nature methods

RESOURCE

https://doi.org/10.1038/s41592-018-0086-z

The GAGOme: a cell-based library of displayed glycosaminoglycans

Yen-Hsi Chen[1], Yoshiki Narimatsu[1], Thomas M. Clausen[2], Catarina Gomes[1,3], Richard Karlsson[1], Catharina Steentoft[1], Charlotte B. Spliid[2], Tobias Gustavsson[2], Ali Salanti[2], Andrea Persson[4], Anders Malmström[4], Daniel Willén[5], Ulf Ellervik[5], Eric P. Bennett[1], Yang Mao[1,6*], Henrik Clausen[1*] and Zhang Yang[1*]

Glycosaminoglycans (GAGs) are essential polysaccharides in normal physiology and disease. However, understanding of the contribution of specific GAG structures to specific biological functions is limited, largely because of the great structural heterogeneity among GAGs themselves, as well as technical limitations in the structural characterization and chemical synthesis of GAGs. Here we describe a cell-based method to produce and display distinct GAGs with a broad repertoire of modifications, a library we refer to as the GAGOme. By using precise gene editing, we engineered a large panel of Chinese hamster ovary cells with knockout or knock-in of the genes encoding most of the enzymes involved in GAG biosynthesis, to generate a library of isogenic cell lines that differentially display distinct GAG features. We show that this library can be used for cell-based binding assays, recombinant expression of proteoglycans with distinct GAG structures, and production of distinct GAG chains on metabolic primers that may be used for the assembly of GAG glycan microarrays.

Glycosaminoglycans are linear polysaccharides that are found on proteoglycans on the cell surface and in the extracellular matrix. They serve a wide range of biological functions in development, homeostasis, and disease, including cell signaling, cell proliferation, tissue morphogenesis, and interactions with various growth factors, cytokines, chemokines, and pathogens[1–3]. More than 40 glycosyltransferases, epimerases, sulfotransferases, and other enzymes orchestrate the assembly of GAGs containing chondroitin sulfate (CS), dermatan sulfate (DS), and/or heparan sulfate (HS). Many of these are homologous isoenzymes with seemingly similar functions; however, they are predicted to serve distinct, important biological functions on the basis of studies with mouse knockout models, as well as congenital diseases associated with most of the genes[4]. Although it is believed that distinct structural motifs of GAGs, such as sulfation and epimerization patterns, serve as specific recognition motifs for biological interactions[5], only a few defined motifs are known today[1,6]. This is due partly to obstacles in the structural analysis of GAG chains, which is generally limited to analysis of disaccharides generated by enzymatic digestion of GAG chains without global sequence information, and partly to the lack of extended GAG chains with defined structures. Appreciation of the distinct structural features of GAGs in specific interactions requires improved technologies for discovery and characterization. Furthermore, although great progress has been made in the chemical and chemoenzymatic synthesis of short GAG oligosaccharides and mimetics, the availability of comprehensive libraries of GAG structures for the analysis of interactions with GAGs is currently limited.

One appealing way to overcome these technical constraints is to use the cellular capacity for GAG biosynthesis, with methods that enable investigators to produce and display GAGs and interrogate the detailed structural features required for specific bioactivities. Here we present a systematic genetic deconstruction and reconstruction of GAG biosynthesis in a mammalian cell line that provides a large library of isogenic cells with distinct CS/DS and HS GAG biosynthetic capacities (Fig. 1). We used a rational combinatorial knockout (KO) and knock-in (KI) stable glycoengineering strategy and demonstrate that the generated cell library can be used for dissection of the fine binding specificities of pathogen adhesions, growth factors, and antibodies. Moreover, we show that the library is useful for the production of recombinant proteoglycans with distinct GAG structures and GAG chains on primers suitable for the development of GAG microarrays. Once fully built, the cell-based GAGOme will be a sustainable resource for diverse applications and methodologies to explore the biology of GAGs.

Results

The GAG glycosylation capacities of CHO cells. The Chinese hamster ovary (CHO) cell line produces both CS/DS and HS. Previously developed mutant CHO cell lines have been instrumental in demonstrating the binding specificities of proteins to GAGs[7,8]. CHO cells are also used for the cellular synthesis of GAG chains on β-xylosides[9]. We therefore chose CHO cells as a model for the development of a cell-based GAG array. We first identified the expressed repertoire of genes encoding enzymes involved in GAG biosynthesis and encoding proteoglycans in wild-type (WT) CHO cells by quantitative RNA-seq (Supplementary Fig. 1). Next we confirmed that the GAG composition in cell lysates from WT CHO cells determined by high-performance liquid chromatography (HPLC)-based

[1]Center for Glycomics, Departments of Cellular and Molecular Medicine and School of Dentistry, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark. [2]Centre for Medical Parasitology at the Institute for Immunology and Microbiology, University of Copenhagen, Copenhagen, Denmark. [3]i3S–Institute for Research and Innovation in Health, Institute of Molecular Pathology and Immunology (IPATIMUP), University of Porto, Porto, Portugal. [4]Department of Experimental Medical Science, Lund University, Lund, Sweden. [5]Centre for Analysis and Synthesis, Centre for Chemistry and Chemical Engineering, Lund University, Lund, Sweden. [6]Present address: School of Pharmaceutical Sciences, Sun Yat-sen University, Guangzhou, China. *e-mail: maoyang3@mail.sysu.edu.cn; hclau@sund.ku.dk; yang@sund.ku.dk

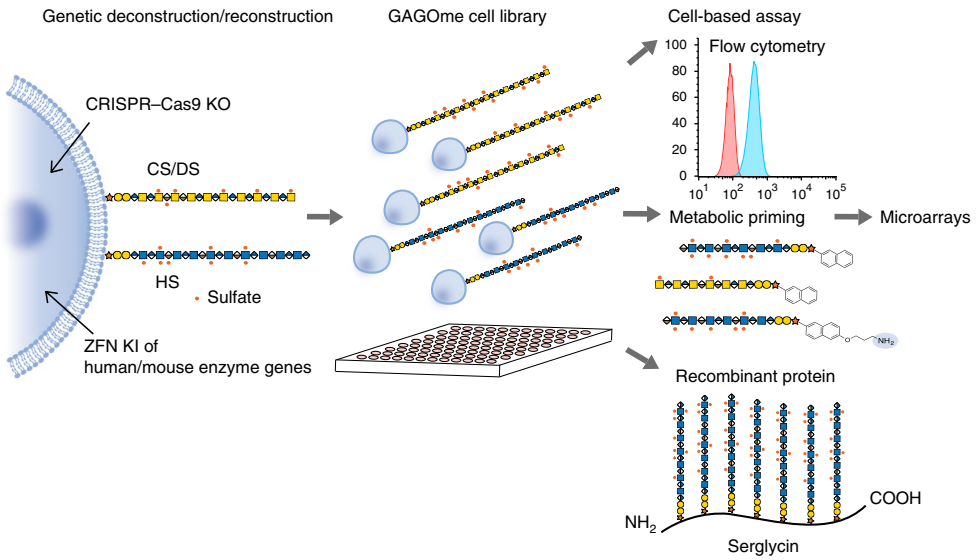

Fig. 1 | Graphic depiction of the GAGOme approach. Genetic engineering by targeted KO with CRISPR–Cas9 or targeted KI with a zinc-finger nuclease (ZFN) generates a library of isogenic cells displaying different repertoires of GAG structures. The GAGOme cell library can be used to dissect the specificities of GAG-binding proteins by flow cytometry, for the production of recombinant proteoglycans with distinct GAG chains, and for metabolic priming with xylosides ($XylNapNH_2$) to generate libraries of distinct GAG chains suitable for microarrays.

disaccharide analysis corresponded to that in previous reports[10] (Supplementary Fig. 2). For comparison, we also performed the disaccharide analysis in the human HEK293 cell line. The GAG compositions of both WT CHO and HEK293 cells correlated well with the predicted glycosylation, indicating that RNA-seq analysis can reasonably predict qualitative GAG biosynthesis.

Genetic deconstruction of GAG biosynthesis in CHO cells. We used the clustered regularly interspaced short palindromic repeat (CRISPR)–CRISPR-associated gene 9 (Cas9) system for targeted KO of genes encoding GAGs. This was done individually and, when required, combinatorially for deconstruction of the endogenous GAG biosynthetic capacity, in a strategy similar to that previously used for the deconstruction of protein N-glycosylation[11]. We systematically targeted the sequential steps in GAG biosynthesis by using prevalidated guide RNAs (gRNAs) (Fig. 2 and Supplementary Tables 1 and 2), and evaluated the effects on the GAG biosynthesis by HPLC-based disaccharide analysis of total cell lysates (Supplementary Figs. 2–6; semiquantification of the HPLC profiles is presented in Supplementary Tables 3–9). The presented engineering does not address keratin sulfate GAGs because the genes involved are less well characterized and the analysis is complex. The engineering also does not address hyaluronic acid, which was not detectable in CHO cells.

Linker assembly. Synthesis of the common tetrasaccharide linker (GlcA-β1-3Gal-β1-3Gal-β1-4Xyl-β1-*O*-Ser) is initiated by xylosyltransferase 1 (XYLT1) and XYLT2 and elongated by galactosyltransferases (β4Gal-T7 and β3Gal-T6) and a glucuronyltransferase (GlcAT-I) (Fig. 2). KO of *Xylt2* eliminated GAG synthesis (Fig. 3a and Supplementary Fig. 3), in agreement with *Xylt1* not being expressed (Supplementary Fig. 1). KO of *B4galt7* and *B3gat3* also eliminated synthesis, whereas KO of *B3galt6* resulted in only the reduction of CS/HS synthesis (Fig. 3a and Supplementary Fig. 3a).

