

Glycan Metabolism

A validated gRNA library for CRISPR/Cas9 targeting of the human glycosyltransferase genome

Yoshiki Narimatsu^{2,3,1}, Hiren J Joshi², Zhang Yang^{2,3}, Catarina Gomes^{2,4}, Yen-Hsi Chen², Flaminia C Lorenzetti², Sanae Furukawa², Katrine T Schjoldager², Lars Hansen², Henrik Clausen², Eric P Bennett^{2,1}, and Hans H Wandall²

²Copenhagen Center for Glycomics, Departments of Cellular and Molecular Medicine and Odontology, Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3, DK-2200 Copenhagen N, Denmark, ³GlycoDisplay Aps, Blegdamsvej 3, DK-2200 Copenhagen N, Denmark, and ⁴Instituto de Investigação e Inovação em Saúde, i3S; Institute of Molecular Pathology and Immunology of University of Porto, Ipatimup, Rua Júlio Amaral de Carvalho, 45, Porto 4200-135, Portugal

¹To whom correspondence should be addressed: Tel: +45-35335528; Fax: +45-35367980; e-mail: yoshiki@sund.ku.dk (Y.N.); Tel: +4535326630; Fax: +45-35367980; e-mail: epb@sund.ku.dk (E.P.B.)

Received 25 September 2017; Revised 20 November 2017; Editorial decision 5 December 2017; Accepted 7 December 2017

Abstract

Over 200 glycosyltransferases are involved in the orchestration of the biosynthesis of the human glyceme, which is comprised of all glycan structures found on different glycoconjugates in cells. The glyceme is vast, and despite advancements in analytic strategies it continues to be difficult to decipher biological roles of glycans with respect to specific glycan structures, type of glycoconjugate, particular glycoproteins, and distinct glycosites on proteins. In contrast to this, the number of glycosyltransferase genes involved in the biosynthesis of the human glyceme is manageable, and the biosynthetic roles of most of these enzymes are defined or can be predicted with reasonable confidence. Thus, with the availability of the facile CRISPR/Cas9 gene editing tool it now seems easier to approach investigation of the functions of the glyceme through genetic dissection of biosynthetic pathways, rather than by direct glycan analysis. However, obstacles still remain with design and validation of efficient gene targeting constructs, as well as with the interpretation of results from gene targeting and the translation of gene function to glycan structures. This is especially true for glycosylation steps covered by isoenzyme gene families. Here, we present a library of validated high-efficiency gRNA designs suitable for individual and combinatorial targeting of the human glycosyltransferase genome together with a global view of the predicted functions of human glycosyltransferases to facilitate and guide gene targeting strategies in studies of the human glyceme.

Key words: gRNA design, IDAA, gene editing, glycoengineering, glyceme, glycosylation

Introduction

The glyceme comprising all glycan structures found on different glycoconjugates including glycosphingolipids, glycoproteins and

proteoglycans of human cells is vast (Cummings 2009). Glycans serve many and diverse roles in the biology of cells and organisms and part of interactions with the environment (Varki 2017).

Studying the glycome requires detailed information on individual glycan structures and their glycoconjugate context, and in part due to analytic constraints, heterogeneity and complexity often present as factors limiting the deciphering and appreciation of the many fundamental roles of glycosylation. Analysis of the glycome of cells is generally limited to profiling strategies of glycans released from individual glycoconjugates, which to a large extent sacrifices information on the origin of individual glycans with respect to proteins and positions on proteins – the glycosites (Thaysen-Andersen and Packer 2014). In contrast, glycoproteome strategies defining glycosites often sacrifice information on the structures of the attached glycans (Levery et al. 2015).

While the glycome of cells and organs is large and heterogeneous, the biosynthetic and genetic basis for the glycome is considerably less complex (Hansen et al. 2015). Some 200–250 human glycosyltransferases orchestrate the biosynthesis of the diverse types of glycoconjugates and glycan structures produced in human cells (Henrissat et al. 2009). These genes are classified in the CAZy database into 44 GT families (Lombard et al. 2014). This is not considering the additional diversity brought by enzymatic modifications of glycans by sulfation, phosphorylation, acetylation and epimerization, which are carried out by another 100 or more enzymes. Nevertheless, the glycosyltransferase genome, the GTf-genome, is quite limited compared to the glycome, and although the individual enzymes have highly specific functions in biosynthetic pathways their functions in common and repeated structural features produce the great structural diversity of the glycome. The current knowledge of the GTf-genome and the roles of individual enzymes is quite advanced, making it possible with a reasonable degree of confidence to assign individual GTf genes to specific biosynthetic steps (Hansen et al. 2015). Thus, a large part of the GTf-genome has unique functions and can be assigned to the biosynthesis of specific glycoconjugates and glycan structures, and these are mainly involved in initiation, immediate core extension and/or branching of distinct types of glycoconjugates (Figure 1). The remainder of the GTf-genome has broader functions with roles in the biosynthesis of several different types of glycoconjugates, and these enzymes are mainly involved in elongation and capping of glycans. Importantly, the GTf-genome consists of a number of families with closely homologous genes shown (or predicted) to encode isoenzymes with related properties such as, e.g., galactosyl and sialyltransferases (Tsuji et al. 1996; Amado et al. 1999; Narimatsu 2006), and while the contributions of these to the glycome is still poorly understood, it is clear that many of these isoenzymes have at least partly overlapping functions and contribute functional redundancy. Thus, with our current understanding of the genetic and biosynthetic regulation of the human glycome it is not possible to assign every single GTf gene unambiguously to a specific glycan structure. However, we can assign a large part of the GTf-genome unambiguously to distinct glycosylation pathways and even structures, while the many families of isoenzymes may be assigned to more broad glycosylation features shared among multiple glycoconjugates.

Genetic dissection of biological functions of glycans has a long history in the glycosylation field (Conzelmann and Kornfeld 1984; Patnaik and Stanley 2006). In particular knockout of GTf genes in mice and other model organisms have greatly advanced appreciation of essential functions of particular types of glycosylation (Lowe and Marth 2003). However, for detailed dissection and identification of specific structure-function relationships it is advantageous to use more simple cell systems as a first step. For a long time genetic dissection of GTf function in higher eukaryotic cell lines was not

possible, but recently, elegant studies with haploid cell lines have uncovered entire glycosylation pathways required for specific biological functions (Jae et al. 2013), and with the advent of precise gene editing including ZFNs, TALENs and CRISPR/Cas9 (Chandrasegaran and Carroll 2016), it is now possible and rather simple to use a gene editing approach as a tool for dissection of glycosylation in cell lines (Steentoft et al. 2014). Evaluating the role of any particular glycosyltransferase gene generally requires complete loss of function, and explains why techniques that do not reliably achieve complete suppression of transcription (such as RNA silencing) have not been favored in the field. Similarly, complete bi-allelic (and sometimes multiallelic) knockout of most GTf genes are needed to abrogate glycosylation features and evaluate glycome and biological effects.

A key factor for broad use of precise gene targeting is efficiency and specificity, and the CRISPR/Cas9 gene editing tool is currently the most efficient and cost-effective (Chandrasegaran and Carroll 2016). However, the efficiency in CRISPR/Cas9 gene targeting relies on design of gRNAs, and although several prediction tools have been developed including Deskgene (Doench et al. 2014, 2016), CHOPCHOP (Montague et al. 2014), E-CRISP (Heigwer et al. 2014) and CRISPR design (Hsu et al. 2013), the performance of predicted gRNAs vary from no/low to high efficiency (Cong et al. 2013; Hart et al. 2017; Shi et al. 2015). Thus, there is still a need for experimental trial and error to identify gRNA designs for efficient mono and/or bi-allelic gene targeting (Yang et al. 2015a; Lonowski et al. 2017; Metzakopian et al. 2017).

In order to facilitate GTf-genome wide dissection and discovery, we have therefore designed and experimentally tested a collection of more than 600 gRNAs for all known human GTf-genes. We tested four gRNAs for each glycogene with a high throughput workflow involving FACS and Indel Detection by Amplicon Analysis (IDAA) (Lonowski et al. 2017), which enabled us to select one validated gRNA for each gene with high multi-allelic cutting efficiency and predictable small indels resulting in coding frameshifts. To facilitate broad use of this resource, we provide a comprehensive protocol for CRISPR/Cas9 gRNA delivery, screening, sorting and selection of gene targeted cells, as well as a predicted framework for interpretation of effects of single and combinatorial targeting of GTf genes on the cellular glycome.

