has been implicated in a study of suicidal thoughts following antidepressant treatment.[20] In addition, *RHEB* is part of the signaling pathway involved in the action of the antidepressant ketamine.[21]

bioRxiv preprint doi: https://doi.org/10.1101/195644; this version posted September 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

The *CDK5* gene is involved in neuronal migration and synaptic plasticity; this gene has shown increased expression in the prefrontal cortex of individuals who died by suicide.[22] The CDK5 pathway has also been implicated in behavioral changes following chronic stressors in a mouse model,[23] and in stress-induced neuronal cell death.[24] Only two of the seven cases giving evidence for the 7q36.1 region had available diagnostic data (Table S4). Both had evidence of previous attempts, and one had depression, anxiety, and alcohol dependence. The other was morbidly obese with associated cardiovascular diagnoses. Missing diagnostic data may reflect a failure to seek treatment or insufficient time for disorders to come to medical attention in the other five young cases (ages 23-33). We are initiating studies of polygenic risk scores in our cases in the hopes of clarifying underlying psychiatric and/or medical risks in cases with no diagnostic data, or when age of death occurred prior to onset.

The *CTNNA3* gene has coding sequence in the chromosome 10q21.3 shared region. A SNP in this gene was one of the most significant findings for suicide attempt in the large genome-wide study of suicidal behavior.[16] This gene has also been implicated in both heart disease and asthma,[25] again suggesting a role of inflammation risk in suicide. Nine of the 12 cases that support this region (Table S4) had diagnostic data. The cases were again young at death, with only two cases over 40. Depression and/or anxiety were common (6/9 cases), as was evidence of previous attempt/ideation (7/9 cases). Five cases showed diagnoses that suggest poor reaction to stress and/or conduct problems. Medical diagnoses included cardiovascular diagnoses, asthma, obesity, and seizures.

The *HTR2A* serotonin receptor gene has regulatory sequence in the shared region on chromosome 13q14.2. Serotonin is a neurotransmitter that has been implicated in both suicide and several psychiatric disorders.[26-27] Suicide risk may be tied to the association between *HTR2A* and response to antidepressant treatment.[28] Three cases from family 27251 (39177, 84752, and 84802) also shared the chromosome 10q21.3 region. These three young male cases had antisocial personality, ADD and anxiety, and PTSD. In addition to these three cases, two other cases also exhibited diagnoses associated with poor stress reaction. Overall, cases for this region were young, with only one over the age of 34, and depression was common (6/8). Medical diagnoses were uncommon, perhaps unsurprising given the young age of the cases.

<u>Chromosome 5q23.3-q31.1</u>. The genome-wide significant region on chromosome 5q23.3–q.31.1 is of interest because it was shared across 21 cases in three families and was our most significant overlapping region ($p < 1E-16$). The genes of possible

bioRxiv preprint doi: https://doi.org/10.1101/195644; this version posted September 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

relevance in this region are associated with both psychopathology and inflammation, a co-occurrence that has been recently recognized.[29] The genes include *HINT1*, which has been associated with social isolation[30] and schizophrenia.[31] The *RAPGEF6* gene has also been associated with schizophrenia,[32] and a knockout mouse has shown changes in amygdala-controlled anxiety responses.[33] The *ACSL6* gene has been linked to schizophrenia[32] and an animal model of depression.[34] Interleukin 3 (*IL3*) is a cytokine with neurotrophic activity; it has been associated with schizophrenia,[35] but also several inflammatory conditions, including asthma and Crohn's Disease.[36-37] The S*LC22A4* gene is involved in the elimination of drugs and environmental toxins. Like *IL3*, it has been implicated in Crohn's disease,[38] and it also affects fibrinogen, a circulating glycoprotein involved in inflammation and autoimmune disease.[39] The *CSF2* gene produces a protein product that controls the function of granulocytes and macrophages. *CSF2RB* which regulates both *IL3* and *CSF2,* was associated with schizophrenia.[35,40] *CSF2* has also been associated with numerous autoimmune disorders and inflammation.[41-42] The *IRF1* gene activates other interferons and is associated with inflammatory conditions[43] and with fibrinogen.[44] Fifteen of the 21 cases supporting this region had diagnostic data. Depression was common (11/15); 6/15 cases had anxiety diagnoses; 5/15 had personality disorders; 2/15 had psychosis or schizophrenia. Five cases had obesity, and one had undergone gastric bypass. Cardiovascular diagnoses, chronic bronchitis, and/or asthma were in multiple cases perhaps reflective of the inflammation risk suggested by the genes in this region.

Conclusions. Our work with these high-risk families has produced evidence for regions with a significant excess of genes previously implicated in suicide risk, and has provided us with a rich resource of additional potential candidate genes. We are currently generating molecular sequence data on cases supporting significant regions to allow for more detailed analyses of potential risk variants, beginning with the genes described in this report.

