**AUTHOR VERSION**

# A GlycoGene CRISPR-Cas9 lentiviral library to study lectin binding and human glycan biosynthesis pathways

Yuqi Zhu, Theodore Groth, Anju Kelkar, Yusen Zhou and Sriram Neelamegham*

*Chemical and Biological Engineering, Biomedical Engineering and Medicine*
*University at Buffalo, State University of New York, Buffalo, NY 14260, USA*

* Correspondence: Sriram Neelamegham, neel@buffalo.edu, 906 Furnas Hall, Buffalo, NY 14260; Ph: 716-645-1200; FAX: 716-645-3822.

**Running title:** gene editing of glycogenes

**Category**: Analytical Glycobiology

**Keywords:** gene editing/ glycoscience/ CRISPR-Cas9/ forward genetic screen/ selectin

**Supplementary Data Included**: Figures S1-S3 and Tables SI – SVI

**ABSTRACT**

Glycan biosynthesis on cell surface proteins and lipids is orchestrated by different classes of enzymes and proteins including: i. glycosyltransferases that add saccharides, ii. glycosidases that trim glycans, iii. Conserved Oligomeric Golgi complex (COG) members that regulate intracellular transport, iv. enzymes aiding the biosynthesis of sugar-nucleotides, and v. sulfotransferases. This manuscript describes a pooled 'glycoGene CRISPR' lentiviral library that targets 347 human genes involved in the above processes. Approximately 10 single guide RNA (sgRNA) are included against each glycogene, with the putative editing site spanning the length of the target. A data analysis scheme is presented in order to determine glycosylation pathways regulating biological processes. As proof of principle, forward genetic screen results are presented to identify penetrating glycogenes that regulate the binding of P-/E-selectin, anti-sialyl Lewis-X mAb HECA-452 and selected lectins (PHA-L, VVA, PNA) to HL-60 promyelocytic cells. Besides validating previously established biology, the study identifies three enzymes, PAPSS1, SLC35B2 and TPST2, as key molecules regulating sulfation of the major P-selectin glycoprotein ligand, PSGL-1 in leukocytes. ~80-90% of the sgRNA used in this study displayed high editing efficiency and the CRISPR library picked up entire gene sets regulating specific biosynthetic pathways rather than only isolated genes. These data suggest that the glycoGene CRISPR library contains high-efficiency sgRNA. Further, this resource could be useful for the rapid screening of glycosylation related genes and pathways that control lectin recognition in a variety of contexts.

## INTRODUCTION

Glycosylation is a common, complex and prominent post-translational modification (PTM) of various extracellular proteins and lipids, and also intracellular macromolecules. Glycan structures formed as a result regulate almost all biological processes including, but not limited to, protein folding, molecular half-life in blood circulation, cell adhesion, cell signaling and transcription, host-pathogen interactions, inflammation, immunity, tumorigenesis and cancer metastasis (Moremen, K.W., et al. 2012, Varki, A. 2017). Underlying the importance of this PTM, defects in glycosylation result in a family of disorders that are collectively called the congenital disorders of glycosylation (Peanne, R., et al. 2018).

Recent years have witnessed an emphasis on the development of molecular tools to study glycosylation from the systems level, as these complement classical biochemical methods that dissect individual steps (Neelamegham, S. and Liu, G. 2011, Neelamegham, S. and Mahal, L.K. 2016). In this regard, forward genetic screens have emerged as a powerful tool to dissect the major players involved in specific biological processes. While there are a number of readily available genome-wide CRISPR libraries (Shalem, O., et al. 2014, Zhou, Y., et al. 2014), none of these specifically focused on dissecting the human glycosylation process. This is however important since genome-wide screens may miss important steps regulating glycosylation as they focus broadly on ~20,000 human genes. Additionally, smaller libraries result in lower cost due to smaller cell culture scales, sorting expenses and reduced next-generation sequencing requirements. Narimatsu and colleagues (Narimatsu, Y., et al. 2018) described a single-guide RNA (sgRNA) plasmid set that targets 183 human glycosyltransferases. While valuable, only a limited number of sgRNA were screened against each of the genes, and the method used to generate this resource is labor-intensive as sgRNA are individually screened. Other efforts to use CRISPR-Cas9 technology for human glycosylation related studies focus on only a few key enzymes (Mondal, N., et al. 2016, Stolfa, G., et al. 2016), and this also does not capture the complexity of the full system.

To address the above gaps, this manuscript describes a new pooled 'glycoGene CRISPR' library targeting 347 glycosylation related genes or 'glycogenes' (**Supplemental Table SI**). Editing is driven by *Streptococcus pyogenes* Cas9 nuclease. This system allows the timed regulation of gene expression, as it is based on our recently described doxycycline regulated Self-inactivating CRISPR (SiC) vectors (Kelkar, A., et al. 2020). It targets most of the human glycosyltransferases involved in the biosynthesis of different types of glycoconjugates including N-glycans, O-glycans, glycolipids, glycosaminoglycans, GPI-anchored carbohydrates, O-fucose, O-GlcNAc, O-Glucose, O-Mannose and EGF domain specific O-linked GlcNAcTs (EOGTs). Additionally, the library includes sgRNA targeting proteins/enzymes involved in intracellular protein transport, sulfotransferases, enzymes involved in nucleotide-sugar biosynthesis, nucleotide-sugar transporters and a range of glycosidases. A unique feature of the library is that it contains ~9-11 sgRNAs per glycogene that are computationally designed to have high on-target and low off-target editing efficiency. These sgRNA span the length of the glycogenes with an average distribution of 40-40-20 spanning the beginning, middle and end of the average molecular target. Thus, in principle, the library may be used to examine the effect of protein knockout vs. protein truncation in a site-specific manner. The merits and disadvantages of using validated sgRNA vs. forward genetic screens is presented in Table I.

In this manuscript, we used the 'glycoGene CRISPR' library to identify enzymes regulating the binding of specific common lectins and carbohydrate binding mAbs. The results showed that >80-90% of the sgRNA targeting individual glycogenes resulted in loss of lectin/mAb binding function in cell-based assays. While we did not individually validate each guides and gene target in this pooled library, the data thus far suggest that the sgRNA present have high gene editing efficiency. Thus, the glycoGene CRISPR resource could be valuable for the study of human glycosylation processes *in vitro* and *in vivo*.

## RESULTS AND DISCUSSION

**Generating glycoGene CRISPR library transduced HL60s:** A high efficiency CRISPR library was designed to target 347 glycogenes regulating various glycan biosynthetic steps, using the Azimuth algorithm to define sgRNA potency scores for each guide ((Doench, J.G., et al. 2016), **Supplemental Table SI**). These sgRNA were cloned into a doxycycline (Dox) regulated lentiviral vector either downstream of the U6 promoter in 'U6-Dox-sgRNA' or following the H1 promoter in 'H1-Dox-sgRNA' (**Fig. 1A**). This library includes ~10 sgRNA against each target gene, with sgRNA editing sites spanning the length of the protein coding sequence (**Supplemental Table SII**). To test this system, an isogenic HL60 clone was established that stably over-expresses *S. pyogenes* Cas9 ('HL60-Cas9', **Fig. 1B**), using the lentiviral Sic-V1-scr (Fig. 1A). This construct contains a fluorescent dTomato reporter, and thus the HL60-Cas9 cells exhibit red fluorescence. Lentivirus generated using the glycoGene CRISPR library ('U6-Dox-sgRNA' construct) was applied to these cells at various concentrations (**Supplemental Fig. S1**), with the final selected cell line being infected at an MOI (multiplicity of infection) of ~0.4 (Fig. 1B). Thus, most glycoGene CRISPR expressing cells contained ~1 sgRNA/cell. Editing was initiated by addition of 1µg/mL Dox 2 days after glycoGene CRISPR library transduction. Cas9 was subsequently knocked out at day-4 by electroporating with a synthetic guide ('Cas9G7') that self-inactivates the nuclease (Kelkar, A., Zhu, Y., et al. 2020). By day 6, four distinct cell populations were evident with 1/3$^{rd}$ of the cells lacking Cas9 activity based on loss of dTomato signal while simultaneously stably expressing sgRNA based on the BFP reporter expressed by U6-Dox-sgRNA. These BFP+ dTomato- HL60 cells were sorted on day 12 to obtain the 'HL60-lib cells' used in this study. sgRNA representation in these cells followed a bell-shaped profile (**Supplemental Fig. S2**).

