



## Supplementary Materials for

### Deciphering the Glycosylome of Dystroglycanopathies Using Haploid Screens for Lassa Virus Entry

Lucas T. Jae,<sup>1</sup> Matthijs Raaben,<sup>2</sup> Moniek Riemersma,<sup>3,4,5</sup> Ellen van Beusekom,<sup>5</sup> Vincent A. Blomen,<sup>1</sup> Arno Velds,<sup>1</sup> Ron. M. Kerkhoven,<sup>1</sup> Jan E. Carette,<sup>6</sup> Haluk Topaloglu,<sup>7</sup> Peter Meinecke,<sup>8</sup> Marja W. Wessels,<sup>9</sup> Dirk J. Lefeber,<sup>3,4</sup> Sean P. Whelan,<sup>2</sup> Hans van Bokhoven,<sup>5</sup> Thijn R. Brummelkamp<sup>1,10</sup>

\*To whom correspondence should be addressed. E-mail: sean\_whelan@hms.harvard.edu (S.P.W.); H.vanBokhoven@gen.umcn.nl (H.V.B.); t.brummelkamp@nki.nl (T.R.B.)

Published 21 March 2013 on *Science Express*  
DOI: 10.1126/science.1233675

#### This PDF file includes

Materials and Methods  
Supplementary Text  
Figs. S1 to S13  
Full References

#### Other Supplementary Material for this manuscript includes the following:

(available at [www.sciencemag.org/cgi/content/full/science.1233675/DC1](http://www.sciencemag.org/cgi/content/full/science.1233675/DC1))

Tables S1 to S4 as Excel tables

**Table S1.** Disruptive gene-trap insertions mapped in unselected control population of HAP1 cells.

**Table S2.** Insertions, by gene, predicted to be inactivating mapped in HAP1 cells that were selected with rVSV-GP-LASV.

**Table S3.** Insertions, by gene, predicted to be inactivating mapped in HAP1 cells that were immunodepleted for the laminin-binding epitope of  $\alpha$ -dystroglycan.

**Table S4.** Insertions, by gene, predicted to be inactivating mapped in HAP1 cells that were immunodepleted for heparan sulfate.

## **Supplementary materials**

### **Experimental procedures**

#### **Cells**

HAP1 cells (11) used for haploid genetic screens and follow-up were cultured in IMDM supplemented with 10% FCS, L-glutamine, and penicillin–streptomycin. Vero cells (American Type Culture Collection, Manassas, Virginia, USA) and primary human dermal fibroblasts were maintained in DMEM supplemented with 10% FCS, L-glutamine and penicillin–streptomycin. Vero cells were used for production of rVSV-GP-LASV and rVSV-GP-EBOV. Human fibroblasts were used to validate dependence of rVSV-GP-LASV on glycosylated  $\alpha$ -DG (the receptor for authentic Lassa virus) to enter the cell.

#### **Viruses**

Recombinant VSV expressing eGFP and the Lassa virus glycoprotein (rVSV-GP-LASV) was cloned and recovered as follows: the open reading frame encoding LASV-GP (strain Josiah, GenBank: HQ688673.1) was amplified using the following primer sequences: 5'-GCGACGCGTACCATGGGACAAATAGTGACATTCT-3' and 5'-GGCGGCCGCTCATCTCTTCCATTTCACAGG-3'. Subsequently, the PCR product was sequenced and cloned into the MluI and NotI sites of pVSV $\Delta$ G-eGFP-MN (26, 27) thereby replacing the native VSV glycoprotein G coding sequence. Recombinant virus was recovered and amplified as described (26). The construction of recombinant VSV expressing eGFP and the Ebola virus glycoprotein (rVSV-GP-EBOV) has been described previously (27).

#### **rVSV infectivity assays**

The infectivity of rVSV-GP-LASV and rVSV-GP-EBOV was assessed microscopically by the number of eGFP-positive cells 6-7 h post infection. Nuclei were stained as a cell marker using Hoechst33342 (Invitrogen, Carlsbad, California, USA).

#### **Flow cytometry**

HAP1 cells were detached with 5mM EDTA in PBS before incubation with mouse anti-human glycosylated alpha-dystroglycan IIIH6-C4 (Millipore, Billerica, Massachusetts, USA) or mouse anti-human heparan sulfate 10E4 (US Biological, Salem, Massachusetts, USA) in PBS 5% BSA. Primary antibodies were labeled using goat anti-mouse Alexa Fluor 568 or goat anti-mouse Cy5 antibodies

(Invitrogen). Measurements were carried out on a BD Fortessa flow-cytometer (BD, Franklin Lakes, New Jersey, USA). Cell sorting was performed on a BD FACS Aria flow-cytometer (BD). Data analysis was completed using BD FACS Diva software.

### **Laminin overlay assay and immunoblotting**

Cells were detached with 5mM EDTA in PBS. Ca.  $2.5 \times 10^7$  cells were subjected to wheat germ agglutinin (WGA)-enrichment, after which laminin overlay and  $\beta$ -dystroglycan western blotting were performed as described previously (3, 28).  $\alpha$ -dystroglycan and  $\beta$ -dystroglycan were detected using the GTX105038 antibody (GeneTex, San Antonio, Texas) directed against amino acids 471-752 of the dystroglycan precursor. LAMP1 was detected using a mouse monoclonal antibody (H4A3, Santa Cruz biotechnology, Dallas, Texas, USA).

### **Generation of null alleles using TALEN technology**

TALENs targeting exonic sequences shared among all protein-coding transcripts of the respective genes were generated as described (22). The targeted sequences are:

*DAG1*: 5'-TGGCCCAGTGAACCCCTCAGA-3' + 5'-AAACCAGCTTGAGGCATCCA-3'

*C3orf39*: 5'-TCTTCTTCATGGAGGGCTGG-3' + 5'-CGACCTCTACAAGCTGCTCA-3'

*SGK196*: 5'-TCCACTGTGGACCCACACA-3' + 5'-TTCAGGATAGGACAGATGAA-3'

*SLC35A1*: 5'-TCAGCAAATTACAGTGGGTT-3' + 5'-GTGCTGGAGTTACGCTTGTA-3'

*TMEM5*: 5'-TCTTAAGAATTACCAAGATG-3' + 5'-TCACATTGTGCCCGGTCGGA-3'

*B3GALNT2*: 5'-TTTCCTCAGTGGAATCTAC-3' + 5'-GGCGTGTTGTCAGCTCGAA-3'

*B3GNT1*: 5'-TGTCGGTGTTCGCGGCCACC-3' + 5'-CTGGCCACGGTGCTGGCCTA-3'

*ST3GAL4*: 5'-TCCCATCCCAGAGAAGAAGG-3' + 5'-GGCAGAGAGCAAGGCCTCTA-3'

*LAMP1*: 5'-TGACCTTTGACCTGCCATCA-3' + 5'-AACCGCAGCTCCTGTGGAAA-3'

HAP1 cells were transfected with TALEN encoding plasmids and a pMX vector expressing GFP-IRES-Blasticidin, briefly selected with blasticidin (from 24 hours after transfection until 48 hours after transfection) and subcloned. For some genes subclones were selected that did not stain for on glycosylated  $\alpha$ -DG using the IIH6 antibody. Subclones were genotyped by PCR and mutations of targeted alleles were sequence verified.

