The American Journal of Human Genetics, Volume *110*

Supplemental information

Epilepsies of presumed genetic etiology show enrichment of rare variants that occur in the general population

Linnaeus Bundalian, Yin-Yuan Su, Siwei Chen, Akhil Velluva, Anna Sophia Kirstein, Antje Garten, Saskia Biskup, Florian Battke, Dennis Lal, Henrike O. Heyne, Konrad Platzer, Chen-Ching Lin, Johannes R. Lemke, Diana Le Duc, and Epi25 Collaborative

Supplemental Figures

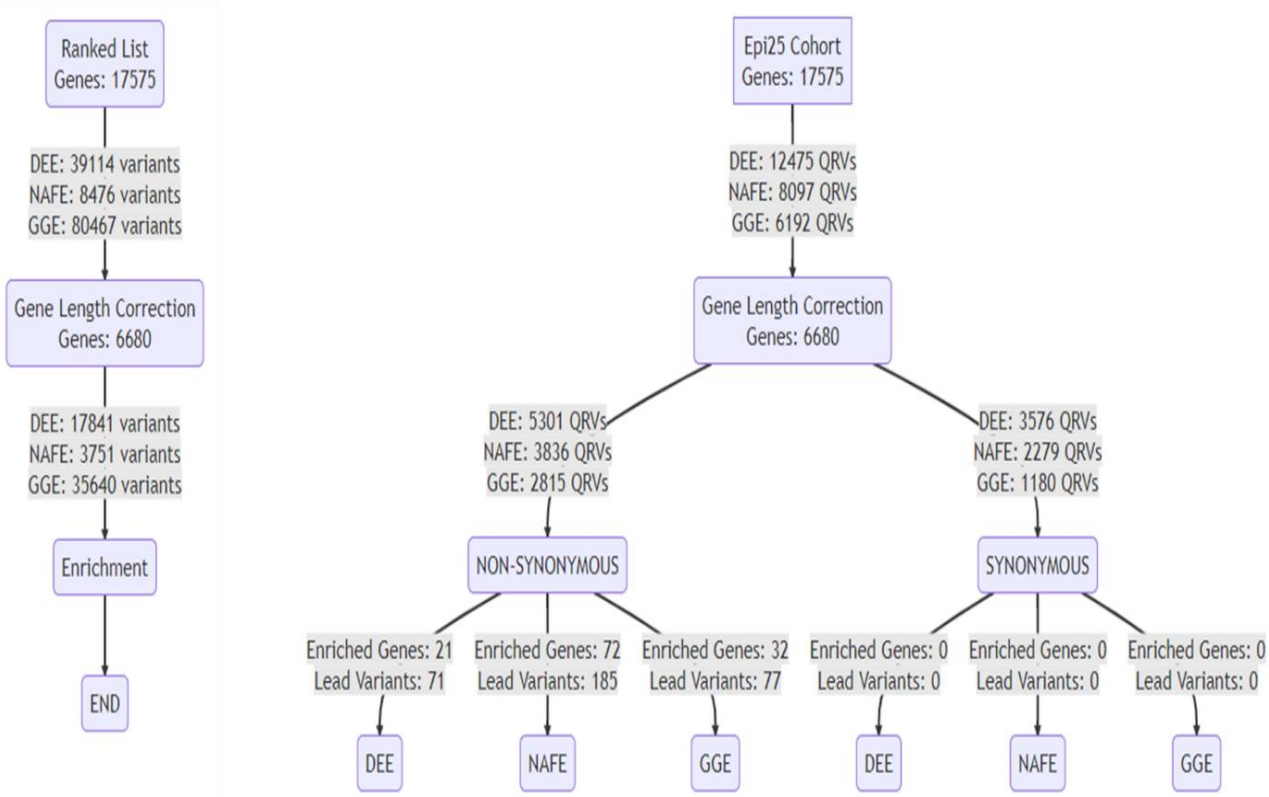

Figure S1. Gene and variant counts through the stages of analysis. The number of genes and variants under consideration changes as we implement different thresholds to filter out genes and variants that will best suit for the analysis. This is to lessen the effect of the influence of gene length for the genes and to increase specificity for variant consideration.

