SUPPLEMENTARY FIGURES AND METHODS

JACS Au

The Human Blood *N*-Glycome: Unraveling Disease Glycosylation Patterns

Tamas Pongracz[1], Oleg A. Maiboroda[1], Manfred Wuhrer[1,*]

[1]Center for Proteomics and Metabolomics, Leiden University Medical Center, Leiden, Netherlands

[*]**Corresponding author:** Manfred Wuhrer; m.wuhrer@lumc.nl

---

This section provides a detailed summary of the re-analysis methodology applied in the current report, which is structured into three main logical segments. Each segment uses a specific set of methods and software tools, namely: a) A Cross-Study Overview of Measured Glyco-Traits, b) Examining the Impact of Demographic Factors on Blood Protein Glycosylation, and c) Analysis of Effect Sizes in Glycosylation Traits Across Nine Clinical Studies. The entire analysis was conducted using the R statistical environment (version 4.3.2).

A. A Cross-Study Overview of Measured Glyco-Traits: This part builds on the UpSet plot tool for a comprehensive visualization of overlaps and unique elements across multiple datasets. The UpSet plot is chosen for its clarity and scalability, especially when comparing multiple datasets. Unlike traditional Venn diagrams that focus on the sets, the UpSet plot emphasizes the intersections of these sets. It arranges the sets along one axis—typically horizontal—and utilizes bars to represent the size of each set. The innovative aspect of the UpSet plot is its representation of intersections through a matrix composed of dots and lines, diverging from the traditional method of using overlapping circles. Each dot represents the intersection between two or more sets, with lines drawing attention to these connections. Next to this matrix, a vertical bar graph presents the size of these intersections, enabling an easy comparison of shared elements among different combinations of sets. For this report, we utilized the UpSetR package (version 1.4.0), adhering closely to its manual guidelines.

Jake R Conway, Alexander Lex, Nils Gehlenborg UpSetR: An R Package for the Visualization of Intersecting Sets and their Properties; doi: https://doi.org/10.1093/bioinformatics/btx364

B. Examining the Impact of Demographic Factors on Blood Protein Glycosylation: Here, we employed the methodology described in a previous report (reference 36 in the main manuscript), which involved aggregating control samples from the studies mentioned. To facilitate comparison across different cohorts, glycosylation traits were re-calculated uniformly **(Table S2)**. The aggregated control sample dataset comprised 1776 observations.

Our analytical strategy was framed within the context of generalized linear models. We utilized standard linear regression to analyze continuous outcomes (such as age) and logistic regression for binary outcomes (such as sex). Additionally, we enriched our analysis by incorporating linear models with interaction terms, which underscored sex-specific and age-related variations in glycosylation, as well as nuanced differences in age associations across sexes. For each glycosylation trait, we built a distinct model, and the most significant associations between glycosylation traits and demographic factors were presented in a forest plot format. The results of these analyses are concisely summarized in **Table S3**. The analysis was executed using basic R syntax for linear and logistic regression models, with the tidyverse package (version 2.0.0) facilitating the organization of the outcomes.

C. Analysis of Effect Sizes in Glycosylation Traits Across Nine Clinical Studies: To quantify the extent of these changes across the studies we used effect sizes as a uniform measure. Specifically, we used

Cohen’s d as the metric for effect size. Cohen’s d is a statistical measure widely used to quantify the effect size, or the magnitude of difference, between two groups. It calculates the difference between two means (such as in treatment vs. control groups) and expresses this difference in standard deviation units. This standardized approach is what makes Cohen’s d exceptionally valuable for comparing effects across different studies or experiments, especially when these studies vary in their size, scale or measurement units. Please note that confidence intervals of the estimates offer insight into statistical significance: if the CI does not encompass zero, the effect can be considered statistically significant at the chosen confidence level (95%).

Within the R statistical environment, Cohen’s d was computed as follows:

1. Calculation Difference in Means: the difference between the means of the two compared groups subtracting the mean of one group from the mean of the other. The order does not change the magnitude of Cohen's d, just the sign, which indicates directionality (which group has the higher mean).

2. Calculation of the Pooled Standard Deviation: To standardize the mean difference, we calculate the pooled standard deviation, a weighted average that merges the standard deviations of both groups. The calculation accounts for the size and variance of each group, returning a singular measure of variability that represents both groups.

