Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A DNA Methylation Signature in a PMM2-CDG Cohort Reveals a Heterogeneous Epigenetic Footprint Beyond Glycosylation Defects

Albert Alegret García
Barcelona Institute for Global Health
Florencia Epifani
Hospital Sant Joan de Déu Barcelona
Christin Johnsen
University of Göttingen
Cristina Carulla
Hospital Sant Joan de Déu Barcelona
Laura Costa- Comellas
Vall d'Hebron Hospital Universitari
Oscar García
Hospital Universitario de Toledo
Salvador Ibáñez
Hospital Universitario Virgen de la Arrixaca
Marta Jimenez-Moreno
Loreto Martorell
Sant Joan de Déu Research Foundation
Pedro J Moreno-Lozano
Hospital Clínic de Barcelona
Mar O' Callaghan
Hospital Sant Joan de Déu Barcelona
Sinziana Stanescu
Hospital Universitario Ramón y Cajal
Cristina Hernando-Davalillo
Hospital Sant Joan de Déu Barcelona
Roser Urreizti
Sant Joan de Déu Research Foundation
Belén Pérez
Centro de Diagnóstico de Enfermedades Moleculares, Centro de Biología Molecular-SO UAM-CSIC, Universidad Autónoma de Madrid, Campus de Cantoblanco
Juan Ramón Gonzalez

Barcelona Institute for Global Health

Mercedes Serrano

mercedes.serrano@sjd.es

Hospital Sant Joan de Déu Barcelona

Research Article

Keywords: Congenital disorders of glycosylation, DNA methylation, Epigenomics, Epimutations, Episignature, Inherited metabolic disorders, Molecular biomarkers, PMM2-CDG.

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11775710/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

A DNA Methylation Signature in a PMM2-CDG Cohort Reveals a Heterogeneous Epigenetic Footprint Beyond Glycosylation Defects

Albert Alegret-García[1,2], Florencia Epifani[3,4], Christin Johnsen[5], Cristina Carulla[3,4], Laura Costa-Comellas[6], Oscar García[7], Salvador Ibáñez[8], Marta Jiménez Moreno[9], Loreto Martorell[4,10], Pedro J Moreno-Lozano[11], Mar O'Callaghan[3,4,12], Sinziana Stanescu[13], Cristina Hernando-Davalillo[4,10], Roser Urreizti[4,12,14], Belén Pérez[12,15], Juan R. Gonzalez[1,2,*], Mercedes Serrano[3,4,12*]

1. Barcelona Institute for Global Health (ISGlobal), Barcelona, Spain.

2. Centre for Biomedical Network Research on Epidemiology and Public Health (CIBERESP), Barcelona, Spain

3. Paediatric Neurology Department, Hospital Sant Joan de Déu, Barcelona, Spain.

4. Sant Joan de Déu Research Institute, Esplugues de Llobregat, Barcelona, Spain.

5. Department of Pediatrics and Adolescent Medicine, University Medical Centre, Göttingen, Germany.

6. Pediatric Neuromuscular Disorders Unit, Pediatric Neurology Department, Vall D'hebron Barcelona Hospital Campus, Barcelona, Spain.

7. Department of Pediatrics, Hospital Universitario Virgen de la Salud, Toledo, Spain.

8. Department of Pediatric Neurology, Hospital Virgen de la Arrixaca, Murcia, Spain .

9. Pediatric Nephrology Department, Hospital Universitario Virgen del Rocío, Sevilla, Spain.

10. Genetic and Molecular Medicine Department, Hospital Sant Joan de Déu, Barcelona, Spain.

11. Internal Medicine Department, Muscular and Inherited Metabolic Disorders Adults Unit, Hospital Clínic de Barcelona, Barcelona, Spain.

12. Centre for Biomedical Network Research on Rare Diseases (CIBERER), ISCIII, Spain

13. Metabolic Diseases Unit, MetabERN, Pediatric Department, Hospital Universitario Ramón y Cajal, Madrid, Spain

14. Clinical Biochemistry Department, Hospital Sant Joan de Déu, Barcelona, Spain

15. Centro de Diagnóstico de Enfermedades Moleculares. Centro de Biología Molecular. Universidad Autónoma de Madrid. CIBERER. IdiPAZ, Madrid, Spain.

*Co-supervised this work.

Corresponding Author:

Mercedes Serrano, MD PhD

Neurology Department, Hospital Sant Joan de Déu, and U-703 Centre for Biomedical Research on Rare Diseases (CIBER-ER), Instituto de Salud Carlos III, Barcelona, Spain.

Passeig Sant Joan de Déu, 2, 08950 Esplugues, Barcelona, Spain

Tel +34 93 253 2100 E-Mail: mercedes.serrano@sjd.es

Correspondence to: juanr.gonzalez@isglobal.org, mercedes.serrano@sjd.es

Number of words in the summary: 339; **Word count:** 4851

Number of figures: 3; **Number of tables:** 1

Supplementary figures: 3; **Supplementary tables:** 8

ABSTRACT

Background: Phosphomannomutase 2 congenital disorder of glycosylation (PMM2-CDG) is the most prevalent congenital disorder of glycosylation and is characterized by marked clinical heterogeneity. Despite extensive research, genotype-phenotype correlations remain difficult to establish, and biomarkers capable of supporting diagnosis or predicting clinical outcome are limited. DNA methylation episignatures have emerged as valuable biomarkers in several Mendelian disorders, yet their relevance to congenital disorders of glycosylation remains largely unexplored. We investigated whether PMM2-CDG is associated with a detectable peripheral blood DNA methylation episignature and evaluated its reproducibility across individuals with molecularly confirmed PMM2-CDG and other *PMM2*-associated phenotypes.

Results: Genome-wide DNA methylation profiling was performed on 29 samples from 28 individuals, including 25 classical PMM2-CDG, two siblings with hyperinsulinemic hypoglycemia and polycystic kidney disease (PMM2-HIPKD), and one individual with discordant clinical and biochemical findings (PMM2-DP), alongside 32 controls. Differential methylation analysis identified 354 CpGs, subsequently refined to a candidate PMM2-CDG episignature comprising 126-CpGs. The episignature classified a subset of affected individuals but showed heterogeneous performance across the validation cohort, yielding an overall sensitivity of 76.9%, specificity of 99.4%, and AUC of 0.961. Episignature detection was not clearly associated with genotype or disease severity. Individuals with PMM2-HIPKD and the atypical PMM2-DP subject clustered with unaffected controls. Differential methylation analysis identified enrichment of immune-related pathways, including antigen presentation and cell adhesion processes.

DMRs involved genes related to immune regulation (*HLA* loci), skeletal development (*ALPL*, *BMPR1B*, *PBX1*), and embryonic patterning (*HOX* clusters). Epimutation analysis revealed substantial inter-individual variability and suggested potential age-related changes in methylation profiles.

Conclusions: This study provides preliminary evidence for a candidate peripheral blood DNA methylation signature associated with PMM2-CDG. Although detection was variable across affected individuals, the identified epigenetic alterations support the existence of molecular heterogeneity beyond the primary glycosylation defect. Integration of methylation profiles with genetic, clinical, transcriptomic, and functional data may help characterize PMM2-CDG heterogeneity, improve understanding of genotype-phenotype relationships, and facilitate the identification of molecular subgroups. Larger independent and longitudinally characterized cohorts will be required to establish the reproducibility, biological significance, and potential clinical utility of these findings.

Keywords

Congenital disorders of glycosylation; DNA methylation; Epigenomics; Epimutations; Episignature; Inherited metabolic disorders; Molecular biomarkers; PMM2-CDG.

BACKGROUND

Congenital disorders of glycosylation (CDG) comprise a heterogeneous group of congenital metabolic disorders caused by defects in the synthesis, processing, or attachment of glycans to proteins and lipids. Phosphomannomutase 2 deficiency (PMM2-CDG, HGNC: 9115; OMIM #212065) is the most frequently reported disorder of N-linked glycosylation [1], with a recently estimated incidence of approximately 1 in 33,600 live births across North America and Europe [2]. PMM2-CDG is caused by biallelic pathogenic variants in *PMM2*, encoding phosphomannomutase 2, which catalyzes the conversion of mannose-6-phosphate to mannose-1-phosphate, a step in GDP-mannose biosynthesis and protein N-glycosylation [3, 4]. Reduced PMM2 activity impairs glycan biosynthesis and contributes to multisystem disease.

PMM2-CDG exhibits considerable clinical heterogeneity, ranging from severe infantile presentations to comparatively mild phenotypes [5]. Common manifestations include developmental delay, hypotonia, cerebellar syndrome, ataxia, peripheral neuropathy, and stroke-like episodes. Extra-neurological involvement is highly variable and may affect the endocrine, hematological, hepatic, cardiac, renal, and skeletal systems, among others [4, 6, 7]. Clinical manifestations also evolve with age, with some systemic and biochemical abnormalities improving while neurological disability often persists. No disease-modifying treatment is currently approved for PMM2-CDG, and management remains multidisciplinary and symptom-directed [8, 9].

Genotype-phenotype correlations in PMM2-CDG remain challenging to establish. While large cohort studies have associated p.Cys241Ser with milder disease and p.Val231Met,

as well as variants affecting PMM2 folding or dimerization, with more severe phenotypes, considerable clinical variability persists [10, 11]. In addition, the high prevalence of compound heterozygosity further complicates the interpretation of individual variant effects [12].

Additional phenotypic complexity is illustrated by PMM2-associated hyperinsulinemic hypoglycemia and polycystic kidney disease phenotype (PMM2-HIPKD), a distinct clinical presentation associated with the recurrent promoter variant c.-167G>T, either in homozygosity or in compound heterozygosity with a pathogenic coding variant [13–15].

Biomarkers that complement molecular diagnosis and help explain the marked clinical variability observed in PMM2-CDG remain limited. Serum transferrin glycoform analysis is widely used as a first-line screening test for N-glycosylation disorders; however, it is not specific for PMM2-CDG, and abnormal profiles may occasionally be absent, atypical, or inconclusive [16, 17]. The diagnosis is therefore primarily established through the identification of pathogenic or likely pathogenic *PMM2* variants, supported by biochemical and functional studies when required [7, 18].

Molecular studies have provided evidence that the cellular consequences of PMM2 deficiency extend beyond the primary glycosylation defect. Transcriptomic analyses of PMM2-CDG fibroblasts have demonstrated dysregulation of pathways involved in extracellular matrix organization, cellular senescence, tissue development, and cytokine-associated responses [19]. In parallel, circulating microRNA profiling has identified altered regulatory RNA signatures associated with neurological, endocrine, and immune-related processes, further supporting the presence of widespread molecular disturbances across multiple biological systems [20].

