Supplementary Information for:

Model-driven engineering of N-linked glycosylation in Chinese Hamster Ovary cells

Christopher S. Stach[1], Meghan G. McCann[2], Conor M. O'Brien[2], Tung S. Le[2], Nikunj Somia[3], Xinning Chen[4], Kyuongho Lee[2], Hsu-Yuan Fu[2], Prodromos Daoutidis[2], Liang Zhao[4], and Wei-Shou Hu[2], Michael Smanski[1]*

Contents:

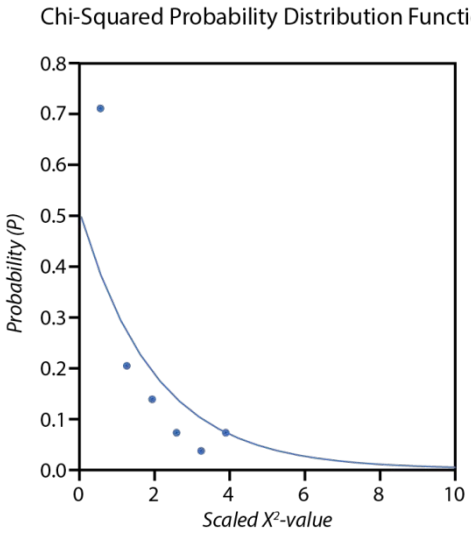
Supplementary Note 1: Statistical Analysis

Differences in galactose incorporation level between the 2C10 cell line and overexpression strains was determined using a modified χ^2-test. Because the glycan profile is measured as a fractional percentage out of 100% and not as raw counts, a standard χ^2-analysis could not be performed (estimation of significance is effected by the multiplicand used to convert percentage into counts, *i.e.* a χ^2-comparison of G0:G1:G2 = 0.2:0.5:0.3 vs 0.3:0.4:0.3 would give a different probability value than a χ^2-comparison of 20:50:30 vs 30:40:30). To circumvent this problem, we used χ^2-values generated pairwise among the nine independent replicates from our CHO-2C10 cell line to fit a standard χ^2-Probability Density Function (PDF) and identified an appropriate scaling factor to convert a χ^2-value to a probability value (p-value).

First, χ^2-tests were performed between each pairwise combination of the nine biological replicates of IgG galactose profile from 2C10. The 36 resulting χ^2-values were binned (bins=6) and fit to a χ^2 probability distribution function for data sets with two degrees of freedom:

$$PDF(x) = \frac{e^{\frac{-x}{2}}}{2} \quad (1)$$

We optimized to χ^2 scaling factor to minimize the sum of square residuals (SSRs) between the histogram of observed χ^2-values and the PDF. The optimal scaling factor of 0.648 resulted in a fit between the PDF and observed χ^2-values shown below in **Supplementary Figure N1**. This scaling factor allowed us to perform the same χ^2-analysis between the galactose-content distribution of IgG produced by the 2C10 cell line and that produced by engineered cells. We used a Bonferonni correction for multiple comparisons to determine which glycan profiles are significantly different than CHO-2C10, with a p-value cutoff of 0.05.



Supplementary Figure N1. Scaled χ^2-values from biological replicates fit to χ^2 probability distribution function (Eq. 1) for comparisons with three degrees of freedom.

Homogeneity metric.

To quantify the homogeneity of glycans with respect to galactose incorporation, we used a scoring metric based on the Shannon entropy. The Shannon entropy of a galactose-content distribution is equal to:

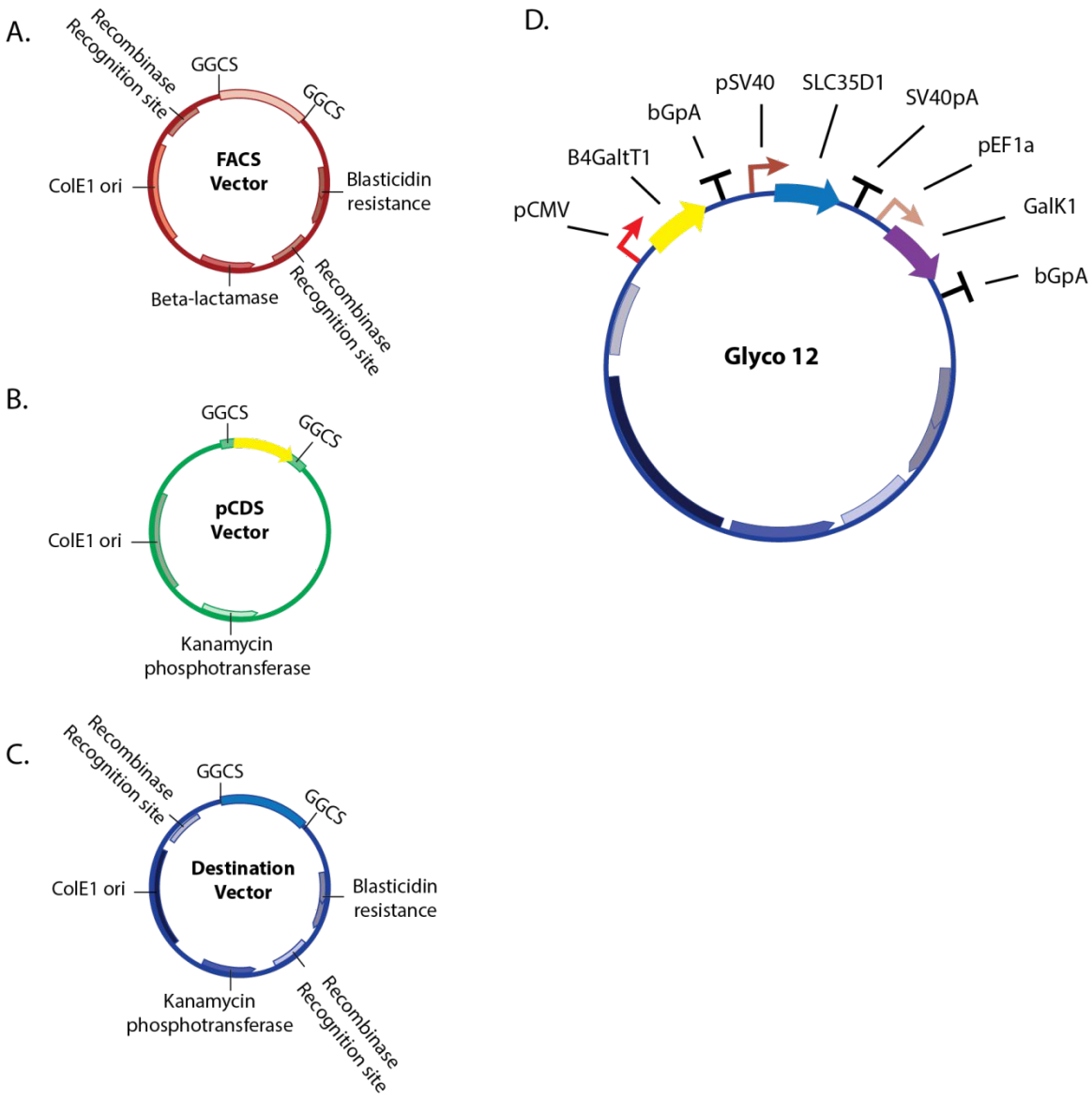
$$H(X) = -\sum_{i=1}^{n} F(x_i) log_{10}(F(x_i)) \quad (2)$$

where $F(x_i)$ is the frequency a particular galactose-content category (G0, G1, G2). To convert Shannon entropy to a scoring metric for homogeneity, we applied the following conversion:

$$S(homogeneity) = 100 * (1 - \frac{H(X)}{0.477}) \quad (3)$$

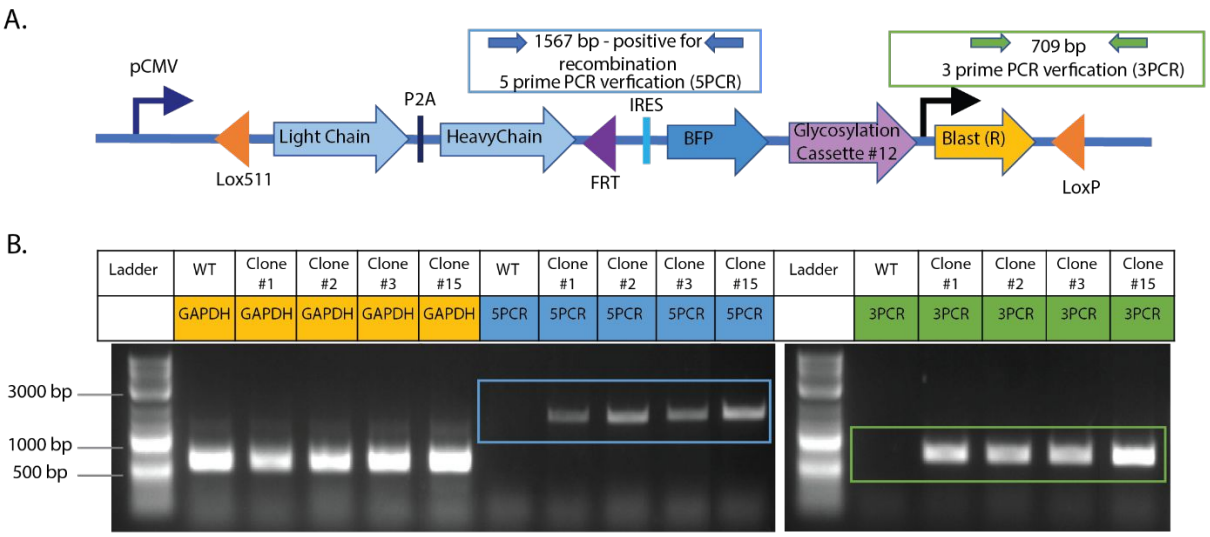
The dividend 0.477 represents the maximal value for $H(X)$ and occurs then the three galactose-content classes are equally abundant. Based on this scoring metric, IgG glycan profiles that have an equal amount of G0, G1, G2 glycans will have a homogeneity score of 0, and profiles that have 100% one species will have a score of 100. Importantly, this scoring metric is independent of which galactose-content class in predominant in a population.

