

# Reanalysis of GNPC Data Uncovers Pervasive Risk of False Positives and False Negatives in Biomarker Discovery from Large, Multi-center Proteomic Studies

Evan Boyle, Katrina Paumier, Ludmila Voloboueva, William Hagstrom, Ferhan Qureshi, David Brazel  
Octave Bioscience, Inc, Menlo Park, CA, USA



P16-014

## Introduction

**Background:** The Global Neurodegeneration Proteomics Consortium (GNPC) V1 harmonized dataset harbors measurements for 7,289 unique proteins from multiple Somalomic assays across >35,000 biofluid samples from Alzheimer's disease, frontotemporal dementia, Amyotrophic Lateral Sclerosis, and Parkinson's disease (PD) patients, plus healthy controls (HCs). Initial associations have been published, but more sophisticated statistical analysis could refine insight into biomarkers for PD.

## Objectives

- To assess reliability of biomarker associations reported from GNPC data
- To improve sensitivity and specificity of biomarker discovery for neurodegeneration

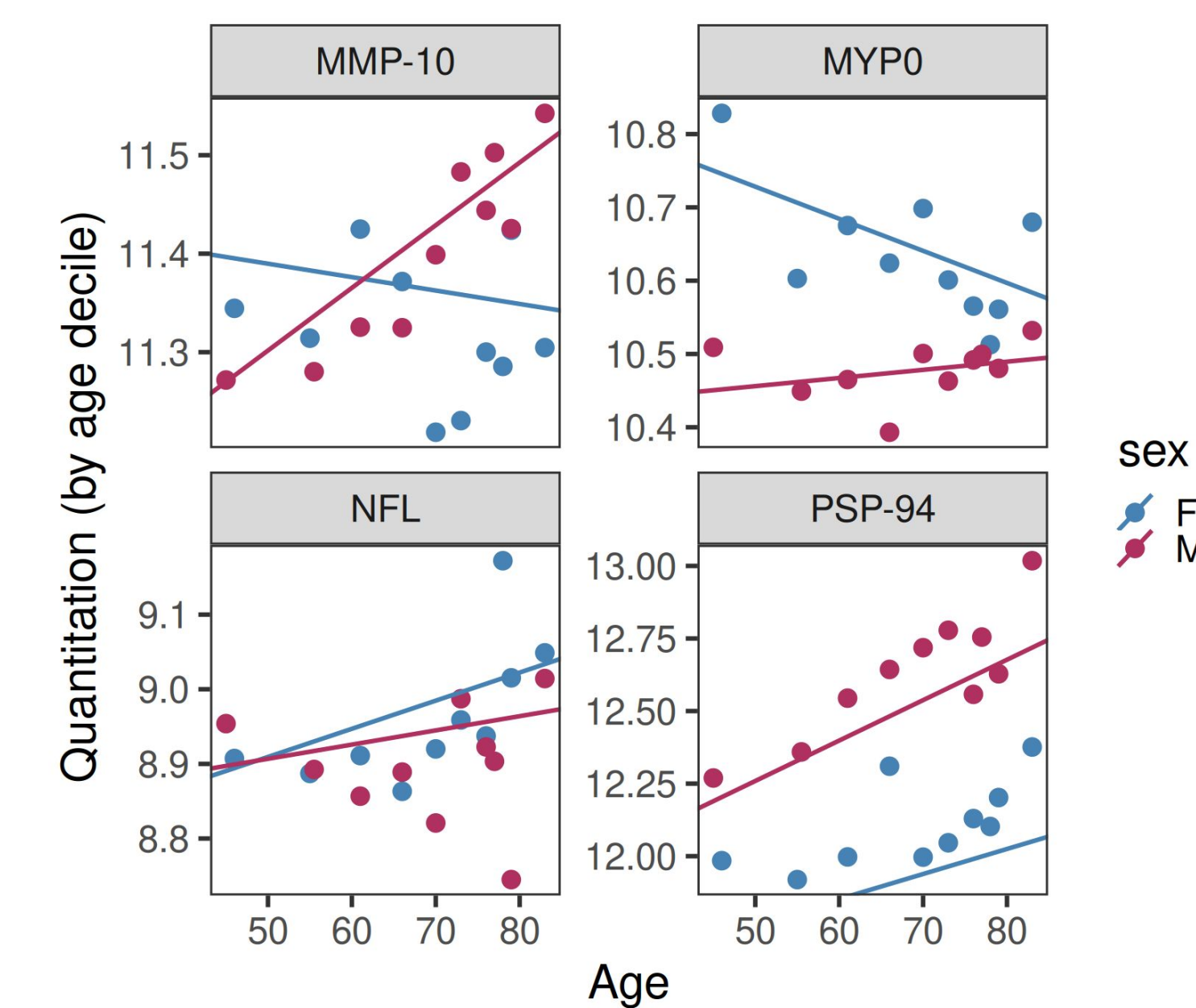
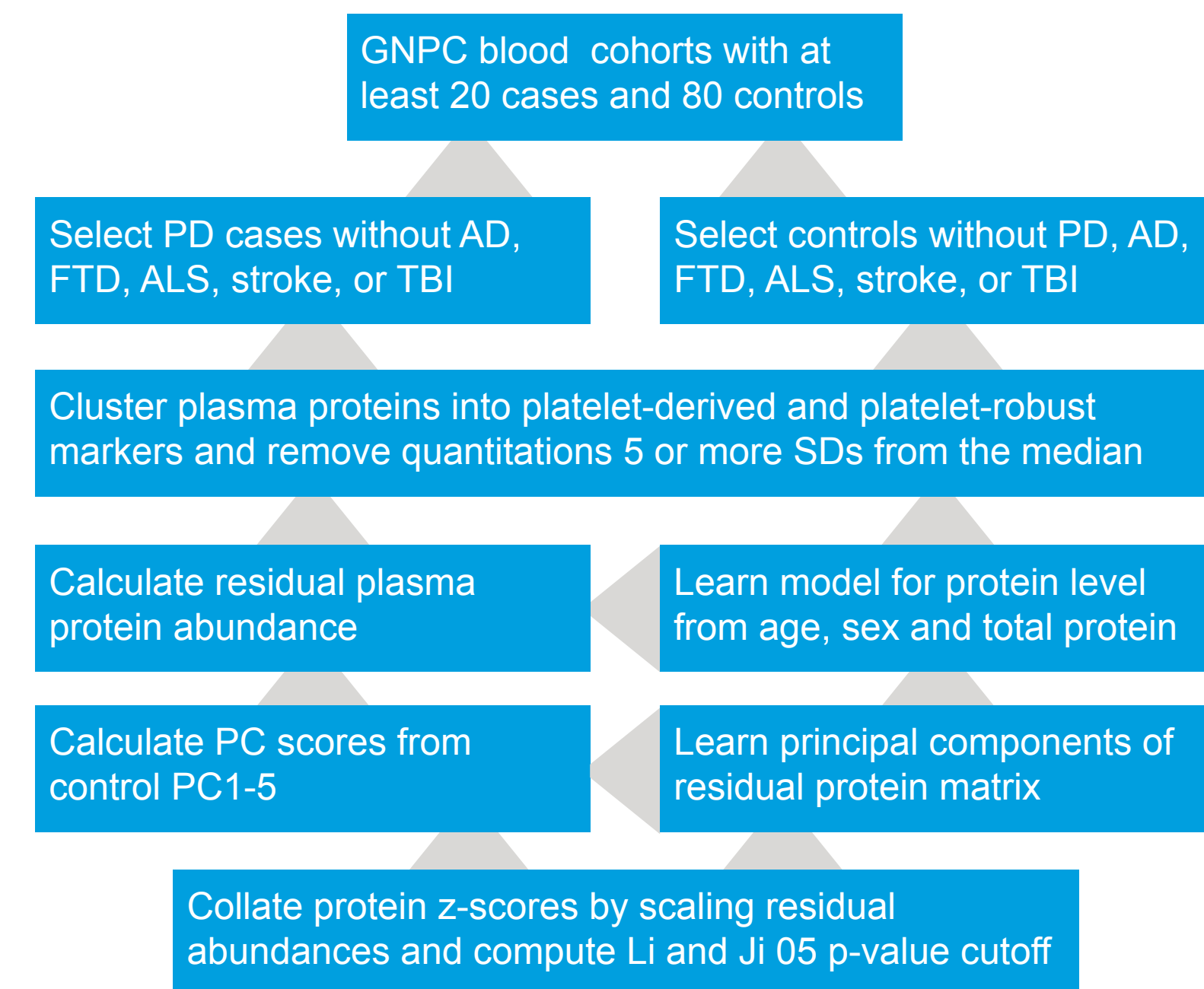


Figure 1: Flow chart for ascertaining GNPC participants with PD and computing protein z-scores and PC scores from control samples from the same contributor

Figure 2: Visualization of diverging age and sex trends for four proteins for contributor C in GNPC. Quantitation is in relative log<sub>2</sub> units (NPX). Observations are binned by age decile.