Results

Overview of the GTf-genome and predicted roles in glycosylation pathways

To date 208 human glycosyltransferase genes have been identified and their catalytic properties characterized to some detail or in a few examples predicted based on close sequence similarities (Table I) (Hansen et al. 2015). In Figure 1 we present a graphic view of 167 human GTf-genes with their predicted functions in the known glycosylation pathways found in human cells. Glycosylation pathways are organized into pathway-specific (vertical colored boxing) and non-pathway-specific (horizontal colored boxing) sequential biosynthetic steps of glycosylation that includes initiation, core extension, branching and elongation, and terminal capping. The underlying glycan structures characteristic for each type of glycosylation and those common to multiple glycosylation pathways are presented in Supplementary data, Figure S1. The organization provides important guidance for predicted consequences of gene editing of human GTfs for the glycosylation capacity of cells. Thus,

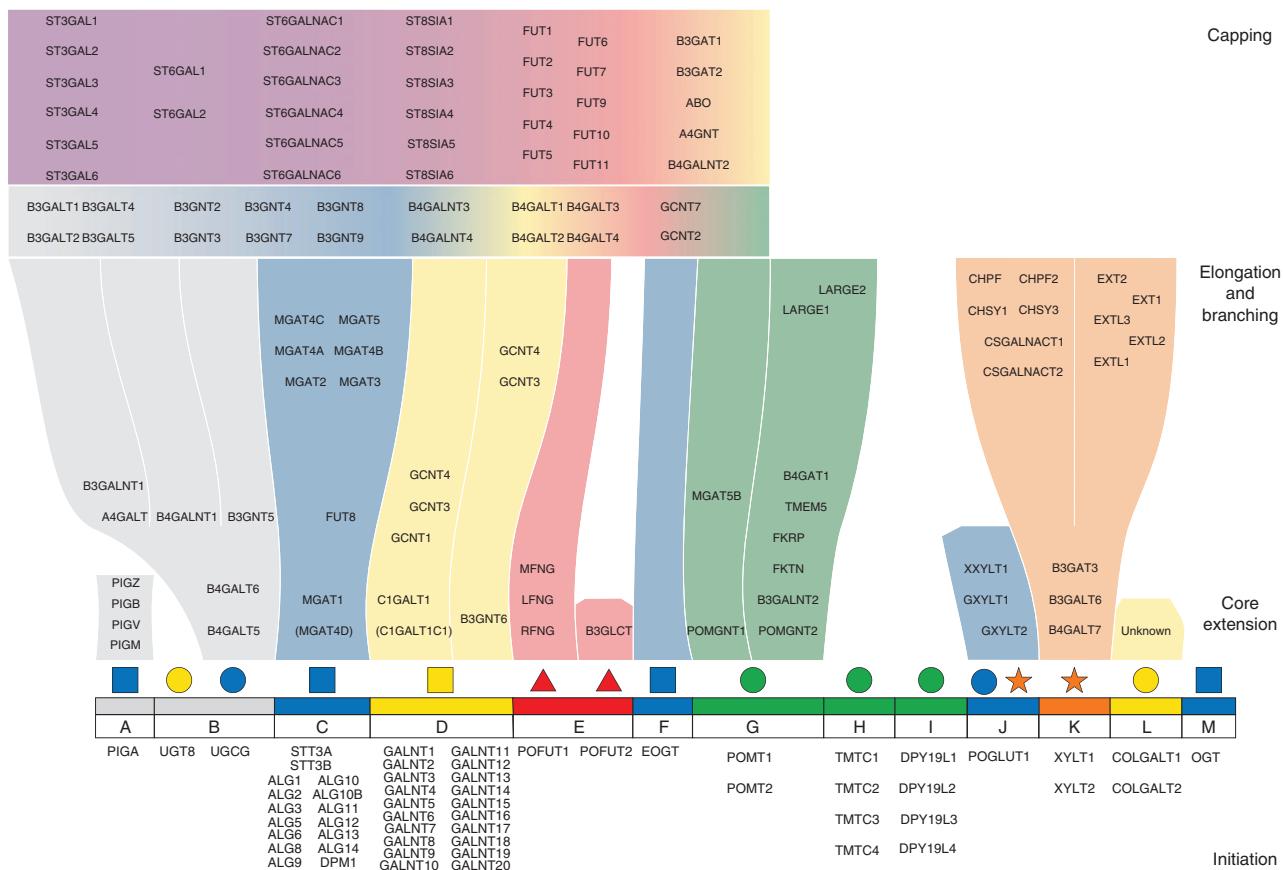


Fig. 1. Graphic rainbow depiction of predicted assignments of the 167 human GTf-genes involved in biosynthesis of protein or lipid glycan structures. Glycosylation pathways from left to right are: (A) GPI-anchor, (B) glycolipids (two pathways), (C) N-linked glycans, (D) O-GalNAc mucin-type, (E) O-Fuc type (two pathways), (F) O-Man type (POMT-directed), (G) O-Man type (TMTC-directed), (H) C-Man type, (I) O-Glc type, (J) O-Xyl type (proteoglycans), (K) O-Gal type (collagen) and (L) two types of O-GlcNAc (extracellular and cytosolic). The horizontal first rainbow level illustrates GTf-genes predicted to perform glycosylation pathway-specific functions in the initiation, core extension and branching/elongation steps of the indicated types of glycosylation. The two horizontal rainbow levels illustrate GTf-genes with predicted pathway-unspecific functions in the elongation or capping steps that are common among the multiple glycosylation pathways beneath. Glycan symbols are drawn according to the SNFG format (Varki et al. 2015).

targeting pathway-specific GTf-genes is predicted to only affect one type of glycoconjugate, while targeting non-pathway-specific GTfs is predicted to affect multiple types. Moreover, the organization illustrates the homologous GTf-gene families encoding isoenzymes with potential overlapping functions, where it is important to consider combinatorial targeting for complete abrogation of glycosylation.

Construction of a human GTf-genome gRNA library

The design of gRNAs is important for the efficiency of gene editing (Yang et al. 2015a; Lonowski et al. 2017; Metzakopian et al. 2017). The target site for indels should obviously be selected to eventually disrupt protein function, and in many large scale screening studies the first coding exon(s) is selected to ensure truncated protein products without having to discriminate individual genes and target functional domains (Doench et al. 2014; Shalem et al. 2014; Wang et al. 2014). Here, we chose the same strategy partly due to the finding that the first exons are the least homologous regions of most GTf-genes including close paralogs. We first tested several gRNA prediction algorithms, and the Desktop Genetics DESKGEN Cloud gRNA design platform (<https://www.deskgen.com/landing/>) was selected for a pilot study with 48 gRNA designs. DESKGEN provides both on-target as well as off-target scores with

assignment of potential off-target sites and similarity with the target sequence (number and position of mismatches). This is valuable for targeting close homologs in isoenzyme gene families, such as the β 4 galactosyltransferase family as previously reported (Duda et al. 2014). As shown in Supplementary data, Figure S2 the pilot study suggested that the on-target scores did influence the cutting-efficiency and the range of 31–50 was optimal.

We proceeded with design and testing of 3–4 gRNAs for all human GTf-genes with a high throughput workflow as illustrated in Figure 2. gRNAs were designed to target the first third of the coding region within an early exon with DESKGEN parameters set for on target score >30, and with off target score set >80 with no predicted off targets with 0 or 1 mismatches and minimal number off targets with 2 or 3 mismatches with gRNA target sequence (Supplementary data, Table SI). gRNAs were tested with human embryonic kidney HEK293T cells and the cutting efficiency and indel profile of each gRNA was characterized by IDAA (Yang et al. 2015a; Lonowski et al. 2017). We and others have shown that an effective gRNA design may induce highly similar indel profiles across most of the commonly used cell lines including iPSC's (Paquet et al. 2016; Lonowski et al. 2017). We used a GFP-tagged CRISPR/Cas9 nuclelease to evaluate Cas9 expression and enrich for different expression

Table I. GlycoCRISPR: the list of validated gRNA target to human glycosyltransferase