### **RT-PCR**

cDNA was generated from total RNA of parental HAP1 cells and TALEN-targeted mutants using Superscript III First Strand Synthesis System (Invitrogen) and probed for presence of wild-type transcripts using the following primer sequences: *C3orf39*: 5'-GGGCTGGGGCGAGGGTGCACA-3' +

5'-GCTAAAGACCAGAATGTACTC-3' *SGK196*: 5'-GAAATTGCAGAGGCCGTCAAC-3' + 5'-  
ATCCTGAAGTGACCATAGGGA-3' *SLC35A1*: 5'-GTGACCTACCAGTTGAAGATT-3' + 5'-  
AACTCCAGCACACAGCATAAA-3' *TMEM5*: 5'-AAGATGCCTTGCTTCAGAGTG-3' + 5'-  
CTTGAAGTGCTGATACCACTG-3' *B3GALNT2*: 5'-GAAATCTACTCACTATGATGT-3' + 5'-  
TGACAACTCGATCCTCAGGCAG-3' *B3GNT1*: 5'-GCAGGAGGAGCAAGACCAATA-3' + 5'-  
CCGTGGCCAGCTGCGCCTCCT-3' *ST3GAL4*: 5'-AAGAAGGAGCCGTGCCTCCAG-3' + 5'-  
CTCTGGATGTTCTTGGGGATG-3'

### **cDNA complementation of TALEN-mutated genes**

cDNAs of TALEN-mutated genes were amplified using the following primer sequences:

*C3orf39*: 5'-GATCGAATTCACCATGCACCTCTCGGCGGTGTTTC-3' + 5'-  
GATCCTCGAGTTACTTGTTCATCGTCGTCCTTGTAGTCCGTGTTGCACACCAGCAC-3' *SGK196*:  
5'-GATCGAATTCACCATGGAAAAGCAGCCCCAGAAC-3' + 5'-  
GATCCTCGAGTTACTTGTTCATCGTCGTCCTTGTAGTCCAGCATCTCTTGCCTGAGA-3'  
*SLC35A1*: 5'-GATCGAATTCACCATGGCTGCCCGAGAGACAAT-3' + 5'-  
GATCCTCGAGTTACTTGTTCATCGTCGTCCTTGTAGTCCACACCAATAACTCTCTCCTT-3'  
*TMEM5*: 5'-GATCGAATTCACCATGCGGCTGACGCGGAAGC-3' + 5'-  
GATCCTCGAGTTACTTGTTCATCGTCGTCCTTGTAGTCACTTTTATTATTATTCATTAATAAATGAGC  
TTTC-3' *B3GALNT2*: 5'-GATCGAATTCACCATGCGAACTGGCTGGTGCTG-3' + 5'  
GATCCTCGAGTTACTTGTTCATCGTCGTCCTTGTAGTCTCTTGCTTGACATCGACAAGGATCAC  
-3' *B3GNT1*: 5'-GATCGAATTCACCATGCAGATGTCCTACGCCATCC-3' + 5'-  
GATCCTCGAGTTACTTGTTCATCGTCGTCCTTGTAGTCGCAGCGTCGGGGAGAGTTGGG-3'  
*ST3GAL4*: 5'-GATCGAATTCACCATGGTCAGCAAGTCCCGCTG-3' + 5'  
GATCCTCGAGTTACTTGTTCATCGTCGTCCTTGTAGTCGAAGGACGTGAGGTTCTTGATAGCT-  
3'

For *B3GALNT2*, the internal EcoRI restriction site was eliminated by introduction of a silent mutation (789A>G) in the coding sequence using overlapping internal primers: 5'-

GGGAGGGTGCATTGCCTCATGAGTTCTTGGAAAGGTGTGGAGGGAG-3' and 5'-  
CTCCCTCCACACCTTCCAAGAACTCATGAGGCAATGCACCCTCCC-3' PCR products were  
cloned into retroviral expression vector pBabe-puro, using EcoRI and SalI restriction sites and virus was  
generated in 293T cells by transfection of the obtained plasmids combined with pAdvantage (Clontech,  
Mountain View, California, USA), CMV-VSV-G and Gag-pol. TALEN-targeted HAP1 cells were  
infected with respective viruses or empty control virus and selected with puromycin.

## Haploid genetic screens

Gene-trap virus was produced in 293T cells by transfection of a gene-trap vector (pGTen2-ACTB) combined with pAdvantage, CMV-VSVG and Gag-pol. For this purpose, the poly-adenylation signal of the pGT-GFP gene-trap vector described previously (10) was replaced with the human  $\beta$ -actin (*ACTB*) poly-adenylation sequence, ACT1 (29), and the murine engrailed 2 (*En2*) splice acceptor (30) was added. The virus was concentrated using ultracentrifugation for 2 h at 21,000 r.p.m. in a Beckman (Indianapolis, Indiana, USA) rotor.  $10^8$  HAP1 cells were seeded for mutagenesis. A proportion of the cells (ca.  $3 \times 10^7$ ) was harvested for genomic DNA isolation to create a control data set. For the screen with rVSV-GP-LASV,  $10^8$  mutagenized cells were exposed to ca. 40,000 PFU/ml of the virus in the presence of 5mM NH<sub>4</sub>Cl. After selection, resistant colonies were expanded to ca.  $3 \times 10^7$  cells and used for genomic DNA isolation and insertion site retrieval. For the screens using magnetic immune-depletion,  $10^8$  mutagenized cells were stained with an antibody against the laminin-binding epitope on  $\alpha$ -DG (IIH6) or heparan sulfate (10E4) and separated on Miltenyi MACS LD Columns (Miltenyi Biotec, Bergisch Gladbach, Germany) according to manufacturer's specifications.

## Sequence analysis of gene-trap insertion sites

Gene-trap insertion sites were recovered by linear amplification of genomic DNA sequences flanking the proviral DNA insertions and mapped to the human genome by deep sequencing. Genomic DNA isolated from selected populations was digested using NlaIII and MseI restriction enzymes. For the generation of the control data set, genomic DNA was additionally digested with MluCI and HaeIII for maximum genome coverage. As described previously (11), digested fragments were subjected to linear amplification, followed by ligation to a 5'-phosphorylated and 3'-modified (dideoxycytidine, ddC) ssDNA linker (5'- ATCGTATGCCGTCTTCTGCTTGA CT CAGTAGTTGTGCGATGGATTGATG-3'). Ligation products were PCR amplified and sequenced on a HiSeq2000 sequencer (Illumina, San Diego, California, USA).

Sequencing reads were cropped to a length of 36bp and aligned to the human genome (hg18) using Bowtie, allowing no mismatches and filtered for repetitive regions as described previously (12). For selected populations, if two reads on the same strand were closer than 3bp from each other, one was removed in order to filter out possible sequencing or PCR artifacts. This criterion was not used for the control dataset in order to prevent underestimating retroviral insertion bias. Insertion sites were intersected with known RefSeq gene locations and filtered for gene inactivating events (sense orientation in intron or present in exon). Per gene, enrichment for inactivating insertions over the control data set (table S1) was assessed using a one-sided Fisher's exact test. All genes with an assigned *P*-value lower than  $10^{-5}$  were considered enriched (tables S2 and S3). For the screen with an antibody against heparan

sulfate (Fig. 2B), a more stringent cut-off was chosen ( $P \leq 10^{-21}$ ) to account for higher complexity of the sorted cell population and thus more measurements per gene (table S4). To further account for retroviral insertion bias, enriched genes were required to display a significant preference ( $P \leq 0.05$ ) for sense orientation insertions versus antisense orientation insertions using a binomial test (see tables S2, S3, and S4). As gene-trap insertions are disruptive in exons regardless of orientation, this test could only be applied to genes for which the majority of insertions ( $\geq 70\%$ ) fell in intronic regions. Genes carrying less than 70% of all insertions in introns, such as *B3GNT1* (see fig. S2), were analyzed for enrichment of disruptive insertions over the control data set as described above, but with stricter cut-off ( $P \leq 10^{-30}$ ) (see tables S2, S3, and S4). All  $P$ -values were false discovery rate (FDR)-corrected. In cases where the  $P$ -value was lower than what R software could report, the numerical value was set to the smallest nonzero normalized floating-point number R could report on the computer used for analysis (ca.  $10^{-314}$ ).

### **Supplementary Clinical Information**

Family 40: The first affected child of non-consanguineous parents was born after a pregnancy complicated with polyhydramnios. At 37 weeks a girl was born by cesarean section because of prolonged second stage labor due to an unexpectedly large head. The birth weight was 3820 grams with a head circumference of 48 cm (+ 9 standard deviations). Directly after birth a cerebral CT-and MRI scan showed a triventricular hydrocephalus with a very thin cortex, compression of the cerebellar hemispheres and brainstem, and an Arnold Chiari malformation. A week after birth a ventriculo-peritoneal drain was placed. Ophthalmological evaluation showed persistent hyperplastic primary vitreous (PHPV) in both eyes, high myopia of the right eye and left microphthalmia. The child presented with severe hypotonia and muscle weakness, and serum creatine kinase levels were elevated with concentration of more than 7000 U/L. The parents were informed of the clinical diagnosis of Walker Warburg syndrome (WWS). A CT scan at the age of 17 months showed dilated cerebral ventricles with agyria, and hypoplasia of the cerebellum and brainstem. Aqueductal stenosis was suspected and the right eye lens showed cataract. A muscle biopsy performed at the age of 18 months demonstrated muscular dystrophy, and immunohistochemistry showed presence of merosin, dystrophin and  $\alpha$ -,  $\beta$ - and  $\gamma$ -sarcoglycans. The girl died at the age of almost 3 years.