The linker assembly and CS/HS priming are coregulated by transient phosphorylation at carbon 2 (C2) of the xylose residue by the xylosylkinase FAM20B[12] and the xylosylphosphatase PXYLP1[13], as well as by sulfation at C6 of both galactose residues and at C4 of the second galactose residue[14–16]. The CS 6-*O*-sulfotransferase 1 encoded by *Chst3* catalyzes sulfation of the linker galactose residues in vitro[17]; however, KI of human *CHST3* (h*CHST3*) did not change the ratio of CS/DS to HS (Supplementary Fig. 4a). The efficiency of β3Gal-T6 and GlcAT-I is regulated by phosphorylation of the xylose residue by FAM20B[18,19]. KO of *Fam20b* led to decreased amounts of both CS and HS; however, KO of *Pxylp1* did not alter the GAG synthesis (Fig. 3a and Supplementary Fig. 3a).

CS/HS initiation and elongation. Two homologous β4-*N*-acetylgalactosaminyltransferases (β4GalNAc-transferases), encoded by *Csgalnact1 and Csgalnact2*, initiate CS synthesis[20,21]. KO of *Csgalnact2*, as well as combinatorial KO of *Csgalnact1* and *Csgalnact2*, markedly reduced but did not eliminate CS synthesis (Fig. 3b and Supplementary Fig. 3b). Four homologous genes—*Chsy1* (previous/alternative gene symbol: *Css1*), *Chsy3* (previous/alternative gene symbol: *Css3* or *Chsy2*), *Chpf* (previous/alternative gene symbol: *Css2*), and *Chpf2* (previous/alternative gene symbol: *Csglcat*)—encode enzymes with two catalytic domains with β4GalNAc-transferase (CAZy GT7)[22] and β3GlcA-transferase (GT31) activities. These enzymes function in heterodimeric complexes that produce the GalNAc-β1-4GlcA-β1-3 repeat backbone and influence the length of CS chains[23,24]. KO of *Chpf* or *Chpf2* did not alter CS synthesis substantially, whereas KO of *Chsy1* eliminated CS formation (Fig. 3b and Supplementary Fig. 3b).

HS biosynthesis is suggested to be initiated by EXTL2 and EXTL3[25,26] and elongated by EXT1 and EXT2[27]. The role of EXTL1 is poorly understood. The enzymes EXTL1 and EXTL3 have two catalytic domains for α4GlcNAc (GT64) and β4GlcA-transferase (GT47) activities, whereas EXTL2 has one catalytic domain with α4GlcNAc and α4GalNAc-transferase activity[25]. EXTL3 is more efficient than EXTL2 in initiating HS synthesis in vitro[26,27], whereas EXTL2 serves as a negative regulator of GAG biosynthesis by adding α4GalNAc to the linker[25]. KO of *Extl3* completely eliminated HS synthesis, whereas KO of *Extl1* slightly decreased HS synthesis, and KO of *Extl2* increased both CS and HS synthesis (Fig. 3b and Supplementary Fig. 3b). The bifunctional EXT enzymes function in a heterodimeric complex[28], and KO of *Ext1* and/or *Ext2* completely eliminated HS synthesis (Fig. 3b and Supplementary Fig. 5).

CS/DS modifications. DS is generated by epimerization of GlcA to L-iduronic acid (IdoA) by two homologous GlcA C5 epimerases,

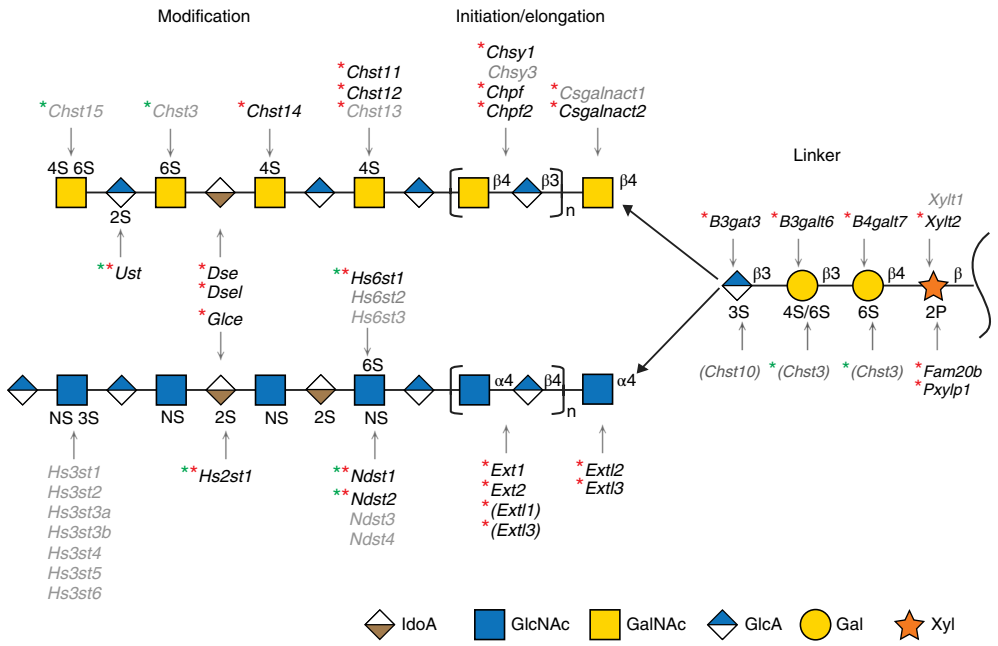

Fig. 2 | Overview of the genetic regulation of GAG biosynthesis. Biosynthetic steps associated with the listed genes are indicated by arrows. Genes in black text were expressed in CHO cells, and genes in gray text were not expressed. Red asterisks indicate genes that were knocked out, and green asterisks indicate genes that were knocked in. Genes in parentheses have unclear functions. The "Linker" region refers to the common tetrasaccharide core linked to a serine of the protein backbone. "Initiation" refers to the fifth biosynthetic step that adds β-GalNAc or α-GlcNAc to the linker region, committing the biosynthesis to CS/DS or HS, respectively. "Elongation" refers to the polymerization process that extends the disaccharide units of the GAG chains. "Modification" refers to epimerization and sulfation steps. Glycan symbols are drawn according to the standardized Symbol Nomenclature for Glycans format.[50]

encoded by *Dse* and *Dsel*[29,30]. CHO cells produced only minor amounts of DS (Supplementary Fig. 2), and KO of *Dsel* did not substantially affect the CS/DS disaccharide composition, whereas KO of *Dse* or of both genes eliminated DS synthesis (Fig. 3c and Supplementary Fig. 3c).

CHO cells mainly produce 4-*O*-sulfated CS/DS. 4-*O*-sulfation of CS/DS is carried out by three homologous CS 4-*O*-sulfotransferases and one DS 4-*O*-sulfotransferase, which are encoded by *Chst11*, *Chst12*, *Chst13*, and *Chst14*, respectively. *Chst13* is not expressed in CHO cells, and KO of *Chst11* slightly decreased 4-*O*-sulfation, whereas KO of *Chst12* increased sulfation. KO of *Chst14* eliminated 4-*O*-sulfated DS. Combinatorial KO of *Chst11*, *Chst12*, and *Chst13* decreased 4-*O*-sulfated CS/DS disaccharides to the same extent as that seen in a *Chst11*-KO line, whereas combinatorial KO of *Chst11*, *Chst12*, *Chst13*, and *Chst14* eliminated detection of 4-*O*-sulfated CS/DS disaccharides (Fig. 3c and Supplementary Fig. 3c).

HS modifications. Four homologous bifunctional *N*-deacetylase–*N*-sulfotransferases (NDST1–NDST4) de-*N*-acetylate and *N*-sulfate GlcNAc residues to start modification of the nascent heparosan chain[6]. KO of *Ndst1* or *Ndst2* did not substantially affect the amount of *N*-sulfated HS disaccharides, whereas KO of both eliminated *N*-sulfated disaccharides, as well as *N*-,2-*O*-disulfated and *N*-,2,6-*O*-trisulfated disaccharides (Fig. 3c and Supplementary Fig. 3d)[6].

Epimerization of GlcA to IdoA is directed by the C-5 epimerase encoded by *Glce*, which is predicted to interact with the 2-*O*-sulfotransferase encoded by *Hs2st1*[31]. KO of *Glce* or *Hs2st1* eliminated detection of 2-*O*-sulfated disaccharides (Fig. 3c and Supplementary Fig. 3d). KO of *Hs6st1*, the only expressed gene encoding 6-*O*-sulfotransferase, eliminated the barely detectable levels of 6-*O*-sulfated, *N*-,6-*O*-disulfated, and *N*-,2,6-*O*-trisulfated HS (Fig. 3c and Supplementary Fig. 3d).