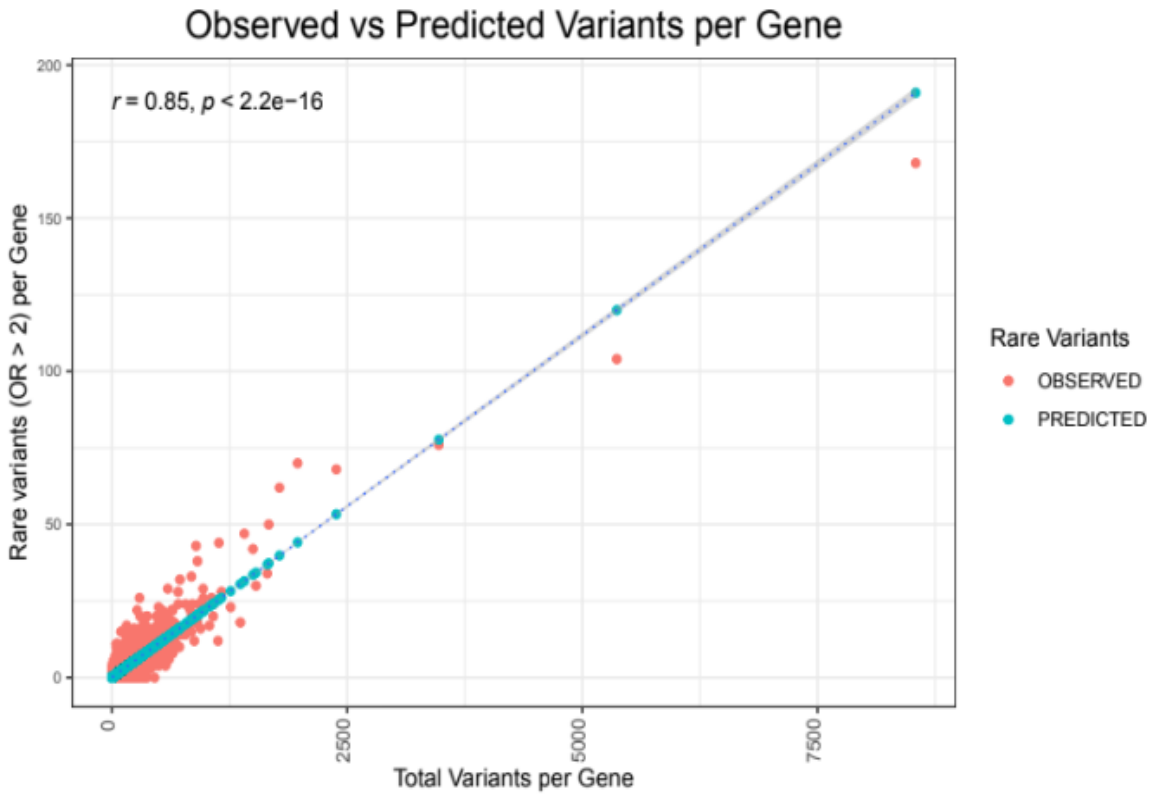

Figure S2. Relationship of the number of rare variants and total number of variants in a gene. The linear regression model (r = 0.85) established that for a given number of variants per gene (total variants), we can estimate the expected number of rare variants. Only those genes which had an excess of expected number variants (i.e. above the dotted slope) were considered for the VSEA.