In the second pregnancy advanced ultrasound examination was performed at 17 weeks because of a previous child with WWS. The fetus showed cerebral ventriculomegaly and a small occipital encephalocele. The pregnancy was terminated at 19 weeks. Pathology confirmed the cerebral anomalies and additionally absence of the falx cerebri and cerebellar tentorium were observed. Both eyes showed

cortical cataract and retinal coloboma, supportive of the diagnosis WWS. The third pregnancy resulted in a healthy male child.

Family 43: The affected boy of this family was born to parents who were second cousins. All clinical features were supportive of the diagnosis WWS. The mother had three pregnancies, including one stillborn child of which no further clinical details are available. An older brother of the proband is healthy. At birth, the affected newborn had a length of 52 cm, weight of 4.280 g, and an increased head circumference of 41.5 cm (+5 standard deviations). The Apgar (**A**ppearance, **P**ulse, **G**rimace, **A**ctivity, **R**espiration) score was 6/8/8. The newborn male showed an occipital encephalocele, right microphthalmia and bilateral opaque cornea. His muscles were markedly hypotonic. Creatine kinase was grossly elevated with concentration of about 3500 U/L at several occasions. Cranial sonography demonstrated a severe hydrocephalus internus and occipital encephalocele. CT or MRI has not been done. Evaluation by an ophthalmologist confirmed the above-mentioned findings and revealed bilateral cataracts in addition. EEG was clearly abnormal with initially slow waves, later a so-called pseudo alpha activity was noted, as seen in patients with lissencephaly. At the age of 6 weeks, the hydrocephalus was shunted to drain the increasing amount of fluid in the rapidly growing head. At the age of 7 months, a muscle biopsy was performed. Histology showed a mixture of dystrophy and myopathy including areas of necrosis and increased connective tissue. Histochemistry demonstrated presence of merosin but lack of  $\alpha$ -dystroglycan. The severely disabled boy passed away at the age of 22 months.

Family 56: The parents of this pedigree are first cousins and had two affected girls. Both girls have severe intellectual disability with autistic features. Small head circumference was evident. They were barely mobile with a broad-based gait denoting ataxia. Both had significantly elevated serum creatine kinase levels of >4000 U/L (n<191). Muscle biopsy was typical for congenital muscular dystrophy, but the biopsy is no longer available for further investigation. A cranial MRI was done from the older sibling at one year of age and showed brainstem atrophy, dilated ventricles, widespread pachygyria and significant white matter involvement with leukodystrophic appearance (Fig. 4B). When the two female siblings were 1 and 3 years of age, respectively, a preliminary diagnosis of WWS was made. Both siblings are alive and currently aged 19 and 21. Both are ambulant, but ataxic. They prefer to walk with assistance and for short durations. Both are still severely intellectually disabled, speaking only a few basic words. Given this longer survival time and the phenotypic presentation, classification as muscle-eye-brain disease appears more appropriate.

## Legends to Supplementary Figures

**Figure S1:** Walker-Warburg patient fibroblasts are refractory to virus entry mediated by the Lassa virus (LASV) glycoprotein. **(A)** Schematic representation of the replication competent vesicular stomatitis virus that expresses the LASV glycoprotein (LASV-GP) instead of VSV-G and an eGFP marker gene. **(B)** Control fibroblasts of healthy individuals and fibroblasts derived from a Walker-Warburg patient carrying a mutation in *ISPD* were infected with rVSV-GP-LASV. Infected cells show viral eGFP expression (middle) and nuclei were stained with Hoechst33342 (right).

**Figure S2:** rVSV-GP-LASV but not rVSV-GP-EBOV infect HAP1 cells in an alpha-dystroglycan ( $\alpha$ -DG)-dependent manner. **(A)** TALEN-induced *DAG1* mutation in HAP1 cells leads to a frameshift and early stop codon in the open reading frame. Polypeptide sequence is indicated below. TALENs are directed to target all protein coding transcripts. Processed *DAG1* transcripts are shown (not drawn to scale). Open reading frames of protein coding transcripts are depicted in red. TALEN-targeted sites are marked by an arrow and shaded in green. **(B)** Wild-type HAP1 cells and HAP1 cells carrying an inactivating mutation in *DAG1* were infected with rVSV-GP-LASV and rVSV-GP-EBOV. Infected cells show viral eGFP expression.

**Figure S3:** Examples of gene-trap insertions in genes containing large introns and genes with few intronic sequences. Gene-trap insertions are mapped to the gene bodies of *TMEM5* and *B3GNT1*. Integrations in the same transcriptional orientation as the gene (sense) are depicted in green, while those in the opposite orientation (antisense) are drawn in red. In the more common gene structure of an intron-rich gene like *TMEM5*, intronic insertions have a strong bias towards sense orientation (disruptive, shown in green), and most of the identified antisense insertions map onto exonic regions and will affect the gene by direct disruption of the coding sequence rather than through redirection of splicing. In intron-poor genes like *B3GNT1*, the majority of the gene-trap insertions fall in exonic regions where they are predicted to disrupt gene function in either orientation and therefore such genes show no strong bias towards gene-trap insertions in the sense orientation.

**Figure S4:** Intersection of haploid screens. Comparison of haploid genetic screens highlights genes important for general glycosylation events, specific glycosylation of  $\alpha$ -DG, entry of the LASV glycoprotein and genes for heparan sulfate biosynthesis. LASV: haploid screen using rVSV-GP-LASV;  $\alpha$ -DG: haploid screen using an antibody specific for glycosylated  $\alpha$ -DG (IIH6); HS: haploid screen using an antibody specific for heparan sulfate (10E4). *COG4*: component of oligomeric golgi complex 4;

*COG5*: component of oligomeric golgi complex 5; *COG7*: component of oligomeric golgi complex 7; *COG8*: component of oligomeric golgi complex 8; *PTARI*: protein prenyltransferase alpha subunit repeat containing 1; *UGDH*: UDP-glucose 6-dehydrogenase; *UXSI*: UDP-glucuronate decarboxylase 1; *MPDU1*: mannose-P-dolichol utilization defect 1; *PMM2*: phosphomannomutase 2; *DPM1*: dolichyl-phosphate mannosyltransferase polypeptide 1; *DAG1*: dystroglycan 1; *LARGE*: like-glycosyltransferase; *ISPD*: isoprenoid synthase domain containing; *FKTN*: fukutin; *FKRP*: fukutin related protein; *DPM3*: dolichyl-phosphate mannosyltransferase polypeptide 3; *POMGNT1*: protein O-linked mannose beta1,2-N-acetylglucosaminyltransferase; *POMT1*: protein-O-mannosyltransferase 1; *POMT2*: protein-O-mannosyltransferase 2; *TMEM5*: transmembrane protein 5; *B3GALNT2*: beta-1,3-N-acetylgalactosaminyltransferase 2; *B3GNT1*: UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 1; *SGK196*: protein kinase-like protein SgK196; *SLC35A1*: solute carrier family 35 member A1; *XYLT2*: xylosyltransferase II; *B3GALT6*: UDP-Gal:betaGal beta 1,3-galactosyltransferase, polypeptide 6; *B3GAT3*: beta-1,3-glucuronyltransferase 3; *FAM20B*: family with sequence similarity 20 member B; *EXT1*: exostosin 1; *EXT2*: exostosin 2; *EXTL3*: exostosin-like 3; *NDST1*: N-deacetylase/N-sulfotransferase 1; *SLC35B2*: solute carrier family 35 member B2; *TM9SF2*: transmembrane 9 superfamily member 2; *LAMP1*: lysosomal-associated membrane protein 1; *FAM3C*: family with sequence similarity 3 member C; *DOT1L*: DOT1-like; *ST3GAL4*: ST3 beta-galactoside alpha-2,3-sialyltransferase 4; *CMAS*: cytidine monophosphate N-acetylneuraminic acid synthetase; *SLC35A2*: solute carrier family 35 member A2; *GNE*: glucosamine (UDP-N-acetyl)-2-epimerase/N-acetylmannosamine kinase; *MGAT1*: mannosyl (alpha-1,3-)-glycoprotein beta-1,2-N-acetylglucosaminyltransferase; *MAN1B1*: mannosidase, alpha, class 1B, member 1; *ALG5*: asparagine-linked glycosylation 5; *ALG6*: asparagine-linked glycosylation 6; *ALG8*: asparagine-linked glycosylation 8; *TMEM165*: transmembrane protein 165.