Engineering expanded GAG biosynthesis capabilities. To expand CS biosynthesis in CHO cells, we introduced *CHST3* to induce 6-*O*-sulfation, *UST* to induce 2-*O*-sulfation, and *CHST15* to induce 4,6-*O*-disulfation. KI of *CHST3* induced substantial synthesis of 6-*O*-sulfated disaccharides, with a corresponding decrease in 4-*O*-sulfation (Fig. 3c and Supplementary Fig. 4a). KI of *CHST15* induced detectable 4,6-*O*-disulfation, and KI of *UST* increased amounts of 2,4-*O*-disulfated disaccharides, as predicted.

To expand HS biosynthesis in CHO cells, we first increased the amount of *N*-sulfation. KI of *NDST2* markedly enhanced the level of *N*-sulfated HS, whereas KI of mouse *Ndst1* was less efficient (Fig. 3c and Supplementary Fig. 4b). KI of *Hs2st1* increased the amount of 2-*O*-sulfated disaccharide, but not of *N*-,2-*O*-disulfated disaccharide, presumably because of limited *N*-sulfation. Similarly, KI of *HS6ST1* resulted in only a slight increase in the amount of 6-*O*-sulfated disaccharide, but stacked KI of *HS6ST1* with *NDST2* generated the *N*-,6-*O*-disulfated disaccharide.

Display and dissection of GAG–protein interactions. The large library of isogenic CHO cells generated here is unique in that it is fairly comprehensive and includes individual cells that display different repertoires of GAG structures, which enables the analysis of GAG features, including larger structural motifs in various bioassays. Further expansion of the current cell library could readily be achieved through additional combinatorial gene engineering of selected sublibraries with the provided validated gene-targeting constructs. We were able to demonstrate the potential of the GAGOme method for the dissection of binding specificities with representative classes of GAG-binding proteins.

Probing GAG specificities of pathogen receptors with the GAGOme. The *Plasmodium falciparum* VAR2CSA protein mediates the adhesion of infected erythrocytes to the human placenta and is proposed to bind a distinct 4-*O*-sulfated CS that is expressed in normal placenta and in most human cancers[32]. We used GAGOme sublibraries to sequentially investigate the role of GAG features for cell binding of recombinant VAR2CSA (rVAR2CSA) by flow cytometry

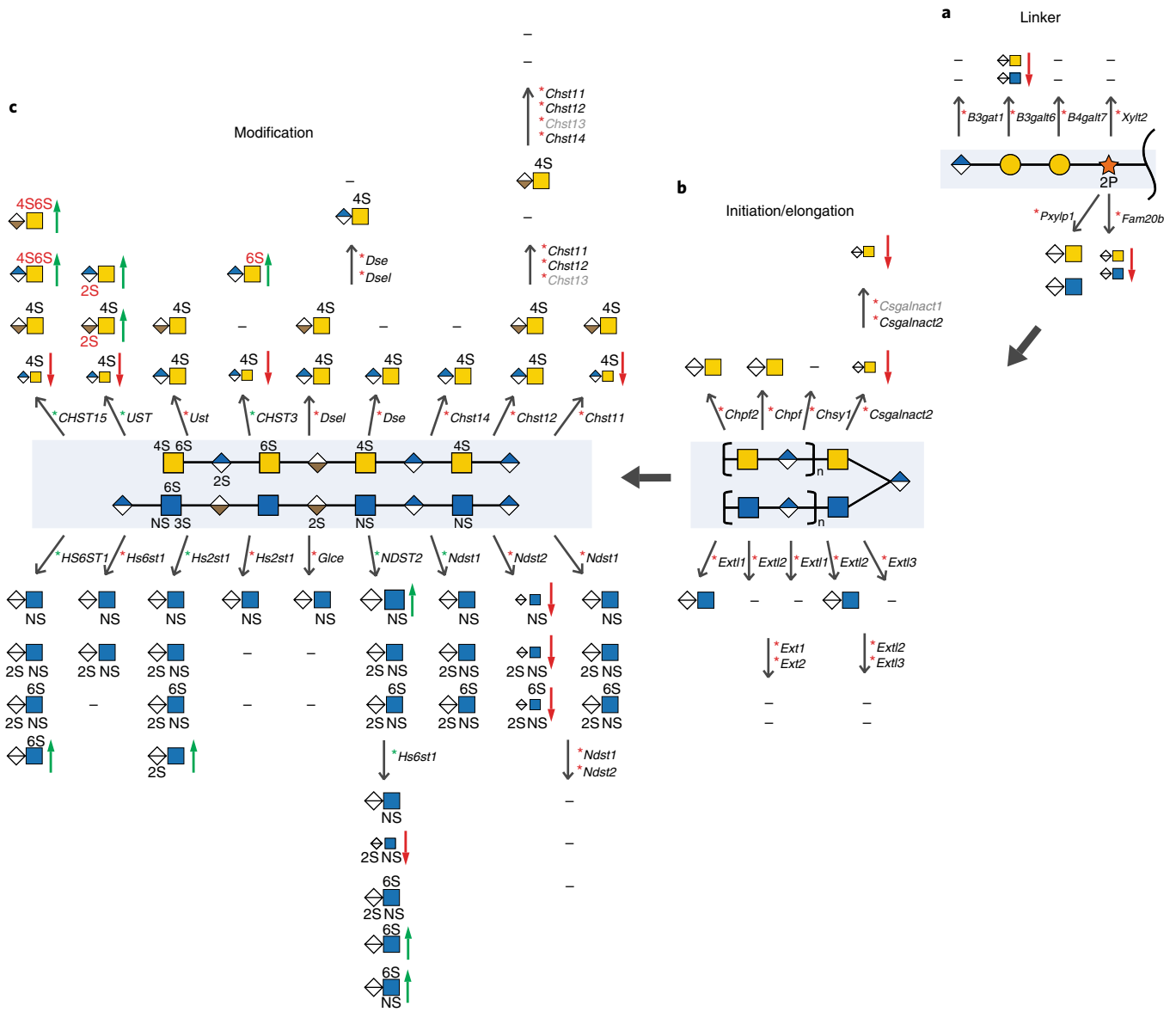

Fig. 3 | Summary overview of effects of KO/KI engineering on HPLC disaccharide profiles of total cell lysates. a, Targeting assembly of the tetrasaccharide linker region. **b**, Targeting of the CS/HS initiation decision and elongation/polymerization steps. **c**, Targeting of modification steps, including sulfation and epimerization. Throughout, colors of asterisks and arrows indicate KO (red) or KI (green) genes, and the main changes observed in the disaccharide analysis of total lysates of mutant cells are indicated by symbol size (smaller or larger) and direction of arrows (down or up) for the key disaccharide observed. Complete loss of detectable disaccharides is indicated by minus signs (–). For CS/DS, both disaccharide variants are shown, with indication of the effects on CS and/or DS. All of the disaccharide HPLC profile analyses are presented in Supplementary Figs. 2–6.

(Fig. 4a). We demonstrated the ultimate requirement of GAGs for binding of rVAR2CSA by targeting the linker assembly, and observed that KO of *Xylt2*, *B4galt7*, and *B3gat3* completely abrogated binding. Targeting of HS initiation did not alter binding, although combinatorial KO of *Extl2* and *Extl3* enhanced binding (Fig. 4a). In contrast, targeting of the CS initiation and elongation enzymes showed that KO of only *Chsy1*, which completely eliminated CS synthesis, abrogated binding, thus demonstrating the requirement for CS. Further targeting of the CS modification steps showed that KO of *Dse* and/or *Dsel* did not affect binding, which indicated that DS was not essential for binding. In contrast, targeting of the 4-*O*-sulfation step revealed that only combined KO of *Chst11*, *Chst12*, and *Chst13* completely abrogated binding. Notably, individual KO of the genes encoding the four 4-*O*-sulfotransferases indicated that KO of *Chst11* selectively reduced binding substantially. This suggests that CHST11 may be a major contributor to the binding motif for VAR2CSA, but additional studies would be needed to confirm this. Similarly, KI of *CHST3* or *CHST15* decreased rVAR2CSA binding to the same extent as that seen after KO of *Chst11*; however, KI of *CHST3* reduced the amount of 4-*O*-sulfated CS disaccharide considerably more than did the KI of *CHST15* (Supplementary Fig. 4a), again suggesting a higher level of complexity in the VAR2CSA binding epitope and confirming the absolute requirement for 4-*O*-sulfated CS. KI of *UST* also decreased rVAR2CSA binding substantially (Fig. 4a), further demonstrating that VAR2CSA requires an isolated 4-*O*-sulfated CS epitope. Although further combinatorial genetic engineering may refine this analysis, the data clearly demonstrate the capacity of the GAGOme for guiding the analysis of specific GAG–pathogen interactions. The method also provides genetic and biosynthetic information for the regulation of the VAR2CSA binding epitope that may

