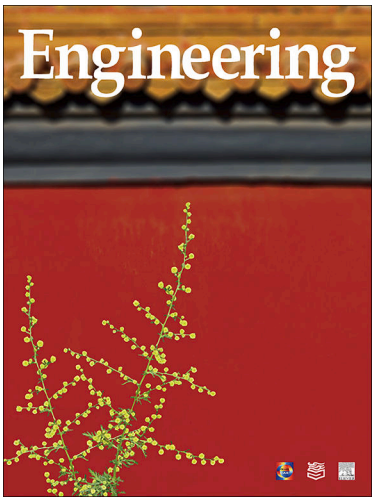
Journal Pre-proofs



Article

Transcription Factors HNF1A, HNF4A, and FOXA2 Regulate Hepatic Cell Protein *N*-Glycosylation

Vedrana Vičić Bočkor, Nika Foglar, Goran Josipović, Marija Klasić, Ana Vujić, Branimir Plavša, Toma Keser, Samira Smajlović, Aleksandar Vojta, Vlatka Zoldoš

PII: S2095-8099(23)00472-1
DOI: https://doi.org/10.1016/j.eng.2023.09.019
Reference: ENG 1407

To appear in: *Engineering*

Received Date: 7 October 2022
Revised Date: 30 August 2023
Accepted Date: 28 September 2023

Please cite this article as: V. Vičić Bočkor, N. Foglar, G. Josipović, M. Klasić, A. Vujić, B. Plavša, T. Keser, S. Smajlović, A. Vojta, V. Zoldoš, Transcription Factors HNF1A, HNF4A, and FOXA2 Regulate Hepatic Cell Protein *N*-Glycosylation, *Engineering* (2023), doi: https://doi.org/10.1016/j.eng.2023.09.019

Research

Glycomedicine—Article

Transcription Factors HNF1A, HNF4A, and FOXA2 Regulate Hepatic Cell Protein *N*-Glycosylation

Vedrana Vičić Bočkor [a,#], Nika Foglar [a,#], Goran Josipović [b], Marija Klasić [a], Ana Vujić [a], Branimir Plavša [c], Toma Keser [c], Samira Smajlović [a], Aleksandar Vojta [a,b,*], Vlatka Zoldoš [a,*]

[a] *Laboratory for Epigenetics, Department of Biology, Faculty of Science, University of Zagreb, Zagreb 10000, Croatia*

[b] *Genos Glycoscience Research Laboratory, Borongajska cesta 83H, Zagreb 10000, Croatia*

[c] *Faculty of Pharmacy and Biochemistry, University of Zagreb, Ante Kovačića 1, Zagreb 10000, Croatia*

[#] These authors contributed equally to this work.

* Corresponding authors.

E-mail addresses: vzoldos@biol.pmf.hr (V. Zoldoš), vojta@biol.pmf.hr (A. Vojta).

ARTICLE INFO

Article history:

Received 7 October 2022

Revised 30 August 2023

Accepted 28 September 2023

Available online

Keywords

Clustered regularly interspaced short palindromic repeats/dead Cas9 (CRISPR/dCas9)

Epigenetics

Hepatocyte nuclear factor 1 alpha (HNF1A)

Hepatocyte nuclear factor 4 alpha (HNF4A)

Forkhead box protein A2 (FOXA2)

N-glycosylation

HepG2 cells

Abstract

Hepatocyte nuclear factor 1 alpha (HNF1A), hepatocyte nuclear factor 4 alpha (HNF4A), and forkhead box protein A2 (FOXA2) are key transcription factors that regulate a complex gene network in the liver, creating a regulatory transcriptional loop. The Encode and ChIP-Atlas databases identify the recognition sites of these transcription factors in many glycosyltransferase genes. Our *in silico* analysis of HNF1A, HNF4A, and FOXA2 binding to the 10 candidate glyco-genes studied in this work confirms a significant enrichment of these transcription factors specifically in the liver. Our previous studies identified HNF1A as a master regulator of fucosylation, glycan branching, and galactosylation of plasma glycoproteins. Here, we aimed to functionally validate the role of the three transcription factors on downstream glyco-gene transcriptional expression and the possible effect on glycan phenotype. We used the state-of-the-art clustered regularly interspaced short palindromic repeats/dead Cas9 (CRISPR/dCas9) molecular tool for the downregulation of the *HNF1A*, *HNF4A*, and *FOXA2* genes in HepG2 cells—a human liver cancer cell line. The results show that the downregulation of all three genes individually and in pairs affects the transcriptional activity of many glyco-genes, although downregulation of glyco-genes was not always followed by an unambiguous change in the corresponding glycan structures. The effect is better seen as an overall change in the total HepG2 *N*-glycome, primarily due to the extension of biantennary glycans. We propose an alternative way to evaluate the *N*-glycome composition via estimating the overall complexity of the glycome by quantifying the number of monomers in each glycan structure. We also propose a model showing feedback loops with the mutual activation of HNF1A–FOXA2 and HNF4A–FOXA2 affecting glyco-genes and protein glycosylation in HepG2 cells.

1. Introduction

Glycosylation significantly modifies the structure and function of proteins, but current knowledge about the molecular regulatory mechanisms that underlie alternative glycosylation is still incomplete. The process of glycosylation is mediated by an array of enzymes catalyzing the sequential addition of monosaccharides to the protein backbone. These enzymes include many different glycosyltransferases (galactosyltransferases, fucosyltransferases, sialyltransferases, etc.), glycosidases (mannosidase, fucosidase, glucosidase, etc.), and enzymes involved in monosaccharide biosynthesis and transport, as well as transcription factors [1]. The regulation of protein glycosylation is highly complex due to the sheer number of enzymes participating in the non-templated process of glycoprotein synthesis and maturation. Despite many studies relating glycosyltransferase activity to glycan composition, little evidence has been generated to support the hypothesis that glycosylation is regulated at the level of glycosyltransferase expression [2,3]. In addition, protein glycosylation is cell-type specific and regulated on a transcriptional, translational, and post-translational level [2,4,5]. The transcription of genes involved in glycan biosynthesis (i.e., glyco-genes) is controlled by the actions of transcription factors on glyco-gene promotors and enhancers, and by epigenetic factors affecting the accessibility of these regions [6–9]. Altered protein glycosylation appears in many types of disease, including chronic inflammatory, autoimmune, and infectious diseases, as well as in cancer [6,7,10,11], and is often the result of epigenetic changes affecting the transcription of glyco-genes or transcription factors themselves.

In the liver, the major transcription factors regulating the processes of blood coagulation, innate immunity, cellular detoxification, and the maintenance of glucose and lipid homeostasis belong to the hepatocyte nuclear factor families (HNFs). There are four families of HNFs: hepatocyte nuclear factor 1 (HNF1), forkhead box protein A (FOXA), hepatocyte nuclear factor 4 (HNF4), and one cut homeobox 1 (ONECUT, OC or HNF6); while they share common features such as DNA binding and trans-activation, they also possess different structural domains, leading to their distinct roles [12–14]. Each family comprises several members, and the isoforms of different families are inter-regulated in transcriptional loops. Hepatocyte nuclear factor 1 alpha (HNF1A) and hepatocyte nuclear factor 4 alpha (HNF4A) are the major transcription factors in the liver, regulating the processes of hepatocyte differentiation, mitochondrial metabolism, ureagenesis, drug transport and metabolism, fatty acid metabolism, blood coagulation, lipid and carbohydrate metabolism, liver regeneration, and many others [15–19]. They have also been proposed to act as tumor suppressors in pancreatic cancer and hepatocellular carcinoma (HCC) [20–24].

According to the Encode database, the HNF1A transcription factor binds more than a thousand gene promoters in hepatocytes, including key glycosyltransferases such as *FUT5*, *FUT6*, *B4GALT1*, *MGAT4A*, *MGAT5*, *ST6GAL1*, and *GMDS*. Given its key role in many hepatic functions, HNF4A is usually referred to as the master regulator of

hepatic function. The forkhead box protein A2 (FOXA2) transcription factor is presumed to act upstream of the HNF1A/HNF4A transcriptional regulatory network regulating HNF1A/HNF4A transcription [14]. FOXA2 null mutants fail to express forkhead box protein A1 (FOXA1) and show reduced levels of HNF1A and HNF4A in mice [25,26]. In humans, mutations in FOXA2 are associated with HCC in a sexually dimorphic manner [27]. Odom and collaborators [22,26] have shown that HNF1A binds to at least 222 and HNF4A to around 1560 gene promoters in hepatocytes, confirming their principal role as regulators of hepatic transcription. HNF1A and HNF4A, together with FOXA2 and forkhead box protein A3 (FOXA3), create a regulatory network responsible for the transcriptional program of the β cells of langerhans islets and hepatocytes. Our previous genome-wide association study (GWAS) of the human plasma *N*-glycome identified HNF1A as a master regulator of the key fucosyltransferases and the fucose biosynthesis genes [28], while ChIP-Atlas analysis has shown HNF1A binding to promoters of the *ST6GAL1* and *MGAT4B* genes. Furthermore, our previous study revealed a correlation between cytosine guanine dinucleotide (CpG) methylation in the first exon of the *HNF1A* gene and glycan branching, and galactosylation in the human plasma *N*-glycome [29]. All the reasons mentioned above guided our selection of the candidate genes to be studied in this work: the genes coding for fucosyltransferases involved in antennary (*FUT3*, *FUT5*, *FUT6)* and core fucosylation (*FUT8*), together with *FUK* and *GMDS* encoding fucokinase and guanosine diphosphate (GDP)-mannose 4,6-dehydratase, respectively; and the *MGAT3*, *MGAT4A*, and *MGAT5* genes coding for the glycosyltransferases responsible for glycan branching, as well as the *B4GALT1* and *ST6GAL1* genes encoding the glycosyltransferases responsible for galactosylation and sialylation, respectively.

The aim of this study was to investigate the role of the major hepatic transcription factors HNF1A, HNF4A, and FOXA2 in the transcriptional activity of the glyco-genes coding for the main glycosyltransferases. For this, we used the state-of-the-art clustered regularly interspaced short palindromic repeats/dead Cas9 (CRISPR/dCas9) molecular tool [8,30]. We were also interested in whether the manipulation of HNF1A, HNF4A, and FOXA2 would affect the whole cell protein *N*-glycosylation. Since most glycoproteins in the human plasma originate from the liver, HepG2 cells of hepatic origin were chosen for these experiments. Even though primary hepatic cells would be a much more relevant model, manipulations using CRISPR/dCas9 methodology require a time frame exceeding the lifespan of primary cells in culture. Studies from other groups have shown that some pathways are still intact in HepG2 cells, regardless of significant differences otherwise, and our previous studies have shown that the main protein glycosylation pathways seem intact when compared with what is expected in healthy hepatocytes [7,8,31], indicating that HepG2 cells represent an adequate model for the human liver. The results of this study show that individual and/or simultaneous downregulation of the *HNF1A*, *HNF4A*, and *FOXA2* genes affected the transcriptional activity of many glyco-genes, even though this change was not always followed by an unambiguous change in the corresponding glycan structures. Rather, the overall complexity of the total HepG2 *N*-glycome increased, primarily due to the extension of biantennary glycans. Here, we propose an alternative way to evaluate the *N*-glycome composition, as well as a model showing feedback loops with the mutual activation of HNF1A–FOXA2 and HNF4A–FOXA2 affecting glyco-genes and protein glycosylation in HepG2 cells.

