

SCIENTIFIC DATA

OPEN Data Descriptor: Whole-exome sequencing data of suicide victims who had suffered from major depressive disorder

Received: 7 September 2018

Accepted: 13 December 2018

Published: 5 February 2019

Dóra Tombácz¹, Zoltán Maróti², Tibor Kalmár², Miklós Palkovits³, Michael Snyder⁴ & Zsolt Boldogkői¹

Suicide is one of the leading causes of mortality worldwide; it causes the death of more than one million patients each year. Suicide is a complex, multifactorial phenotype with environmental and genetic factors contributing to the risk of the forthcoming suicide. These factors first generally lead to mental disorders, such as depression, schizophrenia and bipolar disorder, which then become the direct cause of suicide. Here we present a high quality dataset (including processed BAM and VCF files) gained from the high-throughput whole-exome Illumina sequencing of 23 suicide victims – all of whom had suffered from major depressive disorder - and 21 control patients to a depth of at least 40-fold coverage in both cohorts. We identified ~130,000 variants per sample and altogether 442,270 unique variants in the cohort of 44 samples. To our best knowledge, this is the first whole-exome sequencing dataset from suicide victims. We expect that this dataset provides useful information for genomic studies of suicide and depression, and also for the analysis of the Hungarian population.

Design Type(s)	case-control study design • parallel group design • individual genetic characteristics comparison design
Measurement Type(s)	genetic sequence variation analysis
Technology Type(s)	exome sequencing assay
Factor Type(s)	regional part of brain • Cause of Death
Sample Characteristic(s)	Homo sapiens • occipital cortex • cerebellar cortex • primary motor cortex

¹Department of Medical Biology, Faculty of Medicine, University of Szeged, Somogyi B. u. 4., Szeged, H-6720, Hungary. ²Department of Paediatrics, Faculty of Medicine, University of Szeged, Korányi fasor 14-15., Szeged, H-6720, Hungary. ³Neuromorphological and Neuroendocrine Research Laboratory, Department of Anatomy, Histology and Embryology, Semmelweis University, Üllői u. 26., Budapest, H-1085, Hungary. ⁴Department of Genetics, School of Medicine, Stanford University, 300 Pasteur Dr., Stanford, CA, 94305-5120, USA. Correspondence and requests for materials should be addressed to Z.B. (email: boldogkoi.zsolt@med.u-szeged.hu)

Background & Summary

Suicide is the 10th leading cause of mortality in the world^{1,2}. It is a complex, multifactorial behavioural condition in which proximal and distant risk factors, such as mental disorders, genetic and epigenetic factors, social, cultural and economic environment, gender, age, personality, eating disorders, as well as drug or alcohol abuse^{3–7} play important roles. The presence of early child abuse has also been described as a contributing factor to suicidal behaviour⁸. Studies have repeatedly shown that suicide is most common among patients with mood disorders, such as major depressive disorder and bipolar disorder⁹.

Whole-exome sequencing (WES) is a cost-effective and powerful tool for the analysis of complex and rare genetic diseases¹⁰. WES technique allows a base-pair comparison of exomes and consequently the examination of rare genetic variants, which may play a role in suicide. Alterations in copy number variations (CNVs) and short-tandem repeats (STRs), as well as the viral infections can be also analysed by WES.

We performed high-depth, paired-end, short-read WES on 23 suicide victims suffered from major depressive disorder (MDD; this information is available in EGA sample metadata) and 21 control patients who died from other diseases (EGA sample metadata). Harvesting of brain tissues was approved by the local ethics committee¹¹. Informed consents from the next kin have not been obtained from the victims because the local regulations did not require consent for autopsy. All data were handled anonymously. Both of the cohorts were of Hungarian ethnicity. The Hungarian population is known to have had a very high annual suicide rate in the previous century^{12,13}. The DNA samples derived from post-mortem brains were sequenced using the Agilent SureSelectXT Human All Exon V5 + UTRs kit and Illumina HiSeq 2000 sequencing platform resulting in an average of 117 million paired-end reads.

We obtained very high coverage per base position in both cohorts; in the suicide cohort 95.4–97.5% (quartiles) of target regions had higher than 20-fold and 82.8–90.5% of target regions had higher than 40-fold coverage. In the Hungarian control cohort these values were 96.3–97.5% for 20-fold and 81.6–91.6% for 40-fold coverage, respectively. We called on average ~130,000 variants per sample. Altogether we observed 442,270 unique variants in the control and suicide cohorts. Furthermore, we determined a total of 320,976 single nucleotide variants (SNVs) and 47,313 insertions/deletions (indels).

The library preparation kit used in this study, as well as the obtained very high coverage of sequencing reads that we obtained enabled us to capture information not only about the exons but also about the 5'- and 3'-UTRs, the promoters to a certain length, and even about off-target sequences, such as introns, intergenic regions and infecting viruses. Our dataset had 286,754 high coverage regions (exceeding 20-fold coverage) in all samples that can be suitable for downstream analysis of CNVs.