3. Calculating Cohen's d: Cohen's d is calculated by dividing the difference in means by the pooled standard deviation.

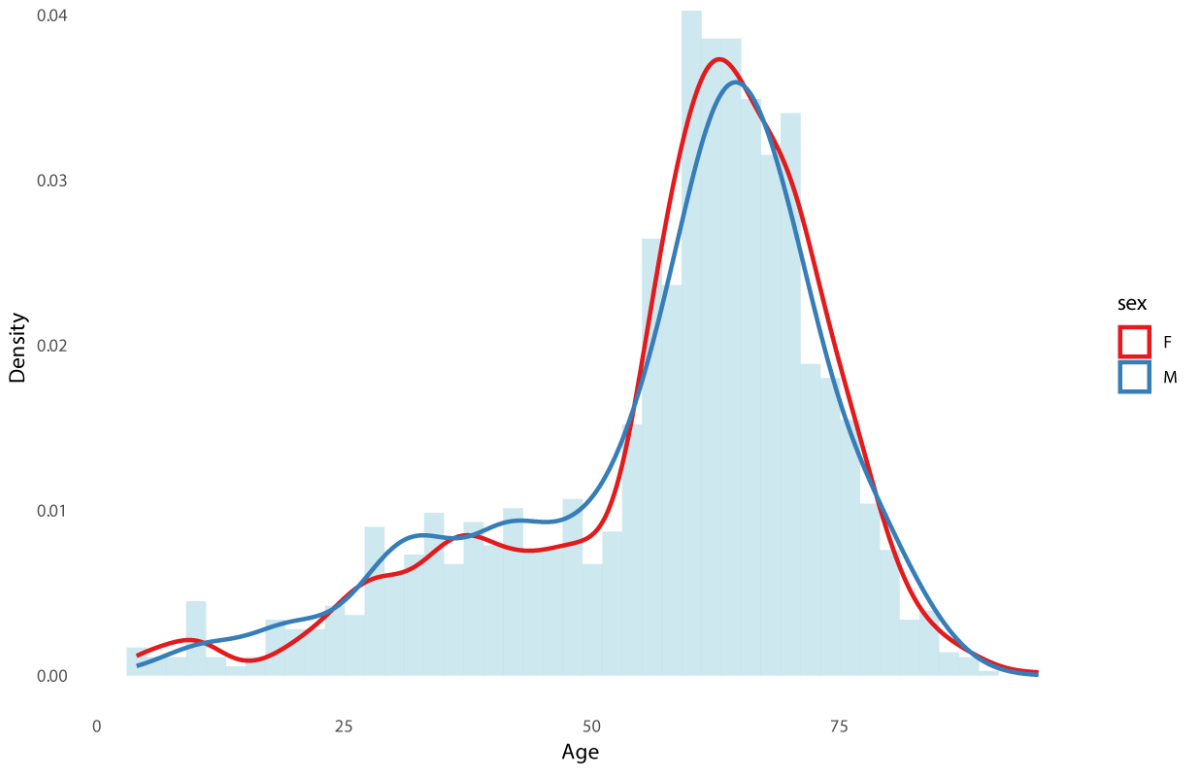

Figure S1. Density plot illustrating a well-balanced age distribution across both sex strata.

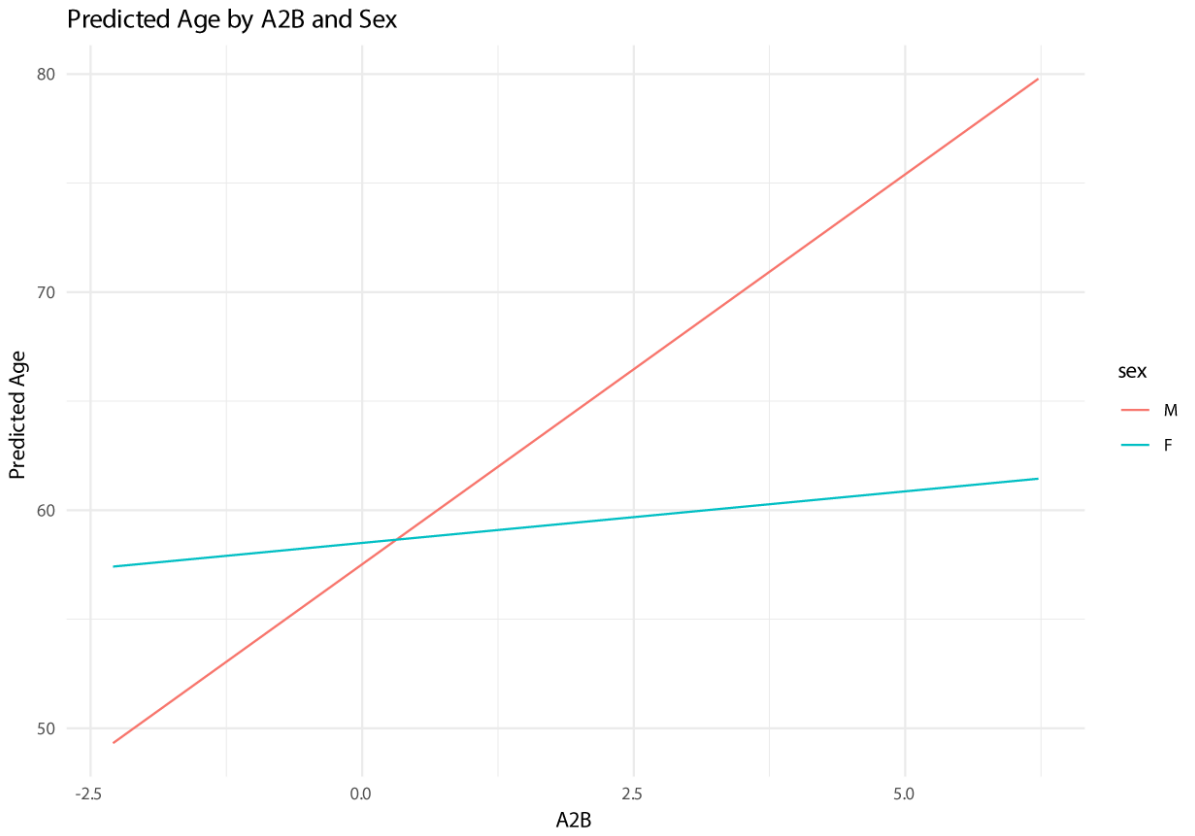

Figure S2. Sex influences the predictive power of the exemplary glycosylation trait A2B (shown are its standardized values) on age, illustrating the steeper increase of A2B in males (M; in red) relative to females (F; in blue) upon aging.