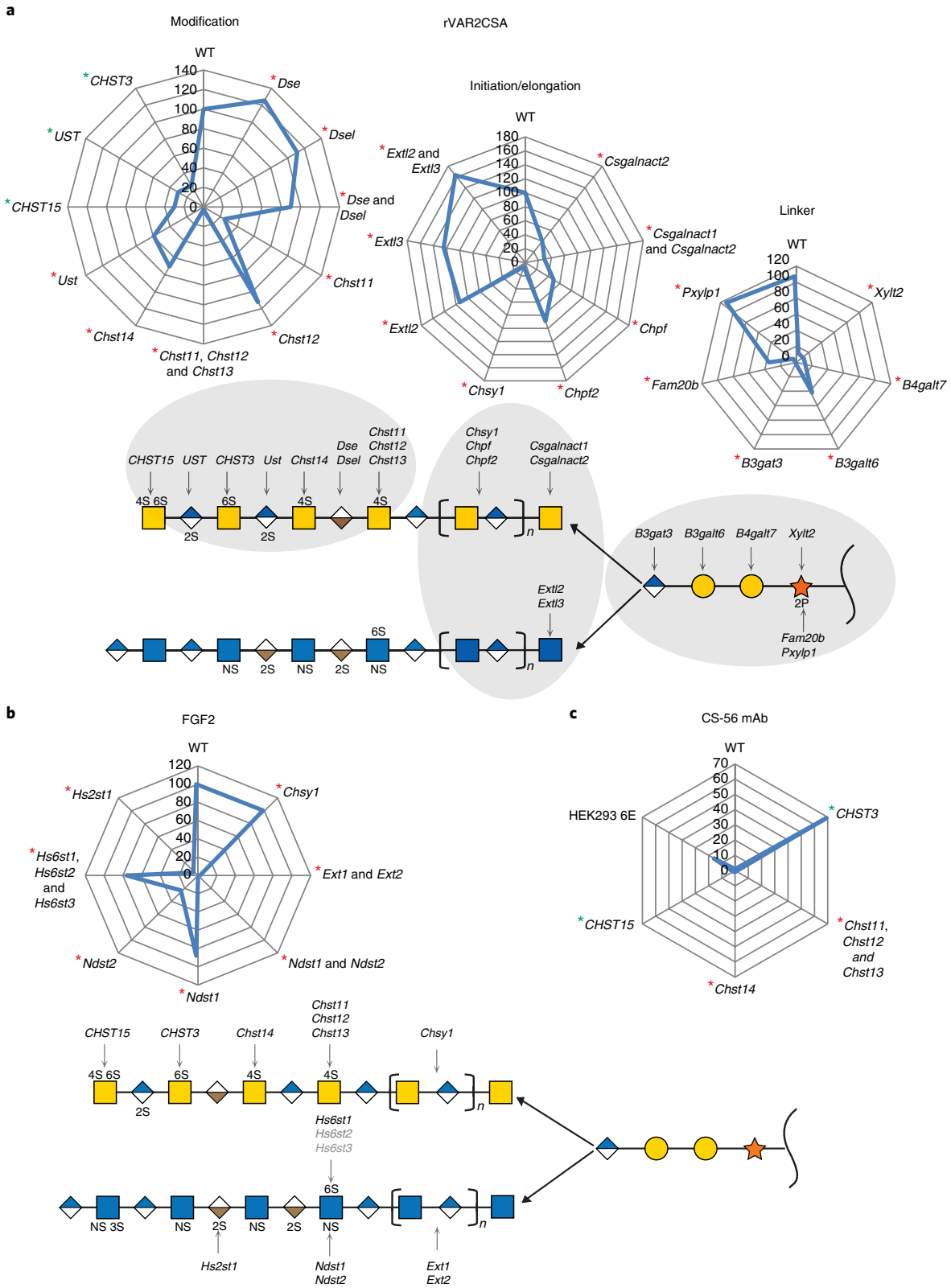

Fig. 4 | Exploring the binding specificities of GAG-binding proteins by using flow cytometry with GAGOme sublibraries. a, rVAR2CSA binding specificity. Three sublibraries probing the linker region, CS/HS initiation/elongation, and CS modification, respectively, were analyzed. Radar charts show the relative mean fluorescence intensity (MFI) (WT cells were assigned a value of 100) after genetic KO or KI of the indicated gene(s). Red or green asterisks indicate genes that were knocked out or knocked in, respectively. The charts represent single experiments that were performed at one concentration of rVAR2 (100 nM). Experiments were performed at least two times and with multiple clones (2–4), with similar results obtained each time. **b**, FGF2 binding specificity. One sublibrary that probed the requirement for CS/HS (*Chsy1*, *Ext1*, and *Ext2*) and HS sulfation was selected for analysis on the basis of the predicted specificity[33]. Radar charts show the relative MFI (WT = 100) of genetic KO of the indicated gene(s). **c**, Binding specificity of the monoclonal antibody CS-56. A sublibrary that probed the requirement for CS sulfation was selected for analysis on the basis of the predicted specificity[36]. Radar charts show the relative MFI (WT = 1) of genetic KO or KI of the indicated gene(s).

be used to explore the molecular basis for the aberrant expression of the VAR2CSA CS epitope on cancer cells.

Probing growth factor interactions. A number of growth factors and their receptors require GAGs for interactions. A classic example is fibroblast growth factor 2 (FGF2), which binds to HS with the requirement for *N*-sulfation and 2-*O*-sulfated IdoA[33]. Evaluation of FGF2 binding by the GAGOme method confirmed the requirement for both *N*-sulfation and 2-*O*-sulfation, as combinatorial KO of *Ndst1* and *Ndst2*, as well as KO of *Hs2st1*, eliminated binding (Fig. 4b). Moreover, KO of *Hs6st1* to eliminate 6-*O*-sulfated HS did not affect binding substantially. This is in agreement with previous studies showing that binding of FGF2 is independent of 6-*O*-sulfation. Of note, FGF2 signaling is dependent on 6-*O*-sulfation, and this may be further explored with the engineered cell lines[34].

Probing antibody specificities. The CS-56 monoclonal antibody reportedly binds to 4-*O*-sulfated and 6-*O*-sulfated CS[35]. In agreement with this, the antibody did not bind WT CHO cells that lacked 6-*O*-sulfated CS, whereas weak binding was observed with WT HEK293 cells (Fig. 4c). Elimination of CS 4-*O*-sulfation (by KO of *Chst11*, *Chst12*, and *Chst13*) or introduction of 4,6-*O*-sulfation (by KI of *CHST15*) did not induce binding; however, the introduction of 6-*O*-sulfation (by KI of *CHST3*) in WT CHO cells combined with the endogenous 4-*O*-sulfated CS induced strong binding (Fig. 4c). Previously, CS-56 was also shown to bind an octasaccharide CS motif containing 4-*O*-sulfated, 6-*O*-sulfated, and 2,6-*O*-disulfated disaccharides[36].

GAGOme method for sustainable production of GAG libraries by xyloside priming of cells. β-xylosides can be used as metabolic primers with cultured cells for the production of GAG chains that are secreted into the culture medium[9,37]. We first used the 2-naphthyl β-D-xylopyranoside (XylNap) as a primer with a representative GAGOme cell library to produce a panel of secreted GAG chains. In general, the GAG structures produced on the XylNap primer reflected the glycosylation capacities of the engineered cell used, but with less sulfation compared with that in the total cell lysates from corresponding nonprimed engineered cells (Supplementary Fig. 5a). Notably, we did observe some retained capacity for priming of GAG synthesis in *B4galt7*-KO cells, which we did not observe in the analysis of GAGs from cell lysates from nonprimed engineered cells (Supplementary Fig. 3). This was also observed previously by Esko and colleagues[38].

We also used an amino derivative of XylNap, 2-(6-((3-aminopropyl) oxy)-naphthyl) β-D-xylopyranoside ($XylNapNH_2$), for priming, in which the amine group may be functionalized and used for coupling chemistry to create openings for direct immobilization of primed GAGs on microarrays. $XylNapNH_2$ required higher concentrations (1 mM) for efficient priming, and we found that the priming patterns were slightly different from those of XylNap (Supplementary Fig. 5b). Nevertheless, $XylNapNH_2$-primed GAG chains probably could be used to immobilize distinct GAGs on activated microarray slides.

Production of recombinant proteoglycans with distinct GAGs. Serglycin is a small secreted proteoglycan expressed in hematopoietic and endothelial cells that carries mainly CS GAGs. An exception is the serglycin expressed in mast cells in connective tissue, which selectively carries heparin and represents a source for isolated heparin[39]. Recombinantly expressed serglycin in WT CHO cells carried CS GAGs almost exclusively (Supplementary Fig. 6). However, when we expressed serglycin in CHO cells that were deficient in CS synthesis, the GAG profile converted to HS with the low-sulfation pattern found in total cell lysates of WT CHO cells (Supplementary Fig. 6). Moreover, KI of *NDST2* induced a substantial increase in *N*-sulfation of HS (Supplementary Fig. 6).

Discussion

The GAGOme cell library method represents an inexhaustible resource for the display and production of GAGs with distinct features. We stably engineered a total of 28 genes individually and in select combinations to generate our GAGOme library with defined capacities for GAG biosynthesis. Our approach allows the construction of comprehensive GAG libraries from total cell lysates, through metabolic priming, or by recombinant expressed proteoglycans such as serglycin, which can be used further to generate traditional printed glycan and proteoglycan arrays (Fig. 1). We also developed a validated gene-targeting gRNA library that can be used to extend the current GAGOme library with additional combinatorial screening strategies needed to dissect higher-order GAG structures.