Here we describe the sample collection methods, the library preparation and sequencing method, the currently available data records, and technical validations for our dataset. A schematic overview of this study, including the experimental procedure and the bioinformatics workflow, is also presented (Fig. 1).

To our knowledge, this dataset is the first individual-level, open-access WES data release targeted to suicide. The dataset described here has been thoroughly analysed in our related manuscript⁶, where it has been shown that the TGF- β signalling pathway and three voltage-gated calcium ion channels may play an important role in the pathogenesis of suicidal behaviour. Our analysis proposed that suicide may be an umbrella-disease like disorder.

Here we provide the processed BAM files, as well as the processed Variant Call format (VCF) files for each of the samples containing the analysis ready variants of GATK HaplotypeCaller pipeline that allows for the testing of different variant filtration strategies that we had used in our original manuscript or for hypothesis testing of candidate genes (covered by SureSelect V5 All Exon plus UTR kit) without tedious bioinformatic processing.

Our dataset provided here in this manuscript can be re-used to investigate suicide and MDD. These data can be a valuable resource for investigating genetic variants, genes and signalling pathways to identify novel factors related to these disorders, and may provide novel information for the investigation of the Hungarian population or in general for studies of genetic polymorphisms of human population.

Methods

The methods described here are an expanded version of those described in our recently published, related publication⁶.

Medical history of the suicide victims

This study is based on whole-genome sequence data from occipital, cerebellar and somatomotor cortex regions of the brains of 23 suicide victims (15 males and 8 females) and 21 control patients (14 males and 7 females) who died suddenly from other causes (this information is available in EGA sample metadata). Samples were obtained by autopsy at the Department of Forensic Medicine of the Semmelweis University Medical School. For all of the suicide victims, a psychiatric diagnosis of major depression was on record. The diagnoses had been made by experienced psychiatrists on the basis of Diagnostic and Statistical Manual of Mental Disorders Fourth Edition (DSM-IV). Patients with a history of schizophrenia, epilepsy and other disorders were excluded from the study. The medical, psychiatric, and drug history of the suicide victims was obtained by psychological autopsy, which included interviews with the attending physician and with family members, as well as data obtained from medical records. The participants included in this study had no history of drug or alcohol abuse, and had not used antidepressant

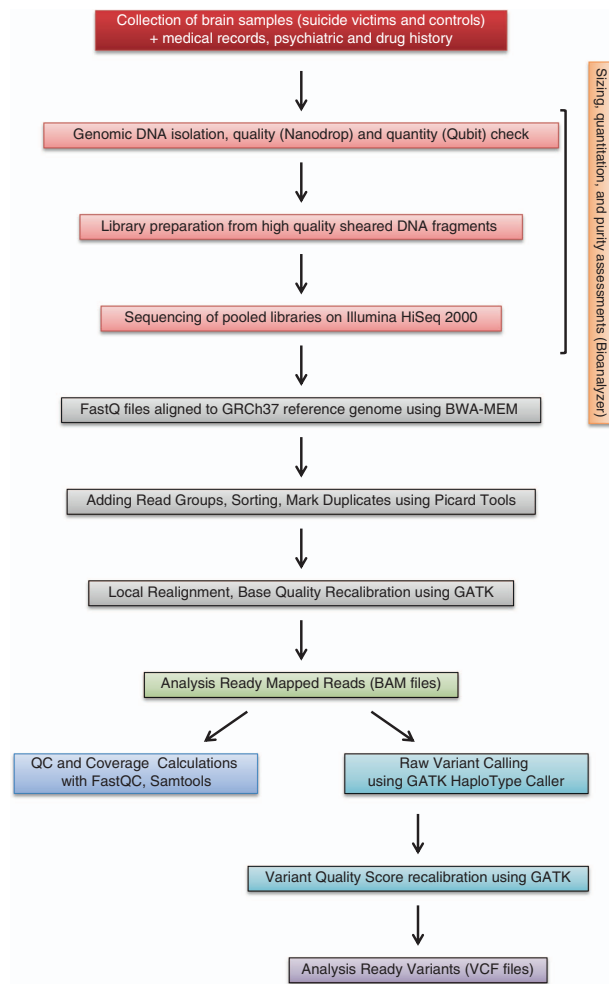


Figure 1. Data flow diagram shows the detailed overview of the study design, wet lab experiments and bioinformatics pipelines.

medication for at least two months prior to death. This information was confirmed by toxicology tests on blood samples. Suicide victims died by hanging ($n = 16$), drug overdose ($n = 6$), or jumping from height ($n = 1$). Causes of death in control subjects are presented in EGA sample metadata. Harvesting of tissues was approved by the local ethics committee.

Tissue collection, dissection, and storage

Brains were obtained 1 to 10 h after death (this information is available in EGA sample metadata). After removal from the skull, the brains were cut in six major pieces (four cortical lobes, basal ganglia – diencephalon, and lower brain stem – cerebellum), rapidly frozen on dry ice, and stored at $-70\text{ }^{\circ}\text{C}$ until dissection (2 days to 2 months). At the time of dissection, the brain samples were sliced into 1 mm-thick coronal sections at a temperature of 0 to $10\text{ }^{\circ}\text{C}$. The cortical areas were cut out of the sections by a fine microdissecting (Graefe’s) knife. In all cases, cortical samples were always taken from the right hemisphere. The samples were stored in airtight containers or plastic tubes at $-70\text{ }^{\circ}\text{C}$ until further use.