¹ CAZy family	Gene name (HGNC)	gRNA sequence	² Target exon	³ Major indel	¹ CAZy family	Gene name (HGNC)	gRNA sequence	² Target exon	³ Major indel	¹ CAZy family	Gene name (HGNC)	gRNA sequence	² Target exon	³ Major indel
N-glycosylation														
GT33	ALG1	GCAATGCTAGGCAGACCTGG	4/13	+1	GT14	XYLT1	ACAAACAGCAACTTCGCACCC	3/12	+1	GT31	B3GA ^{PT} T1	TCAGGCCACCTAACAGTTGCC	1/1	+1
GT4	ALG2	TGTTTCAGGCTGGCTAGACGG	2/2	-12/-8/+1	GT14	XYLT2	GACAGTTCACGAGGGCAGC	2/11	+1	GT31	B3GA ^{PT} T2	CCTTGACATACACTTTCGG	1/1	+1
GT58	ALG3	GATCTATCACCAGACCTGCA	3/9	+1	GT7	B4GALT7	TGACACTGCTCCCTCTCAACG	3/6	+1	GT31	B3GA ^{PT} T4	GGAAGCTTGCAGTGGTCCCG	1/1	-1
GT2	ALG5	GTAGGTGAGTCCCATATGCT	2/10	+1	GT31	B3GALT6	CTTCGAGTTCTGCTCAAGG	1/1	+1	GT31	B3GA ^{PT} T5	TTTCCCCCACGCTCCGGGA	1/1	+1
GT57	ALG6	AGACTCTTCCCGGTGATCG	8/14	+1	GT43	B3GAT3	TCCGAGTTCCGCTTGAGCT	2/5	+1	GT7	B4GA ^{PT} T1	GGAGTCTCCACACCGCTGCA	1/6	+1
GT57	ALG8	GTGCGGTGCCGAGCAATGG	1/13	+1	GT7	CSGALNACT1	GGTACCCCTCTTCCCCGTG	1/7	+1	GT7	B4GA ^{PT} T2	AGTAGAGGATGACGCCACG	2/7	+1
GT22	ALG9	CGTTGATAGCCATGACTGGA	6/15	+1	GT7	CSGALNACT2	GTATTATCAAGCCCTCTAC	1/7	+1	GT7	B4GA ^{PT} T3	TCCCTGATCTCGGCCAAATA	1/6	+1
GT59	ALG10	CTGCCATTGGATCTTGG	2/3	+1	GT7	CHPF	GGAACACCACACGCTCCAGC	2/4	+1	GT7	B4GA ^{PT} T4	ACCCACGAAAGTAGTTACTGG	1/6	+1
GT59	ALG10B	GCAGGAATGGCGCAGCTAGA	1/3	+1	GT7	CHPF2	CAGTGAAGTAGAGTAACCGA	2/4	+1	GT31	B3GN ^{PT} T2	GTTCCAGTATGCCTCGGGAG	1/1	-1
GT4	ALG11	GCAGCAGTCTGATCCCCAA	2/4	-2/+1	GT7	CHSY1	GTACATCAAAGGAGACCGTC	2/3	-11/-1/+1	GT31	B3GN ^{PT} T3	GCAGCCACCGGAGATCCCCG	1/2	-2/-1
GT22	ALG12	GCGAAAGCACGTAACCGCG	2/9	+1	GT7	CHSY3	TCTCGGCTAAAGATCATGCC	2/3	+1	GT31	B3GN ^{PT} T4	GTATCCTTGGAACAGCCTGA	2/2	-1/+1
GT1	ALG13	AATGGCCGAAAGAGGCAAG	4/27	-7/+1	GT64	EXT1	GTAGAACCTGGAGCCCTCGA	1/11	+1	GT31	B3GN ^{PT} T7	GCTCTGCGAGAACTGACCA	1/2	+1
GT1	ALG14	AAGATACTGAGAGACTCCCG	1/4	-2/-1	GT64	EXT2	TCGGCTGGCAGCCTAACAC	1/13	+1	GT31	B3GN ^{PT} T8	GGAGTGTGAGCAGTGTGAC	1/1	+1
GT2	DPM1	GTAGTCCTCCGAGGTCTCGG	1/9	+1	GT64	EXTL1	GTAGAAGCGAGACCCCTCAA	1/11	+1	GT31	B3GN ^{PT} T9	AACCAGCGCACAAGTGCACG	1/1	+1/+2
GT23	FUT8	ACCTTGTGTTTATATAGG	1/9	+1	GT64	EXTL2	CAGCTACCAGTAATAATACG	1/4	+1	GT7	B4GA ^{PT} T3	CATAGATTGACACACCCCG	5/20	-1/+1
GT13	MGAT1	CCCTCAGTCAGCGCTCTGA	1/1	+1	GT64	EXTL3	GGTGGGAAACGAGCTGTGCG	1/5	+1	GT7	B4GA ^{PT} T4	CAGTGGACGACGCCGG	2/20	-1/+1
GT16	MGAT2	GTTCGGCTGCAGCAACGGT	1/1	+1	O-GlcNAc					GT14	GCNT ^{PT} 12	ATAGCAGGTAGCTCATCAA	1/3	+1
GT17	MGAT3	TCCTCGGCCGCGCTGCTGG	1/1	+1	GT61	OGT	GCAACCTATTCTCTCTAAC	12/22	+1	GT14	GCNT ^{PT} 16	GGTCAAATAGATGACGTAGG	1/1	+1
GT54	MGAT4A	CTTGTCTTGTATACTACA	1/15	+1	GT41	EOGT	GTTTCAGCTATGTGACAT	2/15	+1	GT14	GCNT ^{PT} 22	ACTGCTCCAGGATTCTCGG	1/3	+1
GT54	MGAT4B	GGAGAGCCTCAAGCGCTCA	2/15	+1	O-Fucose					Redundant (Capping)				
GT54	MGAT4C	TGTGGCAGCTAGGTAGCGAT	2/3	+1	GT65	POFUT1	AAAGCTGCTAACCGTACCT	2/7	-5/+1	GT11	FUT1 ^{PT} 1	CAGGGTATGCGGAATACCG	1/1	+1
GTnC	MGAT4D	GGTATTTCCACTGTTAACAG	4/11	+1	GT68	POFUT2	GGAAGGTTCAACCTCGC	2/9	+1	GT11	FUT2 ^{PT} 1	AGTGCCTAGCTAACATCAA	1/1	+1
GT18	MGAT5	GCTGTCTGACTCCAGCGTA	1/16	+1	GT31	MFNG	GCCCAAGCGTGGAAAGCCC	1/8	-2	GT10	FUT3 ^{PT} 1	TGTCGTAGCAGGATCAGGA	1/1	+1
GT66	STT3A	ATGTTGTTCTTAGCATAAAGA	6/14	+1	GT31	LFNG	GATGAAGCGGTCTACTCCA	3/8	+1	GT10	FUT4 ^{PT} 1	CCTCCACGCGCTGCGGACGCG	1/1	+1
GT66	STT3B	TTGGGTGTTACTAGCTG	7/16	+1	GT31	RFNG	GCCACCCCTGGACCCCTCGG	4/8	+1	GT10	FUT5 ^{PT} 8	GGCAGTGGAACTGTGACCG	1/1	+1
GT24	UGGT1	CTACTATCATCAATATTGG	4/41	+1	GT31	B3GLCT	GTTCAGCTGATGAAGG	5/15	-1/+1	GT10	FUT6 ^{PT} 10	GGACCCATTAGGTACACAG	1/1	-2/+1
GT24	UGGT2	TTCGCAGCTCGGCCGGGA	1/39	+1/+2	O-Mannose					GT10	FUT7 ^{PT} 10	TAGCGGGTGCAGGTGTCGCT	2/2	+1
O-GalNAc					GT39	POMT1	GAGCTCAAACACTATCTGGT	4/19	+1	GT10	FUT9 ^{PT} 9	GTGAACGGTCCGGTTGTGAGA	1/1	-2
GT27	GALNT1	TCCCCACTGTACACTCACAA	4/11	+1	GT39	POMT2	CTTCGAGGCGGTGCGCTGGT	1/21	+1	GT10	FUT1 ^{PT} 20	GGGACCAACAGACATAATG	2/4	-1/+1
GT27	GALNT2	GTGAAACGTGATCACCACG	4/16	+1	GT13	POMGNT1	GAGGGACACATGGGCTTCG	6/21	-10/-1/+1	GT10	FUT1 ^{PT} 20	CGCCAGCTCTGGGACGCC	1/3	+2
GT27	GALNT3	TATGGAAGTAACCATAACG	4/10	+1	GT61	POMGNT2	ACTGAGGATCGACTACCGA	1/1	-1/+1	GT29	ST3GAL1	TCCAAGTCGATGGTCTTGA	3/6	+1
GT27	GALNT4	AACAGTGGCTATATCTCG	1/1	+1	GT18	MGAT5B	CCACCAAGGTACCTCTGTG	2/17	+1	GT29	ST3GAL2	GTGCGTGTCAAACCGTCG	1/6	+1
GT27	GALNT5	GATGACTTCGATTACTGGAC	4/10	-1	GT31	B3GALNT2	GAAAGCTGATAAGAACACC	2/12	+1	GT29	ST3GAL3	GATTCTAGCCACTTGC	5/11	+1
GT27	GALNT6	GAAGAGCAAGTGGACCCCC	1/10	-2	GTnC	FKTN	GAGTCTATCCGCTAGCCG	4/9	+1	GT29	ST3GAL4	TTACCCGCTTCTTACACTC	7/10	-4/+1
GT27	GALNT7	ATGCCAACCGAGGCGGAA	2/12	+1	GTnC	FKRP	GCACCAAGGACGGTGACACGG	1/1	-8/-1/+1	GT29	ST3GAL5	ATTGAGCACAGGTATAGCG	4/7	+1
GT27	GALNT8	GTAGTCTCGCGTGTGGG	2/11	+1	GT49	B4GAT1	CATCGGCCACAGCATGAGCG	1/2	+1	GT29	ST3GAL6	TGAGAACACTGTGCTTAA	6/9	+1
GT27	GALNT9	CGCACTGTCGCGGATCCGAG	5/11	+1	GT49	LARGE	CCTGGAGGTGCGCATGCG	2/14	+1	GT29	ST6GAL1	TGTATCCCTCAAGCAGCACCC	1/5	+1
GT27	GALNT10	CTCTCTCAGCATCGGTATG	3/12	+1	GT49	LARGE2	CCAGAGCTCCGAGATGTG	4/13	+1	GT29	ST6GAL2	AGCTGGTACAGGCTCAGCG	1/5	+1
GT27	GALNT11	TATGCTTATCAGTGACCG	2/11	+1	GTnC	TMTC1	GACCTGCCAGTCAAGCACA	6/18	+1	GT29	ST6GALNAC1	ACGGTGTAGAGAACACCA	2/9	+1
GT27	GALNT12	TTCTCTAGCAGGATATCG	2/10	+1	GTnC	TMTC2	GCCTGAACCATGCCATTGGA	2/12	+1	GT29	ST6GALNAC2	GAGCCCCCGCAGGCCATAG	3/9	+1
GT27	GALNT13	TTAATACGTGCCGCTTCG	4/11	+1	GTnC	TMTC3	ACTGCTGGACAGTTCTCCG	5/13	+1	GT29	ST6GALNAC3	GTATCCATAGTGAGTTGAA	2/5	+1
GT27	GALNT14	CTTACAGGACTACACGCCG	4/15	+1	GTnC	TMTC4	GCTGCGTGCACAAACACAA	1/17	-2/+1	GT29	ST6GALNAC4	TGTTGAGACGACACGCC	3/5	+1
GT27	GALNT15	GCTGGCTGAGGTGCGTACCG	2/10	+1	C-Mannose					GT29	ST6GALNAC5	CTAGTGTACAGCAGCCTCG	2/5	+1

