# SUPPLEMENTARY DATA

## Material and Methods

### Whole Exome Sequencing (WES)

Exome sequencing was performed at the Centre for Genomic Sciences (CGS) of the Li Ka Shing Faculty of Medicine of the University of Hong Kong on blood or saliva DNA. Illumina’s TruSeq® DNA Sample Prep v.2 and TruSeq® Exome Enrichment Kits (Illumina, San Diego, CA, USA) were used for sample preparation and capture, and enrichment of targeted sequences, respectively. The captured DNA was sequenced as paired-end 100 base reads (PE100) on an Illumina HiSeq 2000, aiming to achieve 50 reads per base (50X) on average.

*Quality control (QC) and variant calling*

The quality assessment of sequencing reads starts with the raw reads. The base quality, duplication levels, GC bias and primer sequences of the raw sequencing reads were evaluated using FastQC[1](#_ENREF_1" \o "Andrews, 2010 #638). All samples passed the sequencing quality thresholds.

Afterwards, sequencing reads were aligned to the human genome reference hg19 by the Burrows-Wheeler Aligner (BWA v 0.7.8) to produce the sequence alignment file. For the detection of contaminated samples, we used the verifyBAMID[2](#_ENREF_2" \o "Jun, 2012 #329). This program takes the sequence alignment file and the 1000 genome reference sites file as input to evaluate the heterogeneity of each given site of our dataset. Individuals exceeding 3% in heterogeneity of supplied sites should be excluded from further analysis. One proband (MTS36) failed the test and was excluded.

Calling and filtering of single nucleotide variants (SNVs) and indels (small insertions/deletions) were done by the Genome Analysis Toolkit (GATK 3.3-0) haplotype-caller and Variant Quality Score Recalibration (VQSR) module respectively[3](#_ENREF_3). VQSR compares known sites in variant databases (1000 Genome and dbsnp137) with novel variants on a number of sequencing parameters in our dataset. This aims to investigate relationships between sensitivity and specificity of variant calling. There are seven sensitivity tranches defined between 90% and 100% in our dataset. After plotting the relationship between sensitivity and specificity, an optimal VQSR sensitivity tranche was determined to be ≤99.5 and therefore variants in tranches with >99.5 sensitivity were excluded. Then we looked at the variants of each individual sequenced. Hard filtering on parameters was performed by KGGSeq on each individual variant with depth of coverage ≤4. Genotyping quality <20 were set as missing.

Principal component analysis (PCA), implemented in PLINK 1.94, was used for the detection of population stratification. Sample relationship checks were performed in pairwise fashion based on identity-by-descent (IBD) sharing (implemented in PLINK). One trio was excluded from all family based analysis because the mother was found to be genetically unrelated to the patient.

For the purpose of association testes, we evaluated the genotypes at “genomic site level” and removed low confidence variations. Additional criteria for good genotyping quality included: <10% miss rate and conforming to Hardy-Weinberg equilibrium (HWE) (p < 0.0001). Then, to enhance the calling accuracy and to facilitate the filtering process, 692 samples from local Chinese individuals participating in a study investigating lumbar disc degeneration were added to the MTLE-HS calling set. These samples had also been sequenced at the Centre for Genomic Sciences (CGS) and processed by a similar exome sequencing pipeline. Principal component analysis (PCA) indicated no observable bias between MTLE-HS samples sequenced and these population control samples (Figure e-1). The resulting variants were subjected to further selection as described in the next section. Sanger sequencing was used for validation of selected non-synonymous de novo and inherited variants.

### Testing enrichment of evolutionarily constrained genes and genetic intolerance

Constraint scores were calculated by a mutation rate table that contains the probability of each triplet mutating to every other combination of triplets by only changing the third nucleotide[4](#_ENREF_4). The calculation was based on intergenic SNPs from the 1000 Genomes project, accounting for gene length, to determine that gene’s probability of mutation. These probabilities of mutation were also corrected for regional divergence and depth of coverage for each base by exome sequencing study included in the calculation. Since there is a high correlation (r = 0.94) between the probability of a synonymous mutation with the number of rare (MAF < 1%) synonymous variants in the same gene seen in NHLBI’s Exome Sequencing Project, a linear model was used to predict the number of rare missense variants expected per gene in the same dataset The difference between observation and expectation was quantified as a signed Z-score under chi-squared distribution. Thus, the calculated missense Z-score can be used as the basis for determining selective constraint.