We used classical HPLC-based disaccharide analysis of isolated GAGs from cell lysates to probe effects on GAG biosynthesis and to guide the combinatorial engineering process[10]. This assay relies on controlled digestion by lyases and has important limitations with respect to exploration of the outcomes of the broad gene engineering performed here. The assay provides only an average estimate of the disaccharide composition, and thus does not reveal potential larger patterns orchestrated at the oligosaccharide level or the length of the GAG chains (for example, isolated DS units are not digested by chondroitinase B into disaccharides). The sulfation pattern may interfere with the digestion efficiency, as, for example, 3-*O*-sulfate in HS inhibits digestion by heparinases[40]. The engineering and changed glycosylation capacities may also affect cells and bias the disaccharide analysis. We note that although we did not observe any consistent gross changes in cell growth and viability with the engineered cell clones, we did observe minor variations in the growth density of cells harvested for analysis. However, these variations did not appear to be consistent for specific engineering events, and we considered them as clonal variations. Moreover, the disaccharide analysis was performed on total cell lysates, which include biosynthetic intermediates, and thus it does not accurately reflect the GAG structures presented on the cell membrane or on secreted primed xylosides and proteoglycans. Finally, the broadness of this high-throughput gene-engineering strategy made it impractical to include further detailed analyses of the structure and size of GAGs and potential cellular effects such as, for example, changes in the expression of specific proteoglycans that might affect GAG biosynthesis. We therefore undertook only semiquantification of the HPLC disaccharide profiles (Supplementary Tables 3–9), the results of which suggest that the total disaccharide content among all of the clones tested with expected GAG biosynthesis varied only two- to threefold.

Nevertheless, the disaccharide analysis provided clear indications and guidance for the effects of the gene engineering, and the effects observed were generally in agreement with the predicted functions of the targeted genes. The real power of the GAGOme cell library is the ability to display complex GAG features, including those that currently are not amenable to structural analysis, and to dissect the specific genes required for the biosynthesis and display of distinct GAG features. The GAGOme method provides direct access to cells for the production of relevant GAG chains as primed xylosides or recombinant proteoglycans that can be used for more detailed analysis. We envision that broader use of the GAGOme library will lead to further refinement and eventually insight into higher-order patterns and binding motifs on GAG chains.

We demonstrated the concept of the GAGOme as a display of GAGs by using a diverse set of proteins known to bind GAGs, and confirmed previous results[32,33,35], but also added further refinement of the epitopes recognized and identified the biosynthetic and genetic basis. The use of cell-based arrays for the display and investigation of biological functions of GAGs is principally different from the way printed arrays are used[41]. Printed glycan arrays

reveal direct binding to individual oligosaccharides as positive signals, and interpretation of the binding specificity is based on knowledge about the printed structures. In contrast, binding specificities in the GAGOme cell library are interpreted through loss or gain of interactions as a result of the gene engineering performed. The GAGOme approach also provides a detailed map of the genetic and biosynthetic regulation of the identified GAG-binding epitopes, as well as a design matrix for engineering cell-based production of the relevant distinct GAG structure. Recently, gene engineering was applied to dissect the requirement of HS in the cellular uptake of tau and α-synuclein[42].

The simple combinatorial KO–KI gene engineering used here might not fully display the physiological regulation of GAG biosynthesis in cells and the true diversity of the 'GAG code'. Reports have shown that the overexpression of individual or multiple GAG enzymes appears to perturb biosynthesis[43], and the formation of complexes among enzymes might provide a different level of regulation[1]. The gene engineering performed here did not uncover such limitations, but it did confirm the existence of essential heterodimeric complexes for EXT1/EXT2 and CHPF/CHPF2/CHSY1, which are involved in the initiation of HS and CS biosynthesis, respectively (Fig. 3).

Glycan arrays have contributed greatly to the understanding of biological interactions with the diversity of the glycome[44]. However, comprehensive arrays with GAGs are still not available for general screening of GAG-binding proteins[45–48]. The use of the GAGOme to produce libraries of distinct xyloside-primed GAG chains and to profile GAG biosynthetic capacities is related to the CORA strategy for profiling and displaying GalNAc-type *O*-glycosylation in cells[49]. The CORA strategy uses a GalNAc primer to amplify and characterize the heterogeneous *O*-glycosylation capacity of cells, and the heterogeneous primed glycans are separated and analyzed. The GAGOme strategy instead relies on a large panel of isogenic cells with more homogeneous and complementary glycosylation capacities, thus eliminating the need for separation of the primed glycans. We further demonstrated that it is possible to use an activated primer suitable for direct printing on arrays[41]. The use of a similar panel of engineered isogenic cells with different *O*-glycosylation capacities would also greatly add to the CORA strategy.

We demonstrated the ability to produce a proteoglycan with distinct GAG structures and manipulate its sulfation pattern. We predict that other proteoglycans could be expressed and their GAGs custom-designed, although further studies are needed to explore potential protein-specific restrictions. The findings with serglycin might represent one step toward the engineering of CHO cells to produce and secrete serglycin with GAGs suitable for the development of CHO-based bioproduction of heparin. We expect that further engineering might eventually enable the production of the anticoagulant heparin, an extensively sulfated form of HS with a high content of IdoA that is produced by mast cells[6], with a sulfation pattern required for antithrombin binding. Overexpression of HS3ST1 alone in CHO cells has resulted in measurable antithrombin binding to the produced HS[40]. Moreover, we envision that heparin could be produced through xyloside-primed GAG chains in such appropriately engineered CHO cells.

In summary, the presented GAGOme method enables the sustainable production and display of GAGs that circumvent past resource and technical constraints. The strategy may provide new insight into the genetic and biosynthetic regulation of GAGs, and potentially open the way for engineering of CHO cells for the production of heparin.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41592-018-0086-z.

Received: 18 December 2017; Accepted: 25 May 2018;
Published online: 13 August 2018

References

1. Esko, J. D. & Selleck, S. B. Order out of chaos: assembly of ligand binding sites in heparan sulfate. *Annu. Rev. Biochem.* **71**, 435–471 (2002).
2. Mizumoto, S., Yamada, S. & Sugahara, K. Molecular interactions between chondroitin-dermatan sulfate and growth factors/receptors/matrix proteins. *Curr. Opin. Struct. Biol.* **34**, 35–42 (2015).
3. Xu, D. & Esko, J. D. Demystifying heparan sulfate-protein interactions. *Annu. Rev. Biochem.* **83**, 129–157 (2014).
4. Mizumoto, S., Yamada, S. & Sugahara, K. Human genetic disorders and knockout mice deficient in glycosaminoglycan. *Biomed. Res. Int.* **2014**, 495764 (2014).
5. Gama, C. I. et al. Sulfation patterns of glycosaminoglycans encode molecular recognition and activity. *Nat. Chem. Biol.* **2**, 467–473 (2006).
6. Fu, L., Suflita, M. & Linhardt, R. J. Bioengineered heparins and heparan sulfates. *Adv. Drug Deliv. Rev.* **97**, 237–249 (2016).
7. Esko, J. D., Stewart, T. E. & Taylor, W. H. Animal cell mutants defective in glycosaminoglycan biosynthesis. *Proc. Natl. Acad. Sci. USA* **82**, 3197–3201 (1985).
8. Zhang, L., Lawrence, R., Frazier, B. A. & Esko, J. D. CHO glycosylation mutants: proteoglycans. *Methods Enzymol.* **416**, 205–221 (2006).
9. Okayama, M., Kimata, K. & Suzuki, S. The influence of *p*-nitrophenyl β-d-xyloside on the synthesis of proteochondroitin sulfate by slices of embryonic chick cartilage. *J. Biochem.* **74**, 1069–1073 (1973).
10. Li, G. et al. Glycosaminoglycanomics of cultured cells using a rapid and sensitive LC-MS/MS approach. *ACS Chem. Biol.* **10**, 1303–1310 (2015).
11. Yang, Z. et al. Engineered CHO cells for production of diverse, homogeneous glycoproteins. *Nat. Biotechnol.* **33**, 842–844 (2015).
12. Koike, T., Izumikawa, T., Tamura, J. & Kitagawa, H. FAM20B is a kinase that phosphorylates xylose in the glycosaminoglycan-protein linkage region. *Biochem. J.* **421**, 157–162 (2009).
13. Koike, T., Izumikawa, T., Sato, B. & Kitagawa, H. Identification of phosphatase that dephosphorylates xylose in the glycosaminoglycan-protein linkage region of proteoglycans. *J. Biol. Chem.* **289**, 6695–6708 (2014).
14. Sugahara, K., Yamashina, I., De Waard, P., Van Halbeek, H. & Vliegenthart, J. F. Structural studies on sulfated glycopeptides from the carbohydrate-protein linkage region of chondroitin 4-sulfate proteoglycans of swarm rat chondrosarcoma. Demonstration of the structure Gal(4-*O*-sulfate)β-1-3Gal-β-1-4XYL-β-1-*O*-Ser. *J. Biol. Chem.* **263**, 10168–10174 (1988).
15. Ueno, M., Yamada, S., Zako, M., Bernfield, M. & Sugahara, K. Structural characterization of heparan sulfate and chondroitin sulfate of syndecan-1 purified from normal murine mammary gland epithelial cells. Common phosphorylation of xylose and differential sulfation of galactose in the protein linkage region tetrasaccharide sequence. *J. Biol. Chem.* **276**, 29134–29140 (2001).
16. de Waard, P., Vliegenthart, J. F., Harada, T. & de Sugahara, K. Structural studies on sulfated oligosaccharides derived from the carbohydrate-protein linkage region of chondroitin 6-sulfate proteoglycans of shark cartilage. II. Seven compounds containing 2 or 3 sulfate residues. *J. Biol. Chem.* **267**, 6036–6043 (1992).
17. Kitagawa, H. et al. Sulfation of the galactose residues in the glycosaminoglycan-protein linkage region by recombinant human chondroitin 6-*O*-sulfotransferase 1. *J. Biol. Chem.* **283**, 27438–27443 (2008).
18. Wen, J. et al. Xylose phosphorylation functions as a molecular switch to regulate proteoglycan biosynthesis. *Proc. Natl. Acad. Sci. USA* **111**, 15723–15728 (2014).
19. Tone, Y. et al. 2-*O*-phosphorylation of xylose and 6-*O*-sulfation of galactose in the protein linkage region of glycosaminoglycans influence the glucuronyltransferase-I activity involved in the linkage region synthesis. *J. Biol. Chem.* **283**, 16801–16807 (2008).
20. Uyama, T., Kitagawa, H., Tamura Ji, J. & Sugahara, K. Molecular cloning and expression of human chondroitin *N*-acetylgalactosaminyltransferase: the key enzyme for chain initiation and elongation of chondroitin/dermatan sulfate on the protein linkage region tetrasaccharide shared by heparin/heparan sulfate. *J. Biol. Chem.* **277**, 8841–8846 (2002).
21. Uyama, T. et al. Molecular cloning and expression of a second chondroitin *N*-acetylgalactosaminyltransferase involved in the initiation and elongation of chondroitin/dermatan sulfate. *J. Biol. Chem.* **278**, 3072–3078 (2003).
22. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495 (2014).
23. Izumikawa, T., Uyama, T., Okuura, Y., Sugahara, K. & Kitagawa, H. Involvement of chondroitin sulfate synthase-3 (chondroitin synthase-2) in chondroitin polymerization through its interaction with chondroitin synthase-1 or chondroitin-polymerizing factor. *Biochem. J.* **403**, 545–552 (2007).
24. Izumikawa, T. et al. Identification of chondroitin sulfate glucuronyltransferase as chondroitin synthase-3 involved in chondroitin polymerization: chondroitin polymerization is achieved by multiple enzyme complexes consisting of chondroitin synthase family members. *J. Biol. Chem.* **283**, 11396–11406 (2008).

