Quality Control metrics of individual genetic variants in the Alzheimer's Disease Sequencing Project are associated with FAVOR annotations.

Naveen Kannan[1], Nicholas Wheeler[1], Genome Center for Alzheimer's Disease, Li-San Wang[2], Yuk Yee Leung[2], William S. Bush[1]

1) Cleveland Institute for Computational Biology, Department for Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, Ohio 44106, USA. 2) Department of Pathology and Laboratory Medicine, Penn Neurodegeneration Genomics Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.

Introduction

- The ADSP (Alzheimer's Disease Sequencing Project) released their 4th dataset consisting of **36,361 sequenced genomes, with 362 million variants.**
- The R4 dataset release also contains **variant level QC (Quality Control)** scores, along with composite **binary VFlags (5 in total) for each variant**, generated from the individual variant level QC scores.

FAVOR annotations

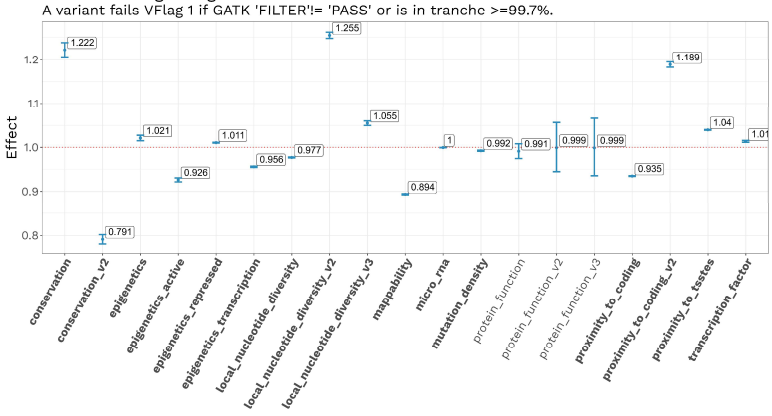
- FAVOR integrates data from multiple databases, including **CADD v1.5, GENCODE v31, Annovar, WGSA, ClinVar, ENCODE, SnpEff, 1000 Genome, TOPMed Bravo Freeze 8 and gnomAD v3.**
- FAVOR functional scores are divided into **17 groups**, along with **annotation Principal Components (aPCs)**, which are the first variant-specific PC calculated from each standardized individual annotation score within these 17 groups.

Annotations and QC metrics

- Approaches for statistical analysis of rare variants **increasingly rely on functional annotations** to weight association test statistics and increase statistical power.
- However, the **impact of variant quality** on these tests is largely unexplored.
- We performed **logistic regression analyses** on chromosomes **21 and 22** of the R4 dataset, with group **aPCs** as predictors, and **VFlags** as outcomes.

Predicted Effects of a change of one unit in the FAVOR aPCs on the odds of failing VFlag 1.

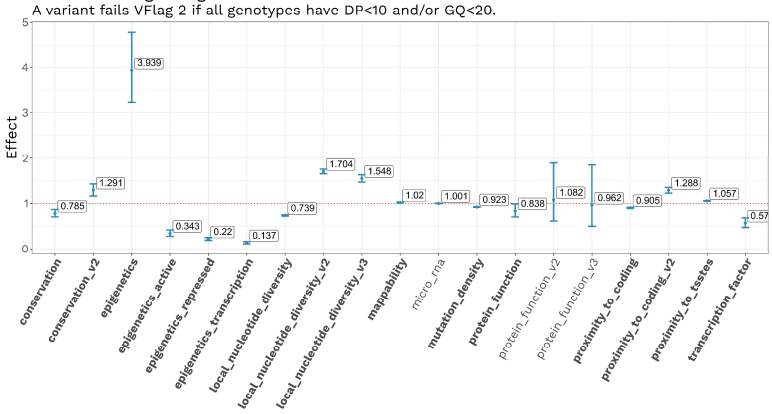
A variant fails VFlag 1 if GATK 'FILTER'!= 'PASS' or is in tranche >=99.7%.



Variants taken from Chromosomes 21 and 22 from the ADSP R4 data release.

Predicted Effects of a change of one unit in the FAVOR aPCs on the odds of failing VFlag 2.

