



## Article

# Concerted Regulation of Glycosylation Factors Sustains Tissue Identity and Function

Daniel Sobral <sup>1,2,\*</sup> , Rita Francisco <sup>1,2,3</sup> , Laura Duro <sup>1,2</sup>, Paula Alexandra Videira <sup>1,2,3,\*</sup> and Ana Rita Grosso <sup>1,2,\*</sup>

<sup>1</sup> Associate Laboratory i4HB—Institute for Health and Bioeconomy, NOVA School of Science and Technology, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal; rab.francisco@campus.fct.unl.pt (R.F.); l.duro@campus.fct.unl.pt (L.D.)

<sup>2</sup> UCIBIO—Applied Molecular Biosciences Unit, Department of Life Sciences, NOVA School of Science and Technology, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal

<sup>3</sup> CDG & Allies—Professionals and Patient Associations International Network (CDG & Allies—PPAIN), Life Sciences Department, NOVA School of Science and Technology (FCT NOVA), Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal

\* Correspondence: dv.sobral@fct.unl.pt (D.S.); p.videira@fct.unl.pt (P.A.V.); argrosso@fct.unl.pt (A.R.G.)

**Abstract:** Glycosylation is a fundamental cellular process affecting human development and health. Complex machinery establishes the glycan structures whose heterogeneity provides greater structural diversity than other post-translational modifications. Although known to present spatial and temporal diversity, the evolution of glycosylation and its role at the tissue-specific level is poorly understood. In this study, we combined genome and transcriptome profiles of healthy and diseased tissues to uncover novel insights into the complex role of glycosylation in humans. We constructed a catalogue of human glycosylation factors, including transferases, hydrolases and other genes directly involved in glycosylation. These were categorized as involved in N-, O- and lipid-linked glycosylation, glypiation, and glycosaminoglycan synthesis. Our data showed that these glycosylation factors constitute an ancient family of genes, where evolutionary constraints suppressed large gene duplications, except for genes involved in O-linked and lipid glycosylation. The transcriptome profiles of 30 healthy human tissues revealed tissue-specific expression patterns preserved across mammals. In addition, clusters of tightly co-expressed genes suggest a glycosylation code underlying tissue identity. Interestingly, several glycosylation factors showed tissue-specific profiles varying with age, suggesting a role in ageing-related disorders. In cancer, our analysis revealed that glycosylation factors are highly perturbed, at the genome and transcriptome levels, with a strong predominance of copy number alterations. Moreover, glycosylation factor dysregulation was associated with distinct cellular compositions of the tumor microenvironment, reinforcing the impact of glycosylation in modulating the immune system. Overall, this work provides genome-wide evidence that the glycosylation machinery is tightly regulated in healthy tissues and impaired in ageing and tumorigenesis, unveiling novel potential roles as prognostic biomarkers or therapeutic targets.

**Keywords:** glycosylation machinery; genomics; transcriptomics; healthy tissues; cancer



**Citation:** Sobral, D.; Francisco, R.; Duro, L.; Videira, P.A.; Grosso, A.R. Concerted Regulation of Glycosylation Factors Sustains Tissue Identity and Function. *Biomedicines* **2022**, *10*, 1805. <https://doi.org/10.3390/biomedicines10081805>

Academic Editor: Thomas Mohr

Received: 31 May 2022

Accepted: 21 July 2022

Published: 27 July 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Glycosylation is a complex multi-enzymatic process that includes assembling monosaccharides into glycans and transferring these onto proteins or lipids [1], with recent evidence suggesting small RNAs can also be glycosylated [2]. Most human proteins are thought to be glycosylated [3], influencing all aspects of cellular activity, including immunity [4,5], cell signaling [6], and cell adhesion [7]. Glycosylation is a ubiquitous post-translational modification present in all domains of life [8], enabling the generation of a large diversity of molecular structures, not directly bound by genetic information, much like the widely studied phosphorylation.

The biosynthesis of distinct glycoconjugates can be classified in different glycosylation pathways depending on the site of the glycosidic linkage and the glycan attached, namely: N-linked glycosylation, O-linked glycosylation; Glycosaminoglycans (GAG); Glypiation (GPI); and lipid glycosylation [9]. In humans, approximately hundreds of glycosylation factors are responsible for this post-translational modification process, including specific transferases (acetylglucosaminyltransferase, sialyltransferase, mannosyltransferase, N-acetylgalactosaminyltransferase, galactosyltransferase, fucosyltransferase, glucosyltransferase, xylosyltransferase), hydrolases (mannosidase, sialidase, glucosidase, fucosidase) and enzymes involved in monosaccharide precursor synthesis and transport [1]. Each glycosylation pathway is accomplished by a specific and distinct set of these enzymes, where some participate in different glycosylation pathways. A precise and tightly regulated action of the different enzymes and proteins are demanded to produce the glycoconjugate profiles specific and distinctive of each cell type and tissue and each development/functional stage [10,11].

Unsurprisingly, dysfunction of glycosylation factors leads to many pathologies [12], including the rare set of congenital disorders of glycosylation (CDG), cancer, metabolic and inflammatory diseases [12–14]. Notably, since aberrant glycosylation plays an important role in the immunosuppression of tumors and in the setting up of inflammatory diseases, they are used in the design of novel immunotherapies [15] or as biomarkers [16]. The most common perturbations in cancer are increased sialylation or fucosylation, O-glycan truncation, and N- and O-linked glycan branching [17]. Increased sialylation, in epithelial cancers, promotes immune evasion [18] and contributes to tumor progression and poor prognosis [19].

Our understanding the role of aberrant glycosylation in dysregulating organismal and cellular homeostasis is significantly progressing. Still, many steps remain largely uncharacterized, hindered by the inherent complexities of glycosylation. Large consortia have been generating a wide diversity of molecular profiles, in both healthy (e.g., GTEX [20], Protein Atlas [21]) and diseased tissues (e.g., TCGA [22]). However, these data have not yet been comprehensively examined to elucidate the role of glycosylation factors. In this study, we aimed at exploring available human molecular profiles, integrating them to obtain novel insights into the complex role of the glycosylation machinery. Our genome/transcriptome integrative analysis of a catalogue of 242 human glycosylation factors revealed that this ancient family of genes have expanded under some evolutionary constraints, capacitating glycosylation factors in defining tissue identity through a tissue-specific glycosylation code. Moreover, our work suggests that glycosylation-driven tumorigenesis is mediated mostly by large genomic amplifications and corresponding transcriptome alterations in glycosylation factors. Overall, our findings reveal novel unappreciated potential roles of glycosylation in human health.

## 2. Materials and Methods

### 2.1. Selection of Glycosylation-Factors and Gene Expression-Related Genes

A list of 242 glycosylation factors (Supplementary Materials Table S1) was obtained by collecting information from: GGDB glycosylation factor Database [23]; Reactome [24] (N-linked glycosylation; O-linked glycosylation); KEGG [25] (N-Glycan biosynthesis; Mucin and Mannose type O-glycan biosynthesis); and selecting human genes annotated in ENSEMBL [26] release 104 (May 2021) with the Gene Ontology term “protein glycosylation” (GO:0006486). The list was then manually curated to correct discontinued official gene symbols and remove genes encoding proteins that may undergo glycosylation but with no evidence for a direct role in the glycosylation process itself.

The genes encoding for regulators of gene expression were retrieved from different sources. Briefly, a list of 287 epigenetics factors was extracted from Boukas et al. [27]. A list of 653 transcription factors was extracted from Jolma et al. [28], including only those genes with a known motif, and excluding the ones only containing C2H2 zinc fingers. Splicing (326 genes), translation (154 genes), phosphorylation (672 genes) and

ubiquitination factors (619 genes) were selected based on the GO annotation in ENSEMBL (GO:0008380, GO:0006417, GO:0006468 and GO:0016567, respectively).

### 2.2. Phylogenetic Information and Expression

Ortholog and paralog information was downloaded from ENSEMBL Biomart. Information on the paralogous first duplication event was compressed to chordates (everything before Vertebrata), vertebrates (everything before Mammalia), and mammals (everything else except humans). To analyze phylogenetic conservation of tissue expression, multi-species expression data were obtained from Brawand et al. [29], comparing six organs across ten species. Namely, we used the normalized RPKM from constitutive exons of 1-to-1 orthologs. For each gene, we scaled the expression and obtained the standard deviation of the tissue with the minimum standard deviation between species.

### 2.3. Transcriptome Profiles of Healthy Tissues

The expression levels for healthy tissues were obtained from Protein Atlas and Genotype-Tissue Expression (GTEx) Project Protein atlas gene expression data were downloaded from [https://www.proteinatlas.org/download/rna\\_tissue\\_hpa.tsv.zip](https://www.proteinatlas.org/download/rna_tissue_hpa.tsv.zip) (downloaded 28 October 2019), and the normalized NX values were used for analyses. The GTEx data comprised transcriptome profiles of 28 healthy tissues and were obtained from the GTEx Portal on 28 October 2019. Only public metadata on age and sex were used, where age was classified in 10-year blocks. Moreover, transformed cells were excluded from the analysis. The normalized TPM values of GTEx samples were used for downstream analyses. We used the UMAP R package with GTEx expression data for UMAP analyses.

### 2.4. Tissue-Specificity and Classification Based on Transcriptome Profiles

To estimate the tissue predictive ability of glycosylation factors and remaining gene sets we used random forest models as implemented in the randomForest R package [30]. Briefly, samples of each tissue were divided into 75% train and 25% test. Then, 100 random forest models were used, each using the GTEx normalized expression data (TPMs) for 30 randomly selected genes (from the specific group of genes of interest). Finally, the Matthews Correlation Coefficient (MCC) was calculated based on the results of the classification of the test data.

To estimate tissue specificity, we used the TissueEnrich R package [31], with protein atlas data and also with GTEx data (Supplementary Materials Table S1). In brief, a Tissue Enriched gene has at least five-fold higher expression levels in a particular tissue compared to all other tissues; a Group Enriched gene has at least five-fold higher expression levels in a group of 2–7 tissues compared to all other tissues; a Tissue Enhanced gene does not fall in the previous categories but has at least five-fold higher expression levels in a particular tissue compared to the average expression of all other tissues.

To estimate age and sex-specific expression, for each tissue and gene, we fitted a linear model of the GTEx normalized TPM gene expression of a given tissue depending on age (in decades) and sex (except for tissues of the reproductive system with samples of only one specific sex, where only age was used). The  $p$ -values derived from each gene's linear model were corrected for multiple testing using the Benjamini–Hochberg procedure. Genes were age or sex-dependent if their coefficient in the linear model had an adjusted  $p$ -value less than 0.05.

### 2.5. Molecular Profiles of Tumor Samples

We downloaded public TCGA data from the GDC portal for all the 33 cancer types: ACC (adrenocortical carcinoma); BLCA (bladder urothelial carcinoma); BRCA (breast invasive carcinoma); CESC (cervical squamous cell carcinoma and endocervical adenocarcinoma); CHOL (cholangiocarcinoma); COAD (colon adenocarcinoma); DLBC (Lymphoid Neoplasm Diffuse Large B-cell Lymphoma); ESCA (Esophageal carcinoma); GBM (Glioblastoma multiforme); HNSC (head and neck squamous cell carcinoma); KICH (kidney Chro-

mophobe); KIRC (kidney renal clear cell carcinoma), KIRP (kidney renal papillary cell carcinoma); LAML (Acute Myeloid Leukemia); LGG (Brain Lower-Grade Glioma); LIHC (liver hepatocellular carcinoma); LUAD (lung adenocarcinoma); LUSC (lung squamous cell carcinoma), MESO (Mesothelioma); OV (Ovarian serous cystadenocarcinoma); PAAD (pancreatic adenocarcinoma); PCPG (pheochromocytoma and Paraganglioma); PRAD (prostate adenocarcinoma); READ (rectum adenocarcinoma); SARC (sarcoma); SKCM (Skin Cutaneous Melanoma); STAD (Stomach adenocarcinoma); TGCT (Testicular Germ Cell Tumors); THCA (thyroid carcinoma); THYM (Thymoma); UCEC (uterine Corpus Endometrial Carcinoma); UCS (Uterine Carcinosarcoma); and UVM (Uveal Melanoma). The data included: clinical metadata, namely the patient and follow up tables; somatic mutations as MuTect2 Variant Aggregation and Masking maf files; copy number as GISTIC copy number focal scores text files; and RNA-Seq read counts as HTSeq count tables.