25. Kitagawa, H., Shimakawa, H. & Sugahara, K. The tumor suppressor EXT-like gene *EXTL2* encodes an α1,4-*N*-acetylhexosaminyltransferase that transfers *N*-acetylgalactosamine and *N*-acetylglucosamine to the common glycosaminoglycan-protein linkage region. The key enzyme for the chain initiation of heparan sulfate. *J. Biol. Chem.* **274**, 13933–13937 (1999).
26. Kim, B. T. et al. Human tumor suppressor *EXT* gene family members *EXTL1* and *EXTL3* encode α1,4- *N*-acetylglucosaminyltransferases that likely are involved in heparan sulfate/ heparin biosynthesis. *Proc. Natl. Acad. Sci. USA* **98**, 7176–7181 (2001).
27. Lind, T., Tufaro, F., McCormick, C., Lindahl, U. & Lidholt, K. The putative tumor suppressors EXT1 and EXT2 are glycosyltransferases required for the biosynthesis of heparan sulfate. *J. Biol. Chem.* **273**, 26265–26268 (1998).
28. McCormick, C., Duncan, G., Goutsos, K. T. & Tufaro, F. The putative tumor suppressors EXT1 and EXT2 form a stable complex that accumulates in the Golgi apparatus and catalyzes the synthesis of heparan sulfate. *Proc. Natl. Acad. Sci. USA* **97**, 668–673 (2000).
29. Pacheco, B., Malmström, A. & Maccarana, M. Two dermatan sulfate epimerases form iduronic acid domains in dermatan sulfate. *J. Biol. Chem.* **284**, 9788–9795 (2009).
30. Maccarana, M. et al. Biosynthesis of dermatan sulfate: chondroitin-glucuronate C5-epimerase is identical to SART2. *J. Biol. Chem.* **281**, 11560–11568 (2006).
31. Pinhal, M. A. et al. Enzyme interactions in heparan sulfate biosynthesis: uronosyl 5-epimerase and 2-*O*-sulfotransferase interact in vivo. *Proc. Natl. Acad. Sci. USA* **98**, 12984–12989 (2001).
32. Salanti, A. et al. Targeting human cancer by a glycosaminoglycan binding malaria protein. *Cancer Cell* **28**, 500–514 (2015).
33. Turnbull, J. E., Fernig, D. G., Ke, Y., Wilkinson, M. C. & Gallagher, J. T. Identification of the basic fibroblast growth factor binding sequence in fibroblast heparan sulfate. *J. Biol. Chem.* **267**, 10337–10341 (1992).
34. Pye, D. A., Vives, R. R., Turnbull, J. E., Hyde, P. & Gallagher, J. T. Heparan sulfate oligosaccharides require 6-*O*-sulfation for promotion of basic fibroblast growth factor mitogenic activity. *J. Biol. Chem.* **273**, 22936–22942 (1998).
35. Avnur, Z. & Geiger, B. Immunocytochemical localization of native chondroitin sulfate in tissues and cultured cells using specific monoclonal antibody. *Cell* **38**, 811–822 (1984).
36. Ito, Y. et al. Structural characterization of the epitopes of the monoclonal antibodies 473HD, CS-56, and MO-225 specific for chondroitin sulfate D-type using the oligosaccharide library. *Glycobiology* **15**, 593–603 (2005).
37. Persson, A., Ellervik, U. & Mani, K. Fine-tuning the structure of glycosaminoglycans in living cells using xylosides. *Glycobiology* **28**, 499–511 (2018).
38. Esko, J. D. et al. Inhibition of chondroitin and heparan sulfate biosynthesis in Chinese hamster ovary cell mutants defective in galactosyltransferase I. *J. Biol. Chem.* **262**, 12189–12195 (1987).
39. Kolset, S. O. & Tveit, H. Serglycin—structure and biology. *Cell. Mol. Life Sci.* **65**, 1073–1085 (2008).
40. Datta, P. et al. Bioengineered Chinese hamster ovary cells with Golgi-targeted 3-*O*-sulfotransferase-1 biosynthesize heparan sulfate with an antithrombin-binding site. *J. Biol. Chem.* **288**, 37308–37318 (2013).
41. Blixt, O. et al. Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proc. Natl. Acad. Sci. USA* **101**, 17033–17038 (2004).
42. Stopschinski, B. E. et al. Specific glycosaminoglycan chain length and sulfation patterns are required for cell uptake of tau *versus* α-synuclein and β-amyloid aggregates. *J. Biol. Chem.* **293**, 10826–10840 (2018).
43. Baik, J. Y. et al. Metabolic engineering of Chinese hamster ovary cells: towards a bioengineered heparin. *Metab. Eng.* **14**, 81–90 (2012).
44. Paulson, J. C., Blixt, O. & Collins, B. E. Sweet spots in functional glycomics. *Nat. Chem. Biol.* **2**, 238–248 (2006).
45. Song, X., Heimburg-Molinaro, J., Cummings, R. D. & Smith, D. F. Chemistry of natural glycan microarrays. *Curr. Opin. Chem. Biol.* **18**, 70–77 (2014).
46. Rogers, C. J. et al. Elucidating glycosaminoglycan–protein–protein interactions using carbohydrate microarray and computational approaches. *Proc. Natl. Acad. Sci. USA* **108**, 9747–9752 (2011).
47. Shipp, E. L. & Hsieh-Wilson, L. C. Profiling the sulfation specificities of glycosaminoglycan interactions with growth factors and chemotactic proteins using microarrays. *Chem. Biol.* **14**, 195–208 (2007).
48. Geissner, A. & Seeberger, P. H. Glycan arrays: from basic biochemical research to bioanalytical and biomedical applications. *Annu. Rev. Anal. Chem.* **9**, 223–247 (2016).
49. Kudelka, M. R. et al. Cellular O-glycome reporter/amplification to explore O-glycans of living cells. *Nat. Methods* **13**, 81–86 (2016).
50. Varki, A. et al. Symbol nomenclature for graphical representations of glycans. *Glycobiology* **25**, 1323–1324 (2015).

Acknowledgements

This work was supported by the Lundbeck Foundation, Læge Sofus Carl Emil Friis og hustru Olga Doris Friis' Legat, the Novo Nordisk Foundation, Lund University, Kirsten og Freddy Johansen Fonden, the European Commission (GlycoImaging H2020-MSCA-ITN-721297; BioCapture H2020-MSCA-ITN-722171), the UCPH Excellence Programme for Interdisciplinary Research (CDO2016), and the Danish National Research Foundation (DNRF107) (all to H.C.). C.G was supported through FCT, POPH (Programa Operacional Potencial Humano) SFRH/BPD/96510/2013. A.S., T.M.C., and T.G. were supported through the ERC MalOnco Program, the Danish Cancer Society, and the Lundbeck Foundation.

Author contributions

Y.-H.C., Y.M., Y.N., T.M.C., H.C., and Z.Y. conceived and designed the study; Y.-H.C., Y.M., Y.N., C.G., R.K., C.S., A.M., E.P.B., and Z.Y. contributed with experimental data and data interpretation; T.M.C., C.B.S., T.G., and A.S. contributed to the VAR2CSA studies; A.P., D.W., and U.E. contributed to the xyloside-priming studies; Y.-H.C., Y.M., H.C., and Z.Y. wrote the manuscript; and all authors edited and approved the final version of the manuscript.