DNA methylation episignatures have emerged as complementary molecular biomarkers for Mendelian disorders, particularly neurodevelopmental disorders caused by pathogenic variants affecting chromatin regulators. To date, more than 100 genetic disorders have been associated with distinct and reproducible DNA methylation profiles, providing additional support for molecular diagnosis, variant interpretation, and disease characterization [21]. However, the sensitivity and specificity of episignatures vary considerably across disorders and among affected individuals [22]. Whether this approach is applicable to congenital disorders of glycosylation remains largely unexplored.

In this study, we investigated whether PMM2-CDG is associated with a detectable peripheral blood DNA methylation pattern and evaluated its performance for the classification of molecularly confirmed individuals. We further examined classification results in two individuals with PMM2-HIPKD and one individual presenting with discordant clinical and biochemical findings. Finally, differential DNA methylation analysis and the characterization of individual epimutations were performed to explore the biological processes associated with the observed methylation changes.

SUBJECTS AND METHODS

Design and sample recruitment

Twenty-five individuals with classical PMM2-CDG phenotype, all compound heterozygotes for biallelic pathogenic or likely pathogenic *PMM2* variants were recruited at Sant Joan de Déu Barcelona Children's Hospital between January 2021 and November 2025, providing 26 blood samples. Two monochorionic twins with a PMM2-

HIPKD phenotype, as well as an individual carrying two previously reported *PMM2* variants whose biochemical, radiological, and clinical findings were discordant (PMM2-DP), were also included. Individuals without informed consent or lacking a confirmed molecular diagnosis were excluded. Additional exclusion criteria included the presence of other non-associated neurological, psychiatric, or genetic conditions.

Clinical, epidemiological, and molecular data were systematically collected. Phenotypic features were annotated using Human Phenotype Ontology (HPO) terms (Supplementary Table 1). Disease severity was assessed using the Nijmegen Pediatric CDG Rating Scale (NPCRS) [23], and cerebellar involvement using the International Cooperative Ataxia Rating Scale (ICARS) [24] and neuroimaging evaluation. Molecular variants were classified according to ACMG/AMP guidelines [25].

Ostensibly unaffected control blood samples were obtained from the Sant Joan de Déu Biobank. The donors attended outpatient clinics for routine laboratory testing and had no known neurological or genetic condition. The control cohort comprised 32 samples (16 females and 16 males).

Ethical considerations

The study protocol was approved by the Research & Ethics Committee of the SJD Research Foundation (Internal Code: PIC-174-24). Written informed consent was obtained from all adult participants or the legal guardians of minors or incapacitated individuals. All procedures were conducted in accordance with the Declaration of Helsinki (1964), as revised in 2013.

Blood processing and DNA methylation profiling

Peripheral blood samples were collected in EDTA tubes, and genomic DNA was extracted using standard protocols. DNA methylation profiling was performed using the Illumina Infinium MethylationEPIC v2.0 BeadChip.

EPICv2 data were preprocessed using the MethylCallR package [26]. Leukocyte proportions (B cells, NK cells, CD4+ T cells, CD8+ T cells, monocytes, neutrophils) were inferred using the FlowSorted.Blood.EPIC package [27].

Samples were excluded if the mean of the $\log_2$-transformed median methylated and unmethylated signal intensities was ≤10.5, or if ≥5% of probes had detection P values >0.01. Probes were excluded if detection P values exceeded 0.01 in ≥5% of samples or bead counts were <3 in ≥5% of samples. Additional exclusions comprised probes with SNVs, cross-reactive probes, probes flagged by the manufacturer for insufficient reproducibility, and probes with mismatched genomic positions in the GRCh38 annotation. Functional normalization was applied to obtain beta-values [28], which were transformed using the base-2 logit conversion into M-values for statistical modelling. EPICv2 CpGs were lifted over to EPICv1 probes for public episignature assessment using MethylCallR. EPICv2 annotations were kept for the remaining analyses.

Differentially Methylation analysis and model training

Before episignature discovery, eleven PMM2-CDG samples were reserved for validation (PPM2_V). The fifteen discovery samples (PMM2) were selected to incorporate all

severity groups and most variants present in the cohort. PMM2-HIPKD and PMM2-DP samples were set aside for separate testing (Figure 1 and Table 1).

Epigenome-wide association analysis was performed using M-values in linear regression models adjusted for age, sex, and blood-cell composition. Age was modelled logarithmically in samples from individuals aged ≤17 years and linearly in older samples, adapted from Horvath 2013 [29]. Two age variables were incorporated into the model, one for the young logarithmic trend and another for the adult linear trend. CpGs were considered significant at a Benjamini-Hochberg-adjusted P value < 0.1 and absolute methylation difference ($|\Delta\beta|$) > 0.05.

For feature selection, M-values at significant CpGs were adjusted for covariates using the limma `removeBatchEffect` function. Random Forest hyperparameters were tuned using repeated five-fold cross-validation with 10 repeats, selecting the combination with the highest classification performance (ROC-AUC). Permutation importance was calculated across the 50 training folds, and CpGs were ranked by their mean importance. The 150 highest-ranked CpGs were flagged in each iteration. Candidate episignature subsets were evaluated incrementally, starting with the top 100 CpGs and adding one CpG at each step up to the total significant CpGs.

For each candidate subset of top-ranked CpGs, features were centered and scaled using parameters estimated from the training data. Principal component analysis (PCA) was performed on the transformed values, retaining the smallest number of components that explained at least 80% of the variance. A PCA separation score was quantified using the class-balanced silhouette coefficient calculated from Euclidean distances between samples in the retained PC space [30]. Hierarchical clustering (HC) was performed on

Euclidean distances across the selected CpGs using Ward's minimum-variance method (Ward.D2). The dendrogram was cut into two clusters, and agreement with case-control labels was quantified by the adjusted Rand index (ARI) [31].

K-nearest-neighbour (KNN) and support vector machine (SVM) classifiers were tuned by inner cross-validation. KNN case scores were derived from class- and distance-weighted neighbour votes; SVM case scores were estimated using Platt scaling [32]. For each subset and classifier, a consensus score was calculated as follows:

$$J = 0.25(\frac{PCA\ score\ +\ 1}{2}) + 0.25max(HC\ score, 0) + 0.25(AUC) + 0.25(1 - BB),$$

Where *PCA score* is the class-balanced PCA silhouette, *HC score* is the ARI, *AUC* is the area under the curve for each ML model, *BB* is the balanced Brier error, defined as half the sum of the mean squared probability error in cases and controls. Higher *J* values indicate better combined separation and classification under this prespecified, study-specific criterion. The CpG subset with the highest consensus score was retained as the candidate episignature. Final KNN and SVM classifiers were trained within the caret framework using five-fold cross-validation with 10 repeats for hyperparameter tuning. Class weights were inversely proportional to the number of samples in each group. Public whole-blood DNA methylation data generated using the EPICv1 array were retrieved from Gene Expression Omnibus (GSE180534, n = 32; GSE193730, n = 23; GSE221864, n = 72) and processed using the minfi R package [33], with preprocessing settings comparable to those used for EPICv2 data. The 127 control samples without interventions were used for external evaluation of classifier specificity (Control_Out).

CpG annotations were retrieved from Illumina Infinium EPIC and EPICv2 Bioconductor R annotation packages. missMethyl was used for Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment [34]. Enrichment terms were kept if they reached nominal P<0.05.

Region-based methylation outliers

Differentially methylated regions were identified using DMRcate, adjusting for the same parameters as the limma model. A minimum of 5 consecutive CpGs, FDR < 0.05 and a region 5% mean methylation difference were necessary for significance.

Epimutations were defined as sample-specific regions with methylation values deviating from the control distribution. Sex, age, and estimated blood-cell proportions were regressed out of M-values. The residuals were transformed using the inverse base-2 logit before analysis with the beta-distribution method implemented in the epimutacions R package [35]. Regions containing at least five outlier CpGs were retained. Epimutation burden, defined as the number of retained regions per sample, was estimated.

CpGs at significant HLA-associated loci were screened for SNV-like methylation patterns using MethylToSNP with default parameters [36].

RESULTS

Clinical and Molecular Characterization

The study included 29 blood samples from 28 individuals with compound heterozygous, biallelic *PMM2* variants: 25 with classical PMM2-CDG, two twins with PMM2-HIPKD, and one individual with discordant clinical and biochemical findings (PMM2-DP). One female

participant provided samples at two different time points (S03 and S10; see Table 1). Cerebellar involvement predominated among individuals with classical PMM2-CDG. Two participants died during the first months of life, and clinical information for these individuals was therefore very limited (Supplementary Table 1). The control group comprised 32 in-house controls. The median age at blood sampling was 6.0 years (interquartile range [IQR], 1.6–13.0; range, 0.1–45) in the PMM2-CDG set, and 6.6 years of age (IQR, 2–12.3; range, 0.1–51) in the control cohort.

Missense variants predominated in the cohort (Supplementary Table 2), with most predicted to affect protein folding or catalytic activity (Figure 1) [10]. The most frequent variant was p.Arg141His (n = 11), followed by p.Pro113Leu and p.Phe207Ser (n = 4 each). Two variants, p.Ile96Ser and p.Trp187Ter, had not previously been reported in the literature.

Clinical severity was assessed, when available, using the NPCRS and ICARS, evealing substantial variability across the cohort (Supplementary Table 1). Among participants with the classical PMM2-CDG phenotype, eight were classified as mild, eight as moderate, and seven as severe according to NPCRS criteria. Carriers of the recurrent p.Arg141His variant spanned the full severity spectrum, from mild to severe disease. Overall, no clear genotype-phenotype correlation was identified. Nevertheless, both individuals harbouring p.Cys241Ser were classified as mild, whereas all four carriers of p.Phe207Ser and the three evaluable carriers of p.Pro113Leu exhibited moderate-to-severe disease. The two individuals who died in early infancy, together with another severely affected participant, carried either p.Val231Met or p.Val129Met, each in trans with p.Arg141His; none of the other individuals in the cohort carried these variants. Two

individuals carrying the p.Arg141His/p.Thr237Met genotype were classified as having moderate disease severity. Finally, p.Phe157Ser was identified in both mildly and moderately affected individuals.

The two individuals with PMM2-HIPKD phenotype (S27 and S28) were preterm monochorionic twins who developed neonatal hypertension requiring pharmacological treatment. Renal ultrasound revealed bilateral cysts in both individuals, who also presented with moderate proteinuria. During the perinatal period, both developed hyperinsulinaemic hypoglycaemia, which was treated with diazoxide and hydrochlorothiazide. Neuroimaging revealed nonspecific findings, with no posterior fossa abnormalities or associated neurological manifestations. At their most recent assessment, aged 2 years, both twins exhibit normal neurodevelopment and preserved renal function. Persistent proteinuria is present in one twin and requires ongoing antihypertensive treatment, while both continue therapy for hyperinsulinaemia. Clinical exome sequencing identified two pathogenic *PMM2* variants in trans: the promoter variant c.-167G>T, previously associated with polycystic kidney disease and hyperinsulinaemic hypoglycaemia, and the recurrent c.422G>A (p.Arg141His), present in compound heterozygosis.