Supplementary Figure S1: Maps of Key Plasmids



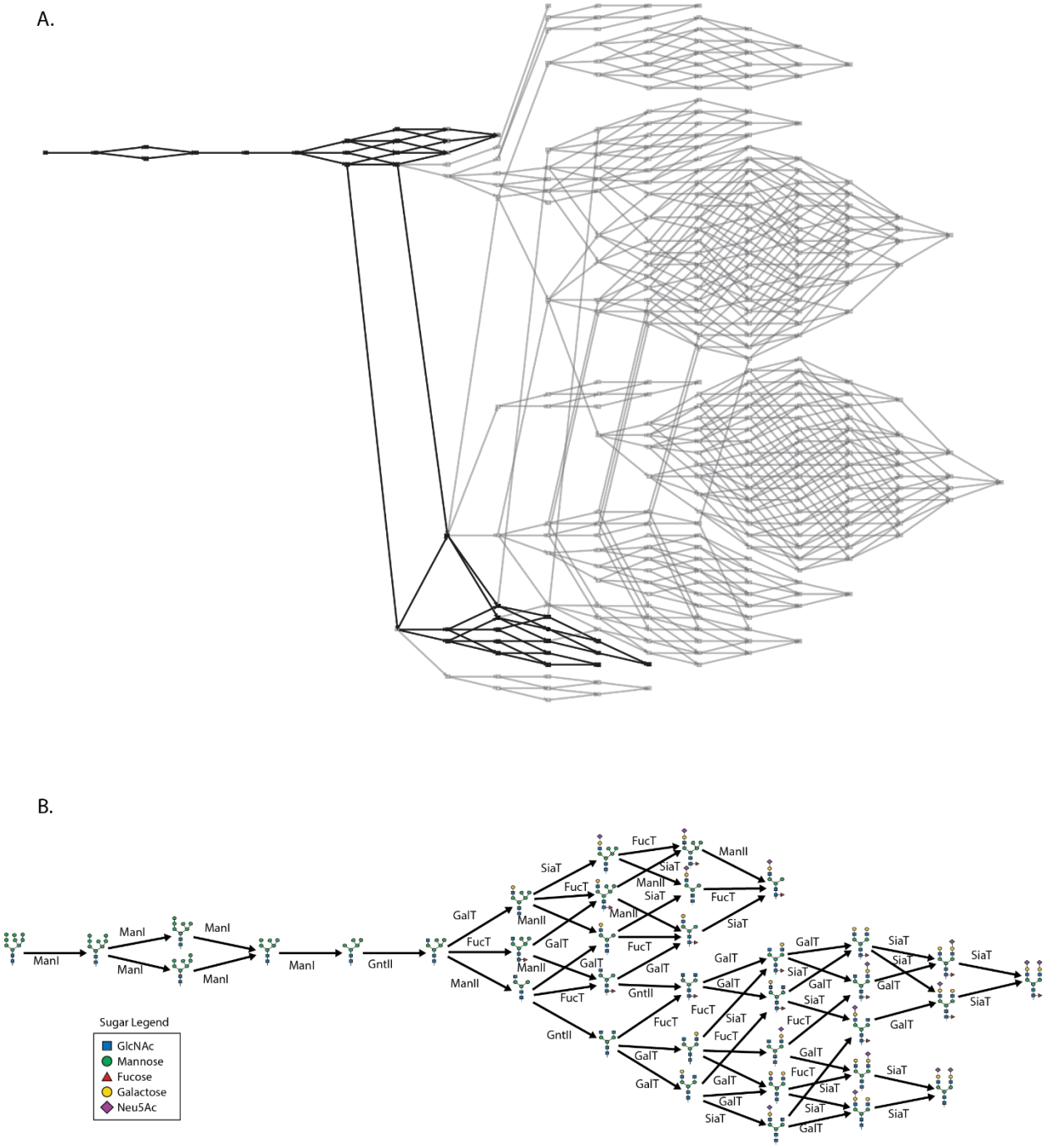
Supplementary Figure S1. Maps of key plasmids, including (A) FACS vector, (B) pCDS vector, (C) Destination vector, and (D) Glyco 12, an example of a three gene overexpression plasmid. GGCS, golden gate cloning site; ori, origin of replication.

Supplementary Figure S2: Confirmation of Site-Specific Integration Events



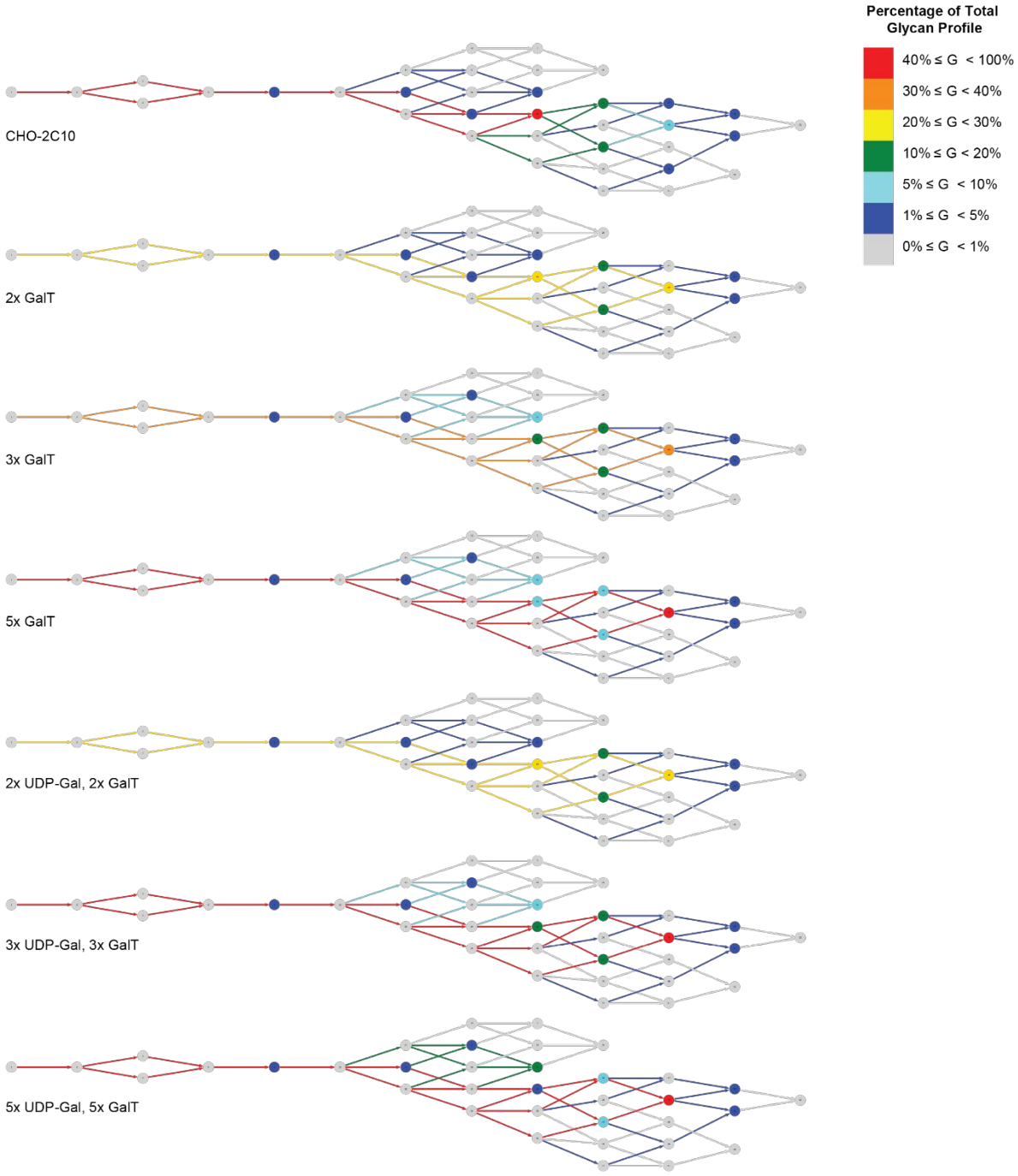
Supplementary Figure S2. Molecular confirmation of site-specific integration events. (A) Map of genomic landing pad with primer binding sites highlighted for 5'-PCR (blue) and 3'-PCR (green). (B) Agarose gel showing PCR products of reactions run with genomic DNA isolated from clonal populations of cells. GAPDH is a housekeeping gene control to demonstrate that gDNA is suitable for PCR amplification.

Supplementary Figure S3: Complete GlycoVis Network Map



Supplementary Figure S3. Complete GlycoVis network map of possible glycans of Human IgG. (A) All possible glycans for human IgG are displayed. The black selection are glycans detailed in (B) below. (B) Glycovis network map displaying glycan structures and the enzymes responsible for each glycan discussed in the main text. This network represents the glycan structure key for abstract network graphs in the main text **Fig. 2A** and **Supplementary Fig. S4** below.

Supplementary Figure S4: Kinetic Model-predicted Glycan Profiles

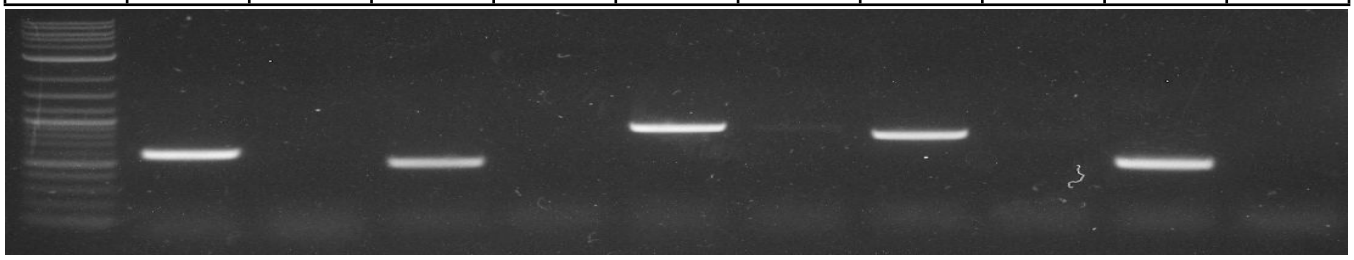


Supplementary Figure S4. Predicted glycan profile for increasing concentrations of GalT and UDP-Gal.