Materials & Methods

Sample. This project is possible because of a 20-year collaboration with the Utah State Office of the Medical Examiner (OME). Since 1997, we have collected de-identified DNA samples from suicides with Institutional Review Board (IRB) permissions. The collection now numbers 4,585 (3,632 males and 953 females). DNA was extracted from blood using the highly reliable Qiagen Autopure LS automated DNA extractor (www.qiagen.com). Identifying information from cases with DNA was linked to data within the UPDB's secure computer servers; identifying data was then stripped before

bioRxiv preprint doi: https://doi.org/10.1101/195644; this version posted September 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

providing data to the research team. As all subjects are referenced by anonymous IDs, no contact is possible with living family members.

In addition to basic demographic and cause of death information, we have access to diagnostic data for psychiatric conditions and selected medical conditions known to be associated with suicide. These diagnostic codes are housed in the electronic medical record from two of the largest clinical health providers in the state. Conditions were defined by groups of diagnostic codes aggregated according to the International Classification of Diseases (ICD) system (www.icd9data.com). To date we have use these data, when available, to determine *post hoc* if subsets of cases who share specific genomic regions may also share notable demographic characteristics or psychiatric conditions. The diagnostic data provides the groundwork for future more detailed studies of potential associations with specific phenotypes.

Determination of familial risk. Genealogical data in the UPDB was used to construct high risk family trees. In addition to suicide cases with DNA, we have knowledge of a total of 14,288 cases from death certificates dating to 1904. All 14,288 cases were used to estimate familial risk of suicide. To determine the extended families at highest risk, we used the Familial Standardized Incidence Ratio (FSIR) statistic,[8] calculated by comparing the incidence of suicide in each extended family to its expected incidence determined by the statewide expected uniform distribution for suicide stratified by sex and age. The FSIR statistic also weights the contribution of each relative to the familial risk by assessing the probability that the related suicides share alleles through a common ancestor. We identified 241 families at high risk according to the FSIR statistic ($p<0.05$) and with at least three suicide cases with DNA. We selected 43 of these families for analysis (see Table 1) based on lowest p-value for the FSIR statistic, family size, and number of cases with DNA.

Molecular data. The SGS analyses used variants from the Illumina Infinium PsychArray platform (https://www.illumina.com/products/by-type/microarray-kits/infinium-psycharray.html) from the 215 suicide cases in the 43 selected families. We first oriented single nucleotide polymorphisms (SNPs) to the forward strand by comparing the PsychArray SNP calls within the 1000 Genomes Project to 1000 Genomes whole genome sequencing data. We deleted SNPs where forward strand status remained ambiguous, and SNPs where all suicide cases were homozygous for the reference allele. Using PLINK,[9] we also removed 17,058 variants with >5% missing calls and 176 variants that failed Hardy-Weinberg equilibrium ($p<0.001$), leaving a total of 237,415 variants for analysis. Additionally, one case from

bioRxiv preprint doi: https://doi.org/10.1101/195644; this version posted September 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

family 553615 was removed due to poor genotyping quality.

Analysis. We used a new analytical method, Shared Genomic Segments (SGS)[7] that is ideal for our extended family data. This method identifies genomic segments shared between distantly related affected cases using a genome-wide SNP map. If an observed shared segment of the genome is significantly longer than expected by chance, then inherited sharing is implied; theoretically, chance inherited sharing in distant relatives is extremely improbable. Chance occurrence for the shared segments is assessed empirically using gene-drop simulations to create a null distribution of expected sharing within each family; we assign haplotypes to family founders according to a publicly available linkage disequilibrium map from 1000 Genomes European data, followed by simulated segregation of these through the family structure. We systematically iterate over subsets of the full case set in each extended family to account for within-family heterogeneity. To perform significance testing, we derive family-specific significance thresholds using distributions fitted to 100,000-300,000 sets of the simulated data. The resulting thresholds account for multiple testing of regions across the genome.

Power of SGS was previously investigated for a range of genetic models using simulated high-risk families with genotyping only in affected cases.[10] Results indicated that if families had at least 15 meioses between cases, 3-10 families were sufficient to gain excellent power (>80%) to see at least one true positive within any given family. For all scenarios considered, genome-wide association studies would have had negligible power. Given these results, we selected only extended families with at least 15 meioses between cases.