**Figure S5:** Cell surface profiling of mutagenized HAP1 cells using magnetic cell separation. **(A)** Mutagenized HAP1 cells were labeled with antibodies directed against the laminin-binding epitope of glycosylated  $\alpha$ -DG (IIH6) and immunoreactive cells were depleted using magnetic cell sorting. The cell population enriched for IIH6-negative cells was expanded and gene-trap insertion sites were mapped using parallel sequencing. **(B)** Cells magnetically depleted for the laminin-binding epitope of glycosylated  $\alpha$ -DG were stained using the IIH6 antibody and analyzed by flow-cytometry (mutagenized cells express GFP). **(C)** Immunoblot analysis of the levels of glycosylated  $\alpha$ -DG in the immunodepleted cell pool using the IIH6 antibody.

**Figure S6:** Mutagenized HAP1 cells were depleted for heparan sulfate-positive cells, immunostained using an antibody specific for heparan sulfate (10E4) and subjected to magnetic depletion. The parental population and the cell population after immunodepletion were cultured, harvested, stained with the heparan sulfate specific antibody and analyzed using flow-cytometry. Mutagenized cells express GFP.

**Figure S7:** TALENs are directed to target all protein coding transcripts. Processed transcripts of TALEN-targeted genes are shown (not drawn to scale). Open reading frames of protein coding transcripts are depicted in red, those of transcripts affected by nonsense-mediated decay are shown in blue. TALEN-targeted sites are marked by an arrow and shaded in green.

**Figure S8:** Wild-type transcripts were undetectable in HAP1 cell lines containing the TALEN-induced mutations. RT-PCR on mutant HAP1 cell lines shows no detectable wild-type transcripts using primers that were designed specifically for the detection of wild-type transcripts. These primers could not amplify transcripts expressed by the TALEN-mutated alleles.

**Figure S9:** Analysis of  $\alpha$ -DG glycosylation in TALEN-targeted HAP1 cells. **(A)** Individual HAP1 clones carrying the described mutations in the corresponding genes were stained with an antibody recognizing the laminin-binding epitope of  $\alpha$ -DG and analyzed using flow-cytometry. **(B)** TALENs were used to generate a mutation in *LAMP1* and identified knockout cells were subjected to immunoblot analysis using LAMP1 antibodies. TALEN-induced *LAMP1* mutation in HAP1 cells leads to a frameshift and early stop codon in the open reading frame. Polypeptide sequence is indicated below. **(C)** LAMP1 deficient cells were stained with an antibody recognizing the laminin-binding epitope of  $\alpha$ -DG and analyzed using flow-cytometry.

**Figure S10:**  $\alpha$ -DG glycosylation, laminin binding and virus infection in cells deficient for the identified host factors. **(A)** Glycosylation of  $\alpha$ -DG in mutant cell lines (upper panel) and binding to laminin (lower panel). Cell lysates of the corresponding mutant HAP1 cell lines were subjected to immunoblot analysis using antibodies directed against the laminin-binding epitope of glycosylated  $\alpha$ -DG (IIH6). Glycoproteins isolated from the mutant HAP1 cells were subjected to a laminin overlay assay. **(B)** Cell lysates of the corresponding HAP1 mutant cell lines were subjected to immunoblot analysis using an antibody recognizing both  $\alpha$ -DG and  $\beta$ -DG. **(C)** HAP1 cell lines with independent TALEN-induced mutations in the corresponding genes and wild-type control cells were infected with rVSV-GP-Lassa as in Fig 3B. Infected cells express viral eGFP.

**Figure S11:** Correction of the TALEN-induced phenotype by re-introduction of the corresponding cDNAs. **(A)** HAP1 cells carrying TALEN-induced mutations in the indicated genes were transduced with retroviruses expressing cDNA of the corresponding genes or with an empty control virus and exposed to rVSV-GP-LASV. rVSV-GP-LASV infected cells show viral eGFP expression. **(B)** Re-expression of *C3orf39*, *SGK196* and *SLC35A1* cDNAs in TALEN-induced mutant HAP1 cell lines restores the  $\alpha$ -DG laminin-binding epitope on the cell surface. HAP1 cells carrying a TALEN-induced mutation in *C3orf39* and 2 independent HAP1 cell lines carrying mutations in *SGK196* or *SLC35A1* were infected with empty vector (upper panel) or with a retrovirus expressing cDNA of the corresponding genes (lower panel), stained using the IIIH6 antibody and analyzed by flow-cytometry. Wild-type HAP1 cells were analyzed as control (upper left).

**Figure S12:** Exostosin family domain present in *TMEM5*. Sequence alignment of the Exostosin domain consensus sequence with *TMEM5*. The amino acid sequence of *TMEM5* was subjected to a PFAM domain search. The exostosin family domain was identified ( $E = 0.0005$ ) as the only hit.

**Figure S13:** Mutations in *TMEM5* and *SGK196* observed in patients do not support virus infection. Cells deficient for *TMEM5* or *SGK196* were complemented with the corresponding wild-type cDNAs or with cDNAs containing the mutations observed in patients. Cells were exposed to rVSV-GP-LASV. Infected cells display eGFP expression.

## Supplementary Tables

Supplementary tables can be found in the supplementary online material accompanying this manuscript.

**Table S1:** List of disruptive gene-trap insertions mapped in unselected control population of HAP1 cells. Chromosome, genomic coordinates and affected gene are indicated.

**Table S2:** Table listing, per gene, insertions predicted to be inactivating mapped in HAP1 cells that were selected with rVSV-GP-LASV; these insertions were compared to insertions predicted to be inactivating mapped in unselected control HAP1 cells and significance of enrichment was calculated using a one-sided Fisher's exact test and FDR correction of  $P$ -values. The percentage of intronic insertion events (as opposed to exonic) is also indicated. Additionally, per gene, the number of sense insertions (i.e. in the same orientation as transcription of the affected gene) is compared to the number of antisense insertions using a binomial test and FDR-correction of  $P$ -values. Genes carrying the majority gene-trap insertions in

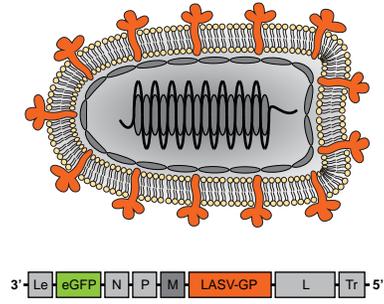
introns were colored (orange) if they passed 2 statistical tests: enrichment of disruptive mutations compared to control cells (one-sided Fisher's exact test,  $P \leq 10^{-5}$ , first tab) and bias for gene-trap insertion events in the transcriptional orientation of the affected gene (binomial test,  $P \leq 0.05$ , second tab). Intron-poor genes were colored (purple) if they passed the former criterion using a stricter cut-off (one-sided Fisher's exact test  $P \leq 10^{-30}$ , third tab).

**Table S3:** Table listing, per gene, insertions predicted to be inactivating mapped in HAP1 cells that were immunodepleted for the laminin-binding epitope of alpha-dystroglycan; these insertions were compared to insertions predicted to be inactivating mapped in unselected control HAP1 cells and significance of enrichment was calculated using a one-sided Fisher's exact test and FDR correction of  $P$ -values. The percentage of intronic insertion events (as opposed to exonic) is also indicated. Additionally, per gene, the number of sense insertions (i.e. in the same orientation as transcription of the affected gene) is compared to the number of antisense insertions using a binomial test and FDR-correction of  $P$ -values. Genes carrying the majority gene-trap insertions in introns were colored (orange) if they passed 2 statistical tests: enrichment of disruptive mutations compared to control cells (one-sided Fisher's exact test,  $P \leq 10^{-5}$ , first tab) and bias for gene-trap insertion events in the transcriptional orientation of the affected gene (binomial test,  $P \leq 0.05$ , second tab). Intron-poor genes were colored (purple) if they passed the former criterion using a stricter cut-off (one-sided Fisher's exact test  $P \leq 10^{-30}$ , third tab).

