

## A complement C4-derived glycopeptide is a biomarker for PMM2-CDG

Kishore Garapati, ... , Eva Morava, Akhilesh Pandey

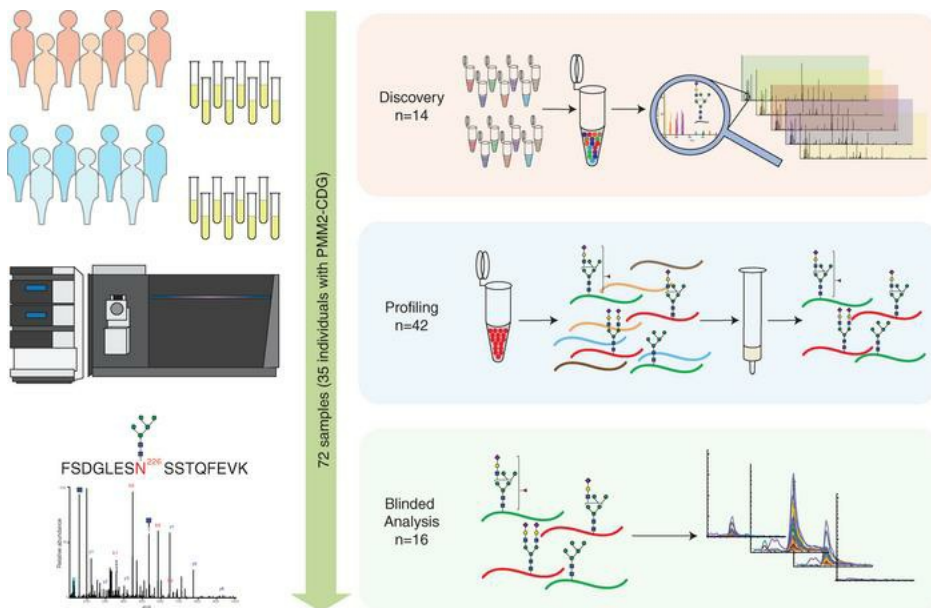
JCI Insight. 2024;9(7):e172509. <https://doi.org/10.1172/jci.insight.172509>.

Clinical Medicine

Genetics

Metabolism

### Graphical abstract



Find the latest version:

<https://jci.me/172509/pdf>



# A complement C4-derived glycopeptide is a biomarker for PMM2-CDG

Kishore Garapati,<sup>1,2,3</sup> Rohit Budhreja,<sup>1</sup> Mayank Saraswat,<sup>1</sup> Jinyong Kim,<sup>1</sup> Neha Joshi,<sup>1,2,3</sup> Gunveen S. Sachdeva,<sup>1,3</sup> Anu Jain,<sup>1</sup> Anna N. Ligezka,<sup>4</sup> Silvia Radenkovic,<sup>4</sup> Madan Gopal Ramarajan,<sup>1,2,3</sup> Savita Udainiya,<sup>1,2,3</sup> Kimiyo Raymond,<sup>5</sup> Miao He,<sup>6</sup> Christina Lam,<sup>7,8</sup> Austin Larson,<sup>9</sup> Andrew C. Edmondson,<sup>10</sup> Kyriakie Sarafoglou,<sup>11,12</sup> Nicholas B. Larson,<sup>13</sup> Hudson H. Freeze,<sup>14</sup> Matthew J. Schultz,<sup>5</sup> Tamas Kozicz,<sup>4,5,15,16</sup> Eva Morava,<sup>4,5,15,16</sup> and Akhilesh Pandey<sup>1,17</sup>

**Authorship note:** KG and RB are co-first authors.

**Conflict of interest:** KG, RB, JK, TK, EM, and AP are inventors on a pending patent application (US 63/529,913) assigned to Mayo Foundation for Medical Education and Research. CL received royalties for an UpToDate article on an overview of congenital disorder of glycosylation (CDG) from Wolters Kluwer; CL received a meeting fee waiver to present at a family conference from CDG CARE, a nonprofit; and CL is chair of the Data Safety Monitoring Board for an epalrestat trial in PMM2-CDG and a member of the Medical Advisory Board for CDG CARE. AL has grants/contracts from Astellas, Stealth BioTherapeutics, Neuren Pharmaceuticals, and Traverre Therapeutics and consulting fees from Illumina. ACE is a member of the CDG Care Medical Advisory Board and of the Data Safety Monitoring Board for Maggie's Pearl for a clinical trial of epalrestat in PMM2-CDG. HHF receives personal consulting fees from Bridge Bio and Glycomine Therapeutics and an honorarium from University of Wisconsin; HHF received travel reimbursement from University of Wisconsin and registration fee support for attending the Gordon Research Conference on Glycobiology; and HHF is a representative for the Society for Glycobiology at the FASEB Science Policy Committee.

**Copyright:** © 2024, Garapati et al. This is an open access article published under the terms of the Creative Commons Attribution 4.0 International License.

**Submitted:** May 24, 2023

**Accepted:** February 15, 2024

**Published:** April 8, 2024

**Reference information:** JCI Insight. 2024;9(7):e172509. <https://doi.org/10.1172/jci.insight.172509>.

<sup>1</sup>Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, USA. <sup>2</sup>Institute of Bioinformatics, International Technology Park, Bangalore, India. <sup>3</sup>Manipal Academy of Higher Education (MAHE), Manipal, India. <sup>4</sup>Department of Clinical Genomics and <sup>5</sup>Biochemical Genetics Laboratory, Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, USA. <sup>6</sup>Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. <sup>7</sup>Center for Integrative Brain Research, Seattle Children's Research Institute, Seattle, Washington, USA. <sup>8</sup>Division of Genetic Medicine, Department of Pediatrics, University of Washington School of Medicine, Seattle, Washington, USA. <sup>9</sup>Colorado Children's Hospital, Denver, Colorado, USA. <sup>10</sup>Division of Human Genetics, Department of Pediatrics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. <sup>11</sup>Division of Pediatric Endocrinology, Department of Pediatrics, University of Minnesota Medical School, Minneapolis, Minnesota, USA. <sup>12</sup>Department of Experimental and Clinical Pharmacology, University of Minnesota School of Pharmacy, Minneapolis, Minnesota, USA. <sup>13</sup>Department of Quantitative Health Sciences, Mayo Clinic, Rochester, Minnesota, USA. <sup>14</sup>Sanford Children's Health Research Center, Sanford Burnham Prebys Medical Discovery Institute, La Jolla, California, USA. <sup>15</sup>Department of Anatomy, University of Pécs Medical School, Pécs, Hungary. <sup>16</sup>Department of Genomics and Genetic Sciences, Icahn School of Medicine at Mount Sinai Hospital, New York, New York, USA. <sup>17</sup>Center for Individualized Medicine, Mayo Clinic, Rochester, Minnesota, USA.

**BACKGROUND.** Diagnosis of PMM2-CDG, the most common congenital disorder of glycosylation (CDG), relies on measuring carbohydrate-deficient transferrin (CDT) and genetic testing. CDT tests have false negatives and may normalize with age. Site-specific changes in protein N-glycosylation have not been reported in sera in PMM2-CDG.

**METHODS.** Using multistep mass spectrometry-based N-glycoproteomics, we analyzed sera from 72 individuals to discover and validate glycopeptide alterations. We performed comprehensive tandem mass tag-based discovery experiments in well-characterized patients and controls. Next, we developed a method for rapid profiling of additional samples. Finally, targeted mass spectrometry was used for validation in an independent set of samples in a blinded fashion.

**RESULTS.** Of the 3,342 N-glycopeptides identified, patients exhibited decrease in complex-type N-glycans and increase in truncated, mannose-rich, and hybrid species. We identified a glycopeptide from complement C4 carrying the glycan Man<sub>5</sub>GlcNAc<sub>2</sub>, which was not detected in controls, in 5 patients with normal CDT results, including 1 after liver transplant and 2 with a known genetic variant associated with mild disease, indicating greater sensitivity than CDT. It was detected by targeted analysis in 2 individuals with variants of uncertain significance in *PMM2*.

**CONCLUSION.** Complement C4-derived Man<sub>5</sub>GlcNAc<sub>2</sub> glycopeptide could be a biomarker for accurate diagnosis and therapeutic monitoring of patients with PMM2-CDG and other CDGs.

**FUNDING.** U54NS115198 (Frontiers in Congenital Disorders of Glycosylation: NINDS; NCATS; Eunice Kennedy Shriver NICHD; Rare Disorders Consortium Disease Network); K08NS118119 (NINDS); Minnesota Partnership for Biotechnology and Medical Genomics; Rocket Fund; R01DK099551 (NIDDK); Mayo Clinic DERIVE Office; Mayo Clinic Center for Biomedical Discovery; IA/CRC/20/1/600002 (Center for Rare Disease Diagnosis, Research and Training); DBT/Wellcome Trust India Alliance)

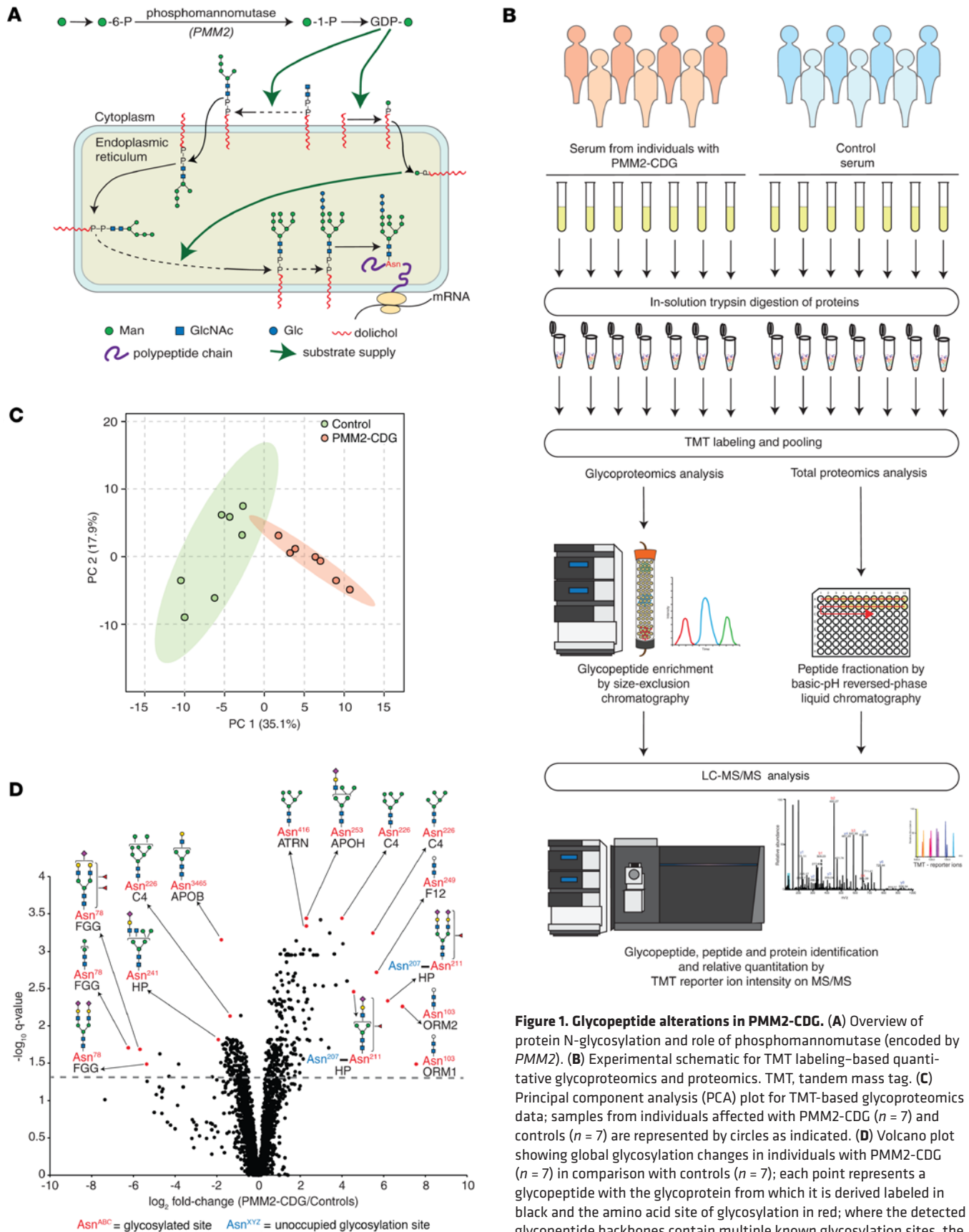
## Introduction

Congenital disorders of glycosylation (CDGs) are rare genetic diseases affecting the enzymatic addition of glycans to proteins and lipids. The most common disorder, PMM2-CDG, is caused by a deficiency of phosphomannomutase encoded by the *PMM2* gene, which converts mannose-6-phosphate into mannose-1-phosphate. Pathogenic variants in this gene reduce the amount and availability of glycosylation precursors, such as GDP-mannose primarily for protein N-linked glycosylation (Figure 1A) (1). This autosomal recessive disorder has an estimated prevalence of 1:20,000 to 1:80,000 (2–4). PMM2-CDG symptoms include global developmental delay, ataxia, seizures, and progressive neuropathy, along with endocrine and coagulation abnormalities. Severe infections, liver insufficiency, or cardiomyopathy cause lethality in 20% of patients by age of 6 years (2, 5, 6). PMM2-CDG is often missed in the diagnostic odyssey because of its nonspecific presentation, e.g., hypotonia and feeding difficulties (5). Molecular genetic diagnosis of PMM2-CDG still remains challenging (7). The first hint is often obtained from biochemical analysis of intact serum transferrin by mass spectrometry (MS) or other methods indicating carbohydrate-deficient transferrin (CDT; transferrin molecules lacking 1 or both N-glycan units) (8–11). However, these methods sometimes show normal or borderline abnormal profiles that may normalize upon repeat measurements in both young and older patients (12–15).