2. Material and methods

2.1. In silico ChIP enrichment analysis

To check the interactions between the transcription factors HNF1A, HNF4A, and FOXA2 and the glyco-genes *B4GALT1*, *ST6GAL1*, *MGAT3*, *MGAT4A*, *MGAT5*, *FUT3*, *FUT5*, *FUT6*, *FUT8*, *FUK*, and *GMDS* in different tissues, we used the ChIP-Atlas database[†] [32,33], "Enrichment Analysis" module, experiment type "ChIP: TFs and others", threshold for significance of 50, and Refseq coding genes as the control dataset. We ran three enrichment analyses for each tissue type: transcription factors only, glyco-genes only, and all genes (listed above). We analyzed enrichment in liver, lung, digestive system, blood, neural, and muscle cell types.

2.2. Plasmids construction

The fusion construct Krüppel associated box (KRAB)-dSpCas9 was used to target the *HNF1A*, *HNF4A*, and *FOXA2* gene locus and to simultaneously target the *HNF4A/FOXA2*, *HNF1A/FOXA2*, and *HNF1A/HNF4A* gene pairs in HepG2 cells. Fusion constructs were made using a modular system with the BsaI assembly reaction, as described in Ref. [30]. The final constructs had the fusion part KRAB-dSpCas9 along with the fluorescence marker mRuby3 and the selection marker for puromycin resistance under the strong chicken β-actin (CBh) promoter. One guide RNA (gRNA) molecule was used to target the *HNF1A* gene, while two different gRNA molecules

[†] https://chip-atlas.org.

were used to target the *FOXA2* or *HNF4A* gene. A construct co-expressing non-targeting gRNA (NT-gRNA), which has no sequence homology in the human genome, was used as a negative control in the experiments (Table S1 in Appendix A).

2.3. Cell culture and transfections

A human HepG2 cell line (ACC 180, DSMZ, Germany) was maintained in RPMI-1640 Medium (Sigma-Aldrich, USA) supplemented with 10% heat-inactivated fetal bovine serum (Sigma-Aldrich). Cells were incubated at 37 °C in a humidified 5% carbon dioxide (CO_2)-containing atmosphere. Transfections of HepG2 cells were done using polyethyleneimine (PEI) MAX 40K (Polysciences, USA), according to the manufacturer's protocol. In brief, HepG2 cells were seeded in 100 mm culture dishes one day before transfection; they were then transfected the next day at around 80% confluency with 8 µg of plasmid in nine biological replicates. The mass ratio of PEI to DNA used was 3 : 1. Cells were screened 24 h post-transfection for expression of the fluorescent protein mRuby3 and were selected with puromycin (Gibco Life Technologies, USA) for 48 h. Cells were then collected on the fourth day after transfection for subsequent DNA, RNA, and protein isolation.

2.4. Gene expression analysis via reverse transcription-quantitative polymerase chain reaction (RT-qPCR)

Total RNA was isolated using an RNeasy Mini Kit (Qiagen, Germany) and treated with DNase using a TURBO DNA-free Kit (Invitrogen, USA), according to the manufacturer's protocol. Reverse transcription was done using PrimeScript Reverse Transcriptase (TaKaRa, Japan) and random hexamer primers (Invitrogen). RT-qPCR was performed according to the manufacturer's protocol using the 7500 Fast Real-Time Polymerase Chain Reaction (PCR) System, TaqMan Gene Expression Master Mix, and the following TaqMan Gene Expression Assays (Applied Biosystems, USA): Hs00167041_m1 (*HNF1A*), Hs00232764_m1 (*FOXA2*), Hs00230853_m1 (*HNF4A*), Hs00382135_m1 (*FUK*), Hs01046865_m1 (*GMDS*), Hs01868572_s1 (*FUT3*), Hs00704908_s1 (*FUT5*), Hs03026676_s1 (*FUT6*), Hs00189535_m1 (*FUT8*), Hs02379589_s1 (*MGAT3*), Hs00923405_m1 (*MGAT4A*), Hs00159136_m1 (*MGAT5*), Hs00155245_m1 (*B4GALT1*), and Hs00949382_m1 (*ST6GAL1*). Gene expression was normalized to the *HMBS* gene (Hs00609297_m1) and analyzed using the comparative cycle threshold (Ct) method [34]. In all experiments, expression was shown as the fold change (FC) relative to non-target-transfected cells.

2.5. Protein extraction and Western blot analysis

For Western blot analysis, cells were lysed in radioimmunoprecipitation assay (RIPA) buffer (50 mmol·L^{-1} tris(hydroxymethyl)aminomethane (TRIS, pH = 7.5), 0.1% Triton X, 1 mmol·L^{-1} ethylenediaminetetraacetic acid (EDTA), 135 mmol·L^{-1} sodium chloride (NaCl)) supplemented with Protease Inhibitor Cocktail (cOmplete™ ULTRA Tablets, EDTA-free, glass vials, Protease Inhibitor Cocktail) (Roche, Switzerland). 25 µg of proteins was separated on 10% polyacrylamide gels and transferred to nitrocellulose membranes (Amersham™Protran® Premium 0.45 µm NC) (GE Healthcare, Germany). Membranes were blocked for 1 h with 0.1% Tween 20 in tris-buffered saline (TBS-T) containing 5% milk (Milchpulver blotting grade) (Roth, Germany) prior to incubation with primary antibodies. The primary antibodies used were diluted in blocking buffer as follows: 1 : 2000 anti-HNF1A (ab272693) (Abcam, UK), anti-HNF4A 1 : 1000 (sc-374229) (Santa Cruz, USA), 1 : 2000 anti-FOXA2 (ab 60721) (Abcam), 1 : 2000 anti-histone H3 (ab1791) (Abcam), and 1 : 1000 anti-β-Actin (sc-69879) (Santa Cruz). The incubations in primary antibodies were performed overnight at 4 °C. The membranes were then incubated in horseradish peroxidase (HRP)-conjugated secondary goat anti-mouse (ab 205719) or goat anti-rabbit antibodies (ab 6721) (Abcam). Signals were developed using Clarity Max™ Western ECL Substrate (BioRad, USA) and photographed using the Alliance Q9 Advanced imaging system (Uvitec, UK).

2.6. Analysis of the total N-glycome of HepG2 cells

The total proteins were precipitated using the methanol and chloroform standard protocol after cell disruption in lysis buffer (50 mmol·L^{-1} TRIS (pH = 7.4), 0.1% Triton X-100, 1 mmol·L^{-1} EDTA, 135 mmol·L^{-1} NaCl), supplemented with Protease Inhibitor Cocktail (cOmplete™ ULTRA Tablets, EDTA-free, glass vials, Protease Inhibitor Cocktail) (Roche). The dried protein pellet was resuspended in 30 µL of 1.33% (w/v) sodium dodecyl sulfate (SDS) (Invitrogen) and incubated at 65 °C for 10 min. Subsequently, 10 µL of 4% Igepal-CA630 (Sigma-Aldrich) and 1.2 U peptide *N*-glycosidase F (PNGase F) (Promega, USA) in 10 µL of 5 × phosphate-buffered saline (PBS) were added. The samples were incubated overnight at 37 °C to permit the release of *N*-glycans. The released *N*-glycans were then labeled with procainamide (Sigma-Aldrich). The labeling mixture was freshly prepared by dissolving procainamide (38.3 mg·mL^{-1}) and 2-picoline borane (44.8 mg·mL^{-1}) (Sigma-Aldrich) in a mixture of

dimethyl sulfoxide (DMSO) (Sigma-Aldrich) and glacial acetic acid (7 : 3, v/v) (Merck, Germany). Labeling mixture (25 µL) was added to each *N*-glycan sample in the 96-well plate. Mixing was achieved by shaking for 10 min, followed by incubation at 65 °C for 2 h. To each sample (75 µL), 700 µL of acetonitrile (CAN) (JT Baker, USA) was added. Free label and reducing agents were removed from the samples using hydrophilic interaction liquid chromatography solid-phase extraction (HILIC-SPE) clean-up. A 0.45 µm hydrophilic polypropylene (GHP) filter plate (Pall Corporation, USA) was used as the stationary phase. All wells were prewashed using 1 × 200 µL of ethanol/water (7 : 3, v/v) and 1 × 200 µL water, followed by equilibration using 1 × 200 µL of acetonitrile (ACN)/water (24 : 1, v/v). The solvent was removed by applying a vacuum using a vacuum manifold (Millipore Corporation, USA). The samples were loaded into the wells, which were subsequently washed five times using 200 µL of ACN/water (24 : 1, v/v). The glycans were eluted with 2 × 50 µL of water, and the combined eluates were stored at −20 °C until usage. The fluorescently labeled and purified *N*-glycans were separated by means of hydrophilic interaction liquid chromatography (HILIC) on a Waters Acquity ultra-performance liquid chromatography (UPLC) instrument (Waters, USA) consisting of a quaternary solvent manager, a sample manager, and a fluorescence detector set with excitation and emission wavelengths of 310 and 370 nm, respectively. The instrument was controlled using Empower 2 software, build 2145 (Waters). The plasma *N*-glycans were separated on a Waters bridged ethylene hybrid (BEH) glycan column sized 150 mm × 2.1 mm, using 1.7 µm BEH particles; 100 mmol·L^{-1} ammonium formate, pH = 4.4, was used as solvent A, and ACN was used as solvent B. The separation method used a linear gradient of 53%–70% ACN (v/v) at a flow rate of 0.561 mL·min^{-1} in a 25 min analytical run. The data was processed using an automated integration method. The chromatograms were all separated in the same manner into separate peaks, where the content of glycans in each peak was expressed as a percentage of the total integrated area. All glycan structures were annotated with a tandem mass spectrometry (MS/MS) analysis via HILIC-UPLC coupled with a Synapt G2-Si electrospray ionization quadrupole time-of-flight mass spectrometry (ESI-QTOF-MS) system (Waters). The instrument was controlled using MassLynx v.4.1 software (Waters). The mass spectrometry (MS) conditions were set as follows: positive ion mode, a capillary voltage of 3 kV, a sampling cone voltage of 30 V, a source temperature of 120 °C, a desolvation temperature of 350 °C, and a desolvation gas flow of 800 L·h^{-1}. Mass spectra were recorded from 500 to 3000 *m/z* at a frequency of 1 Hz. MS/MS experiments were performed in a data-dependent acquisition (DAD) mode. Spectra were first acquired from 500 to 3000 *m/z*; then, two precursors with the highest intensities were selected for collision-induced dissociation (CID) fragmentation (100–3000 *m/z* was recorded). A collision energy ramp was used for the fragmentation (low-mass collision energy (LM CE) Ramp Start 7 V, LM CE Ramp End 13 V, high-mass collision energy (HM CE) Ramp Start 97 V, HM CE Ramp End 108 V).

2.7. Statistical analysis

All *in vitro* experiments were carried out in nine replicates. All data are shown as mean ± standard deviation (SD). Statistical significances between groups were calculated with the two-tailed Mann Whitney test using the GraphPad Prism version 5.0.3 for Windows (GraphPad Software, USA)[†]. The expression and glycan composition/complexity analysis, along with the corresponding data visualization, was done using the R language and statistical environment (R Core Team, Austria).

