



Juszczak, A., Pavić, T., Vučković, F., Bennett, A. J., Shah, N., Medvidović, E. P., Groves, C. J., Šekerija, M., Chandler, K., Burrows, C., Putarek, N. R., Lovrenčić, M. V., Knežević, J. Č., James, T. J., Gloyn, A. L., Lauc, G., McCarthy, M. I., Owen, K. R., & Gornik, O. (2019). Plasma fucosylated glycans and C-reactive protein as biomarkers of HNF1A-MODY in young adult-onset nonautoimmune diabetes. *Diabetes Care*, 42(1), 17-26. <https://doi.org/10.2337/dc18-0422>

Peer reviewed version

Link to published version (if available):
[10.2337/dc18-0422](https://doi.org/10.2337/dc18-0422)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via ADA at <http://care.diabetesjournals.org/content/early/2018/11/08/dc18-0422> . Please refer to any applicable terms of use of the publisher.

Full-text PDF (final published version)

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Plasma fucosylated glycans and C-reactive protein as biomarkers of HNF1A-MODY in young adult onset non-autoimmune diabetes

Short title: Fucosylated glycans and CRP as HNF1A-MODY biomarkers

*Agata Juszczyk^{1,2} and *Tamara Pavić³, Frano Vučković⁴, Amanda J. Bennett¹, Neha Shah⁵, Edita Pape Medvidović⁶, Christopher J. Groves^{1,2}, Mario Šekerija^{7,8}, Kyla Chandler⁹, Carla Burrows¹, Nataša Rojnić Putarek¹⁰, Marijana Vučić Lovrenčić¹¹, Jadranka Čuća Knežević¹¹, Tim J. James⁵, Anna L. Gloyn^{1,2,12}, Gordan Lauc^{3,4}, Mark I. McCarthy^{1,2,12}, *Katharine R. Owen^{1,2} and *Olga Gornik³

*equal authorship

¹Oxford Centre for Diabetes, Endocrinology and Metabolism, Churchill Hospital, University of Oxford, Oxford, UK

²NIHR Oxford Biomedical Research Centre, Oxford

³Faculty of Pharmacy and Biochemistry, University of Zagreb, Zagreb, Croatia

⁴Genos Ltd, Glycobiology Division, Zagreb, Croatia

⁵Department of Biochemistry, Oxford University Hospitals NHS Trust, Oxford, UK

⁶Merkur University Hospital, Vuk Vrhovac University Clinic for Diabetes, Endocrinology and Metabolic Diseases, Zagreb University School of Medicine, Zagreb, Croatia

⁷Croatian Institute of Public Health, Zagreb, Croatia

⁸School of Medicine, Andrija Štampar School of Public Health, University of Zagreb, Zagreb, Croatia

⁹Diabetes and Metabolism, School of Clinical Sciences, University of Bristol

¹⁰University Hospital Center Zagreb, Department of Pediatric Endocrinology and Diabetes, Zagreb University School of Medicine, Zagreb, Croatia

¹¹Department of Clinical Chemistry and Laboratory Medicine, Merkur University Hospital, Zagreb, Croatia

¹²Wellcome Trust Centre for Human Genetics, Oxford, UK

Corresponding authors:

Olga Gornik, University of Zagreb, Faculty of Pharmacy and Biochemistry, Department of Biochemistry and Molecular Biology, A. Kovačića 1, 10000 Zagreb, Croatia, tel.: +38514856201, fax +3856394400, mail: ogornik@pharma.hr

Katharine Owen, University of Oxford, OCDEM, Churchill Hospital, Oxford, UK, OX3 7LJ, tel.: +441865857226, [mail: katharine.owen@drl.ox.ac.uk](mailto:katharine.owen@drl.ox.ac.uk)

Word count: 4092

Number of Tables/Figures: 4

Abstract

Objective

Maturity-onset diabetes of the young due to variants in *HNF1A* is the commonest type of monogenic diabetes. Frequent misdiagnosis results in missed opportunity to use sulphonylureas as first-line treatment. A non-genetic biomarker could improve selection of subjects for genetic testing and increase diagnosis rates. We have previously reported that plasma levels of antennary fucosylated *N*-glycans and hsCRP are reduced in individuals with *HNF1A*-MODY. In this study, we examined the potential use of *N*-glycans and hsCRP in discriminating individuals with damaging *HNF1A* alleles from those without *HNF1A* variants, in an unselected population of young adults with non-autoimmune diabetes.

Research Design and Methods

We analysed plasma *N*-glycan profile, measured hsCRP and sequenced *HNF1A* in 989 individuals with diabetes diagnosed below age 45, persistent endogenous insulin production and absence of pancreatic autoimmunity. Systemic assessment of rare *HNF1A* variants was performed.

Results

We identified 29 individuals harbouring 25 rare *HNF1A* alleles, of which 3 were novel and 12 (in 16 probands) considered pathogenic. Antennary fucosylated *N*-glycans and hsCRP were able to differentiate subjects with damaging *HNF1A* alleles from those without rare *HNF1A* alleles. Glycan GP30 had a ROC curve AUC of 0.90 (88% sensitivity, 80% specificity, cut-off 0.70%), while hsCRP had an AUC of 0.83 (88% sensitivity, 69% specificity, cut-off 0.81 mg/L).

Conclusions

Half of rare *HNF1A* sequence variants do not cause MODY. Both *N*-glycan profile and hsCRP could be used as tools, either alone, or as adjuncts to existing pathways, for identifying individuals at high risk of carrying a damaging *HNF1A* allele.

Introduction

Whilst there are a number of genes implicated in monogenic diabetes, maturity onset diabetes of the young due to variants in *HNF1A* (HNF1A-MODY) is the most frequent form in adults [1] and has a significant impact on management when the diagnosis is made. Common clinical criteria for selecting individuals for genetic testing for MODY include: diabetes onset below 25 years of age, preserved endogenous insulin production, absence of pancreatic autoimmunity and consecutive generations of diabetes [2, 3].

These criteria clearly overlap with features of both type 1 and type 2 diabetes, so many individuals with HNF1A-MODY remain unrecognized, particularly if they do not fit classic MODY criteria [1]. Herein, we defined HNF1A-MODY as non-autoimmune young adult-onset diabetes in individuals carrying deleterious *HNF1A* alleles.

