Evolutionary rate covariation is pervasive between glycosylation pathways and points to potential disease modifiers

ERC identifies potential genetic modifiers of glycosylation

Holly J. Thorpe[1], Raghavendran Partha[2], Jordan Little[1], Nathan L. Clark[2], Clement Y. Chow[1*]

[1] Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, Utah, United States of America

[2] Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA, United States of America

***cchow@genetics.utah.edu**

ABSTRACT

Mutations in glycosylation pathways, such as N-linked glycosylation, O-linked glycosylation, and GPI anchor synthesis, lead to Congenital Disorders of Glycosylation (CDG). CDGs typically present with seizures, hypotonia, and developmental delay but display large clinical variability with symptoms affecting every system in the body. This variability suggests modifier genes might influence the phenotypes. Because of the similar physiology and clinical symptoms, there are likely common genetic modifiers between CDGs. Here, we use evolution as a tool to identify common modifiers between CDG and glycosylation genes. Protein glycosylation is evolutionarily conserved from yeast to mammals. Evolutionary rate covariation (ERC) identifies proteins with similar evolutionary rates that indicate shared biological functions and pathways. Using ERC, we identified strong evolutionary rate signatures between proteins in the same and different glycosylation pathways. Genome-wide analysis of proteins showing significant ERC with GPI anchor synthesis proteins revealed strong signatures with ncRNA modification proteins and DNA repair proteins. We also identified strong patterns of ERC based on cellular sub-localization of the GPI anchor synthesis enzymes. Functional testing of the highest scoring candidates validated genetic interactions and identified novel genetic modifiers of CDG genes. ERC analysis of disease genes and biological pathways allows for rapid prioritization of potential genetic modifiers, which can provide a better understanding of disease pathophysiology and novel therapeutic targets.

AUTHOR SUMMARY

Congenital Disorders of Glycosylation (CDGs) are a group of rare disorders resulting from impaired protein glycosylation. Glycosylation is the addition of sugar chains onto proteins and is required for proper protein function. CDG patients typically present with seizures and hypotonia. However, they can have a large amount of clinical variability, which is likely influenced by modifier genes. Modifier genes are genes that affect a phenotype without causing the disease. Using an evolutionary method that examines proteins that evolve at similar rates, we identified proteins

53 within glycosylation pathways and among other unexpected pathways, such as ncRNA
54 modification and DNA repair, that could be potential genetic modifiers of CDG genes. We also
55 tested top protein pairs using the *Drosophila* eye as a model and identified novel genetic modifiers
56 of CDG genes. Broadening our understanding of CDG modifiers can help us to better understand
57 why loss of glycosylation results in specific patient symptoms and could provide new treatment
58 targets.
59
60 **INTRODUCTION**
61 Phenotypic variation is common in Mendelian diseases, even in patients with the same disease-
62 causing variant (1–3). Understanding the factors underlying variable disease presentation can be
63 difficult despite often knowing the causal disease gene. Genetic modifiers can enhance or
64 suppress the phenotype of a primary disease-causing mutation, but they may not independently
65 lead to disease phenotypes.
66
67 Congenital disorders of glycosylation (CDG) are a group of ultra-rare multisystemic disorders
68 typically characterized by seizures, hypotonia, and neurodevelopmental delays (4,5). However,
69 because of the ubiquity of glycosylation, symptoms of CDGs can affect most organ systems. Loss-
70 of-function mutations in >150 glycosylation genes lead to CDGs (6). Patients with the same CDG
71 and even the same disease-causing variant can show phenotypic variability, indicating that
72 genetic background and environment can influence clinical presentation (7,8). Current treatment
73 options for CDGs are limited, primarily focusing on symptom management (9,10). Identifying
74 genetic modifiers of glycosylation genes could lead to new therapeutic targets for CDG patients.
75
76 Protein glycosylation is essential for proper protein folding, stability, and localization (4,5,11).
77 There are three main types of glycosylation: N-linked glycosylation, O-linked glycosylation, and
78 glycosylphosphatidylinositol (GPI) anchor biosynthesis. N-linked glycosylation is the addition of a

79 glycan onto an asparagine (4,5). The glycan is synthesized on the endoplasmic reticulum (ER)
80 membrane, with the stepwise addition of sugars onto a dolichol-phosphate. The glycan is then
81 transferred from the dolichol-phosphate onto a protein and undergoes further glycan additions
82 and trimming in the ER and Golgi. O-linked glycosylation is the addition of a glycan to the hydroxyl
83 group of a serine, threonine, or hydroxylysine (4). O-linked glycosylation encompasses the most
84 extensive variety of glycans and can occur in the ER, Golgi, or cytoplasm (12). O-linked glycans
85 are classified by the first sugar attached to the protein. The rest of the glycan is built directly onto
86 that sugar.
87
88 GPI anchors are glycolipids that tether proteins to the cell membrane (8,13). GPI anchor synthesis
89 begins by adding an N-acetylglucosamine (GlcNAc) onto a phosphatidylinositol on the
90 cytoplasmic side of the ER membrane. The acyl group is then removed before the glycan is flipped
91 onto the lumen side of the ER membrane. In the ER lumen, three mannoses (four in some
92 species/tissues) and three phosphoethanolamine groups are added before a protein is attached
93 to the GPI anchor. The GPI anchor then undergoes fatty acid remodeling before it is sent to the
94 Golgi for further remodeling. Then, the GPI-anchored glycoprotein is trafficked to the cell surface.
95
96 Standard methods for identifying genetic modifiers typically involve utilizing natural variation in
97 the population or performing large mutagenesis screens (1). Several recent studies reported
98 modifier screens in animal models of specific CDGs (14–16). However, with over 150 CDGs,
99 performing similar studies for each CDG is impractical. PMM2-CDG, the most common CDG, has
100 just over 1000 diagnosed patients worldwide (17). Most CDGs, however, have fewer than 100
101 patients, making it challenging to have enough statistical power to distinguish impactful variation
102 within the population (8,18). It is likely CDGs share common genetic modifiers because different
103 forms can have similar effects on glycosylation and similar clinical presentations. Instead of
104 examining each CDG independently, we sought to identify common modifier genes for all or a

105 subset of CDGs. To do this, we used evolutionary rate covariation (ERC), a computational
106 evolutionary method developed to identify proteins with evolutionary rates that covary across
107 species (19). Protein pairs with high ERC values have similar changes in evolutionary rate and
108 are thought to function together, either in a complex or in a pathway, and may genetically modify
109 each other. ERC allows us to identify proteins with evolutionary rates that covary with CDG
110 proteins, potentially identifying genetic modifiers.
111
112 A number of recent studies suggest that other glycosylation genes modify CDG genes. In 2002,
113 a study examined a small group of PMM2-CDG patients, the most common CDG, and identified
114 a variant in *ALG6*, an N-linked glycosylation gene, segregating with more severe outcomes (20).
115 A modifier screen and ERC analysis of NGLY1, a deglycosylating enzyme associated with a CDG,
116 identified multiple glycosylation genes showing high ERC with NGLY1, implicating them as
117 potential modifiers (15). A CRISPR screen in cells with reduced DPAGT1 function, the first step
118 in N-glycan synthesis, identified multiple glycosylation genes able to rescue glycosylation defects
119 caused by inhibition of DPAGT1 (14). In yet another study, a yeast model of PMM2-CDG was
120 evolved over 1000 generations to identify genetic changes that increased the viability of the
121 *PMM2* mutant yeast strains (16). Several top genes that accumulated compensatory mutations
122 were also involved in glycosylation. These studies indicate that genetic variation in glycosylation
123 genes might potentially modify the outcomes of CDGs.
124
125 Applying a computational evolutionary method allows for more rapid and broad detection of
126 disease modifiers. Identifying genetic modifiers of CDGs will lead to a better understanding of the
127 physiology underlying variable symptoms in patients and could point to new therapeutic targets.
128 Using ERC, we performed an analysis to discover proteins that have correlated evolutionary rates
129 with glycosylation proteins to discover genetic modifiers of CDGs. We found strong intra-pathway
130 ERC signatures among glycosylation proteins in each of the three glycosylation pathways, with

131 the highest average score in GPI anchor synthesis. There were also high inter-pathway ERC
132 values between proteins in different glycosylation pathways. Strikingly, GPI anchor synthesis
133 proteins had many high ERC values with N-linked glycosylation proteins. Protein complexes and
134 proteins with similar functions in glycosylation were also enriched for strong ERC. When
135 examining the genome for non-glycosylation modifiers of GPI anchor synthesis proteins, we found
136 high ERC scores with proteins involved in multiple unrelated pathways, including ncRNA and DNA
137 repair. We found that the strength of these ERC scores is driven by the GPI anchor synthesis
138 enzyme's location in the cell. *In vivo* testing of top scoring protein pairs, including GPAA1 with
139 ALG1 and PIGA with RBSN, showed a genetic interaction in a *Drosophila* eye model. Using
140 computational evolutionary methods, we identified patterns of ERC between CDG proteins and
141 novel potential modifier proteins, many of which validated *in vivo.*
142
143 **RESULTS**
144 **High ERC is pervasive between glycosylation proteins**
145 To find potential candidate genetic modifiers of CDGs, we created a dataset of ERC scores across
146 the genome. Using evolutionary rate data from yeast, nematode worms, *Drosophila*, vertebrates,
147 and mammals (S1 Fig A), we calculated ERC scores for 19,149 proteins across the genome
148 resulting in 180,112,645 pairwise ERC values (S1 Fig B). The genome-wide dataset values range
149 from -10.73 – 19.30, with an average of 0.12 ± 1.47. Higher values indicate stronger covariation
150 of evolutionary rates. In this study, ERC values ≥ 3 are considered notably elevated because they
151 are two standard deviations from the mean. Using this dataset, we compared ERC scores
152 between subsets of proteins.
153
154 We first examined the pairwise correlated evolutionary rates among all glycosylation proteins. We
155 included 296 glycosylation proteins in our analyses (43,660 protein pairs), including all known
156 CDG proteins and glycosylation proteins not yet associated with a disease (S1 Data). These

157 include enzymes and support proteins involved in synthesizing and modifying glycans, enzymes
158 required to form glycosylation building blocks, and proteins involved in transporting glycoproteins
159 throughout the cell. We calculated ERC scores between all 296 glycosylation proteins (S2 Fig,
160 S1 Data). Glycosylation protein pairs showed higher average values of ERC than expected by
161 chance (mean ERC = 0.41, $p < 1x10^{-5}$ Permutation test) (S3 Fig, Table 1). The mean ERC score
162 for glycosylation protein pairs is more than 12 standard deviations above the mean of the
163 permutations.

Table 1: Average pairwise ERC scores in glycosylation pathway

Pathway	Proteins in pathway	Pairwise scores	Average ERC score	P-value
All Glycosylation	296	43,660	0.41	$<1x10^{-5}$
GPI Anchor Synthesis	27	351	1.69	$<1x10^{-5}$
N-linked Glycosylation	71	2,485	0.64	$<1x10^{-5}$
O-linked Glycosylation	122	7,381	0.31	$<1x10^{-5}$
Other	76	2,850	0.47	$<1x10^{-5}$

Average of the pairwise ERC scores between the proteins in each pathway. Significance tested against 100,000 permutations.

164

165 Proteins in a pathway are often under similar evolutionary constraints and typically have high ERC
166 values (19). To determine if each of the individual glycosylation pathways exhibited even higher
167 ERC than when all glycosylation proteins are considered, we analyzed pairwise ERC of proteins
168 in each major glycosylation pathway (GPI anchor synthesis, N-linked glycosylation, and O-linked
169 glycosylation) as well as proteins that contribute to multiple glycosylation pathways (Other) (Table
170 1). As before, permutation testing was used to determine the significance of the ERC values. GPI
171 anchor synthesis had a mean ERC score of 1.69 (27 proteins, $p < 1x10^{-5}$) (Figure 1), N-linked
172 glycosylation had a mean ERC score of 0.64 (71 proteins, $p < 1x10^{-5}$) (S4 Fig A), and O-linked
173 glycosylation had a mean ERC score of 0.31 (122 proteins, $p < 1x10^{-5}$) (S4 Fig B). The proteins
174 involved in multiple pathways (Other) also showed higher than expected ERC with an average of

175 0.47 (76 proteins, $p < 1x10^{-5}$) (S4 Fig C). Strikingly, the average score of the GPI anchor synthesis
176 pathway was nearly four times as high as the average of all glycosylation proteins (GPI average
177 = 1.69, all glycosylation average =
178 0.41). O-linked glycosylation (average
179 ERC = 0.31) is the only pathway with a
180 lower average ERC score than the
181 overall glycosylation average. This
182 lower score is likely because O-linked
183 glycosylation comprises eleven
184 independent pathways that form
185 different types of O-linked glycans,
186 reducing the number of proteins in this
187 category that work directly with each
188 other.