3. Results

3.1 In silico analysis suggests that transcription factors HNF1A, HNF4A, and FOXA2 regulate glyco-genes specifically in the human liver

In this work, we aimed to investigate the role of the HNF1A, HNF4A, and FOXA2 transcription factors in the regulation of glyco-genes in a human liver cell model—HepG2. To verify whether the three key transcription factors bind to the candidate glyco-genes specifically in the human liver, we performed an enrichment analysis using the ChIP-Atlas database [32,33]. We tested candidate glyco-genes coding for the main glycosyltransferases—the *B4GALT1*, *ST6GAL1*, *MGAT3*, *MGAT4A*, *MGAT5*, *FUT3*, *FUT5*, *FUT6*, and *FUT8* genes—as well as the *FUK* and *GMDS* genes encoding fucokinase and GDP-mannose 4,6-dehydratase, respectively. The results of this analysis clearly indicate that the regulation takes place mainly in the liver (Table 1). Virtually no enrichment was found for the blood, neural, and muscle samples used as control tissues (these results are not shown in the table because there was no enrichment at all). There was some regulation by FOXA2 in the lung and HNF4A in the digestive tract (mostly in the CaCo-2, the epithelial human colon adenocarcinoma cell line).

[†] www.graphpad.com.

Table 1

Transcription factors HNF1A, HNF4A, and FOXA2 regulate glycosylation mainly in the liver.

			Liver			Lung (best score)			Digestive tract (best score)		
	Gene	ID (liver)	log *P* value	fold enrichment	*N* targets	log *P* value	fold enrichment	*N* targets	log *P* value	fold enrichment	*N* targets
GG only	*HNF1A*	SRX2636316	−1.6	3.40	4/10	—	—	—	—	—	—
	HNF1A	SRX10478540	−1.6	2.62	5/10	—	—	—	—	—	—
	HNF4A	SRX100505	−2.4	2.31	8/10	—	—	—	−3.1	8.77	4/10
	HNF4A	SRX10475577	−2.2	2.23	8/10	—	—	—	−2.9	7.69	4/10
	FOXA2	SRX100448	−1.6	1.83	8/10	−0.9	8.06	1/10	−0.4	1.53	3/10
	FOXA2	SRX10475774	−1.5	2.19	6/10	−0.9	2.31	3/10	—	—	—
TFs only	*HNF1A*	SRX2636317	−3.5	14.67	3/3	—	—	—	−3.3	54.43	2/3
	HNF1A	SRX2636316	−2.8	8.49	3/3	—	—	—	—	—	—
	HNF4A	SRX10829254	−2.6	7.14	3/3	—	—	—	−2.2	13.94	2/3
	HNF4A	SRX10829250 *	−2.4	19.14	2/3	—	—	—	−2.2	13.65	2/3
	FOXA2	SRX3321879 **	−2.1	4.98	3/3	−2.1	5.07	3/3	−3.5	14.84	3/3

	FOXA2	SRX10475774	−1.8	4.10	3/3	−2.0	4.57	3/3	−2.5	19.34	2/3
TFs+GG	*HNF1A*	SRX2636316	−3.5	4.58	7/13	—	—	—	−2.0	12.55	2/13
	HNF1A	SRX10478540	−2.3	2.83	7/13	—	—	—	—	—	—
	HNF4A	SRX100505	−3.5	2.45	11/13	—	—	—	−4.9	10.14	6/13
	HNF4A	SRX10475577	−3.4	2.36	11/13	—	—	—	−4.6	8.89	6/13
	FOXA2	SRX100448	−2.4	1.93	11/13	−1.5	2.34	6/13	−1.6	2.35	6/13
	FOXA2	SRX10475774	−2.4	2.31	12/13	−1.1	1.65	8/13	−1.3	3.42	3/13

We performed an enrichment analysis for transcription factors using the ChIP-Atlas [32], against all Refseq coding genes with a threshold significance of 50, to test the binding of HNF1A, HNF4A, and FOXA2 to 11 glyco-genes (*B4GALT1*, *ST6GAL1*, *MGAT3*, *MGAT4A*, *MGAT5*, *GMDS*, *FUT8*, *FUT3*, *FUT5*, *FUT6*, *FUK*) (GG only), to themselves (TFs only), and to all 14 genes (TFs+GG). The two most significant (by $\log_{10} P$ value) results were reported. While the liver showed a significant enrichment of the tested transcription factors in all combinations, the transcription factors were less important in the digestive system and relatively unimportant in other tissues, except for the FOXA2 in the lung. All datasets for the liver were from HepG2 cells unless indicated otherwise. Virtually no enrichment was found for the blood, neural, and muscle samples used as control tissues. ID: identification number; *: Hep3B, **: liver.

3.2. Individual manipulations of the HNF1A, HNF4A, or FOXA2 genes affect glyco-gene transcription and the composition of the total HepG2 cell N-glycome

Since HNF1A, HNF4A, and FOXA2 are master regulators of the gene transcriptional network in the human liver and our *in silico* analysis revealed that these transcription factors bind to the candidate glyco-genes, we wanted to validate this experimentally using CRISPR/dCas9 tools. If the expression of glyco-genes is regulated by HNF1A, HNF4A, and FOXA2, we assumed that there would be some effect on the total HepG2 cell protein *N*-glycosylation. First, the KRAB-dSpCas9 fusion construct was guided with specific gRNAs to the promoter region of each of these three genes in separate experiments in order to repress their transcription. A significant decrease in *HNF1A*, *HNF4A*, and *FOXA2* gene expression was detected on both the transcript and protein levels (Figs. 1–5). Significant changes in the transcription of several glyco-genes and in the cell glycome composition were found following manipulation of each of the three genes (Figs. 1–5). Surprisingly, in some cases, the changes in the messenger RNA (mRNA) levels of certain glyco-genes did not result in unambiguous changes in the corresponding glycans. The results are comprehensively presented in Table 2.

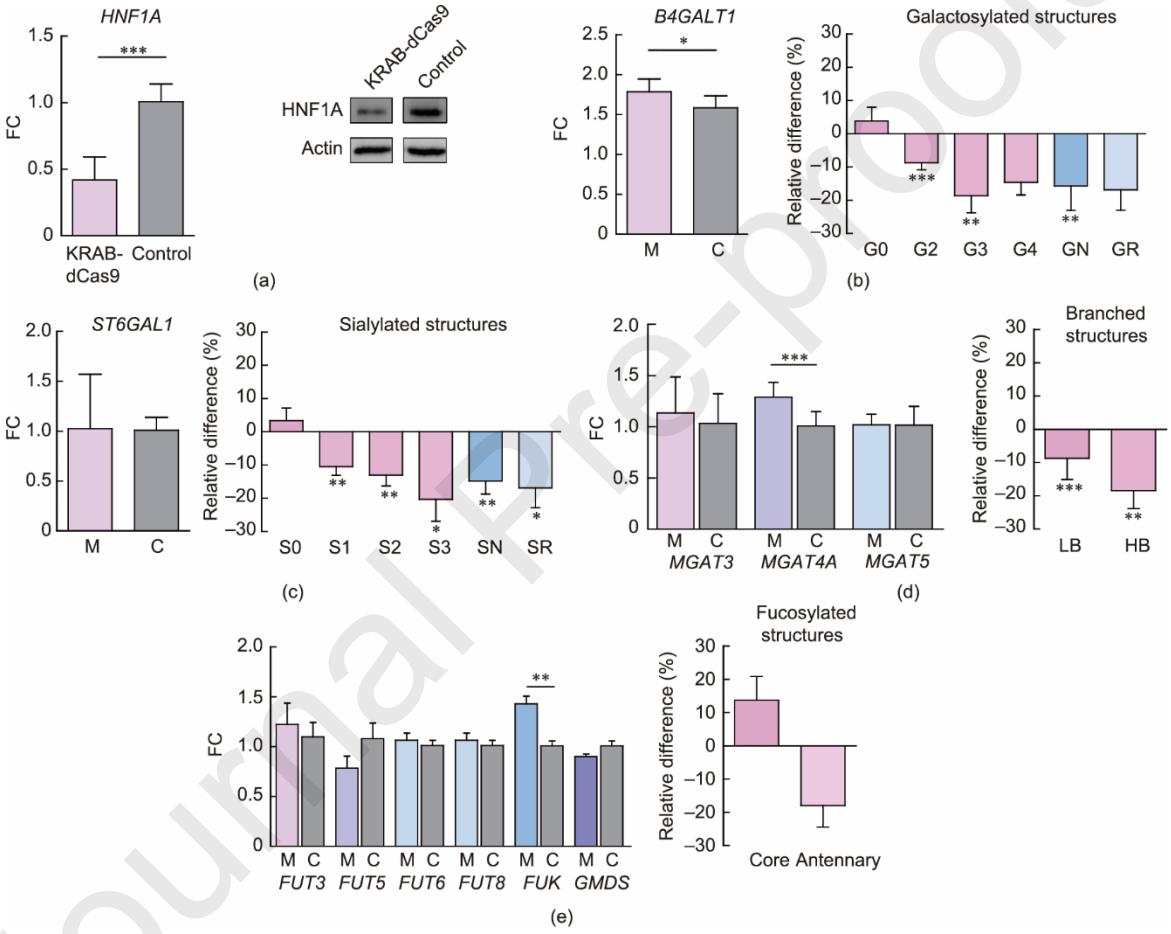


Fig. 1. Effects of HNF1A downregulation using KRAB-dSpCas9 on glyco-gene transcription and total *N*-glycome of HepG2 cells. (a) Following HNF1A downregulation its expression decreased on transcript and protein levels, (b) and this was reflected on increased level of *B4GALT1* transcription and subsequent changes of galactosylated glycan structures. (c) Changes in sialylated glycan structures (despite no significant change in *ST6GAL1* transcription). (d) *MGAT4A* expression was increased, however glycan branching was decreased. (e) No significant change in expression of analysed fucosyltransferases was observed, only a significant increase in FUK transcription, however it did not impact the fucosylated structures. Changes in total *N*-glycome composition are shown as relative difference compared to a control and expressed as percentage with the horizontal line positioned at 0 (indicating no change). Error bar is ± SD ($n = 9$, Mann–Whitney U test $*P < 0.05$, $**P < 0.01$, $***P < 0.001$). G0: agalactosylated glycans; G2–G4: di-, tri- and tetra-galactosylated glycans; GN: total galactosylated glycans; GR: galactosylated ratio; S0: asialylated glycans; S1–S3: mono-, di- and tri-sialylated glycans; SN: total sialylated glycans; SR: sialylated ratio; LB: low-branched glycans; HB: high-branched glycans; M: KRAB-dCas9 manipulation of *HNF1A*; C: control.