## Methods

We reanalyzed GNPC blood cohorts with >20 PD cases and >80 HCs. PD participants with coincident diseases were excluded. Protein markers below the limit of detection in >12.5% of samples or correlated with cell type composition were excluded. Protein abundances were log transformed and standardized relative to HCs accounting for sex, age, plate, and average protein abundance for each of two clusters: apparent platelet-derived and platelet-robust sets of markers. We performed logistic regressions on case status against protein z score, sex, years of education, and the first 5 principal component scores. Cohorts with hits were meta-analyzed via weighted Z-test. Results were compared to Imam 2025, and the ratio of the observed to expected median chi-squared test statistic (proteomic inflation factor) was calculated for all cohorts.

**Disclosure:** E. Boyle, L. Voloboueva, K. Paumier, W. Hagstrom, F. Qureshi, and D. Brazel are employees of Octave Bioscience and may hold equity.

**References:** Imam, F., Saloner, R., Vogel, J.W. et al. The Global Neurodegeneration Proteomics Consortium: biomarker and drug target discovery for common neurodegenerative diseases and aging. *Nat Med* 31, 2556–2566 (2025).  
Li, J., Ji, L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* 95, 221–227 (2005).

## Findings

- Extreme and variable p-value inflation** | Proteomic inflation factors ranged from 1.01 to 26.1 in PD, and up to 34.5 in AD.
- Diverse age and sex trends** | Blood protein levels not uncommonly exhibited differential age trajectories for men and women
- Comorbidity of PD participants** | 21% (200 of 946) PD plasma participants were diagnosed with another neurodegenerative condition