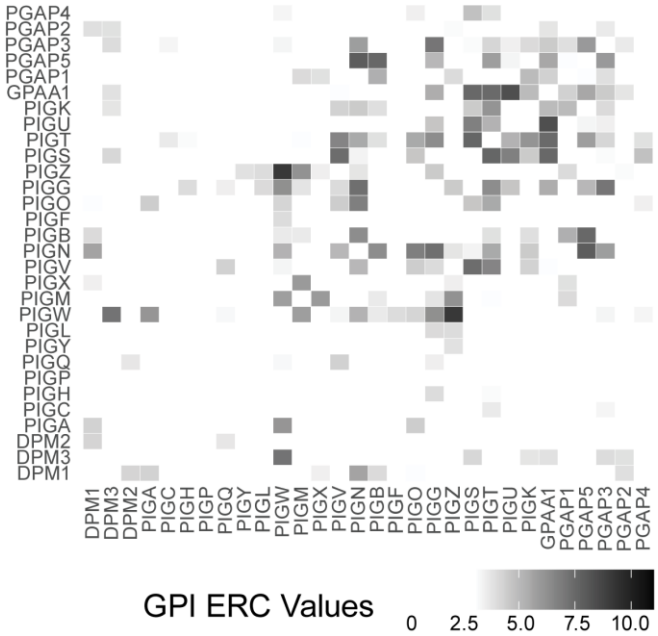

Figure 1: Heatmap showing significant (>3) pairwise ERC scores between proteins involved in GPI anchor synthesis.

189

190 **Strong evolutionary covariation between proteins in different glycosylation pathways**

191 Glycosylation is a broad term encompassing multiple pathways with hundreds of proteins in
192 different organelles. While some proteins associated with CDGs impact different types of glycans,
193 nearly all enzymatic proteins are specific to one of the three main glycosylation pathways. We
194 asked if the high mean ERC score observed across all glycosylation proteins (mean ERC = 0.41,
195 $p < 1x10^{-5}$) was driven primarily by high scores between proteins in the same pathway or if high
196 ERC scores between proteins in different pathways were also observed. Proteins in all three
197 glycosylation pathways showed strong ERC with proteins in different pathways (Table 2, S5 Fig,
198 S6 Fig). As within the pathway, GPI anchor synthesis had the highest scores with other
199 glycosylation pathways. The average ERC score between proteins in GPI anchor synthesis and
200 proteins in N-linked glycosylation is 0.81 ($p < 1x10^{-5}$), O-linked glycosylation is 0.46 ($p < 0.014$),

201 and Other is 0.61 ($p < 1.30^{-4}$). The average scores between GPI anchor synthesis and either N-
202 linked, O-linked glycosylation, or Other are higher than the average of all glycosylation proteins
203 (mean ERC = 0.41). This strong signal between glycosylation proteins in different pathways,
204 especially those in GPI anchor synthesis and N-linked glycosylation, could indicate unexpected
205 interactions between pathways.

Table 2: Average ERC values between glycosylation pathways

	GPI	N-linked	O-linked	Other
GPI	1.69 $p < 1x10^{-5}$	-	-	-
N-linked	0.81 $p < 1x10^{-5}$	0.64 $p < 1x10^{-5}$	-	-
O-linked	0.46 $p < 1.44x10^{-2}$	0.42 $p < 1x10^{-5}$	0.31 $p < 1x10^{-5}$	-
Other	0.61 $p < 1.30\ x10^{-4}$	0.34 $p < 1.80x10^{-3}$	0.26 $p = 4.87x10^{-3}$	0.47 $p < 1x10^{-5}$

Average of the pairwise ERC scores between the proteins in different glycosylation pathways. Significance tested against 100,000 permutations.

206

207 **Not all glycosylation complexes have high correlated evolutionary rates**

208 We next examined whether glycosylation proteins that form physical complexes have high ERC
209 values. The GlcNAc transferase complex (GlcNAc-T) performs the first step in GPI anchor
210 synthesis, attaching a GlcNAc to the phosphatidylinositol base (13). GlcNAc-T consists of seven
211 proteins: PIGA, PIGC, PIGH, PIGP, PIGQ, PIGY, and DPM2. The proteins in the GlcNAc-T
212 complex do not have enriched rate covariation with each other (mean ERC = 0.18, p = 0.41)
213 (Figure 2A, S7 Fig A). Only one pair of proteins shows significant ERC, PIGQ and DPM2 (ERC =
214 3.58). The low ERC scores between proteins in this complex suggest that physical interaction
215 does not drive evolutionary rate changes in these proteins. The other complex in GPI anchor
216 synthesis is the GPI transamidase complex which is comprised of five proteins: PIGS, PIGT,

217 PIGU, PIGK, and GPAA1 (13). The GPI transamidase complex attaches the protein to the
218 synthesized GPI anchor. The mean ERC score among proteins in the GPI transamidase complex
219 is 5.86 ($p < 1x10^{-5}$), which is more than 11 standard deviations higher than the average expected
220 based on permutation testing (permutations mean = 0.12) (Figure 2B, S7 Fig B). All pairwise ERC
221 scores between GPI transamidase proteins show significant rate correlation (ERC ≥ 3), except
222 between PIGU and PIGK (ERC = 0.80).

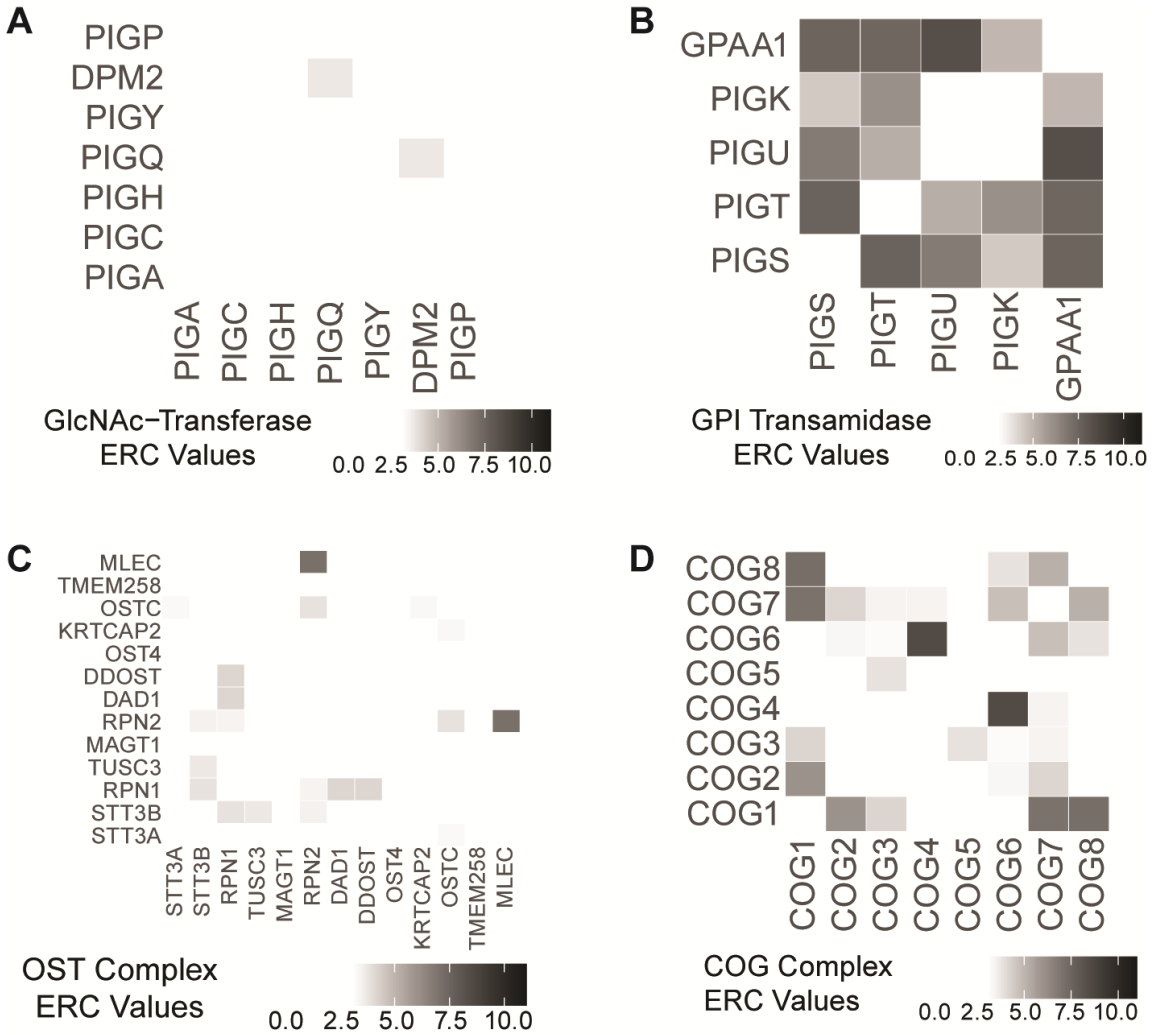

Figure 2: Heatmaps of significant pairwise ERC values for proteins in complexes involved glycosylation. The GlcNAc-T complex (A) does not show significant ERC ($p = 0.41$). The GPI transamidase complex (B), the OST complex (C), and the COG complex (D) all have significant ERC ($p < 1x10^{-4}$).

223 In N-linked glycosylation, the oligosaccharyltransferase (OST) complex attaches the synthesized
224 glycan onto a protein (11). This complex is made up of 13 proteins. The average pairwise ERC
225 score of the OST complex is 1.26 and is higher than expected based on permutation testing ($p =$
226 4.0×10^{-5}) (Figure 2C, S7 Fig C). The highest value is between MLEC and RPN2 (ERC = 7.02).
227 Both MAGT1 and OST4 have no significant ERC scores with any other proteins in the complex.
228 MAGT1 acts as a scaffold protein for STT3B in the OST complex, but it also functions as a
229 magnesium transporter (21). Because MAGT1 has functions outside of this complex, it might have
230 multiple evolutionary pressures that reduce its covariation rate with OST complex proteins. OST4
231 plays a critical role in the OST complex, binding the catalytic subunits STT3A or STT3B to the
232 rest of the complex (22). Despite this, it does not show strong ERC with OST complex components
233 suggesting that it might have another unknown, unrelated function.
234
235 The COG complex is a group of eight proteins (COG1-8) that acts as a tethering factor in intra-
236 Golgi trafficking (23). Pairwise ERC scores between the proteins in the COG complex show
237 significant ERC (mean ERC = 3.00, $p < 1 \times 10^{-5}$) (Figure 2D, S7 Fig D). Only half of the protein
238 pairs show a significant correlation (ERC ≥ 3). The COG complex is organized into two lobes, with
239 COG1-4 making up one lobe and COG5-8 making up the second lobe. Within the first lobe, COG1
240 has significant ERC with both COG2 and COG3. Strikingly, COG4 does not have any significant
241 ERC scores with the other proteins in the same lobe; however, COG4 has the highest ERC value
242 of the whole complex with COG6 (ERC = 8.46), a protein in the opposite lobe. The second lobe
243 has slightly more protein pairs with significant ERC. COG6, COG7, and COG8 all have significant
244 ERC with each other. COG5 has no significant ERC with other proteins in the same lobe. A bond
245 between COG1 and COG8 connects the two lobes. COG1 and COG8 have the second highest
246 ERC score of the complex (ERC = 7.16). The average ERC score between proteins in the first
247 lobe is 2.70, and the average ERC score between proteins in the second lobe is 2.51.
248 Unexpectedly, the average ERC score among inter-lobe protein pairs is 3.29. Strong ERC scores

249 in this complex are not solely driven by direct physical contact between the proteins and are, on
250 average, higher between proteins in different lobes.
251
252 **Glycosylation enzymes with similar functions have increased ERC**
253 Next, we examined whether glycosylation proteins with similar enzymatic functions, but from
254 different pathways display high ERC. There are twelve mannosyltransferases across the three
255 glycosylation pathways. These mannosyltransferases are not known to physically interact with
256 each other, and each one functions in only one specific glycosylation pathway. Pairwise ERC
257 scores between these twelve mannosyltransferases show significant ERC by permutation testing
258 (mean ERC = 3.07, $p < 1x10^{-5}$) (Figure 3, S7 Fig E). Of these 55 ERC values, 27 are above the
259 significance cut-off (≥3). The top score is between
260 ALG12 and ALG2 (ERC = 8.02), which are both in
261 the N-linked glycosylation pathway. The second
262 highest score is between POMT1, involved in O-
263 mannosylation, and ALG2, involved in N-linked
264 glycosylation (ERC = 7.97). The six
265 mannosyltransferases involved in N-linked
266 glycosylation have an average ERC of 4.40. The
267 four GPI anchor synthesis specific
268 mannosyltransferases have an average of 2.32.
269 O-linked glycosylation has only two
270 mannosyltransferases (POMT1 and POMT2), and
271 their average ERC score is 6.86.

