

Counterfactual Analysis in Conditional Variational Autoencoder for disentangling sex effects in high-dimensional protein data

Chiara Zambon^{1,2}, Sara Calhas^{1,3}, Natalia Vilor Tejedor^{1,4}

¹Barcelonaβeta Brain Research Center (BBRC), Pasqual Maragall Foundation, Barcelona, Spain. ² Politecnico di Torino, International University, Torino, Italy

³Centre for Preventive Neurology, Queen Mary University of London. ⁴Institute for Risk Assessment Sciences (IRAS), Utrecht University, Utrecht, Netherlands.

Background

Dementia is projected to affect more than **139 million people worldwide by 2050**, highlighting the need for scalable approaches for early detection and biological stratification [1].

Cerebrospinal fluid (CSF) proteomics provides a **rich molecular characterization of Alzheimer's disease (AD)**, capturing thousands of proteins simultaneously. However, the **high dimensionality of proteomic data and the presence of correlated biological variables complicate the identification of meaningful disease mechanisms** [2,3,4,5,6].

Biological sex is a major source of heterogeneity in AD risk, progression, and biomarker profiles. Despite its biological relevance, sex is often treated solely as a confounding factor rather than as a source of meaningful variation [3].

Aim

To develop a **sex-conditioned Conditional Variational Autoencoder (CVAE)** capable of estimating individual-level counterfactual sex effects in CSF proteomic data while accounting for major biological confounders

Methods

Cohort and Data preprocessing

We analyzed **CSF proteomic data** from the Alzheimer's Disease Neuroimaging Initiative (ADNI), comprising **737 individuals** and **7,032 SomaScan-measured proteins**.

Analytical steps

- Data were randomly divided into training (n=515), validation (n=111), and test (n=111) sets.
- Unwanted biological variation:** we evaluated candidate confounders using multivariable models, partial R² statistics, and false discovery rate correction.
- Protein measurements were adjusted using linear residualization combined with Generalized Additive Models, allowing flexible modeling of non-linear and heavy-tailed protein distributions [2,10,11].

Framework

Conditional Variational Autoencoder

A Conditional Variational Autoencoder (CVAE) was trained to model **CSF proteomic profiles** while explicitly **conditioning the latent representation on biological sex** [15] (Figure 1).

- CVAE learns a continuous **latent representation capturing biological variability** not directly explained by biological sex.
- The model was optimized by maximizing the Evidence Lower Bound, balancing reconstruction accuracy and latent-space regularization [14,15] (Eq. 1, Figure 1)
- The resulting latent space enables the generation of both **factual and counterfactual proteomic profiles** for each individual.

