PRIDE repository (PXD058779) data analysis by Team 5

Introduction

The study titled "Apolipoprotein E abundance is elevated in the brains of individuals with Down syndrome-Alzheimer's disease" investigates how trisomy 21 (the genetic basis of Down syndrome, DS) alters the brain proteome and transcriptome in the context of Alzheimer's disease (AD). Individuals with DS are at high risk of developing early-onset AD (DSAD), largely due to the triplication of the APP gene on chromosome 21. However, the broader molecular consequences of trisomy 21 on AD pathology remain poorly understood.

Methodology

The researchers conducted a comprehensive proteomic and transcriptomic analysis of post-mortem frontal cortex tissue from individuals with Down syndrome-associated Alzheimer's disease (DSAD), early-onset Alzheimer's disease (EOAD), and healthy ageing (HA) controls. Proteomic profiling was performed using quantitative label-free mass spectrometry, identifying 2,855 proteins with at least two unique peptides. Differential protein abundance across groups was assessed using Progenesis QI for Proteomics, with peptide identification based on the UniProt Human Reference Proteome (2022) and stringent filtering criteria. Statistical analyses were carried out using SPSS Statistics 29 (IBM) for ANOVA and Prism 10 (GraphPad). The proteomics dataset was deposited in the PRIDE repository (PXD058779). To complement the proteomic data, single-nuclei RNA sequencing (snRNA-seq) and quantitative PCR were performed on the same samples to identify cell-type-specific gene expression patterns, enabling a multi-layered understanding of molecular alterations in DSAD.

Key Proteomics Findings from this study

- 2,855 proteins with two or more unique peptides were identified.
- Hsa21-encoded proteins: eight proteins encoded on chromosome 21 were significantly elevated in DSAD compared to EOAD and HA - APP, S100B, SYNJ1, CSTB, PDXK, CBR1, CBR3, and PFKL.
- However, not all Hsa21 genes showed increased protein levels, indicating post-transcriptional regulation.
- Non-Hsa21 proteins: thirty-one non-chromosome 21 proteins were significantly upregulated in DSAD.
- Notably, APOE (apolipoprotein E), a key AD risk factor, was significantly elevated in DSAD compared to both EOAD and HA. Other upregulated proteins included SSB, PTN, WASL, and NDE1.
- Downregulated proteins in DSAD: nineteen non-Hsa21 proteins were significantly decreased in DSAD, including KRT77, KIF20B, and ALDH1A1.
- APOE correlations: APOE abundance correlated positively with APP and its cleavage products (APP-CTFs and amyloid-β40), suggesting a link between trisomy 21-driven APP overexpression and APOE upregulation.
- This study highlights that trisomy 21 leads to widespread proteomic changes beyond chromosome 21, with APOE emerging as a key protein elevated in DSAD. These findings underscore the need to consider DS-specific molecular profiles when developing AD therapies.

Key methods and results by Team 5

- The team analysed four .mzid files from the PRIDE repository, focusing on preprocessing of PSM (Peptide Spectrum Match) data, including assessment of decoy hits, rank, and confidence score distributions.
- No decoy-labeled hits were found; however, a subset of PSMs with zero confidence scores was identified. These were associated with very short peptide sequences and, after comparison with high-confidence PSMs, were excluded due to unreliability.
- Following filtering, 72,527 peptides and 5,907 proteins were identified. Analysis of the peptide-protein graph revealed 3,172 connected components (CCs), which were further filtered to retain only proteins connected to two or more peptides, in line with the methodology from the publication.
- To proceed further, we contacted the authors for their quantitative protein-level data. The normalized peptide level quantification file was further used for feature aggregation and analysis for protein level quantification.
- Differential expression (DE) analysis was performed using OmicsQ. It is a useful tool to provide statistical confidence for identifying significantly regulated proteins between experimental conditions.
- The PCA plot did not show sample clustering based on the group conditions, and the main proteins of interest were not significantly dysregulated in the differential results.

Results comparison between the team and the publication

Our comparison revealed several key differences, highlighting the underlying challenge of reproducibility in proteomics. From the initial PSM data, we identified 5,684 proteins, whereas the original study reported only 2,855. Although both analyses used the same quantitative dataset—shared by the original authors—our pipeline, which employed OmicsQ for DE analysis, did not highlight APOE as significantly upregulated in both disease conditions, contrary to the original findings. While the number of peptides mapped to APOE matched exactly (64), likely due to the shared input files, our PCA results failed to replicate the study's trends and appeared uninformative. This discrepancy, despite the use of identical data, emphasizes how variations in analytical tools and workflows can substantially influence results. It highlights the need for transparent and reproducible pipelines in proteomics research, any strategies for template workflows,