GT27	GALNT16	GTAATGGCGGGTGTCCCGGA	2/15	+1	GT98	DPY19L1	TCTCCCTTCCAATGTTGAG	3/22	+1	GT29	ST6 ¹ ALNAC6	TCTTCATTACGGCTCCCTG	3/6	-1/+1
GT27	GALNT17	AGTCACATTGACCTCGCAG	5/12	-2/-1	GT98	DPY19L2	AGCCAGTCTAAGGGCGGCG	1/22	-9/+1	GT29	ST8 ¹ IA1	CGTTGGCAGCCGGTAGACG	1/5	+1
GT27	GALNT18	TGAGATAGAAAGAGTACCCGC	5/11	-1/+1	GT98	DPY19L3	GCTGGCTACTCAGTGGTACA	5/18	+1	GT29	ST8 ¹ IA2	TGCCATCGTGGCAACTCGG	4/6	+1
GT27	GALNT19	AGGTTGGCACCTCGCGCAG	1/11	+1	GT98	DPY19L4	TGTCTTGAGCGTTACTAG	3/19	+1	GT29	ST8 ¹ IA3	GATGAGCATAAAATCAGCA	1/4	+1
GT27	GALNT20	ATACTCTGTTCACCTCACAG	5/8	+1	O-Galactose					GT29	ST8 ¹ IA4	AGATGCGCTCCATTAGGAAG	1/5	+1
GT31	C1GALT1	GTAAAGCAGGGCTACATGAG	2/3	+1	GT25	COLGALT1	GAAGAGTTGTACCATTCCG	2/12	+1	GT29	ST8 ¹ IA5	ATACAGGATCTGTCAGCA	1/8	+1
GT31	C1GALTC1	GTAGGTGATGATGCTCATGG	1/1	+1	GT25	COLGALT2	GTAGAAAAGTCAGCTTGTCCG	5/12	+1	GT29	ST8 ¹ IA6	GGAGCGCTCTAGCGCTGCG	2/8	+1
GT14	GCNT1	TAGTCGTAGGGTGTCCACCG	1/1	+1	Hyaluronan					GT43	B3G ¹ AT1	GCTAGGATGTCGGCTCTT	1/4	+1
GT14	GCNT3	ACTGTTAGGGGTACCCGA	1/1	-1/+1	GT2	HAS1	CATGGTCGACATGTTCCGCG	2/5	+1	GT43	B3G ¹ AT2	AAAAGCAGGGTAAAAGCG	1/4	-1/+1
GT14	GCNT4	GCAGCCATAGGGTTAAACAC	1/1	-2/-1	GT2	HAS2	TGAAAAGGCTAACCTACCC	1/3	-5/-2/-1	GT32	A4G ¹ NT	CTACGGAACAGGAGACAAA	1/2	-1/+1
GT31	B3GNT6	GCGGGGACCTTGGCGCTGG	1/1	-7/+1	GT2	HAS3	GGTGGGTGATGGCAACCGCC	1/3	-1/+1	GT6	AB6 ¹ 6	AATGTGCCCTCCAGACAAT	6/7	+1
Glycosphingolipid (GSL)														
					Glycogen					GT12	B4GALNT2	CCGTGGACTGGTACCCAAA	4/11	-2/-1/+1
GT21	UGCG	TTAGGATCTACCCCTTTCAG	2/9	+1	GT8	GYG1	GACCAGGGCACCTTGGCGT	2/8	-2/-1	Unknown				
GT1	UGT8	TGGTACTGTTAAAGATCCCT	1/5	-1	GT8	GYG2	GAGAGTCCAACAGTGAAGCT	4/11	+1	GT2	B3G ¹ NT1	CAGTCCACAAACGCTGAACCG	2/12	+1
GT7	B4GALT5	TTCGGAGTGCTTATGCCAAG	2/9	+1	GT3	GYS1	GACGAAGGCGAAGGTGACAG	2/16	-7/+1	GT25	CER ¹ CAM	GTCCTCCCTTCAGCCCCCA	5/12	+1
GT7	B4GALT6	CTCTTTATGGTACAAGCTCG	2/9	+1	GT3	GYS2	ACTGTAAGAAGAATGCTTCG	5/16	+1	GT4	GL ¹ TD1	GCTCTTCATCTATAGGGG	2/8	-2/-1
GT32	A4GALT	GTC ¹ TGACCCCTGTCATCAT	1/1	+1	GPI anchor					GT6	GL ¹ TD1	GGGAAGGGACTTTCGACAGG	3/5	+1
GT31	B3GALNT1	CGTTCTATCACATTGAGTG	1/1	+1	GT4	PIGA	CCGTACCCATAATATATGCA	1/5	+1	GT8	GL ¹ TD1	GCTCTCCGACATGCACTAGA	3/9	+1
GT31	B3GNT5	CTCTTAAGCACACCTCAGCG	1/1	-7/+1	GT22	PIGB	TACCTTGACTTCAACACCC	1/12	-2/+1	GT8	GL ¹ TD2	GCTATTGATGGCAGCCATAG	3/9	-12/+1
GT12	B4GALNT1	ACCGGGATGTTGCGCTAGCG	1/10	+1	GT50	PIGM	CCCTTCGGTACCGAACGGCG	1/1	+1	GT4	GT ¹ D1	TCAGAGTGTGATACACACTG	4/9	+1
GT6	GBGT1	GTTGGCGCCCATCGCTCCG	5/6	+1	GT76	PIGV	AGTTGGTCCACAAAGCCTGA	2/3	+1	GT90	KDE ¹ C1	CTGAATATAGAAATAGCGGG	1/10	+1
					GT22	PIGZ	GCACATAGCCCCCTCGCGGA	1/2	+1	GT90	KDE ¹ C2	GATTGACTACGACTTTGAA	2/8	-1/+1
O-Glucose														
GT8	GXYLT2	ACCCGAGCTCTGGATCCACC	2/7	+1										
GT90	POGLUT1	GAGGATCTAACCTTTCCG	3/11	+1/+2										
GT8	XXYLT1	GCAGTGAGCGCAGCGCAGC	1/4	+1										

¹GT family according to the CAZy database (www.cazy.org).

²The targeted exon out of the total numbers of exons assigned according to UniprotKB/Swiss-Prot information.

³Indication of the indel profile obtained by IDAA. Predominant indel present >30% of total is indicated. For lower indel representations, only the three most predominant indel events are indicated.