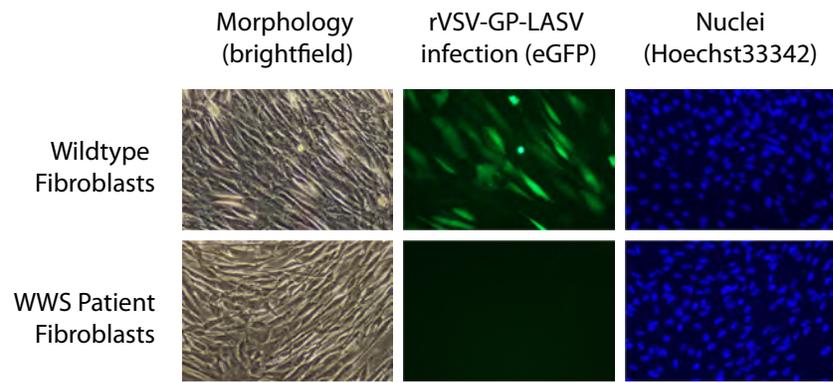
**Table S4:** Table listing, per gene, insertions predicted to be inactivating mapped in HAP1 cells that were immunodepleted for heparan sulfate; these insertions were compared to insertions predicted to be inactivating mapped in unselected control HAP1 cells and significance of enrichment was calculated using a one-sided Fisher's exact test and FDR correction of  $P$ -values. The percentage of intronic insertion events (as opposed to exonic) is also indicated. Additionally, per gene, the number of sense insertions (i.e. in the same orientation as transcription of the affected gene) is compared to the number of antisense insertions using a binomial test and FDR-correction of  $P$ -values. Genes carrying the majority gene-trap insertions in introns were colored (orange) if they passed 2 statistical tests: enrichment of disruptive mutations compared to control cells (one-sided Fisher's exact test,  $P \leq 10^{-21}$ , first tab) and bias for gene-trap insertion events in the transcriptional orientation of the affected gene (binomial test,  $P \leq 0.05$ , second tab). Intron-poor genes were colored (purple) if they passed the former criterion using a stricter cut-off (one-sided Fisher's exact test  $P \leq 10^{-30}$ , third tab).

# Fig S1

**A**

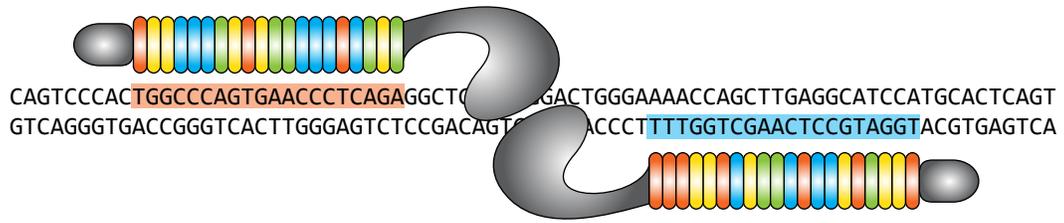


**B**



# Fig S2

A

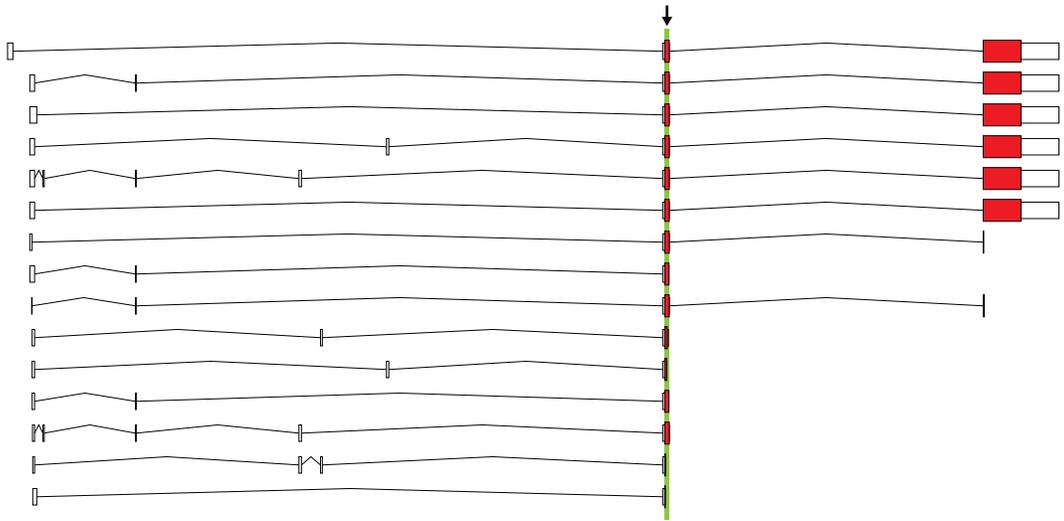


**HAP1 WT**

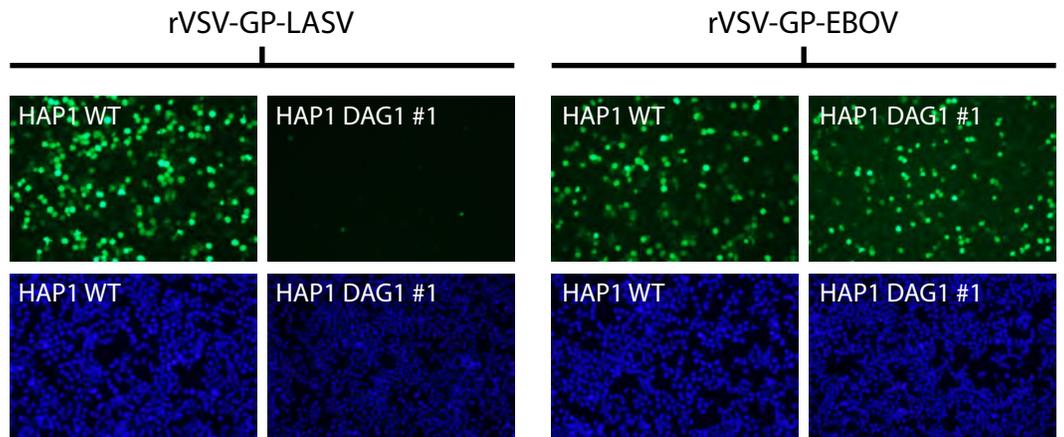
CAGTCCCAC TGGCC CAGTGAACCCTCAGAGGCTGTCAGGGACTGGGAAAACCAGCTTGAGGCATCCATGCACTCAGT  
 -Q--S--H--W--P--S--E--P--S--E--A--V--R--D--W--E--N--Q--L--E--A--S--M--H--S--V

**HAP1 DAG1 #1**

CAGTCCCAC TGGCC CAGTGAACCCTCAG ::::::::::::::: GACTGGGAAAACCAGCTTGAGGCATCCATGCACTCAGT  
 -Q--S--H--W--P--S--E--P--S-----G--L--G--K--P--A--\*-