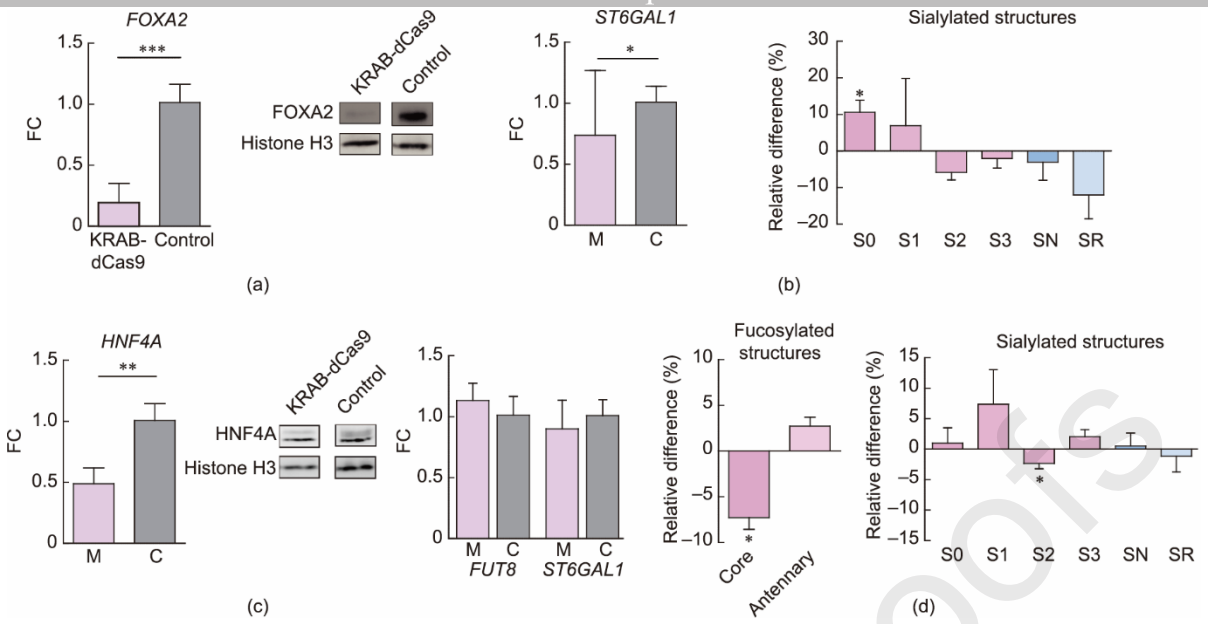


Fig. 2. Effects of downregulation of FOXA2 or HNF4A using KRAB-dSpCas9 on glyco-gene transcription and total *N*-glycome of HepG2 cells. (a) Following the *FOXA2* gene downregulation, (b) its expression decreased on transcript and protein levels and this affected transcription of *ST6GAL1* and resulted in an increase of S0. (c) Downregulation of the *HNF4A* gene resulted in decreased expression on transcript and protein levels and (d) only core fucosylated and S2 structures were significantly decreased, despite no significantly statistical in the *FUT8* and *ST6GAL1* gene transcription. Changes in total *N*-glycome composition are shown as relative difference compared to a control and expressed as percentage, with the horizontal line positioned at 0 (indicating no change). Error bar is ± SD (n = 9, Mann–Whitney U test $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$). M: KRAB-dCas9 manipulation of *FOXA2* or *HNF4A*.

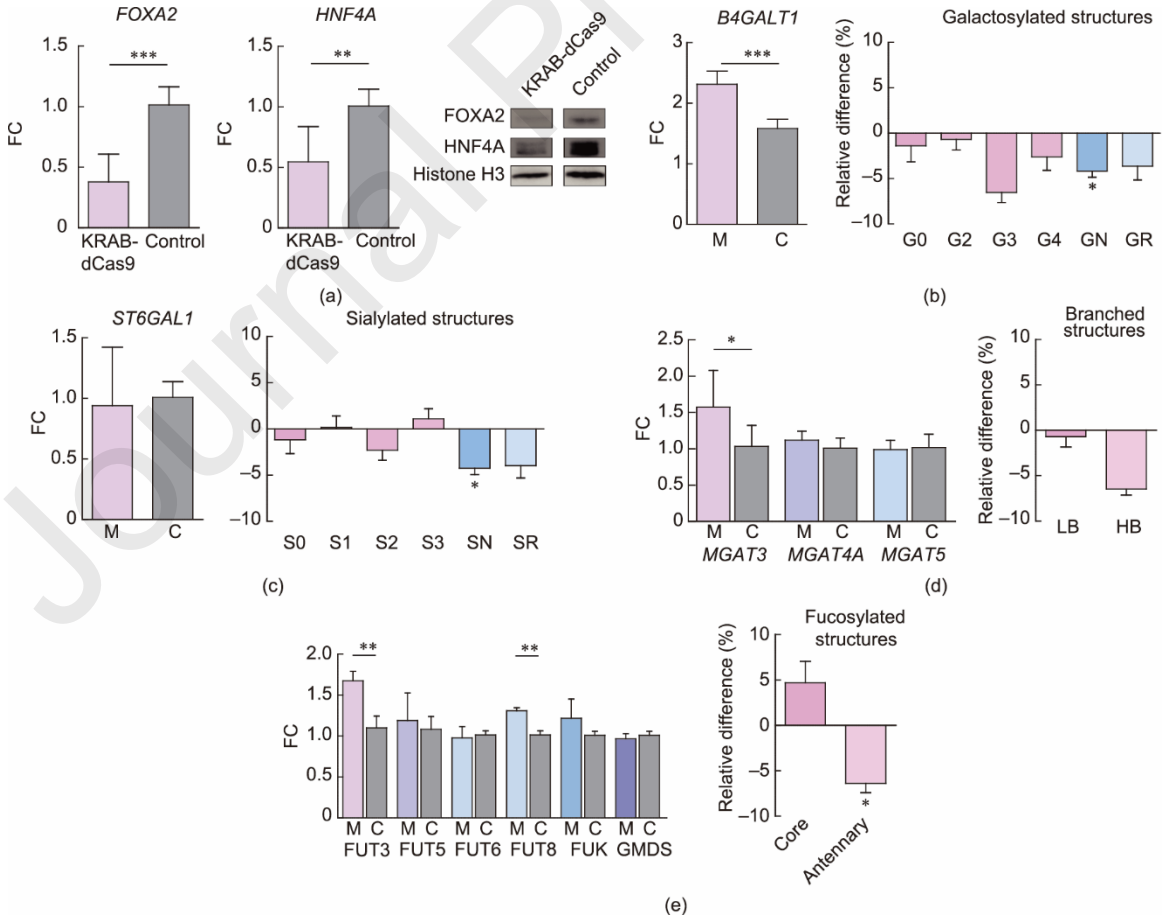


Fig. 3. Effects of simultaneous downregulation of HNF4A and FOXA2 using KRAB-dSpCas9 on glyco-gene transcription and total *N*-glycome of HepG2 cells. (a) Downregulation of HNF4A and FOXA2 decreased transcript and protein levels of both genes. (b) Following this manipulation changes

were observed in transcription level of *B4GALT1* with concomitant changes in GN structures. (c) *ST6GAL1* transcription level was no significant change while the level of SN structures increased. (d) *MGAT3* transcription was significantly increased but no significant change was observed in the quantity of branched glycan structures. (e) Transcription levels of *FUT3* and *FUT8* with concomitant changes in core fucosylated glycans. Changes in total *N*-glycome composition are shown as relative difference compared to a control and expressed as percentage, with the horizontal line positioned at 0 indicating no change. Error bar is ± SD (n = 9, Mann–Whitney U test *P < 0.05, **P < 0.01, ***P < 0.001). M: KRAB-dCas9 manipulation of *HNF4A/FOXA2* gene pair.

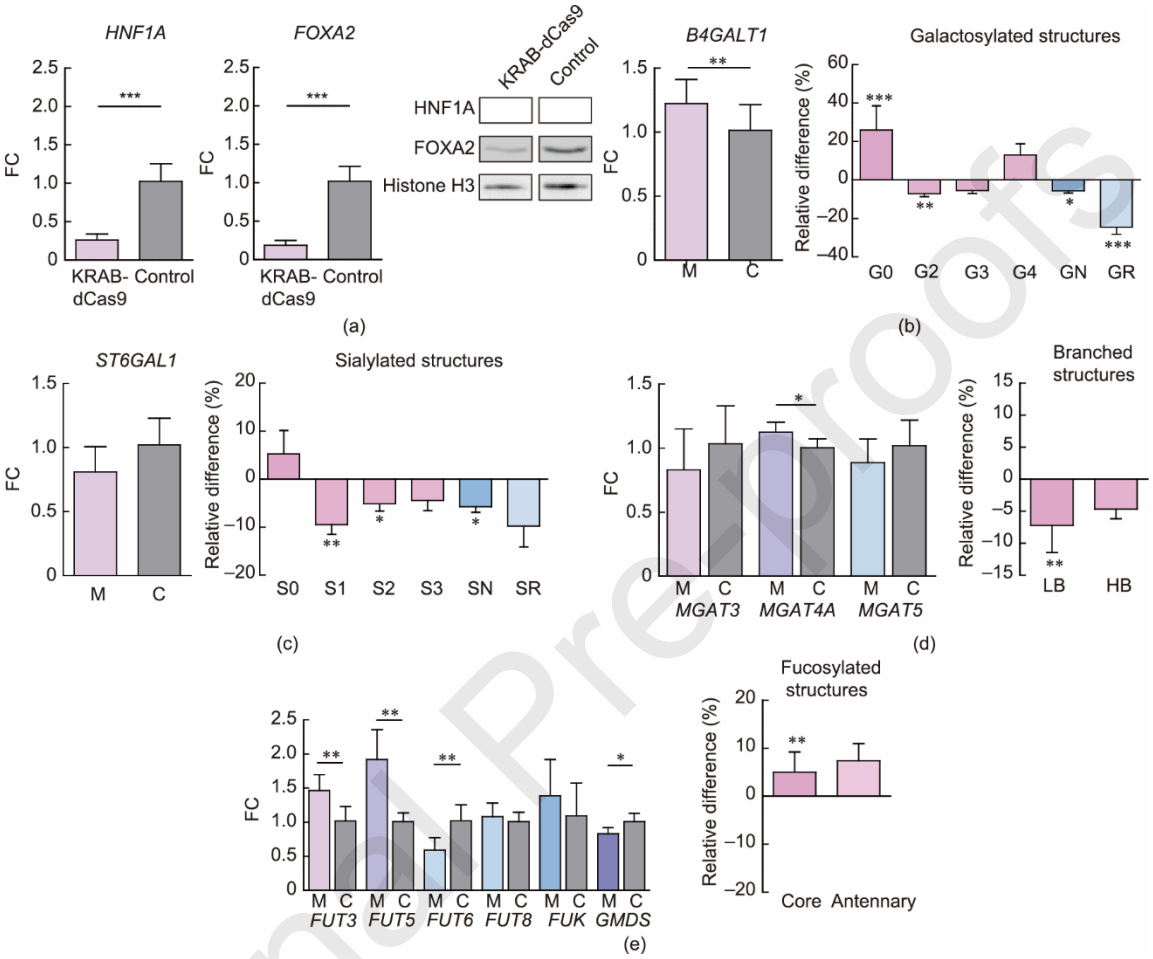


Fig. 4. Effects of simultaneous downregulation of HNF1A and FOXA2 using KRAB-dSpCas9 on glyco-gene transcription and total *N*-glycome of HepG2 cells. (a) Downregulation of HNF1A and FOXA2 decreased transcript and protein levels for both genes. Following this manipulation changes were observed in transcription level of (b) *B4GALT1* with concomitant changes in galactosylated structures, (c) *ST6GAL1* with concomitant changes in sialylated glycan structures, (d) *MGAT4A* with changes in LB, and (e) several *FUT* genes with concomitant changes in core fucosylated glycans. Changes in total *N*-glycome composition are shown as relative difference compared to a control and expressed as percentage, with the horizontal line positioned at 0 indicating no change. Error bar is ± SD (n = 8, Mann–Whitney U test *P < 0.05, **P < 0.01, ***P < 0.001). M: KRAB-dCas9 manipulation of *HNF1A/FOXA2* gene pair.