Participant S29 is an adolescent with a history of developmental delay affecting language, fine motor skills, coordination, and learning, followed by progressive action tremor from the age of four years. Neuropsychological assessment demonstrated overall cognitive abilities within the normal range, with prominent attentional and executive deficits consistent with predominantly inattentive ADHD, as well as motor impairment consistent with developmental coordination disorder. Neurological

examination revealed mild postural and axial tremor, impaired foot tapping, distal hypertonia, and brisk Achilles reflexes in the absence of other pyramidal signs. No dysmorphic facial or limb features, lipodystrophy, or inverted nipples were observed. Brain MRI, video-EEG, visual evoked potentials, electroretinography, nerve conduction studies, and extensive metabolic investigations, including serum transferrin glycoform analysis on two separate occasions, were all normal. Owing to the discordance between the clinical phenotype and the biochemical findings, he was designated PMM2-DP (Supplementary Table 1). Clinical exome sequencing identified c.91T>C [p.(Phe31Leu)] in trans with the canonical donor splice-site variant c.255+2T>C. PMM2 activity in cultured skin fibroblasts was 58% of the control mean, a value within the range expected for a heterozygous carrier [37].

Discovery and Refinement of the PMM2-CDG Episignature

The discovery set comprised 15 samples from individuals with classical PMM2-CDG (PMM2), whereas the validation set comprised 11 (PMM2_V). The two samples obtained from the same individual at different ages were both assigned to the validation set. Samples from the two siblings with PMM2-HIPKD and the individual with PMM2-DP were reserved for separate evaluation.

Epigenome-wide differential methylation analysis comparing classical PMM2-CDG discovery cases with unaffected controls in the discovery cohort identified 354 differentially methylated CpG sites (Benjamini-Hochberg-adjusted $P < 0.10$; $|\Delta\beta| > 0.05$). Of these sites, 104 (29.4%) were hypomethylated and 250 (70.6%) were hypermethylated. All significant CpGs were ranked by Random Forest importance. Candidate subsets were evaluated using a consensus score combining PCA-based

silhouette scores, hierarchical clustering agreement with case-control labels, and SVM or KNN classification performance, as described in the Methods. The highest consensus score was obtained with 126 CpGs, which were retained as the candidate PMM2-CDG episignature (Figure 2A; Supplementary Table 3).

Hierarchical clustering separated most discovery cases from controls and six validation samples. Five of the 11 validation samples clustered with controls (S04, S14, S16, S22, and S26), including three from individuals with mild clinical presentations (Figure 2B). The same validation samples overlapped with the control distribution along the first two principal components (Figure 2C). The two longitudinal samples obtained from the same individual occupied distinct positions in PCA space and did not cluster together. All perinatal samples (S01, S02 and S03) formed a distinct subcluster.

The two siblings presenting PMM2-HIPKD (S27 and S28) as well as the atypical PMM2-DP individual (S29) clustered with controls rather than with classical PMM2-CDG in both the hierarchical clustering and PCA analysis (Figures 2B and 2C).

Classification performance of the PMM2-CDG Episignature

At a classification threshold of 0.5, SVM assigned five of 11 validation PMM2-CDG samples (45.5%) to the PMM2-CDG group; the remaining six scored below 0.5 (Figure 3A; Supplementary Table 4). Discovery sample S19 scored close to the threshold (SVM score, 0.53). Both PMM2-HIPKD and PMM2-DP samples received scores below 0.05. Supplementary KNN analysis showed concordant classifications except for S17 (PMM2_V; SVM score, 0.37; KNN score, 1).

Classification was also evaluated in 127 public control samples generated using the EPICv1 array and retrieved from Gene Expression Omnibus (GSE180534, GSE193730, and GSE221864). Across all evaluated samples, including the discovery cohort, SVM yielded a sensitivity of 76.9%, specificity of 99.4%, and area under the receiver operating characteristic curve (AUC) of 0.961 (Figure 3B; Supplementary Table 5). KNN performance metrics are provided in Supplementary Table 5.

Functional enrichment of PMM2-CDG-associated methylation changes

To investigate the potential biological relevance of the methylation alterations associated with PMM2-CDG, GO and KEGG enrichment analyses were performed using the complete set of 354 differentially methylated CpGs. These analyses identified enrichment of pathways and biological processes related to immune function. KEGG terms “Cell adhesion molecule (CAMs) interaction” (FDR = 0.017) and “Antigen processing and presentation” (FDR = 0.026) were significantly enriched. GO analysis revealed suggestive enrichment for terms related to “MHC class II protein complex assembly” and “peptide antigen assembly with MHC class II protein complex” (FDR = 0.08 for both; Supplementary Table 6).

Differentially Methylated Regions (DMRs) and Epimutations

DMRcate identified 20 DMRs between individuals with classical PMM2-CDG and controls, each comprising at least five CpGs, with FDR <0.05 and an absolute mean methylation difference >0.05. Of these, thirteen regions were hypomethylated and seven were hypermethylated in PMM2-CDG subjects. The identified DMRs were

annotated to several genes, including *ALPL*, *PBX1*, *BMPR1B*, and *BRSK2*, as well as to loci within the *HOX* and *HLA* gene clusters (Supplementary Table 7).

Individual methylation outlier analysis across all 29 samples identified 185 candidate epimutations, comprising 96 hypomethylated and 89 hypermethylated. Epimutations were detected in 26 of 29 samples, with burdens ranging from 0 to 59 regions per sample (Supplementary Figure 1). The same region was not detected in more than three samples at the same time. All three early-infant samples (S01, S02, and S03) shared a hypomethylated region overlapping *LAX1*. This region was not detected in S10, the later sample from the same individual. Consistent with this observation, the epimutation burden decreased from 16 regions in S03 to seven in S10. Three regions were detected in both PMM2-HIPKD siblings. However, only a hypomethylated region overlapping *SVIL/SVIL-AS1*, which was absent from all other samples, was shared between the two siblings.

DISCUSSION

This study provides preliminary evidence of a candidate peripheral blood DNA methylation episignature associated with PMM2-CDG. The identification of a shared epigenetic pattern in a subset of affected individuals provides a framework for investigating the relationship between epigenetic variation and the clinical and genetic heterogeneity of the disorder. In addition to the candidate episignature, regional and individual methylation alterations point to biological pathways that warrant further investigation. Collectively, these findings motivate the integration of methylation

profiling with genetic, biochemical, and clinical assessments to improve the characterization of phenotypic variability in PMM2-CDG.

Given the marked clinical and genetic heterogeneity of PMM2-CDG, the discovery cohort was deliberately selected to capture a broad spectrum of genotypes and disease severities. This approach enabled the identification of a candidate PMM2-CDG episignature, although its detection was inconsistent across validation samples. Notably, two samples from the same individual, collected during the first month of life (S03) and at 5 years of age (S10), exhibited distinct methylation profiles despite both being assigned to the PMM2-CDG group. The neonatal sample clustered with the neonatal discovery samples (S01 and S02), suggesting a possible age-related influence on the methylation profile. However, given that this observation is based on a single longitudinal comparison, the relative contributions of age-related biological variation and technical factors cannot be determined.

The episignature showed no clear relationship with genotype or clinical severity. To evaluate whether classification performance depended on the analytical approach, KNN was assessed alongside SVM. The two models yielded largely concordant results, differing only for S17 and a small number of external controls, suggesting that the observed heterogeneity was independent of classifier choice. Both individuals carrying p.Cys241Ser (S04 and S17) were classified as negative by SVM, although S17 was classified as positive by KNN. SVM Classifier scores also differed between the two individuals sharing the p.Arg141His/p.Thr237Met genotype (S25 and S26; 0.91 vs 0.30 respectively). Notably, three validation individuals with SVM scores below 0.5 had mild disease (S04, S16 and S22), while the mildly affected discovery case S19 scored close to

the threshold (SVM score= 0.53). However, the severely affected validation participant S14 was misclassified as control (SVM score= 0.02), suggesting that clinical severity alone does not explain the classification. Additional factors, such as age, blood cell composition, and technical variation, may have contributed to the observed heterogeneity.

The clinical heterogeneity observed in our cohort provides important context for the variable detection of the episignature. Individuals carrying p.Cys241Ser exhibited mild phenotypes, consistent with consistent with previous reports associating this variant with a comparatively mild disease course [10]. In contrast, p.Arg141His, a catalytically inactive allele, was observed across the full spectrum of disease severity, supporting the notion that clinical outcome is largely influenced by the variant present on the second allele. In our cohort, p.Val129Met and p.Val231Met, each identified in trans with p.Arg141His, were associated with severe disease, including two cases of early mortality. These observations underscore the importance of considering the allelic context when interpreting genotype-phenotype relationships. Experimental studies have shown that interactions between different PMM2 variant subunits can partially compensate for their functional impairment, highlighting the limitations of predicting disease severity on individual variants alone [38]. Nevertheless, phenotypic variability among individuals sharing the same genotypes suggests that additional genetic, environmental, or other modifying factors contribute to disease expression [12, 39].

In our cohort, both PMM2-HIPKD siblings carrying c.-167G>T in trans with p.Arg141His exhibited methylation profiles that clustered with unaffected controls. PMM2-HIPKD predominantly affects the kidneys and pancreatic β cells. One potential contributor to

this specificity is hepatocyte nuclear factor 4 alpha (*HNF4A*), a transcription factor with regulatory functions in the liver, pancreas, and kidney [40]. Pathogenic *HNF4A* variants have previously been associated with neonatal hyperinsulinism and renal Fanconi syndrome [41]. Experimental studies have shown that c.-167G>T promoter variant impairs ZNF143 binding, resulting in reduced PMM2 promoter activity [14]. A proposed mechanistic model links this defect to altered regulatory interactions within CTCF-associated chromatin loops, potentially affecting HNF4A-dependent regulation in specific tissues [14, 42] (Supplementary Figure 2). However, this mechanism has not been experimentally confirmed at the *PMM2* locus. Moreover, recent acute-depletion studies have challenged previous interpretations of the role of ZNF143 in chromatin looping and identified CTCF cross-reactivity in a commonly used anti-ZNF143 antibody, raising questions about earlier findings [43].

The PMM2-HIPKD siblings also shared a hypomethylated region overlapping *SVIL/SVIL-AS1* that was absent from all other samples (Supplementary Figure 3). Its relevance to PMM2-HIPKD is unknown. However, given that the two individuals are monozygotic twins, shared genetic or environmental factors could also account for this finding. Nevertheless, each twin also exhibited unique epimutations that were not detected in the co-twin (Supplementary Table 8).