*Tests on inherited variants: homozygosis and compound heterozygosity*

For the first approach, we queried the Phenolyzer [5](#_ENREF_5) database for three terms, “mesial temporal lobe epilepsy”, “hippocampus”, and“temporal lobe epilepsy”. The query returned 109 seed genes associated with the keywords. Only five double hit events overlapped with the list. Candidates were also fetched from SynaptomeDB, and 1 gene overlapped with that list. The significance of remaining events was tested by Fisher exact test to compare occurrences in cases vs controls.

Rare loss-of-function (LoF) compound heterozygous mutations were analyzed separately from the missense variant analysis.

Finally, the compound heterozygosity of genes carrying *de novo* variants (listed in Table 5) was also investigated in other MTLE-HS patients. Some of the homozygous or compound heterozygous variants were selected for validation by Sanger sequencing.

## Figure e-1: PCA plot of MTLE-HS cases and in-house controls



## **Table e-1:** Candidates focal epilepsy genes tested in gene-set[6](#_ENREF_6)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *CHRNA2**CHRNA4**CHRNB2**CNTNAP2**DEPDC5* | *FLNA**GABRG2**GRIN2A**KCNQ2**KCNQ3* | *KCNT1**LGI1**MECP2**NPRL2**NPRL3* | *PCDH19**POLG**PRIMA1**PRRT2**RELN* | *SCN1A**SCN1B**SCN2A**SLC2A1 SRPX2* | *SYN1**TBC1D24**TSC1**TSC2**VPS13A* |

## **Table e-2**: Gene-sets tested in association

|  |  |  |  |
| --- | --- | --- | --- |
| Set name |  | Number of genes | Description |
| ARC |  | 28 | AIDS-related complex (ARC) |
| ASD\_candidates |  | 112 | Autism spectrum disorder (ASD) candidates |
| ASD\_DENOVO\_NONSYN |  | 743 | Autism spectrum disorder (ASD) *de novo* non synonymous variants |
| ASD-49 |  | 49 | 49-gene PPI network derived from ASD *de novo* mutations |
| ASD-74 |  | 74 | 74-gene PPI network derived from ASD *de novo* mutations |
| ASD\_dnv\_lof |  | 128 | Autism spectrum disorder (ASD) *de novo* Loss of Function (LoF) variants |
| calcium-channel |  | 26 | Calcium ion channel genes |
| CHD8\_Network |  | 6 | CHD8-based PPI network |
| FMRP\_Asc |  | 939 | Fragile X mental retardation protein (FMRP) by Ascano |
| FMRP\_ASD\_overlap |  | 93 | Overlap between FMRP and ASD candidate genes |
| FMRP\_Darnell |  | 788 | Fragile X mental retardation protein (FMRP) by Darnell |
| ID\_Candidates |  | 196 | Intellectual disability (ID) candidates |
| ID\_dnv\_lof |  | 30 | Intellectual disability (ID) *de novo* Loss of Function (LoF) variants |
| ID\_dnv\_nonsyn |  | 132 | Intellectual disability (ID) *de novo* non-synonymous variants |
| kirov-denovo-cnv |  | 234 | *de novo* CNVs from Kirov et al. |
| mGluR5 |  | 39 | mGluR5 complex |
| miR-137 |  | 446 | Targets of miR-137 |
| NMDAR |  | 61 | N-methyl-D-aspartate (NMDA) network |
| PSD-95 |  | 65 | PSD-95 complex |
| PSD |  | 685 | Postsynaptic density genes |
| psych-cnv |  | 346 | Genes in regions of associated CNVs in SCZ/ASD |
| scz-denovo-lof |  | 87 | Schizophrenia (SCZ) *de novo* Loss of Function (LoF) variants |
| scz-denovo-nonsyn |  | 611 | Schizophrenia(SCZ) *de novo* non-synonymous variants |
| scz-gwas |  | 479 | Associated GWAS LD-intervals, Swedish study |
| FMRP\_MERGED |  | 1557 | Merged FMRP set by Ascano and Darnell |