To prioritize results, genes within significant regions were searched for previous associations with suicide or suicidal behavior, or for involvement in psychiatric disorders. We additionally conducted a comprehensive search of the literature for significant genetic associations with suicide or suicidal behavior in order to estimate an expected number genes with these specific phenotypic associations in our identified regions. While genes in all of the statistically significant regions resulting from our work provide a new resource of candidate genes, we have chosen to prioritize the discussion of our results on regions showing overlap across many cases in more than one family. Genetic variants in regions specific to a single family may also reveal important information about potential risk mechanisms; these will be the focus of additional future analyses.

bioRxiv preprint doi: https://doi.org/10.1101/195644; this version posted September 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Acknowledgments

HC had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. HC and LJ are partially supported by a research contract from Janssen Research LLC. No other authors have conflicts of interest relevant to the content of this manuscript, including no financial interest, relationships or affiliations. This work was supported by R01MH099134, the Clark Tanner Foundation, a donation from the Sharon Kae Lehr Endowed Research Fund in Memory of James Raymond Crump, and GCRC M01-RR025764 from the National Center for Research Resources. Partial support for all data sets within the UPDB was provided by the University of Utah Huntsman Cancer Institute. We especially thank University of Utah and OME staff whose hours of work made this study possible.

bioRxiv preprint doi: https://doi.org/10.1101/195644; this version posted September 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

References

1. Centers for Disease Control & Prevention (CDC) Data & Statistics Fatal Injury Report for 2015. 2015: https://webappa.cdc.gov/sasweb/ncipc/mortrate.html.

2. Pederson NL, Fiske A: Genetic influences on suicide and nonfatal suicidal behavior: Twin study findings. European Psychiat. 2010; 25:264-7.

3. McGuffin P, Marusic A, Farmer A: What can psychiatric genetics offer suicidology? Crisis. 2001; 22:61–65.

4. Mirkovic B, Laurent C, Podlipski MA, et al.: Genetic association studies of suicidal behavior: A review of the past 10 years, progress, limitations, and future directions. Front Psychiatry. 2016; 7:158.

5. Younes N, Melchior M, Turbelin C, et al.: Attempted and completed suicide in primary care: not what we expected? J Affect Disord. 2015; 170:150-4.

6. Tabor HK, Risch NJ, Myers RM: Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat Rev Genet. 2002; 5:391-7.

7. Waller RG, Darlington TM, Wei X, et al.: Novel pedigree analysis implicates DNA repair and chromatin remodeling in multiple myeloma risk. BioRxiv. 2017; https://doi.org/10.1101/137000.

8. Kerber RA: Method for calculating risk associated with family history of a disease. Genet Epidemiol. 1995; 12:291–301.

9. Chang CC, Chow CC, Tellier LCAM, et al.: Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015; 4:7.

10. Knight S, Abo RP, Abel HJ, et al.: Shared genomic segment analysis; the power to find rare disease variants. Ann Hum Genet. 2012; 76:500-8.

11. Fisher RA: Statistical Methods for Research Workers. 1925, Edinburgh: Oliver and Boyd, ISBN 0-05-002170-2.

12. Yi G, Sing-Hoi S, Thon M: Identifying clusters in functionally related genes in genomes. Bioinformatics. 2007; 23:1053-60.

13. Ezkurdia I, Juan D, Rodriguez JM, et al.: Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. Hum Molec Genet. 2014; 23:5866-78.

14. Schosser A, Butler AW, Ising M, et al.: Genomewide association scan of suicidal thoughts and behavior in major depression. PLoS One. 2011; 6:e20690.

15. Fullerton JM, Willis-Owen SA, Yalcin B, et al.: Human-mouse quantitative trait locus concordance and the dissection of a human neuroticism locus. Biol Psychiatry, 2008; 63:874-83.

16. Galfalvy H, Haghighi F, Hodgkinson C, et al.: A genome-wide association study of suicidal behavior. Am J Med Genet B Neuropsychiatr Genet. 2015; 168:557-63.

17. Chen X, Chen X, Xu Y, et al.: Association of six CpG-SNPs in the inflammation-related genes with coronary heart disease. Hum Genomics. 2016; 10 Suppl 2:21.

18. Wang J, Campbell IL, Zhang H: Systemic interferon-alpha regulates interferon-stimulated genes in the central nervous system. Mol Psychiatry. 2008; 13:293-301.

19. Hoyo-Becerra C, Huebener A, Trippler M, et al.: Concomitant interferon alpha stimulation and TLR3 activation induces neuronal expression of depression-related genes that are elevated in the brain of suicidal persons. PLoS One. 2013; 8:e83149.

20. Menke A, Domschke K, Czamara D, et al.: Genome-wide association study of antidepressant treatment-emergent suicidal ideation. Neuropsychopharmacology. 2012; 37:797-807.

bioRxiv preprint doi: https://doi.org/10.1101/195644; this version posted September 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

21. Harraz MM, Snyder SH: Antidepressant actions of ketamine mediated by the mechanistic target of rapamycin, nitric oxide and Rheb. Neurotherapeutics. 2017; doi:10.1007/s13311-017-0540-0.