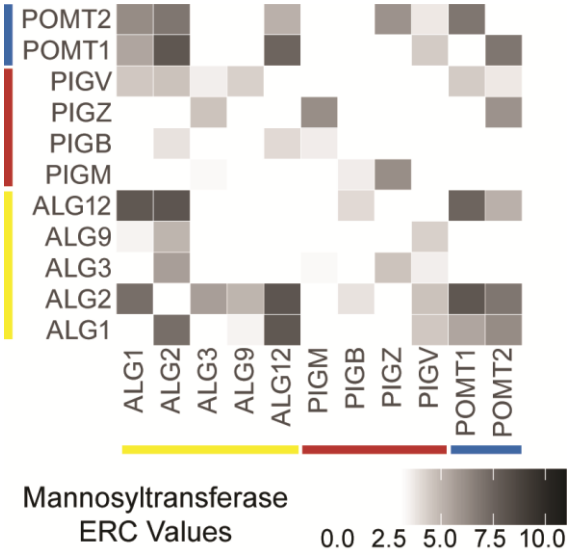

Figure 3: Heatmap showing significant ERC values between mannosyltransferases. N-linked glycosylation (yellow), GPI-anchor synthesis (red), O-linked glycosylation (blue). Mannosyltransferases have significant ERC with each other across pathways ($p < 1x10^{-5}$).

272
273 **Glycosylation protein pairs with strongest ERC**

274 Some of the top ten highest-scoring glycosylation protein pairs show overlapping functions, which
275 could explain the high ERC scores (Table 3). For example, COG3 and TRIP11 (ERC = 9.62) are
276 proteins bound to the outer membrane of the Golgi involved in vesicle tethering and assembly
277 and maintenance of the Golgi. PIGZ and PIGW also have a high ERC score (ERC = 9.10). PIGZ
278 and PIGW are both involved in GPI anchor synthesis in the ER, and they share a common GPI
279 recognition domain in their transmembrane domains (24). While these two pairs have high ERC
280 scores supported by current knowledge, the top scores also include pairs with high scores that
281 are not easily explained based on our current knowledge of their function. FKRP, an enzyme that
282 adds ribitol to O-linked alpha-dystroglycans in the Golgi (25), and MOGS, an enzyme that cleaves
283 glucose from the synthesized N-glycan in the
284 ER (26), have an ERC score of 10.63. EXTL3,
285 a GlcNAc transferase for O-linked
286 glycosaminoglycan synthesis located in the
287 Golgi (27), and GMPPA, a proposed inhibitor of
288 GDP-mannose synthesis localized to the
289 cytoplasm (28), have an ERC score of 9.62.
290 The remaining six pairs in the top 10 show
291 similar patterns of high ERC with no obvious
292 overlap in function, localization, or glycosylation
293 pathway. These unrelated pairs of proteins with
294 high ERC values could signal novel interactions
295 between glycosylation proteins.

Table 3: Top 10 pairwise ERC scores among glycosylation proteins

Protein 1	Protein 2	ERC score
MOGS	FKRP	10.63
EXTL3	GMPPA	9.62
DOLK	PIGG	9.59
TRIP11	PIGG	9.50
TRIP11	COG3	9.48
GNPTAB	ALG6	9.38
ALG1	GPAA1	9.29
B3GNT1	DOLK	9.18
PIGZ	PIGW	9.10
COG2	PIGG	9.00

296

297 **Genome-wide patterns of evolutionary rate correlation for GPI Anchor Synthesis Proteins**

298 We next examined whether proteins with unrelated, non-glycosylation functions might also have
299 high ERC values with glycosylation proteins. To do this, we extended our analyses to a full

genome-wide comparison. Because the GPI anchor synthesis pathway had the highest average ERC score of all glycosylation pathways (GPI mean ERC = 1.61), we focused the genome-wide analyses on the GPI anchor biosynthesis pathway. We included the 27 proteins specific to the GPI anchor synthesis pathway and DPM1, DPM2, and DPM3. DPM2 is a component of the GlcNAc-transferase complex in GPI anchor synthesis and the dolichol-phosphate-mannose (DPM) synthase complex with DPM1 and DPM3. DPM synthase is required to add mannose to the GPI anchor (29). We calculated pairwise ERC values for each of the 30 proteins involved in GPI anchor biosynthesis with all 19,149 proteins in our dataset (Figure 4, S1 Data). This resulted in 568,438 pairwise values ranging from -8.97 to 11.28, with an average of 0.32. 31,483 protein pairs had significant ERC scores (ERC ≥ 3).

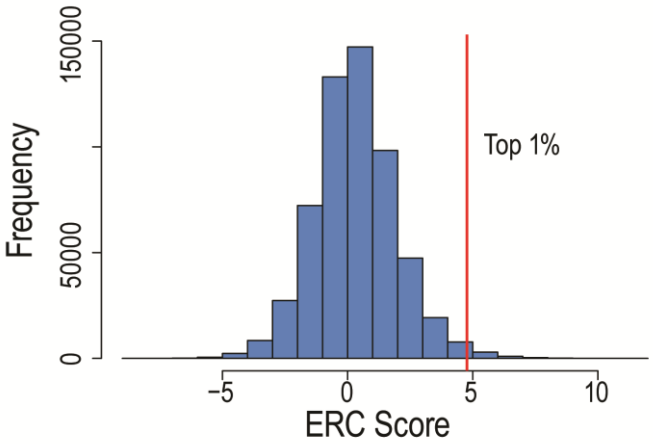

Figure 4: Histogram of genome-wide ERC values for with GPI anchor synthesis proteins. Vertical red line indicates top 1% of values.

Gene Ontology enrichment of proteins showing high ERC with GPI Anchor Synthesis

Gene Ontology (GO) analysis of the top 1% of GPI anchor synthesis genome-wide ERC values (5685 values, ERC > 4.77, 3386 proteins) (Figure 5, S2 Data) revealed enrichment in two glycosylation related pathways, *GPI anchor biosynthetic process GO:0006506* (fold enrichment = 3.86, q = 1.03×10^{-3}) and *dolichol-linked oligosaccharide biosynthetic process GO:0006488* (fold enrichment = 4.70, q= 1.36×10^{-3}), which support the finding described above that GPI anchor synthesis proteins show strong signals with themselves and other glycosylation pathways.

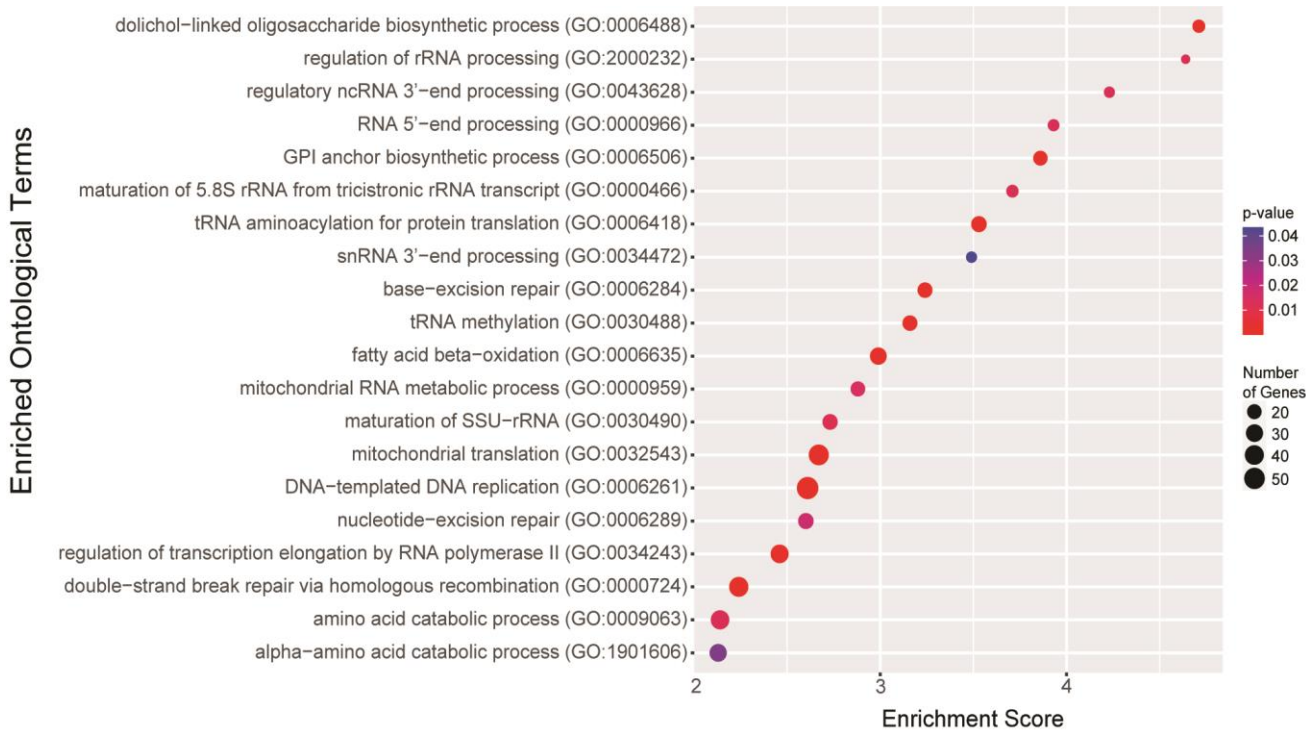

Figure 5: GO terms from analysis of proteins in the top 1% of GPI anchor synthesis genome-wide ERC values. P-values are indicated by red-to-blue gradient, with red the lowest p-values and blue the highest p-values. Gene number identified in each category is indicated by the size of the circle.

GO analysis also returned multiple terms relating to ncRNA modification: *regulation of rRNA processing GO:2000232* (fold enrichment = 4.63, q = 1.37×10^{-2}), *regulatory ncRNA 3'-end processing GO:0043628* (fold enrichment = 4.23, q = 1.34×10^{-2}), and *tRNA methylation GO:0030488* (fold enrichment = 3.16, q = 3.48×10^{-3}) (Figure 5, S2 Data). Previous work from our lab also identified ncRNA processing proteins as potential modifiers of NGLY1-CDG (15). The enrichment of ncRNA processing proteins in this analysis further supports an unexplored connection between glycosylation proteins and ncRNAs. There were also multiple GO terms related to DNA repair: *base excision repair GO:006284* (fold enrichment = 3.24, q = 2.71×10^{-3}), *nucleotide-excision repair GO:0006289* (fold enrichment = 2.60, q = 1.83×10^{-2}), and *double-strand break repair via homologous recombination GO:0000724* (fold enrichment = 2.24, q = 2.27×10^{-3}). This enrichment was unexpected as no previous evidence links DNA repair to GPI anchor synthesis and suggests a possible interaction between the two pathways.