## Table e-3: Enrichment test result for 25 gene-sets over 7 expression studies.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Set name1 | Expression study | Overlaps2 | *n* DEGs | Set size(genes) | *p* | Corrected *p* |
| PSD | Lees *et al.* 2007[11](#_ENREF_11) | 112 | 685 | 722 | 2.37 x10-41 | 4.75x10-39 |
| FMRP\_Darnell | Lees *et al* 2007*.*[11](#_ENREF_11) | 91 | 788 | 722 | 3.13 x10-22 | 6.27x10-20 |
| *PSD* | Van *et al.* 2008[12](#_ENREF_12) | 67 | 685 | 491 | 6.61 x10-22 | 1.32x10-19 |
| PSD | Intersected\* | 41 | 685 | 233 | 5.59 x10-18 | 1.12x10-15 |
| PSD | Ozbas *et al.* 2006[13](#_ENREF_13) | 27 | 685 | 126 | 1.61 x10-14 | 3.22x10-12 |
| FMRP-Ascano+Darnell | Lees *et al.* 2007[11](#_ENREF_11) | 117 | 1557 | 722 | 1.03 x10-13 | 2.07x10-11 |
| NMDAR | Lees *et al.* 2007[11](#_ENREF_11) | 19 | 61 | 722 | 3.89 x10-13 | 7.78x10-11 |
| mGluR5 | Lees *et al.* 2007[14](#_ENREF_14) | 12 | 39 | 722 | 9.88 x10-9 | 1.98x10-6 |
| NMDAR | Intersected\* | 9 | 61 | 233 | 3.61 x10-8 | 7.21x10-6 |
| PSD-95 | Lees *et al.* 2007[14](#_ENREF_14) | 13 | 65 | 722 | 5.99 x10-7 | 1.20x10-4 |
| FMRP\_Darnell | Van *et al.* 2008[12](#_ENREF_12) | 43 | 788 | 491 | 1.03 x10-6 | 0.0002 |
| FMRP\_Darnell | Intersected\* | 24 | 788 | 233 | 1.43 x10-5 | 0.0029 |
| NMDAR | Van *et al.* 2008[12](#_ENREF_12) | 9 | 61 | 491 | 1.78 x10-5 | 0.0036 |
| mGluR5 | Van *et al.* 2008[12](#_ENREF_12) | 7 | 39 | 491 | 4.22 x10-5 | 0.0084 |
| PSD | Arion *et al.* 2006[15](#_ENREF_15) | 10 | 685 | 64 | 5.06 x10-5 | 0.0101 |
| mGluR5 | Intersected\* | 5 | 39 | 233 | 8.46 x10-5 | 0.0169 |
| mGluR5 | Ozbas *et al.* 2006[13](#_ENREF_13) | 4 | 39 | 126 | 1.02 x10-4 | 0.0204 |
| PSD-95 | Intersected\* | 6 | 65 | 233 | 1.05 x10-4 | 0.0210 |
| FMRP-Ascano+Darnell | Van *et al.* 2008[12](#_ENREF_12) | 61 | 1557 | 491 | 1.54 x10-4 | 0.0308 |
| PSD-95 | Van *et al.* 2008[12](#_ENREF_12) | 8 | 65 | 491 | 1.84 x10-4 | 0.0367 |
| FMRP-Ascano+Darnell | Intersected\* | 34 | 1557 | 233 | 1.87 x10-4 | 0.0375 |
| 1: Detail of gene-sets see Supplementary Table 22: Number of differentially expressed genes overlapping with the gene-set*n* DEG: number of differentially expressed genes\*: An intersected DEGs list which requires at least 2 of 7 included studies reporting the same DEG as a callout. |

# Only tests with p<0.05 after correction for multiple comparisons (25x7 tests) are shown.

## **Table e-4:** Ratios of non-synonymous/synonymous *de novo* mutations across studies

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Controls | This study | Schizophrenia [7](#_ENREF_7) | Autism spectrum disorder [8](#_ENREF_8) | Intellectual disability [9](#_ENREF_9), [10](#_ENREF_10) |
| Non-synonymous | 434 | 18 | 482 | 789 | 141 |
| Synonymous | 155 | 3 | 155 | 255 | 25 |
| Ratio | 2.8 | 6.0 | 5.1 | 3.1 | 5.6 |
| P vs controls | − | 0.22 | 0.43 | 0.41 | 0.0027 |