### 2.6. Impact of Genetic Alterations

To assess the genetic alterations associated with pathogenic phenotypes we used the Clinvar variant summary data downloaded from <https://ftp.ncbi.nlm.nih.gov/pub/clinvar/> (release of 2 December 2019). Information on disease description was used with the wordcloud R package to produce a word cloud representing the relative frequency of terms in the Clinvar entries associated with glycosylation factors. The intolerance to loss of function mutations was obtained through the pLI score, that derives from the probability that a loss of function mutation occurs in a large cohort of healthy persons. The pLI scores were downloaded from ExAC ([ftp://ftp.broadinstitute.org/pub/ExAC\\_release/release0.3/functional\\_gene\\_constraint/fordist\\_cleaned\\_exac\\_r03\\_march16\\_z\\_pli\\_rec\\_null\\_data.txt](ftp://ftp.broadinstitute.org/pub/ExAC_release/release0.3/functional_gene_constraint/fordist_cleaned_exac_r03_march16_z_pli_rec_null_data.txt)), accessed on 4 April 2019. We also downloaded the list of cancer drivers from IntOGen release 1 February 2020. To identify genes under selective pressure in cancer, we applied the dNdScv method [32] using somatic mutation data from each of the TCGA cancer types. Briefly, dNdScv compares, for each gene, the observed tumor-specific mutations against a global background of mutations). A gene was considered under selection if it had significant evidence with dNdScv in at least one cancer type.

Influence of genomic perturbations in cancer patient survival was analyzed by Kaplan–Meier curve comparison using a log-rank test and a multivariate Cox proportional hazards analysis, as implemented in the survival R package [33].

### 2.7. Transcriptome Alterations in Tumor Samples

We used the HTSeq count tables to obtain normalized CPM values through TMM normalization in edgeR [34]. To infer gene expression perturbation in tumors, we used expression data of TCGA tumor samples and their paired normal samples (in cancers where both were available, namely BLCA, BRCA, CHOL, COAD, ESCA, HNSC, KICH, KIRC, KIRP, LIHC, LUAD, LUSC, PRAD; READ, STAD, THCA and UCEC) to perform a per-cancer differential expression analysis with the limma-voom R package [35]. Genes with adjusted  $p$ -value lower than 0.05 were considered differentially expressed. Gene set enrichment analysis of gene expression pathways was performed using the fgsea package [36].

### 2.8. Association with Cellular Composition in Healthy and Diseased Tissues

To obtain estimates of the relationship between glycosylation and immune cell-populations in TCGA tumor samples, we downloaded the pre-calculated frequencies from TCIA (quantiseq frequencies) [37]. We evaluated the associations using Pearson's correlation and lasso regression. First, for each cancer, and glycosylation factor, we estimated the Pearson correlation between the normalized gene expression and immune cell population frequencies. We then counted—for each gene—with how many tumor types each immune cell type correlated significantly ( $p < 0.05$ ). We also applied the same method with normal GTEX tissues. For this, we inferred the frequencies of immune cell populations of GTEX samples by applying quantiseq [38] to all GTEX samples using the normalized TPM expression values. Lasso regression analysis was performed as described

previously [39]. Briefly, the estimated frequency of M2 macrophages for each individual cancer sample (grouped by cancer type) was modelled by Lasso regression as implemented in the *glmnet* R package [40], with a 10-fold cross validation to choose the lambda parameter. Statistical significance of the explained variance by each model was assessed for values greater than zero using a margin of more than one standard deviation. Finally, the obtained models were evaluated by assessing the correlation (Pearson method) between the observed and predicted M2 levels based on tumor mutation and expression profiles. A similar approach based on Lasso Models was applied to detect alterations in glycosylation factors associated with the number of tumor subclones, i.e., intra-tumor heterogeneity. Intra-tumoral heterogeneity (ITH) of TCGA samples was defined as the number of clones estimated using EXPANDS [41] on the tumor SNVs and CNAs.

### 2.9. Statistical Analysis

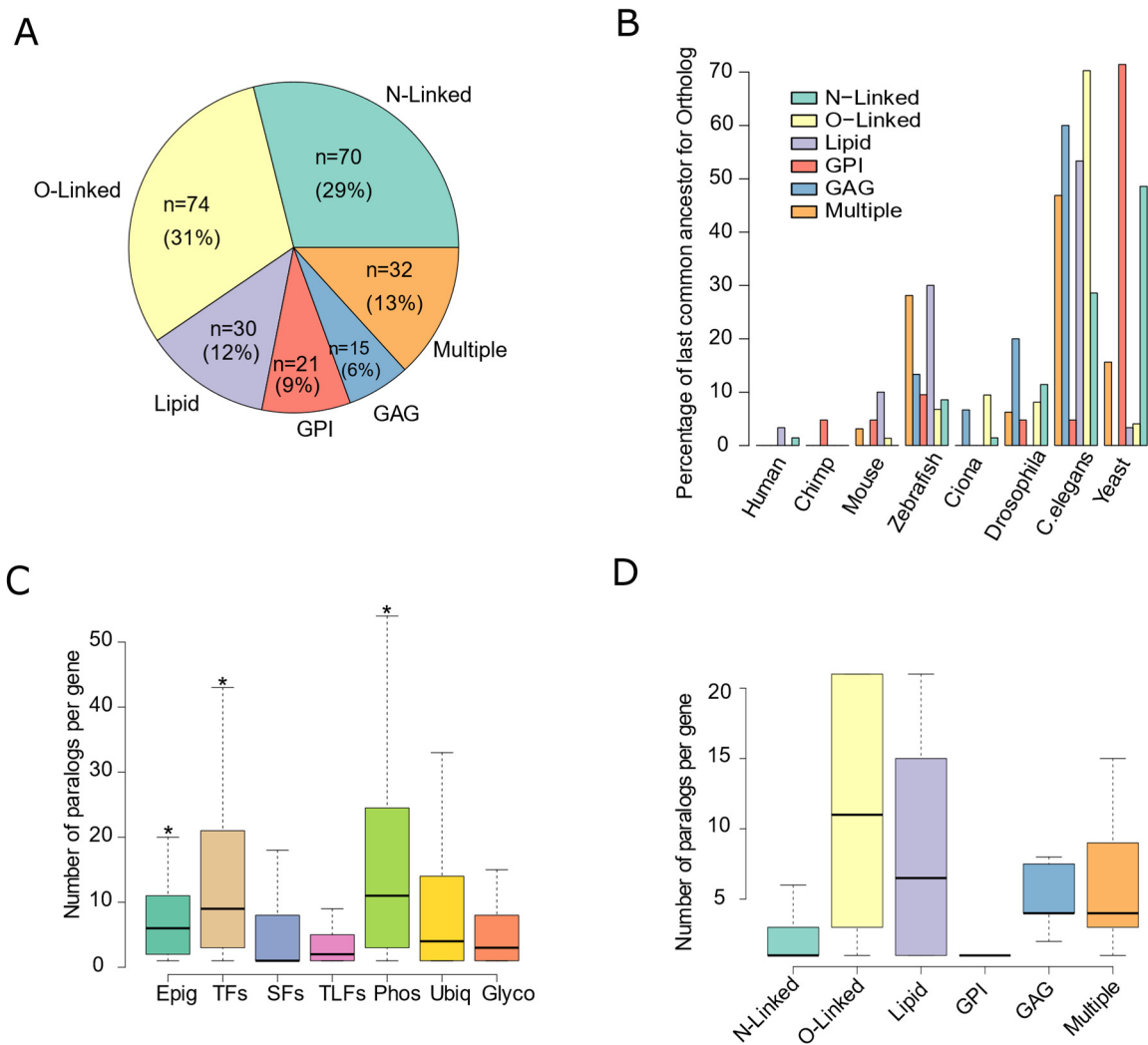
All statistical analysis and figures were generated using R. Besides the already mentioned R packages, we also used the *graph* package for the network plots; the *ComplexHeatmap* package to produce the oncoprint; the *corrplot* package heatmap with correlations; and the *pheatmap* package for the remaining heatmaps.

## 3. Results

### 3.1. Glycosylation Factors Are an Evolutionarily Conserved Family of Genes

We constructed a catalogue of 242 human genes directly involved in the process of glycosylation (see details in methods), herein referred to as glycosylation factors (Supplementary Materials Table S1). Of these, 70 genes (29%) are mainly involved in N-linked glycosylation, 74 (31%) in O-linked glycosylation, 21 (9%) in glypiation (GPI), 30 (12%) in lipid glycosylation, and 15 (6%) in glycosaminoglycan (GAG) glycosylation. The remaining 32 (13%) genes are involved in at least two of the above-mentioned pathways (Figure 1A). Regarding their functions, the vast majority (74%) of the glycosylation factors are glycosyltransferases (Supplementary Materials Table S1), 10% are glycosyl hydrolases, with the remaining genes playing other roles (Supplementary Materials Figure S1A and Table S1).

Our phylogenetic analysis revealed that glycosylation is ubiquitous in the tree of life, with both N- and O-Glycosylation already appearing in archaeal and bacterial species [42,43]. In addition, most human glycosylation factors contain orthologs down to distant organisms such as the nematode worm (Figure 1B and Supplementary Materials Table S1). In fact, some glycosylation subclasses emerged earlier, with 70% of the human glypiation genes and 50% of the N-linked glycosylation factors presenting orthologs already in yeast. Despite their early ancestry, glycosylation factors suffered several gene duplications increasing the number of paralogs present in the human genome (Supplementary Materials Figure S1B). However, glycosylation factors have less paralogs than transcription, phosphorylation, and epigenetic factors (Figure 1C). In fact, 59% of N-linked glycosylation and 90% of glypiation genes are single copy, suggesting a strong control of gene copy number since their ancient origin (Figure 1D). Interestingly, O-linked and lipid glycosylation seem to be an exception, with the average number of paralogs per gene higher than 10 in the human genome (Figure 1D). A striking example is the GALNT gene family, required to initiate O-glycosylation, enclosing 21 human paralogs (Supplementary Materials Table S1). Thus, glycosylation is a conserved and essential cellular process where most direct regulators have emerged early in evolution and were spared from large duplication events.