DNA purification

Genomic DNA (gDNA) samples were isolated from the occipital, cerebellar or somatomotor cortex regions (Table 1 (available online only)) using the DNeasy Blood and Tissue Kit (Qiagen) according to the manufacturer’s recommendations for the spin-column protocol, using 30 mg starting tissue material from each sample. In short, tissue samples were cut into small pieces, and then lysis buffer (provided by the kit) and proteinase K were added. Lysis reactions were carried out at $56\text{ }^{\circ}\text{C}$ until complete lysis was obtained. DNeasy Mini spin columns (kit’s component) were used for the isolation of gDNA from the lysate. Elution was carried out twice to a final volume of $100\text{ }\mu\text{l}$ per elution.

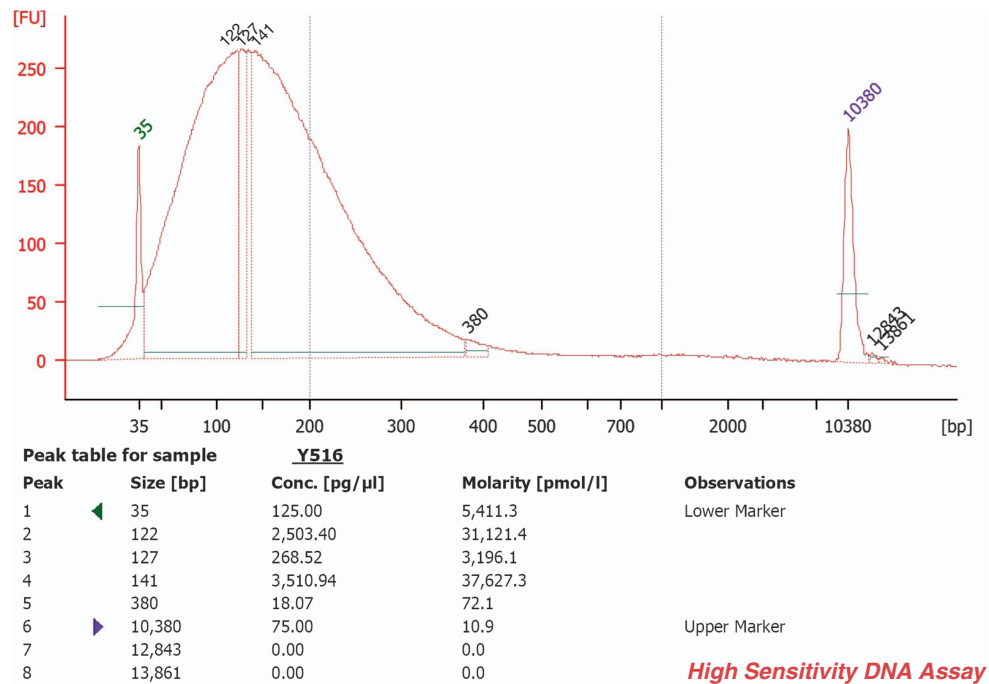


Figure 2. Bioanalyzer (BA) electropherograms of a representative sample (# Y516) at the first step in the library preparation workflow. The figure shows a representative example of the DNA fragment size distribution after sonication. The fragment size peak of this sample is 141 bps.

Quality and quantity check of DNA

The quality of the DNA was checked prior to downstream analysis by NanoDrop (Thermo Fischer Scientific) measurement. High-quality samples with an OD 260/280 ratio ranging from 1.8 to 2.0 were subjected to the next step. Qubit 2.0 fluorometer (Invitrogen) was used for genomic DNA quantification before library preparation (Table 1 (available online only)).