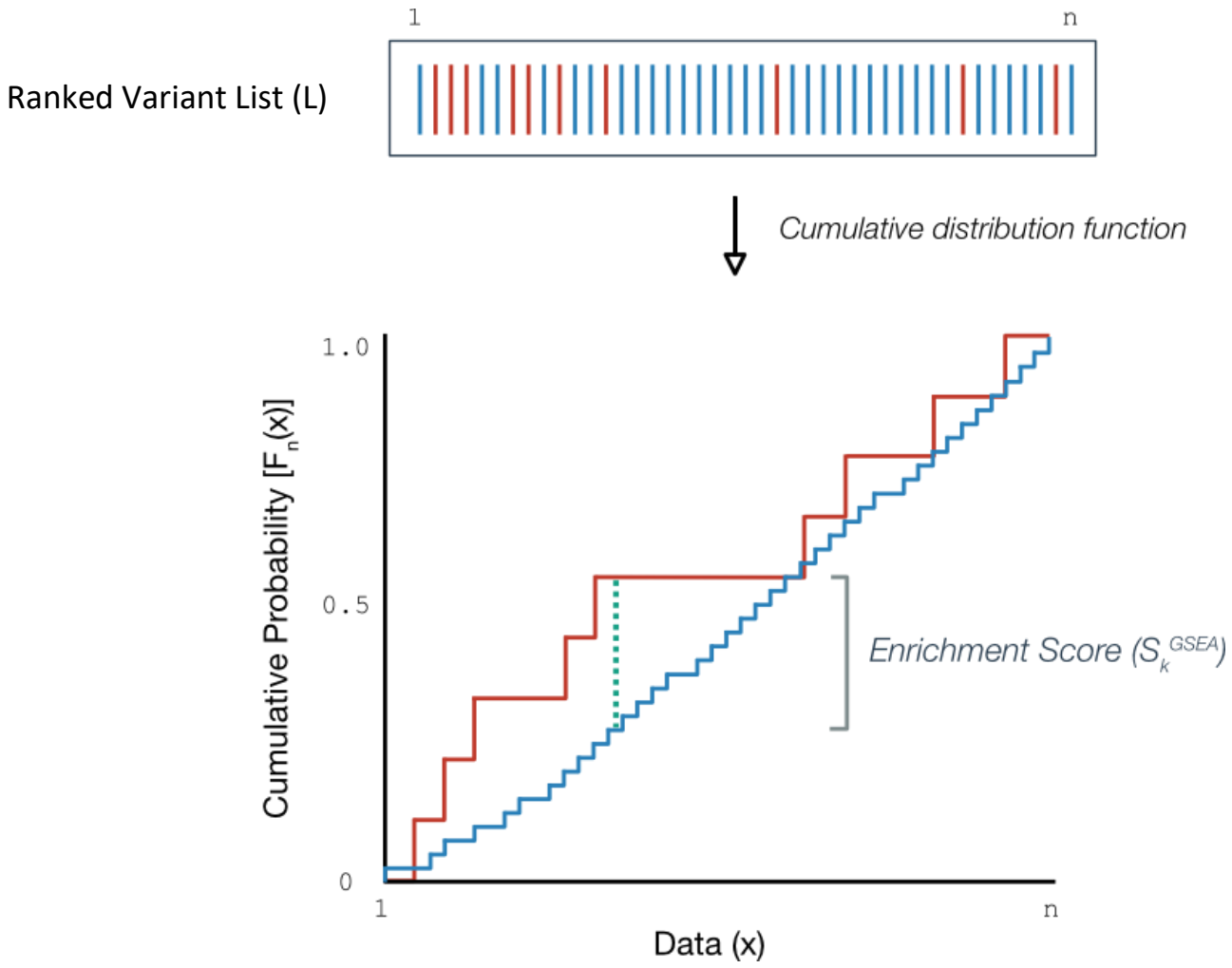

Figure S3. Cumulative ranked sum for hits and miss of QRVs. For every QRVs found in the list of variants having a higher frequency in patients – Data(x), the cumulative HIT score (red) will increment by 1/n otherwise the cumulative MISS score (blue) will be incremented. The normalized difference between the MISS and HIT score yields the ENRICHMENT SCORE. Note: The figure used was based from information and figures found in the Pathway Commons website. (https://www.pathwaycommons.org/guide/primers/data_analysis/gsea/#safe)