To test the efficacy of the sgRNA library, additional 3-color sorts were also performed on day 12 to obtain BFP+ dTomato- HL60 cells that exhibited altered lectin/mAb-binding preference. The specific populations we enriched include cells displaying reduced E- and P-selectin binding **(Fig. 2A, 2B)**, reduced HECA-452 expression, low PHA-L binding **(Fig. 2C, 2D)**, enhanced VVA binding (**Fig. 2E**) and reduced PNA binding (**Fig. 2F**). Since a distinct cell population was not obtained in a single sort due to the low number of genes contributing to the selected function, 1-2 additional sorts were performed within 2-3 weeks to enrich the various 'lectin-selected' cell populations (**Supplemental Fig. S3**).

**Workflow to identify regulators of lectin binding:** Genomic DNA was isolated from HL60-lib cells that exhibited nominal lectin binding properties, and also various enriched populations with altered carbohydrate expression profiles. Regions surrounding the sgRNA sequence in the isolated genomic DNA were PCR amplified, barcoded, mixed and then subjected to deep sequencing. **Supplemental Table SIII** presents raw count data.

Two data analysis methods were applied in order to determine sgRNA over-expressed in each of the 'lectin-selected' cell populations with respect to the control HL60-lib. First, the Model-based Analysis of Genome-wide CRISPR-Cas9 Knockouts package (MAGeCK) identified sgRNAs having significantly higher representation in the lectin-selected pool compared to HL60-lib ((Zhou, Y., Zhu, S., et al. 2014), **Supplemental Table SIV**). Here, 'total sample read count normalization' was performed prior to statistical testing. The False Discovery Rate (FDR<$10^{-3}$) parameter from this analysis identified putative glycogenes altering lectin/mAb binding (Fig. 2 right panels). Second, we computed log fold-change (LFC) values in the selected cell populations, in order to determine sgRNA read counts for target genes with respect to non-target sgRNA control (**Figure 3**). Here, LFC>0 indicates sgRNA enrichment upon lectin/mAb-based selection. Corresponding target genes play a functional role in lectin/mAb binding. LFC$\leq$0 indicates specific guides that are not enriched as their absolute read counts would be similar to non-target control sgRNA. This analysis showed that typically >80-90% of the sgRNA targeting specific glycogenes were increased upon sorting for specific lectin-binding phenotype. Thus, a majority of the sgRNA in the glycoGene CRISPR library displayed high on-target editing efficiency. Further, all the functional genes identified using MAGeCK were also observed to be important in the LFC calculations, thus validating the overall analysis approach (**Supplemental Table SV**).

**Glycogenes regulating selectin recognition**: Among the identified targets, genes involved in GDP-fucose and CMP-Neu5Ac biosynthesis and corresponding nucleotide-sugar transport were critical for E-selectin, P-selectin and HECA-452 binding. This observation is consistent with the well-known roles of fucose and sialic acid during selectin recognition (Cummings, R.D. and McEver, R.P. 2015). Among the glycosyltransferases, the human α(1,3)fucosyltransferase FUT7 and α(2,3)sialyltransferase ST3Gal4 were found to be most penetrating (Buffone, A., Jr., et al. 2013, Mondal, N., et al. 2015). Enzymes involved in

core-1 and core-2 glycan biosynthesis were necessary for P-selectin binding, and this is consistent with the dominant role of such glycans for P-selectin dependent inflammatory leukocyte (Ellies, L.G., et al. 1998, Lo, C.Y., et al. 2013). None of the enzymes involved in LacNAc (N-Acetyllactosamine) and glycolipid biosynthesis were identified to be critical for E-selectin and HECA-452 recognition. However, we noted a role for N-glycan forming glucosidases (GANAB, MOGS), and core-2 GlcNAcT (GCNT1) in regulating HECA-452 epitope expression. It is possible that deep sequencing of cells after the first sort can identify additional partial regulators of selectin-binding, as studies with sequential sorts are only designed to identify critical or 'penetrating' glycogenes.

Importantly, our studies identified the three genes controlling protein tyrosine sulfation that is necessary for P-selectin dependent adhesion in HL60s: PAPSS1, SLC35B2 and TPST2. In this regard, PAPSS1 is one of two PAP synthase isoforms in humans, along with PAPSS2 (van den Boom, J., et al. 2012). It is a bifunctional, cytosolic enzyme that contains both ATP sulfurylase and APS kinase activity (Girard, J.P., et al. 1998). Thus, it both transfers sulfate to ATP to yield APS (adenosine-5'-phosphosulfate) and then transfers the phosphate group from ATP to APS to yield 3'-phosphoadenylylsulfate (PAPS). SLC35B2 is one of two PAPS multi-pass, membrane spanning transporters which, along with SLC35B3, regulates PAPS anti-port into the Golgi (Song, Z. 2013). Finally, TPST2, is also one of two human tyrosylprotein sulfotransferases. Previous work has established an important role for the TPSTs in regulating P-selectin binding to PSGL-1 (Westmuckett, A.D., et al. 2011, Wilkins, P.P., et al. 1995). This study identifies these putative functional genes in human leukocytes.

**Glycogenes regulating lectin binding**: Pilot studies were performed to determine if the glycoGene CRISPR library can be useful for arbitrary lectin binding screens. Here, we confirmed that PHA-L binding is highly dependent on the Galβ1-4GlcNAcβ1-2(Galβ1-4GlcNAcβ1-6)Manα1 arm of N-glycans as knocking out MGAT1, MGAT2 and MGAT5 abolished lectin recognition. MGAT4 and MGAT3 function was not necessary for PHA-L binding. Interestingly, none of the enzymes involved in lipid-linked oligosaccharide biosynthesis were important for PHA-L binding. This warrants investigation as it suggests that the *de novo* synthesis of the dolichol linked $Glc_3$-$Man_9$-$GlcNAc_2$ oligosaccharide via a linear biosynthetic pathway may not be absolutely critical for N-glycan branching in knockouts.

Among the O-glycan binding lectins, the study examined the glycogenes upregulating VVA and reducing PNA binding. Here, PNA binding to HL60s was low to start with and this could be further reduced by abolishing O-glycan elaboration by knocking out core-1 glycan biosynthesis via either the core-1 GalT enzyme C1GALT1, or its chaperone C1GALT1C1. This is consistent with the observation that PNA predominantly binds type III lactosamine chains (Galβ1,3GalNAcα), and that sgRNA targeting O-glycan core-1 biosynthesis work efficiently. As anticipated, VVA binding was dependent on sialic acid biosynthetic pathways, as this lectin recognizes GalNAcα-Ser/Thr, without sialylation. However, surprisingly, cells displaying high VVA binding were not enriched for sgRNA targeting either C1GALT1 or C1GALT1C1 (Chugh, S., et al. 2018, Steentoft, C., et al. 2011, Stolfa, G., Mondal, N., et al. 2016). We speculate that this may be due to the nature of the cell sorting selection method used to obtain VVA-high cells. Thus, the glyco-wide screening approach while efficient in many cases, could sometimes miss key regulators previously identified using more targeted methods.