Some studies of patients with PMM2-CDG suggest that, in addition to unoccupied glycosylation sites, more subtle changes occur on the N-glycans that are added to a broad array of proteins (16, 17). Although glycoproteomic analysis for characterization of N-glycosylation (i.e., MS-based analysis of intact glycopeptides) is increasingly being applied to study biological problems (18–21), no such analysis of serum samples from patients with PMM2-CDG has yet been published. Thus, we applied our recently described liquid chromatography-tandem mass spectrometry-based (LC-MS/MS-based) N-glycoproteomics methods to study site-specific glycosylation changes in serum proteins in a large cohort of individuals with PMM2-CDG (22, 23). Here, we describe our biomarker discovery efforts progressing from discovery to validation using blinded sample sets. First, we obtained untargeted serum glycoproteomics data from a discovery set of 7 well-phenotyped individuals with PMM2-CDG and 7 controls. We confirmed the presence of unoccupied glycosylation sites in affected individuals and identified a diverse set of glycopeptide signatures that could distinguish individuals with PMM2-CDG from controls. These included glycopeptides from complement C4, alpha-1-acid glycoproteins 1 and 2, coagulation factor XII, and haptoglobin. Second, in the validation stage, we used 2 separate sets of samples from affected individuals for single-shot untargeted analysis followed by targeted analysis by parallel reaction monitoring-mass spectrometry (PRM-MS). We observed that a  $\text{Man}_5\text{GlcNAc}_2$  glycopeptide from complement C4 was detected only in individuals with PMM2-CDG and not in controls. Notably, this glycopeptide was also detected in 5 affected individuals with normal CDT, including 2 individuals with a known genetic variant associated with mild disease. Interestingly, this glycopeptide was also detected in an affected individual from whom the sample was obtained after liver transplantation, even though CDT was negative. Further, in another individual enrolled in a clinical trial, the abundance of this glycopeptide was decreased after 6 months of treatment with the aldose reductase inhibitor epalrestat, though CDT results were unchanged. Finally, this glycopeptide was also detected in 2 affected individuals with abnormal CDT results whose genetic testing showed variants of uncertain significance (VUSs) in *PMM2* gene. These data indicate that this glycopeptide has a greater concordance with the disease status than CDT. This glycopeptide could be deployed for the diagnosis of patients suspected to have PMM2-CDG as well as for monitoring of confirmed patients on treatment in the future.

## Results

*Demographics and clinical features of affected individuals.* Thirty-five individuals affected with PMM2-CDG were included in this study; the salient features of this group are summarized in Table 1. The demographic, clinical, and biochemical information of these individuals is shown in detail in Tables 2 and 3, along with severity reported as NPCRS scores (24, 25). All affected individuals had ataxia and global developmental delay. However, 7 individuals were able to walk without support (corresponding to samples 10, 22, 23, 25, 26, 30, and 36; Table 2), and 4 individuals had learning disabilities without intellectual disability (samples 20, 22, 26, and 30). All affected individuals had multisystem disease including peripheral organ involvement, in addition to neurologic involvement. Nine individuals had seizures (samples 6, 7, 11, 13, 18, 19, 27, 28, and 35), and 1 had stroke-like episodes within 6 months of sample collection (sample 1). Two siblings with a shared genotype (*PMM2*, p.R141H in 1 allele and the noncoding variant c.640-23A>G in the



**Table 1. Salient features of affected individuals**

Feature	Number of affected individuals
Total	35
Males	22
Females	13
<b>Age (y)</b>	
Median	10
Range	1–36
<b>Clinical severity (NPCRS) score</b>	
Median	23
Range	9–37
<b>Affected individuals by NPCRS score</b>	
Mild (0–14 points)	6
Moderate (15–25 points)	18
Severe (>26 points)	11
<b>Commonest pathogenic variants in <i>PMM2</i></b>	
p.R141H	20
p.T237R	5
p.F119L	4
p.C241S	4
p.P113L	4

NPCRS, Nijmegen Progression CDG Rating Scale.

second allele) had retinitis pigmentosa and sensorineural hearing loss (samples 27 and 28). However, 10 affected individuals had no signs of neuropathy at the time of investigation (samples 5, 8, 17, 19, 22, 23, 26, 30, 32, and 36). One individual had recurrent deep venous thrombosis of extremities (sample 11). One affected individual was born with a complex cardiac malformation, i.e., transposition of the great arteries (sample 7), and another affected individual had gastrointestinal dysmotility requiring feeding through gastrostomy-jejunostomy tube (sample 12). One affected individual had a severe chest deformity, i.e., pectus carinatum (sample 9), and another had severe scoliosis requiring surgery (sample 16). Twenty-six of the 35 affected individuals had abnormal levels for the activity of N-glycosylated coagulation factors IX, XI, antithrombin, and protein C and endocrine parameters (adrenocorticotrophic hormone, thyroid stimulating hormone, insulin-like growth factor binding protein) and normal liver transaminases (Table 3).

One affected individual donated 2 samples collected at different time points (samples 3 and 15). This individual was enrolled in a single-patient investigational new drug (IND) protocol with epalrestat, an aldose reductase inhibitor previously shown to improve phosphomannomutase enzyme activity and glycosylation in vitro (26). Sample 3 was collected from this individual at the age of 6 years at baseline, i.e., before starting treatment with epalrestat, and sample 15 was collected 6 months later on epalrestat treatment. Another affected individual (sample 24) had an unusual course of disease with early progressive liver failure. This individual developed intermittent vomiting, chronic abdominal pain, and recurrent ascites requiring diuretic therapy and repeated paracenteses. Her liver ultrasound showed cirrhosis. Due to end-stage liver disease, hepatic encephalopathy, and esophageal variceal bleeding, she underwent a deceased donor orthotopic liver transplant at the age of 4 years. Although a pretransplantation serum sample was not available, a sample after liver transplantation was obtained as she was enrolled in a CDG natural history study (sample 24) (27).

The genotypes of all affected individuals have been previously reported as pathogenic except 2: p.T118A (corresponding to sample 29) and p.Y229S (sample 36), which are reported to have “uncertain significance” (ClinVar accessions RCV000310325 and RCV000078596.7, respectively) (26, 28–30). Additionally, no genotype-phenotype correlation was noted in the affected individuals except in cases with the genotype p.C241S, which has previously been described as “mild” because of its association with mild ataxia and variable learning difficulties (28). Two individuals with this variant (samples 25 and 26) had comparable clinical phenotypes with normal CDT results. However, 2 other individuals with this variant (samples 20 and 30) had classic abnormal CDT at the adult ages of 21 and 25, respectively (28).



MS of intact transferrin for detecting CDT was performed as part of the standard of care in all 35 affected individuals (31). Transferrin usually contains 2 biantennary sialylated N-glycans (di-oligo transferrin). Patients with PMM2-CDG typically lack 1 or both of them. The ratio of singly glycosylated transferrin (mono-oligo) to di-oligo is no more than 0.06 in unaffected individuals. The ratio of nonglycosylated (a-oligo) to di-oligo transferrin is no more than 0.011 in unaffected individuals. Mono-oligo to di-oligo transferrin ratios in affected individuals ranged from 0.03 (within reference range) to 1.93. A-oligo to di-oligo ratios ranged from 0.002 (within reference range) to 0.66. Importantly, 5 affected individuals with age in the range 6–23 years had fully normal CDT results, even though they all had genetically and/or enzymatically confirmed PMM2-CDG. Two of these individuals (samples 25 and 26) had the variant p.C241S, which has been associated with “mild” disease (28). Two other individuals with CDT results within the reference range (samples 27 and 28) had “severe” NPCRS scores, while the fifth individual (sample 24) donated the sample after liver transplantation.

*Glycoproteomic analysis of serum proteins.* We chose 7 individuals (discovery set: samples 1 to 7, Table 3) with PMM2-CDG along with 7 age- and sex-matched controls for an in-depth analysis of their serum glycoproteome to identify glycopeptide biomarkers in PMM2-CDG. We first confirmed CDT results by MS-based protein-level measurements with TMT-based relative quantitation: peptides from transferrin with unoccupied N-glycosylation sites were increased in PMM2-CDG (Figure 1B and Supplemental Figure 1A; supplemental material available online with this article; <https://doi.org/10.1172/jci.insight.172509DS1>). We also identified increased abundance of peptides with unoccupied N-glycosylation sites derived from several other serum proteins in PMM2-CDG (Supplemental Figure 1B). For glycoproteomics analysis, serum-derived peptides were multiplexed with tandem mass tags (TMTs) for relative quantitation followed by N-glycopeptide enrichment by size-exclusion chromatography (SEC) and LC-MS/MS analysis (32, 33) (Figure 1B). A total of 3,342 glycopeptides were identified from 284 glycoproteins, and their distribution by class of glycan is shown in Supplemental Figure 2, A–C. Alpha-1-acid glycoproteins 1 and 2 contributed the largest number of glycopeptides, 297, for any protein. The most frequently identified glycans on glycopeptides were the sialylated biantennary glycans Hex<sub>5</sub>HexNAc<sub>4</sub>NeuAc<sub>2</sub> (352 glycopeptides) and Hex<sub>5</sub>HexNAc<sub>4</sub>NeuAc<sub>1</sub> (230 glycopeptides) (Supplemental Table 1). Principal component analysis (PCA) performed using TMT-based relative abundance values of glycopeptides clustered PMM2-CDG and control samples separately (Figure 1C).

*Site-specific alterations in N-glycosylation of serum proteins.* We found significant glycosylation changes ( $q < 0.05$ ) in 371 unique N-glycopeptides from 85 glycoproteins (Figure 1D and Supplemental Figure 3). A total of 136 glycopeptides from 39 proteins were elevated with an average fold-change (PMM2-CDG/controls) of  $\geq 2$ -fold ( $q < 0.05$ ). A total of 14 of these glycopeptides bore paucimannose (3 or fewer hexose residues) or chitobiose-core glycan composition. Notably, glycopeptides from 3 proteins, alpha-1-acid glycoprotein 1 (ORM1, glycopeptide containing amino acid glycosylation site Asn<sup>103</sup>), alpha-1-acid glycoprotein 2 (ORM2, Asn<sup>103</sup>), and coagulation factor XII (F12, Asn<sup>249</sup>) bearing the truncated glycan, Hex<sub>1</sub>HexNAc<sub>2</sub>, were found to be elevated  $\geq 50$ -fold (average fold-change, PMM2-CDG/controls), with undetectable levels in several controls. Glycopeptides bearing the tetrasaccharide, Hex<sub>2</sub>HexNAc<sub>2</sub> from attractin (ATRN, Asn<sup>416</sup>), complement C3 (C3, Asn<sup>85</sup>), immunoglobulin heavy constant mu chain (IGHM, Asn<sup>46</sup>), and F12 (Asn<sup>249</sup>) were also high in PMM2-CDG. Glycopeptides bearing the paucimannose glycan Hex<sub>3</sub>HexNAc<sub>2</sub> from F12 (Asn<sup>249</sup>) and galectin-3-binding protein (LGALS3BP, Asn<sup>551</sup>) were also elevated (Figure 2A and Supplemental Table 1).

Oligomannose glycopeptides containing 6 or fewer mannose residues from several proteins were elevated, including transferrin (Asn<sup>432</sup>: Hex<sub>4</sub>HexNAc<sub>2</sub>), ATRN (Asn<sup>416</sup>: Hex<sub>4</sub>HexNAc<sub>2</sub>), and alpha-1-antitrypsin (SERPINA1, Asn<sup>271</sup>: Hex<sub>5</sub>HexNAc<sub>2</sub>) (Figure 2A). A number of hybrid glycopeptides were increased in affected individuals. The glycopeptide from transferrin (Asn<sup>432</sup>) with glycan composition Hex<sub>5</sub>HexNAc<sub>3</sub>NeuAc<sub>1</sub> was elevated in PMM2-CDG. Glycopeptides from several glycoproteins with the hybrid glycan composition Hex<sub>6</sub>HexNAc<sub>3</sub>NeuAc<sub>1</sub>Fuc<sub>1</sub>, corresponding to a fucosylated hybrid glycan, were elevated: ceruloplasmin (Asn<sup>138</sup>), immunoglobulin heavy chain constant gamma 1 (IGHG1, Asn<sup>180</sup>), and IGHM (Asn<sup>209</sup>) (Figure 2A). Notably, the levels of hybrid glycopeptides with the glycan composition Hex<sub>6</sub>HexNAc<sub>3</sub>NeuAc<sub>1</sub> were increased on 25 glycosylation sites derived from 19 proteins, including ORM1/ORM2 (Asn<sup>72</sup>), prothrombin (F2, Asn<sup>121</sup>), and beta-2-glycoprotein 1 (Asn<sup>253</sup>) (Figure 2A and Supplemental Table 1).