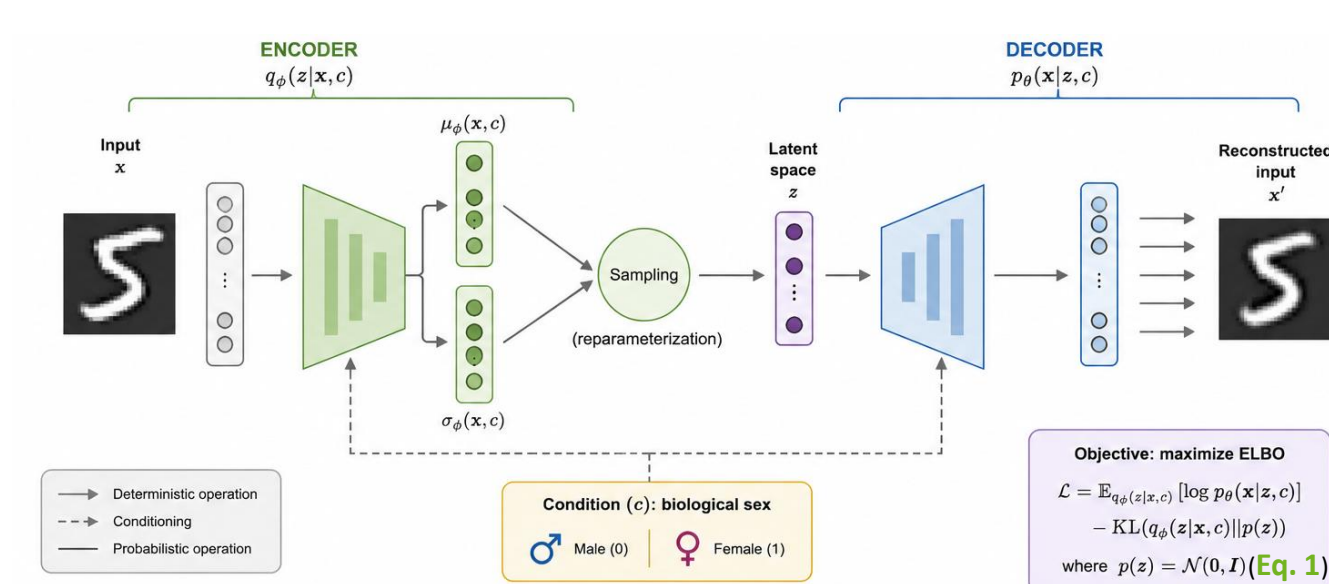


Figure 1. Conditional Variational Autoencoder (CVAE) architecture. The encoder maps the input proteomic profile (x), conditioned on biological sex (c), into a latent representation (z).

Counterfactual Framework

For each participant, the **trained CVAE generates:**

- A **factual reconstruction**, preserving the observed sex label (Figure 2a);
- A **counterfactual reconstruction**, obtained by intervening on the sex conditioning variable while preserving all other latent characteristics (Figure 2b).

The **individual counterfactual effect** was defined as the difference between factual and counterfactual protein profiles:

$$\Delta X = X_{\text{counterfactual}} - X_{\text{factual}} \quad (\text{Eq. 2})$$

This quantity estimates the **contribution of biological sex to each protein at the subject level** [16,17].