22. Choi K, Le T, Xing G, et al.: Analysis of kinase gene expression in the frontal cortex of suicide victims: implications of fear and stress. Front Behav Neurosci. 2011; 5:46.

23. Bavley CC, Fischer DK, Rizzo BK, et al.: Cav1.2 channels mediate persistent chronic stress-induced behavioral deficits that are associated with prefrontal cortex activation of the p25/Cdk5-glucocorticoid receptor pathway. Neurobiol Stress. 2017; 7:27-37.

24. Song H, Kim W, Choi JH, et al.: Stress-induced nuclear translocation of CDK5 suppresses neuronal death by downregulating ERK activation via VRK3 phosphorylation. Sci Rep. 2016; 6:28634.

25. Folmsbee SS, Gottardi CJ: Cardiomyocytes of the Heart and Pulmonary Veins: Novel Contributors to Asthma? Am J Respir Cell Mol Biol. 2017; doi:10.1165/rcmb.2016-0261TR.

26. Bani-Fatemi A, Howe AS, Matmari M, et al.: Interaction between methylation and CpG single-nucleotide polymorphisms in the HTR2A gene: association analysis with suicide attempt in schizophrenia. Neuropsychobiology. 2016; 73:10-15.

27.Hofer P, Schosser A, Calati R, et al.: The impact of serotonin receptor A1 and A2 gene polymorphisms and interactions on suicide attempt and suicide risk in depressed patients with insufficient response to treatment—a European multicentre study. Int Clin Psychopharmacol. 2016; 31:1-7.

28.McMahon FJ, Buervenich S, Charney D, et al.: Variation in the Gene Encoding the Serotonin 2A Receptor Is Associated with Outcome of Antidepressant Treatment. Am J Hum Genet. 2006; 78:804–814.

29. Brundin L, Bryleva EY, Rajamani KT: Role of inflammation in suicide: from mechanisms to treatment. Neuropsychopharmacology. 2017; 42:271-283.

30. Li B-J, Liu P, Chu Z, et al.: Social isolation induces schizophrenia-like behavior potentially associated with HINT1, NMDA receptor 1, and dopamine receptor 2. Neuroreport. 2017; 28:462-469.

31. Varadarajulu J, Schmitt A, Falkai P, et al.: Differential expression of HINT1 in schizophrenia brain tissue. Eur Arch Psychiatry Clin Neurosci. 2012; 262:167-172.

32. Chen X, Wang X, Hossain S, et al.: Haplotypes spanning SPEC2, PDZ-GEF2 and ACSL6 genes are associated with schizophrenia. Hum Mol Genet. 2006; 15:3329-42.

33. Levy RJ, Kvajo M, Li Y, et al.: Deletion of Rapgef6, a candidate schizophrenia susceptibility gene, disrupts amygdala function in mice. Transl Psychiatry. 2015; 5:e577

34. Setnik B, Nobrega JN: Long-chain acyl-CoenzymeA synthetase-2 mRNA: increased cerebral cortex expression in an animal model of depression. Prog Neuropsychopharmacol Biol Psychiatry. 2004; 28:577-82.

35.Chen Q, Wang X, O'Neill FA, et al.: Association study of CSF2RB with schizophrenia in Irish family and case-control samples. Mol Psychiatry. 2008; 13:930-8.

36. Park BL, Kim LH, Choi YH, et al.: Interleukin 3 (IL3) polymorphisms associated with decreased risk of asthma and atopy. J Hum Genet. 2004; 49:517-27.

37. Kenny EE, Pe'er I, Karban A, et al.: A genome-wide scan of Ashkenazi Jewish Crohn's disease suggests novel susceptibility loci. PLoS Genet. 2012; 8:e1002559.

38. De Ridder L, Weersma RK, Dijkstra G, et al.: Genetic susceptibility has a more important role in pediatric-onset Crohn's disease than in adult-onset Crohn's disease. Inflamm Bowel Dis. 2007; 13:1083-92.

39. Davalos D, Akassoglou K: Fibrinogen as a key regulator of inflammation in disease. Semin Immunopathol. 2012; 34:43-62.

bioRxiv preprint doi: https://doi.org/10.1101/195644; this version posted September 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

40. Sun S, Wang F, Wei J, et al.: Association between interleukin-3 receptor alpha polymorphism and schizophrenia in the Chinese population. Neurosci Lett. 2008; 440:35-7.

41. Ghosh D, Curtis AD 2nd, Wilkinson DS, et al.: Depletion of CD4+ CD25+ regulatory T cells confers susceptibility to experimental autoimmune encephalomyelitis (EAE) in GM-CSF-deficient Csf2-/- mice. J Leukoc Biol. 2016; 100:747-760.

42. Okada Y, Terao C, Ikari K, et al.: Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population, Nat Genet. 2012; 44:511-6.