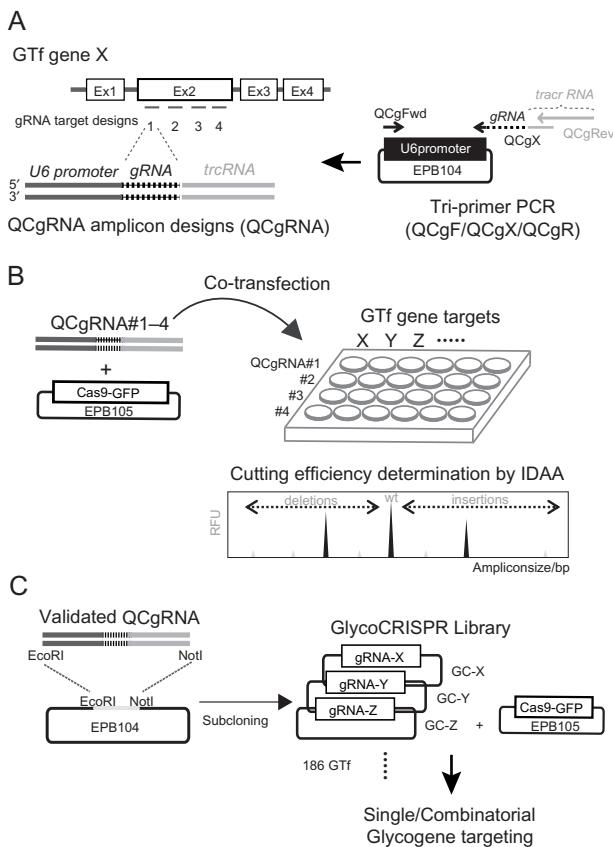


Fig. 2. Schematic workflow for GlycoCRISPR library construction and validation. **(A)** Desktop Genetics DESKGEN Cloud gRNA design platform was used for gRNA design and the most 5'-exons of the GTf-genes were preferentially selected as target site. Amplicons encoding the GTf specific gRNA sequence were amplified using the invariant QCgFwd and QCgRev primers and the variable QCgX primers. The generated QuickChange gRNA amplicon expression cassettes (QCgRNA) were generated by a single tri-primer PCR reaction and a plasmid template only encoding the U6 promoter sequence. **(B)** Individual QCgRNA amplicons were co-transfected with Cas9-GFP plasmid into HEK293 cells and seeded into 24-wells. Two days post-transfection crude lysates were used for amplification across the respective GTf gRNA targeted loci using our tri-primer labeling methodology for single reaction fluorescence labeling of amplicons, followed by indel profiling by IDAA. **(C)** IDAA allows for simultaneous profiling and quantification of indel events, and gRNA designs giving rise to greatest indel formation efficiencies were subcloned into the EPB104 plasmid. The selected constructs were included in the validated GlycoCRISPR toolbox with GC-“gene names”, for use in individual or combinatorial engineering experiments.

levels as a measure of transfection efficiency by FACS (Figure 2B). Cellular transfection of all gRNA designs tested was initially based on our amplicon delivery (QCgRNA) method that only requires co-delivery of Cas9 plasmid with a PCR derived gRNA encoding template to cells (Lonowski et al. 2017). The QCgRNA delivery method avoids laborious and time-consuming cloning, screening and sequencing of gRNA plasmid designs, and allows for fast evaluation of the indel formation efficiency of individual gRNA designs.

IDAA profiles were obtained for unsorted, medium and high GFP positive FACS sorted cells as illustrated in Figure 3. Cutting efficiencies were calculated based on individual indel IDAA peak sizes relative to total accumulated peak sizes (including wildtype-allele). For the validated gRNA library we selected cutting efficiencies >30% derived from medium sorted cells and preferentially with

small out of frame causing indels, and in most cases it was possible to select gRNA designs inducing +/-1 bp indels as these are by far the most common and ensure in-exon frameshifts (Table I). Cutting efficiencies and indel profiles for all gRNAs are listed in Supplementary data, Table SI. All primers and conditions for the target-specific IDAA assays are designed and validated as listed in Supplementary data, Table SII.

A sustainable community resource

To enable wide distribution of the gRNA library to the scientific community, we cloned the selected optimal QCgRNA for each GTf-gene into the EPB104 plasmid backbone (available from Addgene, <https://www.addgene.org/>, see “materials and methods” section for detailed information), and for ease we designated these GC-“gene name” for the GlycoCRISPR library (Figure 2C). We previously demonstrated that delivery of gRNAs as amplicons or plasmids produced similar cutting efficiencies and indel profiles (Lonowski et al. 2017). Moreover, all parameters for the IDAA based screening and clone selection are listed in Supplementary data, Table SII. The pre-designed IDAA assay is a simple protocol that can be used in most labs without major investments, and it is now also becoming commercially available as a custom service by several vendors.

Multiplex gene targeting example

Access to validated plasmid gRNA targeting constructs also enables rational multiplex targeting, which is often required to engineer glycosylation capacities of cells due to isoenzymes with partially overlapping functions as illustrated in Figure 1 (Hansen et al. 2015). The quantitative indel profiling by IDAA further enables easy and fast screening and selection of clones with desirable multiple gene editing events. To illustrate this we used two optimal gRNA plasmids to simultaneously target two polypeptide GalNAc-transferase isoenzyme genes, GALNT2 and T3 in HEK293 cells (Figure 4). We monitored efficiencies at the two target sites by IDAA profiling and immunocytoLOGY after FACS enrichment and single cell cloning. IDAA profiling of the transfected pool clearly demonstrated +1 indel formation for both target sites (~20%), and FACS enrichment resulted in substantial increase in the +1 indel peak. Only one example of a single cloned cell is shown with a clear bi-allelic +1 indel in GALNT2 and compound heterozygote bi-allelic -1/+1 indel in GALNT3, however, most tested single cells in fact contained bi-allelic indels in both targeted genes. While IDAA profiling of heterogeneous cell pools does not provide unambiguous information of the allelic targeting state of the individual cells, we demonstrate by immunocytoLOGY using monoclonal antibodies specific for GalNAc-T2 or T3 that the FACS enriched pool contains about 50% cells without immunoreactivity suggesting that most targeting events in fact are bi-allelic for both genes in a subpopulation of cells.

Discussion

Here we provide a ready to use resource for efficient and validated gene targeting in the GTf-genome with a comprehensive set of plasmid targeting constructs, optimized screening and selection protocols, and a predicted framework to guide engineering of the glycosylation capacity of human cell lines. While gene targeting is becoming a house-hold procedure in most cell biology labs there are still obstacles to overcome and efforts that can be minimized. In particular we have identified the following factors that limit dissection of complex biosynthetic pathways like glycosylation: (i) need for

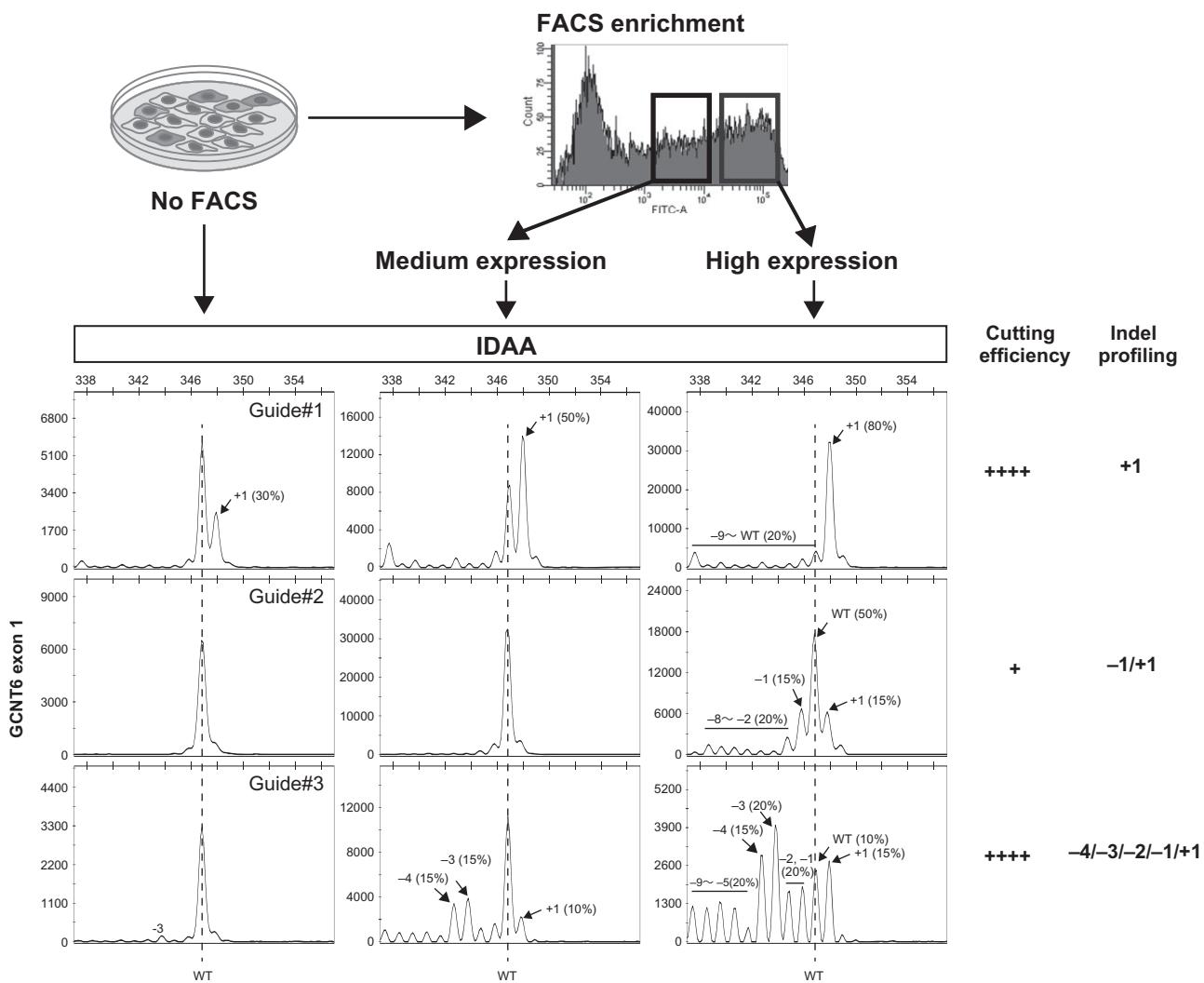


Fig. 3. Schematic illustration of the gRNA selection and validation strategy using FACS enrichment and IDAA monitoring. Shown are IDAA profiles for three different QCgRNAs designs (Guide #1–3) monitored in the original pool of transfected HEK293T cells (48 h postransfection), and after FACS enrichment for cells with medium or high Cas9-GFP expression. The target gene was *GCNT6*, and Guide#1 displayed the highest cutting efficiency with a rather homogenous +1 indel, which was easily detectable even in the original cell pool. Guide#2 displayed low efficiency essentially only detectable after enrichment for high GFP expression, while Guide#3 displayed high efficiency but rather heterogeneous indel formation.

testing multiple guides to identify efficient gRNAs despite improved prediction algorithms; (ii) need for expertise in design of targeting regions in GTf genes to ensure functional inactivation; (iii) need for facile screening and isolation of gene targeted cells with mono and/or bi-allelic inactivating indels; and (iv) need for expertise with translation of individual gene functions to the complex biosynthetic pathways of glycosylation. We believe the developed resource, that we name GlycoCRISPR, meets most of these needs. Importantly the framework for design and interpretation of gene targeting of the complex glycosylation pathways will continue to improve with wider use.