Whole-exome sequencing

Paired-end, indexed libraries for Illumina sequencing were prepared from the DNAs of the 23 suicide victims and the 21 controls using post-mortem brain cortex tissues as a source. Whole-exome sequencing was carried out as previously described¹⁴ with slight modifications. In short, 200 ng from each of the qualified gDNA samples in 100 μl volume were sheared into fragments of approximately 200–450 bps in Covaris microTUBEs (Covaris microTUBE AFA Fiber Pre-Slit Snap-Cap 6 × 16 mm) with Covaris s220 high-performance ultrasonicator. The quality of the DNA was checked prior to downstream analysis, using the Agilent Bioanalyser 2100 (Agilent Technologies) and the High Sensitivity DNA Analysis Kit (Agilent Technologies; Fig. 2). The end-repair, A-tailing (3' adenylation) and ligation of the paired-end adaptors steps were executed with the Agilent SureSelect Human All Exon V5 + UTRs kit (Agilent Technologies) as was described in the following manual: SureSelect^{XT} Target Enrichment System for Illumina Paired-End Sequencing Library Illumina HiSeq and MiSeq Multiplexed Sequencing Platforms Protocol version 1.6 (SureSelect^{XT}). Amplification of the adaptor-ligated libraries was performed using the Herculase II Fusion DNA polymerase (Agilent Technologies) following instructions as set forth in the above mentioned manual. Purification after the enzymatic reactions was carried out by using Agencourt AMPure XP beads (Beckman Coulter). Quality and quantity of the libraries were determined by 2100 Bioanalyser instrument and DNA 1000 Assay (Agilent Technologies; Fig. 3). DNA library samples were hybridized with target-specific Capture Library using the SureSelect^{XT} Reagent Kit, HSQ for the HiSeq platform (Agilent Technologies) based on the SureSelect^{XT} protocol. Incubation of the hybridization mixture was carried out at 65 °C for 24 h. After hybridization, the targeted molecules were captured on streptavidin-coated magnetic beads. The SureSelect enriched libraries were amplified by PCR with 6-bp indexing primers (Table 2 (available online only)) following the SureSelect^{XT} manual's post-capture PCR recommendations using 12 cycles. The amplified libraries were purified using AMPure XP beads. The quality and quantity of the indexed libraries were checked by Agilent Bioanalyser and High Sensitivity DNA Assay (Fig. 4, Table 3). The quantities were also measured by Qubit 2.0 (Table 3). Two or three indexed libraries were pooled (Tables 2 (available online only) and 3) and the pooled samples were sequenced in one lane of a HiSeq 2000 flow cell, generating 101 bp paired-end reads.

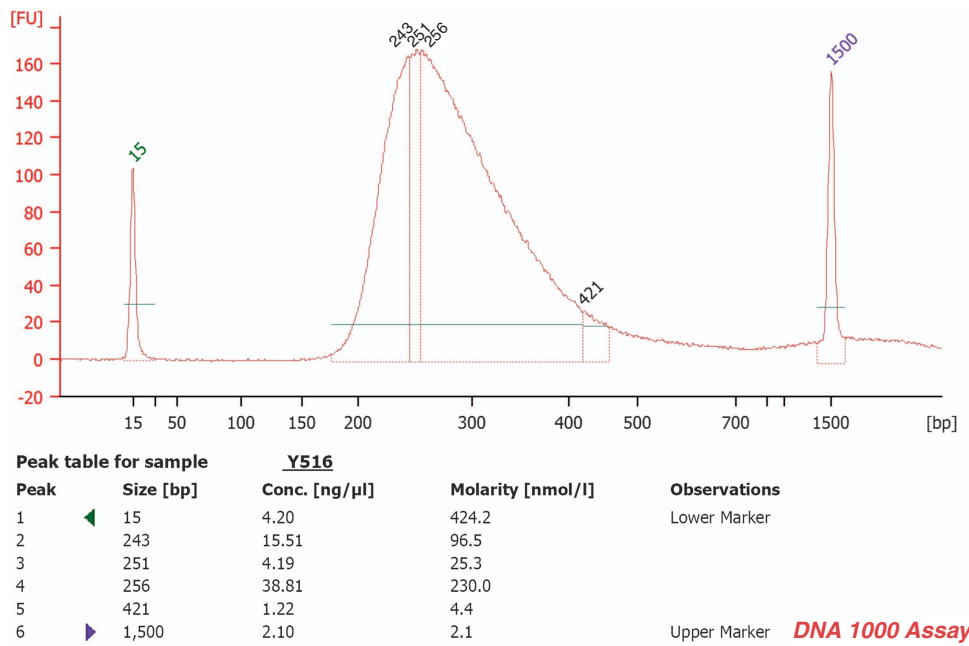


Figure 3. A representative sample peak is shown in this figure, which illustrates the quality of the amplified library. This electropherogram shows a distribution with a DNA fragment size peak of 256 bps for sample Y516.