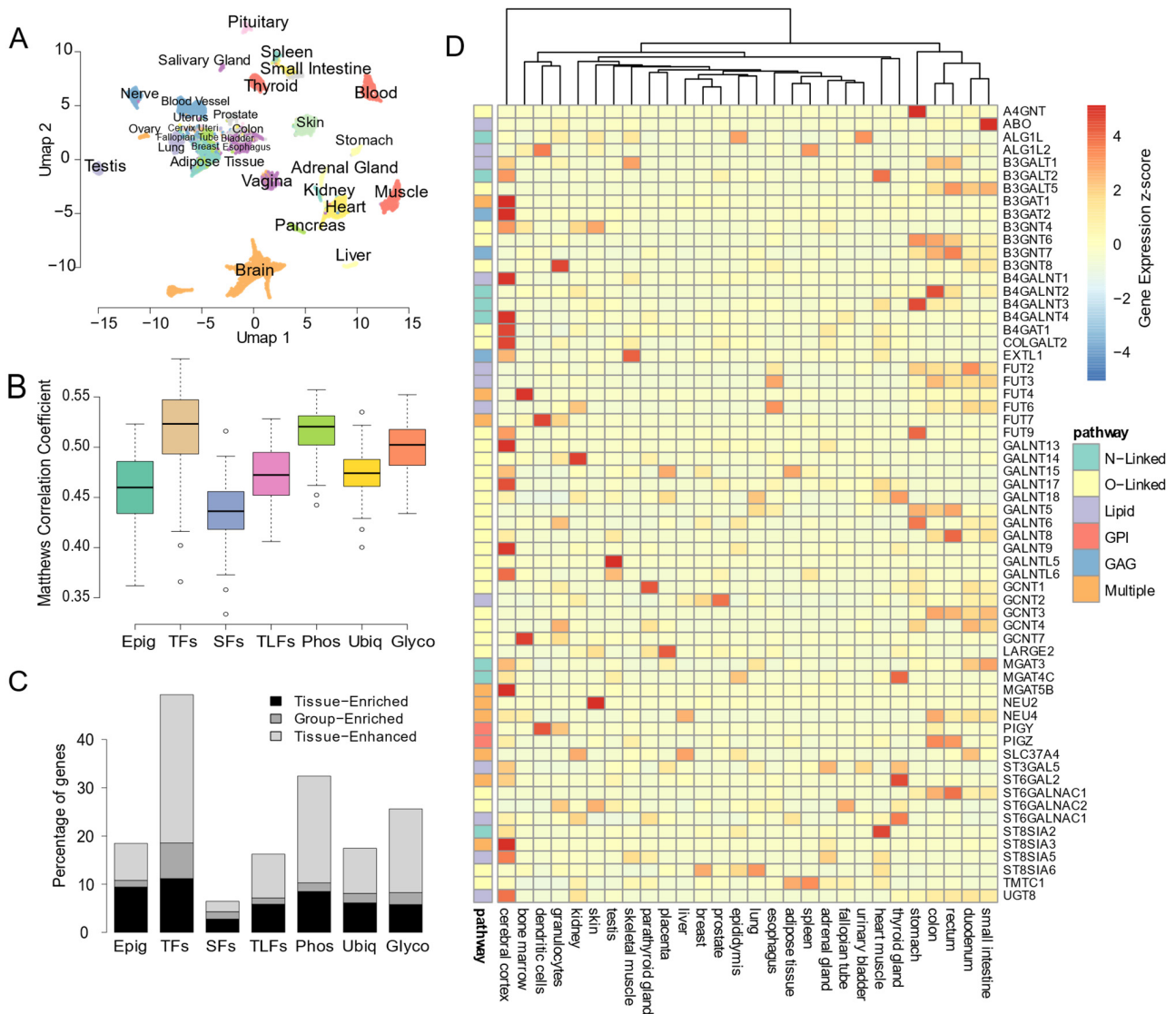


**Figure 1.** Glycosylation Factors are an evolutionarily conserved family of genes. (A) Pie chart with the frequencies of each glycosylation pathway among the 242 genes considered in this study: N-linked; O-Linked; Lipid; Glypiation (GPI); Glycosaminoglycans (GAG); Genes involved in more than one pathway (Multiple). (B) Barplot with the percentages of the organism representing the last common ancestor to the human gene, for each glycosylation subclass. (C) Boxplot of the total number of paralogous copies per gene, for each class of gene expression regulators: epigenetic (Epig); transcription (TF); Splicing (SF); Translation (TLF); Phosphorylation (Phos); Ubiquitination (Ubiq); Glycosylation (Glyco). An asterisk (\*) represents that the number of paralogs per gene in the class is significantly different from glycosylation (Wilcoxon Test  $p < 0.05$ ). (D) Boxplot of the total number of paralogous copies per gene for each glycosylation subclass.

### 3.2. Glycosylation Factors Show Tissue-Specific Expression Preserved across Mammals

Given the complexity of the glycosylation machinery and the existence of cell-specific glycoconjugate profiles, we decided to explore how glycosylation factors expression varied across healthy tissues from the GTEX project [20]. An unsupervised analysis unveiled that human samples could be grouped according to tissue type based solely on the expression levels of glycosylation factors (Figure 2A), reinforcing the fact that each tissue may display different patterns of glycosylation. Hence, transcription profiles of glycosylation factors can distinguish the different tissues significantly better than most regulators of gene expression, except transcription and phosphorylation factors (Figure 2B). In fact, 25% of glycosylation factors present tissue-specificity, enclosing enriched or enhanced expression associated with a single tissue or a small group of tissues

(Figure 2C and Supplementary Materials Table S1). Such specificity is only surpassed by transcription and phosphorylation factor tissue-specific expression (Figure 2C). Interestingly, specificity appears to be most relevant in O-linked and lipid glycosylation pathways, where approximately 40% of the genes showed restricted expression to single or a small group of tissues (Supplementary Materials Figure S2A). Concordantly, these gene subclasses also have the highest ability to distinguish different tissues (Supplementary Materials Figure S2B). Thus, such findings support the existence of a tissue-specific glycosylation code that is sustained by activation of distinct glycosylation factors.



**Figure 2.** Glycosylation Factors show tissue-specific expression. (A) Uniform Manifold Approximation and Projection (UMAP) representation, based on the gene expression profiles of glycosylation factors of GTEX samples, colored according to their tissue of origin. (B) Boxplot representation of the Matthews Correlation Coefficient (MCC) value for the classification of the tissue of each GTEX sample ( $n = 100$  random forest models using 30 randomly selected genes of each gene class). (C) Cumulative bar plot representation of the frequency of genes belonging to each gene class that show tissue-specific expression in protein atlas (colors represent degree of specificity as defined in TissueEnrich from the most specific Tissue-Enriched to the least specific Tissue-Enhanced). (D) Heatmap representation of the protein atlas gene expression of glycosylation factors displaying some degree of specificity (represented expression values are scaled z-scores). Only tissues with at least one gene with z-score  $> 2$  are represented.

Within the different human tissues, the brain and digestive system presented the most distinct expression, with the highest number of tissue-specific glycosylation factors (Figures 2D and S2C). Relevant examples include *B4GAT1* and *B4GALNT1* transferases, showing brain-specificity (Figures 2D and S2D,E), in agreement with the fact that its disruption is known to cause defects in the nervous system [44,45]. Interestingly, the expression of paralogous glycosylation factors alternates and contributes to tissue-specific glycosylation. While *B4GALNT1* and *B4GALNT4* are brain specific, *B4GALNT2* and *B4GALNT3* are specifically expressed in tissues of the digestive system (Figure 2D). Or the case of *GALNT13* and *GALNT14* specifically expressed in the brain and kidney, respectively. Moreover, our analysis also identified signatures that may define distinct glycosylation patterns in tissues, highlighting: the brain (*GALNT9*, *GALNT13*, *ST8SIA3*, *B3GAT1*, *B3GAT2*, *MGAT5B*, *B4GAT1*, *B4GALNT1*, *COLGALT2*, *B4GALNT4* and *GALNT17*); stomach (*GALNT6*, *FUT9*, *A4GNT*, *B4GALNT3*); intestine (*ABO*, *B4GALNT2*, *GALNT8*, *ST6GALNAC1*); thyroid and parathyroid glands (*ST6GAL2*, *MGAT4C*, *GALNT18*, *ST6GALNAC3*, *GCNT1*) (Figure 2D, Supplementary Materials Figure S2F and Table S2).

To reinforce the existence of a tissue-specific code, we investigated how conserved is the expression of glycosylation factors across mammals. Interestingly, glycosylation factors unveiled a low within-tissue expression divergence of orthologs across different mammalian species, only slightly above transcription or phosphorylation factors (Supplementary Materials Figure S2G). Moreover, the glycosylation subclasses with the highest tissue-specific expression, O-linked and lipid glycosylation, also revealed lower within-tissue expression variation across species (Supplementary Materials Figure S2H). Such evolutionarily conserved tissue-specific profiles clearly strengthen the importance of glycosylation to maintain cell identity and tissue homeostasis in complex organisms. Thus, our results indicate that glycosylation factor expression is tightly regulated and preserved across tissues and mammalian species, supporting the existence of a conserved glycosylation code.

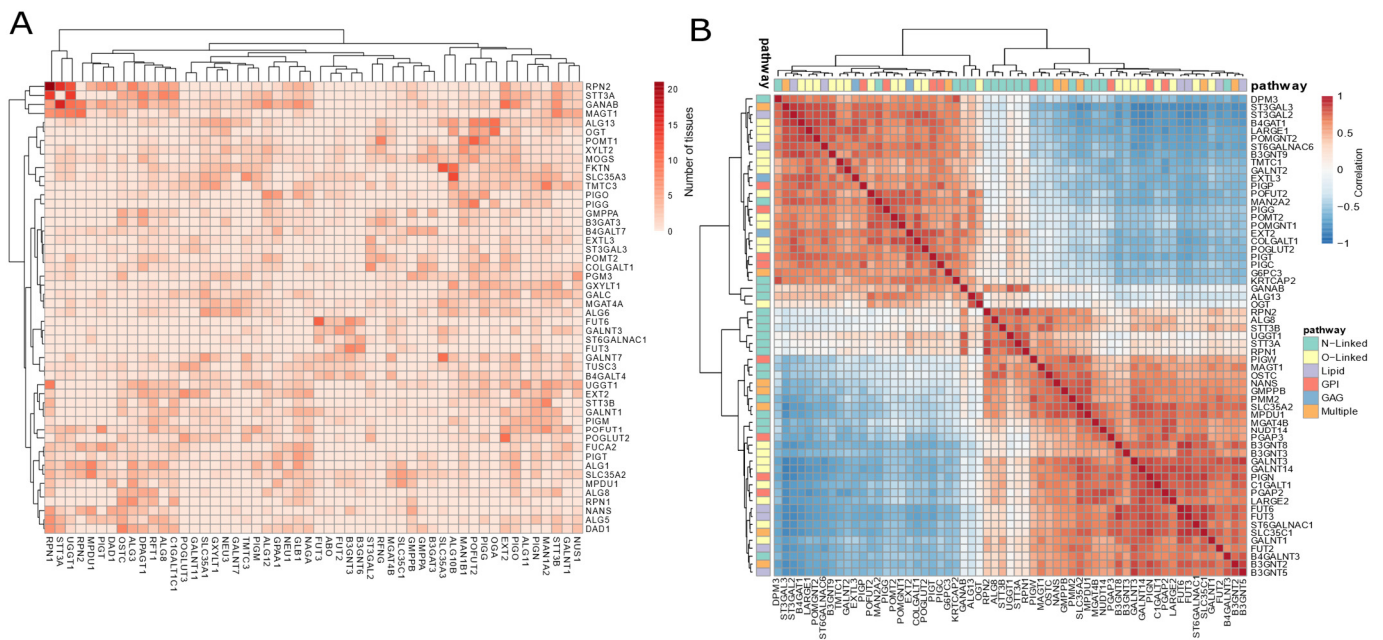
### 3.3. Co-Expression Patterns Strengthen a Tissue-Specific Glycosylation Code

Since glycosylation patterns depend on the concerted activity of multiple glycosylation factors, one could expect some coordinated gene expression within tissues. Indeed, correlation analysis of human transcriptome profiles identified 6491 pairs of glycosylation factors highly correlated (Pearson  $R > 0.8$ ) covering almost all glycosylation factors (Supplementary Materials Table S3). Moreover, we identified glycosylation factors consistently co-expressed across several tissues (Figures 3A and S3A). The most common cluster includes a core of N-linked glycosylation genes *RPN1*, *RPN2*, *STT3A*, *GANAB* and *UGGT1*, which are strongly correlated to each other across over 10 tissues. Other cases include *FUT3* and *FUT6* in 10 tissues, and *OGT*, *OGA* and *ALG13* in nine tissues. Notably, only few glycosylation factors showed strong anti-correlated expressions (Pearson  $R < -0.8$ ), such as *ST3GAL3/GALNT3* and *ST3GAL3/FUT3*, both in the colon and the esophagus (Supplementary Materials Figure S3B). The consistent co-expression of glycosylation factors suggests that these are responsible for the biosynthesis of prevalent glycoconjugates common to several tissues.

Despite the conserved co-expression patterns, our analysis revealed an association between certain glycosylation subclasses and tissues. Indeed, we could depict several N-linked glycosylation genes highly correlated in muscle, and O-linked glycosylation genes in digestive system tissues (Supplementary Materials Figure S3C). Such findings suggest a differential prevalence of glycosylation pathways in distinct tissues. More importantly, we identified two highly anti-correlated ( $R < -0.8$ ) clusters of genes in the esophagus, indicating two opposing glycosylation patterns occurring in that tissue (Figure 3B). These opposing clusters included pairs of paralogous genes (e.g., *LARGE1* and *LARGE2*, *GALNT1* and *GALNT2*, *B3GNT8* and *B3GNT9*), indicating a concerted regulation of paralogous genes in the glycosylation process. Therefore, our findings unveiled the existence of a concerted



expression of the glycosylation factors within tissues, reinforcing the existence of a firmly regulated glycosylation code.