Competing interests

The University of Copenhagen has filed patent application EP/2017/061385 on the basis of this work. Y.N., C.S., E.P.B., H.C., and Y.Z. are named inventors on the PCT application.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41592-018-0086-z.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to Y.M. or H.C. or Z.Y.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Cell culture. All gene targeting was performed in CHOZN GS$^{-/-}$ (glutamine synthetase; WT CHO) (Sigma-Aldrich, St. Louis, MO) and/or CHO-K1 (ATCC) cells. CHO GS$^{-/-}$ cells were maintained as suspension cultures in EX-CELL CHO CD Fusion serum-free medium supplemented with 2% L-glutamine in T25 flasks (NUNC; Denmark) or 50-ml TPP TubeSpin Bioreactors (TPP; Switzerland) at 36.5 °C and 5% CO_2 in air. The CHO medium, supplements, and other reagents used were obtained from Sigma-Aldrich unless otherwise specified. HEK293-6E suspension cells were maintained in serum-free F17 medium supplemented with 4 mM L-glutamine, 0.1% Kolliphor P188, and 25 µg/ml G418 in 50-ml TPP TubeSpin Bioreactors at 36.5 °C and 5% CO_2 in air.

gRNA design and screening. We designed three or four gRNAs for all of the GAG-associated genes expressed in CHO cells by using the CRISPR–Cas9 target online predictor tool CCTop (http://crispr.cos.uni-heidelberg.de), inserted them into the gRNA expression plasmid EPB104 (Addgene plasmid #68369), and tested them with a high-throughput workflow as previously described[51]. 2×10^6 CHO cells were cotransfected with 1 µg of gRNA-encoding plasmid and GFP-tagged Cas9-expressing plasmid (Addgene plasmid #68371) each with Fectopro (Polyplus-transfection) in a six-well plate. GFP-tagged CRISPR–Cas9 nuclease was used to evaluate Cas9 expression, and 48 h after the transfection the 10–15% of the cell pool with the highest GFP labeling was enriched by FACS[52]. The gRNA with the highest cutting efficiency was subsequently identified by indel detection by amplicon analysis (IDAA), as previously described[53]. Cutting efficiencies were estimated on the basis of indel IDAA peak sizes relative to the total accumulated peak sizes, and a library of one validated gRNA design for each GAG-associated gene with cutting efficiency > 30% was selected. Most gRNA designs induced ±1 bp, as previously reported[51] (Supplementary Table 1). In total we developed validated gRNA constructs for >30 CHO GAG-associated genes and generated >100 CHO clones with distinct targeting of GAG-associated genes, including clones with KI of human and mouse GAG-associated genes (see below) (Supplementary Tables 1 and 2).

Gene targeting. Cells were seeded at 0.5×10^6 cells/ml in a T25 flask (NUNC; Denmark) 1 d before transfection. For targeted knockout, 2×10^6 cells were cotransfected with 2 µg of endotoxin-free gRNA and a plasmid expressing GFP-tagged Cas9. Transfections were conducted by electroporation with Amaxa kit V and program U24 with Amaxa Nucleofector 2B (Lonza; Switzerland). Electroporated cells were subsequently placed in 3 ml of fresh medium in a six-well plate. 48 h after electroporation, the 10–15% of cells with the highest GFP labeling were enriched by FACS[52]. After 1 week the FACS-sorted cell pool was further single-cell-sorted in round-bottom 96-well plates to obtain single clones. The single-cell clone plating medium contained 80% EX CELL CHO Cloning Medium (cat. no. C6366) and 20% spent medium from cell culture. KO clones were identified by IDAA[53], and mutations of the selected clones were further verified by Sanger sequencing.

The limited GAG biosynthesis capacity of CHO cells provides an ideal basis for the design and engineering of novel biosynthetic capabilities. We used targeted KI with zinc-finger nucleases (ZFNs) to stably introduce genes into a safe-harbor locus that exhibited stable transcription during the cell-culture process[54]. We used a modified ObLiGaRe strategy[11,55] and full coding constructs of human and mouse genes (Supplementary Table 2).

For targeted KI of the human or mouse sulfotransferase-encoding gene to the CHO safe-harbor locus, full coding codon-optimized genes (Genewiz; USA) or cDNAs (Harvard PlasmID Database) were linked to sequences encoding MYC, V5, T7, HA, or OLLAS as a C-terminal tag, and further constructed into donor plasmid EPB69 (for the first gene KI) or EPB70 plasmid (for a second gene KI). These donor plasmids contained two flanking CHO safe-harbor-locus ZFN-binding sites, two tDNA insulator elements flanking the ZFN-binding sites, and a landing pad for a second gene-targeted KI, as previously described[53]. KI was performed by the same transfection method as for the targeted KO, except that 5 µg of DNA and 2 µg of each ZFN were used. Each ZFN was tagged with GFP and Crimson, and 48 h after transfection the 10–15% of cells with the highest labeling for both GFP and Crimson was enriched by FACS. One week later the FACS-sorted cell pool was further single-cell sorted in round-bottom 96-well plates to obtain single clones. Positive KI clones were screened by immunocytology with antibodies specific for MYC (clone 9E10, ATCC CRL-1729), V5 (Thermo Fisher, R963-25), T7 (Novagen, 69522), HA (Santa Cruz Biotechnology, sc-7392), or OLLAS (Novus Biological, NBP1-06713F), and validated by PCR with primers specific for the junction area between the donor plasmid and the safe-harbor locus. A primer set flanking the targeted KI locus was also used to characterize the allelic insertion status.

Disaccharide analysis of GAGs. Disaccharide composition analysis was performed with a panel of bacterial polysaccharide lyases (R&D Systems) and 2-aminoacridone (AMAC) labeling on a Waters Acquity UPLC system equipped with a fluorescence (FLR) detector, as described previously[56]. Briefly, approximately 5×10^6 cells of each individual CHO clone grown at subconfluency were washed in PBS, and the cell pellet was stored at −20 °C until use. Cell pellets were resuspended in a lysis buffer containing 50 mM Tris-HCl buffer, pH 7.6, 10 mM $CaCl_2$, 0.1% Triton X-100, and 1 mg/ml Pronase (Roche) in a total volume of 1 ml. Reactions were incubated at 37 °C overnight with end-to-end rotation, and heated at 100 °C for 10 min to inactivate Pronase; after centrifugation, supernatants were adjusted to 2 mM $MgCl_2$, mixed with 250 U of Benzonase (Sigma-Aldrich), and incubated at 37 °C for 2 h. Reactions were acidified with acetic acid to pH 5.0 and loaded onto a Q-Sepharose column (0.5 ml) (Sigma-Aldrich) equilibrated with 20 mM sodium acetate, pH 5.0, 100 mM NaCl, and 0.1% Triton X-100. The column was washed first with the equilibration buffer and then with the same buffer without Triton X-100 to remove the detergent. The bound GAG fraction was eluted with 1.5 ml of buffer containing 20 mM sodium acetate, pH 5.0, and 1 M NaCl. The eluate was mixed with ethanol saturated with sodium acetate (1:3 vol/vol) and kept at −20 °C overnight, and the GAGs were pelleted by centrifugation. Pellets were dried in a vacuum microcentrifuge (SpeedVac), dissolved in double-deionized water, and stored at −20 °C until use. Each sample was divided into five aliquots (corresponding to approximately 1×10^6 cells) and subjected to digestion with lyases. For evaluation of total CS/DS, chondroitinase ABC (10 mU) was used for digestion, and for the selective evaluation of DS, chondroitinase B (10 mU) was used. Reactions were performed in 40 mM sodium acetate and 1 µM $CaCl_2$ at 37 °C overnight. For the evaluation of HS, a mixture of heparinases I, II, and III, or of II and III only (10 mU of each), was used, with similar results. Reactions were performed in 40 mM sodium acetate and 5 mM $CaCl_2$ at 37 °C overnight. Released disaccharides were freeze-dried and labeled with AMAC as described previously[56]. Briefly, the dried samples were dissolved in 5 µl of 0.1 M AMAC solution in glacial acetic acid–DMSO (vol/vol 3:17) and incubated at room temperature for 15 min, followed by mixing with 5 µl of 1 M $NaCNBH_3$ and further incubation at 45 °C for 3 h. Excess AMAC was removed by acetone precipitation. Aliquots of labeled disaccharides (1/4, corresponding to approximately 2.5×10^5 cells) were subsequently analyzed on a Waters Acquity UPLC system equipped with a fluorescence (FLR) detector. The separation was optimized on a BEH C18 column (2.1 × 150 mm, 1.7 µm; Waters) at 30 °C for CS/DS and 40 °C for HS, with 80 mM ammonium acetate as mobile phase A (pH 5.5) for CS/DS and 150 mM ammonium acetate as mobile phase A (pH 5.6) for HS, and with 100% acetonitrile as mobile phase B for CS/DS and HS. Separation of the disaccharides was performed with a gradient of mobile phase B increasing from 3 to 13% over 30 min at a flow rate of 0.2 ml/min. Each HPLC run corresponded to an injection of disaccharides from approximately 2.5×10^5 CHO cells. Each series of HPLC runs was preceded with standards (20 pmol AMAC-labeled disaccharides; Iduron) and a WT CHO sample for consistency, calibration, and identification.