**CONCLUSIONS**

The manuscript describes a freely available, comprehensive glycoGene CRISPR library that could be useful for human cell systems, both *in vitro* and *in vivo*. Upon generating knockouts, it may be used to identify glycogenes regulating lectin binding and carbohydrate recognition function in a variety of physiological and disease contexts, including tumorigenesis, metastasis, inflammation, thrombosis, neurological disorders etc. Additionally, this CRISPR library has been specifically designed to measure the effect of mutations along the length of the glycogene. Thus, it may be used to study the nature of multi-glycosyltransferase complexes, and the hierarchy of molecular organization in the Golgi. The availability of multiple guides is also useful, if epigenetic changes render portions of the chromatin inaccessible. Only penetrating glycogenes were studied in this manuscript since we focused on proteins that completely abolish specific lectin/mAb binding. Relaxing the selection criterion may enable the identification of partial regulators. Finally, the library system is designed to study the impact of glycogenes *in vivo* in a temporal manner. In this regard, it should be feasible to transplant human cells expressing spCas9 and the glycogene CRISPR library into appropriate animal systems, and then feeding dox-chow to initiate timed gene editing

(Kelkar, A., Zhu, Y., et al. 2020). Similar timed editing trials may also be possible *ex vivo* in different experimental setups, e.g. 3D-organoid models.

All data collected so far in our laboratory suggest that the sgRNA that compose the glycogene CRISPR library have high editing efficiency. Thus, while we did not quantify the efficiency of individual sgRNA, the data support the possibility that sgRNA listed here could also be individually applied in functional assays. Additionally, ~10 guides are provided for each target gene, and an FDR based bioinformatics pipeline is used to account for variability in gene editing efficiency. Using robust rank aggregation (Fig. 2), it is clear that only a small subset of candidates have *FDR*<0.001 and this gives confidence in the overall approach. Due to the high efficiency, in addition to picking up single glycogenes, the library could pick up entire pathways regulating glycosylation.

While library based screening is ideal for hypothesis generation, secondary validation assays are critical for corroboration. Here, it is necessary both to quantify gene editing efficiency using sequencing methods, and to check function using enzymatic/structure-based assays. To enable the selection of sgRNA against specific gene targets for such secondary screens, we provide an online tool at https://virtualglycome.org/gCRISPR/ with data on sgRNA sequence, target location, and sgNA efficiency data when available. Using this tool, if one wishes to edit multiple, redundant or overlapping genes in a pathway, it should be possible to select sgRNA against the target genes of interest, complex with spCas9 to form ribonucleoprotein particles (RNPs), and electroporate into cells. Functional studies can then be performed with these modified cells, with targeted amplicon sequencing of genomic DNA being used to quantify gene editing efficiency (Zhu *et al*., unpublished data). Alternatively, it is also possible to study the impact of multiple genes in a single cell by applying sgRNA containing virus at MOI>1. Overall, the above examples illustrate the many potential applications of the glycogene CRISPR library. Thus, this could be a valuable resource for systems-level investigations of glycosylation.

**MATERIALS AND METHODS**

Supplemental Material provides detailed sources of materials, stepwise methods for creation of the glycoGene CRISPR library/virus and other common molecular biology techniques.

**Generation of HL60-lib and lectin/mAb-binding selected cells:** 1µL of 200× concentrated glycoGene CRISPR lentivirus was added to $3x10^6$ HL60-Cas9 cells in a 6-well plate containing 3mL media in the presence of 8µg/mL polybrene (Buffone, A., Jr., Mondal, N., et al. 2013). Two-days post transduction, 1µg/mL Dox was added to the culture medium to activate gene editing. On day 4, Dox was removed and the Neon transfection kit was used to electroporate $1.5\text{-}2.5\times10^6$ cells suspended in 100µl Resuspension Buffer R with 10µg Cas9G7 sgRNA ((Kelkar, A., Zhu, Y., et al. 2020), Supplemental Material). The electroporation conditions were: 1600V, 10ms, 3 pulses. Following electroporation, the cells were transferred into 6-well plates containing 3mL of complete IMDM culture medium (Gibco™ IMDM 1640 Medium, HEPES with 10% FBS, 1X Antibiotic-Antimycotic and 1% GlutaMAX™). Cells were scaled up till day-12, when FACS sorting was undertaken to select HL60-lib cells containing sgRNA (BFP-positive) but lacking Cas9 (dTomato-negative).

Independent sorts were also undertaken on day 12 for selection of BFP+dTomato- cells that exhibit specific lectin-binding preference. To this end, 10µg/mL fluorescent lectin, 2µg/mL anti-human Cutaneous Lymphocyte Antigen (CLA) mAb (HECA-452-eFluor® 660), or 3-10µg/mL E-/P-selectin-IgG pre-incubated with 1.5µg/mL Alexa-647 conjugated Goat anti-Human IgG $F(ab')_2$ were added to $2x10^6$/ml HL60s containing the sgRNA library on ice for 20 min. Three color flow sorting was then performed to enrich for BFP+dTomato- cells exhibiting specific lectin/mAb-binding preference. One additional cell sorts was performed during the following week to enrich for the cell population with minimum interaction with E-selectin, P-selectin, HECA, PHA-L and PNA. Two sorts were required in order to obtain a distinct high VVA-binding population. Gates used for each sort are shown in Supplemental Fig. S3.

**Deep sequencing:** Genomic DNA was isolated from $\sim5\times10^6$ cells (HL60-lib and lectin-selected) using PureLink™ Genomic DNA Mini Kit (Invitrogen™). A two-step PCR was applied to generate amplicons for deep sequencing. First, regions flanking the sgRNA were PCR amplified by mixing 1µg of isolated genomic DNA with 1µM gFwd primer, 1µM gRev primer and 25µl NEBNext® Ultra™ II Q5® Master Mix (New England Biolabs) and water to make up a 50µl reaction (primer sequences in **Supplemental Table SVI**). Product size was checked using 2% agarose gel electrophoresis, followed by PCR column purification using NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel, Bethlehem, PA). Second, the Illumina P5 adapter was added to the 5'-end of the above PCR product and P7 adapters along with a sample specific barcode to the 3'end, in a similar PCR as above using ~15µl amplicon from first step as template, 1µM iFwd and 1µM iRev primer (Supplemental Table SVI). All PCR products were gel purified, pooled and sequenced using an illumina MiSeq platform (San Diego, CA): 150bp paired end, $30x10^6$ reads.

**Bioinformatics analysis:** Deep sequencing data were analyzed using two methods. First, the Model-based Analysis of Genome-wide CRISPR-Cas9 Knockouts package (MAGeCK) was used to identify sgRNA responsible for altered lectin/mAb binding (Zhou, Y., Zhu, S., et al. 2014). Briefly, MAGeCK determines the number of sgRNA in the HL60-lib and lectin-selected samples by using the sgRNA amplicon read .fastq data and a reference sgRNA sequence file containing all possible guide sequences. These data are normalized based on the total sample read count: $sgRNA_{i,norm}=10^6\times(sgRNA_i/\Sigma\ sgRNA_i)$, where $sgRNA_i$ is the raw read count of the $i^{th}$ sgRNA in a given cell sample. Statistical analysis was then performed to determine if sgRNA read count differences between samples (lectin-selected vs. HL60-lib) is significantly different. These enrichment likelihoods (*P*-values), along with the corresponding sgRNA $\log_2$-fold change values are used to rank the sgRNAs detected firstly from high to low-fold change, then based on low to high *P*-value. This ranked list is used in an α-robust rank aggregation analysis to identify the gene knockouts that produce the relevant phenotype. The *P*-values computed from this analysis are adjusted for false discovery using the Benjamini Hochberg procedure. These adjusted *P*-values (FDR values) are used to sort genes based on their relevance to the phenotype in ascending order.

Second, in the fold-change calculations, we measured the ratio of target sgRNA counts in each cell type versus non-target sgRNA counts in the same cell. Here, we used sgRNA corresponding to carbohydrate sulfation (all CHSTs) and heparan sulfation (all HS6STs, HS3STs and HS2STs) enzymes as non-target control. Representation of sgRNA corresponding to these genes is not expected to be altered

for the lectins used in this study. These data are plotted as $\log_2$ fold-change. Here, sgRNA enrichment results in $\log_2$ fold-change>0, and depletion in $\log_2$ fold-change$\leq$0.