We noted that 87 glycopeptides derived from 24 proteins had decreased abundance (average fold-change  $\leq 0.5$ ,  $q < 0.05$ ) in PMM2-CDG. A majority of these were of the hybrid/complex type, with high-mannose glycopeptides accounting for a small minority. Complex mono- and di-sialylated biantennary glycans accounted for 20 of these glycopeptides derived from several proteins, including kallistatin (SERPINA4, Asn<sup>157</sup>), CD207 (Asn<sup>180</sup>), and N-acetylgalactosaminyltransferase 5 (Asn<sup>845</sup>). Complex di- and tri-sialylated

**Table 2. Affected individuals and their clinical features**

Analysis set	Sample ID	Age (y)	Sex	NPCRS <sup>A</sup>				Neurologic symptoms		Other unique features and clinical remarks
				Section I	Section II	Section III	Total	Ataxia with GDD <sup>B</sup>	Seizures	
Discovery set	Sample 1	6	M	15	5	16	36	+	-	Stroke-like episodes
	Sample 2	6	M	11	4	14	29	+	-	
	Sample 3	6	F	11	0	13	24	+	-	Epalrestat trial (pretreatment)
	Sample 4	7	F	9	3	12	24	+	-	
	Sample 5	5	M	7	1	10	18	+	-	Absence of neuropathy
	Sample 6	15	M	6	5	9	20	+	+	
	Sample 7	7	M	12	3	13	28	+	+	Congenital heart defect
Profiling set	Sample 8	2	M	7	3	8	18	+	-	Absence of neuropathy
	Sample 9	3	M	15	4	18	37	+	-	Chest deformity
	Sample 10	36	F	6	4	13	23	+	-	Able to walk
	Sample 11	28	M	11	6	15	32	+	+	Recurrent thrombotic episodes
	Sample 12	7	M	17	3	14	34	+	-	Jejunal tube feeding/hearing loss
	Sample 13	4	F	9	4	13	26	+	+	
	Sample 14	24	M	9	1	14	24	+	-	
	Sample 15	7	F	6	0	12	18	+	-	Epalrestat trial (on treatment)
	Sample 16	14	M	9	2	12	23	+	-	Scoliosis
	Sample 17	1	M	8	2	9	19	+	-	Absence of neuropathy
	Sample 18	22	M	13	8	15	36	+	+	
	Sample 19	31	M	3	0	9	12	+	+	Absence of neuropathy
	Sample 20	21	F	5	4	9	18	+, #	-	
	Sample 21	13	M	6	4	11	21	+	-	
	Sample 22	11	F	5	1	4	10	+, #	-	Absence of neuropathy/able to walk
	Sample 23	10	F	9	1	12	22	+	-	Absence of neuropathy/able to walk
	Sample 24	6	F	15	2	13	30	+	-	After liver transplant
	Sample 25	12	M	4	0	10	14	+	-	Able to walk
	Sample 26	23	F	3	3	3	9	+, #	-	Absence of neuropathy/able to walk
	Sample 27	11	F	9	4	13	26	+	+	Hearing loss/retinitis pigmentosa
Sample 28	9	M	10	2	14	26	+	+	Hearing loss/retinitis pigmentosa	
Blinded set	Sample 29	7	M	9	2	13	24	+	-	
	Sample 30	25	F	3	3	3	9	+, #	-	Absence of neuropathy/able to walk
	Sample 31	9	M	6	1	10	17	+	-	
	Sample 32	4	M	4	1	7	12	+	-	Absence of neuropathy
	Sample 33	9	F	6	1	11	18	+	-	
	Sample 34	11	M	5	1	9	15	+	-	
	Sample 35	10	M	7	3	13	23	+	+	
	Sample 36	16	F	5	5	7	17	+	-	Absence of neuropathy/able to walk

List of included individuals with PMM2-CDG with demographic information and clinical findings. <sup>A</sup>Nijmegen Progression CDG Rating Scale. Patients are scored on 3 sections as follows: section I (current function, 21 points), section II (system-specific involvement, 30 points), and section III (current clinical assessment, 31 points) for a total score of 82 points; using these scores, cases are placed into mild (0–14 points), moderate (15–25 points), or severe (>25 points) categories. <sup>B</sup>Global developmental delay. +, present; -, absent; #, learning disability without intellectual disability.

triantennary glycans accounted for an additional 13 glycopeptides decreased in abundance, mainly from plasma protease C1 inhibitor (SERPING1, Asn<sup>253</sup>), clusterin (Asn<sup>86</sup>), and transferrin (Asn<sup>630</sup>) (Figure 2A and Supplemental Table 1). Box plots representing significantly altered glycopeptides between PMM2-CDG and control groups are shown in Figure 2A, and the most significantly elevated and decreased glycopeptides (average fold-change, PMM2-CDG/controls,  $q < 0.05$ ), are shown in Figure 2, B and C. Among the significantly decreased glycopeptides shown in Figure 2C, 3 glycopeptides were from albumin, with glycosylation at the noncanonical motif Asn<sup>123</sup>-Glu-Cys. We discovered this N-linked glycosylation site in albumin in a recent exploratory study (34), and the current finding indicates that glycosylation at this site, which we believe to be novel, is also affected in PMM2-CDG.

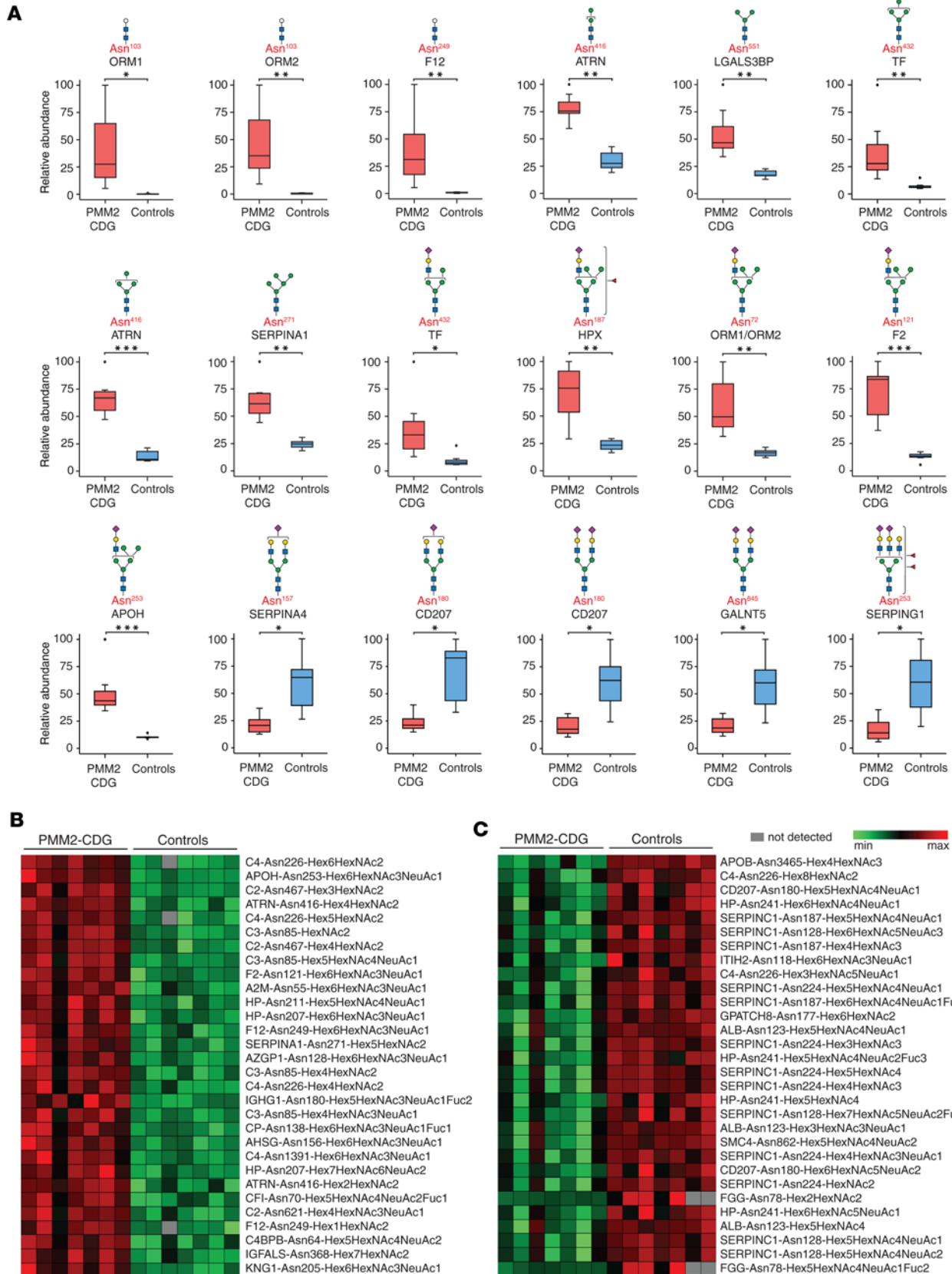
Table 3. Affected individuals and their laboratory features

Analysis set	Sample ID	Genetic variants in PMM2 (predicted protein level, allele 1/allele 2)	Abnormal glycosylated coagulation factors and hormones	CDT by intact mass spectrometry		Result of glycopeptide analysis (Complement C4-derived glycopeptide bearing Man <sub>5</sub> HexNAc <sub>2</sub> at Asn <sup>226</sup> )	
				Mono- to di-oligo ratio (normal [N] ≤ 0.06)	A- to di-oligo ratio (normal [N] ≤ 0.011)		
Discovery set	Sample 1	R141H/F119L	+	0.14 (↑)	0.005	Detected	
	Sample 2	P113L/D148N	+	0.054 (N)	0.123 (↑)	Detected	
	Sample 3	R141H/E139K	-	0.09 (↑)	0.004 (N)	Detected	
	Sample 4	D188G/V231M	+	0.3 (↑)	0.31 (↑)	Detected	
	Sample 5	R141H/F183S	+	0.69 (↑)	0.196 (↑)	Detected	
	Sample 6	R141H/I153T	+	0.85 (↑)	0.367 (↑)	Detected	
	Sample 7	R69C/D148N	-	0.12 (↑)	0.007 (N)	Detected	
Profiling set	Sample 8	Q37X/P113T	+	0.51 (↑)	0.031 (↑)	Detected	
	Sample 9	R141H/V129M	+	0.58 (↑)	0.112 (↑)	Detected	
	Sample 10	F119L/F119L	+	1.06 (↑)	0.520 (↑)	Detected	
	Sample 11	R141H/F119L	+	1.31 (↑)	0.656 (↑)	Detected	
	Sample 12	R141H/F119L	+	0.14 (↑)	0.005 (N)	Detected	
	Sample 13	R141H/V129M	+	0.70 (↑)	0.314 (↑)	Detected	
	Sample 14	C9Y/D148N	+	0.46 (↑)	0.147 (↑)	Detected	
	Sample 15	R141H/E139K	-	0.09 (↑)	0.004 (N)	Detected	
	Sample 16	R141H/I132T	+	0.75 (↑)	0.212 (↑)	Detected	
	Sample 17	R141H/F68C	+	0.61 (↑)	0.190 (↑)	Detected	
	Sample 18	P113L/-	+	0.791 (↑)	0.244 (↑)	Detected	
	Sample 19	R141H/P113L	-	0.81 (↑)	0.60 (↑)	Detected	
	Sample 20	R141H/C241S	+	0.8133 (↑)	0.0639 (↑)	Detected	
	Sample 21	A108V/T237R	+	0.42 (↑)	0.097 (↑)	Detected	
	Sample 22	Y229S/T237R	+	0.79 (↑)	0.376 (↑)	Detected	
	Sample 23	Y229S/T237R	+	1.20 (↑)	0.628 (↑)	Detected	
	Sample 24	P113L/T237R	-	0.04 (N)	0.004 (N)	Detected	
	Sample 25	R141H/C241S	-	0.06 (N)	0.004 (N)	Detected	
	Sample 26	R123Q/C241S	+	0.05 (N)	0.004 (N)	Detected	
	Sample 27	R141H/-	+	0.03 (N)	0.002 (N)	Detected	
	Sample 28	R141H/-	-	0.04 (N)	0.003 (N)	Detected	
	Blinded set	Sample 29	R141H/T118A	-	0.36 (↑)	0.064 (↑)	Detected
		Sample 30	R123Q/C241S	+	0.08 (↑)	0.013 (↑)	Detected
		Sample 31	R141H/I153T	+	0.67 (↑)	0.176 (↑)	Detected
		Sample 32	R141H/A108V	+	0.1069 (↑)	0.0607 (↑)	Detected
		Sample 33	R141H/G186R	-	0.2 (↑)	0.017 (↑)	Detected
		Sample 34	A108V/T237R	-	0.18 (↑)	0.025 (↑)	Detected
		Sample 35	Q33P/S47L	+	0.54 (↑)	0.161 (↑)	Detected
Sample 36		R141H/Y229S	+	1.9322 (↑)	0.5024 (↑)	Detected	

List of included individuals with PMM2-CDG with genetic information and biochemical findings, including results of intact mass analysis testing for carbohydrate-deficient transferrin (CDT) and results of glycopeptide analysis for Man<sub>5</sub>GlcNAc<sub>2</sub> glycopeptide from complement C4 at Asn<sup>226</sup>. Participants corresponding to 3 samples had noncoding pathogenic variants in their second allele: sample 18, IVS3+2T>C; samples 27 and 28, c.640-23A>G.