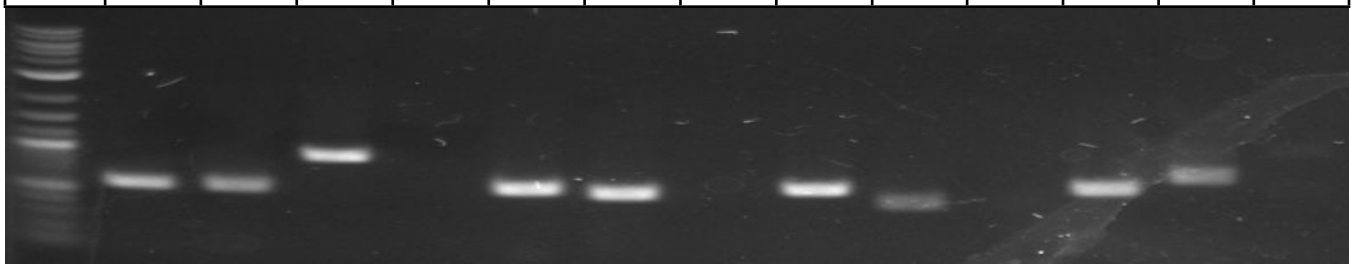
Supplementary Figure S5: RT-PCR Confirmation of Glycosylation Enzyme Transgene Transcription

Supplementary Figure S5. RT-PCR confirmation of glycosylation enzyme transgene transcription in genetically engineered cell pools. Primers used are detailed in **Supplementary Data File 1**.

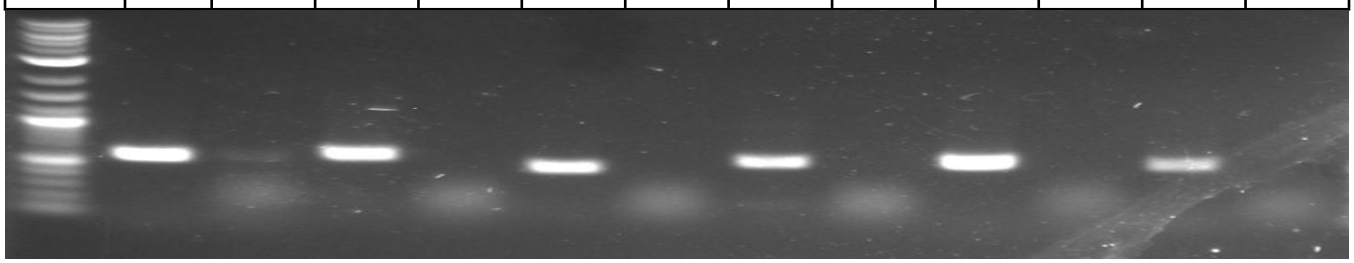
Cell Pool: 2 Log Ladder	B4GalT2	Control	MGAT2	Control	UGP2iA	Control	UGP2iB	Control	B4GalT1	Control
Template Type	cDNA	cDNA	cDNA	cDNA	cDNA	cDNA	cDNA	cDNA	cDNA	cDNA
Gene	B4GalT2	B4GalT2	MGAT2	MGAT2	UGP2iA	UGP2iA	UGP2iB	UGP2iB	B4GalT1	B4GalT1
Expected Band Size:	613	-----	534	-----	926	-----	831	-----	530	-----



Cell Pool: 2 Log Ladder	Control	SLC35A2	SLC35A2	Control	MGAT2	MGAT2	Control	GALE	GALE	Control	GALT	GALT	Control
Template Type	cDNA	cDNA	cDNA	cDNA	cDNA	cDNA	cDNA	cDNA	cDNA	cDNA	cDNA	cDNA	cDNA
Gene	GAPDH	GAPDH	SLC35A2	SLC35A2	GAPDH	MGAT2	MGAT2	GAPDH	GALE	GALE	GALE	GAPDH	GALT
Expected Band Size:	570	570	900	0	570	534	0	570	457	0	570	668	0

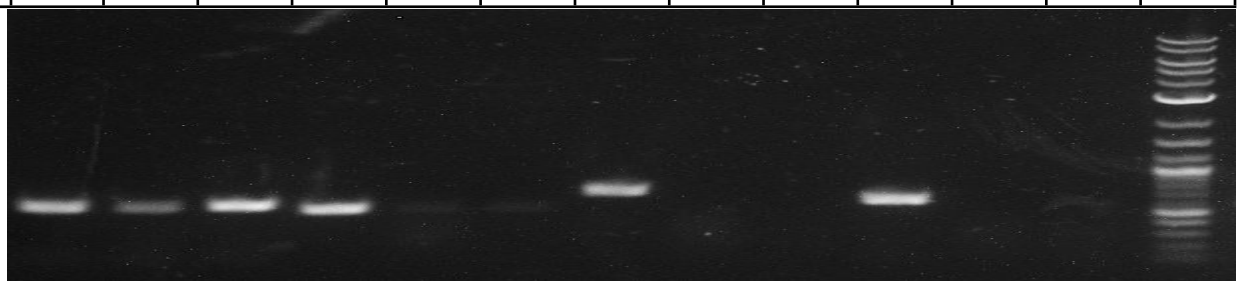


Cell Pool: 2 Log Ladder	Sia3	Sia3	Sia3	Sia3	Sia3	Sia3	Sia3	Sia3	Hyb1	Hyb1	Hyb1	Hyb1
Template Type	cDNA	RNA	cDNA	RNA	cDNA	RNA	cDNA	RNA	cDNA	RNA	cDNA	RNA
Gene	GAPDH	GAPDH	ST6Gal1	ST6Gal1	NANP	NANP	SLC35A1	SLC35A1	GAPDH	GAPDH	MGAT1	MGAT1
Expected Band Size:	570	0	612	0	504	0	574	0	570	0	528	0

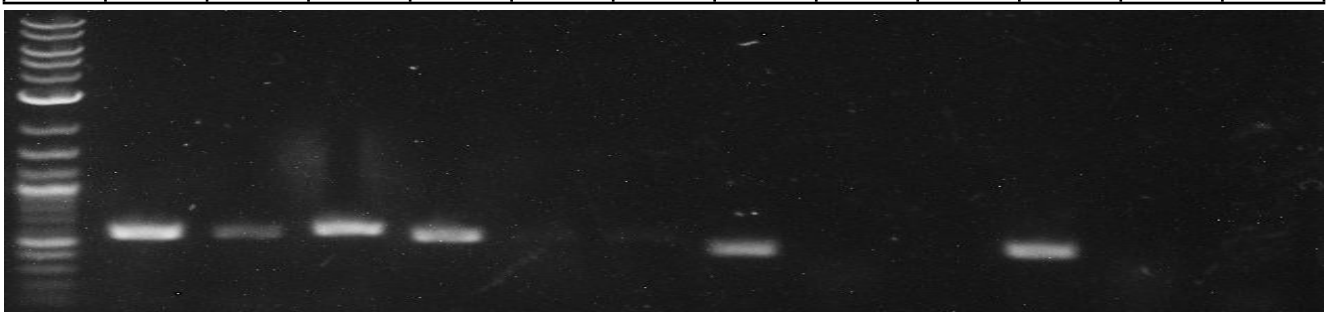


Cell Pool: 2 Log Ladder	Hyb1	Hyb1	Hyb1	Hyb1	Control	Control	Control	Control	Control	Control	Control
Template Type	cDNA	RNA	cDNA	RNA	cDNA	cDNA	cDNA	cDNA	cDNA	cDNA	cDNA
Gene	B4GalT2	B4GalT2	GALT	GALT	GAPDH	ST6Gal1	NANP	SLC35A1	MGAT1	B4GalT2	GALT
Expected Band Size:	613	0	668	0	570	0	0	0	0	0	0

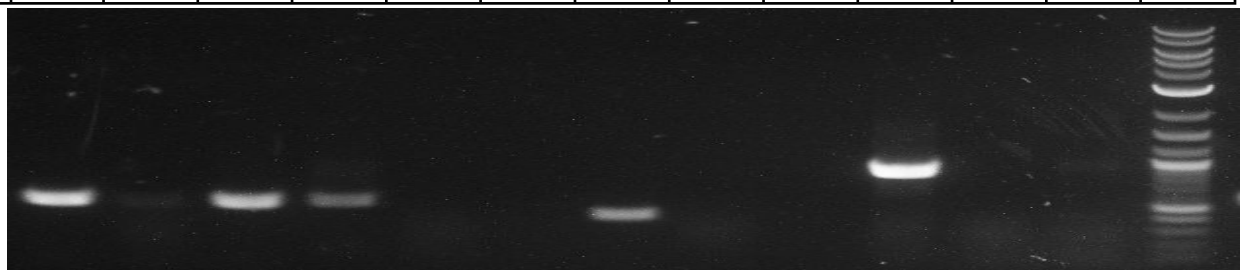
Cell Pool:	Glyco13	Glyco13	Control	Glyco13	Glyco13	Control	Glyco13	Glyco13	Control	Gyco13	Gyco13	Control	2 Log Ladder
Template Type	cDNA	RNA	cDNA	cDNA	RNA	cDNA	cDNA	RNA	cDNA	cDNA	RNA	cDNA	
Gene	GAPDH	GAPDH	GAPDH	B4GalT1	B4GalT1	B4GalT1	SLC35D1	SLC35D1	SLC35D1	GALK2	GALK2	GALK2	
Expected Band Size:	570	0	570	530	0	0	706	0	0	593	0	0	