S29 carries c.91T>C [p.(Phe31Leu)] in trans with the canonical donor splice-site variant c.255+2T>C. Both variants have been previously reported in unrelated individuals (Supplementary Table 2). The p.(Phe31Leu) variant was initially described in trans with p.Arg141His in an individual with classical multisystemic PMM2-CDG [44]. It was subsequently identified in a second individual with an atypical presentation. This

patient, a 14-year-old girl with autism spectrum disorder, hearing loss, and ptosis, carried p.Phe31Leu and p.Arg123Gln inherited from her father and mother, respectively [45]. Similar to S29, her phenotype did not fit the classical PMM2-CDG presentation. However, no glycosylation or enzymatic studies were reported, and the authors did not establish a diagnosis of PMM2-CDG.

In S29, the phenotype was not suggestive of PMM2-CDG, serum transferrin glycoform analysis was repeatedly normal at different ages, neuroimaging revealed no cerebellar abnormalities, and PMM2 enzymatic activity was within the range expected for a heterozygous carrier. The discordance between the genetic findings, clinical presentation, and biochemical results warrants re-evaluation of both the diagnosis and the evidence supporting the pathogenic classification of each variant. Taken together, the reported observations suggest that p.(Phe31Leu) is not a fully deleterious allele and that its phenotypic consequences may depend on the the variant present on the second allele and, possibly, on additional genetic or enviromental modifiers. Nevertheless, no functional data are currently available for p.(Phe31Leu), and the classical PMM2-CDG phenotype reported in combination with p.Arg141His is difficult to reconcile with the carrier-range PMM2 activity observed in S29. Furthermore, given the discordant phenotype and the enzyme activity observed in the carrier range, the pathogenic effect of the c.255+2T>C variant has not been directly assessed at the transcript level in this individual. Functional characterisation of p.(Phe31Leu), together with RNA studies of c.255+2T>C in fibroblasts, would help to clarify the contribution of each allele.

This case highlights the importance of integrating genetic, biochemical, and clinical evidence before establishing a definitive molecular diagnosis or pathogenic

classification. Premature variant classification may delay the identification of an alternative diagnosis and compromise the accuracy of genetic counselling for the family.

Regarding the methylation alterations, enrichment analyses identified pathways related to antigen presentation and cell adhesion. However, the associated GO terms reached only suggestive significance (FDR = 0.08). These enrichment signals were largely driven by *HLA*-related loci, and a DMR was also identified overlapping *HLA-DPA1/HLA-DPB1*. None of the *HLA*-associated CpGs evaluated showed SNV-like methylation patterns according to MethylToSNP [36]. The pathway-level signal enrichment is consistent with previous transcriptomic studies reporting altered cytokine signalling and inflammatory responses in PMM2-CDG [19, 46], as well as with evidence of immune-related dysregulation in circulating miRNA profiles [20]. The relationship between these methylation changes and broader disease mechanisms, including immune dysfunction and susceptibility to recurrent and severe infections, warrants further investigation [47].

Other DMRs overlapped genes involved in skeletal development and cellular differentiation, including *ALPL*, *BMPR1B*, and *PBX1*, as well as loci within *HOX* gene clusters, which are implicated in embryonic patterning. These findings are also consistent with previous transcriptomic studies reporting dysregulation of bone-related pathways and extracellular matrix components in PMM2-CDG fibroblasts [19]. In addition, shared *LAX1*-associated hypomethylation was observed in three early-infant samples but was absent in the later sample obtained from one of these individuals, suggesting a potential age-related effect on the methylation profile.

Several limitations should be acknowledged. The relatively small cohort and substantial clinical and genetic heterogeneity limited the precision of classification performance

estimates and the assessment of genotype-phenotype relationships. The absence of individuals with other congenital disorders of glycosylation or clinically overlapping conditions hindered assessment of disease specificity. Residual confounding by age, blood cell composition, and technical variation also remains possible despite adjustment. Finally, the limited availability of longitudinal samples prevented a systematic assessment of within-individual episignature stability, particularly during adulthood. The PMM2-HIPKD subgroup comprised a single pair of monochorionic twins, who cannot be considered independent observations. Consequently, no conclusions can be drawn regarding the presence or absence of a methylation signature associated with this phenotype. Furthermore, the lack of matched transcriptomic and functional data from the same individuals and sampling time points limited the assessment of the regulatory consequences of the observed methylation alterations and their relationship to metabolic dysfunction. Larger independent cohorts, including individuals with other congenital disorders of glycosylation, longitudinal sampling, and complementary subgroup analyses, will be required to establish reproducibility, clarify biological significance, and evaluate potential clinical utility.

Our study contributes to a largely unexplored area in inherited metabolic disorders. To date, epigenomic investigations in inborn errors of metabolism have been scarce and have predominantly relied on patient-derived fibroblasts [48,49]. While informative, fibroblast sampling is invasive, not routinely available in clinical practice, and poorly suited for longitudinal studies requiring serial assessments. In contrast, peripheral blood is readily accessible and amenable to repeated sampling, making it a more practical substrate for biomarker discovery and longitudinal monitoring. To our knowledge, this

represents the first genome-wide DNA methylation study in a congenital disorder of glycosylation.

CONCLUSIONS

Our work provides preliminary evidence for a candidate peripheral blood DNA methylation episignature associated with PMM2-CDG, with heterogeneous detection across affected individuals. Although the episignature was not consistently detected across all affected individuals, it identified a subset of cases sharing a common epigenetic profile and revealed methylation alterations in pathways previously implicated in PMM2-CDG pathophysiology, including immune regulation, extracellular matrix biology, and developmental processes. Beyond its potential diagnostic value, DNA methylation profiling may provide complementary insights into disease mechanisms, phenotypic variability, and the molecular basis of genotype-phenotype relationships in PMM2-CDG. Integration of methylation data with genetic, clinical, transcriptomic, and functional information may help identify biologically relevant patient subgroups and improve disease stratification. These findings provide a foundation for further investigation of epigenetic variation in congenital disorders of glycosylation and for evaluating the utility of methylation signatures as biomarkers for diagnosis, variant interpretation, patient stratification, disease monitoring, and assessment of therapeutic interventions. Future studies in larger, independent, and longitudinally characterized cohorts will be necessary to establish the reproducibility, biological significance, and clinical applicability of these observations.

REFERENCES

1. Piedade A, Francisco R, Jaeken J, et al (2022) Epidemiology of congenital disorders of glycosylation (CDG)—overview and perspectives. J Rare Dis 1:3. https://doi.org/10.1007/s44162-022-00003-6

2. Edmondson AC, Honzík T, Lam C, et al (2025) Incidence and prevalence of phosphomannomutase 2-congenital disorder of glycosylation: Past, present, and future. Mol Genet Metab 146:109188. https://doi.org/10.1016/j.ymgme.2025.109188

3. Lam C, Krasnewich DM (2021) PMM2-CDG. In: Adam MP, Bick S, Mirzaa GM, et al (eds) GeneReviews®. University of Washington, Seattle, Seattle (WA)

4. He M, Zhou X, Wang X (2024) Glycosylation: mechanisms, biological functions and clinical implications. Signal Transduct Target Ther 9:194. https://doi.org/10.1038/s41392-024-01886-1

5. Vals M-A, Morava E, Teeäär K, et al (2017) Three families with mild PMM2-CDG and normal cognitive development. Am J Med Genet A 173:1620–1624. https://doi.org/10.1002/ajmg.a.38235

6. Muthusamy K, Perez-Ortiz JM, Ligezka AN, et al (2024) Neurological manifestations in PMM2-congenital disorders of glycosylation (PMM2-CDG): Insights into clinico-radiological characteristics, recommendations for follow-up, and future directions. Genet Med Off J Am Coll Med Genet 26:101027. https://doi.org/10.1016/j.gim.2023.101027

7. Altassan R, Péanne R, Jaeken J, et al (2019) International clinical guidelines for the management of phosphomannomutase 2-congenital disorders of glycosylation:

Diagnosis, treatment and follow up. J Inherit Metab Dis 42:5–28. https://doi.org/10.1002/jimd.12024

8. Martínez-Monseny AF, Bolasell M, Callejón-Póo L, et al (2019) AZATAX: Acetazolamide safety and efficacy in cerebellar syndrome in PMM2 congenital disorder of glycosylation (PMM2-CDG). Ann Neurol 85:740–751. https://doi.org/10.1002/ana.25457

9. Serrano M, Epifani F, Marino R, McWilliams P (2026) Results From a Phase 2, Open-Label Study Evaluating the Safety, Tolerability, and Effect on Ataxia of GLM101 in Three Adult Patients With PMM2-CDG. J Inherit Metab Dis 49:e70240. https://doi.org/10.1002/jimd.70240

10. Pajusalu S, Vals M-A, Serrano M, et al (2024) Genotype/Phenotype Relationship: Lessons From 137 Patients With PMM2-CDG. Hum Mutat 2024:8813121. https://doi.org/10.1155/2024/8813121

11. Holubova V, Barone R, Grunewald S, et al (2025) Clinical severity and cardiac phenotype in phosphomannomutase 2-congenital disorders of glycosylation : Insights into genetics and management recommendations. J Inherit Metab Dis 48:e12826. https://doi.org/10.1002/jimd.12826

12. Oliveira T, Ferraz R, Azevedo L, et al (2025) A comprehensive update of genotype-phenotype correlations in PMM2-CDG: insights from molecular and structural analyses. Orphanet J Rare Dis 20:207. https://doi.org/10.1186/s13023-025-03669-5

13. Acharya R, Upadhyay K (2022) Hyperinsulinemic Hypoglycemia Due to PMM2 Mutation in Two Siblings with Autosomal Recessive Polycystic Kidney Disease. Pediatr Rep 14:444–449. https://doi.org/10.3390/pediatric14040052

14. Cabezas OR, Flanagan SE, Stanescu H, et al (2017) Polycystic Kidney Disease with Hyperinsulinemic Hypoglycemia Caused by a Promoter Mutation in Phosphomannomutase 2. J Am Soc Nephrol JASN 28:2529–2539. https://doi.org/10.1681/ASN.2016121312

15. Kiparissi F, Dastamani A, Palm L, et al (2023) Phosphomannomutase 2 (PMM2) variants leading to hyperinsulinism-polycystic kidney disease are associated with early-onset inflammatory bowel disease and gastric antral foveolar hyperplasia. Hum Genet 142:697–704. https://doi.org/10.1007/s00439-023-02523-7

16. Hall PL, Liedke K, Turgeon C, et al (2024) Sensitivity of transferrin isoform analysis for PMM2-CDG. Mol Genet Metab 143:108564. https://doi.org/10.1016/j.ymgme.2024.108564

17. Raynor A, Bruneel A, Vermeersch P, et al (2024) “Hide and seek”: Misleading transferrin variants in PMM2-CDG complicate diagnostics. Proteomics Clin Appl 18:e2300040. https://doi.org/10.1002/prca.202300040