The efficiency and ease of the CRISPR/Cas9 tool for gene targeting makes this the clear choice for broad genome screening strategies, and predesigned gRNA targeting constructs for all human genes as well as entire CRISPR libraries designed for whole genome targeting are now available (Miles et al. 2016). However, the study of glycosylation in cells can be greatly optimized and facilitated by access to individual gRNA targeting constructs for all GTf-genes

that are pre-validated for high efficiency and simple indel formation (+/-1 bp) ensuring loss of function by in-exon frameshifts (Figure 3). Recent studies show that a given gRNA design gives rise to highly similar CRISPR/Cas9 induced indels in most commonly used cell lines, such as K562, HEK293 and iPSC (Paquet et al. 2016; van Overbeek et al. 2016; Lonowski et al. 2017). Thus, the use of validated gRNAs facilitates screening and cloning especially in combinatorial engineering experiments. Genetic targeting strategies to decipher functions in glycosylation pathways present special considerations and limitations. The phenotypic read-out is rarely the enzyme or protein encoded by the targeted gene as, e.g., illustrated in Figure 4, but rather changes in glycosylation and glycan structures. Complete inactivation of targeted genes is almost always required to affect glycosylation detectably (Steentoft et al. 2014), and it is often necessary to target multiple genes to affect glycosylation covered by overlapping isoenzymes and/or alternative pathways such as galactosylation, fucosylation and sialylation (Yang et al. 2015b). Probes to detect loss and gain of specific glycan structures

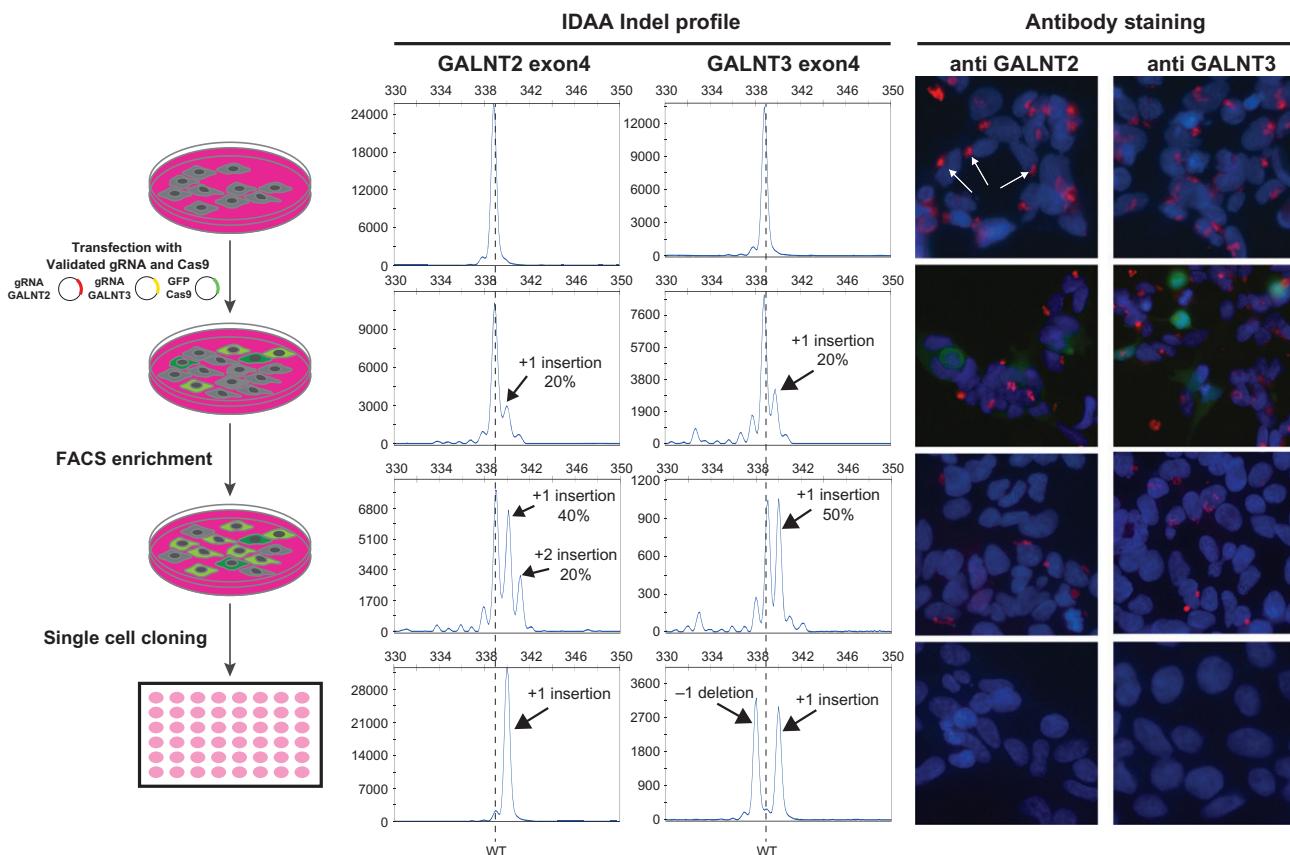


Fig. 4. Multiplex targeting of two GTf-gene genes in HEK293 cells monitored by IDAA and immunocytoLOGY. Simultaneous targeting of *GALNT2* and *T3* with validated GC-*GALNT2/T3* plasmids and Cas9-GFP was monitored before transfection (upper panels), in the transfected cell pool (72 h posttransfection), after FACS enrichment for medium GFP expression, and after single cell cloning (lower panels). IDAA profiling demonstrated clear detection of indels in the cell pool with increase after enrichment, and in one final clone tested absence of wild type alleles. ImmunocytoLOGY at each stage confirmed selective loss of protein expression in subpopulations of cells, and complete loss of expression of both enzymes in the final single cloned cell. GalNAc-T2 and T3 are endogenously expressed in HEK293 wild type cells with supranuclear Golgi-like localization (red color) as indicated by white arrows. Cas9-GFP expression (green) can be seen in the transfected cell pool after 72 h.

greatly simplify the task (Steentoft et al. 2011). This was elegantly illustrated by the whole genome screen for the Lassa virus glycan receptor in the THP-1 haploid cell line, which identified a large number of GTf-genes and other related enzymes acting in successive steps in biosynthesis of the uniquely complex glycan receptor for laminin on alpha-dystroglycan (Jae et al. 2013). In most cases probes are not available though, and in general comprehensive analysis of the glycome and glycoproteomes of cells are too daunting tasks to be used for screening and selection in larger gene engineering experiments (Thaysen-Andersen and Packer 2014; Rudd et al. 2017). Thus, access to a single validated gRNA targeting construct for every GTf-gene with facile IDAA protocols to confirm expected predefined indels that ensure frameshifts and inactivation, facilitates screening and cloning of single cells, and enables multiplex gene targeting as illustrated in Figure 4.

Targeting GTf-genes presents unique problems with prediction and interpretation of the outcome. Human cells have at least 15 distinct glycosylation pathways for the different types of glycoconjugates and the step-wise biosynthesis of these involve complex and often entangled biosynthetic steps. Figure 1 and Supplementary data, Figure S1 present a predicted global framework for the genetic regulation of these glycosylation pathways, which illustrates GTf-genes that are predicted to affect specific glycosylation pathways

and GTf-genes that are predicted to serve multiple pathways, or for which it currently is difficult to assign pathway specificity. The latter group is represented by GTf-genes mainly involved in elongation and capping of glycans. The framework further highlights the many families of paralogous genes encoding isoenzymes with related and potentially overlapping functions.