*Changes in glycan microheterogeneity.* We observed interesting patterns of change in the overall microheterogeneity at some glycosylation sites in PMM2-CDG. Several oligomannose glycopeptides at Asn<sup>226</sup> of complement C4 showed significant changes: Man<sub>4</sub>GlcNAc<sub>2</sub>, Man<sub>5</sub>GlcNAc<sub>2</sub>, and Man<sub>6</sub>GlcNAc<sub>2</sub> glycopeptides were elevated while Man<sub>7</sub>GlcNAc<sub>2</sub>, Man<sub>8</sub>GlcNAc<sub>2</sub>, and Man<sub>9</sub>GlcNAc<sub>2</sub> glycopeptides were reduced in PMM2-CDG. Specifically, Man<sub>5</sub>GlcNAc<sub>2</sub> and Man<sub>6</sub>GlcNAc<sub>2</sub> glycopeptides from C4 (Asn<sup>226</sup>) were elevated 45- and 16-fold, respectively, in PMM2-CDG (average fold-change PMM2-CDG/controls);





**Figure 2. Site-specific alterations in glycopeptide abundance from TMT-based experiments in the discovery set. (A)** Box plots representing levels of significantly altered glycopeptides; the glycoprotein from which the glycopeptide is derived is labeled in black and the amino acid site of glycosylation in red; putative structures are shown using SNFG and represent glycan composition inferred from mass spectrometry data (63). The box plots depict minimum and maximum values (whiskers), upper and lower quartiles, and median. The length of the box represents the interquartile range. \*= $p < 0.05$ , \*\*= $p < 0.01$ ,

\*\*\*= $q < 0.001$ . **(B)** Heatmap showing the most significantly increased ( $q < 0.05$ ) glycopeptides in PMM2-CDG. **(C)** Heatmap showing the most significantly decreased ( $q < 0.05$ ) glycopeptides in PMM2-CDG; glycopeptides are represented by protein followed by the amino acid site of glycosylation and glycan composition. The  $q$  values in **A–C** were calculated by  $t$  test with multiple testing using Benjamini-Hochberg procedure. PMM2-CDG ( $n = 7$ ), controls ( $n = 7$ ). Hex, hexose; HexNAc, N-acetylglucosamine; NeuAc, N-acetylneuraminic acid; Fuc, fucose.

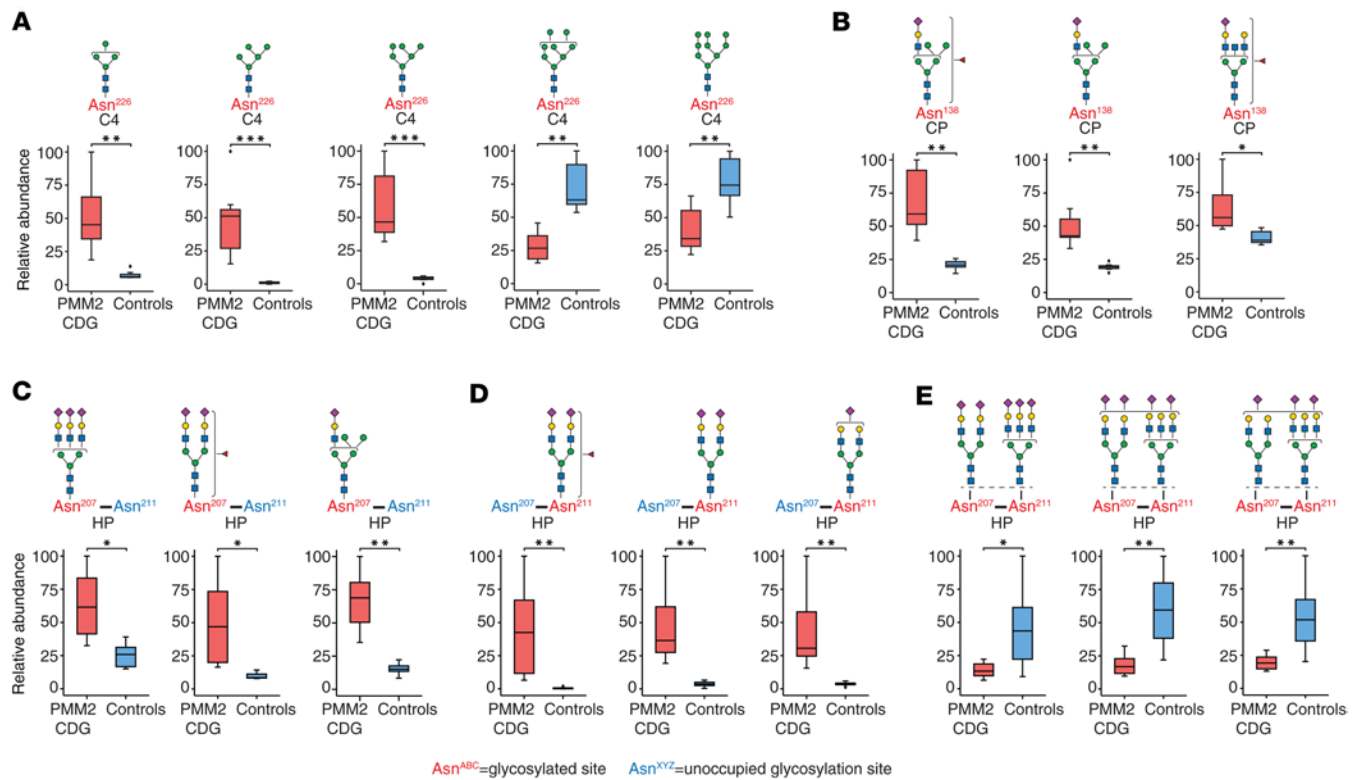
$q = 0.0006$  and  $0.0004$ , respectively) (Figure 3A and Supplemental Table 1). Interestingly, data from the protein-level measurements showed that complement proteins C4A and C4B (both of which share the tryptic peptide sequence that contains this glycosylation site) were not significantly different in their abundance in PMM2-CDG at the protein level (average fold-change PMM2-CDG/controls of 1.2 and 1.1, respectively, with  $q = 0.41$  and  $0.52$ , respectively). However, the nonglycosylated peptide containing the glycosylation site Asn<sup>226</sup> was elevated 6-fold in PMM2-CDG compared with controls ( $q = 0.0099$ ) (Supplemental Figure 1). This suggests that there are 2 changes occurring at this site in PMM2-CDG: the first, an overall decrease in glycosylation site occupancy; and second, where the glycosylation site is occupied, Man<sub>5</sub>GlcNAc<sub>2</sub> and Man<sub>6</sub>GlcNAc<sub>2</sub> glycans are increased and Man<sub>8</sub>GlcNAc<sub>2</sub> or Man<sub>9</sub>GlcNAc<sub>2</sub> glycans are decreased.

Another interesting pattern was observed at Asn<sup>138</sup> of ceruloplasmin where hybrid glycans with the composition Hex<sub>6</sub>HexNAc<sub>3</sub>NeuAc<sub>1</sub>Fuc<sub>1</sub> and Hex<sub>6</sub>HexNAc<sub>3</sub>NeuAc<sub>1</sub> and the complex monosialylated glycan Hex<sub>5</sub>HexNAc<sub>3</sub>NeuAc<sub>1</sub>Fuc<sub>1</sub> were significantly increased in PMM2-CDG (Figure 3B). The peptide from ceruloplasmin with nonglycosylated Asn<sup>138</sup>, however, was elevated 5-fold in PMM2-CDG ( $q = 0.0096$ ) while the protein level was relatively unchanged (average fold-change PMM2-CDG/controls of 1.05,  $q = 0.69$ ) (Supplemental Figure 1). This suggests a decrease in glycosylation at this site along with a significant increase in hybrid glycans where the site is glycosylated.

*Hypoglycosylation of specific sites on haptoglobin.* Because glycosylation changes in haptoglobin have previously been described in the context of PMM2-CDG (35, 36), we took a closer look at this glycoprotein. Haptoglobin-derived tryptic peptide NLFLN<sup>207</sup>HSEN<sup>211</sup>ATAK contains 2 known glycosylation sites. The fully glycosylated peptide (both sites occupied) was decreased while peptides glycosylated on a single site were increased in PMM2-CDG (Figure 3, C–E). To our knowledge, this finding has not been previously described for any haptoglobin-derived glycopeptide with multiple glycosylation sites in PMM2-CDG. We were able to do this because of the peptide- and glycosylation site-specific analysis that is now possible with recent advances in MS and database searching capabilities (22, 23, 37). These findings indicate substantial hypoglycosylation of these sites in haptoglobin in PMM2-CDG.

*Streamlined glycoproteomic profiling.* Although our multiplexed workflow using TMT labeling followed by SEC is excellent for discovery studies, it is not practical for analysis of clinical samples as it requires long hands-on time for sample preparation in addition to longer MS analysis time. To address this, we focused on an independent set of 17 samples from individuals with PMM2-CDG (samples 8 to 24) and 17 controls using a more streamlined method in which glycopeptides were enriched using mixed-mode anion exchange (MAX) cartridges (Figure 4A) and each sample analyzed separately by LC-MS/MS in a single-shot run (Table 3, profiling set).

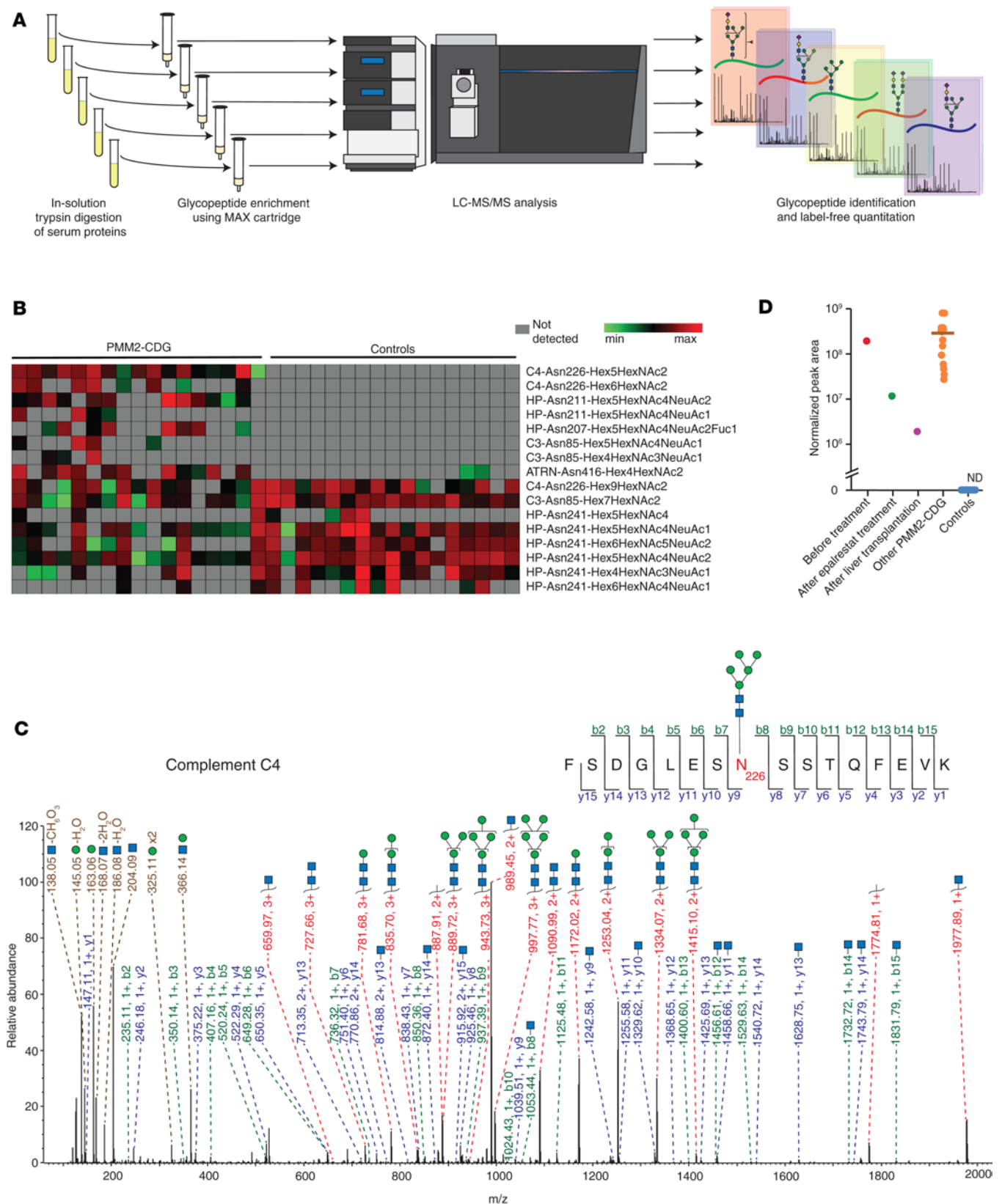
Figure 4B shows a heatmap of the most prominent glycosylation changes in PMM2-CDG. The only glycopeptide that was found in all 17 samples but absent in all controls was the glycopeptide with amino acid sequence and glycan composition FSDGLESN<sup>226</sup>(Man<sub>5</sub>GlcNAc<sub>2</sub>)SSTQFEVK derived from complement C4. Several other glycopeptides were found in many, but not all, samples from individuals with PMM2-CDG and absent from all controls. Haptoglobin-derived glycopeptides with the peptide sequence NLFLN<sup>207</sup>HSEN<sup>211</sup>ATAK, which contain 2 glycosylation sites but are only glycosylated at 1 of them, were only detected in samples from affected individuals; for example, glycopeptides with composition Hex<sub>5</sub>HexNAc<sub>3</sub>NeuAc<sub>2</sub> at Asn<sup>211</sup> and Hex<sub>5</sub>HexNAc<sub>4</sub>NeuAc<sub>2</sub>Fuc<sub>1</sub> at Asn<sup>207</sup> were found in 12 and 8 affected individuals, respectively, but in none of the controls. The glycopeptide from ATRN with Man<sub>4</sub>GlcNAc<sub>2</sub> at Asn<sup>416</sup> was found in samples from 11 affected individuals and with low intensity in only 2 controls. Many complement C4-derived glycopeptides observed to be elevated in the initial discovery set were also identified in this profiling set. Similar to TMT-based discovery data, the profiling set revealed a differential abundance of high-mannose-containing glycopeptides from C4 (Asn<sup>226</sup>) in individuals with PMM2-CDG. The glycopeptides from C4 (Asn<sup>226</sup>) bearing Man<sub>6</sub>GlcNAc<sub>2</sub> and fewer mannose residues were only found in individuals with PMM2-CDG; on the other hand, in control samples, only glycopeptides with the composition Man<sub>7</sub>GlcNAc<sub>2</sub>, Man<sub>8</sub>GlcNAc<sub>2</sub>, and Man<sub>9</sub>GlcNAc<sub>2</sub> were detected at this site. Man<sub>6</sub>GlcNAc<sub>2</sub> glycopeptide was identified in 8 affected individuals but not in any of the control samples. The peak area for glycopeptides