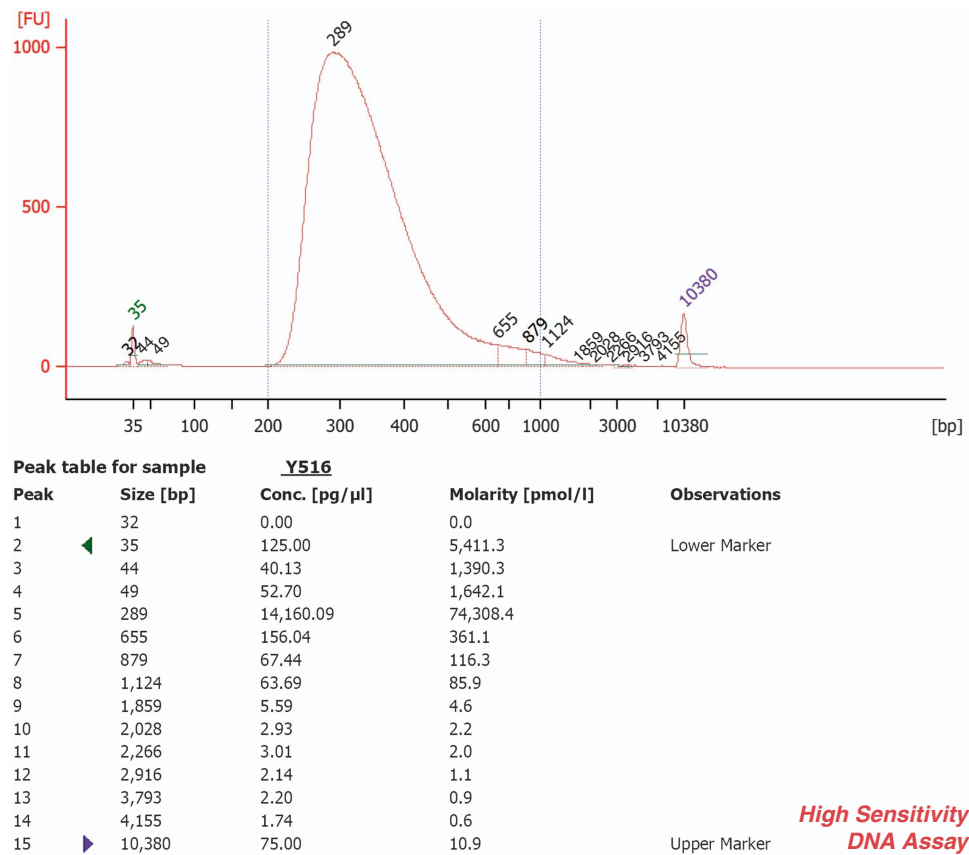


Figure 4. This figure represents the post-capture quality analysis of the PCR-amplified indexed library from the sample Y516. The size peak of this sample was 289 bps, while the average size was 351 bps.

RUN	SAMPLE	C (nM) ^A	Volume (uL) ^B	C (pg/uL) ^C	C (ng/uL) ^D	C (pg/uL) ^E	C (ng/uL) ^F	Library size ^G	Average size ^H
1	Y166	199.2	1.51	41654.8	6.4	68000	8.4	343	356
	Y919	124.8	2.40	26722		26300		358	
	Y645	74.5	4.03	16024		21500		366	
2	Y375	58.781	5.10	12521.73	6.4	8610	4.7	357	350
	Y316	57.027	5.26	11749.08		7200		338	
	Y393	86.632	3.46	18817.57		17100		356	
3	Br857	111.258	2.70	24428.34	6.4	30700	6.7	375	369
	Br333	54.035	5.55	11475.42		13100		375	
	BrA206	131.504	2.28	27907.52		19500		357	
4	Y440	39.245	7.64	8191.29	6.4	13300	7.8	352	363
	Y482	25.05	11.98	5482.86		5100		375	
	Y558	75.605	3.97	16207.71		18000		363	
5	Y331	61.145	4.91	12795.49	6.4	13500	7.2	344	357
	Y724	60.31	4.97	12715.34		20500		362	
	Y292	147.258	2.04	31909.64		23500		365	
6	Y591	103.25	2.91	22871.87	6.6	19500	7.3	376	374
	Y520	60.856	4.93	12793.51		17400		357	
	Y787	110.566	2.71	25104.72		28100		388	
7	Y421	75.488	3.97	16289.39	6.5	14200	5.7	355	353
	Y516	61.276	4.90	13104.89		11000		351	
	Y532	69.348	4.33	14934.17		14300		352	
8	BIB82	139.113	2.16	29501.49	6.0	27400	5.7	350	330
	Y988	68.642	4.37	13380.9		12600		315	
	BI370	43.857	6.84	8660.63		8200		325	
9	BI269	81.736	5.51	15828.78	6.0	11800	4.4	313	329
	Y426	54.168	8.31	11230		8210		345	
10	Y803	43.374	6.92	8970.21	6.4	6420	4.8	341	358
	A438	135.98	2.21	30004		22100		373	
	Y880	115.546	2.60	24433.7		19200		359	
11	YA243	49.12	6.11	10428.1	6.4	9450	5.7	371	361
	Y915	48.499	6.19	10605		8720		363	
	Y967	123.886	2.42	26052		25000		349	
12	Y64	139.449	2.15	29738.07	6.4	30500	7.1	357	357
	Y726	112.212	2.67	23744		28300		357	
	B168	135.267	2.22	29033.01		32100		357	
13	Y872	149.778	2.00	31377.06	6.4	28100	6.5	350	361
	Br708	116.782	2.57	24462.06		33500		355	
	Br969	100.18	2.99	21967.36		17400		377	
14	YB479	127.498	2.35	27282.84	6.3	25500	6.1	356	350
	YA205	61.118	4.91	12889.7		10030		354	
	YB330	95.388	3.15	19830.54		23900		340	
15	Y928	67.261	4.46	14431.43	6.4	14700	6.9	358	356
	Y974	99.194	3.02	21249.16		24900		367	
	Y78	90.024	3.33	18648		19800		343	
16	BI919	113.104	2.65	24609.88	6.4	24100	7.5	377	360
	Y640	161.115	1.86	34615.56		48300		369	
	YA762	119.56	2.51	24368.42		28200		335	
17	YA626	94.57	4.76	20189.16	6.5	22400	6.6	352	360
	YA29	136.8	3.29	29716.84		28000		368	