43. Chiaroni-Clarke RC, Munro JE, Chavez RA, et al.: Independent confirmation of juvenile idiopathic arthritis genetic risk loci previously identified by immunochip array analysis. Pediatr Rheumatol Online J. 2014; 12:53.

44. de Vries PS, Chasman DI, Sabater-Lleal M, et al.: A meta-analysis of 120 246 individuals identifies 18 new loci for fibrinogen concentration. Hum Mol Genet. 2016; 25:358-70.

bioRxiv preprint doi: https://doi.org/10.1101/195644; this version posted September 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure Legends

Figure 1. Example of cases descending from more than one founder.

Supporting Information Legends

Table S1. Genes with coding or regulatory sequence within genomic regions with genome-wide significance evidence for sharing. Genes with more likely associations with suicide risk are highlighted.

Table S2. All regions meeting genome-wide suggestive significance. Regions overlapping in more than one family are noted.

Table S3. All genes with coding or regulatory sequence in regions with at least suggestive genome-wide evidence that overlapped in more than one family. Genes that have been followed up in our initial studies are highlighted.

Table S4. Phenotypic characteristics of cases in families sharing overlapping genomic regions on chromosomes 1q31.2-q31.1 (blue shading), 2q32.2-q32.3 (pink shading), 7q36.1 (red shading), 10q21.3 (yellow shading), 13q14.2 (gray shading), and 5q23.3-q31.1 (green shading).

Figure S1. Extended family structures linking cases used for SGS analyses in the 10 families with genome-wide significant results analyses (each analyzed case is labeled with a numeric ID). Gender is disguised and sibship order is randomized in order to protect the privacy of family members.

Figure S2. Extended family structures in the families with SGS evidence of shared regions overlapping across more than one family discussed in the main text. Only suicide cases in the line of descent to analyzed cases are shown. Each analyzed case is labeled with a numeric ID. Gender is disguised and sibship order is randomized in order to protect the privacy of family members. NOTE: Suicide cases are not evident in upper generations because suicide status from death certificates is only available back to 1904.

bioRxiv preprint doi: https://doi.org/10.1101/195644; this version posted September 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table 1. Characteristics of 43 extended families at high risk for suicide.

Family	FSIR statistic	FSIR P-value	Total N Observed Cases	Total N Expected Cases	N Cases Analyzed	N Meioses Between Analyzed Cases	SGS Threshold: Significant	SGS Threshold: Suggestive	Produced at Least One Significant Region	Produced at Least One Suggestive Region
709	2.69	<0.0001	27	10.04	8	35	6.36E-08	6.41E-07	**YES**	**YES**
1881	3.12	<0.0001	21	9.31	4	17	2.03E-06	3.00E-05	no	no
2082	3.99	<0.0001	19	4.77	3	15	3.93E-06	9.78E-05	no	no
7785	2.78	<0.0001	28	10.08	6	26	1.45E-07	1.82E-06	no	**YES**
8556	2.85	<0.0001	41	14.39	7	37	8.99E-08	1.01E-06	no	**YES**
11593	2.31	0.0002	25	10.82	6	29	1.28E-07	1.70E-06	no	**YES**
12291	2.43	0.0001	23	9.48	6	29	1.59E-07	1.96E-06	no	**YES**
27251	2.43	0.0001	36	16.93	12	52	1.16E-09	1.41E-08	no	**YES**
36667	3.61	0.0039	7	1.61	4	17	7.99E-07	1.56E-05	no	**YES**
37661	2.06	0.0031	19	9.21	5	20	1.79E-06	1.91E-05	no	**YES**
40780	3.61	0.0038	7	1.94	4	19	1.59E-06	2.57E-05	no	**YES**
41469	2.29	0.0013	18	7.86	4	20	4.24E-06	5.86E-05	no	**YES**
43035	2.46	<0.0001	36	14.63	7	31	1.09E-07	1.15E-06	no	**YES**
43580	2.45	0.0004	19	7.75	6	29	2.93E-07	3.08E-06	no	**YES**
46547	2.48	<0.0001	29	11.71	6	28	2.75E-07	2.92E-06	no	**YES**
60205	2.26	0.0007	21	9.31	5	26	5.07E-07	6.52E-06	**YES**	no
66494	2.45	0.0001	24	9.80	7	34	5.45E-08	6.47E-07	**YES**	**YES**
68939	2.21	0.0007	22	9.97	8	41	3.66E-08	3.95E-07	no	**YES**
91500	3.19	0.0003	13	4.07	4	16	1.24E-06	2.18E-05	no	no
129334	2.75	<0.0001	25	9.09	7	36	6.92E-08	8.32E-07	no	**YES**
148039	3.65	0.0002	12	3.29	4	15	5.14E-06	5.73E-05	no	no
176860	2.71	<0.0001	43	15.88	9	39	3.28E-08	3.08E-07	no	**YES**
185855	2.96	0.0003	15	5.06	6	25	2.46E-07	2.24E-06	no	**YES**
209487	2.82	<0.0001	26	9.23	7	32	1.19E-07	1.25E-06	**YES**	**YES**
233769	2.50	<0.0001	33	13.19	11	50	1.58E-08	1.40E-07	**YES**	**YES**
265545	4.05	0.001	8	1.98	5	17	2.74E-06	3.82E-05	no	no
540295	2.59	<0.0001	43	16.59	5	31	1.73E-07	3.10E-06	no	**YES**
540775	2.46	0.003	13	5.28	7	34	5.88E-08	7.45E-07	**YES**	**YES**
544252	3.21	0.0025	9	2.80	4	19	7.48E-07	1.43E-05	no	**YES**
553615	2.04	<0.0001	81	39.70	13	69	2.69E-10	3.74E-09	**YES**	**YES**
554151	2.38	0.003	14	5.89	7	32	9.98E-08	1.05E-06	no	no
587072	2.39	0.003	14	5.86	5	25	1.39E-06	1.55E-05	no	**YES**
590241	2.47	0.0004	19	7.70	8	38	2.86E-08	3.16E-07	no	**YES**
595955	2.48	<0.0001	39	15.73	7	39	5.35E-08	6.62E-07	no	**YES**
601627	2.86	<0.0001	69	24.14	12	70	7.06E-10	9.14E-09	**YES**	**YES**
603481	2.64	0.0003	18	6.83	9	37	5.09E-08	5.35E-07	**YES**	**YES**
622459	2.54	0.0001	22	8.68	4	18	1.75E-06	2.60E-05	no	**YES**
755858	2.28	0.001	19	8.33	5	24	3.77E-07	5.48E-06	no	no
756794	2.70	0.0002	18	6.68	5	24	3.94E-07	5.67E-06	no	**YES**
791533	3.67	0.001	9	2.45	3	17	3.09E-06	8.38E-05	no	**YES**
807334	2.50	0.0001	24	9.58	6	30	3.25E-07	3.57E-06	**YES**	**YES**