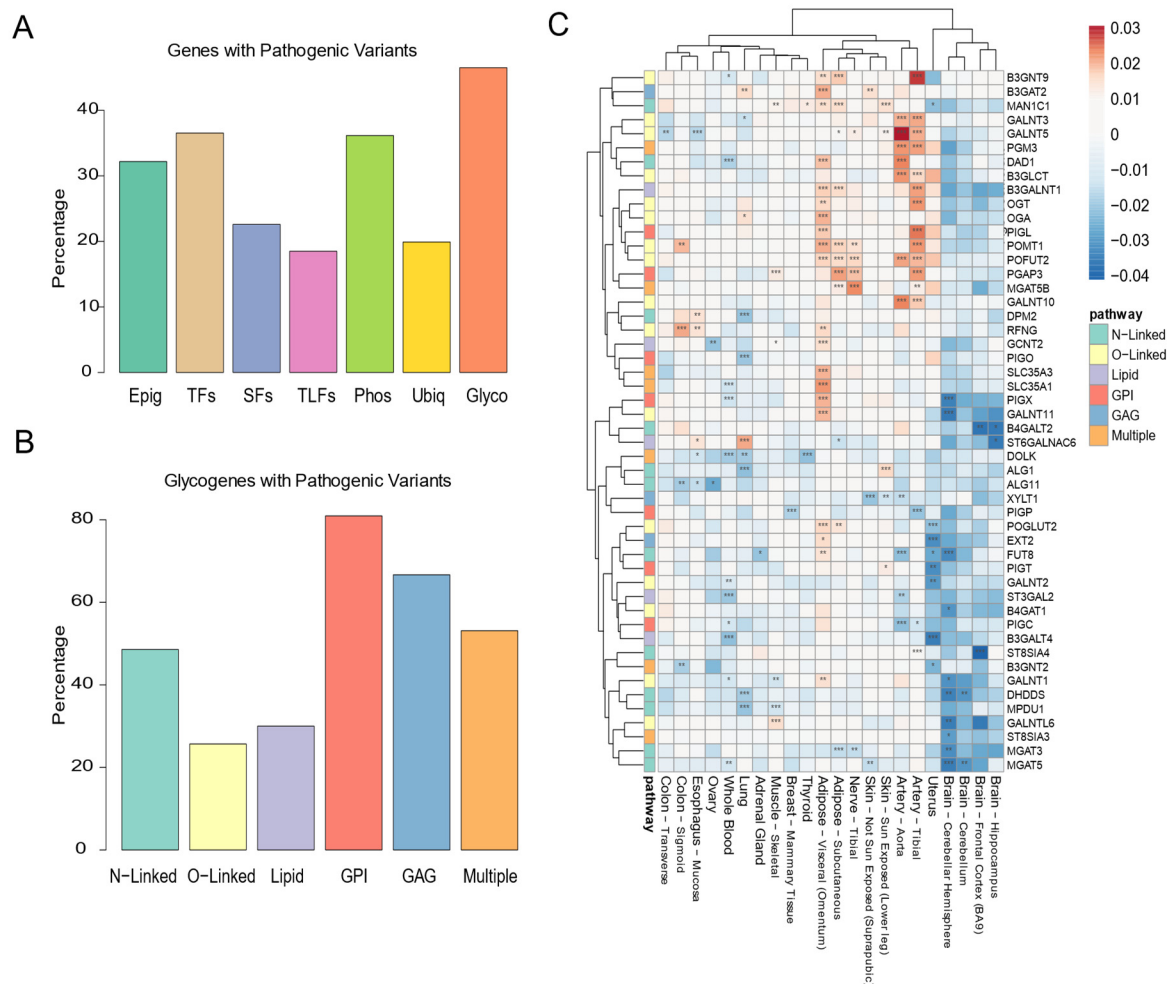
**Figure 3.** Glycosylation factors show co-expression clusters across healthy tissues. (A) Heatmap representation of the number of tissues with high positive correlation between pairs of glycosylation factors (based on the gene expression of GTEX samples,  $R > 0.8$ ,  $p < 0.05$ ). (B) Heatmap representation of the correlation values of highly (anti)correlated pairs of glycosylation factors (based on the gene expression of GTEX esophagus samples,  $|R| > 0.8$ ). Node colors represent glycosylation subclass; color intensity and line thickness represent the number of tissues where the gene pair displays a high degree of correlation.

### 3.4. Maintenance of the Glycosylation Machinery Is Important for Human Health

Due to such highly conserved tissue-specificity, one could foresee that impairment of glycosylation factors would have a significant impact on tissue homeostasis and human health. In fact, according to ClinVar resource [46], over 40% of the glycosylation factors enclose genetic variants associated with clinically pathogenic phenotypes (Figure 4A and Supplementary Materials Table S1), overtaking the other classes of gene expression regulators. However, such proportion differs across the different glycosylation pathways, with pathogenic variants being described for 80% of the glypiation genes but only affecting 23% of the O-linked glycosylation factors (Figure 4B). The pathogenic variants in glycosylation factors are mostly associated with muscle and intellectual disability, usually included in the rare family of congenital disorders of glycosylation (CDGs) (Supplementary Materials Figure S4A). The majority of CDGs are autosomal recessive disorders that manifest from infancy, being homozygous mutations usually associated with lethality to the embryo [47–50].

Due to the higher number of glycosylation factors associated with pathogenic variants, we decided to assess their intolerance to mutations in human samples using the pLI score (probability of being Loss-of-function intolerant). The pLI measures the probability of intolerance of a given gene to the loss of function on the basis of the frequency of protein-truncating variants [51]. Surprisingly, most of the glycosylation factors showed low or intermediated pLI scores, contrasting with other gene regulator classes (Supplementary Materials Figure S4B). This apparent tolerance to protein variation may be due to compensatory co-expression of redundant paralogous glycosylation factors observed in certain tissues. Nevertheless, some factors showed high intolerance to loss of function mutations as it is the case of *STT3A*, *STT3B* and *GANAB* genes, belonging to

the co-expressed cluster of N-glycosylation (Supplementary Materials Table S1). Such findings disclose the genetic diversity, complexity and susceptibility underlying the glycosylation machinery.



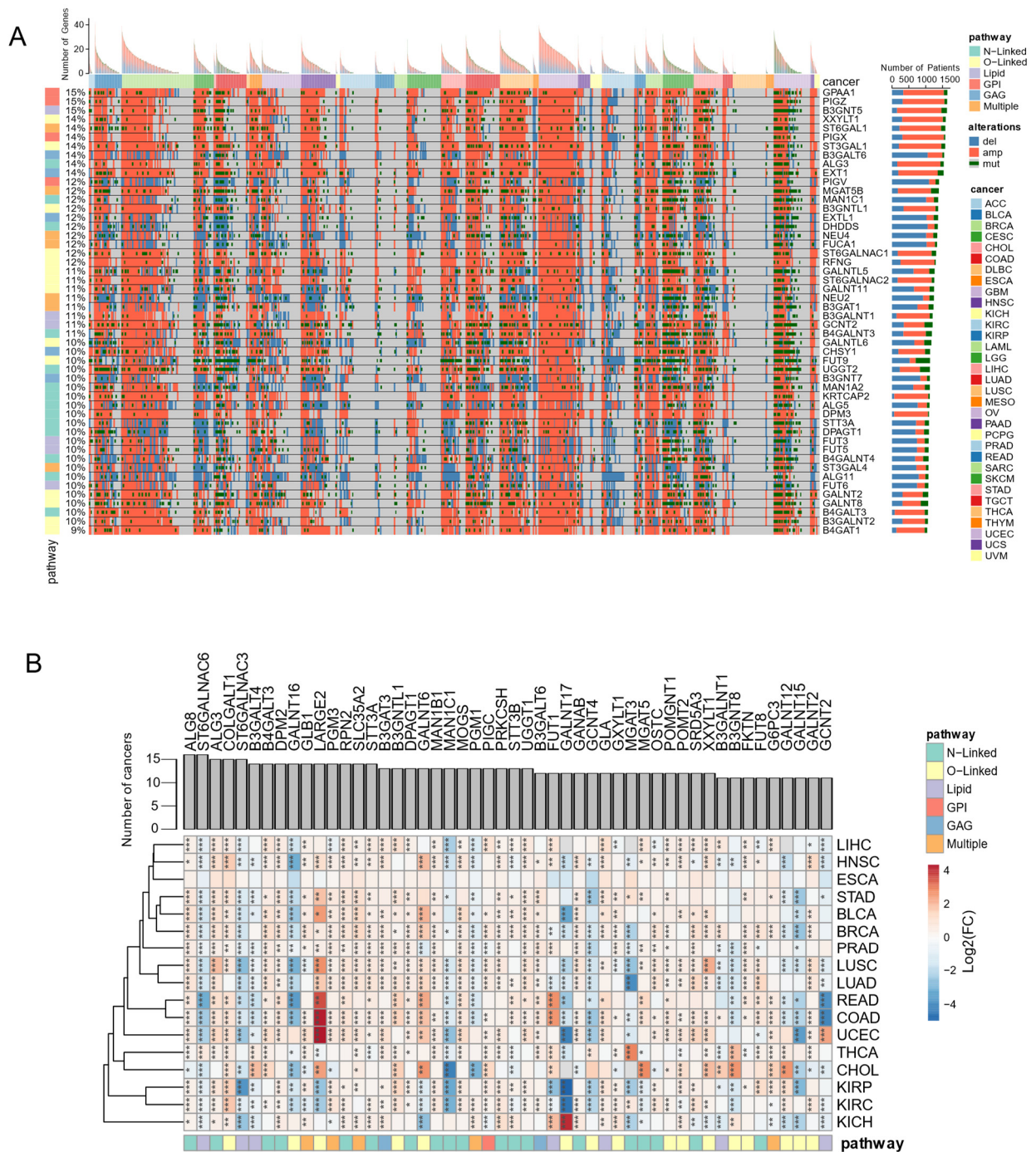
**Figure 4.** Glycosylation machinery is important in human health. (A) Bar plot of the percentage of genes associated with pathogenic variants in Clinvar, according to their gene class. (B) Bar plot of the percentage of glycosylation factors associated with pathogenic variants in Clinvar, according to their glycosylation subclass. (C) Heatmap representation of glycosylation factors whose expression changes in GTEX ageing tissues. Heatmap colors represent the coefficients from a linear model associating age with gene expression, in a given tissue (\*:  $p < 0.05$ ; \*\*:  $p < 0.01$ ; \*\*\*:  $p < 0.001$ ). Red indicates higher and blue lower expression with ageing.

Besides the impact of genetic alterations, glycosylation also changes during ageing, representing both a predisposition to and a functional mechanism involved in disease pathology [52]. In fact, GTEX expression profiles unveiled a significantly higher number of glycosylation factors altered in ageing samples of adipose tissue, artery, brain, and whole blood (Figures 4C and S4C, Table S1). Interestingly, the expression of glycosylation factors seems to predominantly decrease with age in the brain, whole blood, and uterus, while it predominantly increases in adipose tissue and arteries (with a predominance of O-linked glycosylation) (Figure 4C). Apart from changing according to tissue-type and age, a smaller set of glycosylation factors showed sex-specific expression profiles in muscle, skin, thyroid, and adipose tissue (Supplementary Materials Figure S4D,E). Overall, these results suggest that the glycosylation machinery may play a role not only in congenital anomalies but also in ageing-related diseases.

### 3.5. Genome and Transcriptome Alterations of Glycosylation Factors Are Pervasive in Cancer

One of the diseases with increasing prevalence with age is cancer, where aberrant glycosylation plays an important role [17]. However, pathogenic variants in glycosylation factors rarely appear as explicitly associated with cancer in relevant resources such as Clinvar (Supplementary Materials Figure S4A) or IntOGen (Supplementary Materials Figure S5A), and they do not seem to be under strong selective pressure in cancer (Supplementary Materials Figure S5B). Accordingly, the genomic profiles from The Cancer Genome Atlas cohorts (TCGA) unveil that only a median of 0.34% of cancer patients have mutations in glycosylation factors, versus 0.66% and 0.54% for epigenetic and phosphorylation factors, respectively (Supplementary Materials Figure S5C). Within glycosylation factors, glypiation genes emerge as the least mutation prone, with a median of only 0.2% of cancer patients harboring mutations in these genes (Supplementary Materials Figure S5D). However, around 5% of cancer patients have copy number alterations (CNA) in the glycosylation factors, similarly to other gene classes (Supplementary Materials Figure S5E,F). Some glycosylation factors showed genomic alterations (SNVs and CNVs) in 15% of tumor samples (namely *GPAA1*, *PIGZ* and *B3GNT5*) with a predominance of amplifications (Figure 5A). Other genes are mostly affected by deletions, such as *B3GALT6*, *PIGV* and *MAN1C1*. Interestingly, three of the ten most altered glycosylation genes in cancer are glypiation-related genes (*GPAA1*, *PIGZ*, *PIGX*), which are predominantly affected by amplifications (Figure 5A). *GPAA1*, *PIGZ* and *PIGX* amplifications are associated with worse prognosis at a pan-cancer level (Supplementary Materials Figure S5G–I). CNAs are the most prevalent type of mutation even in the genes with most mutations, such as *FUT9* and *UGGT2*. Importantly, 26% of the glycosylation factors show a significant association between being affected by large genomic alterations and overall patient survival (Supplementary Materials Table S1). Thus, disruption of glycosylation factors by large genomic alterations is frequent across several cancer types, where some can be potential prognostic biomarkers of tumor progression.