**Figure 3. Microheterogeneity in glycosylation for selected proteins from TMT-based experiments in the discovery set.** Box plots showing relative abundance of glycopeptides bearing different glycans at the same site in selected proteins (**A**) complement C4 Asn<sup>226</sup> and (**B**) ceruloplasmin Asn<sup>138</sup>. (**C**) Representative haptoglobin-derived glycopeptides containing Asn<sup>207</sup> and Asn<sup>211</sup> but with glycan occupancy only at Asn<sup>207</sup>, with Asn<sup>211</sup> unoccupied. (**D**) Representative haptoglobin-derived glycopeptides containing Asn<sup>207</sup> and Asn<sup>211</sup> but with glycan only at Asn<sup>211</sup>, with Asn<sup>207</sup> unoccupied. (**E**) Representative haptoglobin-derived glycopeptides containing Asn<sup>207</sup> and Asn<sup>211</sup> with glycosylation at both Asn<sup>207</sup> and Asn<sup>211</sup>; putative structures are shown using SNFG and represent total composition of glycan(s) inferred from mass spectrometry data (63). The box plots depict minimum and maximum values (whiskers), upper and lower quartiles, and median. PMM2-CDG (*n* = 7), controls (*n* = 7); \*=*q* < 0.05, \*\*=*q* < 0.01, \*\*\*=*q* < 0.001; the *q* values in **A–E** were calculated by *t* test with multiple testing using Benjamini-Hochberg procedure.

bearing Man<sub>7,9</sub>GlcNAc<sub>2</sub> was lower in all individuals affected with PMM2-CDG. A representative MS/MS spectrum from the sample of an affected individual for the Man<sub>5</sub>GlcNAc<sub>2</sub> glycan-bearing glycopeptide derived from complement C4 (Asn<sup>226</sup>) is shown in Figure 4C. By the streamlined profiling method using MAX-based enrichment, we did not detect some of the paucimannose glycopeptides that were identified to be elevated in affected individuals in the TMT-based discovery experiment. This could be due to single-shot LC-MS/MS runs of MAX-enriched fractions not being as comprehensive as the analysis of multiple fractions in the TMT-based discovery experiment. Another reason could be the differences in the basis of the enrichment methods. The downregulated glycopeptides common to the discovery and profiling sets also included a Man<sub>7</sub>GlcNAc<sub>2</sub> glycopeptide on Asn<sup>85</sup> site of complement C3, which was identified in 15 out of 17 affected individuals and all 17 controls. The peak area of this glycopeptide was lower in all individuals with PMM2-CDG as compared with controls. Eight other glycopeptides that were decreased in samples from individuals with PMM2-CDG were complex and hybrid glycopeptides containing site Asn<sup>241</sup> of haptoglobin, which is concordant with changes observed at this site in the TMT-based experiment.

This cohort included an affected individual under a single-patient IND protocol who was sampled twice. Sample 15 was collected from this individual at 6 months of treatment with the aldose reductase inhibitor epalrestat, which was being tested to improve phosphomannomutase (PMM2) enzyme activity. We had access to an archived sample from this individual at baseline, i.e., before treatment with epalrestat (sample 3), which was also analyzed using this streamlined method. The sample drawn after treatment showed a substantially lower peak area for the Man<sub>5</sub>GlcNAc<sub>2</sub> bearing glycopeptide derived from C4 (Asn<sup>226</sup>) as compared with the sample drawn before treatment (ratio of normalized peak areas, at 6 months of treatment to baseline, 0.06). Another interesting case involved sample 24, which was donated by a different individual after liver transplantation.



**Figure 4. Profiling N-glycosylation using MAX cartridge-based enrichment and single-shot MS analysis.** (A) Strategy for MAX-based enrichment and analysis. (B) Heatmap showing glycopeptides identified by label-free analysis using discovery methods; glycopeptides are represented by protein followed by the amino acid site of glycosylation and glycan composition. PMM2-CDG ( $n = 17$ ), controls ( $n = 17$ ). Hex, hexose; HexNAc, N-acetylglucosamine; NeuAc, N-acetylneuraminic acid. Fuc, fucose. (C) Annotated MS/MS spectrum for the  $\text{Man}_5\text{GlcNAc}_2$  glycopeptide from complement C4 at site Asn<sup>226</sup>; putative structures are shown using SNFG and represent glycan composition inferred from MS data (63). (D) Dot plot representing normalized peak intensity values



for the  $\text{Man}_3\text{GlcNAc}_2$  glycopeptide from complement C4 at site  $\text{Asn}^{226}$  in PMM2-CDG and control samples. Each dot represents abundance in a sample from an individual, and the horizontal bar represents the median; special cases are labeled and shown separately as follows: “Before treatment”: sample 3 ( $n = 1$ ) was donated by an individual affected with PMM2-CDG before treatment with epalrestat; “After epalrestat treatment”: sample 15 ( $n = 1$ ) was donated by the same individual at 6 months of therapy with epalrestat; “After liver transplantation”: sample 24 ( $n = 1$ ) was donated by an affected individual after liver transplantation; other PMM2-CDG ( $n = 15$ ); controls ( $n = 17$ ). ND, not detected.

Although a corresponding pretransplantation sample was unavailable for this individual, the peak area for this glycopeptide was markedly lower than in all other samples from individuals with PMM2-CDG (ratio of normalized peak areas, sample 24 to average of all other affected individuals, 0.007) (Figure 4D).

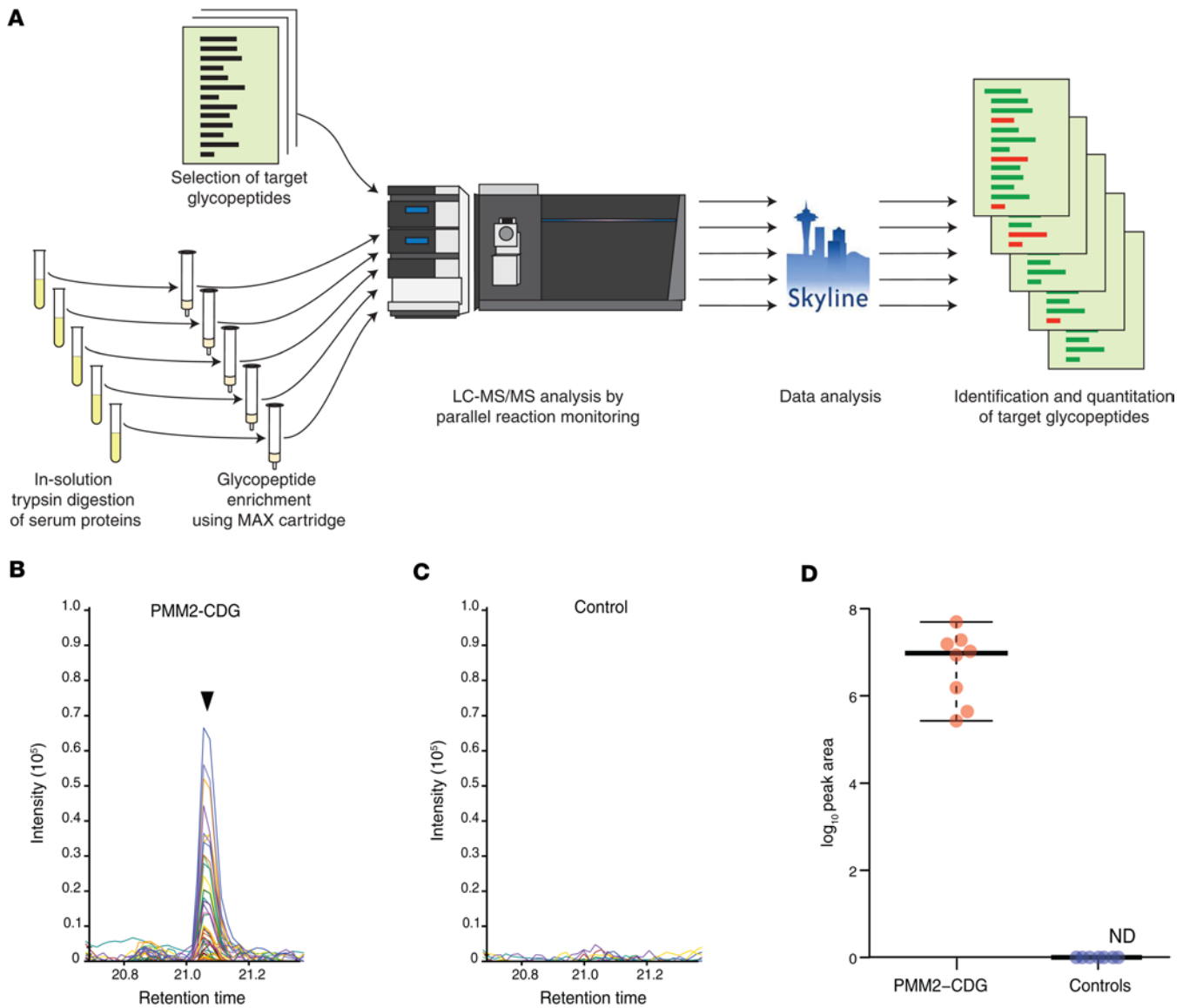
*Development of a targeted MS assay.* Encouraged by the results above, we developed a targeted method where MAX-enriched samples were analyzed by PRM for selected candidate glycopeptides. This method is more suitable for analysis in a clinical setting due to lower MS analysis time as well as enhanced sensitivity (Figure 5A). Four glycopeptide candidates were selected for targeted analysis based on the TMT-based discovery experiment and profiling experiments with MAX-enriched glycopeptides. The targeted glycopeptides are  $\text{EN}^{103}(\text{Hex}_1\text{HexNAc}_2)\text{GTISR}$  (alpha-1-acid glycoprotein 1, ORM1),  $\text{EN}^{103}(\text{Hex}_1\text{HexNAc}_2)\text{GTVSR}$  (alpha-1-acid glycoprotein 2, ORM2),  $\text{N}^{249}(\text{Hex}_1\text{HexNAc}_2)\text{VTAEQAR}$  (coagulation factor XII, F12), and  $\text{FSDGLESN}^{226}(\text{Man}_3\text{HexNAc}_2)\text{SSTQFEVK}$  (complement C4). The PRM assay for the selected peptides was first tested on 8 samples from individuals with a known diagnosis of PMM2-CDG and 8 controls (Table 3, samples 21 to 28). This included a subset of samples in which profiling experiments were already carried out, along with an additional 4 samples from individuals with PMM2-CDG who were reported to have control-like profiles based on their CDT results. The 3 target glycopeptides from ORM1, ORM2, and F12 bearing the paucimannose glycan  $\text{Hex}_1\text{HexNAc}_2$  were not detected in any of the samples. However, the target glycopeptide  $\text{FSDGLESN}^{226}(\text{Man}_3\text{HexNAc}_2)\text{SSTQFEVK}$  from C4 was detected in all 8 samples from affected individuals but in none of the controls. This target was detected as a doubly charged precursor ion with  $m/z$  1,496.1216 at a retention time of 21 minutes. Importantly, this glycopeptide was detected in 5 individuals with PMM2-CDG who had CDT results within the reference range. A representative targeted LC-MS/MS–extracted chromatogram of the target glycopeptide detected in the sample from an affected individual (sample 24) is shown in Figure 5B, and an extracted chromatogram from a control sample at the same retention time is shown in Figure 5C. Levels of this target glycopeptide in samples from affected individuals and controls that were used for testing the targeted assay are shown in Figure 5D.