Frequently, physicians do not prioritise diagnosis of MODY, and also, many countries have limited access to genetic testing. Establishing a correct molecular diagnosis of HNF1A-MODY allows treatment change to sulphonylureas or glinides, which may provide excellent diabetes control for decades [4]. The correct diagnosis also facilitates prompt identification of affected family members.

Widening access to MODY genetic testing is assisted by tools such as the MODY probability calculator [5], however these models rely largely on clinical criteria. Criteria such as absence of β -cell autoantibodies and presence of C-peptide exclude most cases of autoimmune diabetes. Work in this area has focused mainly on the discrimination of MODY presenting in childhood, by selecting β -cell antibody negative children for further investigation [6]. A recent study took a similar approach with young

adults diagnosed before 30 years [7], although most had a clinical label of type 1 diabetes.

The differentiation from type 2 diabetes, particularly in an older age group, where the proportion of MODY is lower, is more challenging. Adding specific HNF1A-MODY biomarkers, which rely on extrapancreatic manifestations of HNF1A, could assist in the differentiation from non-autoimmune diabetes.

HNF1A encodes a transcription factor which regulates the expression of many genes [8]. C-reactive protein (CRP) expression in the liver is regulated by HNF1A [9] and genome-wide association studies (GWAS) showed that plasma CRP level was associated with genetic variation near *HNF1A* [10, 11]. We subsequently reported that hsCRP levels were lower in subjects with typical HNF1A-MODY than in other forms of diabetes, with the best discrimination from young adult-onset type 2 diabetes [12].

Similarly, a GWAS of the plasma *N*-glycome identified HNF1A as a key regulator of plasma protein fucosylation [13]. *N*-glycosylation is a frequent posttranslational modification, characterized by enzymatic attachment of complex sugar moieties (glycans) to the protein. It is essential for proper protein function [14], and has shown to have an important role in many (patho)physiological processes [15, 16]. We also reported that disease-causing *HNF1A* alleles are associated with marked alterations of plasma *N*-glycans bearing antennary fucose [17].

Thus, both plasma *N*-glycans and hsCRP are promising candidates for HNF1A-MODY diagnostic markers. Our previous studies focused on individuals with an established clinical diagnosis based on clinical, biochemical and molecular investigations. These

studies were likely to be subject to spectrum bias, leading to an overestimation of the discriminative properties.

In this study, we aimed to assess the value of *N*-glycans and hsCRP as HNF1A-MODY biomarkers in a relatively unselected population with a young adult-onset non-autoimmune diabetes and evaluate their translational potential. As part of this process it was necessary to evaluate whether the identified rare *HNF1A* alleles were likely to be the main cause of diabetes in the individuals recruited.

Research design and methods

Study participants

Subjects were recruited in UK and Croatia. UK participants (n=523) were recruited via the Young Diabetes in Oxford (YDX) study, which included seven hospital diabetes centres and multiple primary care centres from the Thames Valley. Croatian subjects (n=466) were recruited through the Croatian national diabetes registry (CroDiab) and sampled at Vuk Vrhovac University Clinic in Zagreb.

Inclusion criteria were: current age ≥ 18 years, diabetes diagnosis < 45 years, preserved endogenous insulin production (fasting C-peptide ≥ 0.2 nmol/L) and negative Glutamic Acid Decarboxylase (GAD) antibodies. From 989 subjects included in the study, 84 had diabetes onset before 25 years of age. Four UK subjects had a previously known diagnosis of HNF1A-MODY as a result of previous investigation in the YDX study. All participants signed an informed consent. Table 1 shows clinical characteristics of the recruited individuals and their treatment at the time of recruitment.

DNA sequencing and an assessment of *HNF1A* alleles

DNA was extracted, amplified and sequenced using the Sanger method [18]. Mutation Surveyor version 5.0.1 (Soft Genetics, UK) was used for detection of variants compared to the reference sequence (NM_000545.5). A systematic assessment of rare *HNF1A* alleles (minor allele frequency, MAF <1%) was performed and aligned to the American College of Medical Genetics (ACMG) classification [19]. This included clinical features, co-segregation of the allele with diabetes in the family and *in silico* analysis of missense variants using SIFT, Polyphen2 and Provean. Potential effect on splicing was examined using Human Splicing Finder (HSF). The presence of rare *HNF1A* alleles in the publicly available database of 123,136 exomes and 15,496 whole-genomes in Genome Aggregation Database (GnomAD, Broad Institute, USA) [20] was recorded. The results of laboratory assessment of function, available in the literature, were reviewed and functional studies of five previously uncharacterised *HNF1A* alleles were performed.

Functional assessment of *HNF1A* alleles

cDNA of human *HNF1A* (NM_000545.5) was inserted into the pcDNA 3.1/His C plasmid (gift from KG Jebsen Center for Diabetes Research, University of Bergen) and used in site-directed mutagenesis as a template. HeLa cells were cultured for all functional assessments and transfected with mutagenized plasmids. Each *HNF1A* variant underwent an assessment of transcription activity using dual luciferase reporter system, an analysis of protein expression using Western blotting and an assessment of DNA binding employing electrophoretic mobility shift assay (EMSA). Each experiment included empty plasmid, wild type (WT) *HNF1A*, 2-3 positive controls (p.P112L, p.T260M and p.R203H), 1 synonymous variant (p.H179H) and 1 variant associated with

an increased risk of type 2 diabetes (p.E508K). The choice of control variants was based on an established evidence for causing MODY, such as co-segregation of the allele with the young adult-onset diabetes in multiple families and supporting functional data. Each experiment was repeated twice on further passages of HeLa cells to obtain three biological replicates.

N-glycan analysis

N-glycan release, labelling and clean-up was performed as described previously [21]. Fluorescently labelled glycans were separated by hydrophilic interaction liquid chromatography on a Waters Acquity UPLC instrument (Milford, USA) as previously described [21]. All chromatograms were separated into 42 chromatographic peaks, which enabled reliable quantification. The amount of glycans in each peak was expressed as percentage of total integrated area. The corresponding glycan structures were assigned according to Saldova *et al* [22].

Biochemical and immunological assays

Most of the UK samples (n=495) had CRP measured using a wide range latex-enhanced immunoturbidimetric high-sensitivity assay on ADVIA 2400 analyser (Siemens Healthcare Diagnostics, Germany) with the limit of quantification of 0.01 mg/L. All Croatian samples and remaining UK samples (n=494) had CRP measured with the Abbott hs-CRP method and the lowest quantifiable level of 0.1mg/L. Methods were reproducible, both having the coefficient of variation below 10.5% across the concentration range tested. Comparison of the clinical samples (n=51) measured by

both methods showed agreement to be: Abbott method = $0.26 + 0.99$ [Siemens method], by Passing and Bablock regression.