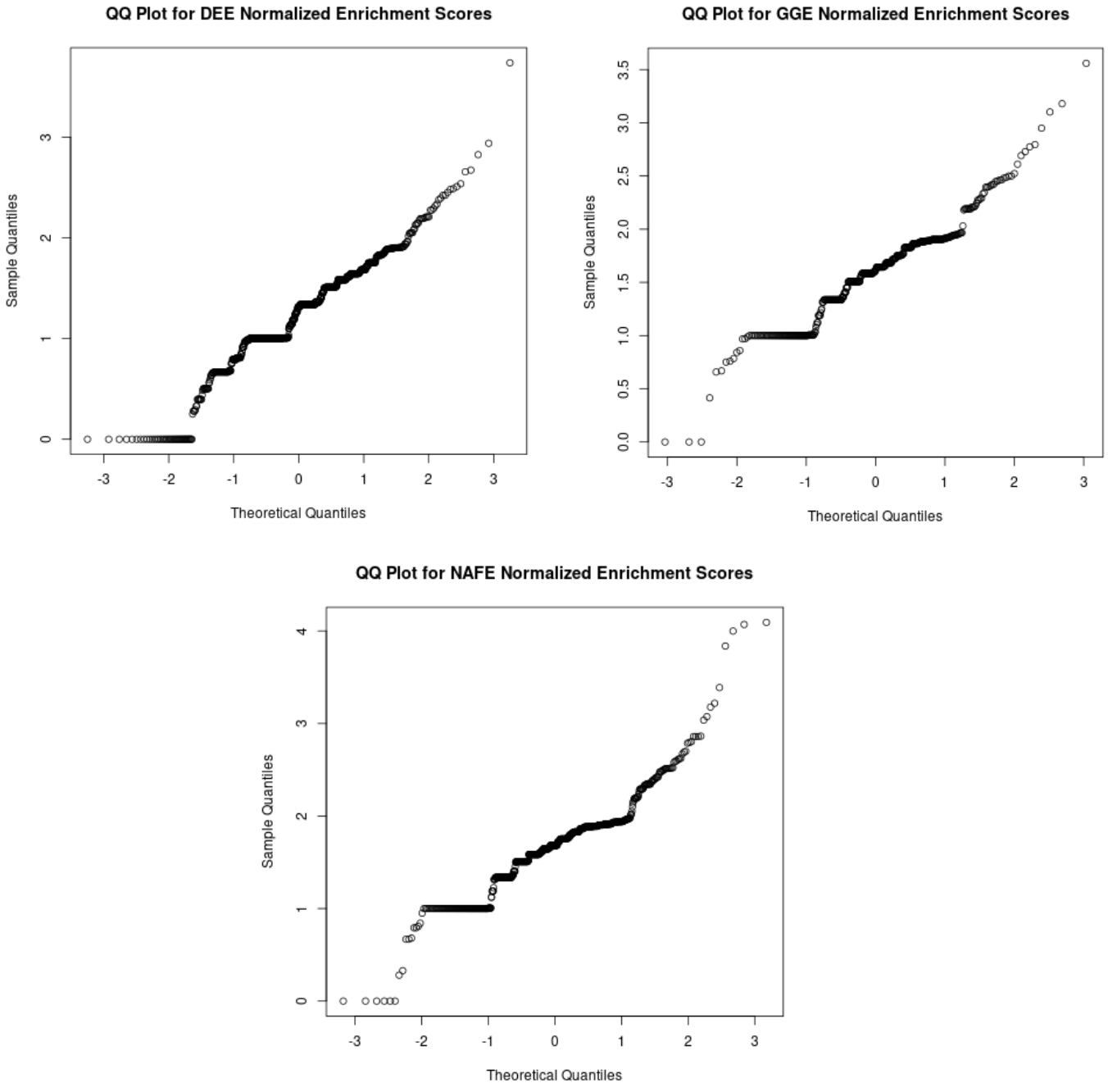

Figure S4. QQ-plot of NES across all the genes per EPI group. The plots show a normal distribution of Normalized Enrichment Scores (NES).