*PRM assay successfully identifies confirmed cases of PMM2-CDG as well as affected individuals with VUSs.* Blinded analysis of 16 samples (PMM2-CDG or control) was performed using this PRM assay. The assay was 100% accurate in assigning individuals to their respective groups based on the targeted detection of the glycopeptide  $\text{FSDGLESN}^{226}(\text{Man}_3\text{GlcNAc}_2)\text{SSTQFEVK}$ . This glycopeptide was not detected in any of the control samples. Representative LC-MS/MS–extracted chromatograms for affected individuals are shown in Figure 6, A–D, and extracted chromatograms for controls at the same retention time are shown in Figure 6, E–H. Most importantly, this glycopeptide was also detected in 2 individuals who had VUSs in *trans* with the common variant R141H in *PMM2*: samples 29 (p.T118A, “Uncertain significance,” ClinVar accession RCV000310325) and 36 (p.Y229S, “Uncertain significance,” ClinVar accession RCV000078596.7) (Table 3). Next, to assess if there is any correlation between levels of the complement C4–derived  $\text{Man}_3$  glycopeptide at site  $\text{Asn}^{226}$  and clinical severity, we analyzed the levels of this glycopeptide along with the NPCRS scores for each affected individual using Kendall’s Tau rank correlation test. We did not observe any correlation between any of the aggregate or component NPCRS scores and the glycopeptide biomarker (Supplemental Tables 3 and 4, respectively).

## Discussion

CDT measured by intact mass analysis of transferrin glycosylation isoforms is the current standard of care in clinical diagnosis of CDG (13), though other methods have also been used (8–11). False negatives are known to occur in several types of CDG, including PMM2-CDG (12). Rare transferrin polymorphisms can also complicate interpretation of CDT assays (38, 39). Other glycoproteins, such as coagulation factors and hormones, have been studied in a similar fashion as transferrin and are sometimes included in CDG phenotyping (13, 40–43). Another diagnostic approach is MS analysis of N-glycans enzymatically released from total serum (44), although this method does not indicate glycosylation site occupancy or identify the proteins with changes in glycosylation (7). Additionally, the measurement of N-glycans released from



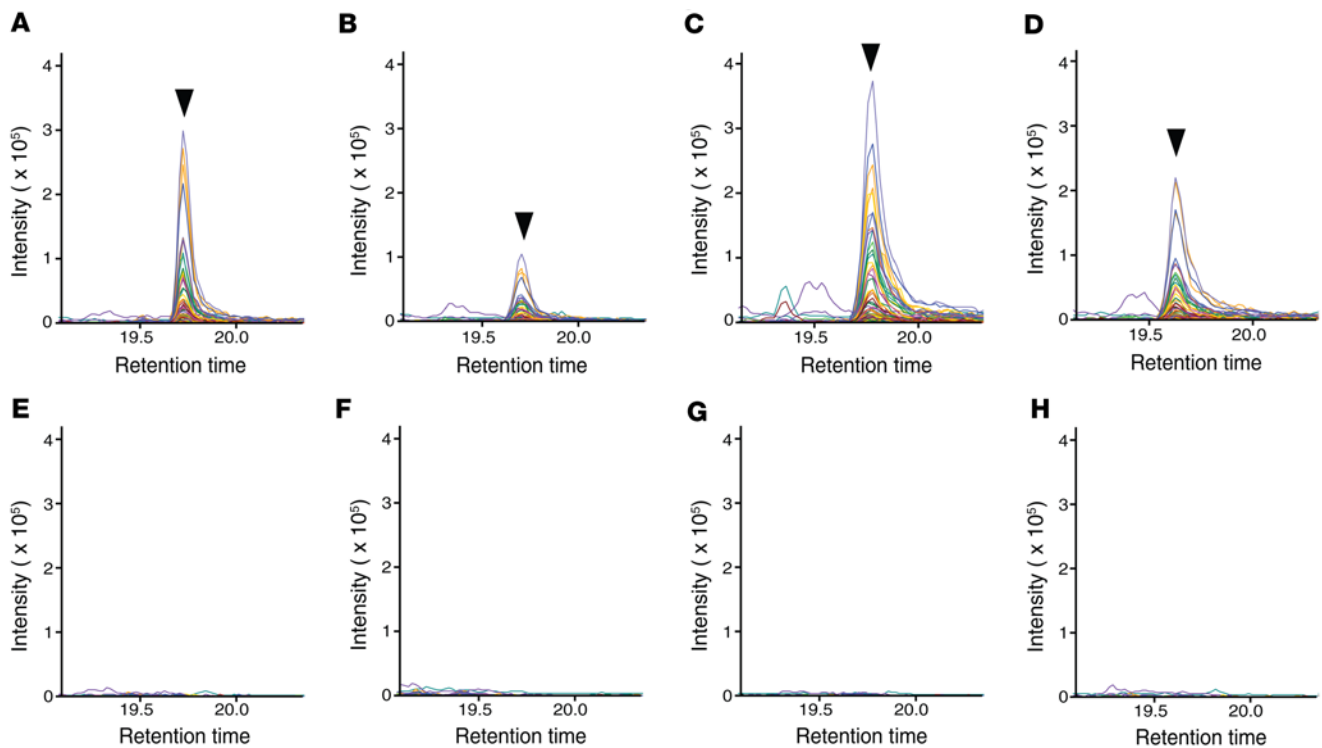


**Figure 5. Detection of  $\text{Man}_5\text{GlcNAc}_2$  glycopeptide from complement C4 by targeted PRM assay.** (A) Strategy for MAX-based enrichment and targeted analysis of glycopeptides. (B) Representative extracted ion chromatogram showing abundance of  $\text{Man}_5\text{GlcNAc}_2$  glycopeptide from complement C4 at Asn<sup>226</sup> in an individual with PMM2-CDG. (C) Representative extracted ion chromatogram at the same retention time in a control sample. (D) Dot plot showing abundance of the  $\text{Man}_5\text{GlcNAc}_2$  glycopeptide from complement C4 at Asn<sup>226</sup> in PMM2-CDG ( $n = 8$ ) and control ( $n = 8$ ) samples used for development of PRM assay. The dots represent abundance in each sample, and the horizontal bars represent, from top to bottom, the maximum value, median, and minimum value, respectively.

serum is actually a composite of glycans derived from many proteins. The resulting loss of information on glycosylation of individual proteins reduces the discriminatory capacity of glycomic analyses.

Recent advances in sample preparation, MS, and software development overcome these limitations (45–47), allowing analysis of thousands of glycopeptides to identify putative glycan structure based on composition and known biosynthetic pathways, as well as the protein-specific glycosylation sites at which the glycan is attached (37, 48, 49). Determining definitive glycan structures requires further analysis (50, 51). These glycoproteomics approaches have already begun to reveal global alterations in cancer, infection, and immune dysfunction (18–21) but have not yet been widely applied to CDG. We have developed methods for unbiased enrichment and deep profiling of intact N-glycopeptides by LC-MS/MS and previously applied them to study cellular models of CDG (22, 23, 26, 52, 53).

*Glycoproteomics reveals patterns of altered protein glycosylation in PMM2-CDG.* We found a remarkable increase in truncated glycans on many N-glycosylation sites in the serum proteome of PMM2-CDG



**Figure 6. Detection of  $\text{Man}_5\text{GlcNAc}_2$  glycopeptide from complement C4 at  $\text{Asn}^{226}$  by blinded analysis correctly identifies individuals with PMM2-CDG.** Representative Skyline plots showing glycopeptide fragment peaks from selected samples in the blinded set. Targeted analysis was performed on 16 samples by PRM by operators blinded to the identity of the samples. (A–D) Samples identified to be from individuals with PMM2-CDG. (E–H) Samples identified to be from controls.

individuals as compared with controls, where mature glycans were more abundant. The increase of paucimannose glycopeptides on several proteins with a concomitant reduction in  $\text{Man}_8\text{GlcNAc}_2$  and  $\text{Man}_9\text{GlcNAc}_2$  bearing glycopeptides suggests aberrant N-glycan processing. The  $\beta$  chain of haptoglobin, which contains 4 N-glycosylation sites, was previously described as altered in CDG (35, 36). We found that a tryptic glycopeptide from haptoglobin containing 2 closely spaced N-glycosylation sites ( $\text{Asn}^{207}$  and  $\text{Asn}^{211}$ ) is usually glycosylated at both sites in control individuals. In PMM2-CDG, a lack of glycosylation on one of these sites was noted, which might indicate frequent skipping of glycosylation at closely spaced acceptor sites on a polypeptide sequence (54, 55) and may also be applicable to other proteins in CDG.

*$\text{Man}_5\text{GlcNAc}_2$  glycopeptide from complement C4 as a biomarker for reliable diagnosis.* Finding better biomarkers in PMM2-CDG is a significant unmet need. CDT and phosphomannomutase enzyme activity are sometimes unreliable, complicating diagnosis especially in cases where genetic testing shows VUSs. The  $\text{Man}_5\text{GlcNAc}_2$  glycopeptide derived from complement C4 at site  $\text{Asn}^{226}$  was detected in all 35 tested individuals with PMM2-CDG, including 5 who had normal CDT results. Specifically, 2 of these 5 individuals carried the “mild” genetic variant (p.C241S) and were scored as clinically mild by the NPCRS but were classified along with other affected individuals based on detection of this glycopeptide. Notably, this marker was also detected in 2 phenotypically affected individuals with VUSs. Our results combined with their clinical course and biochemical abnormalities suggest these variants to be likely pathogenic.

To our knowledge, the cohort of affected individuals included in this study is the largest group with this disorder studied by glycoproteomic methods. However, this biomarker study is still limited by sample size considerations inherent to studying ultra-rare genetic disorders. Another limitation of this study is that this biomarker will need to be tested for its potential applicability in other CDGs. This glycopeptide biomarker after additional validation could be considered for deployment for clinical testing, as well as a part of studying the natural history of PMM2-CDG. Lower levels of this marker detected in an affected individual on treatment with epalrestat and another individual after liver transplant indicate that it could be tested as a potential measure of outcomes in future clinical trials in PMM2-CDG. We also discovered additional glycopeptides and peptides with unoccupied glycosylation sites from abundant plasma proteins that show

coordinate changes in samples from individuals with PMM2-CDG. These glycopeptides can be tested as biomarkers in the future using other methods (e.g., targeted methods with heavy-labeled glycopeptides as internal standards) or be incorporated as part of potential panels of glycopeptide biomarkers. To conclude, glycoproteomic profiling for discovery followed by streamlined profiling and validation experiments, enabled discovery of  $\text{Man}_5\text{GlcNAc}_2$  glycopeptide from C4 at site Asn<sup>226</sup> as a biomarker in PMM2-CDG.

## Methods

*Sex as a biological variable.* Our study included both male and female individuals. Affected individuals and control groups were matched for sex.

*Affected individuals.* Thirty-six samples from 35 affected individuals (22 male and 13 female, age range = 1–36 years, median age = 10 years) with a molecularly confirmed diagnosis of PMM2-CDG were included in the study. All individuals are enrolled in the Frontiers in Congenital Disorders of Glycosylation Consortium (FCDGC) natural history study upon written informed consent (IRB: 16-004682 and IRB: 19-005187; <https://clinicaltrials.gov/ct2/show/NCT04199000?cond=CDG&draw=2&rank=4>). Data were collected prospectively and retrospective data was available for most affected individuals as well, as they have been previously followed up by standard of care. Some of the patients were previously reported in Ligezka et al. (26). The demographic, clinical, and genetic information of the affected individuals is presented in Tables 2 and 3.

*Drug administration.* As part of a single-patient IND protocol, epalrestat was administered orally to an affected individual (corresponding to samples 3 and 15), 3 times per day (tid) before meals in a divided dose (0.8 mg/kg/d; 5 mg tid) (Ono Pharmaceuticals) between day 1 and day 180 of the study (IRB: 19-010017; IND 145262; Protocol PMM2-CDG001/A) (26).

*Controls.* Thirty-six control serum samples were used from unaffected individuals, between the ages of 1–36 (median 10 years) (22 male and 14 female), who donated deidentified residual samples (IRB: 21-012890) (Supplemental Table 2).