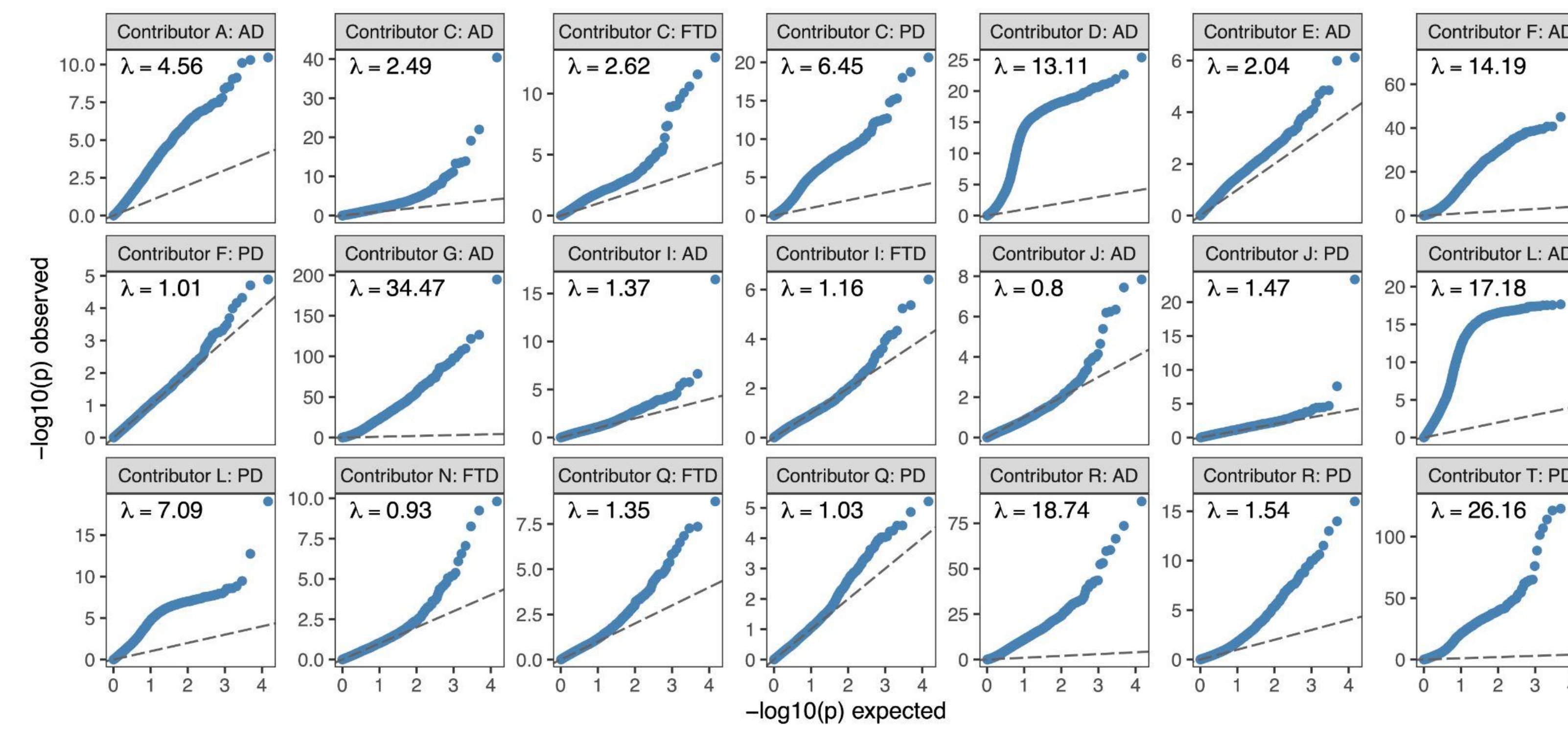


Figure 3: Biomarker quantile-quantile plots for Global Neurodegeneration Proteomics Consortium HDS v1.0 Somalomic panels.

- Proteomic inflation factors (left)** | The median proteomic inflation factor fell from 1.47 to 1.13, with the greatest gains in the most severely affected cohorts
- Biomarker consistency (right)** | Inconsistency in sign of proteins detected as hits in multiple cohorts dropped from 38% to 1%, and overlap of the top 100 meta-analysis hits with a published UK Biobank analysis increased from 5 to 16.