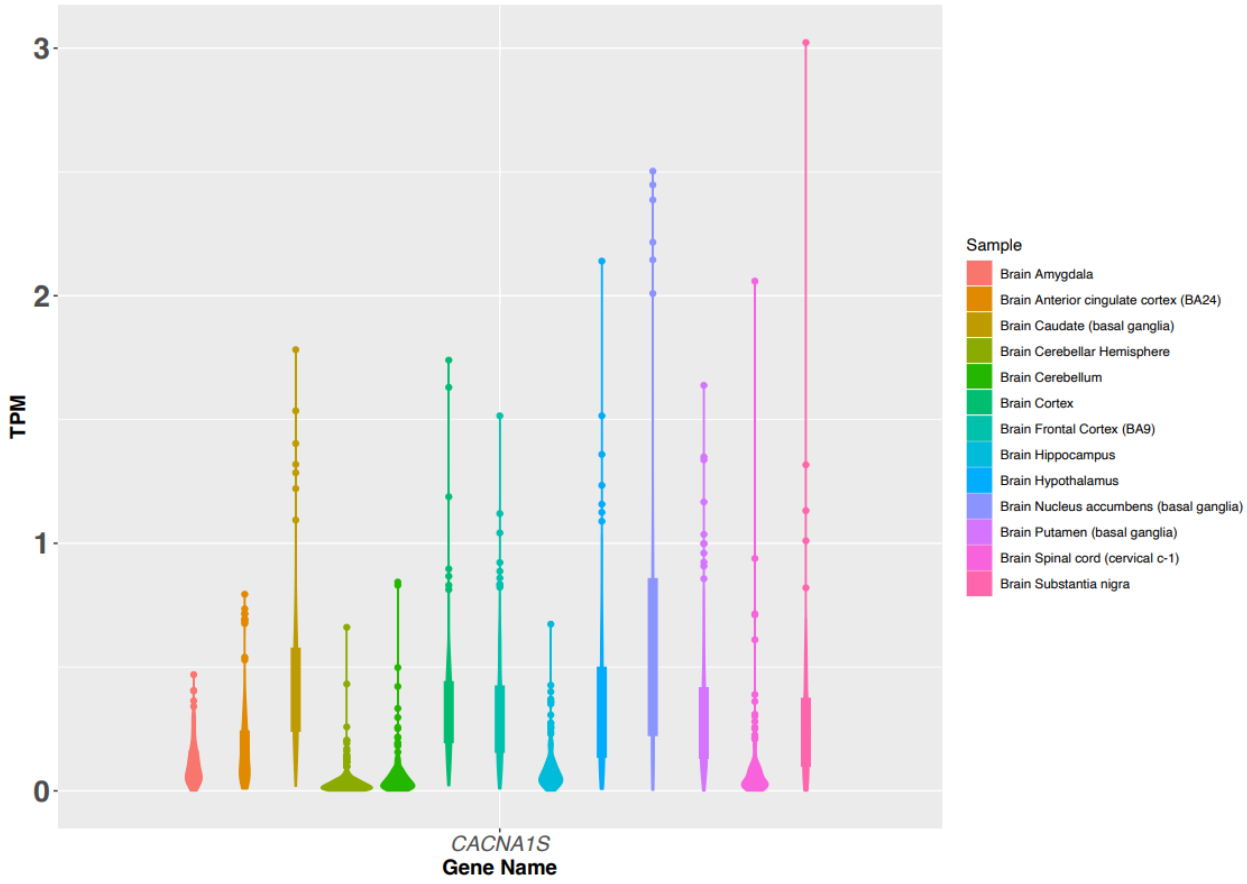

Figure S5. Expression of *CACNA1S* from Single Gene Analysis. *CACNA1S* is not that highly expressed in brain. However, using the *Single Gene Analysis* feature of PTEE (*https://bioinf.eva.mpg.de/PTEE*), we can observe the outliers in across different brain-related tissues. A previous study suggested that the voltage-gated calcium channel encoded by *CACNA1S* could play a role in mediating calcium influx and neuronal excitability[1].

Table S1. Variant set enrichment results for non-synonymous and synonymous variants per epilepsy subtypes. The table contains all the results of the Variant Set Enrichment Analysis for non-synonymous and synonymous variants across 3 epilepsy subtypes, DEE, NAFE and GEE. A gene is considered to be QRV-enriched if it has Z score ≥ 1.96 and FDR ≤ 0.05.

Table S2. GoFuncR overrepresented GO terms for QRV-enriched genes across different epilepsy subtypes. The table contains the resulting overrepresented GO terms for the QRV-enriched genes across 3 epilepsy subtypes in the study. *Note: The analysis was done using GoFuncR with a 95% confidence level. The returned result does not include the confidence interval.*

Table S3. Information of the QRVs enriched in genes for non-synonymous and synonymous variants across different epilepsy subtypes. The table contains the variants that cause the enrichment for the QRV-enriched genes. The annotation of variants was done using the Ensembl's Variant Effect Predictor for human genome assembly GRCh37

Supplemental References

1. Cain SM, Snutch TP. Voltage-Gated Calcium Channels in Epilepsy. (2012). In: Noebels JL, Avoli M, Rogawski MA, Olsen RW, Delgado-Escueta AV, editors. Jasper's Basic Mechanisms of the Epilepsies [Internet]. 4th ed. Bethesda (MD): National Center for Biotechnology Information (US).