Loss of function in most GPI anchor synthesis proteins leads to a CDG. To identify potential proteins that could modify multiple CDG genes across the pathway, we also ran GO analysis on proteins that had a significant ERC score with at least a third (10 or more) of the 30 GPI anchor synthesis proteins (Figure 6, S2 Data). Unsurprisingly, GO analysis again showed high enrichment for glycosylation proteins (*GPI anchor biosynthetic process* GO:0006506, fold enrichment 18.53, q = 1.59×10^{-4}; *dolichol-linked oligosaccharide biosynthetic process* GO:0006488, fold enrichment 15.63, q = 4.48×10^{-2}) as well as terms related to ncRNA (*tRNA aminoacylation for protein translation GO:0006418*, fold enrichment = 15.63, q = 8.10×10^{-5}; *tRNA methylation GO:0030488*, fold enrichment = 11.45, q = 7.69×10^{-3}; *rRNA processing GO:0006364*, fold enrichment = 4.52, q = 7.26×10^{-3}). Unlike what was found when analyzing the top 1% of values, there was also enrichment for ribosome assembly (*ribosomal small subunit biogenesis GO:0042274*, fold enrichment = 8.51, q = 9.95×10^{-5}; *ribosome assembly GO:0042255,* fold enrichment = 7.46, q = 4.73×10^{-2}) and m*itotic recombination GO:0006312* (fold enrichment =

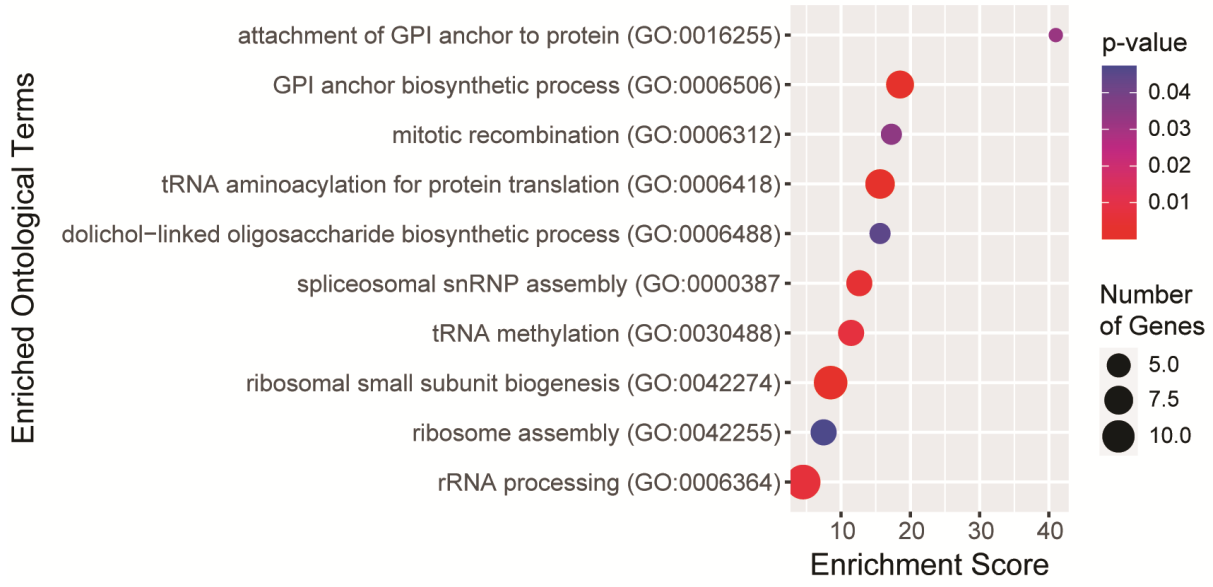

Figure 6: GO terms from analysis of proteins with significant ERC values with at least 10 GPI anchor synthesis proteins. P-values are indicated by red-to-blue gradient, with red the lowest p-values and blue the highest p-values. Gene number identified in each category is indicated by the size of the circle.

351 17.27, q = 3.43×10^{-2}). These pathways interacting
352 with multiple proteins in GPI anchor synthesis may
353 represent compelling therapeutic targets for multiple
354 GPI anchor deficiencies.
355
356 **Top scoring pairs across the genome**
357 Many top-scoring pairs from the genome-wide
358 analysis fall under the categories enriched in the GO
359 analyses. Four of the top 30 scoring pairs were
360 between known glycosylation proteins (Table 4).
361 PIGN, a phosphoethanolamine transferase (13), had
362 high ERC with MAN2B2, a protein involved in glycan
363 recycling (ERC = 10.01) (10). PIGG, a
364 phosphoethanolamine transferase (13), had strong
365 ERC with DOLK, a kinase that creates dolichol-
366 phosphate upon which the N-glycan is built (ERC =
367 9.59) (30), and TRIP11, a protein involved in Golgi
368 vesicle tethering (ERC = 9.50) (31). GPAA1, a
369 member of the GPI transamidase complex (13), had
370 high ERC with ALG1, a mannosyltransferase in N-
371 linked glycosylation (ERC = 9.29) (11). All seven
372 glycosylation proteins in these top-scoring pairs are
373 associated with their own CDGs, further suggesting
374 that known CDG genes are likely candidates for
375 genetic modifiers of other CDGs (14–16,20).
376

Table 4: Top scoring protein pairs from genome-wide analysis of GPI anchor synthesis proteins.

Rank	GPI Protein	Protein 2	ERC score
1	PIGG	HTT	11.28
2	PIGQ	WDR90	10.50
3	PIGG	TUBGCP2	10.49
4	PIGG	ATG7	10.44
5	PIGB	UTP20	10.33
6	PIGW	DHODH	10.32
7	PIGG	TTC8	10.22
8	PIGO	ZNF622	10.19
9	PIGB	AP4E1	10.09
10	PIGG	PRKDC	10.01
11	PIGN	MAN2B2	10.01
12	GPAA1	MCM7	9.97
13	PIGG	ERCC6	9.76
14	PIGN	WDR36	9.74
15	PIGG	DOLK	9.59
16	PIGA	ZFYVE20	9.58
17	PIGZ	C12orf65	9.57
18	PIGB	GEMIN5	9.57
19	PIGG	TRIP11	9.50
20	PIGG	GTF3C5	9.47
21	PIGM	INF2	9.45
22	PIGQ	PKD1	9.40
23	PIGN	CTDP1	9.39
24	PGAP4	DHX40	9.38
25	PIGG	PHRF1	9.36
26	PIGB	MSH6	9.33
27	PIGO	SHPRH	9.31
28	GPAA1	ALG1	9.29
29	PIGG	AP5Z1	9.26
30	PGAP1	SMYD3	9.22

377 Top scores in our dataset revealed high ERC between PIGB and UTP20 (ERC = 10.33) and PIGN
378 and WDR36 (ERC = 9.74). UTP20 and WDR36 are both involved in rRNA processing (32,33).
379 Top scores also included PIGG and PRKDC (ERC = 10.01), a DNA damage sensing protein (34),
380 ERCC6 (ERC = 9.76), a protein important for transcription-coupled excision repair (35), and
381 AP5Z1 (ERC = 9.26), a protein involved in homologous recombination DNA repair (36). GPAA1
382 had high ERC with MCM7 (ERC = 9.97), a protein that signals DNA damage (37). These top pairs
383 further support the hypothesis of a relationship between GPI anchor synthesis genes and the GO
384 terms identified and offer targeted protein pairs to begin testing to explore these pathways.
385
386 **Cytoplasmic GPI anchor synthesis proteins have lower ERC**
387 Among the top 1% of genome-wide GPI anchor scores, nearly 10% are with PIGG (552/5685).
388 To determine if this enrichment was significant, we compared the actual number of ERC values
389 each GPI anchor synthesis protein had in the top 1% to an expected number of values, assuming
390 an equal distribution of all the GPI anchor synthesis proteins (expected ~189 occurrences per
391 protein) (Figure 7A, S1 Table). PIGG (552 occurrences, 2.91 fold change, $p < 2.2x10^{-16}$) was the
392 most overrepresented, with nearly three times as many ERC values in the top 1% as expected,
393 followed by PIGW (526 occurrences, 2.78 fold change, $p < 2.2x10^{-16}$). On the other hand, PIGY
394 (8 occurrences, -23.68 fold change, $p < 2.2x10^{-16}$), PIGH (7 occurrences, -27.06 fold change, p <
395 $2.2x10^{-16}$), and DPM2 (6 occurrences, -31.57 fold change, $p < 2.2x10^{-16}$) all have more than 20
396 fold less than the expected frequency. We tested whether this enrichment/depletion followed the
397 pathway order (Figure 7B, 7C). Analyses of the GPI anchor synthesis proteins by pathway order
398 demonstrated that 8/10 proteins on the cytoplasmic side are underrepresented, and none are
399 overrepresented. In the ER lumen, only 4/17 proteins are underrepresented, and 9/17 of the
400 proteins are overrepresented.
401

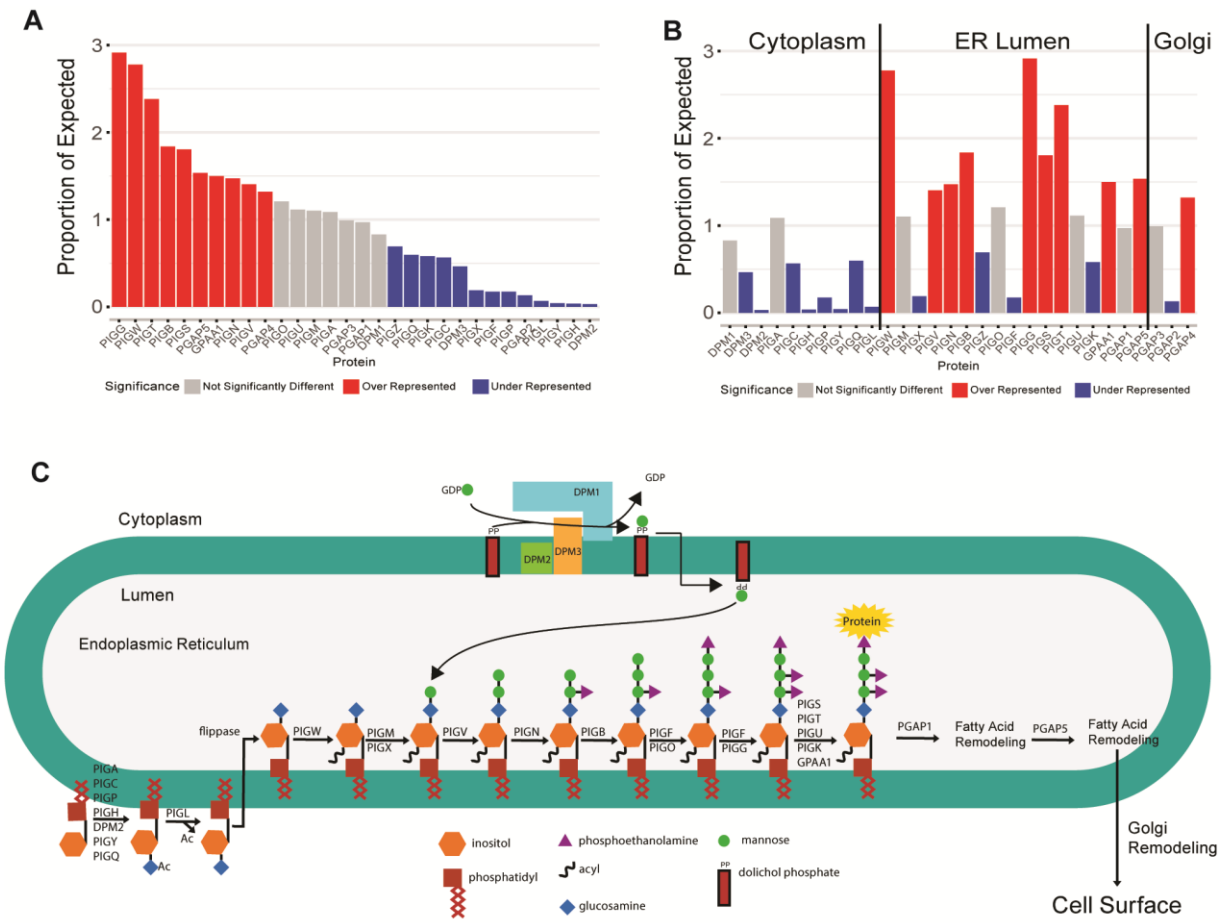

Figure 7: Proportion of expected occurrences of all GPI anchor synthesis proteins in the top 1% of genome-wide scores. Overrepresented proteins are shown in red, underrepresented in blue, and proteins not significantly different from expected are in grey. A) Proteins sorted from most to least represented. B) Proteins sorted in pathway order with lines dividing genes active on the cytoplasmic side of the ER, the lumen side of the ER, and in the Golgi. C) GPI anchor synthesis pathway diagram. Synthesis begins on the cytoplasmic side of the ER. The anchor is flipped to the lumen side. After the protein is attached, the glycan is sent to the Golgi for lipid remodeling.