Clearly, this global view is predicted based on current knowledge and should be considered a project that will have to evolve over time and with additional data. Genetic dissection of glycosylation pathways with ZFNs, TALENs and CRISPR/Cas9 is only at its infancy (Steentoft et al. 2014), but it is clear that these tools are transforming glycosciences and opening up new levels of understanding of the genetic and biosynthetic regulation of glycosylation, new approaches to glycoproteomics, unprecedented capabilities for discovery and molecular dissection of biological functions of glycans, and eventually providing us with design control over cellular glycosylation. Thus, targeting genes that truncate one or more distinct glycosylation pathways is a fruitful strategy to produce homogeneous glycoproteomes suitable for simple enrichment strategies needed for sensitive glycoproteomics studies (Steentoft et al. 2011; Radhakrishnan et al. 2014; Vester-Christensen et al. 2013), as well as to define the involvement of individual types of glycoconjugates or glycosylation pathways in biological processes (Stolfa et al.

2016). Moreover, targeting individual isoenzymes expressed in a cell may provide insight into the non-redundant functions of these by comparative analysis of isogenic cell systems, as recently done with the large family of polypeptide GalNAc-transferase genes that initiate GalNAc-type O-glycosylation (Schjoldager et al. 2015). The strategy is also highly fruitful for discovery of novel GTf-genes, as illustrated by the discovery that the two POMTs do not serve to O-mannosylate cadherins and protocadherins (Larsen et al. 2017b), and further leading to discovery of four genes designated TMTC1-4 as novel protein mannosyltransferases (Larsen et al. 2017a). Finally, the gene engineering tools are making progress in custom engineered host cells like the Chinese hamster ovary (CHO) cell for recombinant production of therapeutic glycoproteins with defined glycosylation (Yang et al. 2015b).

In summary, we present a ready to use CRISPR/Cas9 gene targeting resource with predesigned and validated gRNAs and complete protocols for use. Moreover, we present a first generation global map of the genetic regulation of glycosylation pathways in human cells that will facilitate design and interpretation of genetic engineering of glycosylation.

Materials and methods

gRNA design and gRNA amplicon expression cassettes (QCgRNA)

Three to four gRNA sequences for each of 186 human GTf-genes (Figure 1 and Table I) were designed using DESKGEN (<https://www.deskgen.com/landing/>). The gRNAs were incorporated into amplicon gRNA expression cassettes (QCgRNA) through a tri-primer amplification protocol as outlined in Figure 2. The tri-primer amplification included a universal forward primer (QCgFwd), a reverse primer (QCgRev) and gRNA encoding reverse primer (QCgX). The two downstream reverse primers QCgFwd and QCgRev overlap and in the latter case encode the invariant tracerRNA sequence. The QCgX primer encodes the variable gRNA sequence designed to be specific for the respective GTf gene targets. Importantly, the template used for QCgRNA amplification lacks the complete gRNA and tracerRNA sequences. Thus, QCgRNA tri primer amplification using EPB104 as template allows for specific gRNA design amplification determined by the variable QCgX primer included in the PCR assay. The QCgRNA amplicons include the U6 promoter upstream of the gRNA sequence and scaffolding tracrRNA elements were generated essentially as previously described (Lonowski et al. 2017). Only in a few cases (5) were adjustments needed to the standard conditions. All gRNA sequences and primers used are listed in Supplementary data, Table SI.

Screening gRNA designs by QCgRNA

HEK293T cells were used for gRNA design screening and maintained in Dulbecco's Modified Eagle Medium supplemented with 10% FBS, 2 mM L-glutamine. 1×10^5 cells were seeded into 24 well plates one day before the transfection. Two microliter unpurified QCgRNA amplicon was co-transfected with 1 μ g Cas9-2A-GFP (EPB105) using polyethyleneimine (PEI), as outlined in detail previously (Lonowski et al. 2017). Cells were harvested 48 h post-transfection and a fraction sorted and 50,000 cells with medium or high GFP expression collected as a pool. FACS was performed using a FACSAriaTM III cell sorter, using procedures recommended by the manufacturer (BD Bioscience, USA). The collected cell pool and unsorted pool were lysed with 30 μ L QuickExtractTM (Epicentre), and crude lysates analyzed for presence of indels by IDAA as

described in the following section "IDAA profiling". The best performing gRNA preferentially induced +/-1 bp indels for each GTf-gene target was selected, and the QCgRNA amplicon expression cassette subcloned into the pEPB104 plasmid (Addgene #68369) using EcoRI/KpnI restriction endonuclease sites. The validated GlycoCRISPR gRNA plasmid collection has been deposited at Addgene (<https://www.addgene.org/>, deposit#75008) where individual gRNA plasmids can be obtained (deposit#106682-106867).

IDAA profiling

To evaluate the in-del profile induced by the designed gRNAs we used the IDAA protocol recently reported (Lonowski et al. 2017). In brief, PCR was performed on 1 μ L crude cell pool lysates in 25 μ L volume with 5% DMSO, using TEMPase Hot Start DNA Polymerase (Amplicon, Denmark), and following concentrations 0.5 μ M:0.05 μ M: 0.5 μ M of the three primers (universal 5'-labeled primer FamF:geneXF: geneXR). Primer sequences are supplied in Supplementary data, Table SIII. PCR was performed with a touchdown thermocycling profile based on an initial 72°C annealing temperature ramping down by 1°cycle to 58°C, followed by an additional 25 cycles using 58°C annealing temperature. Denaturation and elongation was performed at 95°C for 45 s and 72°C for 30 s, respectively. The tri-primer principle enables uniform labeling of amplicons, and thus allows for size discrimination based on fragment analysis using standard fragment analytical equipment. Efficient discrimination of fragments down to single base size is only supported by denaturing capillary electrophoretic conditions, which is supported by the ABI3130 instruments and later versions. 0.5 μ L of the PCR reaction or dilutions hereof was mixed with 0.05 μ L LIZ500 size standard (ABI/Life Technologies, USA), formamide and applied to fragment analysis using an ABI3500XL instrument (ABI/Life Technologies, USA) as recommended by the manufacturer. Raw data output was analyzed using Gene Mapper (ABI/Life Technologies, USA).

Multiplex GTf-gene targeting

HEK293 cells (2×10^5) were seeded in six well plates, and one day after 1 μ g of GC-GALNT2 and GC-GALNT3 validated plasmids were co-transfected with 1 μ g of Cas9-PBKS plasmid using Lipofectamine 3000. Cells were harvested after one day, and a fraction sorted for GFP by FACS. After one week of culture the FACS sorted cell pool was further single-sorted into 96-well plates and screened by IDAA and immunocytoLOGY again.

ImmunocytoLOGY

HEK293 cells grown on sterile cover slides for one day were fixed with 4% paraformaldehyde, permeabilized with 0.5% Triton X-100, blocked with 3% BSA, and incubated with monoclonal antibodies to human GalNAc-T2 (UH4 4C4) and T3 (UH5 2D10) followed by rabbit anti-mouse IgG TRITIC-conjugated antibodies (Dako), and mounted with Vectashield with DAPI (Vector labs) as described previously (Mandel et al. 1999; Steentoft et al. 2013). Fluorescence microscopy was performed using a Zeiss Axioskop 2 plus with an AxioCam MR3.

Supplementary data

Supplementary data is available at *Glycobiology* online.

Funding

This work was supported by Læge Sophus Carl Emil Friis og hustru Olga Doris Friis' Legat, The Lundbeck Foundation, The University of Copenhagen Excellence Programme for Interdisciplinary Research [CDO2016], The Danish Research Councils (1331-00133B), The Novo Nordisk Foundation, The Danish National Research Foundation (DNRF107).

Conflict of interest statement

Authors declare conflict of interest related to ownership in GlycoDisplay ApS and patent applications submitted by University of Copenhagen.

Abbreviations

GTf, glycosyltransferase; CRISPR, clustered regularly interspaced short repeats; gRNA, guide RNA; ZFN, zinc finger nucleases; TALEN, transcription activator-like effector nucleases; HEK293T, human embryonic kidney 293T; IDAA, indel detection by amplicon analysis; FACS, fluorescence activated cell sorting; QCgRNA, quick change guide RNA; HEK293 human embryonic kidney 293; iPSC, induced pluripotent stem cell.