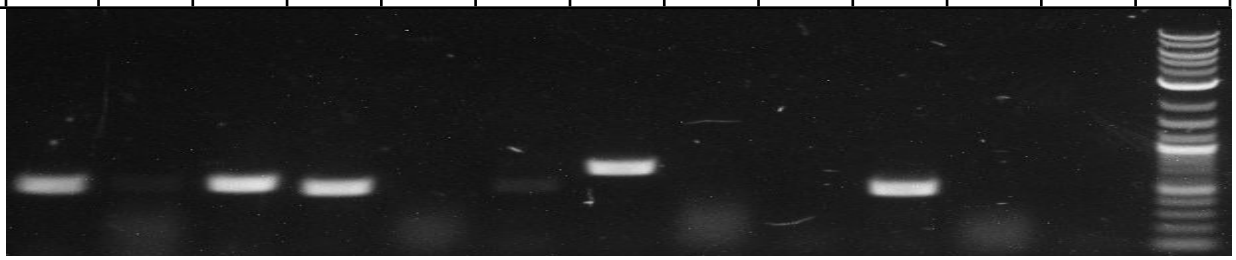
Cell Pool: 2 Log Ladder	Glyco15	Glyco15	Control	Glyco15	Glyco15	Control	Glyco15	Glyco15	Control	Gyco15	Gyco15	Control
Template Type	cDNA	RNA	cDNA	cDNA	RNA	cDNA	cDNA	RNA	cDNA	cDNA	RNA	cDNA
Gene	GAPDH	GAPDH	GAPDH	B4GalT1	B4GalT1	B4GalT1	SLC35A3	SLC35A3	SLC35A3	RMD	RMD	RMD
Expected Band Size:	570	0	570	530	0	0	427	0	0	434	0	0



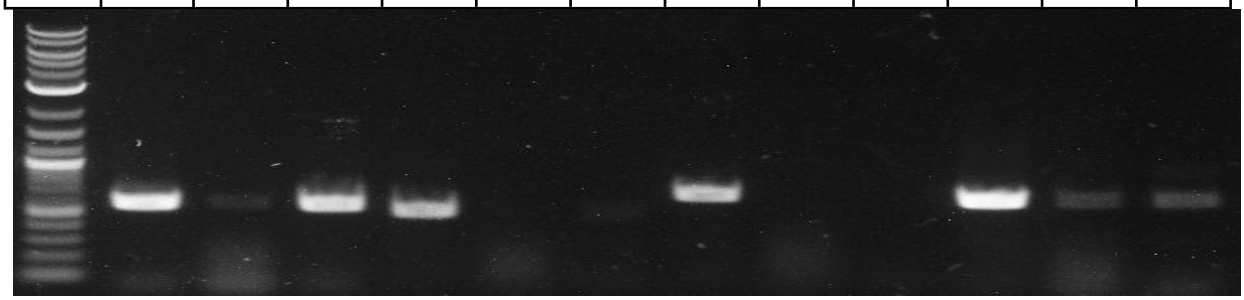
Cell Pool:	Glyco19	Glyco19	Control	Glyco19	Glyco19	Control	Glyco19	Glyco19	Control	Gyco19	Gyco19	Control	2 Log Ladder
Template Type	cDNA	RNA	cDNA	cDNA	RNA	cDNA	cDNA	RNA	cDNA	cDNA	RNA	cDNA	
Gene	GAPDH	GAPDH	GAPDH	MGAT4D	MGAT4D	MGAT4D	Mpdu1	Mpdu1	Mpdu1	DOLK	DOLK	DOLK	
Expected Band Size:	570	0	570	580	0	0	433	0	0	880	0	0	



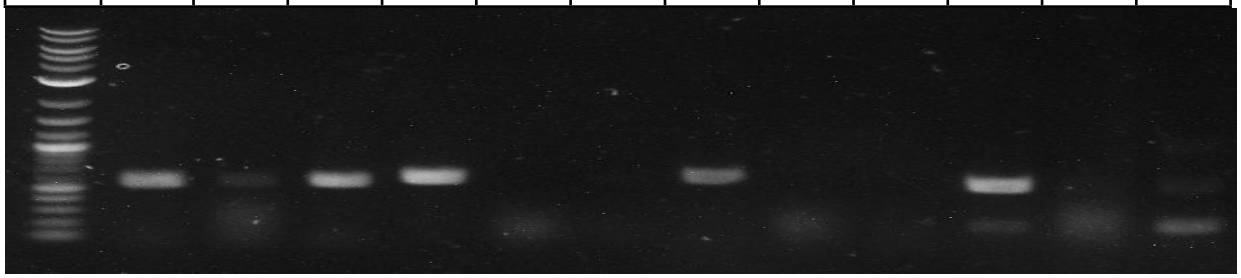
Cell Pool:	Glyco12	Glyco12	Control	Glyco12	Glyco12	Control	Glyco12	Glyco12	Control	Glyco12	Glyco12	Control	2 Log Ladder
Template Type	cDNA	RNA	cDNA	cDNA	RNA	cDNA	cDNA	RNA	cDNA	cDNA	RNA	cDNA	
Gene	GAPDH	GAPDH	GAPDH	B4GalT1	B4GalT1	B4GalT1	SLC35D1	SLC35D1	SLC35D1	GALK1	GALK1	GALK1	
Expected Band Size:	570	0	570	530	0	0	706	0	0	489	0	0	



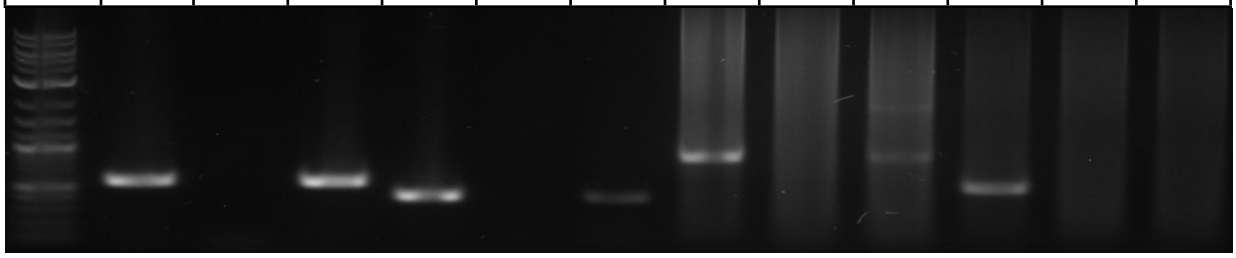
Cell Pool: 2 Log Ladder	Glyco11	Glyco11	Control	Glyco11	Glyco11	Control	Glyco11	Glyco11	Control	Glyco11	Glyco11	Control
Template Type	cDNA	RNA	cDNA	cDNA	RNA	cDNA	cDNA	RNA	cDNA	cDNA	RNA	cDNA
Gene	GAPDH	GAPDH	GAPDH	B4GalT1	B4GalT1	B4GalT1	SLC35D1	SLC35D1	SLC35D1	GALT	GALT	GALT
Expected Band Size:	570	0	570	530	0	0	706	0	0	668	0	0



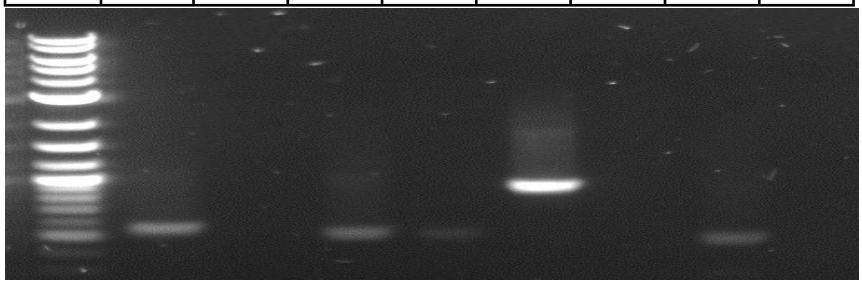
Cell Pool: 2 Log Ladder	Sia2	Sia2	Control	Sia2	Sia2	Control	Sia2	Sia2	Control	Sia2	Sia2	Control
Template Type	cDNA	RNA	cDNA	cDNA	RNA	cDNA	cDNA	RNA	cDNA	cDNA	RNA	cDNA
Gene	GAPDH	GAPDH	GAPDH	ST6Gal1	ST6Gal1	ST6Gal1	CMAS	CMAS	CMAS	SLC35A1	SLC35A1	SLC35A1
Expected Band Size:	570	0	570	612	0	0	646	0	0	574	0	0



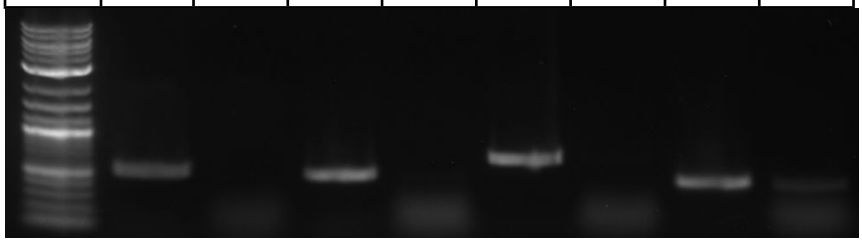
Cell Pool: 2 Log Ladder	Glyco20	Glyco20	Control	Glyco20	Glyco20	Control	Glyco20	Glyco20	Control	Glyco20	Glyco20	Control
Template Type	cDNA	RNA	cDNA	cDNA	RNA	cDNA	cDNA	RNA	cDNA	cDNA	RNA	cDNA
Gene	GAPDH	GAPDH	GAPDH	MPDU1	MPDU1	MPDU1	DOLK	DOLK	DOLK	MGAT1	MGAT1	MGAT1
Expected Band Size:	570	0	570	443	0	0	880	0	0	528	0	0



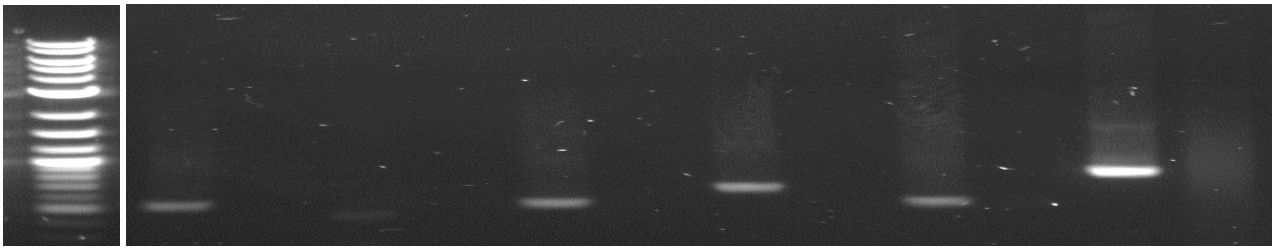
Cell Pool: 2 Log Ladder	Glyco21	Glyco21	Glyco21	Glyco21	Glyco21	Glyco21	Glyco21	Glyco21
Template Type	cDNA	RNA	cDNA	RNA	cDNA	RNA	cDNA	RNA
Gene	GAPDH	GAPDH	B4GalT1	B4GalT1	SLC35A2	SLC35A2	GALK1	GALK1
Expected Band Size:	570	0	530	0	900	0	489	0