Besides genetic alterations, cancer progression is also associated with large transcriptome changes. Hence, similarly to the other regulators of gene expression, glycosylation factors are also shown to be perturbed at the transcriptional level, particularly in colon, lung, kidney and breast cancers from TCGA (Supplementary Materials Figure S5J). Notably, a gene set enrichment analysis shows that glycosylation factors are predominantly upregulated in liver, breast, and lung cancers (Supplementary Materials Figure S5K). Indeed, several glycosylation factors, namely *ALG8*, *ALG3*, *COLGATLT1*, *B4GALT3* and *DPM2*, show increased expression levels in these and other cancers (Figure 5B). In contrast, a limited number of genes are predominantly downregulated throughout all cancer types, such as *ST6GALNAC6*, *ST6GALNAC3*, *GALNT16* and *MAN1C1*. Interestingly, such transcriptome disruption appears to be driven at least in part by increased gene copy numbers, as there is a small but significant correlation ( $R = 0.3$ ,  $p = 4.9 \times 10^{-10}$ ) between CNA prevalence and expression alterations for the same gene (Supplementary Materials Figure S5L). Indeed, the genomic amplification of *ALG3*, *B3GNT5*, *GPAA1*, *ST6GALNAC2*, *EXT1* leads to overexpression in tumors, whereas deletions of *PIGV*, *B3GNT7*, *B3GAT1*, *FUCA1*, *MAN1C1* genes drives to downregulation. Thus, our analysis unveils that glycosylation factors are disrupted in cancer by large genomic and transcriptomic alterations that may drastically affect glycan make-up in tumor cells.



**Figure 5.** Glycosylation Factors are dysregulated in cancer. **(A)** Oncoprint view of the 50 glycosylation factors with highest frequency of genetic alterations (over all TCGA samples). **(B)** Heatmap representation of the transcriptome alterations (log<sub>2</sub>-fold change) between tumor and normal TCGA samples, of the 40 glycosylation factors most frequently deregulated in the different tumor types (\*: *p* < 0.05; \*\*: *p* < 0.01; \*\*\*: *p* < 0.001). ACC (adrenocortical carcinoma); BLCA (bladder urothelial carcinoma); BRCA (breast invasive carcinoma); CESC (cervical squamous cell carcinoma and endocervical adenocarcinoma); CHOL (cholangiocarcinoma); COAD (colon adenocarcinoma); DLBC (Lymphoid Neoplasm Diffuse Large B-cell Lymphoma); ESCA (Esophageal carcinoma); GBM (Glioblastoma multiforme); HNSC (head and neck squamous cell carcinoma); KICH (kidney Chromophobe); KIRC (kidney renal clear cell carcinoma), KIRP (kidney renal papillary cell carcinoma); LAML (Acute Myeloid Leukemia); LGGL (Brain Lower-Grade Glioma); LIHC (liver hepatocellular carcinoma); LUAD (lung adenocarcinoma); LUSC (lung squamous cell carcinoma), MESO (Mesothelioma);

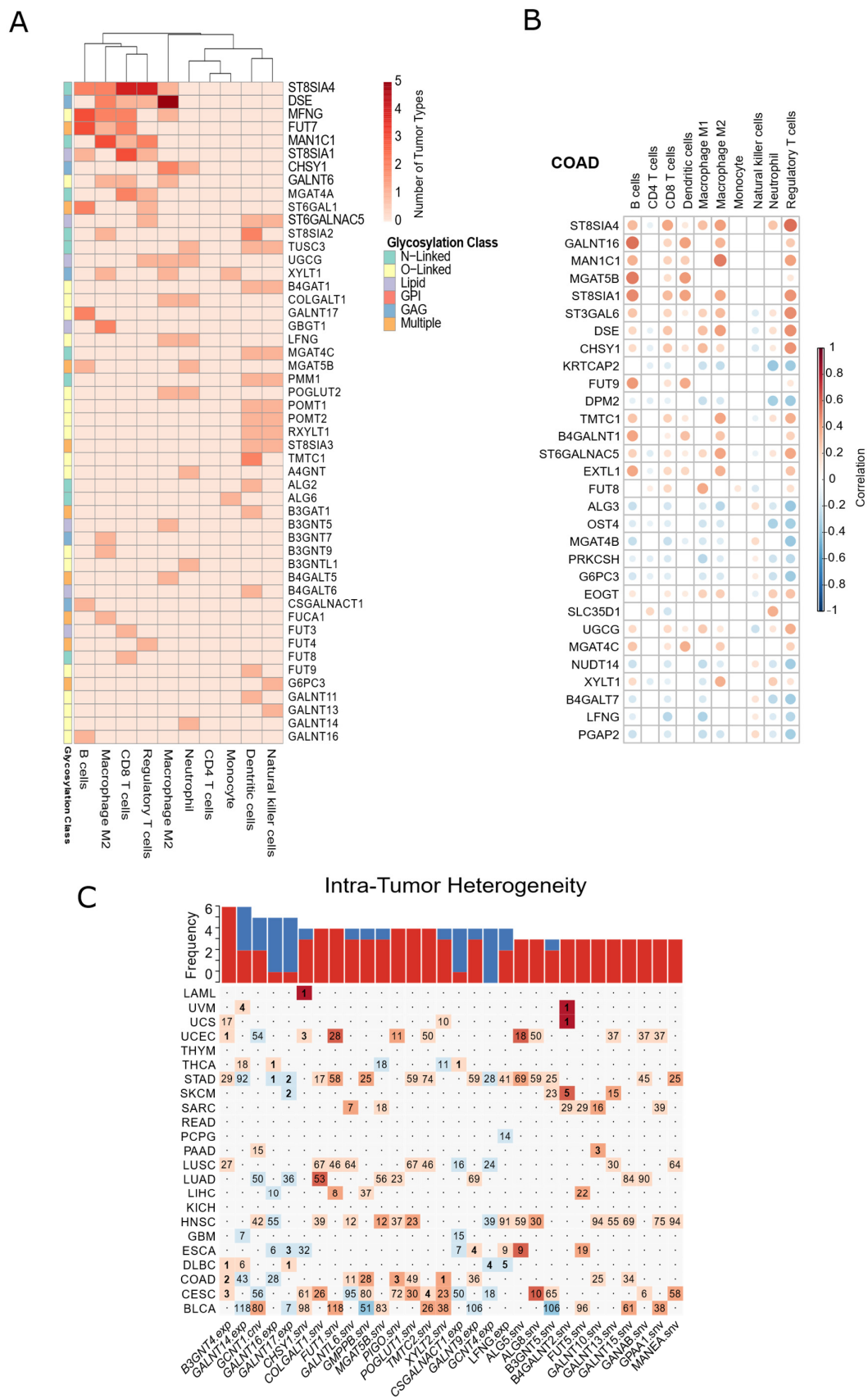
OV (Ovarian serous cystadenocarcinoma); PAAD (pancreatic adenocarcinoma); PCPG (pheochromocytoma and Paraganglioma); PRAD (prostate adenocarcinoma); READ (rectum adenocarcinoma); SARC (Sarcoma); SKCM (Skin Cutaneous Melanoma); STAD (Stomach adenocarcinoma); TGCT (Testicular Germ Cell Tumors); THCA (thyroid carcinoma); THYM (Thymoma); UCEC (uterine Corpus Endometrial Carcinoma); UCS (Uterine Carcinosarcoma); and UVM (Uveal Melanoma).

### 3.6. Glycosylation Machinery Is Associated with Changes in the Cellular Composition of Healthy Tissues and Tumor Microenvironment

Cancer-associated changes in glycosylation may have an impact in cell-to-cell communication, namely between tumor cells and the immune system. For example, *B3GNT3*-mediated glycosylation of the immune checkpoint PD-L1 decreases the capacity of the immune system to suppress tumor progression [53]. Consistently, our analysis unveiled that *B3GNT3* amplification in tumors was associated with the worst patient survival (Supplementary Materials Figure S6A). Thus, to deeply characterize the immune-cell populations associated with glycosylation factor disruption in cancer, we explored cellular compositions inferred by RNA-based deconvolution methods [37]. Combining cellular compositions and gene expression levels within TCGA tumor samples, we depicted glycosylation factors consistently associated with an increase of specific immune-cell types across several cancer types (Figure 6A). The data showed that the presence of dendritic and natural killer cells was associated with the alterations mostly in O-linked glycosylation genes (e.g., *B4GAT1*, *POMT1*, *POMT2*, *RXYLT1*, *TMTC1*, *FUT9*, *GALNT11*, *GALNT13*). In opposition, the presence of macrophages, B and T-cells is favored by alterations in different glycosylation gene subclasses (e.g., *ST8SIA4*, *DSE*, *MFNG*, *FUT7*, *MAN1C1* and *ST8SIA1*). Interestingly, the expression of sialyltransferases, particularly of the *ST8SIA* family, is most often associated with changes in the frequencies of tumor-associated immune cell populations (Figure 6A). For example, in colorectal cancer, an increase in the expression of *ST8SIA4*, *ST3GAL6*, *ST6GALNAC5* is associated with an increase in regulatory T cells and pro-tumoral M2 macrophages (Figure 6B). In addition, a lasso regression analysis revealed that the expression of *ST8SIA4* and other glycosylation factors is positively correlated with pro-tumoral M2 macrophages in various tumor types (Supplementary Materials Figure S6B). Thus, our pan-cancer analysis suggests that disruption of glycosylation factors may lead to specific immune modulation responses in the tumor microenvironment.

Given that glycosylation patterns may also influence the immune cell composition in healthy tissues, we also combined gene expression levels and cellular composition within GTEX tissues. The correlation analysis revealed several significant positive associations between glycosylation factors levels and immune cell populations in healthy tissues, particularly with pro-inflammatory M1 macrophages (Supplementary Materials Figure S6C). Interestingly, we identified a positive association between the *DSE* expression and the relative frequency of M1 macrophages, both in normal (16 distinct GTEX tissues, Supplementary Materials Figure S6C) and tumor tissues (five different TCGA tumor types, Figure 6A). In fact, the *DSE* gene codifies an glycosaminoglycan isomerase, acting as a tumor-rejection antigen and with the potential to stimulate anti-tumoral immunoreactivity [54].

Besides immune-system crosstalk, glycan structures also mediate cell-to-cell communication in the tumor microenvironment influencing tumor progression, where different cancer cell subclones evolve and co-exist (designated as intra-tumor heterogeneity). Thus, we combined the TCGA genomic profiles to infer the number of genetically distinct subclones within each sample (see Materials and Methods). Our lasso regression analysis depicted genomic and transcriptome alterations of several glycosylation factors associated with intra-tumor heterogeneity levels across several cancer types (Figure 6C). Interestingly, some molecular alterations showed the same recurrent outcome in several cancer types. Namely, increased *B3GNT4* expression was associated with subclonal expansion in at least seven tumor types, while *GALNT16* and *GALNT17* levels appeared mostly associated with a decrease in intra-tumor heterogeneity across five cancers. In summary, our analysis revealed that specific alterations in the glycosylation machinery are associated with the presence of distinct immune-cell populations in tumor/normal tissues, and also with tumor subclone diversity.