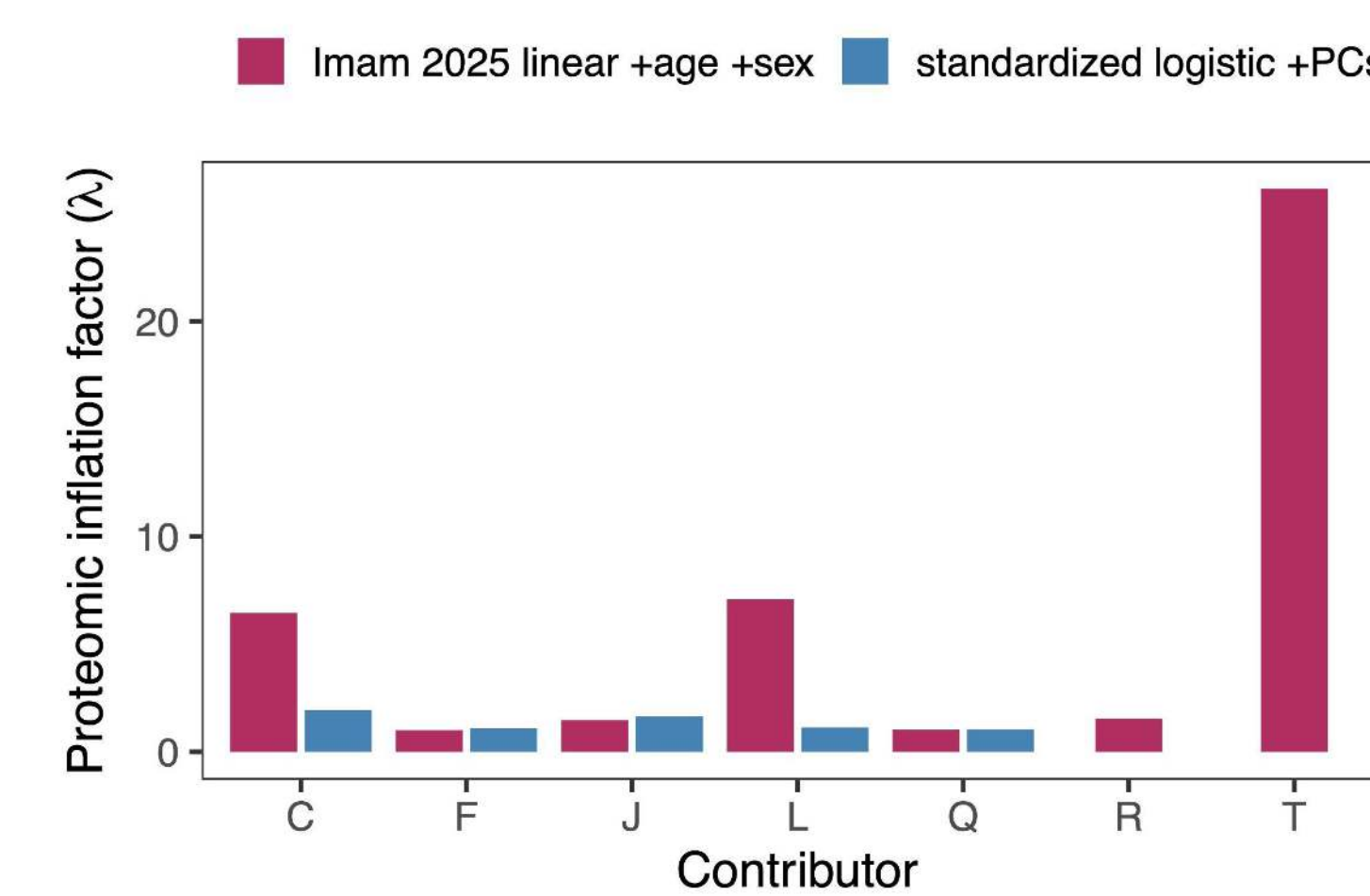


Figure 4: Mean chi-squared statistic for PD vs control associations for the Imam 2025 model and a logistic regression on protein z-scores plus 5 principal component scores as covariates