In UK, GADA were measured by radioimmunoassay using ^{35}S labelled GAD65. The cut-off for the positive result was 13 WHO Units/mL initially using a local assay (samples measured $n=218$, DASP2010 sensitivity 88% at 93% specificity) and changed to 33 DK Units/mL later in the study (standard assay, DASP2010 sensitivity 80%, specificity 97%). In Croatia, GADA were measured by ELISA immunoassay. The cut-off for the positive result was 5 WHO Units/mL (DASP2010 sensitivity 88% at 94% specificity).

Statistical analysis

Subject characteristics and results of the functional work were analysed using SPSS v.23. Continuous data were presented as medians (IQR) and Kruskal-Wallis test was applied to compare groups. P-value <0.05 was considered statistically significant. Differences of frequencies for categorical variables were tested using the Chi-squared test. The results of the functional studies were presented as a percentage of the WT HNF1A. Differences between the studied variants and WT were analysed using analysis of variance with correction for multiple tests.

The Exeter MODY probability calculator [5] was used to compare performance of the biomarkers assessed in this study.

Glycan and hsCRP data were analysed and visualized using R v.3.0.1. Both *N*-glycans and hsCRP had non-parametric distributions. Association analyses between disease status and glycan traits were performed using a general linear model, with age and sex included as additional covariates. False discovery rate was accounted for using

Benjamini-Hochberg procedure. For prediction of disease status, both logistic regression and regularized logistic regression models were applied. Logistic model was applied in bivariate regression classification analyses (one glycan trait used per model). In multiple regression classification analyses (multiple glycan traits used as predictors in model) regularized logistic models were applied. To evaluate performance of regularized logistic model 10-cross validation procedure was used. Predictions from each validation procedure were merged into one validation set, on which model performance was evaluated, based on the receiver operating characteristic (ROC) curve criteria.

Results

Assessment of *HNF1A* alleles

HNF1A sequencing of 989 study participants resulted in identifying 25 rare (MAF<1%) non-synonymous *HNF1A* variants in 29 probands, including 7 protein truncating (PTVs) and 18 missense variants. The identified variants are listed in Table 2. Additional features of all probands are listed in Supplementary Table 1.

The likely phenotypical effect of the *HNF1A* allele (damaging, VUS or benign) was assigned taking into consideration previous reports of the allele causing the MODY phenotype, co-segregation of the allele with diabetes, prediction of bioinformatics, absence of the allele in GnomAD and results of functional studies.

Published data showed that 14 alleles were previously reported as causing the MODY phenotype [24-33], 8 were reported as variants of unknown significance (VUS) or benign [24, 34] and 3 were novel (p.S3C, p.G151S, p.K222N).

PTVs were all located in exons 1-6, affecting all isoforms of the protein and likely to undergo nonsense-mediated decay leading to haploinsufficiency. All PTVs were previously reported to cause the MODY phenotype with evidence of co-segregation of the allele with diabetes (Table 2). We therefore considered all PTVs as deleterious.

Of the missense variants, 7/18 were predicted to be damaging protein function by at least two of the three bioinformatic tools used, while 11/18 were predicted to be benign. HSF predicted that the promoter variant c.-4A>G does not affect splicing. HSF also predicted that the novel coding missense variant c.8C>G, p.S3C, is likely to affect splicing by gaining a new donor site. Also, bioinformatics predicted it as damaging.

Thirteen of 25 alleles were present in individuals from GnomAD with MAF of 0.0008-0.08%, while 12 alleles were not reported in GnomAD database.

We sequenced *HNF1A* in 22 available family members from 8 families. Five rare *HNF1A* alleles were present in 8 individuals. Co-segregation of the *HNF1A* allele and diabetes was reported in this study and/or published literature for 14/25 alleles (Table 2).

Functional assessment of previously uncharacterised HNF1A protein variants

Five HNF1A protein variants underwent functional assessment: two were novel (p.G151S, p.K222N) and three were not previously studied (p.G288W, p.P291T, p.H349Q). During this project, assessment of p.H349Q was published [38] and findings were consistent with ours.

The protein expression of p.G151S, p.G288W, p.P291T and p.H349Q was similar to the WT HNF1A (107-128% of WT, non-significant p-values) and increased for p.K222N (p=0.004, Figure 1B).

The normalised transcription activity (TA) was significantly reduced by p.P291T (52% of WT, p=0.001) and p.G288W (73% of WT, p=0.04), while there was no significant difference in TA for p.G151S and p.K222N (82.8% and 76.1% of the WT respectively).

The transactivation of p.H349Q was similar to the WT (111% WT, Figure 1A).

Finally, we performed EMSA to assess DNA binding of the variants, of which p.G151S and p.K222N are located in the DNA binding domain. The variant p.G151S had significantly reduced DNA binding (19% of the WT, p=2×10⁻⁶, Figure 1C), which remained the same after normalisation to the protein amount (Figure 1D). Normalised DNA binding of the remaining four variants was not different from WT (71-101% of WT).

In summary, the functional assessment of five previously uncharacterised HNF1A protein variants provided support for the novel variant p.G151S to be considered as functionally deleterious. The borderline reduction in TA of variant p.P291T and p.G288W makes the functional results inconclusive.

Summary of the assessment of *HNF1A* allelic variants

Based on the systematic assessment described above and summarised in Table 2, we considered that 12 rare *HNF1A* alleles (present in 16 probands and 3 relatives) are likely to be damaging HNF1A protein function (ACMG classification 1-2)[19], 9 are likely to be benign (ACMG 4-5) and 4 were labelled as VUS (ACMG 3), as there were features both for and against a damaging effect. This corroborates the observation that

the phenotypical effect of the *HNF1A* alleles represent a spectrum without clear borders and shows the complexity of interpretation of the genetic variation. While acknowledging this complexity, we simplified the phenotypical spectrum to likely damaging, VUS and likely benign alleles to enable assessment of the biomarkers.

Plasma *N*-glycans and hsCRP in *HNF1A*-MODY

Two individuals with benign alleles were excluded due to missing data, leaving 27 probands for *N*-glycan data analysis. We also excluded 114 subjects (including 2 with benign alleles) with CRP >10mg/L from hsCRP data analysis, since hsCRP may have been elevated as a result of concomitant inflammation.