B

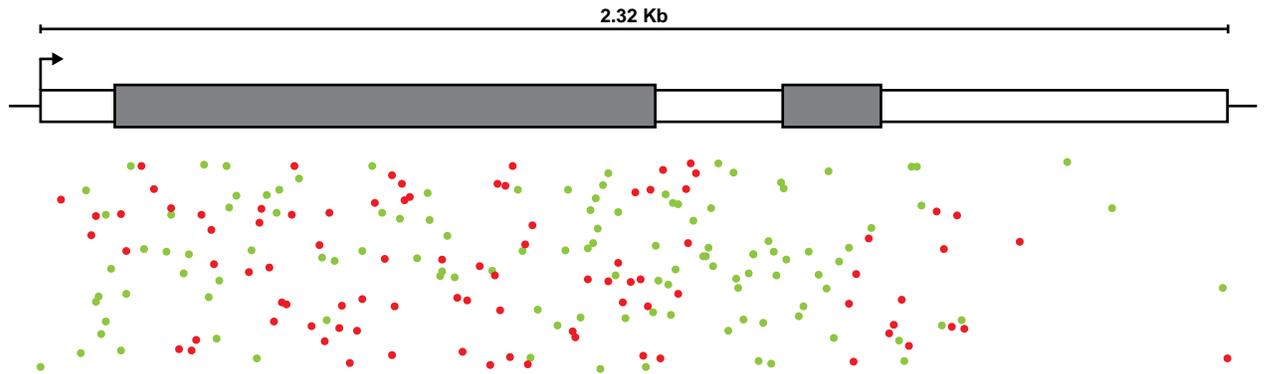


**Fig S3**

***TMEM5***



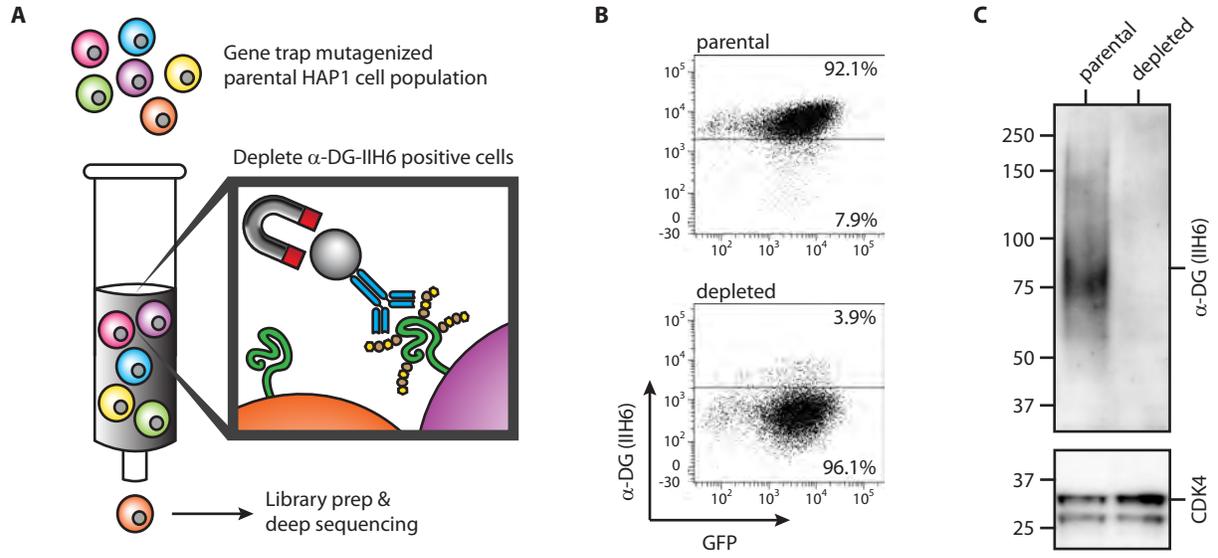
***B3GNT1***



**Fig S4**

Gene(s)	Description	LASV	$\alpha$ -DG	HS
<i>COG4, COG5, COG7, COG8</i>	Conserved Oligomeric Golgi complex; mutations in congenital disorders of glycosylation type II	●	●	●
<i>PTAR1</i>	Uncharacterized protein containing prenyltransferase alpha subunit repeat	●	●	●
<i>UGDH, UXS1</i>	Generation of UDP-glucuronate and UDP-xylose	●	●	●
<i>MPDU1, PMM2, DPM1</i>	Mannose supply; mutated in congenital disorder of glycosylation type I	●	●	
<i>DAG1, LARGE, ISPD, FKTN, FKRP, DPM3, POMGNT1, POMT1, POMT2, C3ORF39</i>	<b>Alpha-dystroglycan (<math>\alpha</math>-DG) and known Walker-Warburg genes</b>	●	●	
<i>TMEM5</i>	<b>Uncharacterized protein with exostosin family domain; mutated in cobblestone lissencephaly; potential role in <math>\alpha</math>-DG glycosylation</b>	●	●	
<i>B3GALNT2</i>	<b>Uncharacterized; potential role in <math>\alpha</math>-DG glycosylation</b>	●	●	
<i>B3GNT1</i>	<b>Uncharacterized; potential role in <math>\alpha</math>-DG glycosylation (reported as WWS gene during preparation of this manuscript (15))</b>	●	●	
<i>SGK196</i>	<b>Uncharacterized kinase; potential role in <math>\alpha</math>-DG glycosylation</b>	●	●	
<i>SLC35A1</i>	<b>CMP-sialic acid transporter; mutated in congenital disorder of glycosylation type II; potential role in <math>\alpha</math>-DG glycosylation</b>	●	●	
<i>XYLT2, B3GALT6, B3GAT3, FAM20B, EXT1, EXT2, EXTL3, NDST1, SLC35B2</i>	Heparan sulfate (HS) biogenesis			●
<i>TM9SF2</i>	Uncharacterized protein, transmembrane 9 superfamily member 2			●
<i>LAMP1</i>	Lysosome-associated membrane protein 1	●		
<i>KIAA1432</i>	Connexin interacting protein	●		
<i>FAM3C</i>	Uncharacterized protein, family with sequence similarity 3, member C	●		
<i>DOT1L</i>	Histone methyltransferase	●		
<i>ST3GAL4</i>	ST3 beta-galactoside alpha-2,3-sialyltransferase 4	●		
<i>CMAS, SLC35A2, GNE</i>	Required for incorporation of sialic acid in glycans	●		
<i>MGAT1</i>	Mannosyl ( $\alpha$ -1,3-)-glycoprotein $\beta$ -1,2-N-acetylglucosaminyltransferase	●		
<i>MAN1B1</i>	Mannosidase alpha, class 1B; mutations lead to intellectual disability	●		
<i>ALG5, ALG6, ALG8</i>	Involved in N-glycosylation; mutations lead to congenital disorder of glycosylation type I	●		
<i>TMEM165</i>	Deficiency causes a congenital disorder of glycosylation		●	●