Proteins acting on the cytoplasmic side of the ER were significantly underrepresented in the top 1% of ERC values, and proteins acting on the lumen side were highly overrepresented. We hypothesized that high ERC with other ER resident proteins may be driving the overrepresentation of lumen proteins. However, GO analysis for cellular compartment of the top 1% of proteins with high ERC with GPI anchor synthesis proteins revealed the top enriched compartments are associated with the exosome, ribosome, and mitochondria (S8 Fig, S2 Data). Some ER-related terms were enriched, but over 50 other terms were more enriched. This indicates that while ER

409 resident proteins may contribute to some of the high scores observed among lumen GPI anchor
410 synthesis proteins, it is unlikely that high scores with ER resident proteins are solely responsible
411 for driving the overrepresentation of lumen GPI anchor synthesis proteins in the top 1% of ERC
412 scores.
413
414 We examined the N-linked glycosylation pathway to determine if lower ERC scores for
415 cytoplasmic enzymes were unique to GPI anchor synthesis or common in glycosylation pathways.
416 The N-linked glycosylation pathway is similar in that the first few steps occur on the cytoplasmic
417 side, and the glycan is completed and attached to a protein on the ER lumen side. We calculated
418 ERC scores for proteins in the N-linked glycosylation pathway across the genome and ran the
419 same analysis as we did for GPI anchor synthesis. We took the top 1% of values and determined
420 the number of scores for each N-linked glycosylation protein and compared this to the expected
421 distribution, as we did above (S9 Fig). Among the eight cytoplasmic proteins, three were under-
422 represented, and four were over-represented. Among the ER lumen proteins, eight of the 23
423 proteins were under-represented, and ten of the 23 lumen proteins were over-represented. The
424 N-linked glycosylation enzymes acting on the cytoplasmic did not have an enrichment of lower
425 scores as observed for GPI anchor synthesis. The ERC pattern for cytoplasmic vs. ER lumen
426 proteins appears specific to the GPI anchor synthesis pathway.
427
428 ***In vivo* testing of top ERC pairs confirms genetic interactions**
429 A strong ERC score is an indicator of shared function between proteins in a complex or across a
430 pathway (19). Therefore, we tested top ERC scoring protein pairs to find new genetic interactions
431 with GPI anchor synthesis proteins (Table 5, S3 Data). Because many GPI anchor synthesis
432 proteins are necessary for survival in *Drosophila*, ubiquitous loss-of-function experiments are
433 challenging. The *Drosophila* eye provides a powerful alternative model for testing genetic
434 interactions with glycosylation genes (14). Using the GAL4-UAS system with the *eyes absent*

435 composite GAL4 driver (*eya*) (38), we expressed RNAi in an eye-specific manner to create
436 knockdown models for each GPI anchor gene tested. Many of these GPI anchor knockdown
437 models have changes in eye size/quality/color. We can observe genetic interactions by creating
438 double knockdowns of the highest scoring protein pairs in our genome-wide list and comparing
439 the effects of each gene's single knockdown with the effects of the double knockdown.

Table 5: *In vivo* testing for genetic interactions of top scoring protein pairs

	GPI protein	Protein 2	Protein 2 function	ERC score	Interaction
GPI x CDG	PIGG	DOLK	Dolichol-phosphate synthesis involved in N-linked glycosylation	9.59	YES
	PIGG	TRIP11	Golgi maintenance and homeostasis	9.50	NO
	GPAA1	ALG1	First mannosyltransferases in N-linked glycosylation	9.29	YES
GPI x Genome-wide	PIGA	RBSN	Early endocytic membrane fusion and membrane trafficking of recycling endosomes	9.58	YES
	PIGW	DHODH	Pyrimidine synthesis	10.32	NO
	PIGM	INF2	Actin polymerization and depolymerization	9.45	YES
	PIGZ	C12orf65	Mitochondrial gene translation	9.57	NO
	PIGO	ZNF622	Ribosomal maturation	10.19	YES
	PIGO	SMPD4	Sphingomyelin catabolism to form phosphorylcholine and ceramide	9.13	YES
	PIGG	HTT	Vesicle transport	11.28	NO
	PIGG	TUBGCP2	Microtubule nucleation	10.49	YES
	PIGG	ATG7	Autophagy	10.44	NO
	PIGG	TTC8	Cilia membrane protein sorting	10.22	YES
	PIGG	PHRF1	Protein ubiquitination	9.36	NO
	PIGG	ZGPAT	Transcription repressor, negatively regulates EGFR expression	9.18	YES
	PIGT	HDHD3	Function unknown	9.20	YES

Top scoring protein pairs tested for interaction in a *Drosophila* eye model. GPI proteins sorted by pathway order. Interactions tested with two-way ANOVA $p < 0.05$.

440

441 This tissue-specific eye model provides an efficient way to screen for genetic interactions.
442 However, there are caveats in using an eye knockdown model. Most genetic interactions are
443 context-specific. Because we are knocking down the genes only in the eye, we may not be testing

444 in the correct tissues for the interactions identified by ERC. All the genes tested have expression
445 in the *Drosophila* eye (39), but the eye does not always have the highest tissue expression.
446 Another major caveat to this model is the use of RNAi to knockdown the genes of interest.
447 Different RNAi transgenes reduce the expression level of their target genes by different levels (S3
448 Data). It is possible that the interactions in our models require specific
449 expression levels that a particular RNAi does not achieve.
450
451 We tested the top sixteen pairs from our genome-wide list with available RNAi lines (Table 5, S10
452 Fig, S3 Data). Of the sixteen pairs tested, 10 showed significant interactions. With over half of the
453 tested pairs validating, this is a higher hit rate than typically seen with ERC (40–42). When multiple
454 RNAi lines were available for a gene, we tested all RNAi lines. qPCR was performed on all RNAi
455 lines to determine knockdown levels for each line. The RNAi line resulting in the lowest expression
456 showed the interaction for almost all pairs with multiple RNAi lines tested. In two interaction pairs,
457 the lower knockdown RNAi line showed interaction while the strongest knockdown line did not.
458 This indicates expression level may play a role in identifying genetic interactions. Here we
459 highlight two interactions: one pair between a GPI anchor synthesis protein and another
460 glycosylation protein, and one pair between a GPI anchor synthesis protein and a non-
461 glycosylation protein.
462
463 GPAA1 and ALG1 had one of the highest scores between a GPI anchor synthesis protein and a
464 glycosylation protein outside of GPI anchor synthesis (ERC = 9.29). GPAA1 is a catalytic
465 component of the GPI-transamidase complex that attaches a protein to the synthesized GPI
466 anchor (13,43). ALG1, located on the cytoplasmic side of the ER, adds the first mannose onto the
467 N-linked glycan (11). Both *GPAA1* and *ALG1* are associated with their own respective CDGs (44).
468 While eye-specific knockdown of *ALG1* alone resulted in no change in eye size, it did result in
469 structural disorganization and occasional necrosis. Eye-specific knockdown of *GPAA1* alone

resulted in a significant reduction in eye size, but there was little qualitative change (Figure 8 A, B). Double knockdown of *GPAA1* and *ALG1* resulted in a significantly smaller eye than the expected additive effect of both single knockdowns ($p = 9.91 \times 10^{-5}$), and the eye was structurally disorganized. This indicates a genetic interaction between *GPAA1* and *ALG1*, two genes involved in distinct glycosylation pathways.

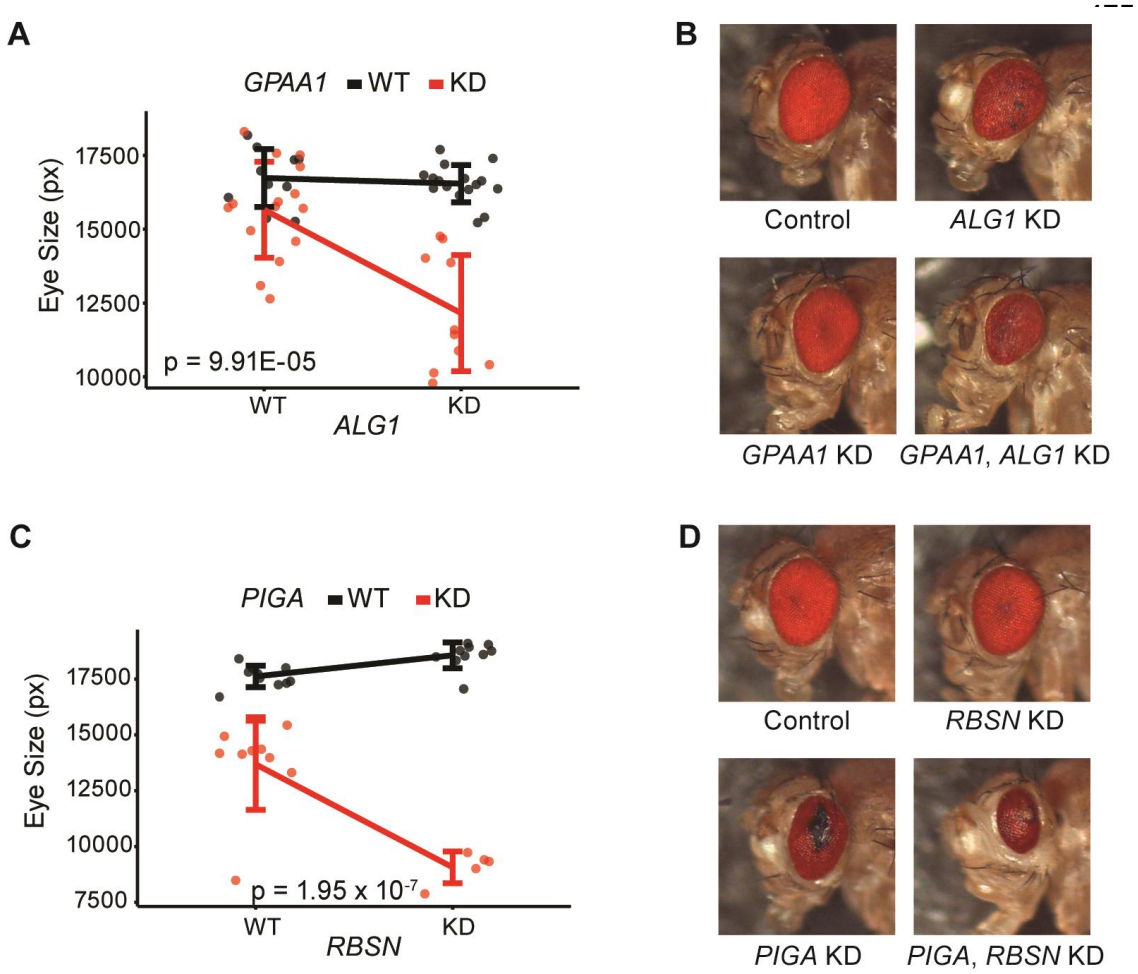

Figure 8: *In vivo* testing for genetic interactions in the eye. A),C) Quantification of eye sizes. Quantification in pixels (px) is along the Y axis. X axis has expression level of Protein 2, Wild Type (WT) or Knockdown (KD), and line color indicates expression level of GPI anchor synthesis protein (WT or KD). Interaction p-value calculated using a two-way ANOVA. B),D) Representative images of *Drosophila* eyes showing WT control, single KDs, and double KD eyes.