Table 3. Summary table of the quantitative data of sample pools for multiplexed sequencing. The final concentration was 30 nM in each pool. A: initial concentration of the libraries; B: volume to use in the 30 nM final pool; C: final concentration of the libraries based on Bioanalyzer (BA) measurements; D: final concentration of the pooled libraries determined by BA; E: final concentration of the single libraries measured by Qubit; F: final concentration of the libraries based on Qubit measurement; G: The length of the single libraries were determined by BA; H: The average length of the mixed pools.

Sequence QC

Illumina HiSeq paired-end BCL files were converted to FASTQ files by the standard Illumina protocol (bcl2fastq) in order to remove the adapters, the known Illumina artefacts.

Read Mapping, BAM file post processing

The trimmed, adapter removed FASTQ files were mapped to the GRCh37 reference genome by the Burrows-Wheeler Aligner (version 0.7.9a-r786) with the BWA MEM paired-end mapping algorithm using the following parameters: -M -w 150 -d 80 -c 2000. This mapper is generally recommended for high-quality queries as it is fast, and it allows gapped alignments which are essential for the accurate identification of SNV and insertion/deletions (indels; http://genestack-user-tutorials.readthedocs.io/tutorials/WES_data_analysis/index.html). Duplicates were marked by the Picard tools (version 1.113) MarkDuplicates algorithm. The raw BAM files (Data Citation 1) were realigned and base quality recalibrated by Genome Analysis Toolkit (GATK version: 3.3-0-g37228af).

Sequence QC

FastQC (version v0.11.2) was used to generate quality reports (Data Citation 2) using the aligned and base quality recalibrated BAM files.

Variant calling

The variants were called by the GATK HaplotypeCaller (version: 3.3-0-g37228af) algorithm using the GVCF mode cohort analysis according to the Best Practice Guidelines¹⁵⁻¹⁷. Raw Variants were recalibrated by GATK VariantRecalibrator according to the best practice recommendations using the HapMap, Omni, dbSNP, 1000 G, Mills datasets for training included in the GATK GRCh37 bundle resulting analysis ready variants (Data Citation 2).

Data records

The raw FASTQ files, the GRCh37 aligned BAM files, and the variants in VCF 4.1 format associated with the samples analysed in this study are available on request at EGA (Data Citation 1). We also deposited the QC metric files of Picard MarkDuplicate, and QC reports of FastQC of the analysis ready Bam files at Figshare (Data Citation 2) and in Supplementary Table 1.

Technical validation

Quantitation of the purified DNA samples

The isolated DNA samples were quantified by Qubit (Life Technologies) fluorometer using Qubit dsDNA HS (High Sensitivity) Assay Kit, which is highly selective for double-stranded DNA over RNA and is designed to be accurate for initial sample concentration ranging from 10 pg/μl to 100 ng/μl. DNA samples were diluted to 4 ng/μl with 1X Low TE Buffer. NanoDrop (Thermo Fischer Scientific) measurements were also performed to assess quantity and quality of DNA, 260:280 and 260:230 ratios greater to 1.8 were accepted.

Quality control of the sheared DNA samples

The quality of the sheared DNA samples (200 ng of each) were checked prior to downstream analysis, using the Agilent Bioanalyser 2100 (Agilent Technologies), and High Sensitivity DNA chip and reagent kit. The electropherogram showed a DNA fragment size peak (for each of the samples) at around 150 bps (Fig. 2).

Quality check of the amplified samples

Agilent Bioanalyser 2100 (Agilent Technologies) and DNA 1000 Assay were used for the quality and quantity control of the libraries after PCR. The sample fragments sizes were between 225 and 275 bps (Fig. 3).

Assessment of the quantity and quality of the indexed library DNAs

The quality of the amplified, indexed libraries were determined before multiplexing using the Agilent Bioanalyser 2100 (Agilent Technologies) and High Sensitivity DNA Assay (Table 3). The peak of the DNA fragment size was between 200–450 bps for each of the samples with an average of 355 bps (Fig. 4, Table 3).

PATIENT		BRAIN REGION
#194	control	occipital cortex
		occipital cortex
#230	control	somatomotor cortex
		occipital cortex
#148	suicide	occipital cortex
		occipital cortex
#143	suicide	occipital cortex
		occipital cortex
#158	suicide	cerebellar cortex
		occipital cortex

Table 4. Summary of the technical replicates. Different brain tissue samples (originating from the same region, but from different tissue pieces or different regions) from five patients (3 suicide victims and 2 controls) were used as technical parallels for sequencing.

Technical replicates

Two independent technical replicate libraries were prepared and sequenced from the same patients (three suicide victims and two controls; Table 4).

Quality control of raw reads

Illumina BCL files were converted to FASTQ files by the standard Illumina protocol to remove low-quality reads and adaptors.