**Fig S5**



**Fig S6**

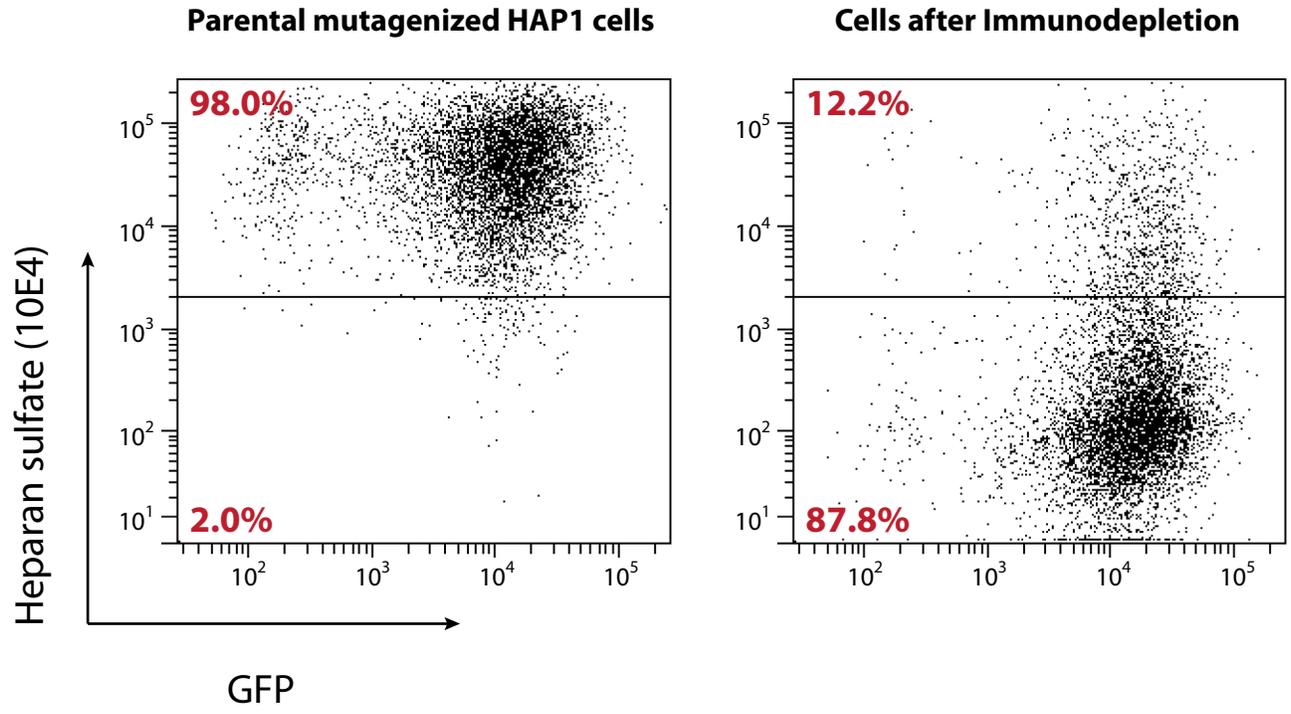
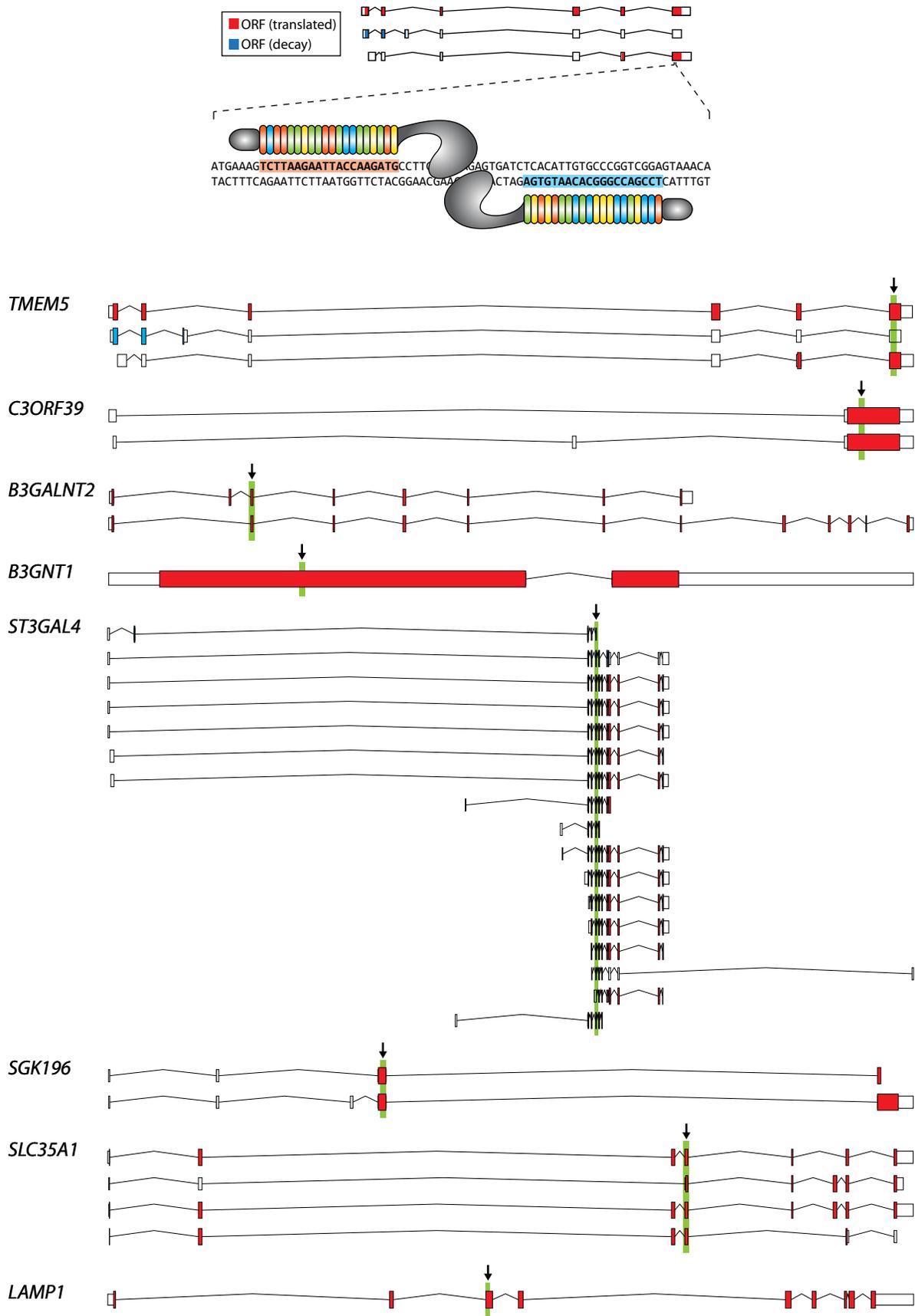
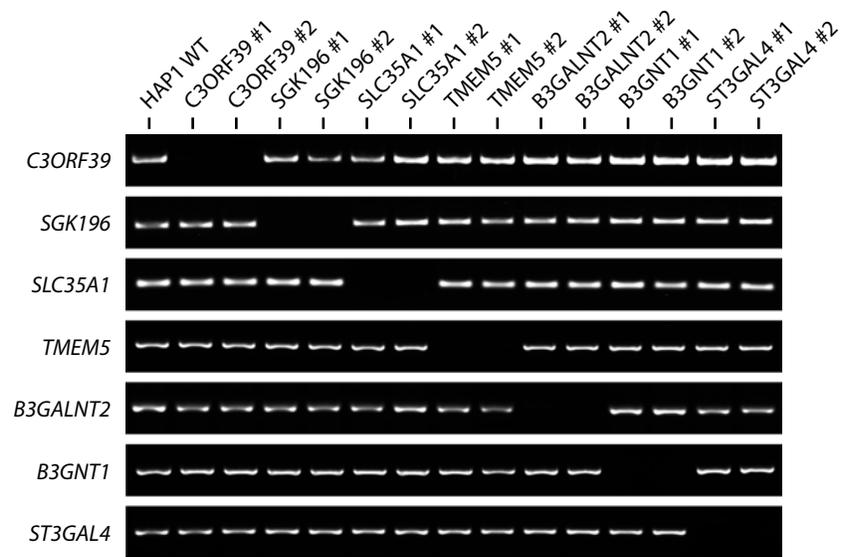