Firstly, we compared probands with likely damaging *HNF1A* alleles (n=16) with individuals without rare *HNF1A* alleles (n=960 for glycan and n=844 for hsCRP analysis). We found that 8 of 42 glycan traits were significantly altered in subjects with likely damaging *HNF1A* alleles compared to subjects without *HNF1A* variants (adjusted $p=1.00\times 10^{-5} - 1.46\times 10^{-2}$) (Supplemental Table 2). Glycan groups GP30, GP36 and GP38 showed the largest differences (Supplemental Figure 1), each of them containing antennary fucosylated glycan (Figure 2A).

Similarly, hsCRP was lower in subjects with likely damaging *HNF1A* alleles than in those without *HNF1A* variants [0.21 (0.07-0.68) vs. 1.70 (0.60-3.91) mg/L; $p=3.09 \times 10^{-5}$] (Supplemental Figure 1).

Secondly, we examined whether GP30 and hsCRP could serve as markers of *HNF1A* allele function. Both biomarkers were significantly lower in probands with likely damaging *HNF1A* alleles (n=16) than in subjects with likely benign *HNF1A* alleles (n=7); median GP30 0.43 (0.34-0.57) vs. 0.95 (0.51-1.79)%, $p=0.012$ and median hsCRP 0.21

(0.07-0.68) vs 1.01 (0.84-2.36) mg/L, $p=0.006$ (Figure 2B-E). Median GP30 and hsCRP in subjects with VUS did not significantly differ from individuals with likely benign or without rare *HNF1A* alleles.

Discriminating *HNF1A*-MODY from young-adult onset non-autoimmune diabetes using plasma *N*-glycans and hsCRP

Examination of the classification performance of plasma *N*-glycans and hsCRP was performed by ROC curve analysis, comparing biomarker values in subjects with likely damaging *HNF1A* alleles against subjects without rare *HNF1A* alleles (Figure 2B-E).

Firstly, the discriminative performance of individual glycan groups was tested, where GP30, GP36 and GP38 showed the best discriminative power among all individual glycans, with AUC of 0.90, 0.87 and 0.90, respectively (Figure 2B-D). Secondly, a model based on the total plasma *N*-glycome (all 42 glycan groups included) was built. It showed a similar discriminative power between early-onset non-autoimmune diabetes and subjects with damaging *HNF1A* alleles, when compared to the GP30, with AUC for total glycome model 0.92 (0.86–0.99) vs 0.90 (0.83–0.97) for GP30. HsCRP also showed a satisfactory performance in distinguishing two groups with AUC of 0.83 (0.71-0.94), Figure 2E. Finally, the joint performance of both GP30 and hsCRP was calculated and resulted in AUC of 0.90 (0.83–0.98), which was again similar to GP30 alone.

Clinical potential of GP30 and hsCRP

The clinical potential of the best performing glycan, GP30, and hsCRP was further evaluated. ROC curve analysis indicated that a diagnostic threshold of 0.70% for GP30 provided optimal discrimination of subjects with likely damaging *HNF1A* alleles from

subjects with early-onset non-autoimmune diabetes and without *HNF1A* variants, showing sensitivity of 88% and specificity of 80%. A GP30 cut-off of 0.70% missed only 2 of 16 probands with likely damaging *HNF1A* alleles. If GP30 was used as a selective tool for stratification of the current cohort, 214 individuals with young-onset non-autoimmune diabetes would have *HNF1A* sequenced (22%).

ROC curve analysis for hsCRP showed that a threshold of 0.81 mg/L provided optimal discrimination between the groups, with 88% sensitivity and 69% specificity. Using the proposed cut-off, 2 of 16 subjects with likely damaging *HNF1A* alleles were also missed, while employing it as a screening tool, it would have resulted in *HNF1A* sequencing of 269 individuals from this study (27%).

In contrast, if we used classical clinical criteria for MODY genetic testing (diagnosis of diabetes <25 years of age, at least 2 generations FH of diabetes, endogenous insulin production and negative GADA), we would have picked-up only 8 of 16 individuals (50%) with likely damaging *HNF1A* alleles, while sequencing 99 individuals from this study (10%).

The Exeter MODY probability calculator gives a pre-test probability of any form of MODY, but is not validated in subjects diagnosed with diabetes at >35 years [5], or of non-white ethnicity. In this study 370 out of the 989 participants could be assessed using the MODY calculator and 136 (37% of those assessed) had an estimated probability of MODY greater than 20%. Fourteen out of sixteen of the probands with damaging MODY variants could be assessed using the MODY calculator and 9 of these had an estimated probability >20%, thus a sensitivity of 56% for detecting *HNF1A*-MODY (Table 2). This is similar to the classic criteria and also to performance in the

UNITED study, where 55% of cases were missed by the calculator (<25% risk) [39]. The calculator would also lead to selection of a higher proportion of the cases (37% compared to 22-27%).

We also examined GP30 and hsCRP levels in subjects with VUS and novel *HNF1A* alleles to estimate if biomarkers assisted in assigning the functional effect of the allele. Among the alleles labelled as VUS, all three individuals (proband and 2 relatives) with variant p.P291T, participant with variant p.T515M and participant with novel variant p.S3C had GP30 and hsCRP above the proposed cut-offs, providing support for their benign effect. In contrast, all three subjects with variant p.A251T (proband and 2 relatives) had hsCRP below 0.30 mg/L, however, two had GP30 above and one had GP30 slightly below the cut-off value, making the results inconclusive. Regarding the remaining two novel *HNF1A* alleles, using GP30 and hsCRP classified p.K222N as damaging (opposite to the functional work results), while p.G151S had discordant biomarker results.

Conclusions

In this study, we found 25 rare *HNF1A* alleles in 29 individuals (2.9 % of the participants). Following the systematic assessment of these alleles, we considered that 12 *HNF1A* alleles (in 16 probands) are likely to be damaging HNF1A protein function, 9 are likely to be benign and 4 remain as variants of unknown significance.

The participants in this study were all found to have HNF1A-MODY as a result of participating in clinical research, demonstrating that many cases are missed in real-life clinical practice. The consequence of the diagnosis was that 10 individuals from the 16

probands and 2 relatives were able to commence SU treatment and 4 discontinued insulin treatment.