18. Granjo P, Pascoal C, Gallego D, et al (2024) Mapping the diagnostic odyssey of congenital disorders of glycosylation (CDG): insights from the community. Orphanet J Rare Dis 19:407. https://doi.org/10.1186/s13023-024-03389-2

19. Gallego D, Serrano M, Cordoba-Caballero J, et al (2024) Transcriptomic analysis identifies dysregulated pathways and therapeutic targets in PMM2-CDG. Biochim Biophys Acta BBA - Mol Basis Dis 1870:167163. https://doi.org/10.1016/j.bbadis.2024.167163

20. Epifani F, Cabus L, Nolasco GA, et al (2025) Exploring a Circulating miRNA Signature for PMM2-CDG: Initial Insights Toward Diagnosis, Stratification, and Monitoring. J Inherit Metab Dis 48:e70104. https://doi.org/10.1002/jimd.70104

21. Alegret-García A, Cáceres A, Sevilla-Porras M, et al (2025) Analysis Methods for Diagnosing Rare Neurodevelopmental Diseases with Episignatures: A Systematic Review of the Literature. Biomedicines 13:3043. https://doi.org/10.3390/biomedicines13123043

22. Husson T, Lecoquierre F, Nicolas G, et al (2024) Episignatures in practice: independent evaluation of published episignatures for the molecular diagnostics of ten neurodevelopmental disorders. Eur J Hum Genet EJHG 32:190–199. https://doi.org/10.1038/s41431-023-01474-x

23. Achouitar S, Mohamed M, Gardeitchik T, et al (2011) Nijmegen paediatric CDG rating scale: a novel tool to assess disease progression. J Inherit Metab Dis 34:923–927. https://doi.org/10.1007/s10545-011-9325-5

24. Trouillas P, Takayanagi T, Hallett M, et al (1997) International Cooperative Ataxia Rating Scale for pharmacological assessment of the cerebellar syndrome. The Ataxia Neuropharmacology Committee of the World Federation of Neurology. J Neurol Sci 145:205–211. https://doi.org/10.1016/s0022-510x(96)00231-6

25. Richards S, Aziz N, Bale S, et al (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med Off J Am Coll Med Genet 17:405–424. https://doi.org/10.1038/gim.2015.30

26. Yang H-H, Han M-R (2024) MethylCallR : a comprehensive analysis framework for Illumina Methylation Beadchip. Sci Rep 14:27026. https://doi.org/10.1038/s41598-024-77914-5

27. Salas LA, Koestler DC, Butler RA, et al (2018) An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina

HumanMethylationEPIC BeadArray. Genome Biol 19:64. https://doi.org/10.1186/s13059-018-1448-7

28. Fortin J-P, Labbe A, Lemire M, et al (2014) Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biol 15:503. https://doi.org/10.1186/s13059-014-0503-2

29. Horvath S (2013) DNA methylation age of human tissues and cell types. Genome Biol 14:R115. https://doi.org/10.1186/gb-2013-14-10-r115

30. Rousseeuw PJ (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 20:53–65. https://doi.org/10.1016/0377-0427(87)90125-7

31. Ward JH (1963) Hierarchical Grouping to Optimize an Objective Function. J Am Stat Assoc 58:236–244. https://doi.org/10.1080/01621459.1963.10500845

32. Platt JC (2000) Probabilities for SV Machines. In: Smola AJ, Bartlett P, Schölkopf B, Schuurmans D (eds) Advances in Large-Margin Classifiers. The MIT Press, pp 61–74

33. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al (2014) Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 30:1363–1369. https://doi.org/10.1093/bioinformatics/btu049

34. Phipson B, Maksimovic J, Oshlack A (2016) missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. Bioinformatics 32:286–288. https://doi.org/10.1093/bioinformatics/btv560

35. Ruiz-Arenas C, Abarrategui L, Hernandez-Ferrer C, et al (2023) Epimutation detection in the clinical context: guidelines and a use case from a new Bioconductor package. Epigenetics 18:2230670. https://doi.org/10.1080/15592294.2023.2230670

36. LaBarre BA, Goncearenco A, Petrykowska HM, et al (2019) MethylToSNP: identifying SNPs in Illumina DNA methylation array data. Epigenetics Chromatin 12:79. https://doi.org/10.1186/s13072-019-0321-6

37. Westphal V, Peterson S, Patterson M, et al (2001) Functional significance of PMM2 mutations in mildly affected patients with congenital disorders of glycosylation Ia. Genet Med Off J Am Coll Med Genet 3:393–398. https://doi.org/10.1097/00125817-200111000-00003

38. Andreotti G, Monti MC, Citro V, Cubellis MV (2015) Heterodimerization of Two Pathological Mutants Enhances the Activity of Human Phosphomannomutase2. PloS One 10:e0139882. https://doi.org/10.1371/journal.pone.0139882

39. Kjaergaard S, Schwartz M, Skovby F (2001) Congenital disorder of glycosylation type Ia (CDG-Ia): phenotypic spectrum of the R141H/F119L genotype. Arch Dis Child 85:236–239. https://doi.org/10.1136/adc.85.3.236

40. Drewes T, Senkel S, Holewa B, Ryffel GU (1996) Human hepatocyte nuclear factor 4 isoforms are encoded by distinct and differentially expressed genes. Mol Cell Biol 16:925–931. https://doi.org/10.1128/mcb.16.3.925

41. Hamilton AJ, Bingham C, McDonald TJ, et al (2014) The HNF4A R76W mutation causes atypical dominant Fanconi syndrome in addition to a β cell phenotype. J Med Genet 51:165–169. https://doi.org/10.1136/jmedgenet-2013-102066

42. Bailey SD, Zhang X, Desai K, et al (2015) ZNF143 provides sequence specificity to secure chromatin interactions at gene promoters. Nat Commun 6:6186. https://doi.org/10.1038/ncomms7186

43. Magnitov MD, Maresca M, Alonso Saiz N, et al (2025) ZNF143 is a transcriptional regulator of nuclear-encoded mitochondrial genes that acts independently of looping and CTCF. Mol Cell 85:24-41.e11. https://doi.org/10.1016/j.molcel.2024.11.031

44. Al Teneiji A, Bruun TUJ, Sidky S, et al (2017) Phenotypic and genotypic spectrum of congenital disorders of glycosylation type I and type II. Mol Genet Metab 120:235–242. https://doi.org/10.1016/j.ymgme.2016.12.014

45. Arteche-López A, Gómez Rodríguez MJ, Sánchez Calvin MT, et al (2021) Towards a Change in the Diagnostic Algorithm of Autism Spectrum Disorders: Evidence Supporting Whole Exome Sequencing as a First-Tier Test. Genes 12:560. https://doi.org/10.3390/genes12040560

46. Pascoal C, Granjo P, Kodríková R, et al (2025) Immunopathology in PMM2-CDG: Defective glycosylation impact in the TNFα -TNFR1 signalling pathway. Front Immunol 16:1655354. https://doi.org/10.3389/fimmu.2025.1655354

47. Pascoal C, Francisco R, Mexia P, et al (2024) Revisiting the immunopathology of congenital disorders of glycosylation: an updated review. Front Immunol 15:1350101. https://doi.org/10.3389/fimmu.2024.1350101

48. Leal AF, Pachajoa H, Tomatsu S (2025) Epigenetic landscape in lysosomal storage disorders: mechanisms and modulation. Front Genet 16:1679497. https://doi:10.3389/fgene.2025.1679497

49. Tejedor JR, Soriano-Sexto A, Beccari L, et al (2025) Integration of multi-omics layers empowers precision diagnosis through unveiling pathogenic mechanisms on maple syrup urine disease. J Inherit Metab Dis 48:e12829. https://doi.org/10.1002/jimd.12829

List of abbreviations

ACMG: American College of Medical Genetics and Genomics

ARI: Adjusted Rand index

AUC: Area under the curve

BH: Benjamini-Hochberg

CDG: Congenital disorder(s) of glycosylation

DMR: Differentially methylated region

DMP: Differentially methylated position

FDR: False discovery rate

GO: Gene Ontology

HC: Hierarchical clustering

HPO: Human Phenotype Ontology

ICARS: International Cooperative Ataxia Rating Scale

IQR: Interquantile range.

KEGG: Kyoto Encyclopedia of Genes and Genomes

KNN: K-nearest neighbours

NPCRS: Nijmegen Pediatric CDG Rating Scale

PCA: Principal component analysis

PMM2-CDG: Phosphomannomutase 2 congenital disorder of glycosylation

PMM2-HIPKD: PMM2-associated hyperinsulinemic hypoglycemia and polycystic kidney disease phenotype (study-specific designation)

PMM2-DP: PMM2-associated discordant phenotype (study-specific designation)

RF: Random Forest

ROC: Receiver operating characteristic

SNV: Single nucleotide variant

SVM: Support vector machine

Ethics approval and consent to participate

Ethical permission for the study was obtained from the Research & Ethics Committee of the SJD Research Foundation (Internal Code PIC-174-24). The cases or their legal guardians gave their written informed consent. Children and adults gave their assent, or written informed consent when possible.

Consent for publication

Written consent for publication of the clinical data and study results was obtained from the families as part of the informed consent procedure.

Availability of data and material

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request. The public datasets supporting the conclusions of this article are available in Gene Expression Omnibus (accession: GSE180534, GSE193730, and GSE221864).

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This study was funded by the Instituto de Salud Carlos III (ISCIII) through projects DTS26/00119, and PI24/00482, the latter co-funded by the European Union; and the

Generalitat de Catalunya through project SLT042/25/000011; and the Spanish Ministry of Science, Innovation and Universities, the Spanish State Research Agency (AEI), and the European Regional Development Fund through project PID2025-172854OB-I00. Additional support was provided by AGAUR STEP (Strategic Technologies for Europe Platform), grant 2025 STEP 00130, co-funded by the European Union under the Catalonia 2021-2027 European Social Fund Plus (ESF+) Programme. Institutional support was provided by the Generalitat de Catalunya through the CERCA Program and the Consolidated Group on HEALTH ANALYTICS (2025 SGR 00709). ISGlobal acknowledges support from the Ministerio de Ciencia, Innovación y Universidades through the Centro de Excelencia Severo Ochoa 2024–2028 Program (CEX2023-001290-S). Dr Epifani was supported by grant FI22/00218, funded by ISCIII and the European Union, NextGenerationEU through the Mecanismo de Recuperación y Resiliencia (MRR).

The funding sources had no role in the design or conduct of the study; collection, management, analysis, or interpretation of the data; preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication.

Authors' contributions

AAG, FE, RU, JRG and MS participated in the conception and design of the study. AAG, FE and MS participated in the acquisition and analysis of data. AAG, FE, RU and MS participated in writing the original draft. All authors have participated in the final writing, review and edition. RU, JRG and MS have supervised the work. JRG and MS have participated in the funding acquisition. Each author has seen, reviewed, and approved this version of the manuscript and takes full responsibility for it.