Quality control of mapped reads

We used Picard tools MarkDuplicate algorithm to identify PCR or optical duplicates. According to MarkDuplicates statistics, we had very low percentages of duplicated reads (on average 0.624%) in our cohort. Genome Analysis Toolkit (GATK version 3.3) BaseRecalibrator base quality score recalibration tool was used to generate the final quality recalibrated BAM files for downstream analysis and variant calling according to best practices¹⁵⁻¹⁷. To check the quality of our mapped reads we used FastQC to generate quality reports of base quality score recalibrated BAM files. According to the FastQC reports we had no adapter contamination or overrepresented sequences indicating contamination in any of the analysed sequences. Furthermore, the majority of sequences were of high quality (average Phred quality score >34) and even the last two base pairs of the sequences had an average Phred quality score >30. On average 99.9% of reads were mapped, 99% were properly paired, and we had less than 0.1% singletons. The average read counts in the cohort were 117 013 694.9 while the average insert sizes were in the range of 157.7–190.0 base pairs (Table 5 (available online only)). According to the main quality indicators of our dataset all data files fall within acceptable parameters as shown in Table 5 (available online only).

Variant calling

The identification of different genomic variants including SNVs, indels, multiple nucleotide variants (MNVs), Genome Analysis Toolkit (GATK version 3.3) HaplotypeCaller algorithm was used with the GVCF mode cohort analysis according best practice¹⁵⁻¹⁷. HaplotypeCaller was run with the “-stand_emit_conf 10 -stand_call_conf 30” parameters. Variants were recalibrated by GATK VariantRecalibrator according to the best practice recommendations using the HapMap, Omni, dbSNP, 1000 G, Mills datasets for training included in the GATK GRCh37 bundle. We used 100.0, 99.9, 99.0, 90.0 tranches for variant recalibration and the 99.0 truth sensitivity level was used at ApplyRecalibration algorithm to identify QC PASS-ed variants.

Usage Notes

Ethical approval for the study was obtained from the following institutional review boards:

The brain samples were collected by the Human Brain Tissue Bank (HBTB; Semmelweis University, Budapest). The activity of the HBTB has been authorized by the Committee of Science and Research Ethic of the Ministry of Health Hungary (ETT TUKEB: 189/KO/02.6008/2002/ETT) and the Semmelweis University Regional Committee of Science and Research Ethic (No. 32/1992/TUKEB) to remove human brain tissue samples, collect, store and use them for research.

The HBTB is a member of the BrainNet Europe II. consortium. The activity of the HBTB is in accordance with the ethic and safety rules and the scientific requirements of the consortium.

Human brain microdissection procedures were approved by the Regional and Institutional Committee of Science and Research Ethics of Scientific Council of Health (the number of the ethical license: 34/2002/TUKEB-13716/2013/EHR, Budapest, Hungary) and the Code of Ethics of the World Medical Association (Declaration of Helsinki).

Conducting genetic testing on tissue samples and then sending them abroad has been authorized by the Semmelweis University Regional Committee of Science and Research Ethics (No. 34/2002/TUKEB).

The Institutional Review Board (IRB) at Stanford has made the following determination about the activity of the study, titled “*Exome sequencing analysis of suicidal behavior using high-throughput DNA sequencing*”, based on Office for Human Research Protections (OHRP) and Food and Drug Administration (FDA) regulations and guidance: this project does not require submission to the Stanford IRB, because the study does not involve human subjects, it is not about living individuals (NOT-H3 1/1; November 26, 2014).

The following explains how to access the dataset provided in this manuscript

The data is made available through the European Genome-phenome Archive (EGA). Search for the Data Access Committee (DAC) ID of our study (EGAD00001004184) at the homepage of EGA (<https://ega-archive.org/>). This will show detailed information on our dataset, data providers, DAC and any other related documentation. You can also find information about who to contact about access to this dataset. Access to data will be allowed to qualified researchers for appropriate health related studies.

A DAC group is responsible for the access of the datasets deposited in EGA.

Request for data access will be referred directly to our Data Access Committee: <https://ega-archive.org/datasets/EGAD00001004184>.

Data access decisions can be passed to the EGA in two ways: 1) By emailing helpdesk@ega-archive.org with the email address of each applicant and confirmation of the dataset/s to provide access. The EGA will then create an EGA account with the relevant access permissions. 2) By using the EGA DAC Admin tools (available to DAC's dealing with more than 5 data access applications/month).

If you need to request access to this data set, please contact:

DAC for Hungarian human exome team (Department of Medical Biology, University of Szeged, Hungary)

Contact person: Dr. Zsolt Boldogkői

Email: boldogkoi.zsolt@med.u-szeged.hu

Applicants will be asked to complete the Data Access Agreement (DAA) (including a brief summary of the proposal, proposed usage of the dataset, the storage of data, so the DAC can determine if the planned usage falls within the consents) and to agree to the terms and conditions of the DAA. The DAA must be signed by the applicant and the relevant Head of Department, or equivalent. If applications include a named collaborator then the collaborator's Institution must sign and submit a separate DAA. A template DAA can be found on the EGA website: <https://www.ebi.ac.uk/ega/sites/ebi.ac.uk/ega/files/documents/Example%20DAA.doc>.