RBSN has the top ERC score with PIGA (ERC = 9.58), the enzymatic component of GlcNAc-transferase in GPI anchor synthesis (13). RBSN is recruited to the early endosome and plays a

478 role in sorting cargo between lysosomal degradation and membrane recycling (45,46). While
479 single knockdown of *RBSN* in the eye yields no change in eye size or structure (Figure 8 C, D),
480 single knockdown of *PIGA* results in a significantly decreased eye size. *PIGA* knockdown also
481 has structural disorganization, necrosis, and glassiness in the eye. Double knockdown of *PIGA*
482 and *RBSN* enhances the small eye *PIGA* phenotype in a non-additive manner indicating a genetic
483 interaction between *PIGA* and *RBSN* ($p = 1.95 \times 10^{-7}$). The double knockdown continues to show
484 necrotic tissue and disorganization. This same interaction was seen with a second RNAi line for
485 *PIGA*, where the loss of *RBSN* enhances the phenotype, in a non-additive manner ($p = 2.70 \times 10^{-}$
486 ^{4}) (S10 Fig D).
487
488 **High ERC scores suggest a connection between GPI anchor synthesis and the endosome**
489 It may not be immediately apparent why GPI anchor synthesis proteins like PIGA might show
490 strong ERC and a genetic interaction with endosomal proteins like RBSN. However, many GPI-
491 anchored proteins are frequently endocytosed and are either recycled back to the cell membrane
492 or sent to the lysosome for degradation (47). RBSN is recruited to the early endosome by Rab5
493 where it aids in vesicle fusion to form the early endosome (46). Interestingly, Rab5 also binds
494 HTT (48). HTT is an essential protein implicated in many cell functions, including cell transport
495 and autophagy, and had the highest genome-wide ERC score with a GPI anchor synthesis protein
496 (HTT and PIGG, ERC = 11.28) (49). Many interactors have been identified for HTT, but none are
497 involved in GPI anchor synthesis. Because of the genetic interaction between RBSN and PIGA
498 and the link between RBSN, Rab5, and HTT, we further examined the evolutionary covariation
499 between RBSN, HTT, and the GPI anchor pathway.
500
501 RBSN and HTT show significant ERC scores with 15 and six of the 30 GPI anchor synthesis
502 proteins, respectively. Strikingly, all six GPI anchor synthesis proteins, including PIGG, that had
503 high ERC scores with HTT also had high scores with RBSN (PIGG with RBSN, ERC = 9.05; PIGG

with HTT, ERC = 11.28) (green proteins in Figure 9). HTT and RBSN also had a very high ERC score with each other (ERC = 8.21). These high ERC scores suggest a previously unknown functional connection between GPI anchor synthesis proteins, HTT, and RBSN, and more broadly, with the endosome pathway. This also highlights the need to examine all proteins in the GPI anchor synthesis pathway together. A high ERC score between HTT and PIGG or between PIGA and RBSN is individually interesting but challenging to contextualize. However, the fact that the majority of GPI anchor synthesis pathway proteins show high ERC with RBSN, which is directly linked to HTT, allows for better contextualization of the top hit in the genome-wide data.

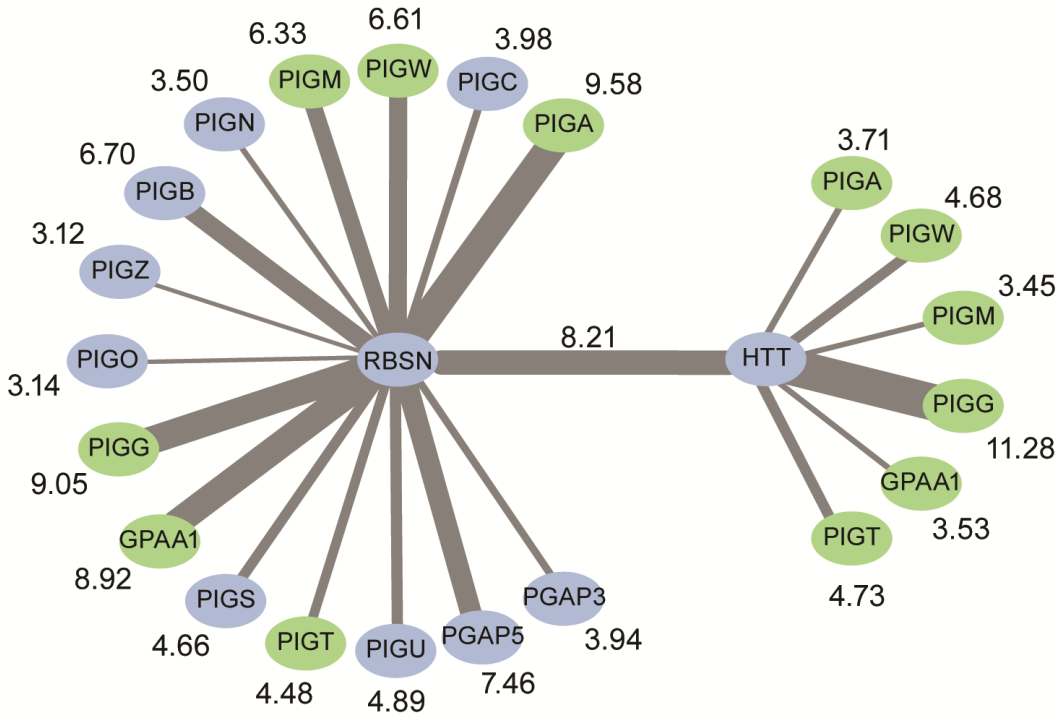

Figure 9: RBSN and HTT show strong ERC with each other and proteins in the GPI anchor synthesis pathway. ERC scores between proteins are indicated by the width of the edges and shown next to the GPI anchor protein. GPI anchor synthesis proteins showing significant ERC with both HTT and RBSN are shown in green. GPI anchor synthesis proteins are listed in pathway order.

514 **DISCUSSION**

515 CDGs are poorly understood, multi-systemic disorders with extensive phenotypic variability,
516 suggesting a role for genetic background and modifier genes. Identifying genetic modifiers of
517 CDGs could help clarify the pathology of these disorders. In this study, we used evolutionary
518 methods to explore genetic modifiers of CDGs and prioritize candidate genes for *in vivo* testing.

519

520 Until now, ERC has been used primarily to characterize unknown proteins or identify new proteins
521 involved in specific pathways (19,40–42,50–53). One previous study used evolutionary
522 similarities between disease genes to prioritize novel disease-causing genes (54). Because
523 modifier genes genetically interact with their disease genes, it is logical that they would also be
524 evolutionarily related to each other. Glycosylation is highly conserved from yeast to mammals
525 allowing for strong evolutionary signals to have developed over time (11,55,56). Previously
526 identified modifier genes for CDGs have high ERC values, including PMM2 – PGM1 (7.08) and
527 DPAGT1 – DPM1 (4.17) (14,16). *In* vivo testing of top hits from our analysis had a hit rate (10 out
528 of 16) higher than previously found in other ERC studies (40–42). Using ERC to identify disease
529 modifiers is a novel application that could be applied to help better understand and treat rare
530 diseases.

531

532 On average, glycosylation protein pairs showed high ERC regardless of whether the two proteins
533 were in the same pathway. However, we found exceptionally high ERC scores between proteins
534 involved in GPI anchor synthesis and N-linked glycosylation. One hypothesis for this might be
535 both pathways' shared reliance on mannose. N-glycans and GPI anchors have multiple
536 mannoses added to their glycan chains, and they both rely on the DPM complex to supply
537 mannose into the lumen of the ER (11,29). Only one type of O-linked glycan incorporates
538 mannose, which could explain why O-linked glycosylation does not have as high of ERC scores
539 with the other glycosylation pathways (57). Another potential explanation could be the necessity

540 of N-linked glycosylation for proper folding of GPI-anchored proteins (58). In the ER, N-glycans
541 on GPI-anchored proteins allow them to interact with chaperone proteins for proper protein
542 folding. This requirement for both N-linked glycosylation and GPI anchor synthesis for protein
543 folding might explain why the two different glycosylation pathways are evolutionarily linked. High
544 ERC between N-linked glycosylation and GPI anchor synthesis could also suggest unexplored
545 connections between these pathways, and they may need to be studied together instead of in
546 isolation.
547
548 High ERC scores occur when two proteins are under similar evolutionary constraints (19,59).
549 When analyzing ERC scores of GPI anchor synthesis proteins across the genome, we found that
550 GPI anchor synthesis proteins acting on the cytoplasmic side of the ER had fewer high ERC
551 scores than proteins acting on the luminal side of the ER. This was unexpected because all the
552 proteins in the GPI anchor synthesis pathway are only known to act in that pathway. It is feasible
553 that cytoplasmic components interact with a wider variety of proteins, which also influence their
554 evolutionary rates in different ways, making it more difficult for them to have strong ERC scores.
555 Cytoplasmic GPI anchor synthesis proteins may perform more functions than are currently known.
556
557 GO analysis of proteins showing high ERC with GPI anchor synthesis proteins resulted in multiple
558 terms related to ncRNA. Until recently it was thought there was no overlap between RNA and
559 glycosylation. In 2021, a group isolated glycosylated ncRNAs (60). It is currently unknown how
560 these glycans are synthesized and attached to RNA (61). It is possible the same enzymes
561 responsible for building the glycans for protein glycosylation also build RNA glycans. This
562 relationship could explain the evolutionary connection between ncRNAs and glycosylation
563 proteins.
564

565 Among the ncRNA GO terms, many were related to rRNA biology. However, it is unclear how
566 glycosylation is involved in rRNA biology. Since glycosylation is a co-translational/post-
567 translational modification, it suggests the evolutionary relationship identified between
568 glycosylation proteins and rRNA processing proteins could be due to glycosylation's dependence
569 on the ribosome functioning properly and translating proteins (62). Further studies need to be
570 performed to identify the specific protein interactions driving the association between these two
571 pathways.
572
573 CDG patients have a broad clinical spectrum influenced by different disease-causing variants,
574 environmental factors, and genetic background (63,64). Previous attempts to identify genetic
575 modifiers with standard methods (i.e., mutagenesis screens and using natural variation in the
576 population) have found promising candidates (14–16,20). These time and effort-intensive
577 methods can only be applied to one disorder at a time. The different CDGs are likely connected
578 because they all affect protein glycosylation and show similar symptoms in patients (65). Because
579 of their relatedness, it is probable that common modifiers affect multiple CDGs. Here, we
580 leveraged evolutionary tools to analyze all glycosylation genes and identify potential modifiers for
581 all CDG genes. Further exploration of these common modifiers could lead to novel therapeutic
582 targets for CDG patients that may be broadly applied to multiple types of CDGs.
583
584 **METHODS**
585 **Evolutionary Rate Covariation**
586 Evolutionary rate covariation (ERC) was calculated independently for orthologous sets of genes
587 from: 62 mammal species, 39 non-mammalian vertebrate species, 22 drosophila species, 18
588 yeast species, and 17 nematode species. The gene-specific rate of evolution was calculated for
589 each gene in all five independent phylogenies using the PAML package (66). The evolutionary
590 rate for each gene was then normalized into relative evolutionary rates (RERs) by comparing the