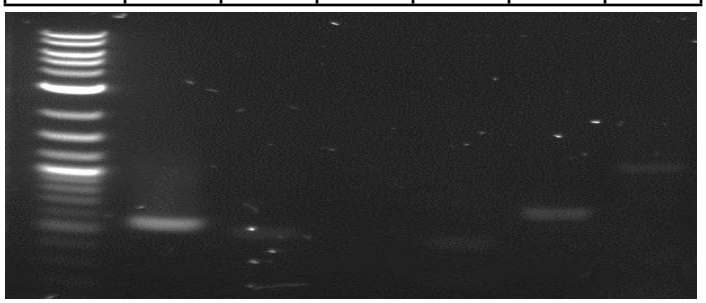
Cell Pool: 2 Log Ladder	Glyco23	Glyco23	Glyco23	Glyco23	Glyco23	Glyco23	Glyco23	Glyco23
Template Type	cDNA	RNA	cDNA	RNA	cDNA	RNA	cDNA	RNA
Gene	GAPDH	GAPDH	B4GalT1	B4GalT1	SLC35D1	SLC35D1	GALE	GALE
Expected Band Size:	570	0	530	0	706	0	570	0



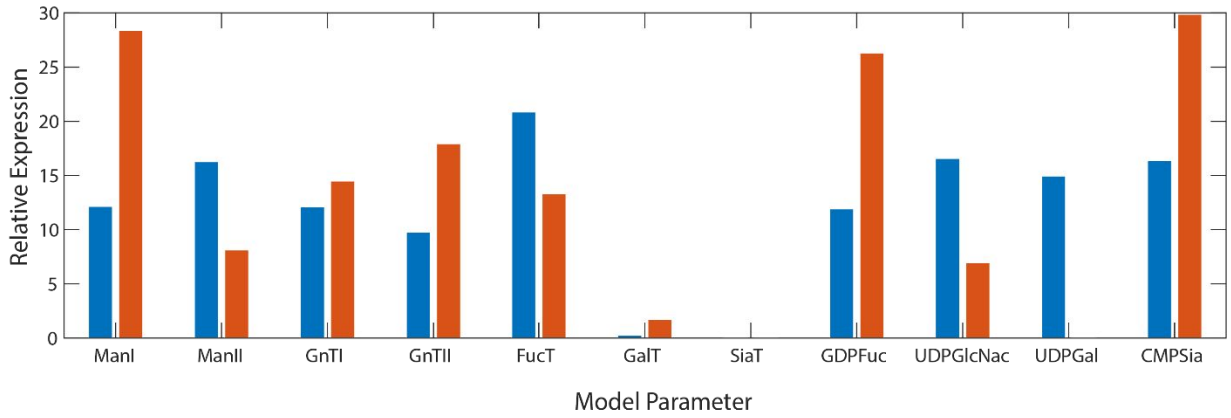
Cell Pool: 2 Log Ladder	GALK1	GALK1	GALK1	GALK1	SLC35D1	SLC35D1	SLC35D1	SLC35D1	DOLK	DOLK	DOLK	DOLK
Template Type	cDNA	RNA	cDNA	RNA	cDNA	RNA	cDNA	RNA	cDNA	RNA	cDNA	RNA
Gene	GAPDH	GAPDH	GALK1	GALK1	GAPDH	GAPDH	SLC35D1	SLC35D1	GAPDH	GAPDH	DOLK	DOLK
Expected Band Size:	570	0	489	0	570	0	706	0	570	0	880	0



Cell Pool: 2 Log Ladder	Control	Control	Control	Control	Control	Control
Template Type	cDNA	cDNA	cDNA	cDNA	cDNA	cDNA
Gene	GAPDH	B4GalT1	SLC35A2	GALK1	SLC35D1	DOLK
Expected Band Size:	570	0	0	0	0	0

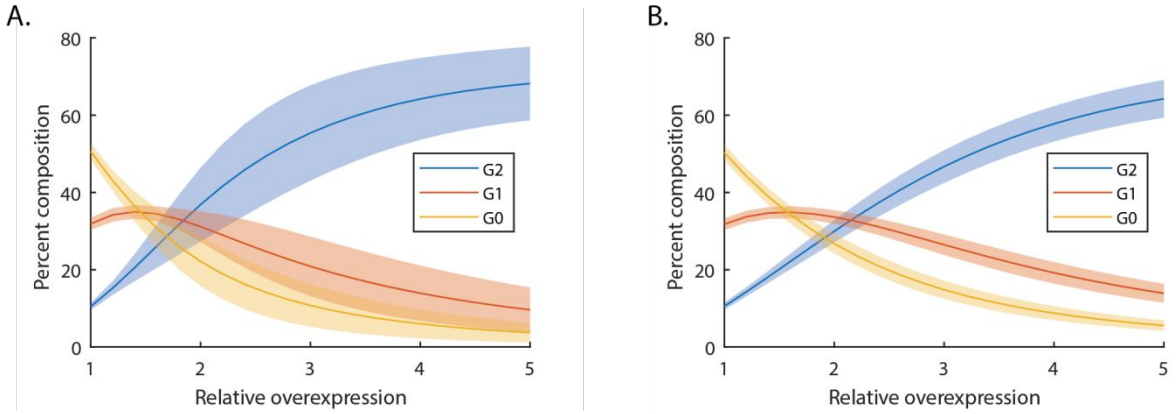


Supplementary Figure S6: Parameter values in retained versus removed sets



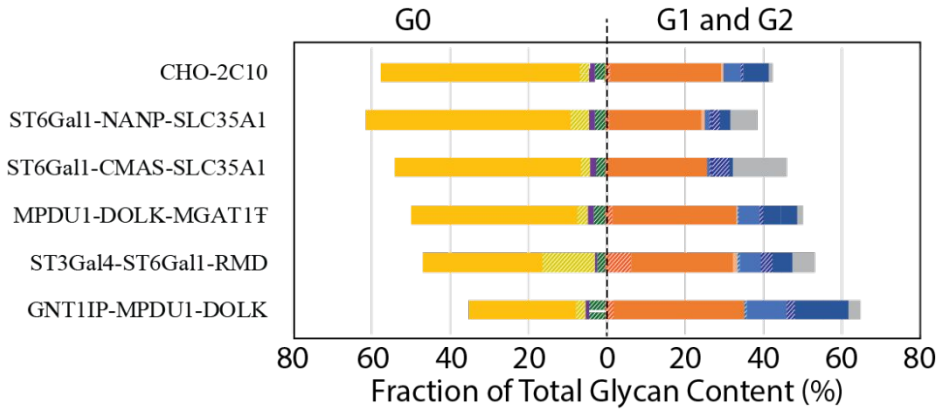
Supplementary Figure S6. Average parameter values in retained parameter sets (blue, N=30) and removed parameter sets (red, N=20).

Supplementary Figure S7: Kinetic modeling of three-gene overexpression constructs



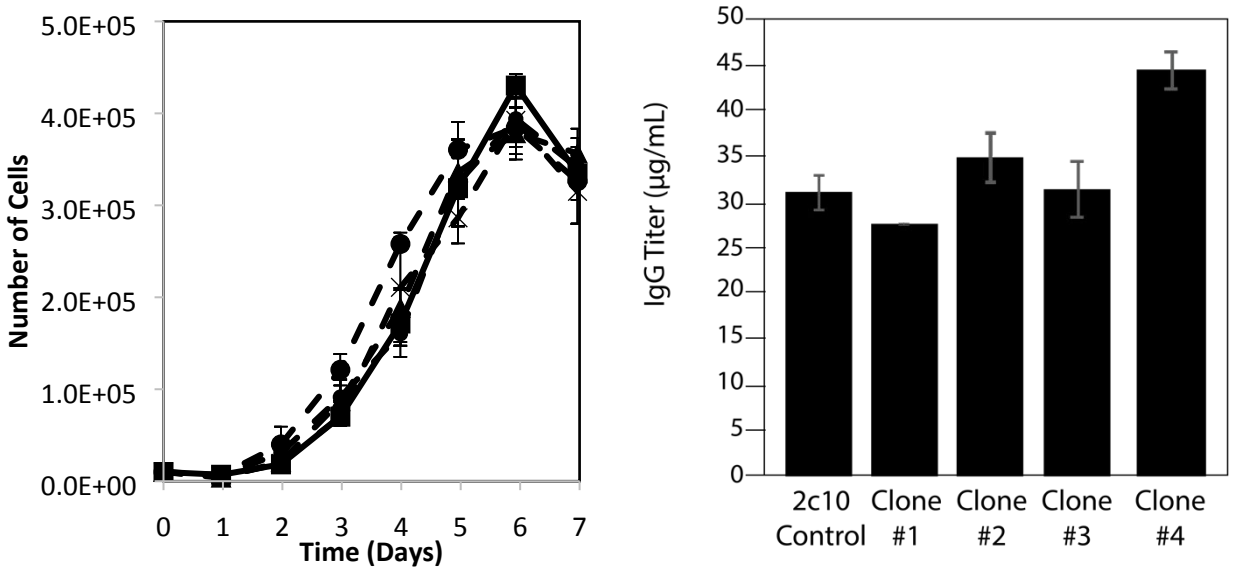
Supplementary Figure S7. Comparison of kinetic modeling of galactose incorporation with initial 50 parameter combinations (A) or down-selected to parameter combinations (B). Relative overexpression refers to the simultaneous increase of both UDP-gal concentration and GalT Vmax (i.e. tracing the x=y diagonal line from a 3d plot as in **Fig. 4**). Shaded region denotes one standard deviation above or below the mean. G0, agalactosylated; G1, monogalactosylated; G2, bigalactosylated.