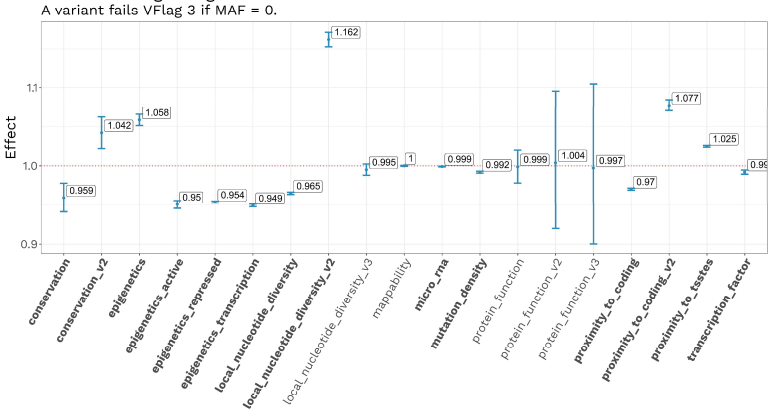
A variant fails VFlag 2 if all genotypes have DP<10 and/or GQ<20.



Variants taken from Chromosomes 21 and 22 from the ADSP R4 data release.

Predicted Effects of a change of one unit in the FAVOR aPCs on the odds of failing VFlag 3.

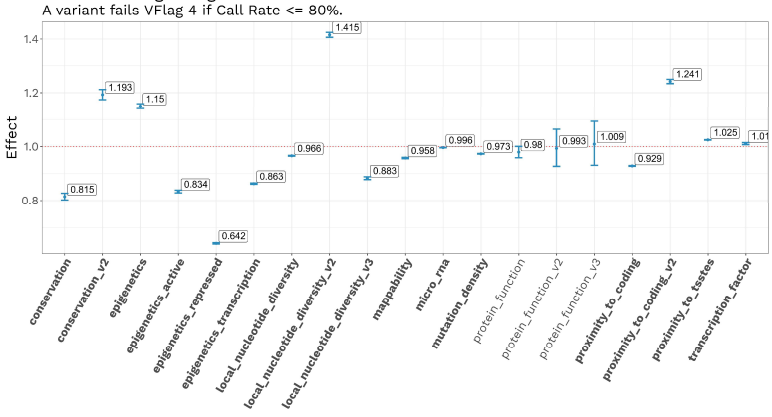
A variant fails VFlag 3 if MAF = 0.



Variants taken from Chromosomes 21 and 22 from the ADSP R4 data release.

Predicted Effects of a change of one unit in the FAVOR aPCs on the odds of failing VFlag 4.

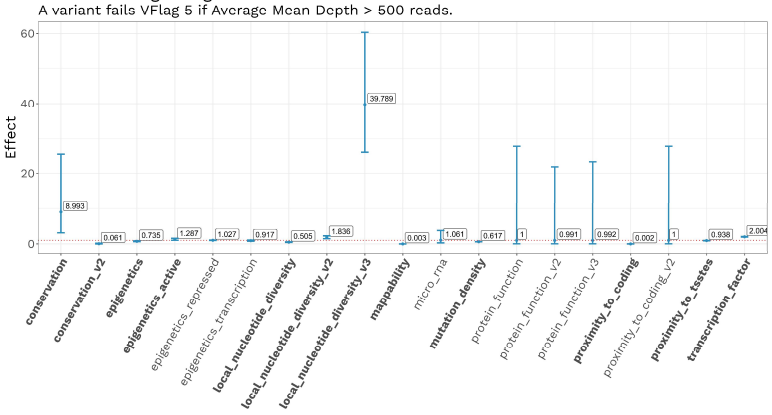
A variant fails VFlag 4 if Call Rate <= 80%.



Variants taken from Chromosomes 21 and 22 from the ADSP R4 data release.

Predicted Effects of a change of one unit in the FAVOR aPCs on the odds of failing VFlag 5.

A variant fails VFlag 5 if Average Mean Depth > 500 reads.



Variants taken from Chromosomes 21 and 22 from the ADSP R4 data release.

Results

- The aPCs were found to have a **statistically significant relationship with the odds of a variant failing the VFlags**, with some aPCs having different effects on different VFlags.
- For example, an increase in the aPC for the **"Epigenetics"** block by one unit for a variant increases the odds of failing VFlag 2 (All Genotypes have DP < 10 or GQ < 20) **by nearly four fold (95%CI=3.24 to 4.78, p < 10^-8)**, and the aPC for **"Local Nucleotide Diversity"** dramatically increases odds of failing VFlag 5 (Average Mean Depth > 500 reads) **by 40 fold (95%CI=26.22 to 60.36, p < 10^-8).**
- These relationships demonstrate the critical importance of variant quality filtering when using annotation weights in association testing.

Acknowledgements: This work was supported by funding from the National Institute on Aging **U54 AG052427 (Schellenberg, Wang).**