Acknowledgements

We are indebted to the “Biobanc de l’Hospital Infantil Sant Joan de Déu per a la Investigació” integrated in the Spanish Biobank Network of ISCIII for biological sample processing and procurement. We would like to thank the participants and their families who kindly donated samples. We thank the Institut Josep Carreras Genomics Unit and staff for their contribution to this publication.

Tables

Table 1. Samples included in the study.

Sample ID	Genotype*	Phenotype	Analysis set
S01	c.385G>A / c.422G>A	PMM2-CDG	Discovery
S02	c.422G>A / c.691G>A	PMM2-CDG	Discovery
S03 / S10	c.385G>A / c.422G>A	PMM2-CDG	Validation
S04	c.368G>A / c.722G>C	PMM2-CDG	Validation
S05	c.338C>T / c.[353C>G;550C>A]	PMM2-CDG	Validation
S06	c.256-1G>C / c.199G>A	PMM2-CDG	Discovery
S07	c.287T>G / c.422G>A	PMM2-CDG	Discovery
S08	c.415G>A / c.470T>C	PMM2-CDG	Validation
S09	c.317A>T / c.620T>C	PMM2-CDG	Discovery
S11	c.338C>T / c.620T>C	PMM2-CDG	Discovery
S12	c.338C>T / c.422G>A	PMM2-CDG	Discovery
S13	c.317A>T / c.620T>C	PMM2-CDG	Discovery
S14	c.205C>T / c.561G>A	PMM2-CDG	Validation
S15	c.357C>A / c.422G>A	PMM2-CDG	Discovery
S16	c.523+3A>G / c.484C>T	PMM2-CDG	Validation
S17	c.359T>C / c.722G>C	PMM2-CDG	Validation
S18	c.97C>T / c.131T>C	PMM2-CDG	Discovery
S19	c.682G>T / c.710C>G	PMM2-CDG	Discovery
S20	c.131T>C / c.620T>C	PMM2-CDG	Discovery
S21	c.470T>C / c.484C>T	PMM2-CDG	Discovery
S22	c.278A>C / c.422G>A	PMM2-CDG	Validation
S23	c.205C>T / c.470T>C	PMM2-CDG	Discovery
S24	c.338C>T / c.368G>A	PMM2-CDG	Discovery
S25	c.422G>A / c.710C>T	PMM2-CDG	Validation

S26	c.422G>A / c.710C>T	PMM2-CDG	Validation
S27	c.-167G>T / c.422G>A	PMM2-HIPKD	Separate testing
S28	c.-167G>T / c.422G>A	PMM2-HIPKD	Separate testing
S29	c.91T>C / c.255+2T>C	PMM2-DP	Separate testing

* Variants are described according to NM_000303.3.

Variants enclosed in square brackets lie in cis. Further information on each variant is provided in Supplementary Table 2.

S03 and S10 are two samples from the same individual, collected at different ages.

Discovery: samples used for episignature discovery. Validation: samples not used for discovery and reserved for independent evaluation. Separate testing: individuals with non-classical PMM2-associated phenotypes, evaluated separately.

PMM2-CDG: classical phosphomannomutase 2 congenital disorder of glycosylation; PMM2-HIPKD: PMM2-associated hyperinsulinemic hypoglycemia and polycystic kidney disease; PMM2-DP: PMM2-associated discordant phenotype.

Figure legends

Figure 1: Study design, *PMM2* variants, and clinical characteristics.

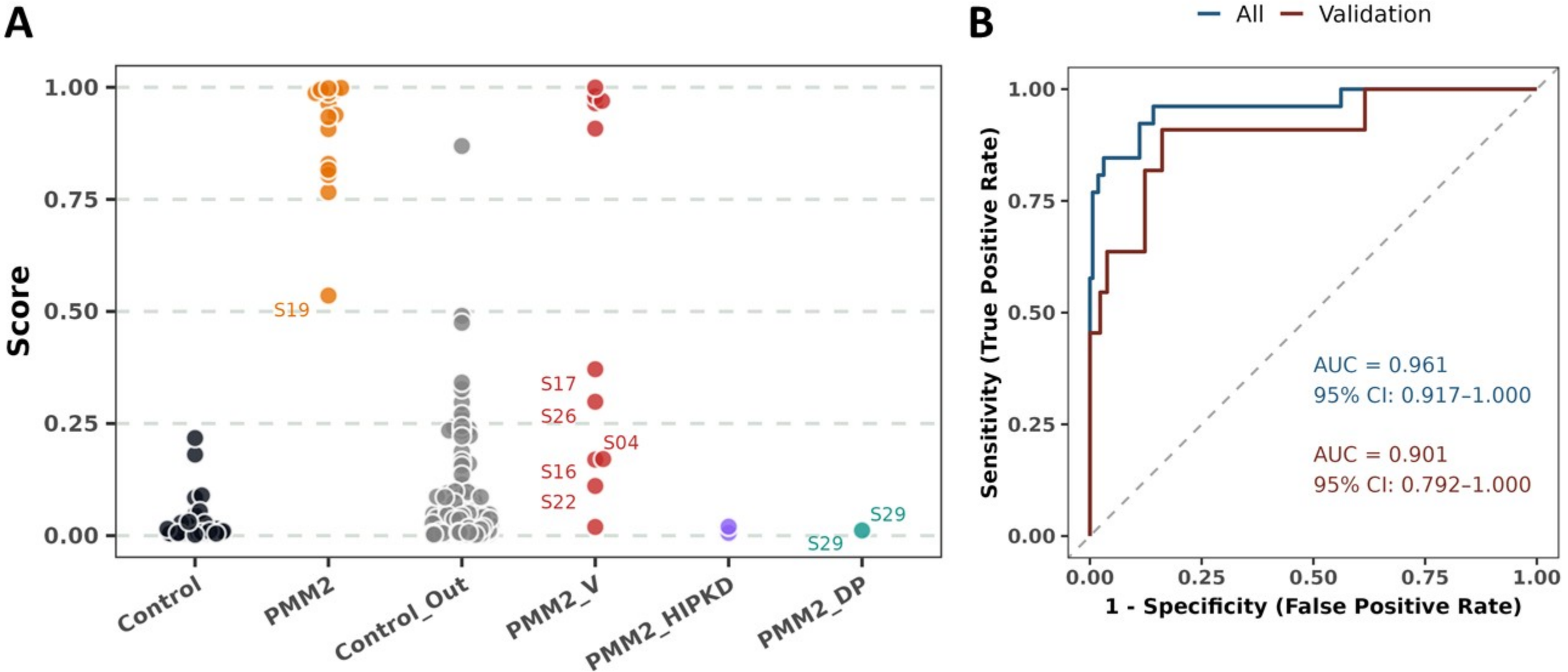
(A) Study design and sample allocation. Counts represent samples rather than unique individuals; S03 and S10 are samples collected from the same individual at different ages and are counted separately. (B) *PMM2* variants identified in all participants with classical PMM2-CDG and PMM2-associated hyperinsulinemic hypoglycemia and polycystic kidney disease (PMM2-HIPKD). Variants identified in the individual with discordant clinical and biochemical findings (PMM2-DP) were excluded. The variant plot was generated using ProteinPaint (accessed 2 November 2026). (C) Clinical characteristics of the recruited participants. Group labels indicate discovery cases (PMM2-CDG), validation cases [PMM2-CDG (V)], and individuals with alternative PMM2-associated phenotypes (PMM2-HIPKD and PMM2-DP). Severity categories were based on Nijmegen Pediatric CDG Rating Scale (NPCRS) scores: mild, 0–14; moderate, 15–25; severe, ≥26. NR, not reported.

Figure 2: Discovery and optimization of the candidate PMM2-CDG episignature.

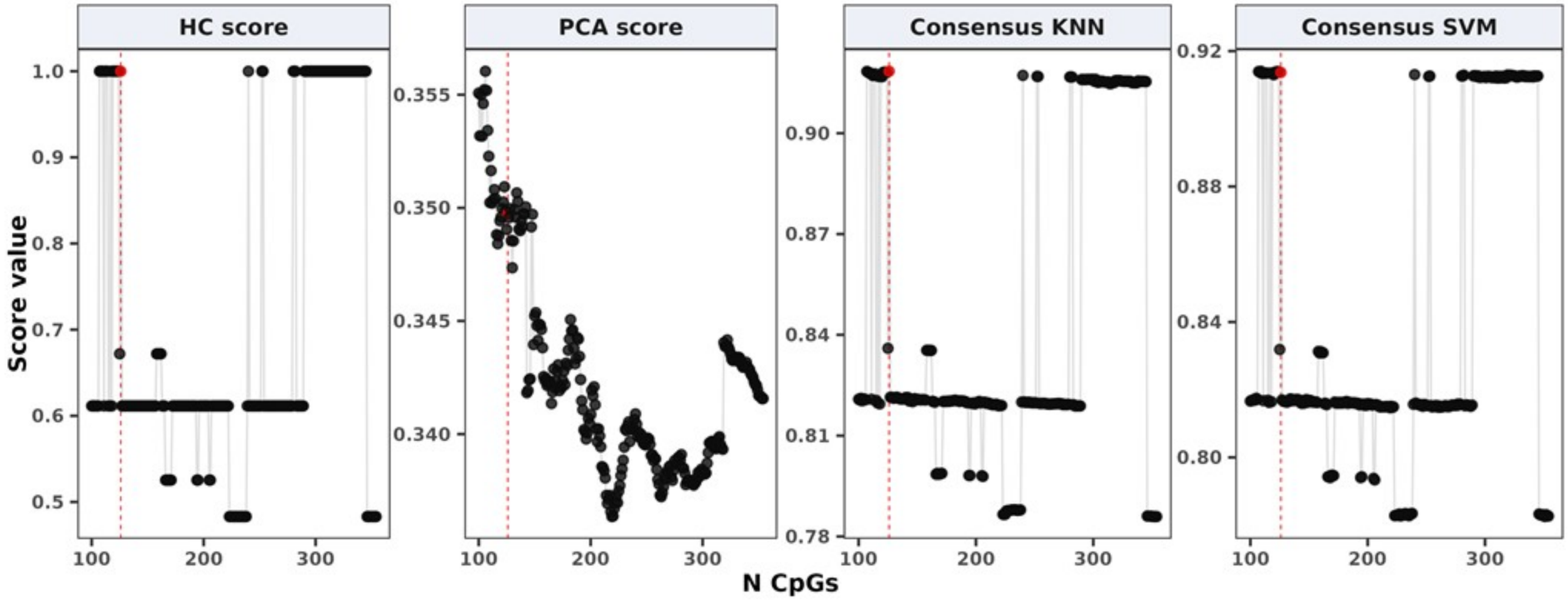
(A) Optimization of Random Forest-ranked CpG subsets using a combined criterion integrating principal component analysis separation, hierarchical clustering agreement, and classification performance using k-nearest-neighbour and support vector machine models. The red point indicates the selected subset of 126 CpGs. (B) Heatmap of centered DNA methylation M-values across the 126 selected CpGs. Hierarchical clustering was performed using Euclidean distances and Ward's minimum-variance method (Ward.D2). The heatmap was generated using the ComplexHeatmap R package. Sample annotations indicate discovery controls (Control), discovery cases (PMM2),

validation cases (PMM2_V), and individuals with alternative PMM2-associated phenotypes (PMM2_HIPKD and PMM2_DP). Severity categories were based on Nijmegen Pediatric CDG Rating Scale (NPCRS) scores: mild, 0–14; moderate, 15–25; severe, ≥26. NR, not reported. (C) Principal component analysis based on the 126 selected CpGs, illustrating the distribution of discovery cases, validation cases, controls, and individuals with alternative PMM2-associated phenotypes.