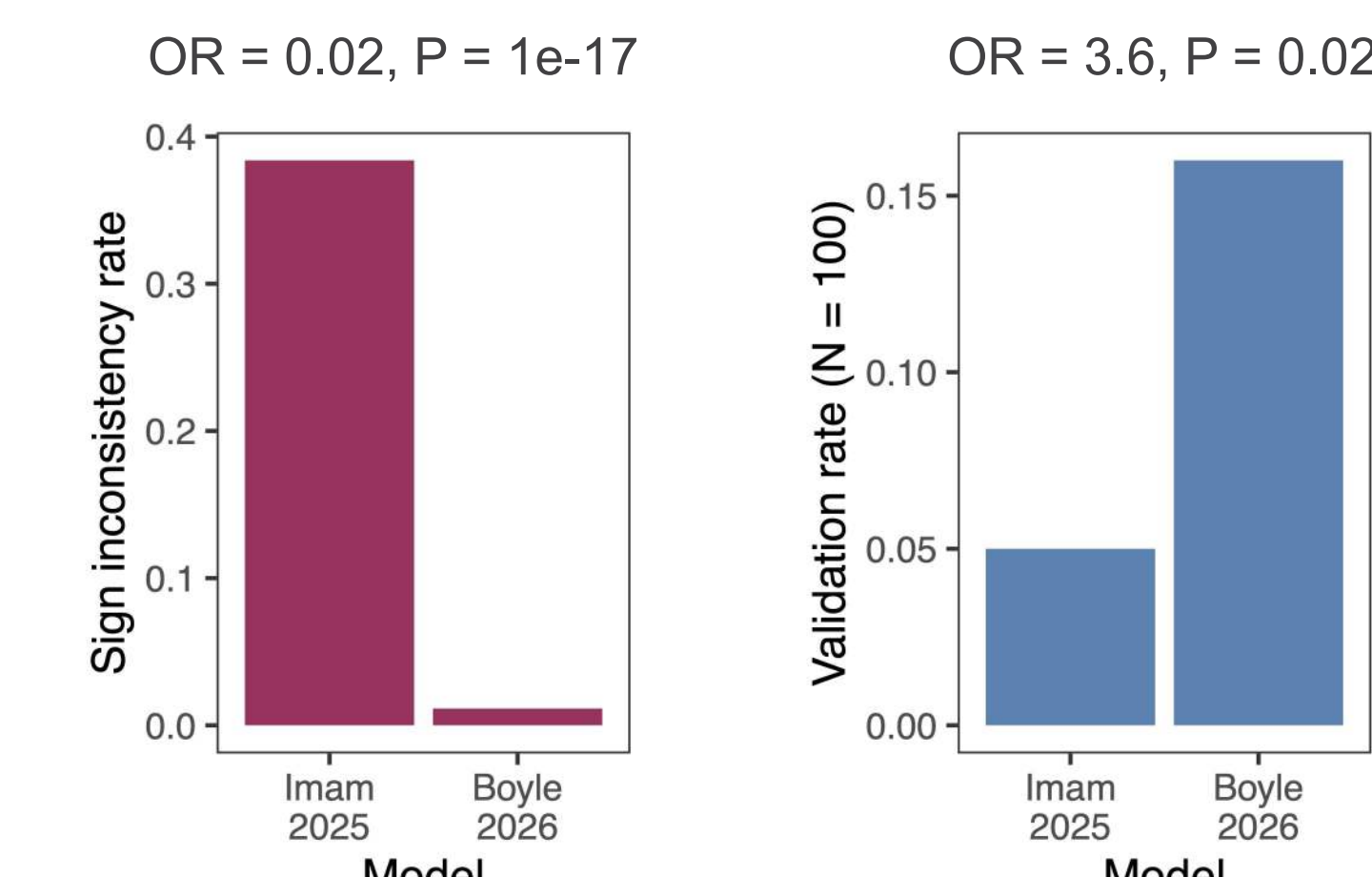


Figure 5: Performance of Imam 2025 and our model framework as shown by the rate of the sign change across studies among reported hits (left) and rate of nonzero importance in UKB (Garg 2024, right)

## Interpretation

- Statistical modeling choices have a profound impact on reported biomarkers: the number of Bonferroni significant meta-analyzed markers decreased from 2,251 (Imam 2025) to 28
- Stratified analysis within cohorts with abundant control samples substantially reduces variation in test association results and produces better behaved statistical distributions
- Our learnings in PD apply to other conditions in the GNPC including Alzheimer's disease, which also exhibited substantial change in estimated effect sizes after incorporating standardization and PC covariates (proteins below Li and Ji 05 cutoff decreased from 112 to 58)