bioRxiv preprint doi: https://doi.org/10.1101/195644; this version posted September 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

923763	4.41	<0.0001	14	3.18	3	15	3.33E-06	8.86E-05	no	no
957634	3.70	0.0001	12	3.24	3	15	3.22E-06	8.40E-05	no	**YES**

bioRxiv preprint doi: https://doi.org/10.1101/195644; this version posted September 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table 2. Basic phenotypic characteristics of analyzed cases within the 43 extended families at high risk for suicide.

Family	Analyzed Cases	Male (%)	Average Age	Method of Death
709	8	7 (87.50)	21.88	5 gun; 3 other violent[1]
1881	4	4 (100.00)	43.75	2 gun; 2 other violent
2082	3	3 (100.00)	19.33	1 gun; 2 other violent
7785	6	4 (66.67)	44.17	4 gun; 2 other violent
8556	7	6 (85.71)	37.71	6 gun; 1 other violent
11593	6	3 (50.00)	37.00	3 gun; 2 other violent; 1 overdose
12291	6	5 (83.33)	37.67	5 gun; 1 other violent
27251	12	9 (75.00)	30.58	5 gun; 4 other violent; 3 overdose
36667	4	4 (100.00)	35.50	2 gun; 2 other violent
37661	5	3 (60.00)	29.00	1 other violent; 4 overdose
40780	4	3 (75.00)	33.00	2 gun; 2 overdose
41469	4	3 (75.00)	26.00	1 gun; 2 other violent; 1 overdose
43035	7	7 (100.00)	35.86	4 gun; 2 other violent; 1 overdose
43580	6	5 (83.33)	46.83	3 gun; 1 other violent; 2 overdose
46547	6	5 (83.33)	42.00	6 gun
60205	5	4 (80.00)	26.80	4 gun; 1 overdose
66494	7	7 (100.00)	31.00	5 gun; 2 other violent
68939	8	5 (62.50)	23.88	3 gun; 4 other violent; 1 overdose
91500	4	4 (100.00)	49.00	2 gun; 2 other violent
129334	7	5 (71.43)	29.14	4 gun; 3 other violent
148039	4	4 (100.00)	32.25	2 gun; 2 other violent
176860	9	7 (77.78)	33.56	6 gun; 2 other violent; 1 overdose
185855	6	5 (83.33)	39.43	4 gun; 2 overdose
209487	7	6 (85.71)	25.53	3 gun; 4 other violent
233769	11	8 (72.73)	26.64	2 gun; 7 other violent; 1 overdose
265545	5	3 (60.00)	38.40	1 gun; 4 other violent
540295	5	5 (100.00)	25.60	1 gun; 2 other violent; 2 overdose
540775	7	7 (100.00)	39.29	3 gun; 4 other violent
544252	4	3 (75.00)	27.50	2 gun; 2 other violent
553615	13	9 (69.23)	39.82	7 gun; 5 other violent; 1 overdose
554151	7	6 (85.71)	34.29	1 gun; 5 other violent; 1 overdose
587072	5	4 (80.00)	31.60	3 gun; 2 other violent
590241	8	6 (75.00)	26.00	6 gun; 1 other violent; 1 overdose
595955	7	4 (57.14)	27.86	3 gun; 2 other violent; 2 overdose
601627	12	9 (75.00)	27.92	9 gun; 2 other violent; 1 overdose
603481	9	7 (77.78)	39.33	7 gun; 1 other violent; 1 overdose
622459	4	3 (75.00)	40.00	1 gun; 3 other violent
755858	5	4 (80.00)	30.80	5 gun
756794	5	4 (80.00)	27.60	1 gun; 4 other violent
791533	3	1 (33.33)	33.67	3 gun
807334	6	6 (100.00)	31.00	3 gun; 3 other violent
923763	3	3 (100.00)	24.67	1 gun; 2 other violent
957634	3	3 (100.00)	44.00	2 gun; 1 other violent