This study showed that antennary fucosylated plasma glycans (GP30, GP36 and GP38) and hsCRP levels were significantly lower in subjects harbouring likely damaging *HNF1A* alleles compared to individuals without rare *HNF1A* alleles. This finding is consistent with the role of HNF1A, which acts as transcription factor for *CRP* (9) and genes encoding fucosyltransferases (13) in hepatocytes. Results of this study also confirmed our previous findings where we examined the plasma *N*-glycans and hsCRP in groups of individuals already known to have MODY [12,17]. However, herein we examined the performance of the biomarkers in a relatively unselected population of subjects with young adult-onset non-autoimmune diabetes, which better reflects the situation encountered by clinicians while assigning a diagnosis.

Both biomarkers (GP30 and hsCRP) were equally successfully in recognizing subjects with likely deleterious *HNF1A* alleles (88% sensitivity), however, GP30 showed better specificity than hsCRP (80% vs. 69%). Moreover, compared to both classic clinical criteria and the MODY probability calculator, both biomarkers were superior in selecting subjects to be referred for genetic testing.

Thus, incorporating biomarkers into clinical use of such prediction models may assist the successful stratification of individuals with young adult-onset diabetes carrying potentially deleterious *HNF1A* alleles.

In many countries panel testing of many genes for monogenic diabetes is the first line genetic test. Biomarkers may still be useful as an estimate of pre-test risk of HNF1A-MODY in the context of interpreting panel results. In other counties, including Croatia

currently, very little genetic testing is routinely available so the biomarkers may continue to have a role in identifying those at greatest risk of HNF1A-MODY for single gene Sanger sequencing approaches.

Large scale sequencing studies in both healthy and disease populations have shown that many variants initially thought to be disease-causing are present in population samples in frequencies greater than it would be expected for a rare monogenic condition [40]. GP30 and hsCRP could provide an additional value in assigning disease causality of identified *HNF1A* alleles, as individuals with likely damaging *HNF1A* alleles had significantly lower levels of antennary fucosylated glycans and hsCRP than subjects with benign *HNF1A* alleles. Both biomarkers were consistent in assigning a direction of the functional effect in 8 individuals harboring VUS in this study (4 probands and 4 relatives), placing the variant p.A251T as a damaging, while p.S3C, p.P291T and p.T515M as likely benign ones. The biomarkers are likely to be particularly useful in assessment of variants where there is most doubt over the functional consequences, e.g. novel missense variants which have been found in population sequencing databases at an allele frequency of $\leq 0.005\%$. It seems likely that most variants with a higher % MAF will be benign.

Since initiating this study, it has become clear that the phenotypical spectrum of *HNF1A* alleles is much wider than originally thought, so that while functionally deleterious alleles frequently cause MODY, this can by no means be assumed for every individual who possesses that allele [40]. This shows the complexity of the interpretation of genetic variation which we consider a limitation of our, or any similar studies. In clinical practice, every case needs to be assessed on the basis of the individual phenotype and the

predicted functional effect of the *HNF1A* allele. Co-segregation of allele with diabetes in a family and clinical response to sulphonylureas will help confirm the diagnosis.

In conclusion, we found that the biomarkers GP30 and hsCRP could differentiate individuals with early onset diabetes and likely damaging *HNF1A* alleles from those with young adult-onset non-autoimmune diabetes and without rare *HNF1A* alleles. A diagnostic protocol combining clinical features with biomarkers could improve the selection of subjects for genetic testing for HNF1A-MODY, the commonest form of monogenic diabetes in adults. Currently, easier availability of the hsCRP assay makes it a more immediate prospect, while for wider use of *N*-glycans, a simpler assay for determining antennary fucose levels would have to be developed.

Acknowledgements

AJ, TP, KRO, and OG wrote the manuscript, EPM, NRP and KRO enrolled patients, FV, TP, AJ, OG, and KRO analysed and interpreted data, AJ, TP, AJB, NS, CJG, MŠ, KC, CB, MVL, JČK and TJJ acquired data, KRO, ALG, GL, MIM and OG designed the study. All authors read, critically revised and approved the final manuscript.

OG and KRO are the guarantors of this work and, as such, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

This work was supported by European Foundation for Study of Diabetes/Novo Nordisk research grant, European Community's Seventh Framework Programme IntegraLife (contract number 315997), Croatian Science Foundation (UIP-2014-09-7769), Diabetes UK and the NIHR Oxford Biomedical Research Centre. The views expressed are those

of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

ALG is a Wellcome Trust Senior Fellow in Basic Biomedical Science (095101/Z/10/Z and 200837/Z/16/Z). AJ is a Diabetes UK George Alberti fellow.

GL declares that he is a founder and owner, and FV declares that he is employee of Genos Ltd, which offers commercial service of glycomic analysis and has several patents in this field. Other authors have no relevant conflict of interest to disclose.

This study was presented in abstract form at the Annual Meeting of the Society for Glycobiology San Francisco, CA, 1–4 December, 2015, 23rd International Symposium on Glycoconjugates, Split, Croatia, 15-20 September, 2015, 6th Meeting of the EASD Study Group on Genetics of Diabetes, Leiden, the Netherlands and 53rd Annual Meeting of the EASD, Lisbon, Portugal, 11-15 September 2017.

The authors would like to thank all of the participants, research nurses, local investigators and other clinical staff for their contribution to this study. We would also like to thank Dr Beverley Shields at the University of Exeter for providing the MODY probability model.