Supplementary Figure S8: Three-gene Constructs for Increased Sialic Acid or Mannose Branching



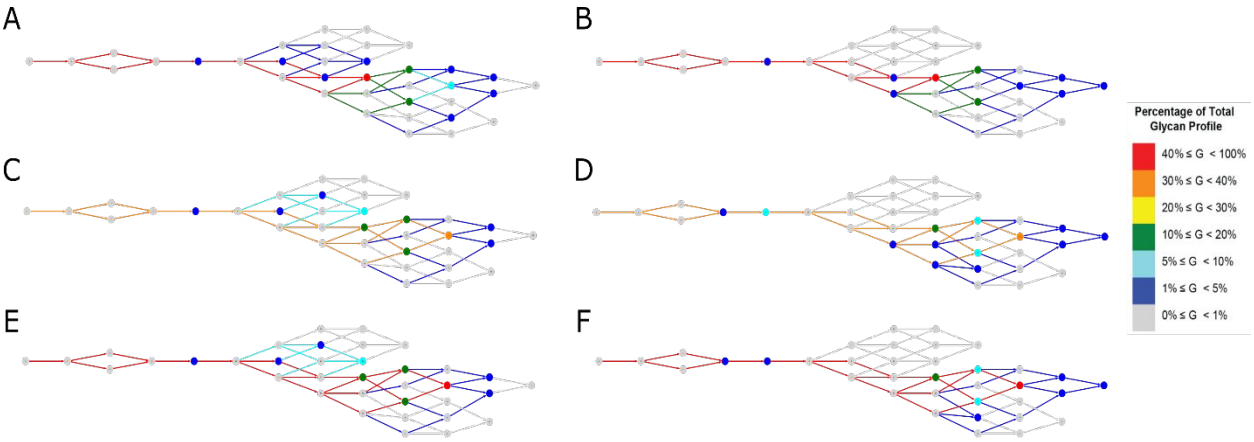
Supplementary Figure S8. Glycan profile of additional three-gene overexpression constructs. Bars are colored as in the main text **Figs. 3 and 4**. CHO-2C10 is the starting IgG overproduction cell line, engineered cell pools are named according to their three overexpression genes.

Supplementary Figure S9: Growth and IgG Titer of Engineered Cell Lines



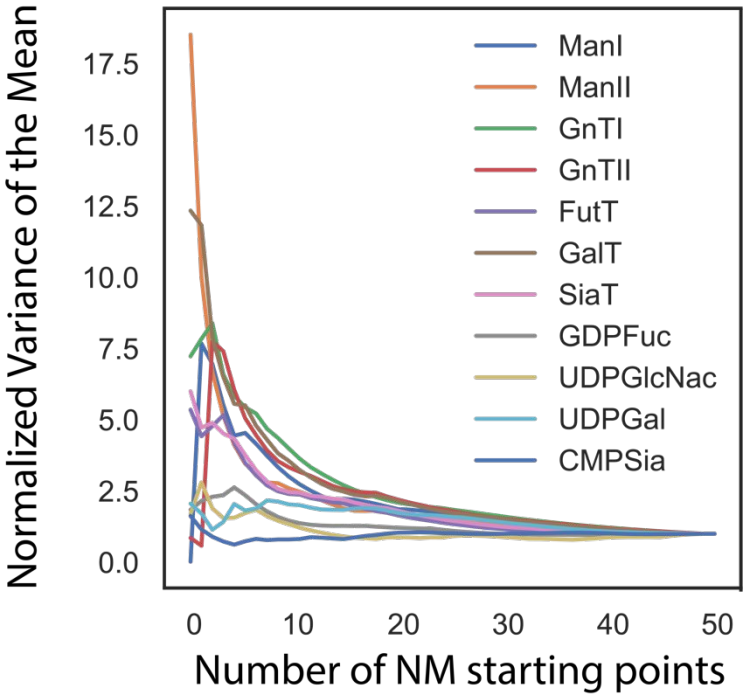
Supplementary Figure S9. Performance of targeted three-gene integration cell lines. (A) Cell growth for adherent culture of 2C10 (solid line) and clones (dashed lines) in 24-well plate over 7 days, with media replaced on day 4, and (B) IgG titer at day 7, 3 days post-media change, as measured by ELISA. N = 3 wells per cell line.

Supplementary Figure S10: Comparison of Model and Experimental Profiles using Glycovis



Supplementary Figure S10. Glycan profiles resulting from model simulation and experimental measurement for CHO-2C10 (A and B), GalT overexpression (C and D), and three gene results for increasing GalT and UDP-Gal supply (E and F), corresponding to the three gene overexpression of B4GalT1-SLC35D1-GALK1. Model results are shown in A, C, and E for these cases, whereas the corresponding experimental results are shown in B, D, and F, respectively. The key to node and edge identities is found is Supplementary Figure S3b.

Supplementary Figure S11: Variance of Mean Galactosylation Sensitivity



Supplementary Figure S11: The impact of the number of starting points representing the best initial agreement with the 2C10 glycan profile on the variance of the mean Galactosylation sensitivity prior to the Nelder-Mead (NM) optimization.

Supplementary Table S1: Key plasmids used in this study

Plasmid	Description	Reference
pCDS	Plasmid for holding CDSs	This study
pSG-DV _AB	Plasmid for generating site 1 promoter and terminator constructs	This study
pSG-DV _BC	Plasmid for generating site 2 promoter and terminator constructs	This study
pSG-DV _CD	Plasmid for generating site 3 promoter and terminator constructs	This study
Cis-Reg_3-AB_pCMV_bGpA	FACS-AB promoter and terminator plasmid for position 1 monocistrons	This study
Cis-Reg_18-BC_pSV40_SV40pA	FACS-BC promoter and terminator plasmid for position 2 monocistrons	This study
Cis-Reg_9-CD_pEF1a_bGpA	FACS-CD promoter and terminator plasmid for position 3 monocistrons	This study
pMG-DV (CHO-DV)	Destination vector for all 3 gene constructs	This study
CHO-DV-BFP-Switch	Destination vector with BFP expression for selection of targeted integration cell lines	This study
Glyco12	Library construct with greatest increase in galactosylation	This study

Supplementary Table 2: Key primers used in this study

Primer Name	Sequence	Description
CSS_Vector_BB_ScarAR	5'- CACCTGCTTACCTCCCTGTCTTCGAAGTTCCTATTCTCTAGAAAG-3'	Assembly of pSG-DV vectors
CSS_ScarAF_LacZ	5'- GAAGACAGGGAGGTAAGCAGGTGTGCACCATATGCGGTGTGAAATACC -3'	Assembly of pSG-DV vectors
CSS_LacZ_ScarBR	5'- GAAGACAGAGTAGTAAGCAGGTGCCCGCGCGTTGGCCGATTCATTAATG -3'	Assembly of pSG-DV vectors
CSS_ScarB_UBCpromF	5'- CACCTGCTTACTACTCTGTCTTCGATCTGGCCTCCGCGCCGGG -3'	Assembly of pSG-DV vectors
CSS_UBCR_RemAarIR	5'- CCTACCGGCACGTGGCCCCACCCTGCATTATAAGC -3'	Assembly of pSG-DV vectors
CSS_UBCF_RemAarIF	5'- GTGGGGCCACGTGCCGGTAGGTGTGCGGTAGGC -3'	Assembly of pSG-DV vectors
CSS_UBCR_Blast	5'- CGAATTAATTCCCGGGCCCAAGCTTCGTCTAACAAAAAAGCCAAAAACG -3'	Assembly of pSG-DV vectors
CSS_BlastF	5'- GAAGCTTGGGCCCGGGAATTAATTCGCCACC -3'	Assembly of pSG-DV vectors
CSS_Blast_R_SynPolyA	5'- GATGTAATGAAAATAAAGATATTTTATTTGAGCTGAAGGTACGCTGTATCTCAG -3'	Assembly of pSG-DV vectors
CSS_SynPolyAF	5'- AATAAAATATCTTTATTTTCATTACATCTGTGTG -3'	Assembly of pSG-DV vectors
CSS_SpacerR_RemSapI	5'- CGTTGGCTACCCGTGATATTGCTGAAGGGCTTGGCGGCGAATGGGCTGACC -3'	Assembly of pSG-DV vectors
CSS_RemSapIF_Spacer	5'- GCCGCCAAGCCCTTCAGCAATATCACGGGTAGCCAACGCTATGTCCTGATAGC -3'	Assembly of pSG-DV vectors
CSS_SpacerR_Beta-lac	5'- GCCTCACTGATTAAGCATTGGTAACCCATGGCGATGCCTGCTTGCCGAATATC -3'	Assembly of pSG-DV vectors
CSS_Beta-LacF	5'- TTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCG-3'	Assembly of pSG-DV vectors
CSS_Beta_LacR-Spacer	5'- CTGATCAAGAGACAGGATGAGGATCGTTTCGCATGAGTATTCAACATTTCCGTGTCGC -3'	Assembly of pSG-DV vectors
SpacerF	5'- GCGAAACGATCCTCATCCTGTCTCTTGATCAG -3'	Assembly of pSG-DV vectors
CSS_ColE1_RemSapIR	5'- GGAGAAAATACCGCATCAGGCGCTCCTCCGCTTCCTCGCTCACTGACTCG -3'	Assembly of pSG-DV vectors
CSS_Rem_SapIF_ColE1	5'- GGAAGCGGAGGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGG -3'	Assembly of pSG-DV vectors
CSS_ColE1R_ISce	5'- AACAGGGTAATGTATCTAGGGATAACAGGGTAATGTATCGCCAGCGTCGCACAGCATG TTTG -3'	Assembly of pSG-DV vectors
CSS_ISceIF	5'- ACATTACCCTGTTATCCCTAGATACATTACCCTGTTATCCCAGATGACATACCCTG -3'	Assembly of pSG-DV vectors
CSS_Vec_ScarBR	5'- CACCTGCTTACAGTACTGTCTTCGAAGTTCCTATTCTCTAGAAAG -3'	Assembly of pSG-DV vectors
CSS_ScarB_LacZF	5'- GAAGACAGTACTGTAAGCAGGTGTGCACCATATGCGGTGTGAAATACCG -3'	Assembly of pSG-DV vectors
CSS_LacZ_ScarCR	5'- GAAGACAGCCAAGTAAGCAGGTGCCCGCGCGTTGGCCGATTCATTAATG -3'	Assembly of pSG-DV vectors
CSS_ScarC_UBCF	5'- CACCTGCTTACTTGGCTGTCTTCGATCTGGCCTCCGCGCCGGGTTTTGGCGCC -3'	Assembly of pSG-DV vectors
CSS_Vec_ScarCR	5'- CACCTGCTTACCCAACTGTCTTCGAAGTTCCTATTCTCTAGAAAGTATAGG -3'	Assembly of pSG-DV vectors
CSS_ScarC_LacZF	5'- GAAGACAGTTGGGTAAGCAGGTGTGCACCATATGCGGTGTGAAATACCGC -3'	Assembly of pSG-DV vectors
CSS_LacZ_ScarDR	5'- GAAGACAGACCTGTAAGCAGGTGCCCGCGCGTTGGCCGATTCATTAATG -3'	Assembly of pSG-DV vectors
CSS_ScarD_UBCF	5'- CACCTGCTTACAGGTCTGTCTTCGATCTGGCCTCCGCGCCGGGTTTTGGC -3'	Assembly of pSG-DV vectors
CSS_G-Block Blasticidin	5'- GAAGCTTGGGCCCGGGAATTAATTCGCCACCATGGAAACCTTCAACATCTCTCAGCAG GATCTGGAGCTGGTGGAGGTCGCCACTGAGAAGATCACCATGCTCTATGAGGACAACA AGCACCATGTCGGGGCGGCCATCAGGACCAAGACTGGGGAGATCATCTCTGCTGTCCA CATTGAAGCCTACATTGGCAGGGTCACTGTCTGTGCTGAAGCCATTGCCATTGGGTCT GCTGTGAGCAACGGGCAGAAGGACTTTGACACCATTGTGGCTGTCAGGCACCCCTACT CTGATGAGGTGGACAGATCCATCAGGGTGGTCAGCCCCTGTGGCATGTGTAGAGAGCT GATCTCTGACTATGCTCCTGACTGCTTTGTGCTCATTGAGATGAATGGCAAGCTGGTC AAAACCACCATTGAGGAACTCATCCCCCTCAAGTACACCAGGAACTAAACCTGAGCTA GCTTGACTGACTGAGATACAGCGTACCTTCAGCTCA -3'	Assembly of pSG-DV vectors
CSS_CMVF_ScarA	5'- CTAGCACCTGCTTTTGGAGACTAGTATTATGCCCAGTACATGACC -3'	Assembly of pFACS plasmids
CSS_CMVR_AATG	5'- CTAGCACCTGCAAAACATTTCTTCTATGGAGGTCAAAACAGCGTGG -3'	Assembly of pFACS plasmids
CSS_SV40promF_ScarB	5'- CTAGCACCTGCTTTTTACTTGTGTCAGTTAGGGTGTGGAAAGTC -3'	Assembly of pFACS plasmids
CSS_SV40promR_AATG	5'- CTAGCACCTGCAAAACATTTTTGCAAAAGCCTAGGCCTCCAAAAAAGC -3'	Assembly of pFACS plasmids
CSS_EF1F_ScarC	5'- CTAGCACCTGCTTTTTTGGAAGGATCTGCGATCGCTCCGGTGC -3'	Assembly of pFACS plasmids
CSS_EF1R_AATG	5'- CTAGCACCTGCAAAACATTGTAGGCGCCGGTCACAGCTTGGATC -3'	Assembly of pFACS plasmids