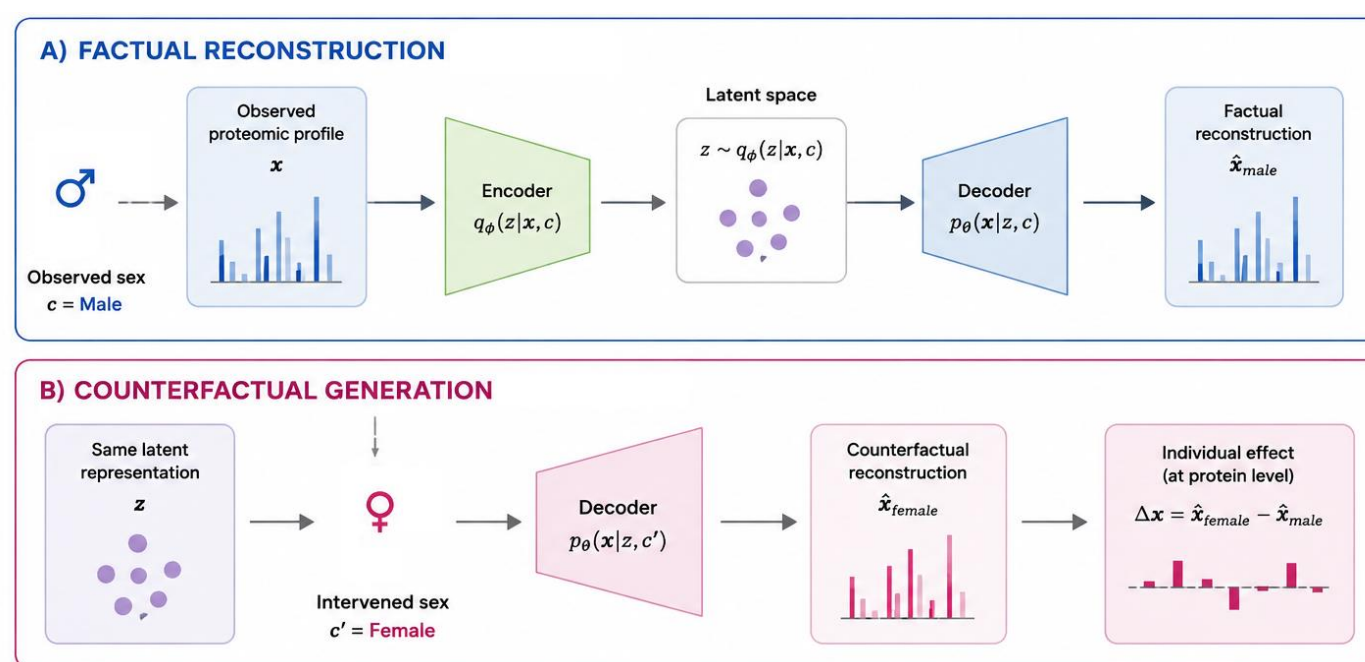


Figure 2. Counterfactual generation framework using a Conditional Variational Autoencoder (CVAE). A) Factual reconstruction: the encoder maps the observed proteomic profile (x), conditioned on the observed biological sex (c), into a latent representation (z), which is subsequently decoded to reconstruct the original profile. B) Counterfactual generation: the same latent representation (z) is combined with the intervened sex condition (c') and decoded to generate a counterfactual profile. The individual counterfactual effect is defined as the difference between the counterfactual and factual reconstructions, providing an estimate of the contribution of biological sex to each protein while preserving all other latent characteristics of the individual.

Results

Latent Space Structure

- The Hopkins statistic and multiple clustering algorithms (K-means, Gaussian Mixture Models, and Spectral Clustering) **failed to identify robust sex-specific clusters** (Figure 3).
- Sex-related molecular variation seems to be predominantly continuous rather than discretely separated.**

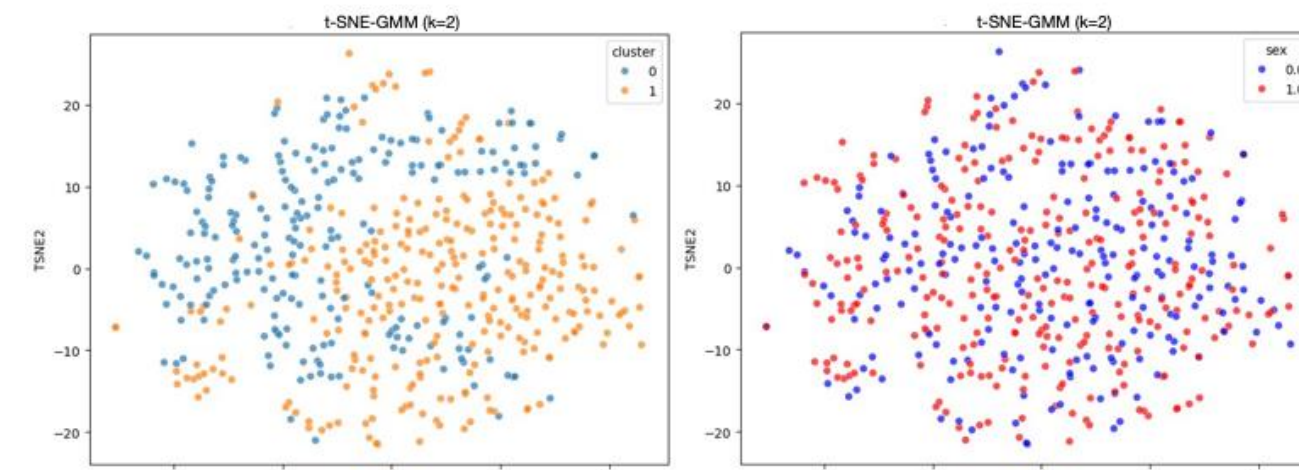


Figure 3. t-SNE colored by sex + GMM overlay: Latent space shows mixed sex distribution, indicating continuous rather than discrete biological variation.

Counterfactual Protein Modulation

- Subject-level counterfactual interventions **revealed widespread protein changes associated with biological sex.**
- Paired statistical testing of counterfactual differences (ΔX)** identified significantly modulated proteins (Figure 4) showing that the model captured substantial sex-dependent molecular variation.