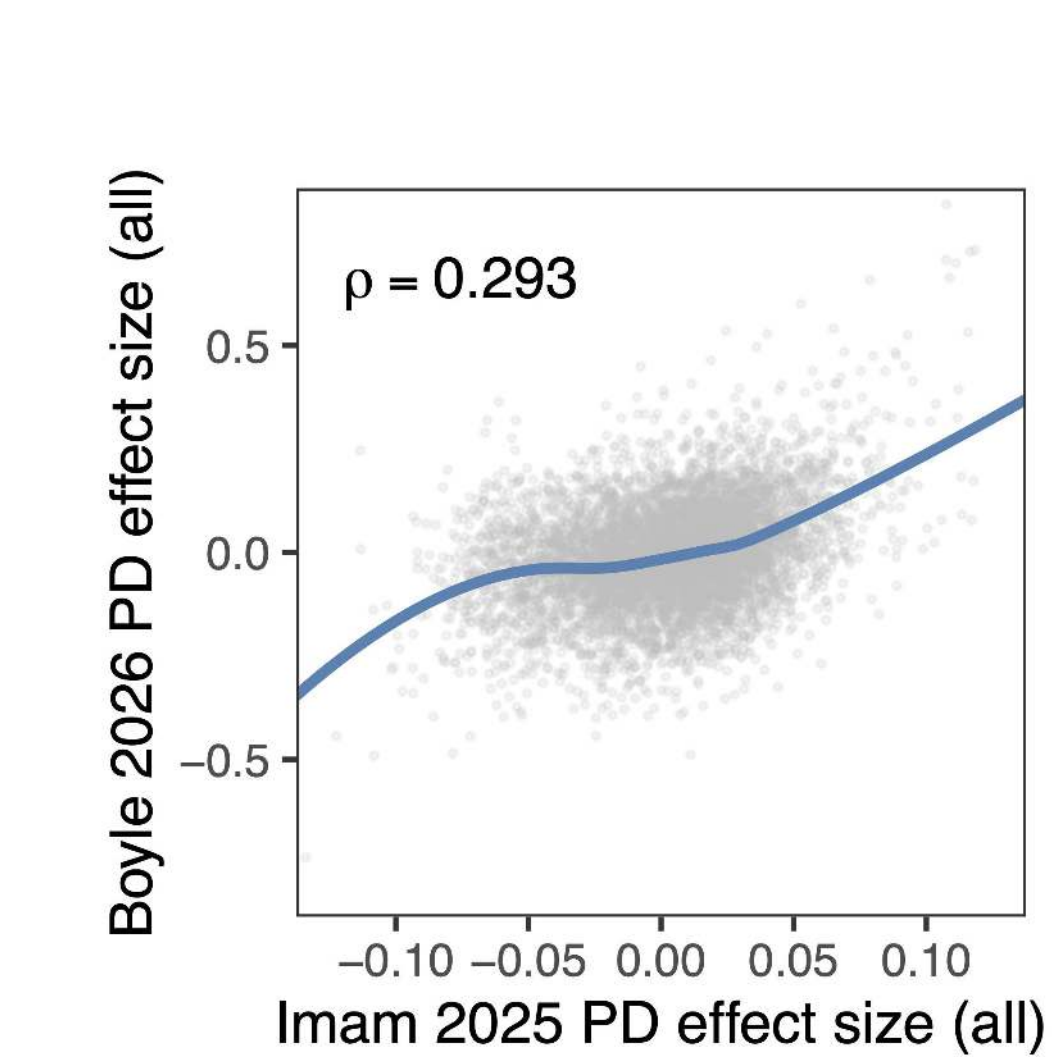


Figure 6: Meta-analysis agreement with Imam 2025 weighted average, GAM trend shown

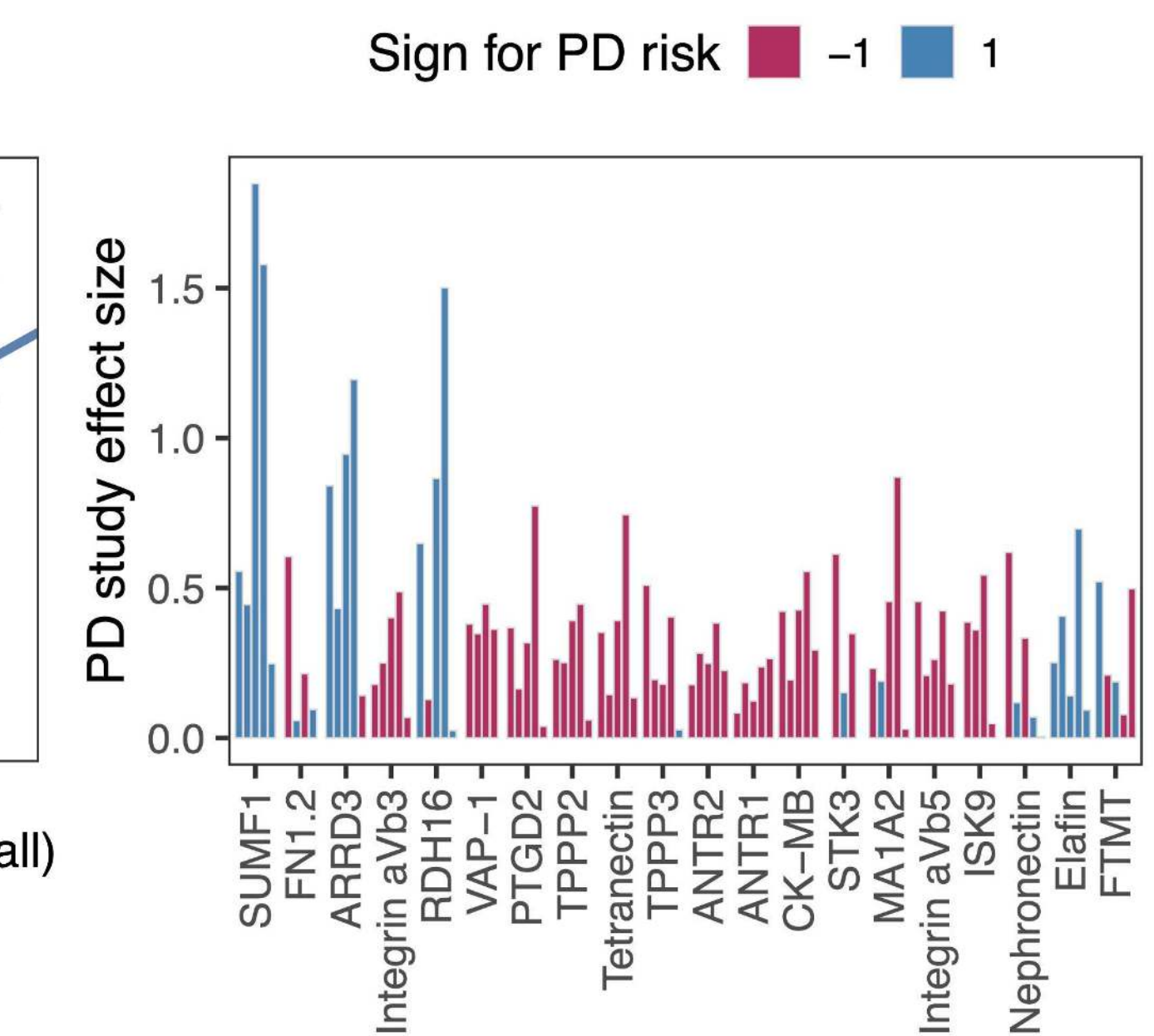


Figure 7: The effect sizes of the 20 markers most significantly associated with PD by meta-analysis across contributors