CSS_AarIF_eGFP	5'- CTAGCACCTGCTACTAATGAGAAGAGCGTCCAAGGGCGAGGAGCTG -3'	Assembly of pFACS plasmids
CSS_AarIR_eGFP	5'- CTAGCACCTGCCCATATCATGAAGAGCTCACTTGTACAGCTCGTCCATGC -3'	Assembly of pFACS plasmids
CSS_rbglobTermAF_tgat	5'- CTAGCACCTGCTTTTTGATTTCACTCCTCAGGTGCAGGCTGCC -3'	Assembly of pFACS plasmids
CSS_rbglobTermAR_ScarB	5'- CTAGCACCTGCAAAAAGTAGAGAAGAGGGACAGCTATGACTGG -3'	Assembly of pFACS plasmids
CSS_rbglobTermAR_ScarD	5'- CTAGCACCTGCAAAAACCTGAGAAGAGGGACAGCTATGACTGG -3'	Assembly of pFACS plasmids
CSS_SV40TermAF_tgat	5'- CTAGCACCTGCTTTTTGATAACTTGTTTATTGCAGCTTATAATGG -3'	Assembly of pFACS plasmids
CSS_SV40TermAR_ScarC	5'- CTAGCACCTGCAAAACCAACCAGACATGATAAGATACATTGATG -3'	Assembly of pFACS plasmids
#1pMC_fwd	5'- TCGACCCATGGGGGCCCG -3'	Assembly of pMG-DV and CHO-DV-BFP-SWITCH
#1 pMC_rev	5'- GGGCTCCCCGGGCGCGAC -3'	Assembly of pMG-DV and CHO-DV-BFP-SWITCH
#3 UbcP(a)_fwd	5'- TAAGCAGGTGGATCTGGCCTCCGCGCCG -3'	Assembly of pMG-DV and CHO-DV-BFP-SWITCH
#4 Ubcp(b)-gcg_rev	5'- GGCCCGGGAATTAATTCGCCACCATGGAA -3'	Assembly of pMG-DV and CHO-DV-BFP-SWITCH
LacZ Fragment G_block	5'- GCTACGCTGACTGCAGTGCAATAACTTCGTATAAAGTATCCTATACGAAGTTATGAAG TTCCTATACTTTCTAGAGAATAGGAACTTCCACCTGCTTACGGAGCTGTCTTCTGCAC CATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCC ATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCT ATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCA GGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTCGAGCTCGGTACCC GGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGC TGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAG CATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTG CGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCG GCCAACGCGCGGGGAAGACAGAGGTGTAAGCAGGTGGATCTGGCCTCCGCGCCG -3'	Assembly of pMG-DV and CHO-DV-BFP-SWITCH
Blast(R)_G-Block	5'- ATTAATTCGCCACCATGGAAACCTTCAACATCTCTCAGCAGGATCTGGAGCTGGTGGA GGTCGCCACTGAGAAGATCACCATGCTCTATGAGGACAACAAGCACCATGTCGGGGCG GCCATCAGGACCAAGACTGGGGAGATCATCTCTGCTGTCCACATTGAAGCCTACATTG GCAGGGTCACTGTCTGTGCTGAAGCCATTGCCATTGGGTCTGCTGTGAGCAACGGGCA GAAGGACTTTGACACCATTGTGGCTGTCAGGCACCCCTACTCTGATGAGGTGGACAGA TCCATCAGGGTGGTCAGCCCCTGTGGCATGTGTAGAGAGCTGATCTCTGACTATGCTC CTGACTGCTTTGTGCTCATTGAGATGAATGGCAAGCTGGTCAAAACCACCATTGAGGA ACTCATCCCCCTCAAGTACACCAGGAACTAATGAAATAAAATATCTTTATTTTCATTA CATCTGTGTGTTGGTTTTTTGTGTGATAACTTCGTATAATGTGTACTATACGAAGTTA TGAAGTTCCTATACTATTTGAAGAATAGGAACTTCTCGACCCATGGGGGCCCGCC - 3'	Assembly of pMG-DV and CHO-DV-BFP-SWITCH
CSS_BFPswtchBFPF	5'- AAAAAGCTCTTCA ATGAGCGAGCTGATTAAGGAGAACATGC -3'	Assembly of pMG-DV and CHO-DV-BFP-SWITCH
CSS_BFPswtchBFPR	5'- TTTTTGCTCTTCATCATTAATTAAGCTTGTGCCCCAGTTTGC -3'	Assembly of pMG-DV and CHO-DV-BFP-SWITCH
CSS_BFPswtchCHDVF	5'- AAAAAGCTCTTCA TGATAACTTGTTTATTGCAGCTTATAATGG -3'	Assembly of pMG-DV and CHO-DV-BFP-SWITCH
CSS_BFPswtchCHDVR	5'- TTTTTGCTCTTCACATGGTTGTGGCCATATTATCATCGTG -3'	Assembly of pMG-DV and CHO-DV-BFP-SWITCH
pCDS_vecF	5'- ATGCGAAGACAGGCTTCCTCGCTCACTGACTCGCTGCGCTC-3'	Assembly of pCDS
pCDS_vecR	5'- ATGCGAAGACAGTTGTGTCTCAAAATCTCTGATGTACATTGCACAAG-3'	Assembly of pCDS
pCDS_lacZF	5'- ATGCGAAGACAGACAAGCTCTTCAATGAAAAGCAGGTGTGCACCATATGCGGTGTGAA ATA -3'	Assembly of pCDS
pCDS_lacZR	5'- ATGCGAAGACAGAAGCGCTCTTCATCATATAGCAGGTGCCCGCGCGTTGGCCGATTCA TT -3'	Assembly of pCDS