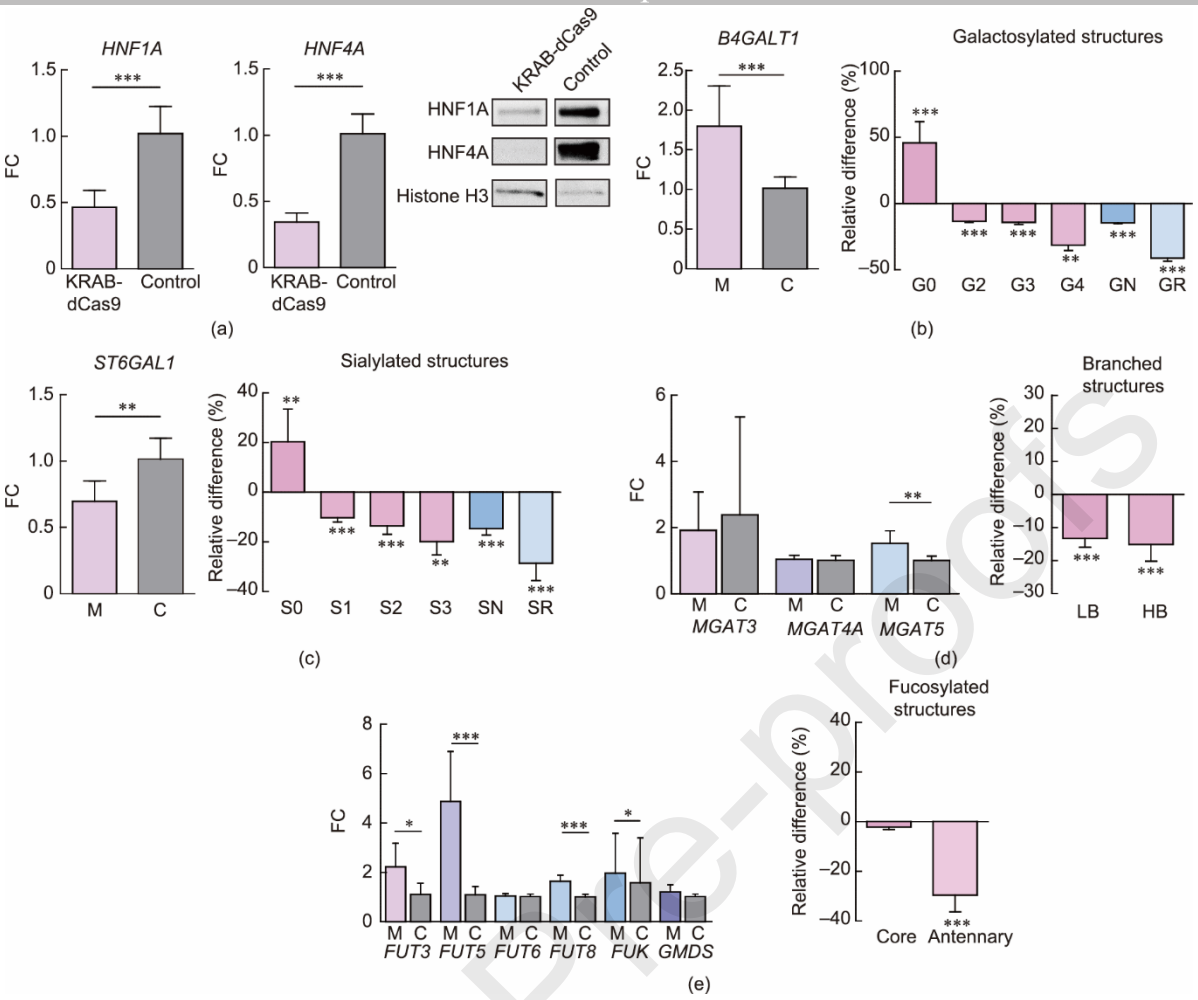


Fig. 5. Effects of simultaneous downregulation of HNF1A and HNF4A using KRAB-dSpCas9 on glyco-gene transcription and total *N*-glycome of HepG2 cells. (a) Downregulation of HNF1A and HNF4A resulted in a decrease of both transcripts and proteins. Following this manipulation changes were observed in transcription level of (b) *B4GALT1* with concomitant changes in all galactosylated structures, (c) *ST6GAL1* with concomitant changes in all sialylated glycan structures, (d) *MGAT5* with changes in both LB and HB, and (e) several *FUT* genes with concomitant changes in anntenary fucosylation. Changes in total *N*-glycome composition of HepG2 cells are shown as relative difference compared to a control and expressed as percentage, with the horizontal line positioned at 0 indicating no change. Error bar is ± SD ($n = 9$, Mann–Whitney U test $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$). M: KRAB-dCas9 manipulation of *HNF1A/HNF4A* gene pair.

Table 2

Changes in glyco-gene transcription following individual or simultaneous downregulation of HNF1A, HNF4A, and FOXA2 using KRAB-dCas9.

Silenced gene	Analyzed gene												
	FOXA2	*HNF1A*	*HNF4A*	*B4GALT1*	*ST6GAL1*	*FUT3*	*FUT5*	*FUT6*	*FUT8*	*FUK*	*GMDS*	*MGAT3*	*MGAT4A*	*MGAT5*
FOXA2	***	**	***		*			*	*				*	*
HNF1A	**	***		*						**			***	
HNF4A	**	**	***	**				*		*		**		
HNF1A/FOXA2	***	***	***	**		**	**	**			*		*	
HNF1A/HNF4A		***	***	***	**	*	***		***	*				**
HNF4A/FOXA2	***		**	***		**			**			*		

Red boxes indicate downregulation, while green boxes indicate upregulation of transcription. The majority of the analyzed glyco-genes changed transcriptional activity in the same direction in a single experiment, with the exception of *FUK* and *MGAT5*, which showed changes in transcription in opposite directions. *P< 0.05, **P< 0.01, ***P< 0.001.

We found that the downregulation of HNF1A, on both the transcript (FC = 0.42) and protein levels (Fig. 1(a), Table 2), resulted in a significant increase (FC = 1.79) in the transcription of *B4GALT1*, which encodes beta-1,4-galactosyltransferase, the enzyme responsible for the addition of galactose to *N*-acetylglucosamine residues. On the other hand, we observed a decrease in di-galactosylated (−8.73%), tri-galactosylated (−18.6%), and overall galactosylated structures (−14.59%) (Fig. 1(b)). Also, transcript levels increased for the *FUK* gene (FC = 1.43), which is a core component of the fucose-salvage pathway (Fig. 1(e)), even though the fucosylation remained unchanged. Despite an increase in the transcription of *MGAT4A* (FC = 1.29), which encodes a glycosyltransferase that regulates glycan branching, we did not detect any increase in high-branched glycan structures, but rather a decrease in both high-branched (−18.44%) and low-branched (−8.73%) complex glycans (Fig. 1(d)). At the glycome level, we also observed a decrease in mono-, di-, and tri-sialylated structures (−10.49%, −13.05%, and −20.30%, respectively) and total sialylated structures (−14.81%), even though *ST6GAL1* transcription did not significantly change (Fig. 1(c)). *FOXA2* gene transcription was decreased (FC = 0.73), while *HNF4A* transcription did not change significantly following HNF1A downregulation (Fig. S1(a) in Appendix A).

Downregulation of FOXA2 was seen on the transcript (FC = 0.19) and protein levels (Fig. 2(a), Table 2), and was followed by a change in the transcription level of both the *HNF1A* (FC = 0.54) and *HNF4A* (FC = 0.54) genes. FOXA2 downregulation affected the transcription of several glyco-genes as well (Fig. S1(b) in Appendix A). The *FUT6* gene transcription decreased (FC = 0.69), while that of *FUT8*, which is responsible for core fucosylation, increased (FC = 1.33). The transcription level of *MGAT4A* increased (FC = 1.26), while the transcription of *MGAT5* and *ST6GAL1* (sialyltransferase, which catalyzes the transfer of sialic acid from cytidine 5'-monophosphate (CMP)-sialic acid to galactose-containing substrates) decreased (FC = 0.76 and 0.74, respectively). At the glycome level, only asialylated glycan structures were increased (10.75%) in concordance with the decreased *ST6GAL1* expression, while quantities of other glycan structures were not significantly changed (Fig. 2(b)).

Downregulation of the HNF4A gene on the transcript (FC = 0.49) and protein level (Fig. 2(c), Table 2) resulted in decreased transcription of *HNF1A* (FC = 0.71) and *FOXA2* (FC = 0.61) (Fig. S1(c) in Appendix A). This manipulation also significantly affected the gene expression of several glyco-genes. The transcription level of *MGAT3* (encoding *N*-acetylglucosaminyltransferase-III) and *B4GALT1* increased (FC = 1.52 and 1.96, respectively), while the transcription of *FUT6* (FC = 0.79), which is responsible for antennary fucosylation, and *FUK* (FC = 0.84) decreased (Fig. S1(c) in Appendix A). Regarding the change in the total cell glycome composition, there were reduced quantities of glycan structures with core fucose (−7.28%) and di-sialylated structures (−2.33%), even though the corresponding genes did not change their expression (Fig. 2(d)). The data for individual peaks and derived traits is provided in Table S2 in Appendix A.

3.3. Simultaneous downregulation of the HNF4A and FOXA2 genes affects the transcription of B4GALT1, FUT3, FUT8, and MGAT3, which is not followed by unambiguous change in corresponding glycan structures

To further dissect the regulatory network between HNF1A, HNF4A, and FOXA2, we simultaneously manipulated HNF4A and FOXA2 with KRAB-dSpCas9 fusion, using specific gRNAs. This resulted in a significant downregulation of the mRNA levels of both *HNF4A* (FC = 0.55) and *FOXA2* (FC = 0.38); the effect was also visible on their protein levels (Fig. 3(a), Table 2). Simultaneous *HNF4A/FOXA2* gene manipulation resulted in a downstream increase in the mRNA levels of several glyco-genes—namely, *B4GALT1* (FC = 2.31), *FUT3* (FC = 1.67), *FUT8* (FC = 1.31), and *MGAT3* (FC = 1.57)—while *HNF1A* did not change expression significantly (Fig. S1(d) in Appendix A). At the *N*-glycome level, a decreased level of total galactosylated glycan structures (−3.99%) was observed, even though *B4GALT1* transcription increased (Fig. 3(b)). The quantity of total sialylated structures decreased (−4.165%), even though there was no significant change in *ST6GAL1* transcription (Fig. 3(c)). No significant changes in the quantities of low- and high-branched glycans were found (Fig. 3(d)). Antennary fucosylation was decreased (−6.39%), while the transcript level of the corresponding gene *FUT3* was increased (Fig. 3(e)). The data for individual peaks and derived traits is provided in Table S2 in Appendix A.