**Acknowledgements:** We gratefully thank Edward J. Sobczak and Srirangaraj Setlur for systems administration of the virtualGlycome server. This work was supported by NIH grants HL103411 and GM133195, and a New York Stem Cell training program award DOH01-C30290GG-3450000. The glycoGene CRISPR libraries is distributed via Addgene. For the sgRNA selection tool, please visit: https://virtualGlycome.org/gCRISPR/.

**Conflicts of interest**: None declared

## REFERENCES


Buffone A, Jr., Mondal N, Gupta R, McHugh KP, Lau JT, Neelamegham S. 2013. Silencing alpha1,3-fucosyltransferases in human leukocytes reveals a role for FUT9 enzyme during E-selectin-mediated cell adhesion. *J Biol Chem*, 288:1620-1633.

Chugh S, Barkeer S, Rachagani S, Nimmakayala RK, Perumal N, Pothuraju R, Atri P, Mahapatra S, Thapa I, Talmon GA*, et al.* 2018. Disruption of C1galt1 Gene Promotes Development and Metastasis of Pancreatic Adenocarcinomas in Mice. *Gastroenterology*, 155:1608-1624.

Cummings RD, McEver RP. 2015. C-Type Lectins. In: rd, Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Darvill AG, Kinoshita T, Packer NH, et al. editors. *Essentials of Glycobiology*. Cold Spring Harbor (NY). p. 435-452.

Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg 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.

Ellies LG, Tsuboi S, Petryniak B, Lowe JB, Fukuda M, Marth JD. 1998. Core 2 oligosaccharide biosynthesis distinguishes between selectin ligands essential for leukocyte homing and inflammation. *Immunity*, 9:881-890.

Girard JP, Baekkevold ES, Amalric F. 1998. Sulfation in high endothelial venules: cloning and expression of the human PAPS synthetase. *FASEB J*, 12:603-612.

Kelkar A, Zhu Y, Groth T, Stolfa G, Stablewski AB, Singhi N, Nemeth M, Neelamegham S. 2020. Doxycycline-Dependent Self-Inactivation of CRISPR-Cas9 to Temporally Regulate On- and Off-Target Editing. *Mol Ther*, 28:29-41.

Lo CY, Antonopoulos A, Gupta R, Qu J, Dell A, Haslam SM, Neelamegham S. 2013. Competition between core-2 GlcNAc-transferase and ST6GalNAc-transferase regulates the synthesis of the leukocyte selectin ligand on human P-selectin glycoprotein ligand-1. *J Biol Chem*, 288:13974-13987.

Mondal N, Buffone A, Stolfa G, Antonopoulos A, Lau JTY, Haslam SM, Dell A, Neelamegham S. 2015. ST3Gal-4 is the primary sialyltransferase regulating the synthesis of E-, P-, and L-selectin ligands on human myeloid leukocytes. *Blood*, 125:687-696.

Mondal N, Stolfa G, Antonopoulos A, Zhu Y, Wang SS, Buffone A, Jr., Atilla-Gokcumen GE, Haslam SM, Dell A, Neelamegham S. 2016. Glycosphingolipids on Human Myeloid Cells Stabilize E-Selectin-Dependent Rolling in the Multistep Leukocyte Adhesion Cascade. *Arterioscler Thromb Vasc Biol*, 36:718-727.

Moremen KW, Tiemeyer M, Nairn AV. 2012. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol*, 13:448-462.

Narimatsu Y, Joshi HJ, Yang Z, Gomes C, Chen YH, Lorenzetti FC, Furukawa S, Schjoldager KT, Hansen L, Clausen H*, et al.* 2018. A validated gRNA library for CRISPR/Cas9 targeting of the human glycosyltransferase genome. *Glycobiology*, 28:295-305.

Neelamegham S, Liu G. 2011. Systems glycobiology: biochemical reaction networks regulating glycan structure and function. *Glycobiology*, 21:1541-1553.

Neelamegham S, Mahal LK. 2016. Multi-level regulation of cellular glycosylation: from genes to transcript to enzyme to structure. *Curr Opin Struct Biol*, 40:145-152.

Peanne R, de Lonlay P, Foulquier F, Kornak U, Lefeber DJ, Morava E, Perez B, Seta N, Thiel C, Van Schaftingen E*, et al.* 2018. Congenital disorders of glycosylation (CDG): Quo vadis? *Eur J Med Genet*, 61:643-663.

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.

Song Z. 2013. Roles of the nucleotide sugar transporters (SLC35 family) in health and disease. *Mol Aspects Med*, 34:590-600.

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 YQ, Yu XH, Buffone A, Neelamegham S. 2016. Using CRISPR-Cas9 to quantify the contributions of O-glycans, N-glycans and Glycosphingolipids to human leukocyte-endothelium adhesion. *Scientific Reports*, 6.

van den Boom J, Heider D, Martin SR, Pastore A, Mueller JW. 2012. 3'-Phosphoadenosine 5'-phosphosulfate (PAPS) synthases, naturally fragile enzymes specifically stabilized by nucleotide binding. *J Biol Chem*, 287:17645-17655.

Varki A. 2017. Biological roles of glycans. *Glycobiology*, 27:3-49.

Westmuckett AD, Thacker KM, Moore KL. 2011. Tyrosine sulfation of native mouse Psgl-1 is required for optimal leukocyte rolling on P-selectin in vivo. *PLoS One*, 6:e20406.

Wilkins PP, Moore KL, McEver RP, Cummings RD. 1995. Tyrosine sulfation of P-selectin glycoprotein ligand-1 is required for high affinity binding to P-selectin. *J Biol Chem*, 270:22677-22680.

Zhou Y, Zhu S, Cai C, Yuan P, Li C, Huang Y, Wei W. 2014. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature*, 509:487-491.

**Table I: Two approaches to study glycogene function**

| | Validated guides | Forward genetic screen |
|---|---|---|
| **Advantage** | • Requires less bioinformatics knowledge<br>• Good approach when there is one or a few targets to study. Well suited for hypothesis validation | • sgRNA libraries can be rapidly designed using established bioinformatics pipelines<br>• Same library can be reused for multiple screens. Ideal for hypothesis generation |
| **Disadvantage** | • Guide validation can be expensive and time consuming as individual genes have to be screened.<br>• Guide validated in one system may not work in others due to epigenetic differences | • Quality of results depends on the quality of sgRNA<br>• Some genes may be missed when library size is large |