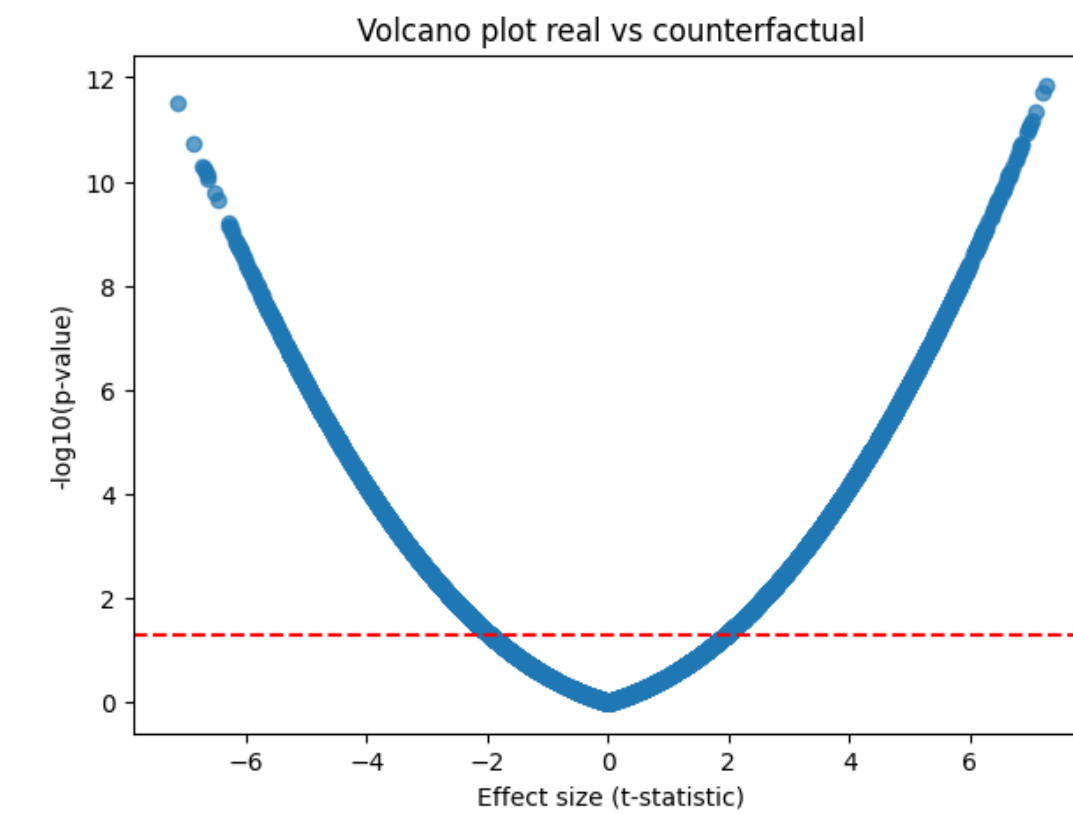


Figure 4. Volcano plot of counterfactual protein effects. Each point represents a protein, plotted according to its counterfactual effect size and statistical significance obtained from paired comparisons between factual and counterfactual protein profiles. The symmetric distribution of significant effects around zero suggests that biological sex is associated with both increases and decreases in protein abundance across the CSF proteome.

Identification of Sex-Sensitive Proteins

Ranking proteins according to the magnitude of their counterfactual effect identified **several biologically relevant candidates** (Table 1).

Top-ranked proteins were involved in:

- endocrine and reproductive regulation** (LEP, CGA, FSHB),
- cardiovascular function** (MYL4), and
- neurodevelopmental or synaptic signaling** processes (LINGO1, DCLK3, SLITRK3, SEMA3E).

Many of these proteins have previously been associated with sex-dependent biological pathways and AD-related mechanisms [18-21].

Rank	Gene	Biological Function (Main Pathway)	[Aproteins] (mean abs. effect)	t-stat	p-value
1	MYL4	Cardiovascular function	0.097119	3.815160	7.08 × 10 ⁻⁴
2	LEP	Endocrine regulation (adiposity)	0.092847	-3.777546	7.99 × 10 ⁻⁴
3	LEP	Endocrine regulation (adiposity)	0.090023	-4.042520	3.54 × 10 ⁻⁴
4	CGA/FSHB	Endocrine / reproductive regulation	0.088737	-3.954550	4.58 × 10 ⁻⁴
5	LINGO1	Neurodevelopment / axon guidance	0.078129	-3.387160	2.57 × 10 ⁻³
6	DCLK3	Neurodevelopment / synaptic signaling	0.077931	3.814508	7.09 × 10 ⁻⁴
7	SLITRK3	Synaptic signaling	0.077331	-3.055416	6.30 × 10 ⁻³
8	SEMA3E	Neurodevelopment / axon guidance	0.076452	-3.756969	8.49 × 10 ⁻⁴
9	PEAR1	Axon guidance / neuronal adhesion	0.075435	-4.207347	2.04 × 10 ⁻⁴
10	SEMA7A	Synaptic signaling / cell signaling	0.074994	-3.425728	2.29 × 10 ⁻³

Cardiovascular Endocrine / Reproductive Neurodevelopment Synaptic signaling / Axon guidance

Table 1. Top 10 proteins most sensitive to sex according to counterfactual analysis. Entrez Gene Symbol (gene) indicates the gene that encodes the corresponding protein. Proteins are ranked by the absolute mean counterfactual effect. [Aproteins] denotes the absolute mean value of the counterfactual difference across subjects. The t-statistic and p-value were obtained from paired comparisons between factual and counterfactual protein values. Higher statistic values indicate greater sensitivity of protein abundance to the sex intervention.