References

- Hawton, K. & van Heeringen, K. Suicide. *The Lancet* **373**, 1372–1381 (2009).
- National Institute of Mental Health. *Statistics/Suicide*, <https://www.nimh.nih.gov/health/statistics/suicide/leading-causes-of-death-ages-1-85-in-the-us.shtml> (2018).
- Haw, C. & Hawton, K. Suicide is a complex behaviour in which mental disorder usually plays a central role. *Aust. N. Z. J. Psychiatry* **49**, 13–15 (2015).
- Platt, S. Suicide in men: what is the problem? *Trends - Urology & Men's Health* **8**, 9–12 (2017).
- Haw, C. *et al.* Suicide Clusters: A Review of Risk Factors and Mechanisms. *Suicide Life Threat. Behav.* **43**, 97–108 (2013).
- Tombácz, D. *et al.* High-Coverage Whole-Exome Sequencing Identifies Candidate Genes for Suicide in Victims with Major Depressive Disorder. *Sci. Rep.* **7**, 7106, <https://doi.org/10.1038/s41598-017-06522-3> (2017).
- Gradus, J. L. *et al.* Acute stress reaction and completed suicide. *Int. J. Epidemiol.* **39**, 1478–1484 (2010).
- Brent, D. What family studies teach us about suicidal behaviour: Implications for research, treatment, prevention. *Eur. Psychiatry* **25**, 260–263 (2010).
- Harris, E. C. & Baraclough, B. Suicide as an outcome for mental disorders. A meta-analysis. *Br. J. Psychiatry* **170**, 205–228 (1997).
- Tang, D. *et al.* Reference genotype and exome data from an Australian Aboriginal population for health-based research. *Sci. Data.* **3**, 160023, <https://doi.org/10.1038/sdata.2016.23> (2016).
- Merali, Z. *et al.* Dysregulation in the suicide brain: mRNA expression of corticotropin-releasing hormone receptors and GABA(A) receptor subunits in frontal cortical brain region. *J. Neurosci.* **24**, 1478–1485 (2004).
- Laszlo, A. *et al.* The use of regression methods for the investigation of trends in suicide rates in Hungary between 1963 and 2011. *Soc. Psychiatry Psychiatr. Epidemiol.* **50**, 249–256 (2015).
- Rihmer, Z., Gonda, X., Kapitány, B. & Dome, P. Suicide in Hungary-epidemiological and clinical perspectives. *Annals of General Psychiatry* **12**, 21 (2013).
- Chen, R. *et al.* Whole-exome sequencing identifies tetratricopeptide repeat domain 7 A (TTC7A) mutations for combined immunodeficiency with intestinal atresias. *J. Allergy Clin. Immunol.* **132**, 656–664.e17 (2013).
- McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

16. DePristo, M. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
17. Van der Auwera, G. A. *et al.* From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Curr. Protoc. Bioinformatics.* **43**, 11.10.1–11.10.33 (2013).

Data Citations

1. *European Genome-phenome Archive* EGAD00001004184 (2018).
2. Tombacz, D. *et al.* *Figshare* <https://doi.org/10.6084/m9.figshare.6803726.v1> (2018).

Acknowledgements

This work was funded by the European Union and the Hungarian State, co-financed by the European Social Fund under the framework of TÁMOP [4.2.2/B-10/1- 2010-0012] and the Swiss-Hungarian Cooperation Programme [SH/7/2/8] to ZBo. The work also received support from the Bolyai János Scholarship of the Hungarian Academy of Sciences to DT, and was supported by the NIH Centers of Excellence in Genomic Science (CEGS) Center for Personal Dynamic Regulomes [5P50HG00773502] to MS and by the High-throughput sequencer for multi-scale genomic studies [1S10OD02014101] to MS.

Author Contributions

Conceptualization: Z.B., T.K., Z.M., M.P., M.S., D.T. Analysis and statistics: Z.M., Z.B., D.T., T.K. Funding acquisition: M.S., Z.B. Investigation, sequencing: D.T. Resources: M.P., M.S., Z.B. Supervision: Z.B., M.S. Visualization: D.T., Z.M., T.K., Z.B. Writing – Original Draft Preparation: D.T., Z.B., Z.M., T.K. All authors reviewed and approved the final version of the manuscript.

Additional Information

Tables 1, 2 and 5 are only available in the online version of this paper.

Supplementary information accompanies this paper at <http://www.nature.com/sdata>

Competing interests: The authors declare no competing interests.

How to cite this article: Tombacz, D. *et al.* Whole-Exome Sequencing Data of Suicide Victims who had Suffered from Major Depressive Disorder. *Sci. Data.* **6**:190010 <https://doi.org/10.1038/sdata.2019.10> (2019).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files made available in this article.

© The Author(s) 2019