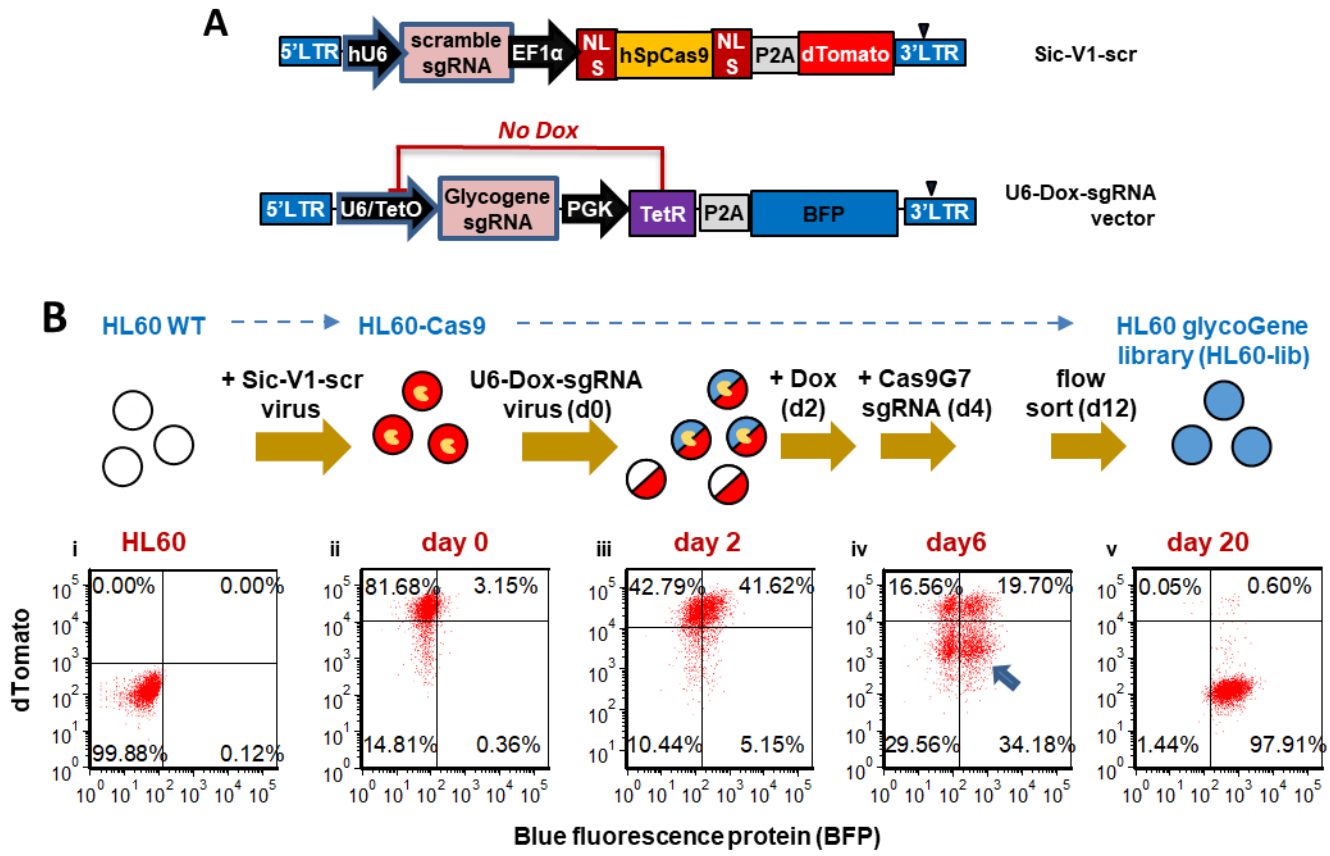



**Figure 1. Creation of GlycoGene CRISPR library transduced cells**. **A.** Schematic of lentiviral vectors used in the CRISPR screen. dTomato acts as a surrogate reporter of Cas9 nuclease activity in SiC-V1-scr. GlycoGene CRISPR library was cloned in a doxycycline-inducible vector with BFP fluorescence reporter. **B.** An HL60-Cas9 cell line was established by transduction with SiC-V1-scr vector, and single cell flow sorting for isogenic clone expressing high levels of dTomato reporter (panel **Bi-ii**). A virus pool synthesized from the glycoGene CRISPR sgRNA library (U6-Dox-sgRNA) was applied to these cells. Two days later, 42% of the cells were BFP positive (panel **Biii**). 1µg/mL Dox was added starting at day 2 to initiate gene editing. Dox was removed and Cas9G7 sgRNA electroporated on day 4 to inactivate Cas9 activity. dTomato- BFP+ cells (arrow) were FACS sorted at day 12 in order to establish the stable HL-60 glycoGene cell library (called 'HL60-lib'). Flow cytometry plots represent: **i.** wild-type HL60; **ii.** HL60-Cas9; **iii.** HL60-Cas9 with glycoGene library; **iv.** Cells after Cas9 nuclease editing on day 6; **v.** HL60-lib cells on day 20. dTomato fluorescence of HL60-lib on day 20 (**Bv**) is similar to wild-type HL60s (**Bi**) confirming complete knockdown of Cas9 activity.

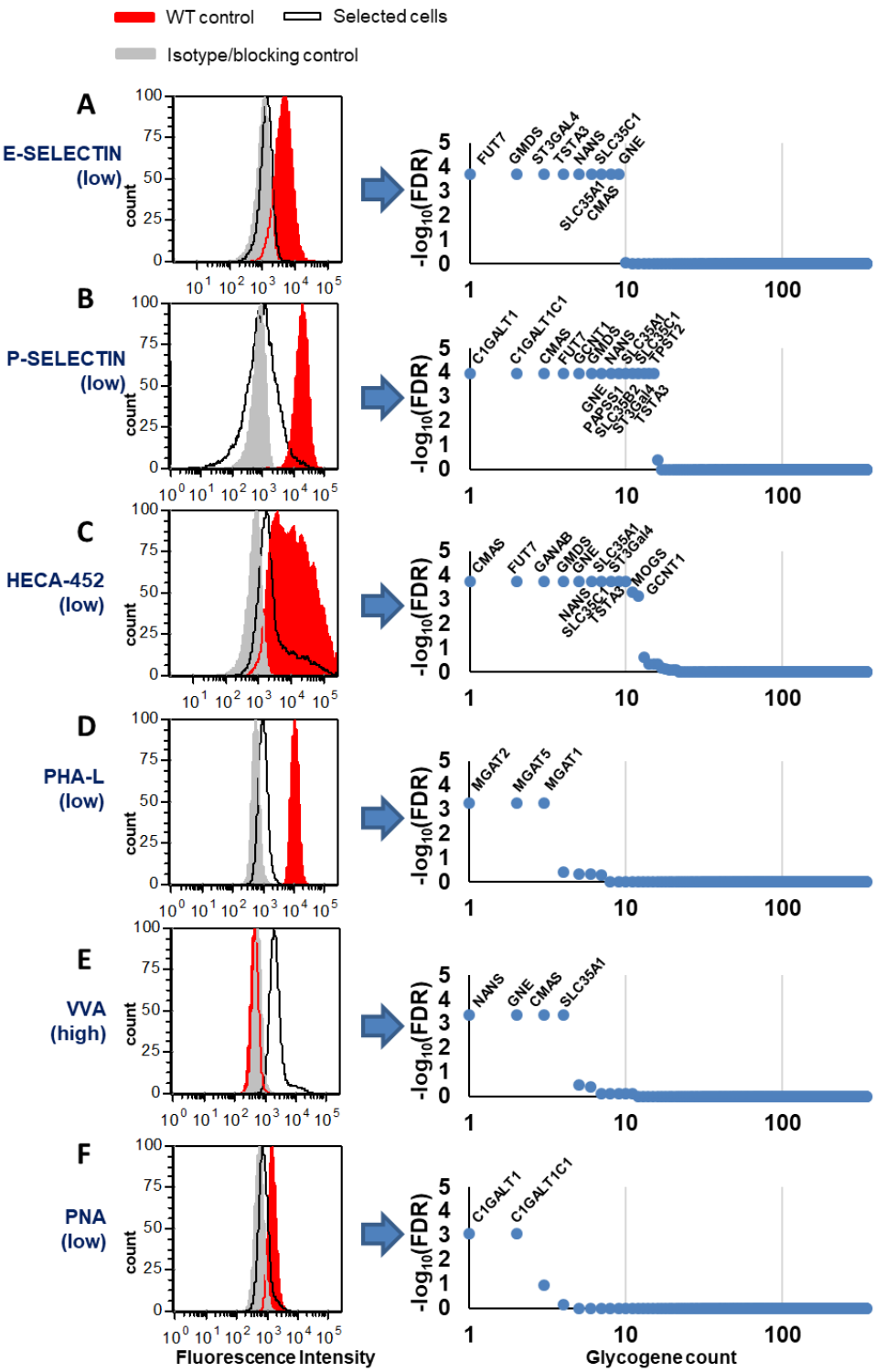



**Figure 2. GlycoGene screening for altered lectin binding using MAGeCK**. **A.** HL60 cells containing the full glycoGene library (i.e. HL60-lib, $BFP^{+}dTomato^{-}$) were sorted on day 12. On the same day, three color sorts were also performed to select for $BFP^{+}dTomato^{-}$ cells exhibiting low E-selectin IgG binding (panel **A**), low P-selectin-IgG binding (panel **B**), low HECA-452/sialyl Lewis-X expression (panel **C**), low PHA-L binding (panel **D**), high VVA binding (panel **E**) and low PNA binding (panel **F**). Either one or two additional sorts were performed to obtain 'lectin-selected' cell populations (left panels). Red histogram marks the binding of HL60-Cas9 control, grey histogram marks the negative control (isotype/blocking antibody), and black line marks the final lectin-selected population. sgRNA enriched in the lectin-selected population were identified using MAGeCK following deep sequencing (right panels). Genes identified with FDR (false discovery rate) $<10^{-3}$ are marked in the individual panels, next to the relevant data point.