## Table e-5: Enrichment test of the 25 gene-sets and genes carrying *de novo* variants

|  |  |  |  |
| --- | --- | --- | --- |
| Set | Overlapping genes | Set size | *p*-values |
| FMRP-Ascano+Darnell | 6 | 1557 | 0.0013 |
| ARC | 1 | 28 | 0.0237 |
| FMRP-Darnell | 3 | 788 | 0.0249 |
| FMRP-Asc | 3 | 939 | 0.0376 |
| PSD-95 | 1 | 65 | 0.0534 |
| ASD-candidates | 1 | 112 | 0.0885 |
| PSD | 2 | 685 | 0.0971 |
| scz-gwas | 1 | 479 | 0.279 |
| scz-denovo-nonsyn | 1 | 611 | 0.318 |

# The other 16 gene-sets were not tested, as no *de novo* variants were found in any genes in those sets.

## Table e-6: *CEP170B* variants

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Patient | Type |  | Variant type |  | Pos | Change | RSID | Family history of epilepsy |
| 40 | *Comp Het* |  | M |  | 105349504 | C/T | rs187049858 | Y (sib + son) |
| M |  | 105353799 | C/T | rs565981186  |
| 25 | *Homo* |  | M |  | 105349504 | C/T | rs187049858 | N |
| 12 | *de novo* |  | M |  | 105353332 | C/T |  | N |
| ASD:14076[1] | *de novo* |  | M |  | 105354158 | C/G |  |  |
| M: missense [1]: proband of a ASD family based study |

## REFERENCES

1. Andrews S. FastQC: A quality control tool for high throughput sequence data. Reference Source 2010.

2. Jun G, Flickinger M, Hetrick KN, et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. American journal of human genetics 2012;91:839-848.

3. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome research 2010;20:1297-1303.

4. Samocha KE, Robinson EB, Sanders SJ, et al. A framework for the interpretation of de novo mutation in human disease. Nature genetics 2014;46:944-950.

5. Yang H, Robinson PN, Wang K. Phenolyzer: phenotype-based prioritization of candidate genes for human diseases. Nature methods 2015;12:841-843.

6. Perucca P, Scheffer IE, Harvey AS, et al. Real-world utility of whole exome sequencing with targeted gene analysis for focal epilepsy. Epilepsy Res 2017;131:1-8.

7. Fromer M, Pocklington AJ, Kavanagh DH, et al. De novo mutations in schizophrenia implicate synaptic networks. Nature 2014;506:179-184.

8. Neale BM, Kou Y, Liu L, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature 2012;485:242-245.

9. Rauch A, Wieczorek D, Graf E, et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet 2012;380:1674-1682.

10. de Ligt J, Willemsen MH, van Bon BW, et al. Diagnostic exome sequencing in persons with severe intellectual disability. The New England journal of medicine 2012;367:1921-1929.

11. Lee TS, Mane S, Eid T, et al. Gene expression in temporal lobe epilepsy is consistent with increased release of glutamate by astrocytes. Mol Med 2007;13:1-13.

12. van Gassen KL, de Wit M, Koerkamp MJ, et al. Possible role of the innate immunity in temporal lobe epilepsy. Epilepsia 2008;49:1055-1065.

13. Ozbas-Gerceker F, Redeker S, Boer K, et al. Serial analysis of gene expression in the hippocampus of patients with mesial temporal lobe epilepsy. Neuroscience 2006;138:457-474.

14. Jais JP, Knebelmann B, Giatras I, et al. X-linked Alport syndrome: natural history and genotype-phenotype correlations in girls and women belonging to 195 families: a "European Community Alport Syndrome Concerted Action" study. Journal of the American Society of Nephrology : JASN 2003;14:2603-2610.

15. Arion D, Sabatini M, Unger T, et al. Correlation of transcriptome profile with electrical activity in temporal lobe epilepsy. Neurobiology of disease 2006;22:374-387.