**Figure 6.** Glycosylation Factors have prognostic value in cancer. (A) Heatmap representation of the number of TCGA cancer types where there is a significant correlation (FDR < 0.05) between gene and

expression and frequencies of immune cell populations (obtained from TCIA). **(B)** Heatmap representation of Pearson correlation values between gene expression values of glycosylation factors and estimated frequency of immune cell populations (from TCIA), in TCGA COAD samples (only significant— $p < 0.05$ —correlation values are displayed, and only top 30 genes with highest (anti)correlation are displayed). **(C)** Heatmap representation of the significant coefficients for gene perturbation events in a lasso-regression model correlating events with the number of genetic clones (estimated using EXPANDS [41]) in TCGA samples. Red color indicates positive coefficients and blue color, negative coefficients. Color intensity indicates value of the coefficient, with darker tones indicating higher values for the coefficient. Numbers indicate relative importance of the event in the regression model. Only cancers with an overall  $R > 0.3$  for the model are displayed.

#### 4. Discussion

Here, we used computational approaches and genome/transcriptome profiling to deeply characterize the evolution and relevance of the glycosylation machinery in human health. Our study unveiled that glycosylation factors are an ancient and conserved family of genes, with an apparent evolutionary pressure to keep a low number of copies, particularly in GPI and N-linked glycosylation, with most glypiation genes being single copies. Despite the low copy number, glycosylation factors also showed high tolerance of loss-of-function mutations, suggesting that a single functional allele suffices for their function or there is a functional compensation from paralogous genes. This is consistent with the etiology of the congenital disorders of glycosylation (CDG), where most tend to be autosomal recessive and enclose less than 100 cases [55]. Accordingly, CDG animal models tend to be embryonic lethal, with the rare instances of viability resulting from hypomorphic variants [56,57]. Noteworthy, most genes associated with known CDG are ubiquitously expressed in healthy human tissues, in line with the observation that most CDGs display complex pleiotropic clinical phenotypes [47]. O-linked and lipid glycosylation are exceptions to this evolutionary path, with a few gene families enclosing large numbers of copies. Perhaps unsurprisingly, these genes display the greatest degree of tissue-specificity expression, likely contributing to the formation of distinct glycosylation patterns in tissues.

Our results support the existence of a tissue-specific glycosylation code established by the coordinated expression of distinct glycosylation factors. Indeed, the gene expression pattern of glycosylation factors enables the correct classification of different tissues, almost similar to that of transcription and phosphorylation factors; already known to be essential for tissue identity [58]. Within glycosylation factors, O-linked and lipid glycosylation showed the highest ability to distinguish different tissues, suggesting an important role for such specific post-translational modifications in tissue definition. Moreover, we unveiled that such specific patterns are phylogenetically conserved across mammals' tissues, reinforcing the existence of tissue-specific glycosylation code. Nevertheless, we acknowledge that the transcriptome may only partially correlate with the activity of glycosylation factors and effective changes in glycan profiles [59]. Despite vast progress in the field of glycoproteomics, these technologies still do not afford a very detailed topology of the glycostructures [1], making the link with gene expression even more complicated. To our knowledge, there is no map of glycostructures simultaneously generated for several human tissues that would enable us to systematically compare with expression data from protein atlas or GTEx at a larger scale, and even less quantitatively. A study in zebrafish has shown that the presence of tissue-specific glycosylation patterns such as sialylation could be associated with gene expression of enzymes associated with sialylation [60]. A recent study in the mammalian brain also observes a correlation between abundances of some glycostructures and gene expression [61]. Finally, recent studies have simultaneously assess the presence of glycan structures and transcripts at the single-cell level, combining scRNA-Seq with one [62] or several [63] lectin-bound DNA-barcodes covering different type of glycosylated proteins. These technologies enabled establishing correlation between gene expression and glycan structures as it is case of ST6Gal1 which showed

the highest correlation with  $\alpha$ 2-6Sia-binding lectin rPSL1a [63]. Yet, the accessibility and resolution of these methods is still quite low. Regarding the different tissues, brain showed the highest number of specific glycosylation factors, in agreement with neuromuscular manifestations being the most frequently described clinical phenotypes derived from glycosylation defects [47]. Recent studies show a very distinctive restricted glycosylation repertoire in the brain, suggesting tight regulation [61,64]. In our analysis, we observed several brain-specific GALNTs (GALNT13/15/17), in agreement with the observation of the predominance of O-GalNAc structures in the brain [61]. Furthermore, we depicted tissue-specific clusters of strongly correlated O-linked and lipid glycosylation factors in stomach, intestine, and colon, providing further evidence of the potential importance of these pathways in the digestive system. Indeed, O-glycosylation of intestinal mucins are among the most well studied examples of the biological relevance of glycosylation, where mucins play a fundamental role in host-microbiota interactions [65,66]. Therefore, we unveiled tissue-specific clusters of co-expressed glycosylation factors, suggesting a synergistic and/or independent role of different glycosylation pathways in the various tissues. More strikingly, we could also detect in the esophagus two strongly antagonistic clusters of glycosylation factors (mostly associated with O-linked and lipid glycosylation), both sharing a common, mostly independent, core of N-linked glycosylation. One of these clusters contain the *LARGE1/B4GAT1* unit responsible for the glycosylation of dystroglycan [44], and several elements of the canonical and non-canonical O-linked glycosylation pathway, including *COLGALT1*, *POMT2*, *POFUT2*, and *POGLUT2* [67]. The other cluster contains several elements of the Golgi-associated O-linked glycosylation of mucins, including several B3GNTs, GALNTs, and *C1GALT1*.

The existence of a tissue-specific glycosylation code reinforces the importance of this post-translational modification process in tissue homeostasis throughout all developmental stages, including in adulthood. Ageing is among the factors most associated with increased cellular dysfunction and emergence of morbidities [68]. In fact, alterations in immunoglobulin G glycosylation have been associated with age-related diseases such as diabetes and hypertension [52]. Interestingly, a commercial test (GlycoAgeTest©) uses the logarithm of the ratio of two N-glycans (NGA2F and NA2F) to infer the glycosylation age of a person [69,70], similarly to the methylation clock [71]. Our exploration of transcriptome profiles of GTEX project unveiled a significant down-regulation of the sialic acid transporter *SLC35A1* in blood samples from older people, in agreement with the reported decrease of sialylation with age [68]. This profile also showed down-regulation of the galactosyltransferase *B3GALT4* and the sialyltransferase *ST3GAL2*, suggesting a decrease in galactosylation and sialylation in O-glycan and GSL with age, yet to be identified. Therefore, our results suggest a downregulation of the glycan structures, derived from *B3GALT4*, *ST3GAL2* and/or *SLC35A1* genes, with ageing, a feature so far only described in neurodegenerative diseases [72]. Moreover, we detected a significant increase in the expression of the mannosidase *MAN1C1* gene in the ageing-skin, an alteration that has also been reported in a previous study [73]. Interestingly, we also observed an increased expression of several O-linked glycosylation factors in the arteries and adipose tissue of old people, suggesting a role of this pathway in ageing-related alterations in these tissues. Previous works have also shown that ageing is associated with a decreased expression of several glycosylation factors in the brain [74]. Since most studies assessing ageing-related changes in glycosylation use blood derivatives (plasma, serum, immunoglobulin fractions), their applicability may be limited and overlook tissue-specific effects. Thus far, our findings further indicate that age-associated changes in glycosylation are tissue-specific.

Alterations in the glycosylation machinery have been previously detected in cancer [17]. Our pan-cancer analysis of TCGA cohorts supports the idea that short variants directly affecting gene function are not likely to be a major mechanism underlying defects in glycosylation leading to cancer [67,75]. Instead, they suggest that copy number alterations of glycosylation factors are quite pervasive across all cancer types, with a predominance of genomic amplifications. These findings are consistent with previous global analysis



showing that large genomic alterations are more recurrent in cancer than small genetic alterations [76]. In addition, our analysis unveiled that such genomic alterations are reflected at the transcriptome level and may affect the glycan make-up in tumor cells. More importantly, we can depict several genomic alterations of specific glycosylation factors associated with changes in patient survival, which can be potential biomarkers for prognosis. In agreement, glycosyltransferases have been shown to be useful in the classification and prognosis of certain cancers [77], where differences in the expression of sialyltransferases seem to be among the most relevant to distinguish different cancer types. An increase in sialylation has been reported in some tumor types, such as colorectal cancer [17,74].

Alteration in the glycosylation patterns of tumor cells plays a very important role in the modulation of immune responses to cancer and have been shown to be major factors in the design of novel immunotherapies [78]. When we combined cancer transcriptome profiles and estimates of immune-cell abundance (through RNA-based deconvolution approaches), we observed a recurrent association between the expression levels of specific glycosylation subclasses and the presence of dendritic, natural killer cells, macrophages, B and T-cells. In fact, we detected an increase in the expression of sialylation genes and the presence of regulatory T-cells and pro-tumoral M2 macrophages, in agreement with recent reports linking sialylation to poorer cancer outcomes [18,79,80]. We could also detect a recurrent association between changes in the gene expression of some glycosylation factors and intra-tumoral heterogeneity, the relevance of which is yet to be determined.

In conclusion, through exploration of molecular profiles from healthy, ageing and diseased tissues, we showed the potential of glycosylation factors in defining tissue identity through a tissue-specific glycosylation code, analogous to transcription and phosphorylation. Moreover, our work unveiled that significant alterations in the gene expression patterns of the glycosylation machinery occur in ageing-tissues, suggesting that cell-specific glycoconjugate profiles may change throughout adulthood during the aging process. Finally, large copy number amplifications and corresponding changes in the expression of glycosylation factors emerged as the main associators to cancer-related alterations.

Overall, this work provides a rich set of information, derived from multiple and integrative sources, that will be useful for the glycosciences community in expanding the knowledge on the function of glycosylation in human health.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/biomedicines10081805/s1>, Figure S1: Glycosylation Factors are an evolutionarily conserved family of genes; Figure S2: Glycosylation Factors show tissue-specific expression; Figure S3: Glycosylation Factors seem to be organized in clusters of co-expression; Figure S4: Glycosylation Factors are clinically relevant; Figure S5: Glycosylation Factors are altered in cancer; Figure S6: Glycosylation Factors have prognostic value in cancer; Table S1: List of glycosylation factors used in the current study and respective explored features; Table S2: List of top 5 glycosylation factors that are markers of GTEx tissues; Table S3: Glycosylation factors highly correlated in GTEx tissues.

**Author Contributions:** D.S., L.D. and A.R.G. conducted the bioinformatics analyses. R.F. and P.A.V. were responsible for the selection and curation of the final list of glycosylation factors. D.S., R.F., P.A.V. and A.R.G. were responsible for the critical revision of the manuscript and for important intellectual content. D.S., P.A.V. and A.R.G. conceived and designed the study, interpreted the data and wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was financed by national funds from Fundação para a Ciência e Tecnologia (FCT), in the scope of the project UIDP/04378/2020 and UIDB/04378/2020 of the Research Unit on Applied Molecular Biosciences—UCIBIO and the project LA/P/0140/2020 of the Associate Laboratory Institute for Health and Bioeconomy—i4HB. Moreover, this research was funded by: PTDC/MED-ONC/28660/2017 from Fundação para a Ciência e Tecnologia (FCT) to A.R.G. A.R.G. is recipient of Researcher Grant CEECIND/02699/2017 from FCT. R.F. is recipient of a research grant (SFRH/BD/124326/2016) from FCT, Portugal; European Union's Horizon 2020 research and innovation program under the EJP RD COFUND-EJP N° 825575, EJPRD/0001/2020. This work had also the support of INCD funded by FCT and FEDER under the project 22153-01/SAICT/2016.

The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All data used in this work are publicly accessible and referred to in the methods or text as appropriate.