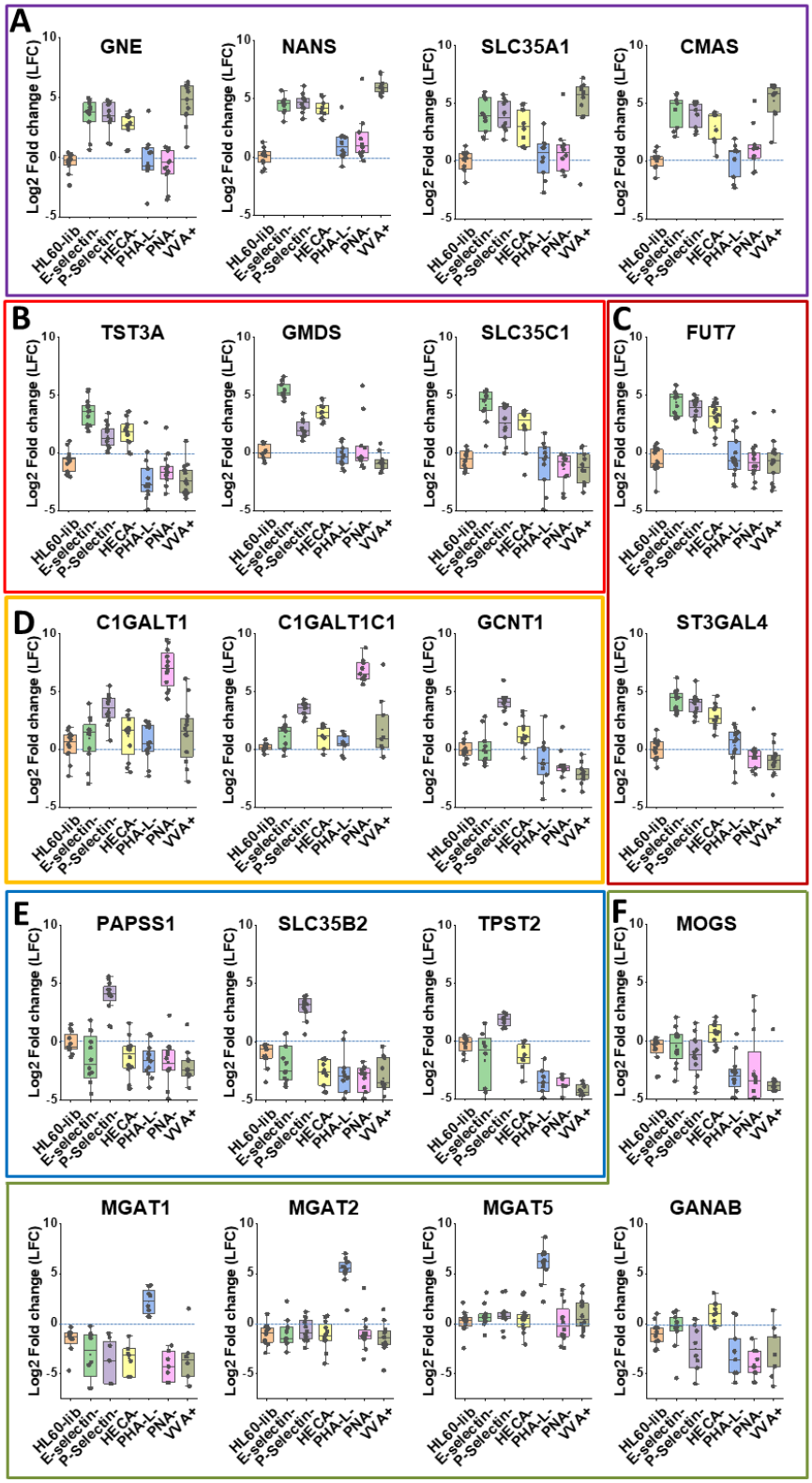



**Figure 3. Log fold change (LFC) analysis**. Log2-fold change sgRNA data are presented for target gene vs. non-target control for the unselected HL60-lib cells, and also the six lectin-selected cell types, i.e. LFC= $\log_2$[sgRNA count for target gene/ sgRNA count for non-target gene]. These data are presented for 20 different glycogenes that are involved in: **A.** CMP-sialic acid biosynthesis; **B.** GDP-Fucose biosynthesis; **C.** α(1,3)fucosyltransferase FUT7 and α(2,3)sialyltransferase ST3Gal-4 activity; **D.** O-glycan construction; **E.** protein tyrosine sulfation; and **F.** N-glycan elaboration. Each dot represents one sgRNA. Enriched sgRNA display LFC>0. Unaffected sgRNA have LFC≤0, similar to the HL60-lib control.

**SUPPLEMENTAL MATERIAL**

**Contents**

**SUPPLEMENTAL METHODS**

**Materials**: All lectins were from Vector labs (Burlingame, CA), primary antibodies from ThermoFisher (Waltham, MA), recombinant human selectin-IgG fusion proteins from R&D Systems (Minneapolis, MN) and secondary Abs from Jackson Immunoresearch (West Grove, PA). HL60 cells were purchased from American Type Culture Collection (Manassas, VA). These cells were transduced with a virus generated using SiC-V1-scr (Kelkar, A., et al. 2020), and single cell sorted to obtain an isogenic clone called 'HL60-Cas9'. These cells express uniform high levels of the nuclease based on expression of the surrogate dTomato fluorescence reporter.

**Dox-inducible sgRNA vector construction:** The SiC-V2-Cas9G7 vector (Addgene 133043) was available from our previous study (Kelkar, A., Zhu, Y., et al. 2020). This vector expresses a tetracycline/Dox-regulated H1 promoter that drives sgRNA expression, a Tet-repressor to regulate this Pol III promoter, and a Cerulean (monomeric Cyan Fluorescence Protein/CFP) fluorescence reporter. Site directed mutagenesis was performed on SiC-V2-Cas9G7 to introduce a silent mutation that deactivates the single BsmBI site. This new plasmid is called "SiC-V2 mutBsmBI". The BsmBI-Stuffer-ChRNA-BsmBI region from LentiCRISPR V2 plasmid (Addgene 52961, (Sanjana, N.E., et al. 2014)) was amplified using AgeI/EcoRI handles and cloned into the AgeI/EcoRI site of "SiC-V2 mutBsmBI" just downstream of H1. Next, BFP (blue fluorescence protein) was PCR amplified from pKLV2-U6gRNA5(BbsI)-PGKpuro2ABFP-W (Addgene 67974, (Tzelepis, K., et al. 2016)), and it replaced the Cerulean encoding cassette in "SiC-V2 mutBsmBI". The resulting vector allows doxycycline-inducible expression of sgRNA from H1 promoter. A variant was also created where the H1 was replaced by the U6 promoter. These vectors are called 'H1-Dox-sgRNA', and 'U6-Dox-sgRNA'.