MGM_HC_1F	5′ - CGGGATGAGCTGACCAAGAA – 3′	Confirmation of 5′ side of targeted integration
MGM_tagBFP_1R	5′ - GCTAGGGAGGTCGCAGTATC – 3′	Confirmation of 5′ side of targeted integration
MGM_BlastR_2F	5′ - CACCATGGAAACCTTCAACATC – 3′	Confirmation of 3′ side of targeted integration
MGM_WPRE_2R	5′ - AAAGGCATTAAAGCAGCGTATC – 3′	Confirmation of 3′ side of targeted integration
MGM_GAPDH_1F	5′ - ACGGATTTGGCCGTATTGGA – 3′	GAPDH Control for PCR
MGM_GAPDH_1R	5′ - CATCACGCCACAGCTTTCC – 3′	GAPDH Control for PCR
MGM_B4GALT1_1F	5' - CTGCGTCTCTCCTCACAAGG - 3'	RT-PCR Confirmation of Transcription
MGM_B4GALT1_1R	5' - TCTTGTCTCTTGAGTGGCGG - 3'	RT-PCR Confirmation of Transcription
MGM_B4GALT2(iA)_1F	5' - AGGAGCAGTGGCCTTGTTTG - 3'	RT-PCR Confirmation of Transcription
MGM_B4GALT2(iA)_1R	5' - ATGACATAGACGCCGTAGCG - 3'	RT-PCR Confirmation of Transcription
MGM_CMAS_1F	5' - TGCAGCGCAACTCTCGC - 3'	RT-PCR Confirmation of Transcription
MGM_CMAS_1R	5' - GCTCGCATTTCGTAGTATGCC - 3'	RT-PCR Confirmation of Transcription
MGM_DOLK_1F	5' - GTGCGCCGTGGCCCTCGCAG - 3'	RT-PCR Confirmation of Transcription
MGM_DOLK_1R	5' - TCGGGCGATGGTGGGGGCCTG - 3'	RT-PCR Confirmation of Transcription
MGM_GALE_1F	5' - TGCGCTATTTCAACCCCACA - 3'	RT-PCR Confirmation of Transcription
MGM_GALE_1R	5' - CGCCAGAGATCCTCACACAT - 3'	RT-PCR Confirmation of Transcription
MGM_GALK1_1F	5' - TGTCCAGCTCAGCATCCTTG - 3'	RT-PCR Confirmation of Transcription
MGM_GALK1_1R	5' - CTCTGTAGTCGCCACGTCTC - 3'	RT-PCR Confirmation of Transcription
MGM_GALK2(i1)_1F	5' - TGGTCTGTTGTGCTGGCTTG - 3'	RT-PCR Confirmation of Transcription
MGM_GALK2(i1)_1R	5' - CTTCTTAAACTGGAGCACTCGC - 3'	RT-PCR Confirmation of Transcription
MGM_GALT(i1)_1F	5' - ATATCCGCTACAACCCGCTG - 3'	RT-PCR Confirmation of Transcription
MGM_GALT(i1)_1R	5' - TGTCTGGTAGGGCCATGTTG - 3'	RT-PCR Confirmation of Transcription
MGM_MGAT1_1F	5' - CCTCACCCGGGAAGTGATTC - 3'	RT-PCR Confirmation of Transcription
MGM_MGAT1_1R	5' - GTGGCCCGAAAGTACTCGAA - 3'	RT-PCR Confirmation of Transcription
MGM_MGAT2_1F	5' - GCGACAAAGGAAGAACGAGG - 3'	RT-PCR Confirmation of Transcription
MGM_MGAT2_1R	5' - TTCAAAGCGGCATTCTTCGG - 3'	RT-PCR Confirmation of Transcription
MGM_MGAT4D_1F	5' - CTTGCTGATCACCCTGGTCG - 3'	RT-PCR Confirmation of Transcription
MGM_MGAT4D_1R	5' - GGATCCAGACCTCACTTGCC - 3'	RT-PCR Confirmation of Transcription
MGM_MPDU1_1F	5' - GCTGCTCGTGCCGATTCTTT - 3'	RT-PCR Confirmation of Transcription
MGM_MPDU1_1R	5' - GGTGACTACAGTCAAGGGCG - 3'	RT-PCR Confirmation of Transcription
MGM_NANP_1F	5' - AACACTCTCATCGACACGGC - 3'	RT-PCR Confirmation of Transcription
MGM_NANP_1R	5' - CCCAGGTTGTACTCCGAGAAG - 3'	RT-PCR Confirmation of Transcription
MGM_RMD_1F	5' - TCAAGCTTATCTGGCAGCGG - 3'	RT-PCR Confirmation of Transcription
MGM_RMD_1R	5' - CGATATGGTTGAACGGTCGC - 3'	RT-PCR Confirmation of Transcription
MGM_SLC35A1(iA)_1F	5' - TGGATGCAGCAGTGTACCAG - 3'	RT-PCR Confirmation of Transcription
MGM_SLC35A1(iA)_1R	5' - GTACCCAGGGCAAAGGTGAG - 3'	RT-PCR Confirmation of Transcription
MGM_SLC35A2(iA)_1F	5' - TTCCGCGGGTGCATTGGA - 3'	RT-PCR Confirmation of Transcription
MGM_SLC35A2(iA)_1R	5' - AAGCCAAAGAGGCGAATGGA - 3'	RT-PCR Confirmation of Transcription
MGM_SLC35A3(i1)_1F	5' - AGCAGTGGTTGTTGCTGAACT - 3'	RT-PCR Confirmation of Transcription
MGM_SLC35A3(i1)_1R	5' - ACATGCTGTGAGAACTGCCAT - 3'	RT-PCR Confirmation of Transcription
MGM_SLC35D1_1F	5' - GCGAAATCCTCCACACTCCG - 3'	RT-PCR Confirmation of Transcription
MGM_SLC35D1_1R	5' - TGTCAGCCCAGCCTTCAAAC - 3'	RT-PCR Confirmation of Transcription
MGM_ST3Gal4(i1)_1F	5' - AGAGCAAGGCCTCTAAGCTC - 3'	RT-PCR Confirmation of Transcription
MGM_ST3Gal4(i1)_1R	5' - AGCTACCAGGACGAGGAGTG - 3'	RT-PCR Confirmation of Transcription
MGM_ST6Gal1(iA)_1F	5' - GGCAGATGTGCTGTTGTGTC - 3'	RT-PCR Confirmation of Transcription
MGM_ST6Gal1(iA)_1R	5' - GCCCTGGTTGAGATGCTTCA - 3'	RT-PCR Confirmation of Transcription
MGM_UGP2(iA)_1F	5' - TACTCACCACAGCATCATCACA - 3'	RT-PCR Confirmation of Transcription
MGM_UGP2(iA)_1R	5' - GGCATTTTGCTCCTGCAGTC - 3'	RT-PCR Confirmation of Transcription
MGM_UGP2(iB)_1F	5' - CAACTCTGGATTGCTTGATACCT - 3'	RT-PCR Confirmation of Transcription
MGM_UGP2(iB)_1R	5' - TGGTCCAAGATGCGAAGGTT - 3'	RT-PCR Confirmation of Transcription

Supplementary Tables S3-S7 show model parameters used in this study. Parameters that were varied to fit experimental data are highlighted in blue with their allowable ranges denoted.

Supplementary Table S3: Enzyme kinetic parameters

Enzyme (e_m)	Rmax (µM/min*cell)	# of Glycan substrates	Dissociation constant of glycans from enzyme (µM)	Dissociation constant of nucleotide-sugar from enzyme (µM)
Man I	9.204 - 92,040	8	17	N/A
Man II	12.196 - 121,960	6	200	N/A
GnT I	9.504 - 95,040	1	260	170
GnT II	18.929 - 189,290	7	190	960
GnT III	0	7	21	420
GnT IV	0	13	800	220
GnT V	0	12	87	11000
FucT	3.222 - 32,220	11	25	46
GalT	3.449 - 34,490	232	660	22
SiaT	7.1 - 71,000	234	67	28

Supplementary Table S4: Golgi nucleotide sugar concentrations

Nucleotide-Sugar	Starting Concentration of nucleotide sugar (µM)
GDP-Fuc	49.1 - 491,000
UDP-GlcNAc	30.35 - 303,500
UDP-Gal	1.2 - 12,000
CMP-Sia	50 - 500,000

Supplementary Table S5: Enzyme localization

Enzyme	Compartment 1	Compartment 2	Compartment 3	Compartment 4
Man I	0.15	0.4	0.3	0.15
Man II	0.15	0.4	0.3	0.15
GnT I	0.15	0.4	0.3	0.15
GnT II	0.2	0.45	0.2	0.15
GnT III	0.2	0.45	0.2	0.15
GnT IV	0.2	0.45	0.2	0.15
GnT V	0.2	0.45	0.2	0.15
FucT	0.2	0.45	0.2	0.15
GalT	0	0.05	0.2	0.75
SiaT	0	0.05	0.2	0.75

Supplementary Table S6: Model inter-compartment transport rates

From/To	Anterograde	Retrograde
ER/cis	0.1	0
cis/medial	0.1	0
medial/trans	0.1	0
trans/TGN	0.1	0
TGN/cytosol	0.1	0

Supplementary Table S7: Other model parameters

Name:	Value	Unit	Description
qp	185.2	molecules/sec	Incoming protein flux into golgi

Description of Supplementary Datafile 1.

Supplementary Datafile 1 contains parts-level description, complete sequence information, and characterization data for all constructs generated as part of this study. The .xlsx file is organized into four worksheets, titled 'Parts', 'Single Gene Constructs', 'Three Gene Constructs', and 'Glycan Profile', which are described in detail below.

Parts.

Column1: Part Name.

Column2: Part Type. (Promoter, CDS, Terminator, Backbone vector, or Destination vector for Golden Gate assembly)

Column3: Characterization data. (Relative expression values for promoter parts, annotated gene encoded for CDS parts)

Column4: Accession numbers. Transcript accession numbers for CDS parts, GenBank accession numbers for newly generated plasmids.

Column 5: Part Sequence.

Single Gene Constructs.

Column 1: Construct Name. ("AB", "BC", "CD" indicates the Golden Gate scars flanking the construct, followed by the gene name.)

Column 2: Part Composition. (Composition of each single gene construct with "p" identifying the promoter part (example pCMV = CMV promoter), "scar" denoting the 4 bp Golden Gate scar, "pA" indicating the polyadenylation sequence, and "FACS_XX" identifying the vector backbone into which the construct was assembled.

Column 3: Construct Sequence.

Three Gene Constructs.

Column 1: Construct Name.

Column 2: Part Composition – Single gene constructs. (The single gene constructs that constitute the final 3 gene construct)

Column 3: Part Composition – Three gene constructs in destination vector. (Annotation of the final assembly of each three gene construct in the final destination vector CHO_DV)

Column 4: Construct Sequence.

Glycan Profile.

Row 1: Sample.

Row 2: Sample Type. (Control = background cell line, Swap = targeted integration, Three = three gene constructs randomly integrated into the genome, Single = single gene constructs randomly integrated into the genome.

Rows 3-25 and 27-30: Glycan structures and their percent abundance in each sample.