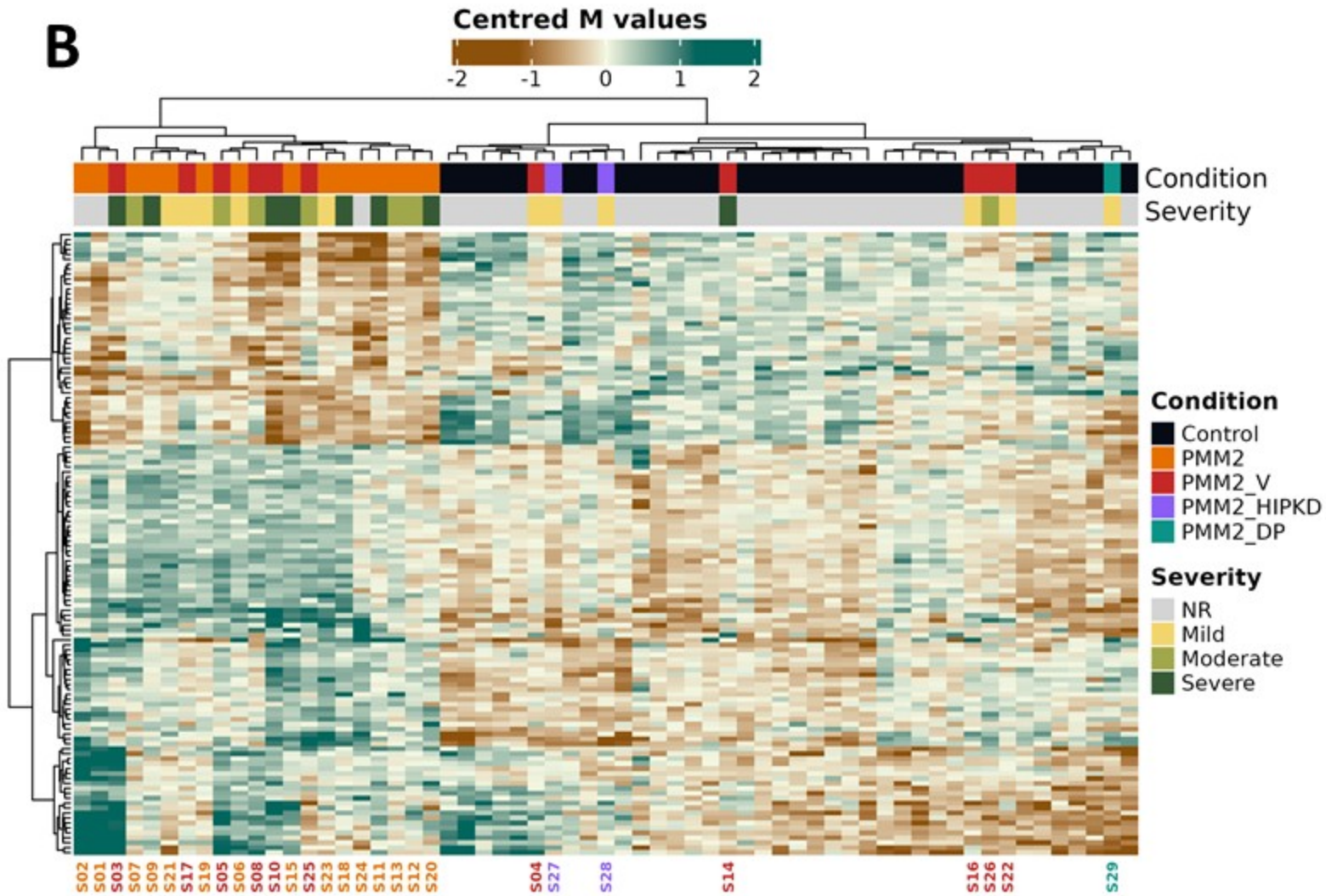
Figure 3: Classification performance of the candidate PMM2-CDG episignature.
(A) Sample-level PMM2-CDG classification scores for discovery controls (Control) and cases (PMM2), external EPICv1 controls (Control_Out), validation cases (PMM2_V), and individuals with alternative PMM2-associated phenotypes (PMM2_HIPKD and PMM2_DP). (B) Receiver operating characteristic (ROC) curves illustrating the discriminative performance of both classifiers across the evaluated samples. Areas under the ROC curves (AUCs) and 95% confidence intervals are indicated. Blue curves include all evaluated samples, including discovery cases and controls, and all validation samples. Red curves exclude discovery samples. Validation cases with classical PMM2-CDG as reference-positive samples and individuals with PMM2-HIPKD or PMM2-DP, together with external controls, as reference-negative samples.


A
Score
1.00
0.75
0.50
0.25
0.00
S19
S17
S26
S04
S16
S22
S29
S29
Control
PMM2
Control_Out
PMM2_V
PMM2_HIPKD
PMM2_DP
B
All
Validation
Sensitivity (True Positive Rate)
1 - Specificity (False Positive Rate)
AUC = 0.961
95% CI: 0.917–1.000
AUC = 0.901
95% CI: 0.792–1.000

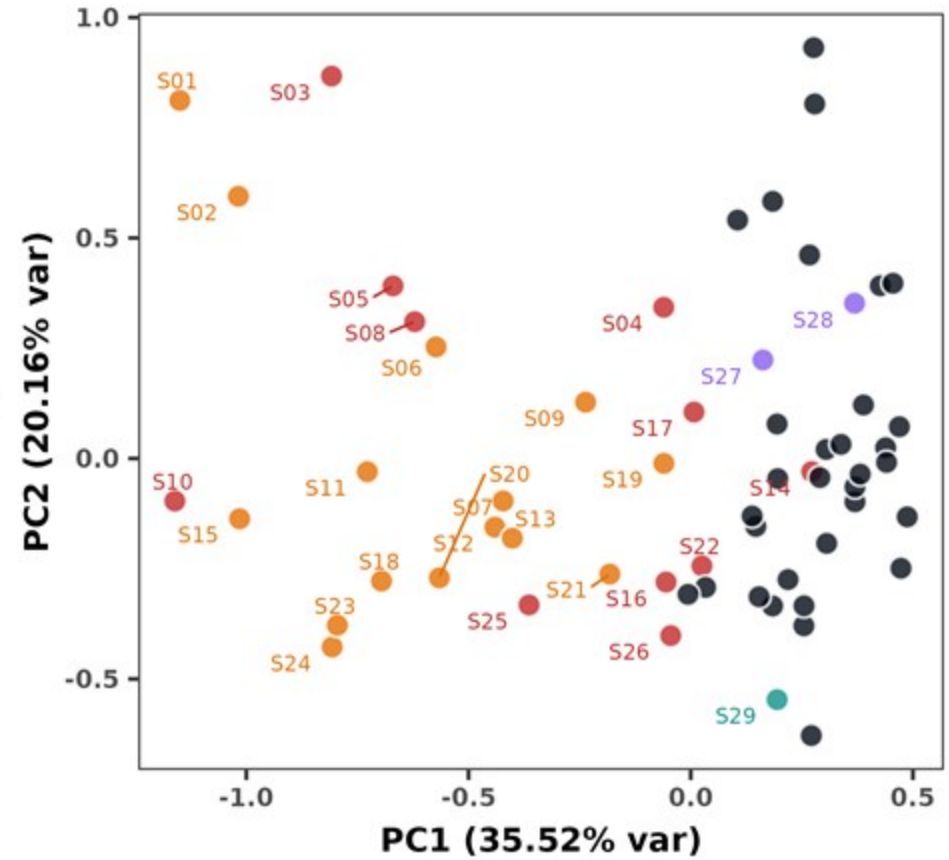
A


HC score
PCA score
Consensus KNN
Consensus SVM
Score value
N CpGs

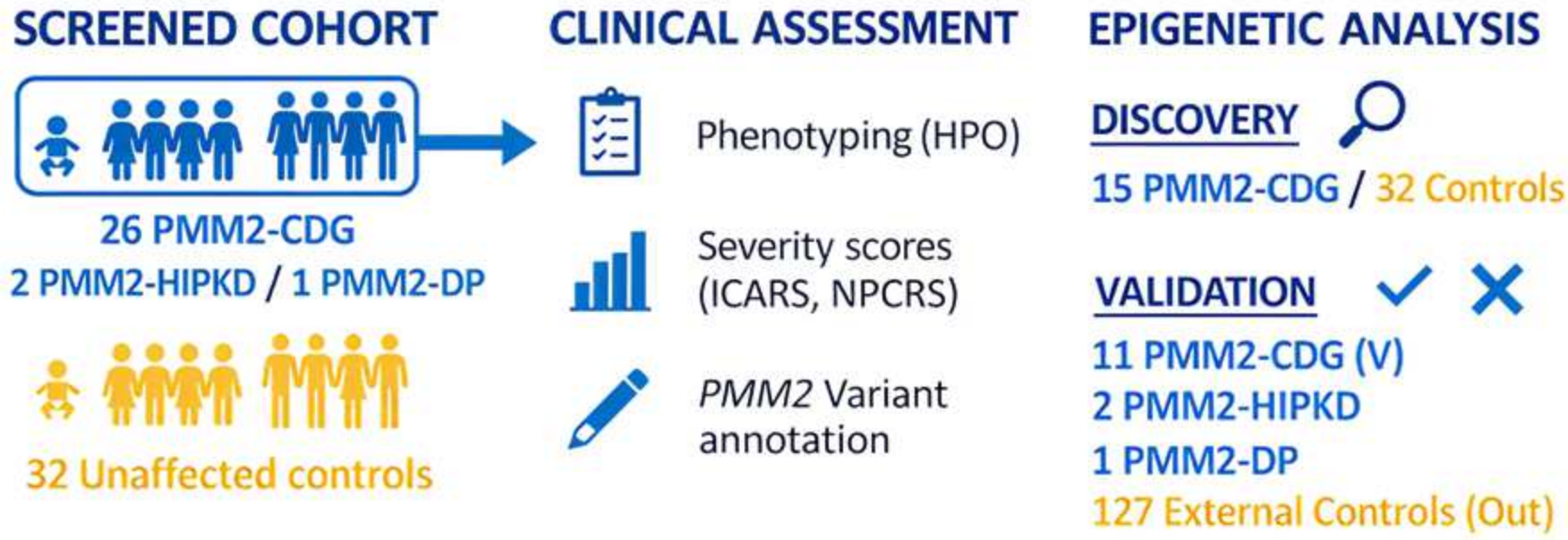
B


Centred M values
Condition
Severity
Control
PMM2
PMM2_V
PMM2_HIPKD
PMM2_DP
NR
Mild
Moderate
Severe