591 gene-specific branch length to the genome-wide average. The RERs for each pair of orthologous
592 genes in each of the five datasets were correlated using a Pearson correlation. The correlation
593 (i.e. the ERC value) was then Fisher transformed to adjust for the number of branches contributing
594 to a score. This allowed us to sum all five datasets into one integrated FtERC dataset. The
595 orthologs from drosophila, yeast, and nematodes were mapped to the hg19 human orthologs
596 using InParanoid (67). Only 1:1 orthologs were retained, but we did not require that an ortholog
597 be present in every independent phylogeny. This resulted in a dataset of FtERC values for 19,149
598 orthologous genes.
599
600 **Statistics**
601 Significance of elevated ERC scores was determined using proteome-wide permutation testing.
602 For significance testing in one group of proteins, the mean ERC value observed between all pairs
603 in a group is calculated. Then it is compared to the mean ERC values of 100,000 groups with the
604 same number of proteins chosen at random from the entire proteome. A p-value is calculated as
605 the proportion of random groups of proteins that had a mean ERC value of equal or greater value
606 than the mean of the tested group. Using 100,000 permutations, the most significant p-value
607 possible is $<1x10^{-5}$.
608
609 ERC significance between two different groups of proteins also used proteome-wide permutation
610 testing. The mean ERC of scores between proteins in two different groups is calculated. Then it
611 is compared to the mean ERC value between proteins where one of the testing groups is
612 maintained, and the other group is substituted by the same number of proteins chosen at random
613 from the entire genome. This is repeated 100,000 times, and a p-value is calculated as the
614 proportion of random groups of proteins that had a mean ERC value of equal or greater value
615 than the mean of the tested group. This process is repeated while maintaining the second group

616 of proteins and replacing the first group with random proteins. The p-value for that testing is
617 calculated as described above, and the higher p-value between the two tests is used.
618
619 Enrichment of each GPI anchor synthesis protein in the genome-wide dataset was determined
620 using chi-squared testing to compare the number of observed occurrences of each protein in the
621 dataset compared to number of expected occurrences if they were evenly distributed.
622
623 Genetic interactions between knockdown groups were determined using a two-way ANOVA run
624 in RStudio. GPI anchor gene level (wild type or knockdown) and gene 2 gene level (wild type or
625 knockdown) were the variables compared in the analysis.
626
627 **Gene Ontology Analyses**
628 Gene Ontology (GO) analysis was performed using the GO enrichment analysis
629 (geneontology.org) powered by the PANTHER Classification System. Enriched pathways were
630 identified using the "GO biological process complete" annotation dataset, and enriched cell
631 compartments were identified using the "GO cellular compartment complete" annotation dataset
632 with the *Homo sapiens* gene reference list using the default parameters (Fisher's Exact, Calculate
633 False Discovery Rate). Figures 5 and 6 and S8 Fig show most specific subclass term as sorted
634 by Panther hierarchical soring.
635
636 **Fly stocks and maintenance**
637 Stocks were maintained on standard Archon Scientific glucose fly food at 25°C on a 12-h
638 light/dark cycle. The *w-;; eya composite-GAL4* and *w-;eya composite-GAL4/CyO;+* lines were a
639 gift from Justin Kumar (Indiana University Bloomington) and were characterized previously (38).
640 *Drosophila* orthologues of candidate genes were determined using DIOPT, requiring a score ≥6

641 (68). RNAi lines used in this study were ordered from the Bloomington Drosophila Stock Center
642 and the Vienna Drosophila Resource Center (Line IDs in S3 Data).
643
644 To create double knockdown flies, either *w-; GPI RNAi/CyO;eya-GAL4/TM3, Sb* or *w-;eya-*
645 *GAL4/CyO; GPI RNAi/TM3, Sb* was crossed to either the candidate modifier RNAi lines or the
646 equivalent control lines (*attP40, attP2, attP, w1118)* selecting progeny that were negative for both
647 *CyO* and *TM3, Sb* balancers.
648
649 **Eye imaging and quantification**
650 Adult female flies aged 3-5 days were collected under CO_2 anesthesia then frozen at -80°C for
651 later imaging. Eyes were imaged at 3x magnification using the Leica EC3 Camera. Eye area was
652 measured as previously described (14,69).
653
654 **qPCR**
655 Knockdown efficiency for each RNAi line used was tested using qPCR. RNAi lines were crossed
656 to a *Tubulin-GAL4*/*Sb, Dfd-YFP* driver (BDSC: 5138) to test full-body knockdown amounts. 5-8
657 adult female flies aged 3-5 days old were collected, and RNA was extracted using a Direct-zol
658 RNA Miniprep (Zymo Research R2061) using TRIzol Reagent (ThermoFisher Cat # 15596026)
659 and including the DNAse step. RNA was converted to cDNA using a ProtoScript® II First Strand
660 cDNA Synthesis Kit (NEB Cat # E6560L). RT-qPCR was performed using a QuantStudio 3 96-
661 well 0.2 ml block instrument and PowerUp SYBR Green Master Mix (ThermoFisher Cat #
662 A25741). If *Tubulin* driving the RNAi line was adult lethal, 5-10 YFP-negative larvae were used.
663 If available, we used primers from the FlyPrimerBank (70) located at
664 *http://www.flyrnai.org/flyprimerbank*. Other primers were designed using Primer3Plus (71) located
665 at *www.primer3plus.com*. All primer sequences listed in S3 Data.
666

FUNDING

This work was supported by NHGRI grant R01 HG009299 received by NLC and NIGMS R35 GM124780 received by CYC. HJT was supported by NIDDK grant T32 DK1109660. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

693 **References**

694 1. Chow CY. Bringing genetic background into focus. Nat Rev Genet. 2016 Feb;17(2):63–4.

695 2. Nadeau JH. Modifier genes in mice and humans. Nat Rev Genet. 2001 Mar;2(3):165–74.

696 3. Rahit KMTH, Tarailo-Graovac M. Genetic Modifiers and Rare Mendelian Disease. Genes.
697 2020 Feb 25;11(3):239.

698 4. Chang IJ, He M, Lam CT. Congenital disorders of glycosylation. Ann Transl Med. 2018 Aug
699 11;6(24):12.

700 5. Frappaolo A, Sechi S, Kumagai T, Karimpour-Ghahnavieh A, Tiemeyer M, Giansanti MG.
701 Modeling Congenital Disorders of N-Linked Glycoprotein Glycosylation in Drosophila
702 melanogaster. Front Genet [Internet]. 2018 [cited 2020 Jun 16];9. Available from:
703 https://www.frontiersin.org/articles/10.3389/fgene.2018.00436/full

704 6. Piedade A, Francisco R, Jaeken J, Sarkhail P, Brasil S, Ferreira CR, et al. Epidemiology of
705 congenital disorders of glycosylation (CDG)—overview and perspectives. J Rare Dis. 2022
706 Dec 7;1(1):3.

707 7. Bayat A, Knaus A, Pendziwiat M, Afenjar A, Barakat TS, Bosch F, et al. Lessons learned
708 from 40 novel PIGA patients and a review of the literature. Epilepsia. 2020;61(6):1142–55.

709 8. Bellai-Dussault K, Nguyen TTM, Baratang NV, Jimenez-Cruz DA, Campeau PM. Clinical
710 variability in inherited glycosylphosphatidylinositol deficiency disorders. Clin Genet.
711 2019;95(1):112–21.

712 9. Brasil S, Pascoal C, Francisco R, Marques-da-Silva D, Andreotti G, Videira PA, et al. CDG
713 Therapies: From Bench to Bedside. Int J Mol Sci. 2018 May;19(5):1304.

714 10. Verheijen J, Tahata S, Kozicz T, Witters P, Morava E. Therapeutic approaches in
715 Congenital Disorders of Glycosylation (CDG) involving N-linked glycosylation: an update.
716 Genet Med. 2020 Feb;22(2):268–79.

717 11. Aebi M. N-linked protein glycosylation in the ER. Biochim Biophys Acta BBA - Mol Cell Res.
718 2013 Nov 1;1833(11):2430–7.

12. Schjoldager KTBG, Clausen H. Site-specific protein O-glycosylation modulates proprotein processing — Deciphering specific functions of the large polypeptide GalNAc-transferase gene family. Biochim Biophys Acta BBA - Gen Subj. 2012 Dec 1;1820(12):2079–94.
13. Almeida A, Layton M, Karadimitris A. Inherited glycosylphosphatidyl inositol deficiency: A treatable CDG. Biochim Biophys Acta BBA - Mol Basis Dis. 2009 Sep;1792(9):874–80.
14. Dalton HM, Viswanatha R, Jr RB, Zuno JS, Berman AR, Rushforth R, et al. A genome-wide CRISPR screen identifies DPM1 as a modifier of DPAGT1 deficiency and ER stress. PLOS Genet. 2022 Sep 27;18(9):e1010430.
15. Talsness DM, Owings KG, Coelho E, Mercenne G, Pleinis JM, Zuberi AR, et al. A Drosophila natural variation screen identifies NKCC1 as a substrate of NGLY1 deglycosylation and a modifier of NGLY1 deficiency. bioRxiv. 2020 Apr 14;2020.04.13.039651.
16. Vignogna RC, Allocca M, Monticelli M, Norris JW, Steet R, Perlstein EO, et al. Evolutionary rescue of phosphomannomutase deficiency in yeast models of human disease. Shou W, Perry GH, Freeze HH, Shou W, editors. eLife. 2022 Oct 10;11:e79346.
17. Gámez A, Serrano M, Gallego D, Vilas A, Pérez B. New and potential strategies for the treatment of PMM2-CDG. Biochim Biophys Acta BBA - Gen Subj. 2020 Nov 1;1864(11):129686.
18. Pajusalu S, Vals MA, Mihkla L, Šamarina U, Kahre T, Õunap K. The Estimated Prevalence of N-Linked Congenital Disorders of Glycosylation Across Various Populations Based on Allele Frequencies in General Population Databases. Front Genet [Internet]. 2021 [cited 2023 Jul 16];12. Available from: https://www.frontiersin.org/articles/10.3389/fgene.2021.719437
19. Clark NL, Alani E, Aquadro CF. Evolutionary rate covariation reveals shared functionality and coexpression of genes. Genome Res. 2012 Apr 1;22(4):714–20.

744 20. Westphal V, Kjaergaard S, Schollen E, Martens K, Grunewald S, Schwartz M, et al. A
745 frequent mild mutation in ALG6 may exacerbate the clinical severity of patients with
746 congenital disorder of glycosylation Ia (CDG-Ia) caused by phosphomannomutase
747 deficiency. Hum Mol Genet. 2002 Mar 1;11(5):599–604.
748 21. Blommaert E, Péanne R, Cherepanova NA, Rymen D, Staels F, Jaeken J, et al. Mutations
749 in MAGT1 lead to a glycosylation disorder with a variable phenotype. Proc Natl Acad Sci.
750 2019 May 14;116(20):9865–70.
751 22. Chaudhary BP, Zoetewey D, Mohanty S. 1H, 13C, 15N resonance assignments and
752 secondary structure of yeast oligosaccharyltransferase subunit Ost4 and its functionally
753 important mutant Ost4V23D. Biomol NMR Assign. 2020 Oct;14(2):205–9.
754 23. Blackburn JB, D'Souza Z, Lupashin VV. Maintaining order: COG complex controls Golgi
755 trafficking, processing, and sorting. FEBS Lett. 2019;593(17):2466–87.
756 24. Kinoshita T. Biosynthesis and biology of mammalian GPI-anchored proteins. Open Biol.
757 2020 Mar 11;10(3):190290.
758 25. Bailey EC, Alrowaished SS, Kilroy EA, Crooks ES, Drinkert DM, Karunasiri CM, et al. NAD+
759 improves neuromuscular development in a zebrafish model of FKRP-associated
760 dystroglycanopathy. Skelet Muscle. 2019 Aug 7;9(1):21.
761 26. Anzai R, Tsuji M, Yamashita S, Wada Y, Okamoto N, Saitsu H, et al. Congenital disorders
762 of glycosylation type IIb with MOGS mutations cause early infantile epileptic
763 encephalopathy, dysmorphic features, and hepatic dysfunction. Brain Dev. 2021 Mar
764 1;43(3):402–10.
765 27. Oud MM, Tuijnenburg P, Hempel M, van Vlies N, Ren Z, Ferdinandusse S, et al. Mutations
766 in EXTL3 Cause Neuro-immuno-skeletal Dysplasia Syndrome. Am J Hum Genet. 2017 Feb
767 2;100(2):281–96.
768 28. Franzka P, Henze H, Jung MJ, Schüler SC, Mittag S, Biskup K, et al. GMPPA defects cause
769 a neuromuscular disorder with **α**-dystroglycan hyperglycosylation. J Clin Invest [Internet].