**Acknowledgments:** We thank João Neto and remaining Computational Multi-Omics Lab members at UCIBIO for their helpful comments and suggestions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Varki, A.; Cummings, R.D.; Esko, J.D.; Stanley, P.; Hart, G.W.; Aebi, M.; Darvill, A.G.; Kinoshita, T.; Packer, N.H.; Prestegard, J.H.; et al. (Eds.) *Essentials of Glycobiology*, 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2015.
2. Flynn, R.A.; Pedram, K.; Malaker, S.A.; Batista, P.J.; Smith, B.A.H.; Johnson, A.G.; George, B.M.; Majzoub, K.; Villalta, P.W.; Carette, J.E.; et al. Small RNAs Are Modified with N-Glycans and Displayed on the Surface of Living Cells. *Cell* **2021**, *184*, 3109–3124.e22. [[CrossRef](#)]
3. Apweiler, R. On the Frequency of Protein Glycosylation, as Deduced from Analysis of the SWISS-PROT Database. *Biochim. Biophys. Acta BBA-Gen. Subj.* **1999**, *1473*, 4–8. [[CrossRef](#)]
4. Li, C.-W.; Lim, S.-O.; Xia, W.; Lee, H.-H.; Chan, L.-C.; Kuo, C.-W.; Khoo, K.-H.; Chang, S.-S.; Cha, J.-H.; Kim, T.; et al. Glycosylation and Stabilization of Programmed Death Ligand-1 Suppresses T-Cell Activity. *Nat. Commun.* **2016**, *7*, 12632. [[CrossRef](#)] [[PubMed](#)]
5. Singh, S.; Thompson, J.A.; Yilmaz, B.; Li, H.; Weis, S.; Sobral, D.; Truglio, M.; Aires da Silva, F.; Aguiar, S.; Carlos, A.R.; et al. Loss of  $\alpha$ -Gal during Primate Evolution Enhanced Antibody-Effector Function and Resistance to Bacterial Sepsis. *Cell Host Microbe* **2021**, *29*, 347–361.e12. [[CrossRef](#)] [[PubMed](#)]
6. Harvey, B.M.; Haltiwanger, R.S. Regulation of Notch Function by O-Glycosylation. In *Molecular Mechanisms of Notch Signaling*; Borggrefe, T., Giaimo, B.D., Eds.; Advances in Experimental Medicine and Biology; Springer International Publishing: Cham, Switzerland, 2018; Volume 1066, pp. 59–78. ISBN 978-3-319-89511-6.
7. Lasky, L.A. Selectin-carbohydrate interactions and the initiation of the inflammatory response. *Annu. Rev. Biochem.* **1995**, *64*, 113–140. [[CrossRef](#)] [[PubMed](#)]
8. Lombard, J. The Multiple Evolutionary Origins of the Eukaryotic N-Glycosylation Pathway. *Biol. Direct* **2016**, *11*, 36. [[CrossRef](#)]
9. Ohtsubo, K.; Marth, J.D. Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* **2006**, *126*, 855–867. [[CrossRef](#)]
10. Bagdonaite, I.; Pallesen, E.M.; Ye, Z.; Vakhrushev, S.Y.; Marinova, I.N.; Nielsen, M.I.; Kramer, S.H.; Pedersen, S.F.; Joshi, H.J.; Bennett, E.P.; et al. O-glycan Initiation Directs Distinct Biological Pathways and Controls Epithelial Differentiation. *EMBO Rep.* **2020**, *21*, e48885. [[CrossRef](#)]
11. Matsuda-Lennikov, M.; Biancalana, M.; Zou, J.; Ravell, J.C.; Zheng, L.; Kanellopoulou, C.; Jiang, P.; Notarangelo, G.; Jing, H.; Masutani, E.; et al. Magnesium Transporter 1 (MAGT1) Deficiency Causes Selective Defects in N-Linked Glycosylation and Expression of Immune-Response Genes. *J. Biol. Chem.* **2019**, *294*, 13638–13656. [[CrossRef](#)]
12. Reily, C.; Stewart, T.J.; Renfrow, M.B.; Novak, J. Glycosylation in Health and Disease. *Nat. Rev. Nephrol.* **2019**, *15*, 346–366. [[CrossRef](#)]
13. Lauc, G.; Pezer, M.; Rudan, I.; Campbell, H. Mechanisms of Disease: The Human N-Glycome. *Biochim. Biophys. Acta* **2016**, *1860*, 1574–1582. [[CrossRef](#)]
14. Videira, P.A.Q.; Castro-Caldas, M. Linking Glycation and Glycosylation With Inflammation and Mitochondrial Dysfunction in Parkinson's Disease. *Front. Neurosci.* **2018**, *12*, 381. [[CrossRef](#)]
15. Fuster, M.M.; Esko, J.D. The Sweet and Sour of Cancer: Glycans as Novel Therapeutic Targets. *Nat. Rev. Cancer* **2005**, *5*, 526–542. [[CrossRef](#)] [[PubMed](#)]
16. Kirwan, A.; Utratna, M.; O'Dwyer, M.E.; Joshi, L.; Kilcoyne, M. Glycosylation-Based Serum Biomarkers for Cancer Diagnostics and Prognostics. *BioMed. Res. Int.* **2015**, *2015*, 490531. [[CrossRef](#)] [[PubMed](#)]
17. Pinho, S.S.; Reis, C.A. Glycosylation in Cancer: Mechanisms and Clinical Implications. *Nat. Rev. Cancer* **2015**, *15*, 540–555. [[CrossRef](#)] [[PubMed](#)]
18. Carrascal, M.A.; Severino, P.F.; Cabral, M.G.; Silva, M.; Ferreira, J.A.; Calais, F.; Quinto, H.; Pen, C.; Ligeiro, D.; Santos, L.L.; et al. Sialyl Tn-Expressing Bladder Cancer Cells Induce a Tolerogenic Phenotype in Innate and Adaptive Immune Cells. *Mol. Oncol.* **2014**, *8*, 753–765. [[CrossRef](#)]
19. Lise, M.; Belluco, C.; Perera, S.P.; Patel, R.; Thomas, P.; Ganguly, A. Clinical Correlations of A2,6-Sialyltransferase Expression in Colorectal Cancer Patients. *Hybridoma* **2000**, *19*, 281–286. [[CrossRef](#)]
20. Lonsdale, J.; Thomas, J.; Salvatore, M.; Phillips, R.; Lo, E.; Shad, S.; Hasz, R.; Walters, G.; Garcia, F.; Young, N.; et al. The Genotype-Tissue Expression (GTEx) Project. *Nat. Genet.* **2013**, *45*, 580–585. [[CrossRef](#)]

21. Uhlén, M.; Fagerberg, L.; Hallström, B.M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, Å.; Kampf, C.; Sjöstedt, E.; Asplund, A.; et al. Tissue-Based Map of the Human Proteome. *Science* **2015**, *347*, 1260419. [[CrossRef](#)]
22. The Cancer Genome Atlas Research Network; Weinstein, J.N.; Collisson, E.A.; Mills, G.B.; Shaw, K.R.M.; Ozenberger, B.A.; Ellrott, K.; Shmulevich, I.; Sander, C.; Stuart, J.M. The Cancer Genome Atlas Pan-Cancer Analysis Project. *Nat. Genet.* **2013**, *45*, 1113–1120. [[CrossRef](#)]
23. Togayachi, A.; Dae, K.-Y.; Shikanai, T.; Narimatsu, H. A Database System for Glycogenes (GGDB). In *Experimental Glycoscience*; Taniguchi, N., Suzuki, A., Ito, Y., Narimatsu, H., Kawasaki, T., Hase, S., Eds.; Springer: Tokyo, Japan, 2008; pp. 423–425, ISBN 978-4-431-77921-6.
24. Jassal, B.; Matthews, L.; Viteri, G.; Gong, C.; Lorente, P.; Fabregat, A.; Sidiropoulos, K.; Cook, J.; Gillespie, M.; Haw, R.; et al. The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* **2019**, *48*, gkz1031. [[CrossRef](#)] [[PubMed](#)]
25. Kanehisa, M.; Sato, Y.; Kawashima, M.; Furumichi, M.; Tanabe, M. KEGG as a Reference Resource for Gene and Protein Annotation. *Nucleic Acids Res.* **2016**, *44*, D457–D462. [[CrossRef](#)]
26. Howe, K.L.; Achuthan, P.; Allen, J.; Allen, J.; Alvarez-Jarreta, J.; Amode, M.R.; Armean, I.M.; Azov, A.G.; Bennett, R.; Bhai, J.; et al. Ensembl 2021. *Nucleic Acids Res.* **2021**, *49*, D884–D891. [[CrossRef](#)] [[PubMed](#)]
27. Boukas, L.; Havrilla, J.M.; Hickey, P.F.; Quinlan, A.R.; Bjornsson, H.T.; Hansen, K.D. Coexpression Patterns Define Epigenetic Regulators Associated with Neurological Dysfunction. *Genome Res.* **2019**, *29*, 532–542. [[CrossRef](#)] [[PubMed](#)]
28. Jolma, A.; Yan, J.; Whittington, T.; Toivonen, J.; Nitta, K.R.; Rastas, P.; Morgunova, E.; Enge, M.; Taipale, M.; Wei, G.; et al. DNA-Binding Specificities of Human Transcription Factors. *Cell* **2013**, *152*, 327–339. [[CrossRef](#)]
29. Brawand, D.; Soumillon, M.; Necsulea, A.; Julien, P.; Csárdi, G.; Harrigan, P.; Weier, M.; Liechti, A.; Aximu-Petri, A.; Kircher, M.; et al. The Evolution of Gene Expression Levels in Mammalian Organs. *Nature* **2011**, *478*, 343–348. [[CrossRef](#)]
30. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
31. Jain, A.; Tuteja, G. TissueEnrich: Tissue-Specific Gene Enrichment Analysis. *Bioinformatics* **2019**, *35*, 1966–1967. [[CrossRef](#)]
32. Martincorena, I.; Raine, K.M.; Gerstung, M.; Dawson, K.J.; Haase, K.; Van Loo, P.; Davies, H.; Stratton, M.R.; Campbell, P.J. Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **2017**, *171*, 1029–1041.e21. [[CrossRef](#)]
33. Lin, H.; Zelterman, D. Modeling Survival Data: Extending the Cox Model. *Technometrics* **2002**, *44*, 85–86. [[CrossRef](#)]
34. Robinson, M.D.; Oshlack, A. A Scaling Normalization Method for Differential Expression Analysis of RNA-Seq Data. *Genome Biol.* **2010**, *11*, R25. [[CrossRef](#)] [[PubMed](#)]
35. Law, C.W.; Chen, Y.; Shi, W.; Smyth, G.K. Voom: Precision Weights Unlock Linear Model Analysis Tools for RNA-Seq Read Counts. *Genome Biol.* **2014**, *15*, R29. [[CrossRef](#)]
36. Korotkevich, G.; Sukhov, V.; Budin, N.; Shpak, B.; Artyomov, M.N.; Sergushichev, A. Fast Gene Set Enrichment Analysis. *BioRxiv* **2021**. [[CrossRef](#)]
37. Charoentong, P.; Finotello, F.; Angelova, M.; Mayer, C.; Efremova, M.; Rieder, D.; Hackl, H.; Trajanoski, Z. Pan-Cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell Rep.* **2017**, *18*, 248–262. [[CrossRef](#)] [[PubMed](#)]
38. Plattner, C.; Finotello, F.; Rieder, D. Deconvoluting Tumor-Infiltrating Immune Cells from RNA-Seq Data Using QUANTISEQ. In *Methods in Enzymology*; Elsevier: Amsterdam, The Netherlands, 2020; Volume 636, pp. 261–285. ISBN 978-0-12-820667-6.
39. de Matos, M.; Posa, I.; Carvalho, F.; Morais, V.; Grosso, A.; de Almeida, S. A Systematic Pan-Cancer Analysis of Genetic Heterogeneity Reveals Associations with Epigenetic Modifiers. *Cancers* **2019**, *11*, 391. [[CrossRef](#)] [[PubMed](#)]
40. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **2010**, *33*, 1–22. [[CrossRef](#)]
41. Andor, N.; Harness, J.V.; Müller, S.; Mewes, H.W.; Petritsch, C. EXPANDS: Expanding Ploidy and Allele Frequency on Nested Subpopulations. *Bioinformatics* **2014**, *30*, 50–60. [[CrossRef](#)]
42. Nothaft, H.; Szymanski, C.M. Protein Glycosylation in Bacteria: Sweeter than Ever. *Nat. Rev. Microbiol.* **2010**, *8*, 765–778. [[CrossRef](#)]
43. Tomono, T.; Kojima, H.; Fukuchi, S.; Tohsato, Y.; Ito, M. Investigation of Glycan Evolution Based on a Comprehensive Analysis of Glycosyltransferases Using Phylogenetic Profiling. *Biophys. Phys.* **2015**, *12*, 57–68. [[CrossRef](#)]
44. Praissman, J.L.; Live, D.H.; Wang, S.; Ramiah, A.; Chinoy, Z.S.; Boons, G.-J.; Moremen, K.W.; Wells, L. B4GAT1 Is the Priming Enzyme for the LARGE-Dependent Functional Glycosylation of  $\alpha$ -Dystroglycan. *eLife* **2014**, *3*, e03943. [[CrossRef](#)]
45. Wakil, S.M.; Monies, D.M.; Ramzan, K.; Hagos, S.; Bastaki, L.; Meyer, B.F.; Bohlega, S. Novel B4GALNT1 Mutations in a Complicated Form of Hereditary Spastic Paraplegia: Letter to the Editor. *Clin. Genet.* **2014**, *86*, 500–501. [[CrossRef](#)] [[PubMed](#)]
46. Landrum, M.J.; Lee, J.M.; Benson, M.; Brown, G.; Chao, C.; Chitipiralla, S.; Gu, B.; Hart, J.; Hoffman, D.; Hoover, J.; et al. ClinVar: Public Archive of Interpretations of Clinically Relevant Variants. *Nucleic Acids Res.* **2016**, *44*, D862–D868. [[CrossRef](#)] [[PubMed](#)]
47. Ferreira, C.R.; Altassan, R.; Marques-Da-Silva, D.; Francisco, R.; Jaeken, J.; Morava, E. Recognizable Phenotypes in CDG. *J. Inher. Metab. Dis.* **2018**, *41*, 541–553. [[CrossRef](#)] [[PubMed](#)]
48. Vaes, L.; Tiller, G.E.; Pérez, B.; Boyer, S.W.; Berry, S.A.; Sarafoglou, K.; Morava, E. PMM2-CDG Caused by Uniparental Disomy: Case Report and Literature Review. *JIMD Rep.* **2020**, *54*, 16–21. [[CrossRef](#)]
49. Matthijs, G.; Schollen, E.; Van Schaftingen, E.; Cassiman, J.-J.; Jaeken, J. Lack of Homozygotes for the Most Frequent Disease Allele in Carbohydrate-Deficient Glycoprotein Syndrome Type 1A. *Am. J. Hum. Genet.* **1998**, *62*, 542–550. [[CrossRef](#)]