*Data extraction.* Clinical data were collected and extracted from the electronic clinical files and electronic laboratory data system. The NPCRS was used to assess disease severity. The NPCRS rating was performed in every affected individual at the time of the laboratory sample collection. The NPCRS is a clinical tool developed to determine the clinical severity and disease progression of individuals affected with CDG in an objective way and is validated for all age groups. Depending on the NPCRS score, cases are scaled in a mild (0–14 points), moderate (15–25 points), or severe (>26 points) category (24–26). CDT analysis was performed by MS analysis of intact transferrin. Mono-oligo/di-oligo and A-oligo/di-oligo ratios of glycosylated transferrin were collected for each affected individual (31, 56).

*Serum protein digestion.* Bicinchoninic acid protein assay (Pierce, Thermo Fisher Scientific) was used to determine protein concentration of serum samples. Equal amount of total protein from each serum sample was vacuum-dried and reconstituted in 8 M urea in 50 mM triethylammonium bicarbonate (TEAB), pH 8.5. Proteins were reduced by incubation with dithiothreitol at a final concentration of 10 mM (MilliporeSigma) at 37°C for 45 minutes. Samples were then cooled to room temperature and alkylated by incubation with 40 mM iodoacetamide (MilliporeSigma) for 15 minutes in the dark at room temperature. The samples were diluted with 50 mM TEAB to a final urea concentration of less than 1 M, and digestion was carried out with trypsin (Worthington, TPCK treated) at a ratio of 1:20 (trypsin/total protein) at 37°C overnight.

*TMT labeling of peptides and glycopeptide enrichment using SEC.* Peptides derived from serum samples of affected individuals ( $n = 7$ ) and controls ( $n = 7$ ) were labeled with TMT (TMTPro, Thermo Fisher Scientific) reagents as per the manufacturer's instructions. After a label check, samples were pooled and the sample was split into 2 aliquots for glycoproteomics and total proteomics experiments. About 90% of dried peptides were resuspended in 100  $\mu\text{L}$  of 0.1% formic acid and loaded onto Superdex peptide 10/300 column (GE Healthcare, now Cytiva) for SEC. Isocratic flow with 0.1% formic acid was used, and 48 fractions were collected over a run time of 130 minutes. Early fractions from SEC collected over 55 minutes starting at 40 minutes after injection were selected based on the UV profile (214 nm, 220 nm) and used for LC-MS/MS. The remaining aliquot (about 10%) of total peptides was cleaned up by C18 TopTips (Glygen) and fractionated by basic reversed phase liquid chromatography (bRPLC) on a C18 column (4.6  $\times$  100 mm column) using an UltiMate 3000 UHPLC System (Thermo Fisher Scientific) for the total proteomics experiment. We used 5 mM ammonium formate in water, pH 9, and 5 mM ammonium formate in 90% acetonitrile, pH 9, as solvent A and B, respectively. Ninety-six fractions collected over 120 minutes were concatenated into

12 fractions. These fractions were dried down in a speed vacuum system and resuspended in 0.1% formic acid for LC-MS/MS analysis.

*Enrichment of glycopeptides using MAX cartridges.* For analysis of the profiling and blinded sets, glycopeptides were enriched from peptide samples using OASIS MAX cartridges (Waters) (57). The MAX cartridges were conditioned with acetonitrile (ACN) thrice, then thrice with 100 mM triethylammonium acetate buffer, followed by thrice with water, and finally thrice with 95% ACN/1% trifluoroacetic acid (TFA). Dried tryptic peptides from sera of affected individuals and controls were reconstituted in 95% ACN/1% TFA. Equal peptide amounts (400 µg) from each sample were used for loading thrice onto the cartridges. The cartridge was then washed thrice with 95% ACN/1% TFA. Glycopeptides were eluted with 50% ACN/0.1% TFA and dried down. The enriched glycopeptides from 18 affected individuals (sample 3 and samples 8 to 24) and 17 matched control serum samples were used for validation experiments using single-run nontargeted analysis. For development and testing of the targeted glycopeptide assay, 16 samples from affected individuals (samples 21 to 36) and 16 matched control samples enriched using MAX cartridges were analyzed using PRM-based targeted analysis.

*LC-MS/MS analysis for discovery and profiling experiments.* Previously published LC-MS/MS parameters (22, 23, 52) were used with some modifications. Briefly, for the TMT-based experiments, 11 early fractions from SEC and 12 concatenated fractions from bRPLC were analyzed by Orbitrap Exploris 480 mass spectrometer (Thermo Fisher Scientific) for glycoproteomics and proteomics, respectively. Peptides were separated by liquid chromatography on an EASY-Spray column (75 µm × 50 cm, PepMap RSLC C18, Thermo Fisher Scientific) packed with 2 µm C18 particles, maintained at 50°C. We used 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B) as solvents. Peptides were trapped on a trap column (100 mm × 2 cm, Acclaim PepMap100 Nano-Trap, Thermo Fisher Scientific) at a flow rate of 20 µL/min. LC separation was performed at a flow rate of 300 nL/min, and the following gradient was used: equilibration at 3% solvent B from 0 to 4 minutes, 3% to 10% sol B from 4 to 10 minutes, 10% to 35% sol B from 10.1 to 125 minutes, 35% to 80% sol B from 125 to 145 minutes, followed by equilibration for next run at 5% sol B for 5 minutes. TMT-based experiments were done in data-dependent acquisition mode with top 15 ions isolated at a window of  $m/z$  0.7 and default charge state of +2. Precursors with charge states ranging from +2 to +7 were considered for MS/MS events. Normalized stepped collision energy was applied to fragment precursors at energies of 15%, 25%, and 40% for glycoproteomics and of 34% for proteomics. Precursor ions were acquired in the Orbitrap mass analyzer in  $m/z$  range of 350–2,000 for glycoproteomics and  $m/z$  350–1,800 for proteomics at a resolution of 120,000 (at  $m/z$  200). Automatic gain control (AGC) for MS and MS/MS were  $10^6$  and  $1 \times 10^5$ , and injection times to reach AGC were 50 ms and 250 ms, respectively. Exclude isotopes feature was set to “on,” and 60-second dynamic exclusion was applied. Data acquisition was performed with option of “lock mass” ( $m/z$  441.12002) for all data.

MAX-enriched glycopeptides were analyzed on an Orbitrap Eclipse mass spectrometer connected to UltiMate 3000 liquid chromatography system. Glycopeptides were trapped on a trap column (100 mm × 2 cm, Acclaim PepMap100 Nano-Trap) at a flow rate of 20 µL/min. LC separation was performed on an analytical column (EASY-Spray 75 µm × 50 cm, C<sub>18</sub> 2 µm, 100 Å) with a flow rate of 300 nL/min with a linear gradient of solvent B (100% ACN, 0.1% formic acid) over a 150-minute gradient. All label-free experiments were done in data-dependent acquisition mode at an isolation window of  $m/z$  1.6 and default charge state of +2. Precursor ions were acquired at a resolution of 120,000 (at  $m/z$  200) and at a resolution of 30,000 (at  $m/z$  200) for fragment ions. Precursor fragmentation was carried out using normalized stepped higher-energy collisional dissociation at 15%, 25%, and 40%. The scans were acquired in top-speed method with 3-second cycle time between MS and MS/MS.

*Targeted LC-MS/MS analysis.* The targeted LC-MS/MS analysis was carried out on an Orbitrap Eclipse mass spectrometer connected to UltiMate 3000 RSLC nano (Thermo Fisher Scientific). MAX-enriched serum glycopeptide samples were injected into trapping column (PepMap C18, 2 cm × 100 µm, Thermo Fisher Scientific) at the flow rate of 10 µL/min. After sample loading, samples were transferred to analytical column (PepMap RSLC C<sub>18</sub> 2 µm, 75 µm × 50 cm, Thermo Fisher Scientific) at the flow rate of 300 nL/min. The analytical gradient started at 3% mobile phase B, and then mobile phase B was increased to 40% over 24 minutes, increased to 95% over 2 minutes, and maintained at 95% for another 7 minutes. Thereafter, mobile phase B was decreased to 2%, and the analytical column was reconditioned for 7 minutes. While an electrospray voltage of 2.0 kV was fixed in positive ion mode, MS spectra at a resolution of 120,000 and MS/MS spectra at a resolution of 30,000 in PRM mode were obtained with stepped collision energy set at



15%, 30%, and 45%. Skyline was used to build the inclusion list, which incorporated the targeted precursor  $m/z$  of glycopeptides and charge state. All targeted PRM raw files were processed in Skyline to generate extracted ion chromatogram and perform peak area integration (58).

*Database searching and data analysis.* Database searching was performed using publicly available software pGlyco Version 2.2.0; pGlyco version 3.0 was used to search for glycosylation at multiple glycosylation sites (37, 59). Glycan databases already available with the software were used, and Uniprot Human Reviewed protein sequences (20,432 entries, downloaded February 1, 2021) were used as protein sequence FASTA file. Fully tryptic cleavage specificity with 2 missed cleavages was used, and precursor and fragment tolerance were set to 10 and 20 ppm, respectively. Carbamidomethylation of cysteine was set as fixed modification and oxidation of methionine as a variable modification. For TMT-based experiments, TMTPro at lysine was set as a fixed modification, and TMTPro at peptide N-terminal was set as variable modification. The results were filtered to retain entries that had 5% FDR at glycopeptide level. Reporter ion quantitation was performed on Proteome Discoverer 2.5 (reporter ion quantifier node), and IDs were matched with quantitation on a scan-to-scan (MS/MS) basis. Glycopeptide-spectrum matches were combined to reflect only unique glycopeptides per search by glycan composition, and peptide sequence and summed-up reporter ion intensities were calculated. Spectra were manually verified for several glycan oxonium ions and quality. All sialic acid glycopeptides' spectra were verified for presence of sialic acid-specific oxonium ions at  $m/z$  274.09, 292.10 and/or 657.23. Multiply glycosylated glycopeptides were verified for evidence of glycosylation (peptide b or y ion+glycan fragment) at both sites. The proteomics data set was searched using Sequest in Proteome Discoverer 2.5.

*Statistics.* Fold-changes of glycopeptides, peptides, and proteins were calculated as average values from affected individuals over average values from control samples. The  $P$  values were obtained using unpaired 2-tailed  $t$  test. The  $P$  values were adjusted (reported  $q$  values) by Benjamini-Hochberg procedure for multiple comparisons on Perseus, and publicly available MetaboAnalyst 5.0 was used to draw PCA score plots (60, 61). A  $q$  value less than 0.05 was considered significant. Perseus was used to draw heatmaps.

Levels of Man<sub>3</sub>GlcNAc<sub>2</sub> glycopeptide from C4 at site Asn<sup>226</sup> were analyzed to determine if they correlated with the NPCRS scores of the affected individuals. Given the differences in protein expression data measurement methodology across study phases along with the ordinal nature of the NPCRS scores, we considered a meta-analytic strategy in which individual nonparametric rank correlation testing was performed for data within a given phase, and the results were combined using  $P$  value combination approaches. Specifically, Kendall's Tau rank correlation test was performed to test the null hypothesis of zero correlation under a 2-sided alternative, with  $P$  values derived using the normal approximation with continuity correction. The  $P$  values resulting from testing across the 3 phases were then combined using a modified form of the Stouffer's weighted Z approach. Weights were defined based on the square root of the case sample size of the corresponding phases. To account for directionality, sign of the corresponding tau estimate was applied to the weighting in the calculation of the numerator of the  $z$ -score (e.g., moderate evidence that is consistently in the same direction will have a compounding effect on the result). For analyses of individual scoring items, similar rank correlation testing was performed as described above. Given the small number of participants per study phase, there is elevated risk of lack of variation for the score item values, preventing hypothesis testing. Under these conditions, usage of Stouffer's combined Z approach was only considered if at least 2 study phases were able to have hypothesis testing performed.

*Study approval.* All affected individuals included in this work are enrolled in the FCDGC natural history study (Mayo Clinic IRB 19-005187; <https://clinicaltrials.gov/ct2/show/NCT04199000?cond=C-DG&draw=2&rank=4>). Written informed consent was obtained from the legally authorized representatives of the individuals prior to study initiation. Control serum samples were used from unaffected individuals who donated deidentified residual samples (IRB: 21-012890).

*Data availability.* The underlying values for data presented in the graphs and supplemental figures are available in the Supporting Data Values file. The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (62) partner repository with the data set identifier PXD042446.

### Author contributions

KG, RB, MS, JK, TK, EM, and AP conceived and designed the research study. KG, RB, MS, JK, NJ, GSS, AJ, MGR, and SU performed the experiments. KG, RB, MS, JK, NJ, GSS, AJ, ANL, SR, MGR, SU, MH, NBL, HHF, MJS, TK, EM, and AP analyzed or interpreted the data. KR, CL, AL, ACE, KS, MJS, and



EM provided samples and clinical data. KG, RB, MS, JK, HHF, TK, EM, and AP wrote the manuscript. KG and RB are co–first authors. The authorship order among co–first authors was agreed upon after discussion by the authors.