References

1. Shields, B.M., et al., *Maturity-onset diabetes of the young (MODY): how many cases are we missing?* Diabetologia, 2010. **53**(12): p. 2504-8.
2. Ellard, S., et al., *Best practice guidelines for the molecular genetic diagnosis of maturity-onset diabetes of the young.* Diabetologia, 2008. **51**(4): p. 546-53.
3. Murphy, R., S. Ellard, and A.T. Hattersley, *Clinical implications of a molecular genetic classification of monogenic beta-cell diabetes.* Nat Clin Pract Endocrinol Metab, 2008. **4**(4): p. 200-13.
4. Pearson, E.R., et al., *Genetic cause of hyperglycaemia and response to treatment in diabetes.* Lancet, 2003. **362**(9392): p. 1275-81.
5. Shields, B.M., et al., *The development and validation of a clinical prediction model to determine the probability of MODY in patients with young-onset diabetes.* Diabetologia, 2012. **55**(5): p. 1265-72.
6. Pihoker, C., et al., *Prevalence, characteristics and clinical diagnosis of maturity onset diabetes of the young due to mutations in HNF1A, HNF4A, and glucokinase: results from the SEARCH for Diabetes in Youth.* J Clin Endocrinol Metab, 2013. **98**(10): p. 4055-62.
7. Shepherd, M., et al., *Systematic Population Screening, Using Biomarkers and Genetic Testing, Identifies 2.5% of the U.K. Pediatric Diabetes Population With Monogenic Diabetes.* Diabetes Care, 2016. **39**(11): p. 1879-1888.
8. Odom, D.T., et al., *Control of pancreas and liver gene expression by HNF transcription factors.* Science, 2004. **303**(5662): p. 1378-81.
9. Toniatti, C., et al., *Synergistic trans-activation of the human C-reactive protein promoter by transcription factor HNF-1 binding at two distinct sites.* EMBO J, 1990. **9**(13): p. 4467-75.
10. Reiner, A.P., et al., *Polymorphisms of the HNF1A gene encoding hepatocyte nuclear factor-1 alpha are associated with C-reactive protein.* Am J Hum Genet, 2008. **82**(5): p. 1193-201.
11. Ridker, P.M., et al., *Loci related to metabolic-syndrome pathways including LEPR, HNF1A, IL6R, and GCKR associate with plasma C-reactive protein: the Women's Genome Health Study.* Am J Hum Genet, 2008. **82**(5): p. 1185-92.
12. Thanabalasingham, G., et al., *A large multi-centre European study validates high-sensitivity C-reactive protein (hsCRP) as a clinical biomarker for the diagnosis of diabetes subtypes.* Diabetologia, 2011. **54**(11): p. 2801-10.
13. Lauc, G., et al., *Genomics meets glycomics-the first GWAS study of human N-Glycome identifies HNF1alpha as a master regulator of plasma protein fucosylation.* PLoS Genet, 2010. **6**(12): p. e1001256.
14. Gornik, O., T. Pavic, and G. Lauc, *Alternative glycosylation modulates function of IgG and other proteins - implications on evolution and disease.* Biochim Biophys Acta, 2012. **1820**(9): p. 1318-26.
15. Varki, A., R. Kannagi, and B.P. Toole, *Glycosylation Changes in Cancer, in Essentials of Glycobiology, A. Varki, et al., Editors. 2009: Cold Spring Harbor (NY).*
16. Lauc, G., et al., *Loci associated with N-glycosylation of human immunoglobulin G show pleiotropy with autoimmune diseases and haematological cancers.* PLoS Genet, 2013. **9**(1): p. e1003225.
17. Thanabalasingham, G., et al., *Mutations in HNF1A result in marked alterations of plasma glycan profile.* Diabetes, 2013. **62**(4): p. 1329-37.
18. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chain-terminating inhibitors.* Proc Natl Acad Sci U S A, 1977. **74**(12): p. 5463-7.
19. Richards, S., et al., *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, 2015. **17**(5): p. 405-24.

20. Lek, M., et al., *Analysis of protein-coding genetic variation in 60,706 humans*. Nature, 2016. **536**(7616): p. 285-91.
21. Huffman, J.E., et al., *Comparative Performance of Four Methods for High-throughput Glycosylation Analysis of Immunoglobulin G in Genetic and Epidemiological Research*. Molecular & Cellular Proteomics, 2014. **13**(6): p. 1598-1610.
22. Saldova, R., et al., *Association of N-glycosylation with breast carcinoma and systemic features using high-resolution quantitative UPLC*. J Proteome Res, 2014. **13**(5): p. 2314-27.
23. Schlosser, M., et al., *Diabetes Antibody Standardization Program: evaluation of assays for insulin autoantibodies*. Diabetologia, 2010. **53**(12): p. 2611-20.
24. Colclough, K., et al., *Mutations in the genes encoding the transcription factors hepatocyte nuclear factor 1 alpha and 4 alpha in maturity-onset diabetes of the young and hyperinsulinemic hypoglycemia*. Hum Mutat, 2013. **34**(5): p. 669-85.
25. Ellard, S., et al., *Partial and whole gene deletion mutations of the GCK and HNF1A genes in maturity-onset diabetes of the young*. Diabetologia, 2007. **50**(11): p. 2313-2317.
26. Moller, A.M., et al., *Mutations in the hepatocyte nuclear factor-1 alpha gene in Caucasian families originally classified as having Type I diabetes*. Diabetologia, 1998. **41**(12): p. 1528-1531.
27. Frayling, T.M., et al., *beta-cell genes and diabetes - Molecular and clinical characterization of mutations in transcription factors*. Diabetes, 2001. **50**: p. S94-S100.
28. Kaisaki, P.J., et al., *Mutations in hepatocyte nuclear factor 1 alpha gene in maturity onset diabetes of the young (MODY3)*. Diabetologia, 1997. **40**: p. 614-614.
29. Lehto, M., et al., *Characterization of the MODY3 phenotype - Early-onset diabetes caused by an insulin secretion defect*. Journal of Clinical Investigation, 1997. **99**(4): p. 582-591.
30. Vaxillaire, M., et al., *Identification of nine novel mutations in the hepatocyte nuclear factor 1 alpha gene associated with maturity-onset diabetes of the young (MODY3)*. Human Molecular Genetics, 1997. **6**(4): p. 583-586.
31. Kawasaki, E., et al., *Identification and functional analysis of mutations in the hepatocyte nuclear factor-1 alpha gene in anti-islet autoantibody-negative Japanese patients with type 1 diabetes*. Journal of Clinical Endocrinology & Metabolism, 2000. **85**(1): p. 331-335.
32. Domínguez-López A, M.-G.A., Segura-Kato YX, Riba L, Esparza-López R, Ramírez-Jiménez S, Rodríguez-Torres M, Canizales-Quinteros S, Cabrera-Vásquez S, Fragoso-Ontiveros V, Aguilar-Salinas CA, Altamirano-Bustamante N, Calzada-León R, Robles-Valdés C, Bravo-Ríos LE, Tusié-Luna MT., *Mutations in MODY genes are not common cause of early-onset type 2 diabetes in Mexican families*. JOP, 2005. **6**(3): p. 238-45.
33. Xu, J.Y., et al., *Genetic and clinical characteristics of maturity-onset diabetes of the young in Chinese patients*. European Journal of Human Genetics, 2005. **13**(4): p. 422-427.
34. Galan, M., et al., *Differential Effects of HNF-1 alpha Mutations Associated with Familial Young-Onset Diabetes on Target Gene Regulation*. Molecular Medicine, 2011. **17**(3-4): p. 256-265.
35. Yamada, S., et al., *Identification of mutations in the hepatocyte nuclear factor-1alpha gene in Japanese subjects with early-onset NIDDM and functional analysis of the mutant proteins*. Diabetes, 1999. **48**(3): p. 645-8.
36. Yamagata, K., et al., *Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset diabetes of the young (MODY3)*. Nature, 1996. **384**(6608): p. 455-8.
37. Bjorkhaug, L., et al., *Hepatocyte nuclear factor-1 alpha gene mutations and diabetes in Norway*. J Clin Endocrinol Metab, 2003. **88**(2): p. 920-31.
38. Najmi, L.A., et al., *Functional Investigations of HNF1A Identify Rare Variants as Risk Factors for Type 2 Diabetes in the General Population*. Diabetes, 2016.