3.4. Simultaneous downregulation of the HNF1A and FOXA2 genes affects the transcription of HNF4A, B4GALT1, ST6GAL1, FUT3, FUT5, FUT6, MGAT4, and GMDS, which is not followed by unambiguous change in the corresponding glycan structures

Furthermore, KRAB-dSpCas9 fusion was simultaneously targeted to the promoter regions of *HNF1A* and *FOXA2* to induce their transcriptional repression. A significant decrease in transcript and protein levels was induced in both genes (FC = 0.26 and 0.19) (Fig. 4(a), Table 2). This manipulation resulted in significant transcription changes of several glyco-genes, as well as *HNF4A* (FC = 0.41, Fig. S1(e) in Appendix A). The transcript level of *B4GALT1* was significantly increased (FC = 1.22), while the quantity of the total galactosylated glycan structures decreased (−5.68%) and that of the agalactosylated glycans increased (26.04%) (Fig. 4(b)). The transcript level of *ST6GAL1* appeared to be slightly decreased, although the change was not statistically significant. Nevertheless, there was a significant decrease in total sialylated structures (−5.72%), while the quantity of asialylated glycan structures did not change significantly (Fig. 4(c)). Of the three genes involved in glycan branching, only *MGAT4A* transcription was increased (FC = 1.12), while no significant change in high-branched glycan structures was observed (Fig. 4(d)) and low-branched glycan structures decreased in quantity (−7.21%). Furthermore, the transcription of the glyco-genes involved in antennary fucosylation changed: the *FUT3* (FC = 1.46) and *FUT5* (FC = 1.92) transcript levels increased, whereas *FUT6* transcription decreased (FC = 0.59) (Fig. 4(e)). However, while there was no significant change in *FUT8* transcription, a significant increase in core fucosylated structures was observed (+5.04%) (Fig. 4(e)). The transcription of the *GMDS* gene, which encodes the enzyme that catalyzes the first step in the synthesis of GDP-fucose from GDP-mannose, was significantly decreased (FC = 0.83), while the transcription of *FUK* was not changed significantly (Fig. 4(e)).

3.5. Simultaneous downregulation of the HNF1A and HNF4A genes affects the transcription of most glyco-genes and is not followed by unambiguous changes in corresponding glycan structures

Following simultaneous targeting of the *HNF1A* and *HNF4A* promoters by the KRAB-dSpCas9 fusion construct using specific gRNAs, the expression of both genes decreased on the transcript (FC = 0.45 and FC = 0.34, respectively) and protein levels (Fig. 5(a), Table 2). We observed effects on the transcription level of most of the selected downstream genes (except *GMDS*, *FUT6*, *MGAT3*, and *MGAT4A)* (Fig. 5, Table 2), while silencing of this gene pair did not affect *FOXA2* transcription significantly (Fig. S1(f) in Appendix A). The transcription levels were increased for *B4GALT1* (FC = 1.73), *FUT3* (FC = 2.02), *FUT5* (FC = 4.5), *FUT8* (FC = 1.61), *MGAT5* (FC = 1.48), and *FUK* (FC = 1.6), while the *ST6GAL1* transcript level decreased (FC = 0.68) (Fig. 5, Table 2). A decrease in mono- (−10.22%), di- (−13.53%), and trisialylated (−19.84%) structures, as well as total sialylated structures, and an increase in asialylated glycan structures (+20.33%) aligned with the decreased *ST6GAL1* transcription (Fig. 5(c)). However, changes in other glycan traits did not show a positive correlation with the transcriptional changes of the corresponding glyco-genes. A decrease in total galactosylated glycans (−14.29%), accompanied by an increase in agalactosylated structures (+45.86%), was not in line with increased transcription of *B4GALT1* (Fig. 5(b)). The proportions of high- (−15.12%) and low-branched (−13.28%) structures were decreased, while the transcription of *MGAT5* was increased (Fig. 5(d)). Similarly, despite the increase in the transcript levels of *FUT3* and *FUT5*, we observed a decrease in structures with antennary fucose (−25.28%), while an increase in *FUT8* transcription had no significant effect on the amount of core-fucosylated glycans (Fig. 5(e)).

3.6. Overall complexity of the total HepG2 N-glycome increases following HNF1A, HNF4A, and FOXA2 downregulation, primarily due to the extension of biantennary glycans

Changes in individual glycan structures and narrowly defined glycan traits have a limited ability to capture changes at the level of the total cell *N*-glycome. Therefore, we defined the “glycan grade” metric to indicate the number of monosaccharide units in a glycan structure. Since we measured glycan structures synthesized after the oligomannose trimming step, the addition to each unit corresponds to one enzymatic reaction. At the level of the whole glycome, we found a consistent increase in the average glycan complexity by about one unit when the *HNF1A/FOXA2* and *HNF1A/HNF4A* gene pairs were silenced (Fig. 6). To gain a better insight into the changes in the total cell *N*-glycome level, we assessed the contribution of individual glycan structures to the overall

glycome complexity. The most important contribution to the complexity increase came from biantennary glycans, with a magnitude sufficient to compensate for a relative decrease in tri- or tetra-antennary structures (Fig. 7). This result complements and further explains the observed changes in glycosyltransferase expression in both experiments, indicating that the most significant contribution of the studied glyco-genes is their action on extending the biantennary structures, even though more complex structures appeared to be less abundant after the gene manipulations. This finding indicates that the increase in *B4GALT1* and *ST6GAL1* gene products primarily acted upon the biantennary structures.

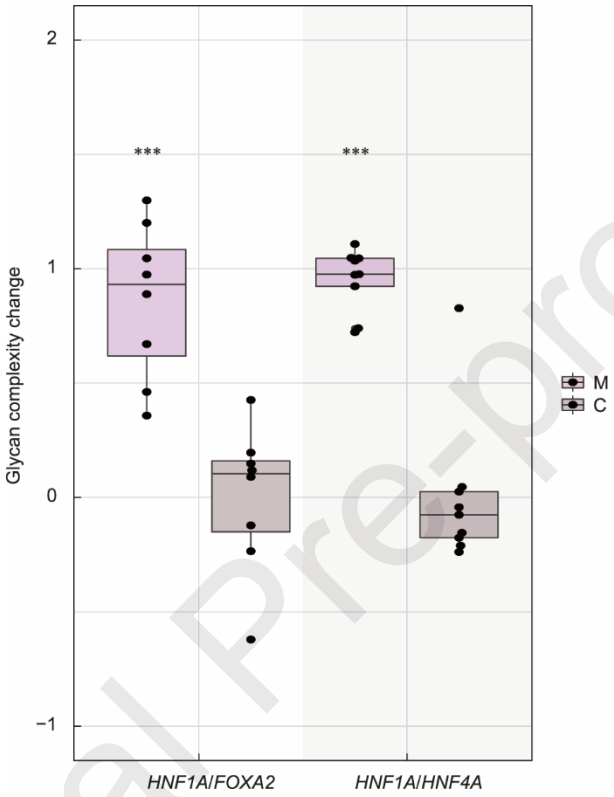


Fig. 6. Complexity of total cell *N*-glycome increases following simultaneous downregulation of *HNF1A/FOXA2* (left) and *HNF1A/HNF4A* (right). Each glycan unit (simple sugar monomer) is counted as one unit of complexity assuming that one enzymatic reaction adds one sugar unit. This assumption holds because the measured structures are on the biochemical pathway after the oligomannose trimming steps, essentially representing modifications of the core glycan structure. Compared to the matching negative control with non-targeting gRNA (C), manipulation using active KRAB-dSpCas9 with specific gRNAs (M) shows an increase of one sugar monomer on average. Average glycan complexity is calculated by multiplying the number of monomers in each glycan structure with its relative abundance. ***$P < 0.001$.

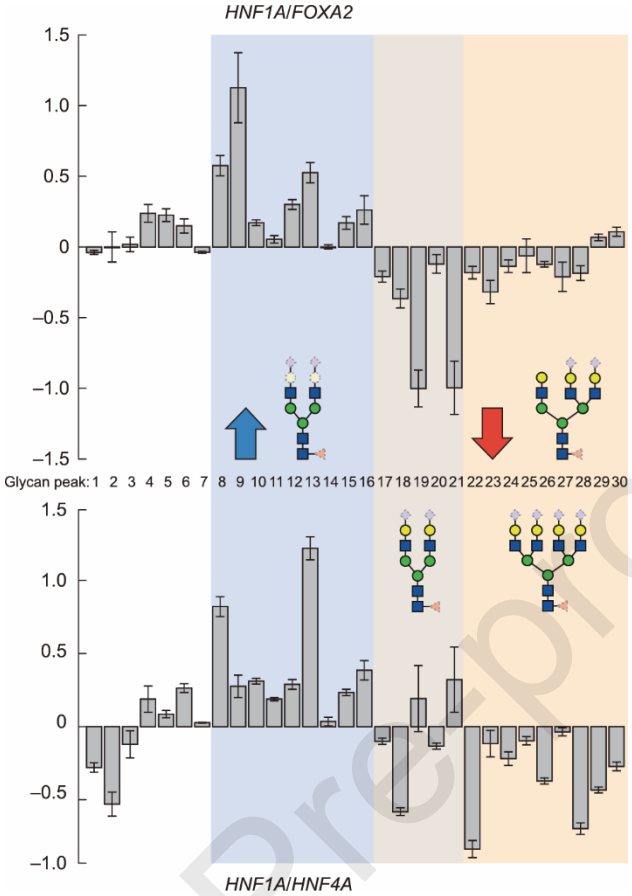


Fig. 7. Changes in the average glycan complexity (number of monomers) come from similar sources in experiments of simultaneous manipulations of *HNF1A/FOXA2* (top panel) and *HNF1A/HNF4A* (bottom panel). Although the total glycan complexity increases (Fig. 6), the increase comes mostly from the increase in the abundance of highly galactosylated, and in part sialylated biantennary glycans (shaded light blue, glycan peaks 8 to 16), while changes in galactosylated glycans are inconclusive (shaded light brown, glycan peaks 17 to 21). A general decrease in tri- and tetra-antennary structures (shaded light orange, glycan peaks 22 to 30) had less impact on the overall glycan complexity. This shows that the increase in galactosylated and sialylated structures is the main driver of the increase in glycan complexity. Scale on the left shows the absolute change in chromatographic peaks expressed as percentage. Composition of the glycan present in all peaks in a category is drawn in solid shapes, while monosaccharides present in some glycans, and absent in other glycans in a category are drawn with dotted lines. Error bars represent standard error.

4. Discussion

In this study, we used state-of-the-art CRISPR/dCas9-based molecular tools [8,30] to investigate whether the glyco-genes encoding for the main glycosyltransferases and corresponding glycans are regulated by the HNF1A, HNF4A, and FOXA2 transcription factors in HepG2 cells, a model for the human liver. The approach of downregulating the *HNF1A*, *HNF4A*, and *FOXA2* genes using the CRISPR/dCas9-KRAB molecular tool was chosen because HNF1A, HNF4A, and FOXA2 are the key transcription factors at the intersection of several key cellular pathways in the human liver [14,26]. Therefore, the total knock-down of these transcription factors was expected to be lethal or to reduce cell viability significantly, up to the point of interfering with the experimental goal of assessing the magnitude and direction of change in the *N*-glycosylation profile of HepG2 cells after changing the expression of the main glyco-genes, presumably regulated by these three transcription factors. We also focused on fine adjustments of the protein glycosylation process under physiological conditions. Therefore, the *HNF1A*, *HNF4A*, or *FOXA2* genes were downregulated individually or simultaneously in pairs (*HNF1A/HNF4A*, *HNF1A/FOXA2*, and *HNF4A/FOXA2*) using the KRAB-dCas9 fusion construct targeting their native promoters.

Downregulation of FOXA2 resulted in a decrease in the transcription of both the *HNF1A* and *HNF4A* genes, which confirms that FOXA2 is a pioneer transcription factor crucial for liver development [35]. Downregulation of HNF4A resulted in a decrease in *HNF1A* and *FOXA2* gene transcription; however, HNF1A downregulation resulted in decreased *FOXA2* but not *HNF4A* transcription. This result is in agreement with those of previous studies showing that HNF4A transcription factor is above HNF1A in the hierarchy of the regulatory network essential for hepatocytes. Previous studies have shown that HNF4A is required for *HNF1A* gene regulation; they have also elucidated the important role of HNF1A in the reciprocal regulation of HNF4A [36–38]. In conclusion, the strong effect of each of the transcription factors HNF1A, HNF4A, and FOXA2 on the expression of others from the group confirms the existence of a control regulatory loop, whose main factors have been identified while the exact mechanisms remain to be explained fully. It is interesting to note that simultaneous downregulation of HNF1A and HNF4A affected the transcription of most glyco-genes, compared with downregulation of the other two pairs, HNF4A/FOXA2 and HNF1A/FOXA2, which affected only some of the candidate genes. This result strongly confirms the already established role of HNF1A and HNF4A as the key regulators of hepatic transcription [26] and suggests that their regulatory loop controls protein glycosylation in the human liver (Fig. 8).