**CRISPR guide selection and library construction**: The glycoGene CRISPR library includes ~8-11 high potency sgRNA against 347 gene targets related to cellular glycosylation. These sgRNA were selected both from previously designed high on-target/low off-target 'second-generation sgRNA' that are part of other whole-genome CRISPR libraries (Doench, J.G., et al. 2016, Read, A., et al. 2017, Shalem, O., et al. 2014, Tzelepis, K., Koike-Yusa, H., et al. 2016, Wang, T., et al. 2015) and additional sequences generated using the Genetic Perturbation Platform (GPP) service (https://portals.broadinstitute.org/gpp/public/analysis-

tools/sgrna-design-help). The fitness of all sgRNA was evaluated using the stand-alone sgRNA potency scoring algorithm, Azimuth (Doench, J.G., Fusi, N., et al. 2016). This algorithm computes a score between 0 and 1, where 1 indicates a high likelihood of potency, and 0 indicates a low likelihood of potency. The score is computed by weighting the influence of base positions within the protospacer as well as its context bases (the bases preceding and succeeding the protospacer). Guides with a potency score >0.35 were considered to be active, and included in the glycoGene CRISPR library.

The library was synthesized as a pool of oligos (91 bases long) by CustomArray Inc. Each oligo consists of the 20-mer sgRNA sequence (Supplemental Table S2) flanked by BsmBI cutting sites, and a pair of PCR primer sequences at the termini (Primer 1F: TAAAGGGACGGTTAATCGCTCACC and Primer 105R: GGTCTGAACGTAAGGAAGTGTGGA). The library was amplified using Accuprimer PfX supermix with 1µl each of 100 µM 1F and 100 µM 105R in a 50 µl reaction. 5 µl amplicon was loaded on a 3% agarose gel to confirm product size and remaining amplicon was column purified using NucleoSpin gel and PCR clean-up kit (Macherey-Nagel).

The cloning vector plasmids (H1-Dox-sgRNA, and U6-Dox-sgRNA) were digested with BsmBI and the vector backbone was gel extracted. Golden gate ligation of sgRNA oligos with both vector plasmid backbones ('H1-Dox-sgRNA', and 'U6-Dox-sgRNA') was performed in two independent reactions, in the presence of Esp31, T7 DNA ligase and ATP using the following cycling conditions: 37ºC 2h, 9 cycles (20ºC, 20 min; 37ºC, 5 min) and 4ºC hold. After column purification, 2 µl of the ligation product was electroporated into NEB-10 beta electrocompetent cells using Bio-Rad MicroPulser. Electroporated cells were plated on a 24.5 $cm^2$ LB agar (+100 µg/ml Carbenicillin) plate and incubated overnight at 37ºC. The bacterial lawn was gently scraped from the plate using 10 ml LB medium and the process was repeated 2 more times to retrieve all the bacterial growth into a 30 ml pool that was incubated at 37ºC for 2 h for bacterial outgrowth. 2/3rd of the culture was used for making Glycerol stocks of the library and 1/3rd diluted into 150 ml of LB medium containing Carbenicillin. Following overnight incubation at 200 rpm, 37ºC, library plasmid was prepared from the bacterial culture using the NuceloBond Xtra Midi Kit (Macherey-Nagel).

**Glycogene library virus preparation**: Lentivirus was made by mixing 17 µg pooled library plasmid, 21 µg of psPAX2 packaging plasmid and 8.7 µg of pMD2.G envelope plasmid with 125mM $Ca^{2+}$, 1mL 2X HBS

(HEPES-buffered saline, 50mM HEPES+140mM NaCl, 10mM KCl, 1.5 mM $Na_2HPO_4$, pH 7.05), and sterile water to make a 2mL reaction mix. Reaction mix was added drop by drop onto a 150mm culture dish plated with Lenti-X 293T cells (Takara Bio, Mountain View, CA) at 60-70% confluence. 8h later, cell culture medium was changed to 20mL Opti-MEM™ (ThermoFisher). 18hrs later, the first batch of virus was harvested. 15mL fresh Opti-MEM™ was fed to the cells with 10mM sodium butyrate (MilliporeSigma, St. Louis, MO). 18hrs later, a second batch of virus was harvested. After spinning down at 1,000g for 2 min, supernatant was filtered using 0.45µm Nalgene™ Rapid-Flow™ Sterile Disposable Filter Units with PES Membrane (Thermofisher). The filtered virus was pelleted at 50,000g for 2h at 4°C. The supernatant was removed and the final product was resuspended in 175µl volume (i.e. 200X concentrated virus). Aliquots of concentrated virus were stored at -80°C until use.

**Synthesis of Cas9G7 sgRNA**: A custom DNA oligo, 5'-**TTCTAATACGACTCACTATAG**TACGCCGGCTACATTGACGG**GTTTTAGAGCTAGA,** was purchased that contained the T7 promoter (red) followed by a 20bp guide sequence against Cas9 (green, called 'Cas9G7' (Kelkar, A., Zhu, Y., et al. 2020)) and the first several bases of the sgRNA scaffold (blue). The EnGen sgRNA Synthesis Kit (New England Biolabs) was used to produce the corresponding synthetic sgRNA. To this end, the above oligonucleotide was mixed with T7 enzyme and reaction mix provided with the kit for 2 h at 37ºC. The product formed was DNAase treated, and the RNA was purified using the RNA Clean & Concentrator kit (Zymo Research, Irvine, CA). The final product was stored at -80ºC until use.

**REFERENCES**

Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg 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.

Kelkar A, Zhu Y, Groth T, Stolfa G, Stablewski AB, Singhi N, Nemeth M, Neelamegham S. 2020. Doxycycline-Dependent Self-Inactivation of CRISPR-Cas9 to Temporally Regulate On- and Off-Target Editing. *Mol Ther*, 28:29-41.

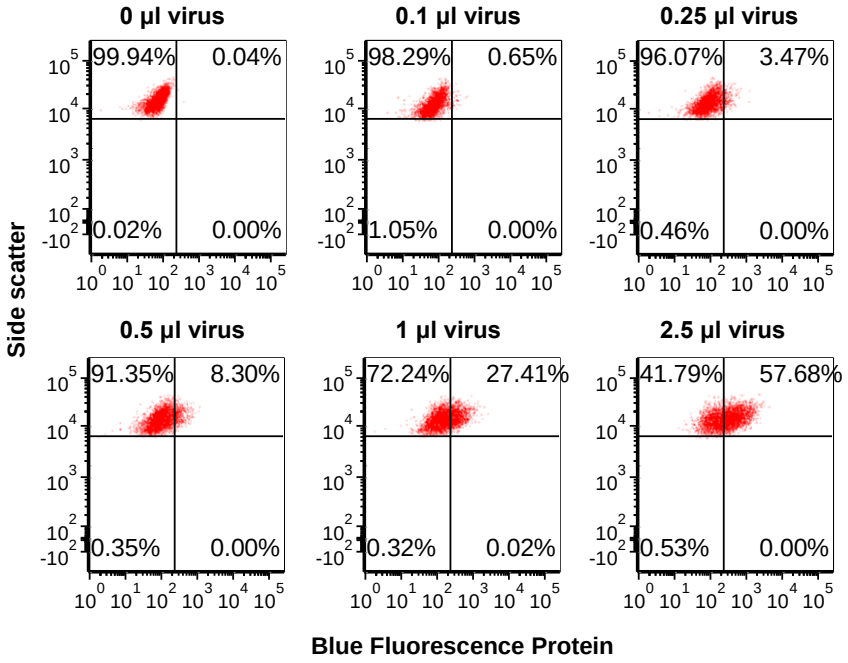
Read A, Gao SJ, Batchelor E, Luo J. 2017. Flexible CRISPR library construction using parallel oligonucleotide retrieval. *Nucleic Acids Research*, 45.

Sanjana NE, Shalem O, Zhang F. 2014. Improved vectors and genome-wide libraries for CRISPR screening. *Nature Methods*, 11:783-784.

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.