770 2021 May 3 [cited 2023 Nov 22];131(9). Available from:
771 https://www.jci.org/articles/view/139076#B6
772 29. Maeda Y, Kinoshita T. Dolichol-phosphate mannose synthase: Structure, function and
773 regulation. Biochim Biophys Acta BBA - Gen Subj. 2008 Jun 1;1780(6):861–8.
774 30. Rush ET, Baker CV, Rizzo WB. Dolichol kinase deficiency (DOLK-CDG): Two new cases
775 and expansion of phenotype. Am J Med Genet A. 2017;173(9):2428–34.
776 31. Ondruskova N, Cechova A, Hansikova H, Honzik T, Jaeken J. Congenital disorders of
777 glycosylation: Still “hot” in 2020. Biochim Biophys Acta BBA - Gen Subj. 2021 Jan
778 1;1865(1):129751.
779 32. Skarie JM, Link BA. The Primary open-angle glaucoma gene WDR36 functions in ribosomal
780 RNA processing and interacts with the p53 stress–response pathway. Hum Mol Genet.
781 2008 Aug 15;17(16):2474–85.
782 33. Wang Y, Liu J, Zhao H, Lü W, Zhao J, Yang L, et al. Human 1A6/DRIM, the homolog of
783 yeast Utp20, functions in the 18S rRNA processing. Biochim Biophys Acta BBA - Mol Cell
784 Res. 2007 Jun 1;1773(6):863–8.
785 34. Chen Y, Li Y, Xiong J, Lan B, Wang X, Liu J, et al. Role of PRKDC in cancer initiation,
786 progression, and treatment. Cancer Cell Int. 2021 Oct 26;21(1):563.
787 35. Kuang L, Liu B, Xi D, Gao Y. A novel heterozygous ERCC6 variant identified in a Chinese
788 family with non-syndromic primary ovarian insufficiency. Mol Genet Genomic Med.
789 2022;10(10):e2040.
790 36. Corsello T, Kudlicki AS, Liu T, Casola A. Respiratory syncytial virus infection changes the
791 piwi-interacting RNA content of airway epithelial cells. Front Mol Biosci [Internet]. 2022 [cited
792 2023 Sep 10];9. Available from:
793 https://www.frontiersin.org/articles/10.3389/fmolb.2022.931354
794 37. Tsao CC, Geisen C, Abraham RT. Interaction between human MCM7 and Rad17 proteins is
795 required for replication checkpoint signaling. EMBO J. 2004 Nov 24;23(23):4660–9.

796 38. Weasner BM, Weasner BP, Neuman SD, Bashirullah A, Kumar JP. Retinal Expression of
797 the Drosophila eyes absent Gene Is Controlled by Several Cooperatively Acting Cis-
798 regulatory Elements. PLOS Genet. 2016 Dec 8;12(12):e1006462.
799 39. Krause SA, Overend G, Dow JAT, Leader DP. FlyAtlas 2 in 2022: enhancements to the
800 Drosophila melanogaster expression atlas. Nucleic Acids Res. 2022 Jan 7;50(D1):D1010–5.
801 40. Brunette GJ, Jamalruddin MA, Baldock RA, Clark NL, Bernstein KA. Evolution-based
802 screening enables genome-wide prioritization and discovery of DNA repair genes. Proc Natl
803 Acad Sci. 2019 Sep 24;116(39):19593–9.
804 41. Findlay GD, Sitnik JL, Wang W, Aquadro CF, Clark NL, Wolfner MF. Evolutionary Rate
805 Covariation Identifies New Members of a Protein Network Required for Drosophila
806 melanogaster Female Post-Mating Responses. PLOS Genet. 2014 Jan 16;10(1):e1004108.
807 42. Raza Q, Choi JY, Li Y, O'Dowd RM, Watkins SC, Chikina M, et al. Evolutionary rate
808 covariation analysis of E-cadherin identifies Raskol as a regulator of cell adhesion and actin
809 dynamics in Drosophila. PLOS Genet. 2019 Feb 14;15(2):e1007720.
810 43. Kawaguchi K, Sato T, Kondo S, Yamamoto-Hino M, Goto S. Stability of the transamidase
811 complex catalyzing GPI anchoring of proteins. Biochem Biophys Res Commun. 2019 May
812 7;512(3):584–90.
813 44. Freeze HH, Ng BG, Wolfe LA. 5 - Disorders Affecting Glycosylation. In: Pyeritz RE, Korf BR,
814 Grody WW, editors. Emery and Rimoin's Principles and Practice of Medical Genetics and
815 Genomics (Seventh Edition) [Internet]. Academic Press; 2021 [cited 2024 Jan 30]. p. 157–
816 81. Available from:
817 https://www.sciencedirect.com/science/article/pii/B9780128125359000054
818 45. Eathiraj S, Pan X, Ritacco C, Lambright DG. Structural basis of family-wide Rab GTPase
819 recognition by rabenosyn-5. Nature. 2005 Jul;436(7049):415–9.

820 46. Morrison HA, Dionne H, Rusten TE, Brech A, Fisher WW, Pfeiffer BD, et al. Regulation of
821 Early Endosomal Entry by the Drosophila Tumor Suppressors Rabenosyn and Vps45. Mol
822 Biol Cell. 2008 Oct;19(10):4167–76.
823 47. Verheijen J, Wong SY, Rowe JH, Raymond K, Stoddard J, Delmonte OM, et al. Defining a
824 new immune deficiency syndrome: MAN2B2-CDG. J Allergy Clin Immunol. 2020 Mar
825 1;145(3):1008–11.
826 48. Pal A, Severin F, Lommer B, Shevchenko A, Zerial M. Huntingtin–HAP40 complex is a
827 novel Rab5 effector that regulates early endosome motility and is up-regulated in
828 Huntington's disease. J Cell Biol. 2006 Feb 13;172(4):605–18.
829 49. Saudou F, Humbert S. The Biology of Huntingtin. Neuron. 2016 Mar 2;89(5):910–26.
830 50. Böhm S, Szakal B, Herken BW, Sullivan MR, Mihalevic MJ, Kabbinavar FF, et al. The
831 Budding Yeast Ubiquitin Protease Ubp7 Is a Novel Component Involved in S Phase
832 Progression *. J Biol Chem. 2016 Feb 26;291(9):4442–52.
833 51. Godin SK, Meslin C, Kabbinavar F, Bratton-Palmer DS, Hornack C, Mihalevic MJ, et al.
834 Evolutionary and Functional Analysis of the Invariant SWIM Domain in the Conserved
835 Shu2/SWS1 Protein Family from Saccharomyces cerevisiae to Homo sapiens. Genetics.
836 2015 Apr 1;199(4):1023–33.
837 52. Huang JW, Acharya A, Taglialatela A, Nambiar TS, Cuella-Martin R, Leuzzi G, et al.
838 MCM8IP activates the MCM8-9 helicase to promote DNA synthesis and homologous
839 recombination upon DNA damage. Nat Commun. 2020 Jun 11;11(1):2948.
840 53. Ziegler AB, Augustin H, Clark NL, Berthelot-Grosjean M, Simonnet MM, Steinert JR, et al.
841 The Amino Acid Transporter JhI-21 Coevolves with Glutamate Receptors, Impacts NMJ
842 Physiology and Influences Locomotor Activity in Drosophila Larvae. Sci Rep. 2016 Jan
843 25;6(1):19692.

844 54. Priedigkeit N, Wolfe N, Clark NL. Evolutionary Signatures amongst Disease Genes Permit
845 Novel Methods for Gene Prioritization and Construction of Informative Gene-Based
846 Networks. PLOS Genet. 2015 Feb 13;11(2):e1004967.
847 55. Lommel M, Strahl S. Protein O-mannosylation: Conserved from bacteria to humans.
848 Glycobiology. 2009 Aug 1;19(8):816–28.
849 56. Orlean P, Menon AK. *Thematic review series: Lipid Posttranslational Modifications.* GPI
850 anchoring of protein in yeast and mammalian cells, or: how we learned to stop worrying and
851 love glycophospholipids. J Lipid Res. 2007 May;48(5):993–1011.
852 57. Van den Steen P, Rudd PM, Dwek RA, Opdenakker G. Concepts and principles of O-linked
853 glycosylation. Crit Rev Biochem Mol Biol. 1998;33(3):151–208.
854 58. Liu YS, Guo XY, Hirata T, Rong Y, Motooka D, Kitajima T, et al. N-Glycan–dependent
855 protein folding and endoplasmic reticulum retention regulate GPI-anchor processing. J Cell
856 Biol. 2018 Feb 5;217(2):585–99.
857 59. Little J, Chikina M, Clark NL. Evolutionary rate covariation is a reliable predictor of co-
858 functional interactions but not necessarily physical interactions. Weigel D, editor. eLife. 2024
859 Feb 28;12:RP93333.
860 60. Flynn RA, Pedram K, Malaker SA, Batista PJ, Smith BAH, Johnson AG, et al. Small RNAs
861 are modified with N-glycans and displayed on the surface of living cells. Cell. 2021
862 Jun;184(12):3109-3124.e22.
863 61. Caldwell RM, Flynn RA. Discovering glycoRNA: Traditional and Non-Canonical Approaches
864 to Studying RNA Modifications. Isr J Chem. 2023;63(1–2):e202200059.
865 62. Chavan M, Lennarz W. The molecular basis of coupling of translocation and N-
866 glycosylation. Trends Biochem Sci. 2006 Jan 1;31(1):17–20.
867 63. Bortot B, Cosentini D, Faletra F, Biffi S, De Martino E, Carrozzi M, et al. PMM2-CDG:
868 Phenotype and genotype in four affected family members. Gene. 2013 Dec 1;531(2):506–9.
869 64. Jaeken J, Péanne R. What is new in CDG? J Inherit Metab Dis. 2017 Jul 1;40(4):569–86.

870 65. Lefeber DJ, Freeze HH, Steet R, Kinoshita T. Congenital Disorders of Glycosylation. In:
871 Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al., editors. Essentials of
872 Glycobiology [Internet]. 4th ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory
873 Press; 2022 [cited 2023 Aug 23]. Available from:
874 http://www.ncbi.nlm.nih.gov/books/NBK579928/
875 66. Yang Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol Biol Evol. 2007 Aug
876 1;24(8):1586–91.
877 67. Sonnhammer ELL, Östlund G. InParanoid 8: orthology analysis between 273 proteomes,
878 mostly eukaryotic. Nucleic Acids Res. 2015 Jan 28;43(D1):D234–9.
879 68. Hu Y, Flockhart I, Vinayagam A, Bergwitz C, Berger B, Perrimon N, et al. An integrative
880 approach to ortholog prediction for disease-focused and other functional studies. BMC
881 Bioinformatics. 2011 Dec;12(1):357.
882 69. Chow CY, Kelsey KJP, Wolfner MF, Clark AG. Candidate genetic modifiers of retinitis
883 pigmentosa identified by exploiting natural variation in Drosophila. Hum Mol Genet. 2016
884 Feb 15;25(4):651–9.
885 70. Hu Y, Sopko R, Foos M, Kelley C, Flockhart I, Ammeux N, et al. FlyPrimerBank: an online
886 database for Drosophila melanogaster gene expression analysis and knockdown evaluation
887 of RNAi reagents. G3 Bethesda Md. 2013 Sep 4;3(9):1607–16.
888 71. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3—
889 new capabilities and interfaces. Nucleic Acids Res. 2012 Aug;40(15):e115.
890