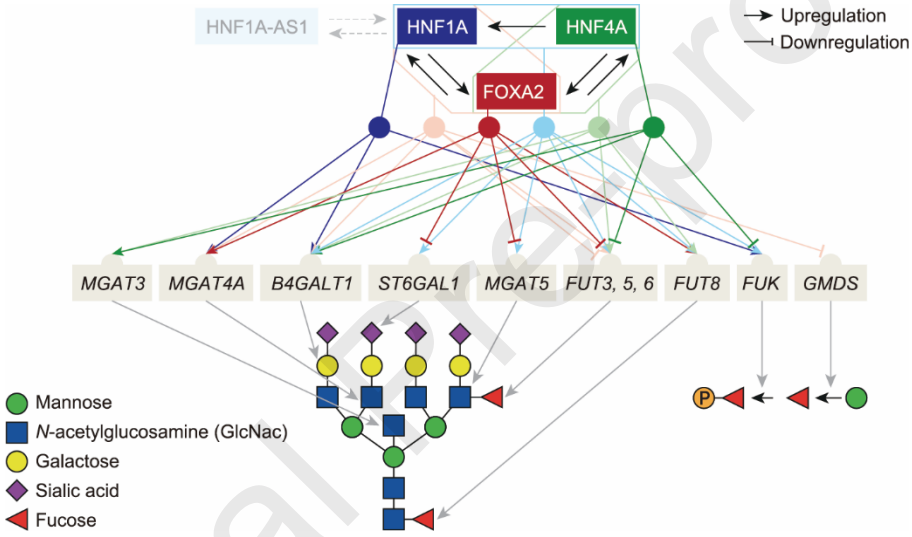


Fig. 8. Model of glyco-gene regulation by transcription factors HNF1A, HNF4A and FOXA2 in the liver. We tested the transcription factors both individually (dark colors) and in pairs (lighter colors, pairing indicated by polygons), and represented their interactions with glyco-genes based on expression profiling. We found feedback loops with mutual activation of HNF1A and FOXA2, and similarly HNF4A and FOXA2. The well-known regulatory feedback between HNF1A and its antisense RNA (HNF1A-AS1) is also depicted. Activation pattern for each combination is indicated in the upper part of the figure. The lower part indicates the reactions catalyzed by the enzymes encoded by glyco-genes. Arrows pointing toward a sugar unit indicate that the glycosyltransferase catalyzes addition of that type of subunit (*B4GALT1* and *ST6GAL1* can act at any branch) to the core glycan.

Our experiments on HNF1A, HNF4A, and FOXA2 individual and combinatorial downregulation using KRAB-dCas9 clearly demonstrate an effect of these transcription factors on downstream glyco-genes and the *N*-glycome of HepG2 cells. However, the gene transcriptional changes were not always positively correlated with the changes in corresponding glycan structures as one would expect. FOXA2 downregulation and simultaneous HNF1A/HNF4A downregulation decreased *ST6GAL1* transcription, and this was accompanied by a decrease in sialylated glycan structures (Figs. 2 and 5). Interestingly, following HNF1A downregulation, a significant decrease in mono-, di-, and tri-sialylated structures was found, regardless of the unchanged *ST6GAL1* transcription level. Downregulation of HNF1A alone or downregulation of the HNF1A/HNF4A and HNF4A/FOXA2 gene pairs resulted in a significant increase in the expression of *B4GALT1*, the main enzyme that adds galactose to *N*-linked glycans; at the same time, however, the levels of galactosylated glycans decreased. Furthermore, in all experiments involving HNF1A, HNF4A, and FOXA2 manipulation, a strong effect on the

transcriptional activity of several glycosyltransferases responsible for glycan branching was found. Thus, the relationship between the expression of glycosyltransferases and the composition of the HepG2 *N*-glycome is not straightforward. This type of discrepancy between the transcript level of a specific glyco-gene and its corresponding glycan structure has been reported previously [2,3]. For example, during the maturation of murine embrionic stem cells to embryonic endodermal (ExE) cells, the increase in the expression of *B3GNT1*, *B3GNT4*, and *B4GALT1* had no effect on the corresponding polylactosamine extension of the glycan structures. Moreover, the proportion of core fucosylated and bisected glycans in ExE cells increased even though the levels of *FUT8* and *MGAT3* were not elevated [2]. Recently, Nguyen and collaborators [39] reported that the overexpression of only 10 out of 42 analyzed glyco-genes had an impact on immunoglobulin G (IgG) glycosylation in Chinese hamster ovary cells.

Despite numerous research attempts, consistent data showing strong and straightforward effects of glycosyltransferase expression on the composition of the glycome is still missing and, under physiological conditions, changes in glycosyltransferase expression are often not reflected in the observed changes in glycome composition [3]. This is actually unsurprising when considering that the same glycosyltransferase produces glycans attached to hundreds of different proteins in a cell. Furthermore, if the expression of glycosyltransferases were the key regulatory mechanism, then thousands of proteins would be regulated in the same way, which is not biologically plausible. Indeed, a series of genome-wide association studies (GWAS) have revealed that glycosyltransferases make up only a small fraction of the gene network associated with *N*-glycome composition [28,40–43]. Recent parallel GWAS of IgG and transferrin clearly showed that very similar glycans, attached to these two different proteins, are regulated by different gene networks despite the fact that the glycans are generated by the same glycosyltransferases [44]. The results obtained in this study strengthen previous findings and strongly suggest that the expression of glycosyltransferases is not the main regulatory mechanism leading to the final glycosylation phenotype.

Even though it would be interesting to identify some key hepatic proteins and their glycosylation to gain more biological information about regulation by HNF1A, HNF4A, and FOXA2 in the human liver, glycopeptide analysis is less quantitative than the UPLC analysis used in this work, which quantifies glycans based on the presence of the introduced fluorescent tag. In this study, we quantified relatively small effects on the total cell *N*-glycome, which is not possible at the glycoprotein level due to technical limitations and greater measurement error. Furthermore, we analyzed general effects on the glycosylation machinery, and it has been clearly demonstrated in a series of previous GWA studies (of plasma or the IgG glycome) that the effects are much more visible when integrated across different proteins and/or glycosylation traits [40,41,43–45].

Here, we propose an alternative way to evaluate the *N*-glycome composition by estimating the overall complexity of the glycome via quantifying the number of individual monomeric building blocks in each of the glycan structures. When analyzed in this way, the silencing of pairs of the key transcription factors HNF1A/HNF4A and HNF1A/FOXA2 has very strong and consistent effects on the complexity of the total HepG2 cell *N*-glycome. The results confirmed our initial line of reasoning that the combinations of transcription factors have a synergistic effect and act together (which was especially seen for the pair HNF1A/HNF4A). Based on our transcription analysis, we also developed a model of glyco-gene regulation by these three liver transcription factors (Fig. 8). Our data supports the existence of positive feedback loops between HNF1A/FOXA2 and HNF4A/FOXA2, which reinforce each other's expression. The densely connected network of transcriptional interactions confirms that the three transcription factors profoundly influence protein glycosylation in HepG2 cells. However, we have observed that changes in glycome composition do not always correlate with changes in the transcriptional expression of relevant glyco-genes, which suggests the presence of an additional regulatory layer on top of glycosyltransferase expression. This is intuitive, since harmonized changes in the structures of thousands of proteins are not biologically meaningful; it is also in accordance with the very limited associations between glycome composition and glycosyltransferase gene expression observed in different studies [2,3]. Furthermore, regulation could also occur at a post-transcriptional level or by mechanisms unrelated to glyco-gene transcription (i.e., by microRNAs) [4,5]. Still, we were able to confirm the key effect of upstream regulatory factors on protein glycosylation in HepG2 and explain most of the observed changes in the glycosylation pattern.

Previous cohort studies have already provided evidence of the association between glycome composition and HNF1A activity [29,46]. We have shown that increased methylation of the HNF1A promoter region positively correlates with an increase in tetra-antennary glycan structures in the plasma *N*-glycome [29], and that the four CpG sites in the first exon of the HNF1A gene region are functionally relevant for its transcription [30]. HNF1A has been identified as one of the GWAS hits for the human plasma *N*-glycome [28], and mutations in this gene substantially alter the plasma *N*-glycome [47], all of which suggest the important role of HNF1A in the regulation of human plasma protein *N*-glycosylation. Therefore, while pathway enrichment analysis is a powerful tool to identify the roles of genes and their products in cellular metabolism, we focused on an approach starting with candidate genes identified in our previous GWA studies, where the putative pathways already represented a hypothesis that we then further tested by specifically targeting the candidate genes. To the best of our knowledge, this work is the first study to use CRISPR/dCas9-based molecular tools to directly manipulate the *HNF1A*, *HNF4A*, and *FOXA2* genes and to explore both glyco-gene transcriptional expression and expression of the final products—glycan structures in the total *N*-glycome within the hepatic model cell line HepG2. We have demonstrated by this study and our previous studies that the CRISPR/dCas9 method can be useful for the functional analysis of candidate genes when exploring the complex pathway of glycosylation in health and in disease [48–50]. Since glycome complexity is also implicated in the development of type 2 diabetes, where changes in protein glycosylation occur at least 7 years before the diagnosis of insulin resistance [51], an understanding of the molecular mechanisms linking HNF1A to protein glycosylation [29] may help in the identification of potential new drug targets for diabetes and related metabolic diseases.

Acknowledgments

This work was supported by the European Structural and Investment Funded Grant “CardioMetabolic” (#KK.01.2.1.02.0321), Croatian National Centre of Research Excellence in Personalized Healthcare Grant (#KK.01.1.1.01.0010), and European Regional Development Fund Grant, project “CRISPR/Cas9-CasMouse” (#KK.01.1.1.04.0085). Glycosylation analysis was performed in the Genos Glycoscience Research Laboratory and was supported by the European Structural and Investment Funded Project of Centre of Competence in Molecular Diagnostics (#KK.01.2.2.03.0006) and Croatian National Centre of Research Excellence in Personalized Healthcare Grant (#KK.01.1.1.01.0010).

Compliance with ethics guidelines

Vedrana Vičić Bočkor, Nika Foglar, Goran Josipović, Marija Klasić, Ana Vujić, Branimir Plavša, Toma Keser, Samira Smajlović, Aleksandar Vojta, and Vlatka Zoldoš declare that they have no conflict of interest or financial conflicts to disclose.

Appendix A. Supplementary data

Supplementary data to this acticle can be found online at

References

[1] Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al, editors. Essentials of Glycobiology. 3rd ed. New York: Cold Spring Harbor Laboratory Press; 2015.

[2] Nairn AV, Aoki K, dela Rosa M, Porterfield M, Lim JM, Kulik M, et al. Regulation of glycan structures in murine embryonic stem cells: combined transcript profiling of glycan-related genes and glycan structural analysis. J Biol Chem 2012;287(45):37835–56.