[1] The "other violent" category included hanging, other suffocation (such as plastic bag), jumping from a high place or in front of a vehicle, drowning, cutting, or intentional motor vehicle accident. Carbon monoxide poisoning was included in the overdose category.

bioRxiv preprint doi: https://doi.org/10.1101/195644; this version posted September 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table 3. Genome-wide significant SGS regions.

Family	Chromosome	Start	End	N Cases	Significance Threshold	P-Value
601627	2p16.3	50902522	51820543	6	7.06E-10	1.94E-10
601627	2q36.3 – q37.1	230899765	231454354	6	7.06E-10	2.39E-10
553615	3p14.1	64735531	65289530	8	2.69E-10	8.87E-11
603481	4q26	117379825	118257841	7	5.09E-08	3.08E-08
807334	4q28.3	131561136	132902055	5	3.25E-07	2.02E-07
553615	5q23.3 – q31.1	129199151	131819921	9	2.69E-10	2.39E-10
601627	5q33.3 – q34	159633484	160328128	7	7.06E-10	5.47E-10
553615	6q11.1 – q12	62563817	64139997	8	2.69E-10	1.34E-10
60205	6q24.3	148162328	148621930	5	5.07E-07	4.02E-07
601627	7p21.2	14144663	15001308	6	6.97E-10	2.04E-10
233769	10p15.3	2408852	2881331	7	1.58E-08	1.11E-09
209487	11p11.2 – q12.1	47312689	56518769	6	1.19E-07	6.60E-08
540775	11q13.3	69482091	69933696	7	5.88E-08	6.20E-09
709	13q12.3	29886987	30492217	7	6.36E-08	1.86E-08
709	15q21.3 – q22.2	58601804	59646991	6	6.36E-08	2.74E-08
66494	15q22.2	62914165	63686327	6	5.45E-08	5.44E-08

bioRxiv preprint doi: https://doi.org/10.1101/195644; this version posted September 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Table 4. SGS regions with at least genome-wide suggestive evidence in > 1 extended family.

Families	Chromosome	Start	End	Thresholds (suggestive)	P-Values	N Cases	Combined Probability[a]
709, 8556	1p34.2	40433771	40555321	6.41E-07, 1.01E-06	3.20E-07, 3.52E-07	6, 5	3.47E-12
791533, 540775	1q31.1 – q31.2	190694813	191590362	8.38E-05, 7.45E-07	4.83E-04, 4.12E-07	3, 7	4.63E-09
176860, 11593	2q32.2 – q32.3	191029604	192020729	3.08E-07, 1.70E-06	1.79E-07, 7.87E-07	5, 6	4.31E-12
129334, 11593	3q26.33	181074751	181229833	8.32E-07, 1.70E-06	2.81E-07, 9.47E-07	4, 5	7.97E-12
8556, 66494	4q35.1 – q35.2	187072383	187513585	1.01E-06, 6.47E-07	4.45E-07, 1.32E-07	4, 5	1.85E-12
553615, 603481, 176860	5q23.3 – q31.1	129684909	131819921	2.06E-08[b], 5.35E-07, 3.08E-07	1.48E-08[b], 3.51E-07, 2.05E-07	7[b], 7, 7	<1E-16
957634, 595955	7q36.1	150239676	151123529	8.40E-05, 6.62E-07	1.13E-05, 2.05E-07	3, 4	6.44E-11
587072, 595955	8p23.1	9157884	10032894	1.55E-05, 6.62E-07	5.24E-06, 3.20E-07	4, 5	4.71E-11
11593, 8556	10p12.33	17391660	17576227	1.70E-06, 1.01E-06	1.64E-06, 6.66E-07	4, 5	3.11E-11
27251, 233769	10q21.3	67735584	68057063	1.41E-08, 1.40E-07	1.66E-09, 1.34E-07	7, 5	8.22E-15
209487, 66494	12q12	41899312	42298882	1.25E-06, 6.47E-07	3.92E-07, 1.74E-07	5, 5	2.14e-12
27251, 41469	13q14.2	48526833	49283795	1.41E-08, 5.86E-05	1.23E-08, 7.67E-06	5, 4	2.93E-12
27251, 233769	18q11.2	24414687	24494344	1.41E-08, 1.40E-07	6.65E-09, 2.10E-08	8, 6	5.22E-15