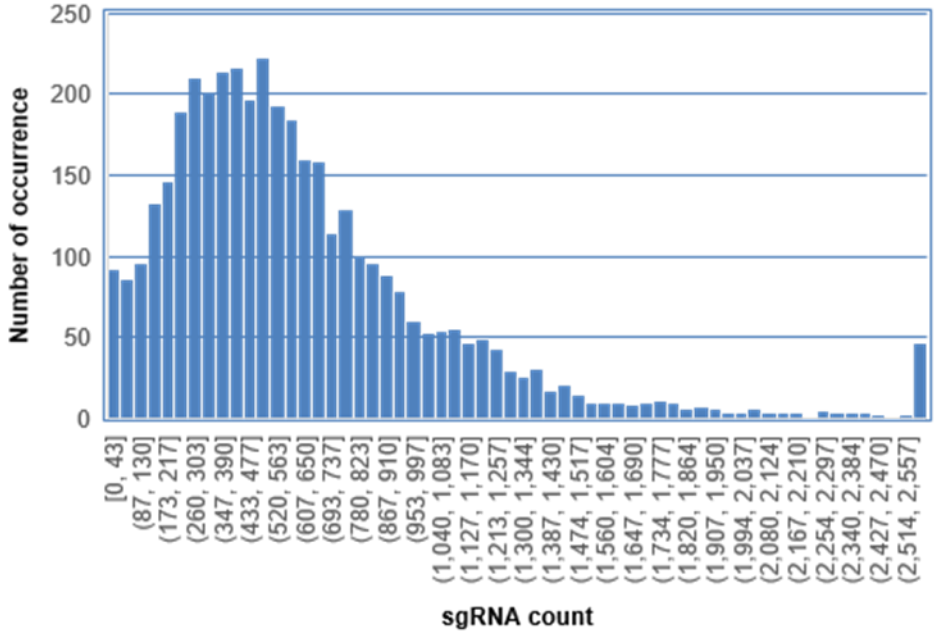
Tzelepis K, Koike-Yusa H, De Braekeleer E, Li Y, Metzakopian E, Dovey OM, Mupo A, Grinkevich V, Li M, Mazan M*, et al.* 2016. A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. *Cell Rep*, 17:1193-1205.

Wang T, Birsoy K, Hughes NW, Krupczak KM, Post Y, Wei JJ, Lander ES, Sabatini DM. 2015. Identification and characterization of essential genes in the human genome. *Science*, 350:1096-1101.



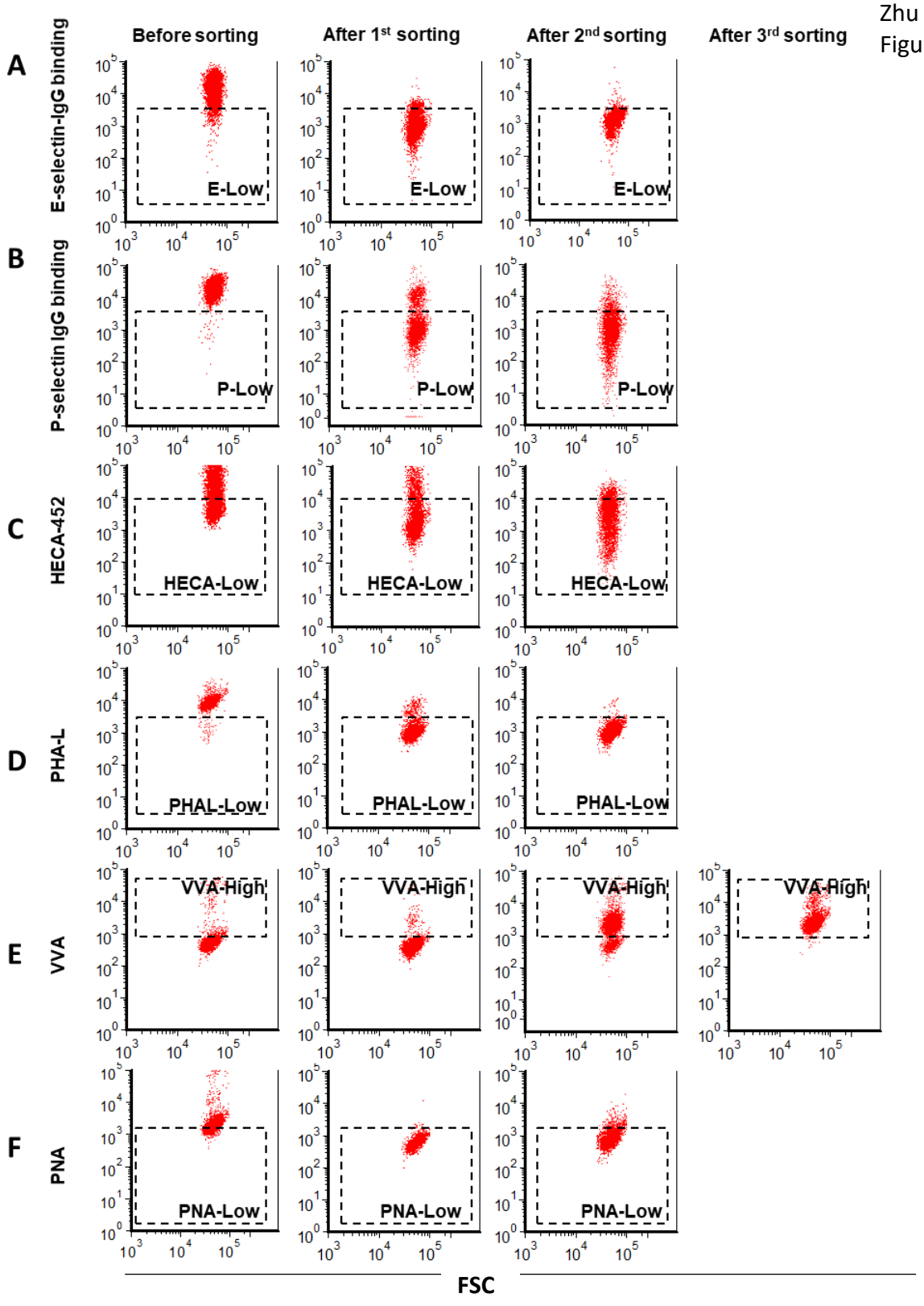

**Supplementary Fig S1. Glycogene CRISPR virus titration**. In a pilot run, $3\times10^6$ HL-60 cells were transduced in 6-well plates (3mL/well) using the glycoGene CRISPR library lentivirus that carries a BFP reporter. GlycoGene CRISPR virus concentrate volume was varied from 0-2.5µL, as shown in individual panels. % transduced cells varied from 0-57.68% on day 2. Based on this pilot study, 1µL virus (MOI~0.27) was applied in the final transduction that was scaled up and used to create the HL60-lib and enriched cell populations described in this manuscript.

Zhu et al.
Figure S2




**Supplemental Fig. S2. Guide representation in HL60-lib**. The full glycoGene CRISPR library contains 3976 sgRNA targeting 347 genes. The figure provides the sgRNA distribution in the HL60-lib cells based on Next Generation Sequencing (NGS). 80 of the sgRNA had low representation (<43 counts, left corner), while 45 genes had high representation (>2557 counts, piled at the right corner). Raw counts are presented in Supplemental Table S3.

Zhu et al.
Figure S3




**Supplementary Fig S3. Serial sorting of transduced HL60s**. HL60 glycogene transduced cell populations exhibiting low E-selectin binding (**panel A**), low P-selectin binding (**panel B**), low HECA-452 expression (**panel C**) and low PHA-L binding (**panel D**) were FACS sorted by gating on the box indicated using dashed line. Two serial sorts were performed as shown to enrich for these cell populations, with the first sort being performed on day 12 after viral transduction. Similar serial sorts were performed to enrich for cell populations exhibiting high VVA binding (**panel E,** 3-sorts) and low PNA recognition (**panel F**). In each case, a distinct low/high binding cell population was evident after the first sort, and this was progressively enriched in subsequent sorts.