[3] Nairn AV, York WS, Harris K, Hall EM, Pierce JM, Moremen KW. Regulation of glycan structures in animal tissues: transcript profiling of glycan-related genes. J Biol Chem 2008;283(25):17298–313.

[4] Thu CT, Mahal LK. Sweet control: microRNA regulation of the glycome. Biochemistry 2020;59(34):3098–110.

[5] Agrawal P, Kurcon T, Pilobello KT, Rakus JF, Koppolu S, Liu Z, et al. Mapping posttranscriptional regulation of the human glycome uncovers microRNA defining the glycocode. Proc Natl Acad Sci USA 2014;111(11):4338–43.

[6] Neelamegham S, Mahal LK. Multi-level regulation of cellular glycosylation: from genes to transcript to enzyme to structure. Curr Opin Struct Biol 2016;40:145–52.

[7] Klasić M, Krištić J, Korać P, Horvat T, Markulin D, Vojta A, et al. DNA hypomethylation upregulates expression of the MGAT3 gene in HepG2 cells and leads to changes in N-glycosylation of secreted glycoproteins. Sci Rep 2016;6:24363.

[8] Vojta A, Samaržija I, Bočkor L, Zoldoš V. Glyco-genes change expression in cancer through aberrant methylation. Biochim Biophys Acta Gen Subj 2016;1860(8):1776–85.

[9] Klasić M, Markulin D, Vojta A, Samaržija I, Biruš I, Dobrinić P, et al. Promoter methylation of the MGAT3 and BACH2 genes correlates with the composition of the immunoglobulin G glycome in inflammatory bowel disease. Clin Epigenetics 2018;10:75.

[10] Reily C, Stewart TJ, Renfrow MB, Novak J. Glycosylation in health and disease. Nat Rev Nephrol 2019;15(6):346–66.

[11] Stowell SR, Ju T, Cummings RD. Protein glycosylation in cancer. Annu Rev Pathol Mech Dis 2015;10:473–510.

[12] Costa RH, Kalinichenko VV, Holterman AXL, Wang X. Transcription factors in liver development, differentiation, and regeneration. Hepatology 2003;38(6):1331–47.

[13] Lee CS, Sund NJ, Behr R, Herrera PL, Kaestner KH. Foxa2 is required for the differentiation of pancreatic α-cells. Dev Biol 2005;278(2):484–95.

[14] Lau HH, Ng NHJ, Loo LSW, Jasmen JB, Teo AKK. The molecular functions of hepatocyte nuclear factors – in and beyond the liver. J Hepatol 2018;68(5):1033–48.

[15] Inoue Y, Hayhurst GP, Inoue J, Mori M, Gonzalez FJ. Defective ureagenesis in mice carrying a liver-specific disruption of hepatocyte nuclear factor 4α (HNF4α): HNF4α regulates ornithine transcarbamylase in vivo. J Biol Chem 2002;277(28):25257–65.

[16] Inoue Y, Peters LL, Yim SH, Inoue J, Gonzalez FJ. Role of hepatocyte nuclear factor 4α in control of blood coagulation factor gene expression. J Mol Med 2006;84(4):334–44.

[17] Inoue Y, Yu AM, Inoue J, Gonzalez FJ. Hepatocyte nuclear factor 4α is a central regulator of bile acid conjugation. J Biol Chem 2004;279(4):2480–9.

[18] Kamiyama Y, Matsubara T, Yoshinari K, Nagata K, Kamimura H, Yamazoe Y. Role of human hepatocyte nuclear factor 4α in the expression of drug-metabolizing enzymes and transporters in human hepatocytes assessed by use of small interfering RNA. Drug Metab Pharmacokinet 2007;22(4):287–98.

[19] Walesky C, Apte U. Role of hepatocyte nuclear factor 4α (HNF4α) in cell proliferation and cancer. Gene Expr 2015;16(3):101–8.

[20] Hoskins JW, Jia J, Flandez M, Parikh H, Xiao W, Collins I, et al. Transcriptome analysis of pancreatic cancer reveals a tumor suppressor function for HNF1A. Carcinogenesis 2014;35(12):2670–8.

[21] Pelletier L, Rebouissou S, Paris A, Rathahao-Paris E, Perdu E, Bioulac-Sage P, et al. Loss of hepatocyte nuclear factor 1α function in human hepatocellular adenomas leads to aberrant activation of signaling pathways involved in tumorigenesis. Hepatology 2010;51(2):557–66.

[22] Luo Z, Li Y, Wang H, Fleming J, Li M, Kang Y, et al. Hepatocyte nuclear factor 1A (HNF1A) as a possible tumor suppressor in pancreatic cancer. PLoS One 2015;10(3):e0121082.

[23] Teeli AS, Łuczyńska K, Haque E, Gayas MA, Winiarczyk D, Taniguchi H. Disruption of tumor suppressors HNF4α /HNF1α causes tumorigenesis in liver. Cancers 2021;13(21):5357.

[24] Bluteau O, Jeannot E, Bioulac-Sage P, Marqués JM, Blanc JF, Bui H, et al. Bi-allelic inactivation of TCF1 in hepatic adenomas. Nat Genet 2002;32(2):312–5.

[25] Duncan SA, Navas MA, Dufort D, Rossant J, Stoffelt M. Regulation of a transcription factor network required for differentiation and metabolism. Science 1998;281(5377):692–5.

[26] Odom DT, Zizlsperger N, Gordon DB, Bell GW, Rinaldi NJ, Murray HL, et al. Control of pancreas and liver gene expression by HNF transcription factors. Science 2004;303(5662):1378–81.

[27] Li Z, Tuteja G, Schug J, Kaestner KH. Foxa1 and Foxa2 are essential for sexual dimorphism in liver cancer. Cell 2012;148(1-2):72–83.

[28] Lauc G, Essafi A, Huffman JE, Hayward C, Knežević A, Kattla JJ, et al. Genomics meets glycomics-the first GWAS study of human N-glycome identifies HNF1α as a master regulator of plasma protein fucosylation. PLoS Genet 2010;6(12):e1001256.

[29] Zoldoš V, Horvat T, Novokmet M, Cuenin C, Mužinić A, Pučić M, et al. Epigenetic silencing of HNF1A associates with changes in the composition of the human plasma N-glycome. Epigenetics 2012;7(2):164–72.

[30] Josipović G, Tadić V, Klasić M, Zanki V, Bečeheli I, Chung F, et al. Antagonistic and synergistic epigenetic modulation using orthologous CRISPR/dCas9-based modular system. Nucleic Acids Res 2019;47(18):9637–57.

[31] Tyakht AV, Ilina EN, Alexeev DG, Ischenko DS, Gorbachev AY, Semashko TA, et al. RNA-Seq gene expression profiling of HepG2 cells: the influence of experimental factors and comparison with liver tissue. BMC Genomics 2014;15(1):1108.

[32] Oki S, Ohta T, Shioi G, Hatanaka H, Ogasawara O, Okuda Y, et al. ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. EMBO Rep 2018;19(12):e46255.

[33] Zou Z, Ohta T, Miura F, Oki S. ChIP-Atlas 2021 update: a data-mining suite for exploring epigenomic landscapes by fully integrating ChIP-seq, ATAC-seq and Bisulfite-seq data. Nucleic Acids Res 2022;50(W1):W175–82.

[34] Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative CT method. Nat Protoc 2008;3(6):1101–8.

[35] Berasain C, Arechederra M, Argemí J, Fernández-Barrena MG, Avila MA. Loss of liver function in chronic liver disease: an identity crisis. J Hepatol 2023;78(2):401–14.

[36] Kuo CJ, Conley PB, Chen L, Sladek FM, Darnell JE Jr, Crabtree GR. A transcriptional hierarchy involved in mammalian cell-type specification. Nature 1992;355(6359):457–61.

[37] Tian JM, Schibler U. Tissue-specific expression of the gene encoding hepatocyte nuclear factor 1 may involve hepatocyte nuclear factor 4. Genes Dev 1991;5(12A):2225–34.

[38] Zhong W, Mirkovitch J, Darnell JE Jr. Tissue-specific regulation of mouse hepatocyte nuclear factor 4 expression. Mol Cell Biol 1994;14(11):7276–84.

[39] Nguyen NTB, Lin J, Tay SJ, Mariati, Yeo J, Nguyen-Khuong T, et al. Multiplexed engineering glycosyltransferase genes in CHO cells via targeted integration for producing antibodies with diverse complex-type N-glycans. Sci Rep 2021;11(1):12969.

[40] Wahl A, van den Akker E, Klaric L, Štambuk J, Benedetti E, Plomp R, et al. Genome-wide association study on immunoglobulin G glycosylation patterns. Front Immunol 2018;9:277.

[41] Lauc G, Huffman JE, Pučić M, Zgaga L, Adamczyk B, Mužinić A, et al. Loci associated with N-glycosylation of human immunoglobulin G show pleiotropy with autoimmune diseases and haematological cancers. PLoS Genet 2013;9(1):e1003225.

[42] Shadrina AS, Zlobin AS, Zaytseva OO, Klarić L, Sharapov SZ, Pakhomov ED, et al. Multivariate genome-wide analysis of immunoglobulin G N-glycosylation identifies new loci pleiotropic with immune function. Hum Mol Genet 2021;30(13):1259–70.

[43] Klarić L, Tsepilov YA, Stanton CM, Mangino M, Sikka TT, Esko T, et al. Glycosylation of immunoglobulin G is regulated by a large network of genes pleiotropic with inflammatory diseases. Sci Adv 2020;6(8):eaax0301.

[44] Landini A, Trbojević-Akmačić I, Navarro P, Tsepilov YA, Sharapov SZ, Vučković F, et al. Genetic regulation of post-translational modification of two distinct proteins. Nat Commun 2022;13(1):1586.

[45] Trbojević-Akmačić I, Lageveen-Kammeijer GSM, Heijs B, Petrović T, Deriš H, Wuhrer M, et al. High-throughput glycomic methods. Chem Rev 2022;122(20):15865–913.

[46] Lauc G, Huffman JE, Pučić M, Zgaga L, Adamczyk B, Mužinić A, et al. Loci Associated with N-glycosylation of human immunoglobulin G show pleiotropy with autoimmune diseases and haematological cancers. PLoS Genet 2013;9(1):e1003225.

[47] Thanabalasingham G, Huffman JE, Kattla JJ, Novokmet M, Rudan I, Gloyn AL, et al. Mutations in HNF1A result in marked alterations of plasma glycan profile. Diabetes 2013;62(4):1329–37.

[48] Mijakovac A, Miškec K, Krištić J, Bočkor VV, Tadić V, Bošković M, et al. A transient expression system with stably integrated CRISPR-dCas9 fusions for regulation of genes involved in immunoglobulin G glycosylation. CRISPR J 2022;5(2):237–53.

[49] Mijakovac A, Jurić J, Kohrt WM, Krištić J, Kifer D, Gavin KM, et al. Effects of estradiol on immunoglobulin G glycosylation: mapping of the downstream signaling mechanism. Front Immunol 2021;12:680227.

[50] Frkatović-Hodžić A, Miškec K, Mijakovac A, Nostaeva A, Sharapov SZ, Landini A, et al. Mapping of the gene network that regulates glycan clock of ageing. medRxiv. In press.

[51] Cvetko A, Mangino M, Tijardović M, Kifer D, Falchi M, Keser T, et al. Plasma N-glycome shows continuous deterioration as the diagnosis of insulin resistance approaches. BMJ Open Diabetes Res Care 2021;9(1):e002263.