## Acknowledgments

We thank Roman Zenka for help with data visualization and Richard K. Kandasamy for helpful discussions. We thank Kaitlin E. Schwartz for support with sample acquisition. This work was supported by the grant titled “Frontiers in Congenital Disorders of Glycosylation” (1U54NS115198–01) from the National Institute of Neurological Disorders and Stroke, National Center for Advancing Translational Sciences, Eunice Kennedy Shriver National Institute of Child Health and Human Development, and Rare Disorders Consortium Disease Network, at the NIH (RB, ANL, SR, KR, CL, ACE, KS, MJS, TK, and EM), and by the Minnesota Partnership for Biotechnology and Medical Genomics research award at the University of Minnesota and Mayo Clinic, Rochester (KS and EM). ACE is supported by grant K08NS118119 from the NIH to Children’s Hospital of Philadelphia. HHF is supported by philanthropic research funding from Rocket Fund and research grant NIH R01DK099551. We also thank the Mayo Clinic DERIVE Office and Mayo Clinic Center for Biomedical Discovery for financial support. This study was also supported by a grant from DBT/Wellcome Trust India Alliance entitled “Center for Rare Disease Diagnosis, Research, and Training” (IA/CRC/20/1/600002) to AP.

Address correspondence to: Akhilesh Pandey, Department of Laboratory Medicine and Pathology, Mayo Clinic, 200 First Street SW, Rochester, Minnesota 55905, USA. Phone: 507.293.9564; Email: pandey.akhilesh@mayo.edu. Or to: Eva Morava, Department of Genomics and Genetic Sciences, Icahn School of Medicine at Mount Sinai Hospital, 1 Gustave L. Levy Place, New York, New York 10029, USA. Phone: 212.659.6841; Email: eva.morava@mssm.edu.

MS’s present address is: Sanford Burnham Prebys Medical Discovery Institute, La Jolla, California, USA.

SU’s present address is: Breast Cancer Now, Institute of Cancer Research, London, United Kingdom.

- Altassan R, et al. International clinical guidelines for the management of phosphomannomutase 2-congenital disorders of glycosylation: diagnosis, treatment and follow up. *J Inherit Metab Dis.* 2019;42(1):5–28.
- Matthijs G, et al. Mutations in PMM2, a phosphomannomutase gene on chromosome 16p13, in carbohydrate-deficient glycoprotein type I syndrome (Jaeken syndrome). *Nat Genet.* 1997;16(1):88–92.
- Vaes L, et al. Genotype-phenotype correlations in PMM2-CDG. *Genes (Basel).* 2021;12(11):1658.
- Schollen E, et al. Lack of Hardy-Weinberg equilibrium for the most prevalent PMM2 mutation in CDG-Ia (congenital disorders of glycosylation type Ia). *Eur J Hum Genet.* 2000;8(5):367–371.
- Ferreira CR, et al. Recognizable phenotypes in CDG. *J Inherit Metab Dis.* 2018;41(3):541–553.
- Matthijs G, et al. Mutations in PMM2 that cause congenital disorders of glycosylation, type Ia (CDG-Ia). *Hum Mutat.* 2000;16(5):386–394.
- Van Scherpenzeel M, et al. Clinical diagnostics and therapy monitoring in the congenital disorders of glycosylation. *Glycoconj J.* 2016;33(3):345–358.
- Jaeken J, et al. Sialic acid-deficient serum and cerebrospinal fluid transferrin in a newly recognized genetic syndrome. *Clin Chim Acta.* 1984;144(2–3):245–247.
- Helander A, et al. Improved HPLC method for carbohydrate-deficient transferrin in serum. *Clin Chem.* 2003;49(11):1881–1890.
- Carchon HA, et al. Diagnosis of congenital disorders of glycosylation by capillary zone electrophoresis of serum transferrin. *Clin Chem.* 2004;50(1):101–111.
- Sanz-Nebot V, et al. Characterization of transferrin glycoforms in human serum by CE-UV and CE-ESI-MS. *Electrophoresis.* 2007;28(12):1949–1957.
- Francisco R, et al. The challenge of CDG diagnosis. *Mol Genet Metab.* 2019;126(1):1–5.
- Lefeber DJ, et al. How to find and diagnose a CDG due to defective N-glycosylation. *J Inherit Metab Dis.* 2011;34(4):849–852.
- Vermeer S, et al. Cerebellar ataxia and congenital disorder of glycosylation Ia (CDG-Ia) with normal routine CDG screening. *J Neurol.* 2007;254(10):1356–1358.
- Witters P, et al. Spontaneous improvement of carbohydrate-deficient transferrin in PMM2-CDG without mannose observed in CDG natural history study. *Orphanet J Rare Dis.* 2021;16(1):102.
- Chen J, et al. Increased clinical sensitivity and specificity of plasma protein N-glycan profiling for diagnosing congenital disorders of glycosylation by use of flow injection-electrospray ionization-quadrupole time-of-flight mass spectrometry. *Clin Chem.* 2019;65(5):653–663.
- Thiesler CT, et al. Glycomic characterization of induced pluripotent stem cells derived from a patient suffering from phosphomannomutase 2 congenital disorder of glycosylation (PMM2-CDG). *Mol Cell Proteomics.* 2016;15(4):1435–1452.
- Byeon SK, et al. Development of a multiomics model for identification of predictive biomarkers for COVID-19 severity: a retrospective cohort study. *Lancet Digit Health.* 2022;4(9):e632–e645.

19. Li Y, et al. Histopathologic and proteogenomic heterogeneity reveals features of clear cell renal cell carcinoma aggressiveness. *Cancer Cell*. 2023;41(1):139–163.
20. Yao Y, et al. Mucus sialylation determines intestinal host-commensal homeostasis. *Cell*. 2022;185(7):1172–1188.
21. Zhao P, et al. Virus-receptor interactions of glycosylated SARS-CoV-2 spike and human ACE2 receptor. *Cell Host Microbe*. 2020;28(4):586–601.
22. Saraswat M, et al. Extensive heterogeneity of glycopeptides in plasma revealed by deep glycoproteomic analysis using size-exclusion chromatography. *Mol Omics*. 2021;17(6):939–947.
23. Saraswat M, et al. TMT-based multiplexed quantitation of N-glycopeptides reveals glycoproteome remodeling induced by oncogenic mutations. *ACS Omega*. 2022;7(13):11023–11032.
24. Achouitar S, et al. Nijmegen paediatric CDG rating scale: a novel tool to assess disease progression. *J Inherit Metab Dis*. 2011;34(4):923–927.
25. Ligezka AN, et al. Patient-reported outcomes and quality of life in PMM2-CDG. *Mol Genet Metab*. 2022;136(2):145–151.
26. Ligezka AN, et al. Sorbitol is a severity biomarker for PMM2-CDG with therapeutic implications. *Ann Neurol*. 2021;90(6):887–900.
27. Tahata S, et al. Liver transplantation recovers hepatic N-glycosylation with persistent IgG glycosylation abnormalities: three-year follow-up in a patient with phosphomannomutase-2-congenital disorder of glycosylation. *Mol Genet Metab*. 2023;138(4):107559.
28. Vals MA, et al. Three families with mild PMM2-CDG and normal cognitive development. *Am J Med Genet A*. 2017;173(6):1620–1624.
29. Yuste-Checa P, et al. The effects of PMM2-CDG-causing mutations on the folding, activity, and stability of the PMM2 protein. *Hum Mutat*. 2015;36(9):851–860.
30. Kjaergaard S, et al. Carbohydrate-deficient glycoprotein syndrome type 1A: expression and characterisation of wild type and mutant PMM2 in *E. coli*. *Eur J Hum Genet*. 1999;7(8):884–888.
31. Biffi S, et al. Carbohydrate-deficient transferrin (CDT) as a biochemical tool for the screening of congenital disorders of glycosylation (CDGs). *Clin Biochem*. 2007;40(18):1431–1434.
32. Vanderboom PM, et al. Proteomic signature of host response to SARS-CoV-2 infection in the nasopharynx. *Mol Cell Proteomics*. 2021;20:100134.
33. Sathe G, et al. Quantitative proteomic analysis of the frontal cortex in Alzheimer's disease. *J Neurochem*. 2021;156(6):988–1002.
34. Garapati K, et al. Defining albumin as a glycoprotein with multiple N-linked glycosylation sites. *J Transl Med*. In press.
35. Bruneel A, et al. Two-dimensional electrophoresis highlights haptoglobin beta chain as an additional biomarker of congenital disorders of glycosylation. *Clin Chim Acta*. 2017;470:70–74.
36. Ferens-Sieczkowska M, et al. Haptoglobin glycoforms in a case of carbohydrate-deficient glycoprotein syndrome. *Glycoconj J*. 1999;16(10):573–577.
37. Zeng WF, et al. Precise, fast and comprehensive analysis of intact glycopeptides and modified glycans with pGlyco3. *Nat Methods*. 2021;18(12):1515–1523.
38. Zuhlsdorf A, et al. Transferrin variants: pitfalls in the diagnostics of congenital disorders of glycosylation. *Clin Biochem*. 2015;48(1–2):11–13.
39. Guillard M, et al. Transferrin mutations at the glycosylation site complicate diagnosis of congenital disorders of glycosylation type I. *J Inherit Metab Dis*. 2011;34(4):901–906.
40. Van Geet C, Jaeken J. A unique pattern of coagulation abnormalities in carbohydrate-deficient glycoprotein syndrome. *Pediatr Res*. 1993;33(5):540–541.
41. Mohamed M, et al. Thyroid function in PMM2-CDG: diagnostic approach and proposed management. *Mol Genet Metab*. 2012;105(4):681–683.
42. Janssen MC, et al. Successful liver transplantation and long-term follow-up in a patient with MPI-CDG. *Pediatrics*. 2014;134(1):e279–e283.
43. Bruneel A, et al. Two-dimensional gel electrophoresis of apolipoprotein C-III and other serum glycoproteins for the combined screening of human congenital disorders of O- and N-glycosylation. *Prot Clin Appl*. 2007;1(3):321–324.
44. Wuhrer M. Glycomics using mass spectrometry. *Glycoconj J*. 2013;30(1):11–22.
45. Thaysen-Andersen M, Packer NH. Advances in LC-MS/MS-based glycoproteomics: getting closer to system-wide site-specific mapping of the N- and O-glycoproteome. *Biochim Biophys Acta*. 2014;1844(9):1437–1452.
46. Riley NM, et al. A pragmatic guide to enrichment strategies for mass spectrometry-based glycoproteomics. *Mol Cell Proteomics*. 2021;20:100029.
47. Delafield DG, Li L. Recent advances in analytical approaches for glycan and glycopeptide quantitation. *Mol Cell Proteomics*. 2021;20:100054.
48. Bagdonaitė I, et al. A strategy for O-glycoproteomics of enveloped viruses—the O-glycoproteome of herpes simplex virus type 1. *PLoS Pathog*. 2022;11(4):e1004784.
49. Kawahara R, et al. Community evaluation of glycoproteomics informatics solutions reveals high-performance search strategies for serum glycopeptide analysis. *Nat Methods*. 2021;18(11):1304–1316.
50. Reiding KR, et al. The benefits of hybrid fragmentation methods for glycoproteomics. *Trends Analyt Chem*. 2018;108:260–268.
51. Thaysen-Andersen M, et al. Maturing glycoproteomics technologies provide unique structural insights into the N-glycoproteome and its regulation in health and disease. *Mol Cell Proteomics*. 2016;15(6):1773–1790.
52. Budhraj R, et al. N-glycoproteomics reveals distinct glycosylation alterations in NGLY1-deficient patient-derived dermal fibroblasts. *J Inherit Metab Dis*. 2023;46(1):76–91.
53. Balakrishnan B, et al. AAV-based gene therapy prevents and halts the progression of dilated cardiomyopathy in a mouse model of phosphoglucomutase 1 deficiency (PGM1-CDG). *Transl Res*. 2023;257:1–14.
54. Shrimal S, Gilmore R. Glycosylation of closely spaced acceptor sites in human glycoproteins. *J Cell Sci*. 2013;126(pt 23):5513–5523.
55. Malaby HL, Kobertz WR. The middle X residue influences cotranslational N-glycosylation consensus site skipping. *Biochemistry*. 2014;53(30):4884–4893.

56. Magalhaes A, et al. Transferrin isoelectric focusing for the investigation of congenital disorders of glycosylation: analysis of a ten-year experience in a Brazilian center. *J Pediatr (Rio J)*. 2020;96(6):710–716.
57. Yang W, et al. Comparison of enrichment methods for intact N- and O-linked glycopeptides using strong anion exchange and hydrophilic interaction liquid chromatography. *Anal Chem*. 2017;89(21):11193–11197.
58. MacLean B, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics*. 2010;26(7):966–968.
59. Liu MQ, et al. pGlyco 2.0 enables precision N-glycoproteomics with comprehensive quality control and one-step mass spectrometry for intact glycopeptide identification. *Nat Commun*. 2017;8(1):438.
60. Tyanova S, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods*. 2016;13(9):731–740.
61. Pang Z, et al. MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res*. 2021;49(w1):W388–W396.
62. Perez-Riverol Y, et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res*. 2022;50(d1):D543–D552.
63. Varki A, et al. Symbol nomenclature for graphical representations of glycans. *Glycobiology*. 2015;25(12):1323–1324.