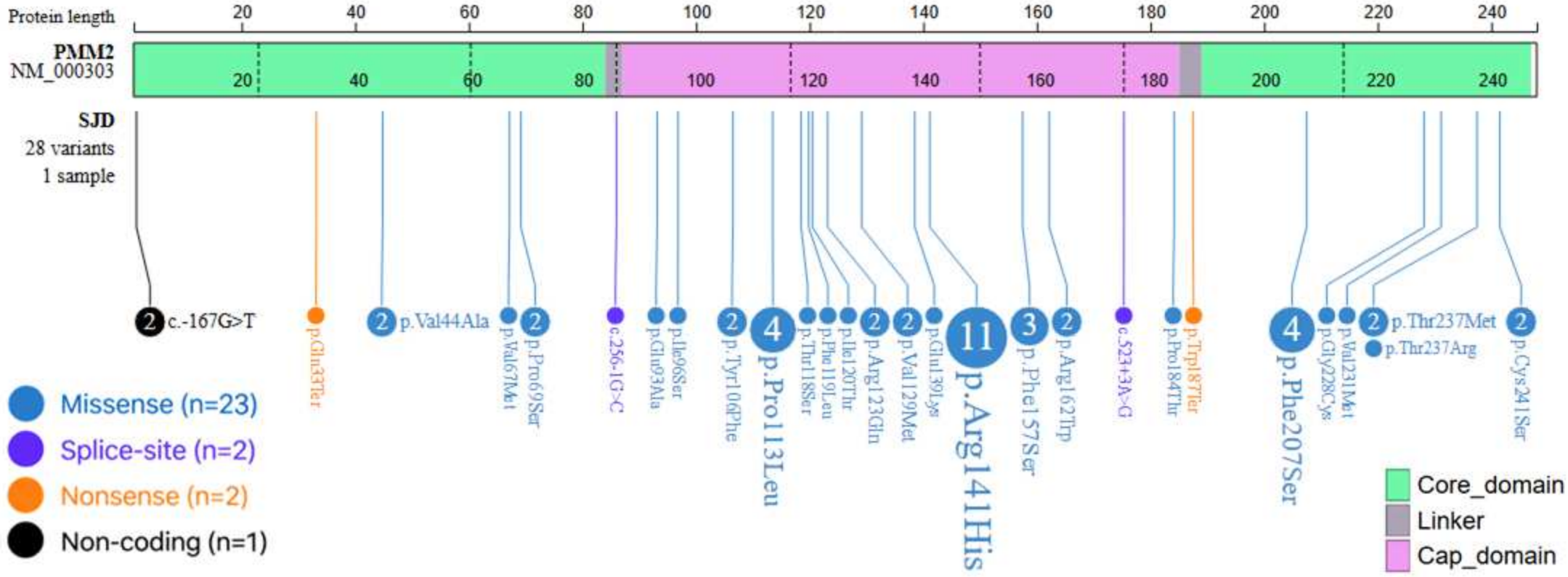
C


PC2 (20.16% var)
PC1 (35.52% var)

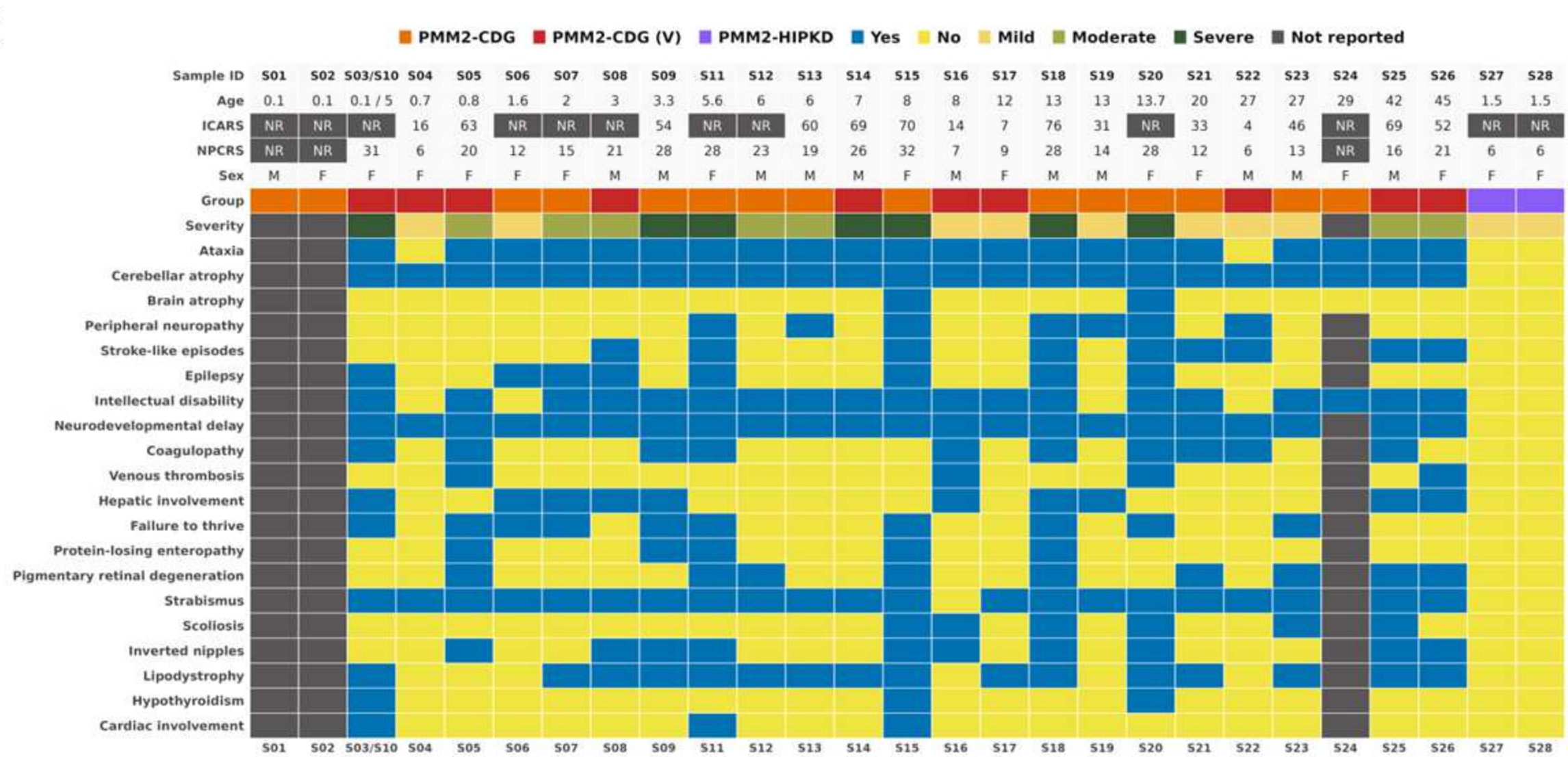
A


SCREENED COHORT
26 PMM2-CDG
2 PMM2-HIPKD / 1 PMM2-DP
32 Unaffected controls
CLINICAL ASSESSMENT
Phenotyping (HPO)
Severity scores (ICARS, NPCRS)
PMM2 Variant annotation
EPIGENETIC ANALYSIS
DISCOVERY
15 PMM2-CDG / 32 Controls
VALIDATION
11 PMM2-CDG (V)
2 PMM2-HIPKD
1 PMM2-DP
127 External Controls (Out)

B


Protein length
PMM2
NM_000303
SJD
28 variants
1 sample
c.-167G>T
p.Gln33Ter
p.Val44Ala
p.Val67Met
p.Pro69Ser
c.256-1G>C
p.Glu93Ala
p.Ile96Ser
p.Tyr106Phe
p.Pro113Leu
p.Thr118Ser
p.Phe119Leu
p.Ile120Thr
p.Arg123Gln
p.Val129Met
p.Glu139Lys
p.Arg141His
p.Phe157Ser
p.Arg162Trp
c.523+3A>G
p.Trp187Ter
p.Pro184Thr
p.Phe207Ser
p.Gly228Cys
p.Val231Met
p.Thr237Met
p.Thr237Arg
p.Cys241Ser
Missense (n=23)
Splice-site (n=2)
Nonsense (n=2)
Non-coding (n=1)
Core_domain
Linker
Cap_domain

C


PMM2-CDG
PMM2-CDG (V)
PMM2-HIPKD
Yes
No
Mild
Moderate
Severe
Not reported
Sample ID S01 S02 S03/S10 S04 S05 S06 S07 S08 S09 S11 S12 S13 S14 S15 S16 S17 S18 S19 S20 S21 S22 S23 S24 S25 S26 S27 S28
Age 0.1 0.1 0.1 / 5 0.7 0.8 1.6 2 3 3.3 5.6 6 6 7 8 8 12 13 13 13.7 20 27 27 29 42 45 1.5 1.5
ICARS NR NR NR 16 63 NR NR NR 54 NR NR 60 69 70 14 7 76 31 NR 33 4 46 NR 69 52 NR NR
NPCRS NR NR 31 6 20 12 15 21 28 28 23 19 26 32 7 9 28 14 28 12 6 13 NR 16 21 6 6
Sex M F F F F F F M M F M M M F M F M M F F M M F M F F F
Group
Severity
Ataxia
Cerebellar atrophy
Brain atrophy
Peripheral neuropathy
Stroke-like episodes
Epilepsy
Intellectual disability
Neurodevelopmental delay
Coagulopathy
Venous thrombosis
Hepatic involvement
Failure to thrive
Protein-losing enteropathy
Pigmentary retinal degeneration
Strabismus
Scoliosis
Inverted nipples
Lipodystrophy
Hypothyroidism
Cardiac involvement
S01 S02 S03/S10 S04 S05 S06 S07 S08 S09 S11 S12 S13 S14 S15 S16 S17 S18 S19 S20 S21 S22 S23 S24 S25 S26 S27 S28

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplFig1EpimutationBurdenPMM2BETA.jpg
- SuppFig2HIPKD.jpg
- SupplFig3EpimutationHIPKD.jpg
- SupplementaryFiguresLegends.docx
- SupplEpisignPMM2tables.xlsx