Sex-specific and shared molecular signatures

- Sex-stratified counterfactual analyses revealed substantial asymmetry in protein responses.**
- Several proteins exhibited strong modulation in only one sex, whereas others demonstrated shared responses across both sexes (Figure 5).

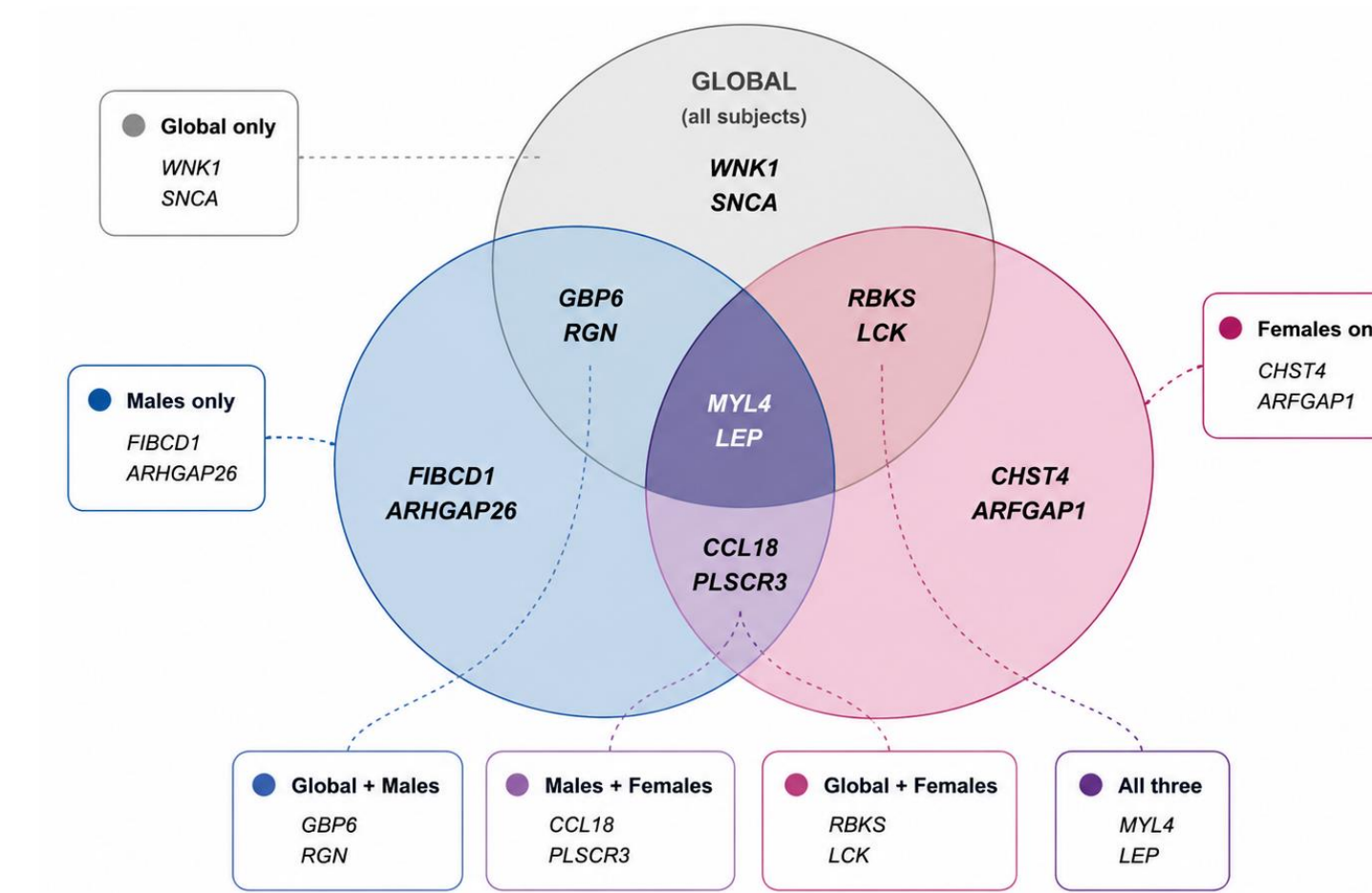


Figure 5. Representative proteins identified across global, male-specific, and female-specific counterfactual analyses. Filled circles indicate the analyses in which each protein category was significant (FDR < 0.05). Representative genes are shown for each overlap group. Proteins classified as "All three" (MYL4 and LEP) were consistently detected across global, male-specific, and female-specific analyses, suggesting robust sex-associated molecular effects.

Comparison with Conventional Analysis

To assess **whether counterfactual effects reflected conventional sex differences**, we compared CVAE-derived rankings with rankings obtained from an observed-space two-sample t-test.

- The resulting Spearman correlation was modest (ρ = 0.29), indicating limited agreement between the two approaches.**

This suggests that the **counterfactual framework captures nonlinear and multivariate sex-associated effects that remain largely undetected by traditional univariate analyses.**

Conclusions

- A **sex-conditioned CVAE successfully learned latent representations that disentangle sex-associated effects from broader biological variability in CSF proteomic data.**
- **Counterfactual inference enabled the estimation of individual-level sex effects without relying on conventional group comparisons.**
- **Sex-related proteomic variation appeared continuous rather than organized into discrete male and female clusters.**
- **Limited concordance with traditional statistical analyses suggests that counterfactual modeling captures additional nonlinear and multivariate biological signals.**
- **These findings support counterfactual deep learning as a promising approach for investigating biological heterogeneity and biomarker discovery in Alzheimer's disease.**

→ References