[a] Estimated using Fisher's combined probability test (Fisher, 1925).

[b] Two cases were omitted from family 553615 to satisfy the independence requirement for computing a combined p-value. Significance thresholds were re-computed for family 553615 eliminating these cases. Person 112304 is a descendant of both 553615 and 603471, and person 95765 is a descendant of both 553615 and 176860.

bioRxiv preprint doi: https://doi.org/10.1101/195644; this version posted September 28, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.
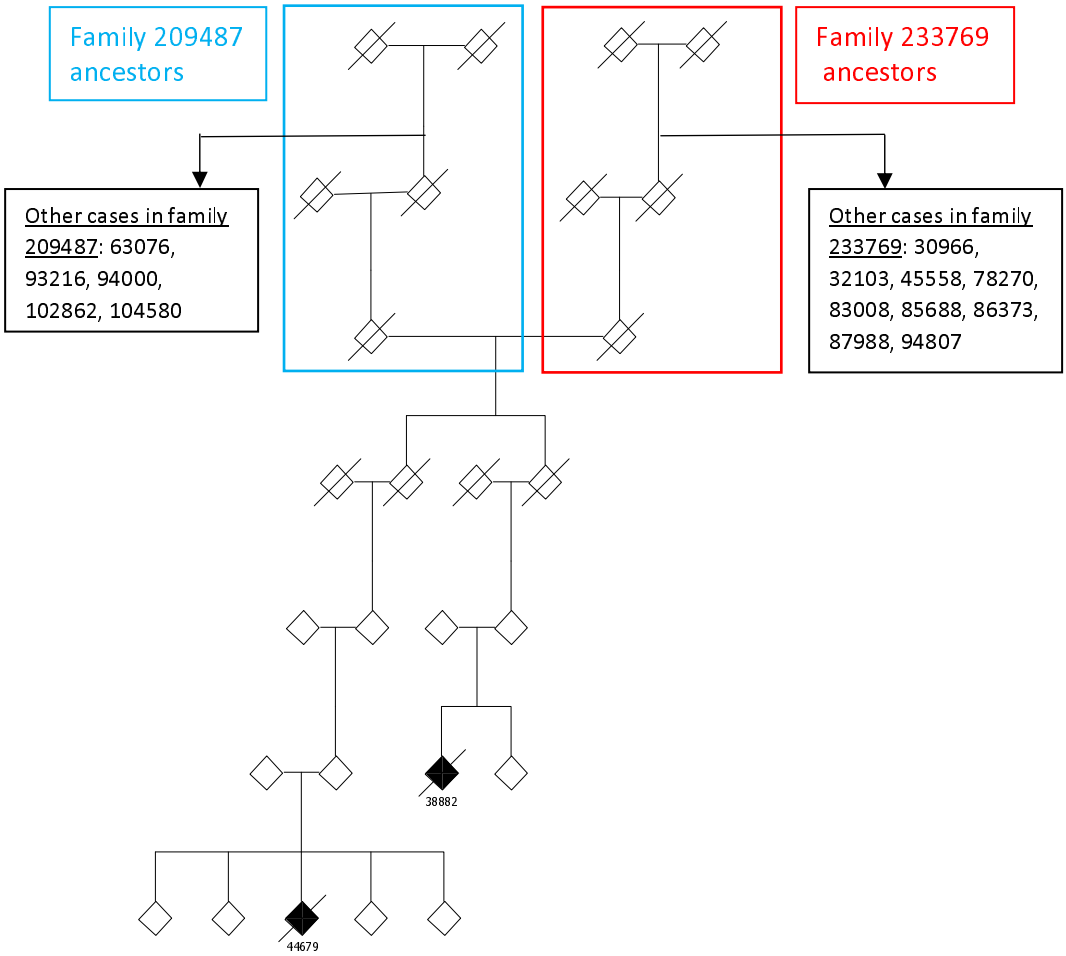
Figure 1.