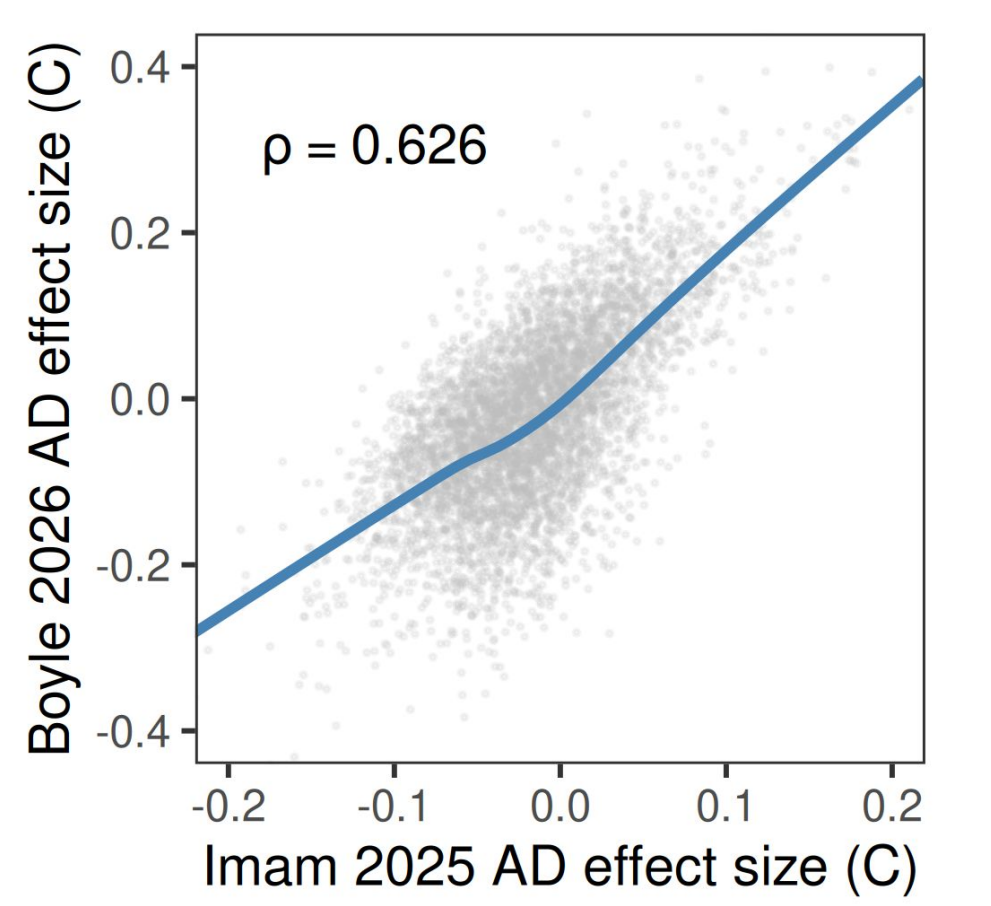


Figure 8: Agreement with Imam 2025 contributor C AD effect sizes. GAM trend shown

## Conclusions

We confirm that SUMF1, ITGAV, TPPP2, and PI3 (Elafin) are top plasma protein biomarkers discriminating PD from healthy controls featuring consistent direction of effect across all GNPC cohorts. Plasma protein exhibit diverse trajectories with respect to age sex, and principal components. Controlling for these factors radically changes the yield and replicability of biomarker associations. We conclude that large, multi-center biofluid protein biomarker datasets require careful consideration of matched control samples to prevent population stratification or matrix effects from dominating discovery based on differential protein abundance.

Data used in the preparation of this article were obtained on September 8, 2025 from the Global Neurodegeneration Proteomics Consortium (GNPC) Harmonized Data Superset Version 1 (<https://www.neuroproteome.org/v1-harmonized-data-set-hds>). While the investigators within the Global Neurodegeneration Proteomics Consortium contributed data to the Global Neurodegeneration Proteomics Consortium, they did not necessarily participate in the analysis or writing of this report