- [1] Alzheimer's Disease International. *World Alzheimer Report* (2025).
- [2] Li, Z. et al. *Briefings in Bioinformatics* (2024).
- [3] Wiesenhagen, K. E. J. et al. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* (2022).
- [4] Jager, K. J. et al. *Kidney International* (2008).
- [5] Seo, D. et al. *Nature Aging* (2025).
- [6] Vromen, E. M. et al. *Alzheimer's & Dementia* (2025).
- [7] Montgomery, D. C. et al. *Introduction to Linear Regression Analysis* (2021).
- [8] Storey, J. D. et al. *Journal of the Royal Statistical Society Series B* (2002); Storey, J. D. et al. *Proceedings of the National Academy of Sciences* (2003).
- [9] Joanes, D. N. et al. *Journal of the Royal Statistical Society Series D* (1998).
- [10] Leek, J. T. et al. *PLoS Genetics* (2007).
- [11] Hastie, T. et al. *Statistical Science* (1986).
- [12] Abdelali, K. et al. *Journal of Lightwave Technology* (2022).
- [13] Mierny, L. D. et al. *Archives of Computational Methods in Engineering* (2025).
- [14] Kingma, D. P. et al. *Auto-Encoding Variational Bayes* (2013).
- [15] Sohn, K. et al. *Advances in Neural Information Processing Systems* (2015).
- [16] Kim, Y. et al. *PLoS Digital Health* (2022).
- [17] Wu, S. et al. *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (2024).
- [18] Christen, T. et al. *Peptides* (2018).
- [19] Santi, D. et al. *Journal of Clinical Medicine* (2020).
- [20] Wang, T. Y. et al. *Scientific Reports* (2019).
- [21] Belonwu, S. A. et al. *Molecular Neurobiology* (2022).

→ Acknowledgements

NV-T was supported by the Spanish Research Agency MICIU/AEI/10.13039/501100011033 (grant RYC2022-038136-I cofunded by the European Union FSE+, and grant PID2022-143106OA-I00 cofunded by the European Union FEDER). Additionally, NV-T was supported by William H. Gates Sr. Fellowship Cohort from the Alzheimer's Disease Data Initiative, and the Alzheimer's Association, AD Strategic Fund (Project #VQID-UMD-26-1514428). Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.



Contact
Chiara Zambon
chiarazambon2001@gmail.com
github.com/GeneticNeuroStats

www.barcelonabeta.org

www.uu.nl