50. Francisco, R.; Marques-da-Silva, D.; Brasil, S.; Pascoal, C.; dos Reis Ferreira, V.; Morava, E.; Jaeken, J. The Challenge of CDG Diagnosis. *Mol. Genet. Metab.* **2019**, *126*, 1–5. [[CrossRef](#)]
51. Exome Aggregation Consortium; Lek, M.; Karczewski, K.J.; Minikel, E.V.; Samocha, K.E.; Banks, E.; Fennell, T.; O'Donnell-Luria, A.H.; Ware, J.S.; Hill, A.J.; et al. Analysis of Protein-Coding Genetic Variation in 60,706 Humans. *Nature* **2016**, *536*, 285–291. [[CrossRef](#)]
52. Gudelj, I.; Lauc, G.; Pezer, M. Immunoglobulin G Glycosylation in Aging and Diseases. *Cell. Immunol.* **2018**, *333*, 65–79. [[CrossRef](#)]
53. Li, C.-W.; Lim, S.-O.; Chung, E.M.; Kim, Y.-S.; Park, A.H.; Yao, J.; Cha, J.-H.; Xia, W.; Chan, L.-C.; Kim, T.; et al. Eradication of Triple-Negative Breast Cancer Cells by Targeting Glycosylated PD-L1. *Cancer Cell* **2018**, *33*, 187–201.e10. [[CrossRef](#)]
54. Sasatomi, T.; Suefuji, Y.; Matsunaga, K.; Yamana, H.; Miyagi, Y.; Araki, Y.; Ogata, Y.; Itoh, K.; Shirouzu, K. Expression of Tumor Rejection Antigens in Colorectal Carcinomas. *Cancer* **2002**, *94*, 1636–1641. [[CrossRef](#)]
55. Chang, I.J.; He, M.; Lam, C.T. Congenital Disorders of Glycosylation. *Ann. Transl. Med.* **2018**, *6*, 477. [[CrossRef](#)] [[PubMed](#)]
56. Thiel, C.; Lübke, T.; Matthijs, G.; von Figura, K.; Körner, C. Targeted Disruption of the Mouse Phosphomannomutase 2 Gene Causes Early Embryonic Lethality. *Mol. Cell. Biol.* **2006**, *26*, 5615–5620. [[CrossRef](#)] [[PubMed](#)]
57. Chan, B.; Clasquin, M.; Smolen, G.A.; Histen, G.; Powe, J.; Chen, Y.; Lin, Z.; Lu, C.; Liu, Y.; Cang, Y.; et al. A Mouse Model of a Human Congenital Disorder of Glycosylation Caused by Loss of PMM2. *Hum. Mol. Genet.* **2016**, *25*, 2182–2193. [[CrossRef](#)] [[PubMed](#)]
58. Yu, X.; Lin, J.; Zack, D.J.; Qian, J. Computational Analysis of Tissue-Specific Combinatorial Gene Regulation: Predicting Interaction between Transcription Factors in Human Tissues. *Nucleic Acids Res.* **2006**, *34*, 4925–4936. [[CrossRef](#)] [[PubMed](#)]
59. Nairn, A.V.; Aoki, K.; dela Rosa, M.; Porterfield, M.; Lim, J.-M.; Kulik, M.; Pierce, J.M.; Wells, L.; Dalton, S.; Tiemeyer, M.; et al. Regulation of Glycan Structures in Murine Embryonic Stem Cells: Combined Transcript Profiling of Glycan-Related Genes and Glycan Structural Analysis. *J. Biol. Chem.* **2012**, *287*, 37835–37856. [[CrossRef](#)]
60. Yamakawa, N.; Vanbeselaere, J.; Chang, L.-Y.; Yu, S.-Y.; Ducrocq, L.; Harduin-Lepers, A.; Kurata, J.; Aoki-Kinoshita, K.F.; Sato, C.; Khoo, K.-H.; et al. Systems Glycomics of Adult Zebrafish Identifies Organ-Specific Sialylation and Glycosylation Patterns. *Nat. Commun.* **2018**, *9*, 4647. [[CrossRef](#)]
61. Williams, S.E.; Noel, M.; Lehoux, S.; Cetinbas, M.; Xavier, R.J.; Sadreyev, R.I.; Scolnick, E.M.; Smoller, J.W.; Cummings, R.D.; Mealer, R.G. Mammalian Brain Glycoproteins Exhibit Diminished Glycan Complexity Compared to Other Tissues. *Nat. Commun.* **2022**, *13*, 275. [[CrossRef](#)]
62. Kearney, C.J.; Vervoort, S.J.; Ramsbottom, K.M.; Todorovski, I.; Lelliott, E.J.; Zethoven, M.; Pijpers, L.; Martin, B.P.; Semple, T.; Martelotto, L.; et al. SUGAR-Seq Enables Simultaneous Detection of Glycans, Epitopes, and the Transcriptome in Single Cells. *Sci. Adv.* **2021**, *7*, eabe3610. [[CrossRef](#)]
63. Minoshima, F.; Ozaki, H.; Odaka, H.; Tateno, H. Integrated Analysis of Glycan and RNA in Single Cells. *iScience* **2021**, *24*, 102882. [[CrossRef](#)]
64. Lee, J.; Ha, S.; Kim, M.; Kim, S.-W.; Yun, J.; Ozcan, S.; Hwang, H.; Ji, I.J.; Yin, D.; Webster, M.J.; et al. Spatial and Temporal Diversity of Glycome Expression in Mammalian Brain. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 28743–28753. [[CrossRef](#)]
65. Kudelka, M.R.; Stowell, S.R.; Cummings, R.D.; Neish, A.S. Intestinal Epithelial Glycosylation in Homeostasis and Gut Microbiota Interactions in IBD. *Nat. Rev. Gastroenterol. Hepatol.* **2020**, *17*, 597–617. [[CrossRef](#)]
66. Bergstrom, K.S.B.; Xia, L. Mucin-Type O-Glycans and Their Roles in Intestinal Homeostasis. *Glycobiology* **2013**, *23*, 1026–1037. [[CrossRef](#)] [[PubMed](#)]
67. Schjoldager, K.T.; Narimatsu, Y.; Joshi, H.J.; Clausen, H. Global View of Human Protein Glycosylation Pathways and Functions. *Nat. Rev. Mol. Cell Biol.* **2020**, *21*, 729–749. [[CrossRef](#)] [[PubMed](#)]
68. Niccoli, T.; Partridge, L. Ageing as a Risk Factor for Disease. *Curr. Biol.* **2012**, *22*, R741–R752. [[CrossRef](#)]
69. Vanhooren, V.; Dewaele, S.; Libert, C.; Engelborghs, S.; De Deyn, P.P.; Toussaint, O.; Debacq-Chainiaux, F.; Poulain, M.; Glupczynski, Y.; Franceschi, C.; et al. Serum N-Glycan Profile Shift during Human Ageing. *Exp. Gerontol.* **2010**, *45*, 738–743. [[CrossRef](#)]
70. Gudelj, I.; Keser, T.; Vučković, F.; Škaro, V.; Goreta, S.Š.; Pavić, T.; Dumić, J.; Primorac, D.; Lauc, G.; Gornik, O. Estimation of Human Age Using N-Glycan Profiles from Bloodstains. *Int. J. Legal Med.* **2015**, *129*, 955–961. [[CrossRef](#)]
71. Horvath, S. DNA Methylation Age of Human Tissues and Cell Types. *Genome Biol.* **2013**, *14*, R115. [[CrossRef](#)] [[PubMed](#)]
72. Schneider, J.S. Altered Expression of Genes Involved in Ganglioside Biosynthesis in Substantia Nigra Neurons in Parkinson's Disease. *PLoS ONE* **2018**, *13*, e0199189. [[CrossRef](#)]
73. Waldera Lupa, D.M.; Kalfalah, F.; Safferling, K.; Boukamp, P.; Poschmann, G.; Volpi, E.; Götz-Rösch, C.; Bernerd, F.; Haag, L.; Huebenthal, U.; et al. Characterization of Skin Aging-Associated Secreted Proteins (SAASP) Produced by Dermal Fibroblasts Isolated from Intrinsically Aged Human Skin. *J. Investig. Dermatol.* **2015**, *135*, 1954–1968. [[CrossRef](#)]
74. Paton, B.; Suarez, M.; Herrero, P.; Canela, N. Glycosylation Biomarkers Associated with Age-Related Diseases and Current Methods for Glycan Analysis. *Int. J. Mol. Sci.* **2021**, *22*, 5788. [[CrossRef](#)]
75. Joshi, H.J.; Narimatsu, Y.; Schjoldager, K.T.; Tytgat, H.L.P.; Aebi, M.; Clausen, H.; Halim, A. SnapShot: O-Glycosylation Pathways across Kingdoms. *Cell* **2018**, *172*, 632–632.e2. [[CrossRef](#)] [[PubMed](#)]
76. Smith, J.C.; Sheltzer, J.M. Systematic Identification of Mutations and Copy Number Alterations Associated with Cancer Patient Prognosis. *eLife* **2018**, *7*, e39217. [[CrossRef](#)]
77. Ashkani, J.; Naidoo, K.J. Glycosyltransferase Gene Expression Profiles Classify Cancer Types and Propose Prognostic Subtypes. *Sci. Rep.* **2016**, *6*, 26451. [[CrossRef](#)] [[PubMed](#)]

- 
78. Costa, A.F.; Campos, D.; Reis, C.A.; Gomes, C. Targeting Glycosylation: A New Road for Cancer Drug Discovery. *Trends Cancer* **2020**, *6*, 757–766. [[CrossRef](#)]
  79. Silva, M.; Silva, Z.; Marques, G.; Ferro, T.; Gonçalves, M.; Monteiro, M.; van Vliet, S.J.; Mohr, E.; Lino, A.C.; Fernandes, A.R.; et al. Sialic Acid Removal from Dendritic Cells Improves Antigen Cross-Presentation and Boosts Anti-Tumor Immune Responses. *Oncotarget* **2016**, *7*, 41053. [[CrossRef](#)] [[PubMed](#)]
  80. Silva, Z.; Ferro, T.; Almeida, D.; Soares, H.; Ferreira, J.A.; Deschepper, F.M.; Hensbergen, P.J.; Pirro, M.; van Vliet, S.J.; Springer, S.; et al. MHC Class I Stability Is Modulated by Cell Surface Sialylation in Human Dendritic Cells. *Pharmaceutics* **2020**, *12*, 249. [[CrossRef](#)] [[PubMed](#)]