For analysis of xyloside-primed GAGs, conditioned medium (6 ml) from cells was treated sequentially with Pronase and Benzonase as cell lysates, and GAGs were purified by Q-Sepharose, followed by sequential digestion with lyases, as described for cell lysates. Each HPLC analysis of AMAC-labeled disaccharides was performed with 1/30 of the starting sample (equivalent to 0.2 ml of medium). For analysis of recombinant secreted serglycin, conditioned medium (3 ml) from cells grown for 48 h to approximately 1×10^6 cells/ml was collected, and GAGs were purified and analyzed in a similar manner, except that no Pronase pretreatment was included, and the elution of the proteoglycan was carried out with 2.5 M NaCl. HPLC analysis was performed with 1/100 of the starting sample (equivalent to 0.03 ml of medium). Disaccharide representation codes were used as previously defined[57].

Recombinant expression. The full coding sequence of human serglycin fused to the Fc portion of human IgG[58] was cloned into a pEE-IgG vector (Lonza Biologics) to establish a clone with stable expression in CHOZN GS$^{-/-}$ cells with glutamine selection (Sigma-Aldrich). Transfection to obtain a serglycin clone was performed in the same way as for a KO, except that 8 µg of plasmid was used. Electroporated cells were subsequently placed in 3 ml of fresh medium in a six-well plate. 72 h later, 1,000 cells/well were seeded in 96-well plates. Stable transfectants were selected in glutamine-free medium. Clones with the highest protein expression were screened by direct enzyme-linked immunosorbent assay (ELISA) using biotinylated anti-human IgG (Sigma-Aldrich) as the primary antibody and horseradish-peroxidase-conjugated streptavidin (Dako) as a secondary antibody.

Cell binding assays. 1×10^5 CHO cells were incubated with rVAR2CSA (100 nM), His-tagged mouse FGF2 (6.7 nM) (Sino Biological Inc.), or monoclonal antibody CS-56 (1:100) (Abcam, ab11570) in PBS with 2% FBS for 30 min at 4 °C in 96-well plates. Cells were washed twice with PBS with 2% FBS and incubated with FITC-conjugated anti-V5 (1:500) (Thermo Fisher, R963-25) for rVAR2CSA, anti-His (1:1,000) (Santa Cruz Biotechnology, sc-53073) for FGF2, or FITC-labeled rabbit anti-mouse IgG (1:200) (Dako, F0261) for the CS-56 binding assay, for 30 min at 4 °C. For the FGF2 binding assay, cells were further incubated with FITC-labeled rabbit anti-mouse IgG (1:200) for 30 min at 4 °C. Fluorescence intensity was analyzed on flow cytometers FACSCalibur (BD Biosciences) for rVAR2CSA and CS-56, and LSR (BD Biosciences) for FGF2 and CS-56 binding assays.

Xyloside priming. XylNap and $XylNapNH_2$ were synthesized as described (Supplementary Note), dissolved in DMSO, and further diluted in EX-CELL CHO CD Fusion serum-free medium to obtain a final concentration of 100 µM XylNap and 1 mM $XylNapNH_2$. 2×10^6 cells were treated with 3 ml of the XylNap- or $XylNapNH_2$-containing medium in a six-well plate. The medium was collected 24 h later.

Resource availability. All cell lines are available on request under a standard material transfer agreement with the University of Copenhagen for academic research purposes.

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability. All data generated or analyzed during this study are included in this article and/or associated Supplementary Information files. Raw data files are available upon request.

References

51. Narimatsu, Y. et al. A validated gRNA library for CRISPR–Cas9 targeting of the human glycosyltransferase genome. *Glycobiology* **28**, 295–305 (2018).
52. Lonowski, L. A. et al. Genome editing using FACS enrichment of nuclease-expressing cells and indel detection by amplicon analysis. *Nat. Protoc.* **12**, 581–603 (2017).
53. Yang, Z. et al. Fast and sensitive detection of indels induced by precise gene targeting. *Nucleic Acids Res.* **43**, e59 (2015).
54. Bahr, S., Cortner, L., Ladley, S. & Borgschulte, T. Evaluating the effect of chromosomal context on zinc finger nuclease efficiency. *BMC Proc.* **7**, 3 (2013).
55. Maresca, M., Lin, V. G., Guo, N. & Yang, Y. Obligate ligation-gated recombination (ObLiGaRe): custom-designed nuclease-mediated targeted integration through nonhomologous end joining. *Genome Res.* **23**, 539–546 (2013).
56. Volpi, N., Galeotti, F., Yang, B. & Linhardt, R. J. Analysis of glycosaminoglycan-derived, precolumn, 2-aminoacridone-labeled disaccharides with LC-fluorescence and LC-MS detection. *Nat. Protoc.* **9**, 541–558 (2014).
57. Lawrence, R., Lu, H., Rosenberg, R. D., Esko, J. D. & Zhang, L. Disaccharide structure code for the easy representation of constituent oligosaccharides from glycosaminoglycans. *Nat. Methods* **5**, 291–292 (2008).
58. Czajkowsky, D. M., Hu, J., Shao, Z. & Pleass, R. J. Fc-fusion proteins: new developments and future perspectives. *EMBO Mol. Med.* **4**, 1015–1028 (2012).

nature research

Corresponding author(s): Henrik Clausen

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☒	☐	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	HPLC data were collected with Empower 3 Chromatography Data Software (Waters), in which data analysis was also performed. In flow cytometry experiments, we have employed the BD Bioscience software for data collection.
Data analysis	We used Empower 3 Chromatography Data Software (Waters) for HPLC data analysis and FlowJo for flow cytometry.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data generated or analysed during this study are included in this article and its supplementary information files.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences

Study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample sizes.
Data exclusions	Only data excluded are repeats obtained with multiple engineered CHO clones.
Replication	Multiple CHO clones (1-4) were obtained for each gene engineering event. HPLC based disaccharide analysis of engineered CHO clones performed 2-4 times for several clones with similar results, and data for one clone each shown. FACS analysis of protein binding to engineered CHO clones performed multiple times (2-) and with multiple clones per same gene engineering event (2-3). FACS data shown represents average of 2-3 independent experiemts at a single concentration. All attempts at replication were successful.
Randomization	Not applicable, but for cell binding assays in Fig. 4 clones (1-3) with same genetic engineering were randomly selected from available number of clones.
Blinding	No blinding was performed during this study.

Materials & experimental systems

Policy information about availability of materials

n/a	Involved in the study
☐	☒ Unique materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Research animals
☒	☐ Human research participants

Unique materials

Obtaining unique materials	All materials developed (gRNA constructs, knockin constructs, and engineered CHO cell lines) are available to academic researchers upon request under a standard MTA with University of Copenhagen. ZFNs used were obtained from Sigma-Aldrich, and a separate 3-way MTA have to be established for access.

Antibodies

Antibodies used	Primary antibodies used: FITC-conjugated anti-V5 (Invitrogen, lot #1921451, 1:500), anti-His (Santa Cruz Biothechnology, lot# E2313, 1:1000), anti-CS-56 (Abcam, lot# 11570, 1:100), anti-HA (Santa Cruz Biothechnology, lot# F2513, 1:200), anti-Myc, anti-T7 (Novagen, lot# 69522, 1:1000), anti-OLLAS (Novus biologicals, lot# F-6-082217-F, 1:100), biotinylated anti-human IgG (Sigma, Lot# SLBN8668V, 1:1000) Secondary antibodies: FITC-labelled rabbit anti-mouse IgG (Dako, lot# 20035868, 1:200), HRP-conjugated streptavidin (Dako, lot# 20040879, 1:10000)
Validation	Commercially purchased antibodies against standard protein tags were tested on cell lines transfected with proteins with or without the relevant tag.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	CHOZN® GS-/- ZFN-modified CHO cell line (Sigma-Aldrich), CHO-K1 (ATCC) and HEK293 6E cells (National Research Council, Canada) were used.
Authentication	No specific authentication of cell lines used apart from separate handling of original obtained vials throughout entire project.

Authentication	Each individual engineered CHO clones were however confirmed multiple times by CHO gene specific IDAA and Sanger sequencing in the target gene area(s).
Mycoplasma contamination	A representative set of growing cell lines in the lab selected randomly is subjected to mycoplasma screening bi-monthly, and within the last 10 yrs no infected cells have been found..
Commonly misidentified lines (See ICLAC register)	None of the cell lines used are listed in the ICLAC database.

Method-specific reporting

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ Magnetic resonance imaging

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	CHO and HEK cells were suspended in FACS buffer (PBS plus 2% FBS) containing V5-tagged rVAR2 protein, His-tagged FGF2 or mAb CS-56 and incubated for 30 minutes at 4 degree in 96 well plate followed by FITS-conjugated secondary antibody. After wash the cells were resuspended in FACS buffer and fluorescence intensity were immediately measured.
Instrument	FACSCallibur (BD Biosciences) was used for binding assay with rVAR2 and mAb CS-56. FACS LSR-III (BD Biosciences) was used for mAB CS-56 and FGF2 binding assay .
Software	FlowJo Version 10 was used.
Cell population abundance	Not applicable. Gating was performed only to exclude dead cells and doublets.
Gating strategy	The dead cells were excluded based on forward and side scatter area (FSC-A and SSC-A) parameter. Doublets were excluded based on FSC-H (height) and FSC-W (width) parameter.

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.