39. Shields, B.M., et al., *Population-Based Assessment of a Biomarker-Based Screening Pathway to Aid Diagnosis of Monogenic Diabetes in Young-Onset Patients*. *Diabetes Care*, 2017. **40**(8): p. 1017-1025.
40. Flannick, J., et al., *Assessing the phenotypic effects in the general population of rare variants in genes for a dominant Mendelian form of diabetes*. *Nat Genet*, 2013. **45**(11): p. 1380-5.
41. Kaisaki, P.J., et al., *Mutations in the hepatocyte nuclear factor-1 alpha gene in MODY and early-onset NIDDM: Evidence for a mutational hotspot in exon 4 (vol 46, pg 528, 1997)*. *Diabetes*, 1997. **46**(7): p. 1239-1239.
42. Pavic, T., et al., *Maturity onset diabetes of the young due to HNF1A variants in Croatia*. *Biochem Med (Zagreb)*, 2018. **28**(2): p. 020703.

Table 1 Clinical characteristics of the recruited subjects (probands only) with young adult-onset non-autoimmune diabetes*

Group depending on rare <i>HNF1A</i> allele status	(Likely) damaging allele	Allele variant of unknown significance	(Likely) benign allele	No rare <i>HNF1A</i> allele variant	p-value
Number of individuals	16	4	9	960	
Gender (N (%) of M)	5 (31.3%)	1 (25.0%)	6 (66.7%)	575 (59.9%)	0.057
Age at recruitment (years)	34.0 (19.5)	62.5 (15.3)	47.0 (16.5)	47.0 (12.0)	0.001
Age at diagnosis (years)	24.5 (11.5)	35.0 (19.0)	37.0 (7.5)	37.0 (8.0)	<0.001
Duration of diabetes (years)	8.5 (12.0)	30.5 (28.8)	9.0 (14.0)	10.0 (12.0)	0.136
BMI (kg/m ²)	25.5 (9.4)	27.6 (6.5)	31.0 (10.1)	30.4 (8.3)	0.007
FPG (mmol/L)	7.20 (3.00)	7.15 (4.20)	9.90 (5.10)	8.10 (3.80)	0.180
HbA1c (mmol/mol (%))	57 (7.36)	67 (8.25)	77 (9.20)	58 (7.50)	0.107
C-peptide (nmol/L)	0.33 (0.39)	0.26 (0.13)	0.67 (0.46)	0.70 (0.55)	<0.001
Total cholesterol (mmol/L)	4.84 (1.15)	3.91 (2.01)	4.50 (1.45)	4.60 (1.50)	0.783
HDL (mmol/L)	1.33 (0.51)	1.51 (0.50)	1.07 (0.40)	1.15 (0.40)	0.017
Triglycerides (mmol/L)	1.10 (0.41)	1.20 (0.53)	1.60 (1.53)	1.60 (1.26)	0.020
Treatment[†]					
					0.001
Insulin + SU/glinide	0 (0.0%)	1 (25.0%)	2 (22.2%)	33 (3.4%)	
Insulin + other OHA	1 (6.3%)	1 (25.0%)	2 (22.2%)	210 (21.9%)	
SU/glinides monotherapy	3 (18.8%)	0 (0.0%)	0 (0.0%)	21 (4.3%)	
SU/glinides + other OHA	3 (18.8%)	0 (0.0%)	1 (11.1%)	190 (38.8%)	
Other OHA	0 (0.0%)	0 (0.0%)	0 (0.0%)	279 (56.9%)	
Insulin	5 (31.2%)	1 (25.0%)	3 (33.3%)	137 (14.3%)	
Diet	4 (25.0%)	1 (25.0%)	1 (11.1%)	90 (9.4%)	

*Continuous variables are given as median (IQR), while categorical variables are given as

percentages. Kruskal-Wallis test was applied to compare groups for continuous data, while the

differences of frequencies for categorical variables were tested using the Chi-squared test. †

treatment at the time of inclusion in the study. BMI - body mass index; FPG – fasting plasma

glucose; HDL – high-density lipoprotein; OHA - oral hypoglycaemic agents +/- GLP1 analogue;

SU – sulphonylurea derivatives. Statistically significant p-value in bold (<0.05).

Table 2 Rare *HNF1A* allele variants identified in the study

Coding DNA variant	Protein variant	Variant type	Reported as causing MODY	Bioinformatics prediction	Allele frequency in GnomAD [%]	Functional work in this or previous studies	Co-segregation of the variant with DM	Probability of MODY using MODY Calculator	Current prediction
c.-4A>G	n/a	splice site	no	n/a	0.08	not performed	not available	S Asian and too old	benign
c.1-326del	del exon 1	exon del	yes [25]	protein truncating	0	not performed as PTV	Yes, 2 relatives with variant & DM [25]	4.6%	damaging
c.8C>G*	S3C	missense	novel	damaging	0	not performed	not available	Too old	VUS
c.139G>C	G47R	missense	yes [24]	neutral	0.001	TA 80-112% of WT, WB 110% of WT‡	Yes, 1 relative with variant & DM [17]	Too old	benign
c.142G>A	E48K	missense	yes [26]	neutral	0.009	TA 63% WT, CNF = WT [38]	Yes, 2 relatives with variant & DM [26]	Too old	benign
c.404delA	D135fs	frameshift	yes [27]	protein truncating	0	not performed as PTV	yes, 2 relatives with variant & DM [27]	>75.5%	damaging
c.451G>A*	G151S	missense	novel	damaging	0	TA 82% WT, WB=WT, DNA binding 15% WT (this study)	not available	>15%	likely damaging
c.586A>G	T196A	missense	no [34]	neutral	0.027	DNA binding & TA=WT [34]	no [34]	Too old	benign
c.666G>T	K222N	missense	novel	damaging	0	TA 76% WT, WB, DNA binding = WT (this study)	3 DM gen., only proband sequenced	>75.5%	likely damaging
c.685C>T*	R229*	nonsense	yes [41]	protein truncating	0.0008	TA 0-7% of WT [35]	yes, 2 relatives with variant & DM [41]	>75.5% & relative too old	damaging
c.751G>A*	A251T	missense	no	neutral	0	TA 80-92% of WT, WB 98% of WT‡	Yes, 2 relatives with variant & DM (this study)	2 subjects too old	VUS