References

- Amado M, Almeida R, Schwientek T, Clausen H. 1999. Identification and characterization of large galactosyltransferase gene families: Galactosyltransferases for all functions. *Biochim Biophys Acta*. 1473:35–53.
- Chandrasegaran S, Carroll D. 2016. Origins of programmable nucleases for genome engineering. *J Mol Biol*. 428:963–989.
- Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science*. 339:819–823.
- Conzelmann A, Kornfeld S. 1984. Beta-linked N-acetylgalactosamine residues present at the nonreducing termini of O-linked oligosaccharides of a cloned murine cytotoxic T lymphocyte line are absent in a Vicia villosa lectin-resistant mutant cell line. *J Biol Chem*. 259:12528–12535.
- Cummings RD. 2009. The repertoire of glycan determinants in the human glycome. *Mol Biosyst*. 5:1087–1104.
- Doench JG, Fusi N, Sullender M, Hegde M, Vainberg EW, Donovan KF, Smith I, Tothova Z, Wilen C, Orchard R et al. 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol*. 34:184–191.
- Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, Sullender M, Ebert BL, Xavier RJ, Root DE. 2014. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol*. 32:1262–1267.
- Duda K, Lonowski LA, Kofoed-Nielsen M, Ibarra A, Delay CM, Kang Q, Yang Z, Pruitt-Miller SM, Bennett EP, Wandall HH et al. 2014. High-efficiency genome editing via 2A-coupled co-expression of fluorescent proteins and zinc finger nucleases or CRISPR/Cas9 nickase pairs. *Nucleic Acids Res*. 42:e84.
- Hansen L, Lind-Thomsen A, Joshi HJ, Pedersen NB, Have CT, Kong Y, Wang S, Sparso T, Grarup N, Vester-Christensen MB et al. 2015. A glycogene mutation map for discovery of diseases of glycosylation. *Glycobiology*. 25:211–224.
- Hart T, Tong AHY, Chan K, Van Leeuwen J, Seetharaman A, Aregger M, Chandrashekhar M, Hustedt N, Seth S, Noonan A et al. 2017. Evaluation and design of genome-wide CRISPR/SpCas9 knockout screens. *G3 (Bethesda)*. 7:2719–2727.
- Heigwer F, Kerr G, Boutros M. 2014. E-CRISP: Fast CRISPR target site identification. *Nat Methods*. 11:122–123.
- Henrissat B, Surolia A, Stanley P. 2009. A genomic view of glycobiology. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME, editors. *Essentials of Glycobiology*. 2nd ed. [(New York) Cold Spring Harbor], Chapter 7.
- Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O et al. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*. 31:827–832.
- Jae LT, Raaben M, Riemersma M, van Beusekom E, Blomen VA, Velds A, Kerkhoven RM, Carette JE, Topaloglu H, Meinecke P et al. 2013. Deciphering the glycosylome of dystroglycanopathies using haploid screens for lassa virus entry. *Science*. 340:479–483.
- Larsen ISB, Narimatsu Y, Joshi HJ, Siukstaite I, Harrison OJ, Brasch J, Goodman K, Hansen L, Shapiro L, Honig B et al. 2017a. Discovery of an O-mannosylation pathway selectively serving cadherins and protocadherins. *Proc Natl Acad Sci USA*. 114(42):11163–11168, In press.
- Larsen ISB, Narimatsu Y, Joshi HJ, Yang Z, Harrison OJ, Brasch J, Shapiro L, Honig B, Vakhrushev SY, Clausen H et al. 2017b. Mammalian O-mannosylation of cadherins and plexins is independent of protein O-mannosyltransferases 1 and 2. *J Biol Chem*. 292: 11586–11598.
- Levery SB, Steentoft C, Halim A, Narimatsu Y, Clausen H, Vakhrushev SY. 2015. Advances in mass spectrometry driven O-glycoproteomics. *Biochim Biophys Acta*. 1850:33–42.
- Lombard V, Golaonda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res*. 42:D490–D495.
- Lonowski LA, Narimatsu Y, Riaz A, Delay CE, Yang Z, Niola F, Duda K, Ober EA, Clausen H, Wandall HH et al. 2017. Genome editing using FACS enrichment of nuclelease-expressing cells and indel detection by amplicon analysis. *Nat Protoc*. 12:581–603.
- Lowe JB, Marth JD. 2003. A genetic approach to Mammalian glycan function. *Annu Rev Biochem*. 72:643–691.
- Mandel U, Hassan H, Therkildsen MH, Rygaard J, Jakobsen MH, Juhl BR, Dabelsteen E, Clausen H. 1999. Expression of polypeptide GalNAc-transferases in stratified epithelia and squamous cell carcinomas: Immunohistological evaluation using monoclonal antibodies to three members of the GalNAc-transferase family. *Glycobiology*. 9:43–52.
- Metzakopian E, Strong A, Iyer V, Hodgkins A, Tzelepis K, Antunes L, Friedrich MJ, Kang Q, Davidson T, Lamberth J et al. 2017. Enhancing the genome editing toolbox: genome wide CRISPR arrayed libraries. *Sci Rep*. 7:2244.
- Miles LA, Garippa RJ, Poirier JT. 2016. Design, execution, and analysis of pooled in vitro CRISPR/Cas9 screens. *FEBS J*. 283:3170–3180.
- Montague TG, Cruz JM, Gagnon JA, Church GM, Valen E. 2014. CHOPCHOP: A CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res*. 42:W401–W407.
- Narimatsu H. 2006. Human glycogene cloning: Focus on beta 3-glycosyltransferase and beta 4-glycosyltransferase families. *Curr Opin Struct Biol*. 16:567–575.
- Paque D, Kwart D, Chen A, Sproul A, Jacob S, Teo S, Olsen KM, Gregg A, Noggle S, Tessier-Lavigne M. 2016. Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature*. 533:125–129.
- Patnaik SK, Stanley P. 2006. Lectin-resistant CHO glycosylation mutants. *Methods Enzymol*. 416:159–182.
- Radhakrishnan P, Dabelsteen S, Madsen FB, Francavilla C, Kopp KL, Steentoft C, Vakhrushev SY, Olsen JV, Hansen L, Bennett EP et al. 2014. Immature truncated O-glycophenotype of cancer directly induces oncogenic features. *Proc Natl Acad Sci USA*. 111:E4066–E4075.
- Rudd P, Karlsson NG, Khoo KH, Packer NH. 2017. Glycomics and glycoproteomics. *Essentials of Glycobiology* In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Darvill AG, Kinoshita T, Packer NH, Prestegard JH et al (Eds.), 3rd ed. [(New York) Cold Spring Harbor], Chapter 51.
- Schjoldager KT, Joshi HJ, Kong Y, Goth CK, King SL, Wandall HH, Bennett EP, Vakhrushev SY, Clausen H. 2015. Deconstruction of O-glycosylation-GalNAc-T isoforms direct distinct subsets of the O-glycoproteome. *EMBO Rep*. 16:1713–1722.
- Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelson T, Heckl D, Ebert BL, Root DE, Doench JG et al. 2014. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*. 343:84–87.

- Shi J, Wang E, Milazzo JP, Wang Z, Kinney JB, Vakoc CR. 2015. Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat Biotechnol.* 33:661–667.
- Steentoft C, Bennett EP, Schjoldager KT, Vakhrushev SY, Wandall HH, Clausen H. 2014. Precision genome editing: a small revolution for glycobiology. *Glycobiology.* 24:663–680.
- Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KT, Lavrsen K, Dabelsteen S, Pedersen NB, Marcos-Silva L et al. 2013. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J.* 32:1478–1488.
- Steentoft C, Vakhrushev SY, Vester-Christensen MB, Schjoldager KT, Kong Y, Bennett EP, Mandel U, Wandall H, Levery SB, Clausen H. 2011. Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nat Methods.* 8:977–982.
- Stolfa G, Mondal N, Zhu Y, Yu X, Buffone A Jr., Neelamegham S. 2016. Using CRISPR-Cas9 to quantify the contributions of O-glycans, N-glycans and glycosphingolipids to human leukocyte-endothelium adhesion. *Sci Rep.* 6:30392.
- Thaysen-Andersen M, Packer NH. 2014. Advances in LC-MS/MS-based glycoproteomics: Getting closer to system-wide site-specific mapping of the N- and O-glycoproteome. *Biochim Biophys Acta.* 1844:1437–1452.
- Tsuji S, Datta AK, Paulson JC. 1996. Systematic nomenclature for sialyltransferases. *Glycobiology.* 6:v–vii.
- van Overbeek M, Capurso D, Carter MM, Thompson MS, Frias E, Russ C, Reece-Hoyes JS, Nye C, Gradia S, Vidal B et al. 2016. DNA repair profiling reveals nonrandom outcomes at Cas9-mediated breaks. *Mol Cell.* 63: 633–646.
- Varki A. 2017. Biological roles of glycans. *Glycobiology.* 27:3–49.
- Varki A, Cummings RD, Aebi M, Packer NH, Seeberger PH, Esko JD, Stanley P, Hart G, Darvill A, Kinoshita T et al. 2015. Symbol nomenclature for graphical representations of glycans. *Glycobiology.* 25: 1323–1324.
- Vester-Christensen MB, Halim A, Joshi HJ, Steentoft C, Bennett EP, Levery SB, Vakhrushev SY, Clausen H. 2013. Mining the O-mannose glycoproteome reveals cadherins as major O-mannosylated glycoproteins. *Proc Natl Acad Sci USA.* 110:21018–21023.
- Wang T, Wei JJ, Sabatini DM, Lander ES. 2014. Genetic screens in human cells using the CRISPR-Cas9 system. *Science.* 343:80–84.
- Yang Z, Steentoft C, Hauge C, Hansen L, Thomsen AL, Niola F, Vester-Christensen MB, Frodin M, Clausen H, Wandall HH et al. 2015a. Fast and sensitive detection of indels induced by precise gene targeting. *Nucleic Acids Res.* 43:e59.
- Yang Z, Wang S, Halim A, Schulz MA, Frodin M, Rahman SH, Vester-Christensen MB, Behrens C, Kristensen C, Vakhrushev SY et al. 2015b. Engineered CHO cells for production of diverse, homogeneous glycoproteins. *Nat Biotechnol.* 33:842–844.