c.779C>T	T260M	missense	yes [29]	damaging	0	TA, WB * DNA binding 10-20% WT (this study)	Yes, 7 relatives with variant & DM [29], this study (2 relatives with variant & DM)	>45.5 and 62.4% (2 subjects)	damaging
c.862G>T*	G288W	missense	no	neutral	0.007	TA 73% WT, WB * DNA binding=WT (this study)	no (this study)	Too old, no age of the 2 nd subject	likely benign
c.871C>A	P291T	missense	yes	neutral	0.0008	TA 76% WT, WB, DNA binding =WT (this study)	yes, 2 relatives with variant & DM [12], this study	>45.5, >4.6% and >4.6% (3 subjects)	VUS
c.872delC	P291fs	frameshift	yes [30]	protein truncating	0	not performed as PTV	yes, 7 relatives with variant & DM [30]	>75 & >4.6% (2 probands)	damaging
c.872dupC†	G292fs	frameshift	yes [29]	protein truncating	0	TA <10% of WT, <5% mRNA expression [36]	yes, 6-25 relatives with variant & DM [29]	>75.5, >2.6 & >2.6% (3 probands)	damaging
c.1015G>A	G339S	missense	No	neutral	0.02	TA 85-100% of WT, WB=WT‡	yes, 1 relative with variant & DM (Mughal, unpublished)	>32.9%	benign
c.1047C>A	H349Q	missense	no	neutral	0.006	TA, WB =WT, DNA binding 76% WT (this study)	not available	>4.6%	benign
c.1129delC	L377fs	frameshift	yes [24]	protein truncating	0	not performed as PTV	2 DM gen., only proband sequenced	>45.5%	damaging
c.1136_1137delCT†	P379fs	frameshift	yes [31]	protein truncating	0	TA 6-62%, DNA binding 37% of WT [37]	yes, 2 relatives with variant & DM [31]	>32.9 & 4.6% (2 probands)	damaging
c.1136C>A	P379H	missense	yes [32]	damaging	0.005	TA 38-58% of WT [32]	yes, 2 relatives with variant & DM [32]	Too old	likely damaging

c.1136C>G*	P379R	missense	yes [33]	damaging	0	TA 70-80% of WT, DNA binding=WT [34]	yes, 1 relative with variant & DM [33], this study, 1 relative with variant & DM	>49.4% & no age available (2 subjects)	damaging
c.1165T>G	L389V	missense	no	neutral	0.06	TA 70% WT, CNF = WT [38]	no co-segregation provided	Too old	benign
c.1544C>T†	T515M	missense	no	damaging	0.002	TA 70-80% of WT, WB 95% of WT‡	not available	>15.1%	VUS
c.1816G>A	G606S	missense	yes [24]	neutral	0.005	TA, WB * DNA binding=WT‡	2 DM gen., only proband sequenced	>4.6%	benign

*rare allele variants found in the Croatian individuals previously reported by our group [42], †allele present in two unrelated probands from UK and Croatian cohort, ‡unpublished data from our group; WT - wild type HNF1A; PTV - protein truncating variant; WB - western blot; CNF - cytosol-nuclear fractionation; TA - transcription activity; DM gen. - generations with diabetes.

Figure legends

Figure 1 Functional assessment of previously uncharacterised HNF1A allele variants. A) transcription activity using luciferase reporter assay, n=3; B) protein expression from the western blot [representative blot image aligned at the bottom with HNF1A band at the top and B-tubulin (technical control) at the bottom] quantified by densitometry, n=3; C) DNA binding of HNF1A protein variants performed by electrophoretic mobility shift assay (EMSA) with a representative gel image at the bottom of the graph (the top arrow pointing to HNF1A antibody-HNF1A protein “super-shift” and bottom arrow to HNF1A bands), HNF1A protein variant bound to the probe quantified from densitometry, n=3; D) DNA binding by EMSA corrected for protein amount. All presented as a mean percentage of the WT HNF1A (n=3) with error bars; p-value is obtained by ANOVA and corrected for multiple comparisons (*p <0.05, **p<0.01, ***p<0.0001); empty vector in cream, WT and synonymous variant p.H179H in green, positive control variants in red, variant increasing risk of type 2 diabetes in orange, tested variants in blue.

Figure 2 A) Representative HILIC-UPLC chromatographic profile of *N*-glycans released from total plasma proteins. Glycan peaks which exhibited the best discriminative power between HNF1A-MODY and early-onset type 2 diabetes are colour-coded as follows: GP30 in green, GP38 in yellow and GP36 in pink. *N*-glycan structures contained in the listed above peaks are also depicted as per legend. Part B-E illustrates the level of GP30 (B), GP36 (C), GP38 (D) and hsCRP (E) according to the type of *HNF1A* allele variants, alongside ROC curves illustrating the performance of the particular biomarker in differentiating subjects with damaging *HNF1A* alleles from all other subjects. Subjects are divided into 4 groups: subjects without the rare *HNF1A* allele variant (NV, in red), subjects with benign alleles (B, in green), subjects with allele variants of unknown significance (VUS, in blue) and subjects with damaging *HNF1A* alleles (D, in purple). Differences in glycan groups and hsCRP are shown as box plots. Each box represents the 25th to 75th percentile. Lines inside the boxes represent the median. The upper whisker extends from the hinge to the highest value that is within 1.5 x IQR of the hinge, where IQR is the inter-quartile range, or distance between the

first and third quartiles. The lower whisker extends from the hinge to the lowest value within 1.5 x IQR of the hinge. Circles indicate outliers. AUC, area under the curve; GP, glycan peak.