Fig S7

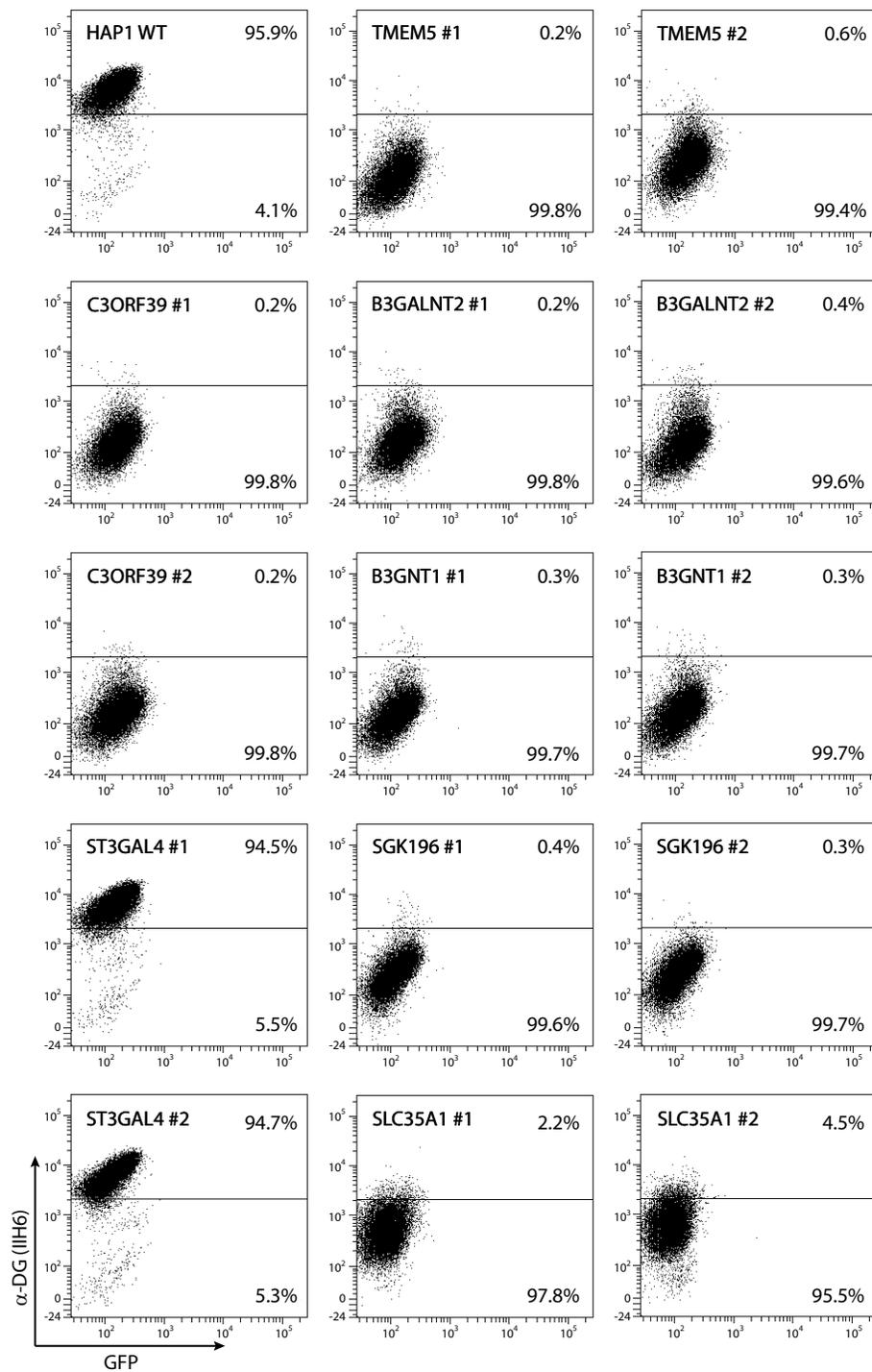


**Fig S8**

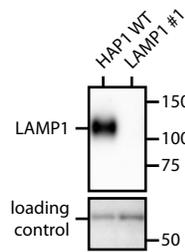


**Fig S9**

**A**



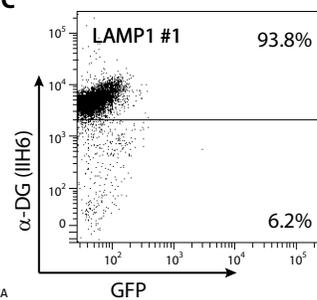
**B**



**HAP1 WT**  
 AACATGACC TTTGACCTGCCATCAGATGCCACAGTGGTCTCAACCGCAGCTCTGTGGAAAAGAGAACACTTCTGA  
 -N-M-T-F-D-L-P-S-D-A-T-V-V-L-N-R-S-S-C-G-K-E-N-T-S-D

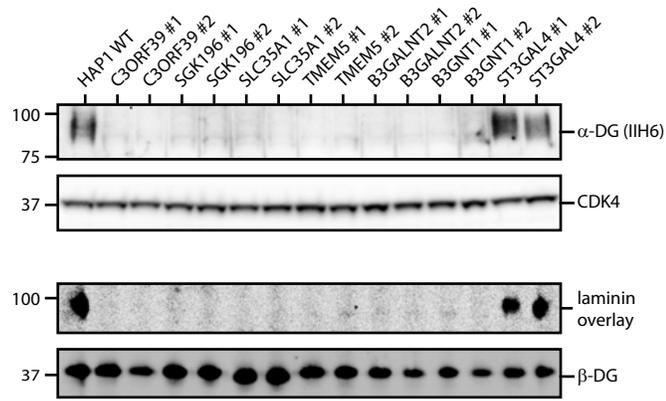
**HAP1 LAMP1 #1**  
 AACATGACC TTTGACCTGCCATCAGATGCCACAG::GTGCTCAACCGCAGCTCTGTGGAAAAGAGAACACTTCTGA  
 -N-M-T-F-D-L-P-S-D-A-T---G-A-Q-P-Q-L-L-W-K-R-E-H-F--

**C**

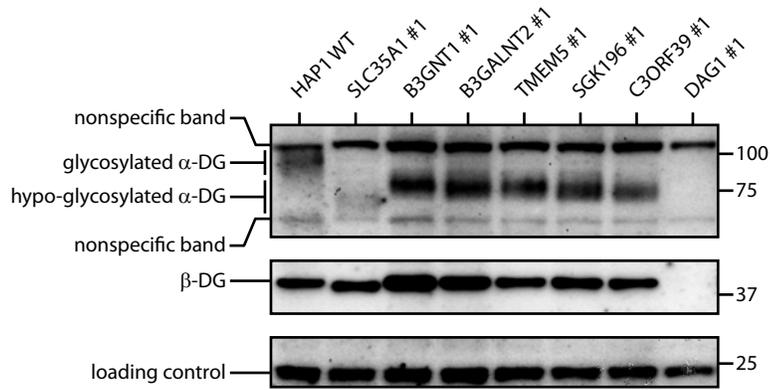


**Fig S10**

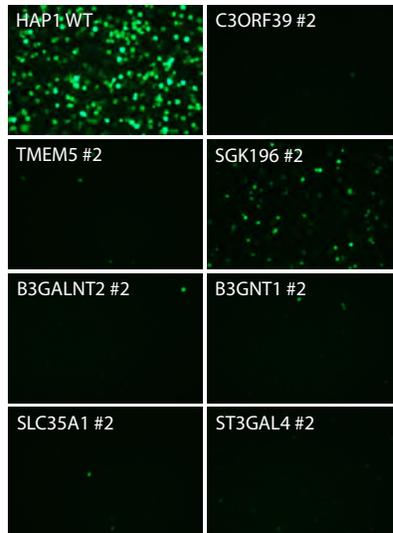
**A**



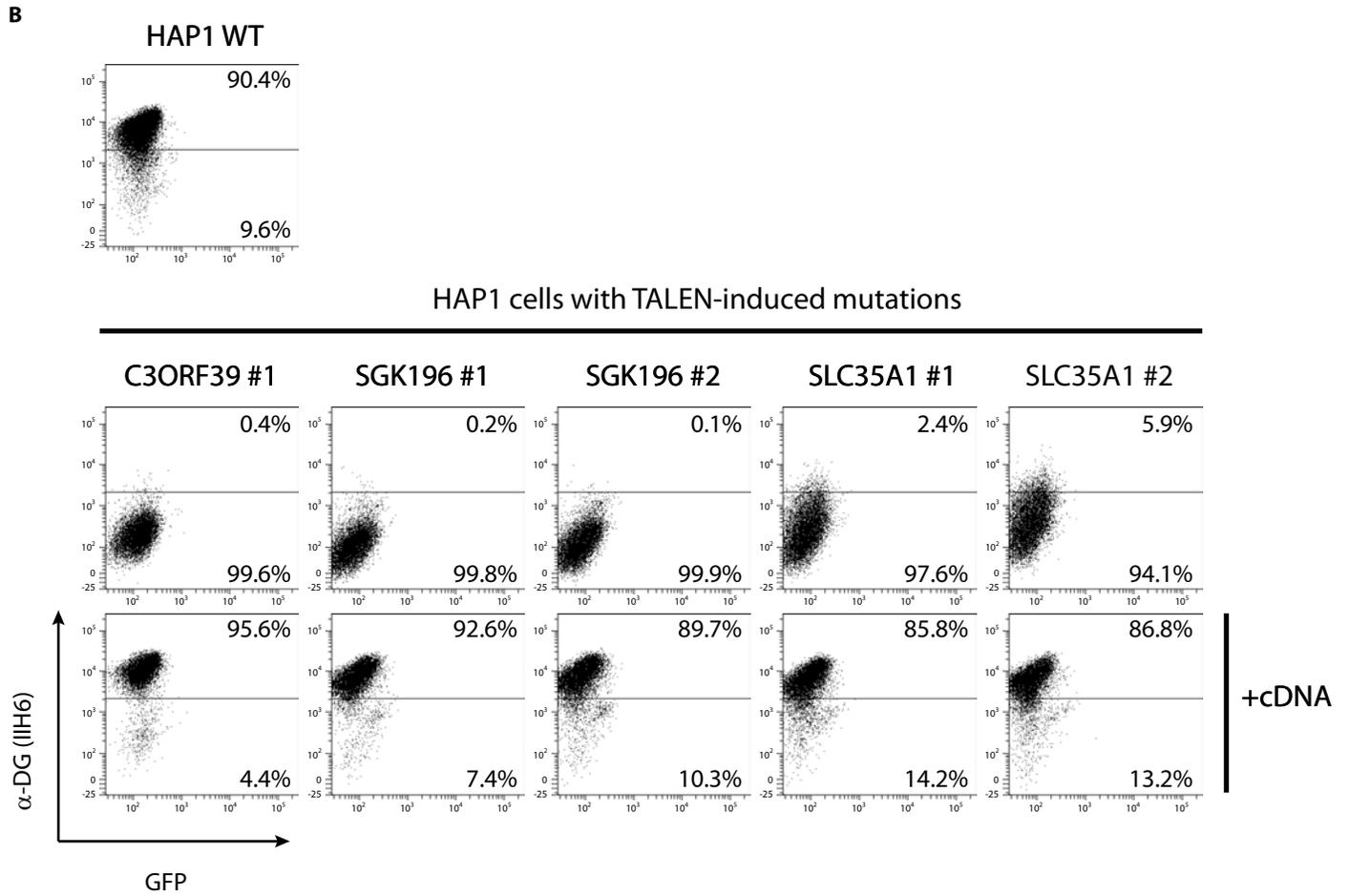
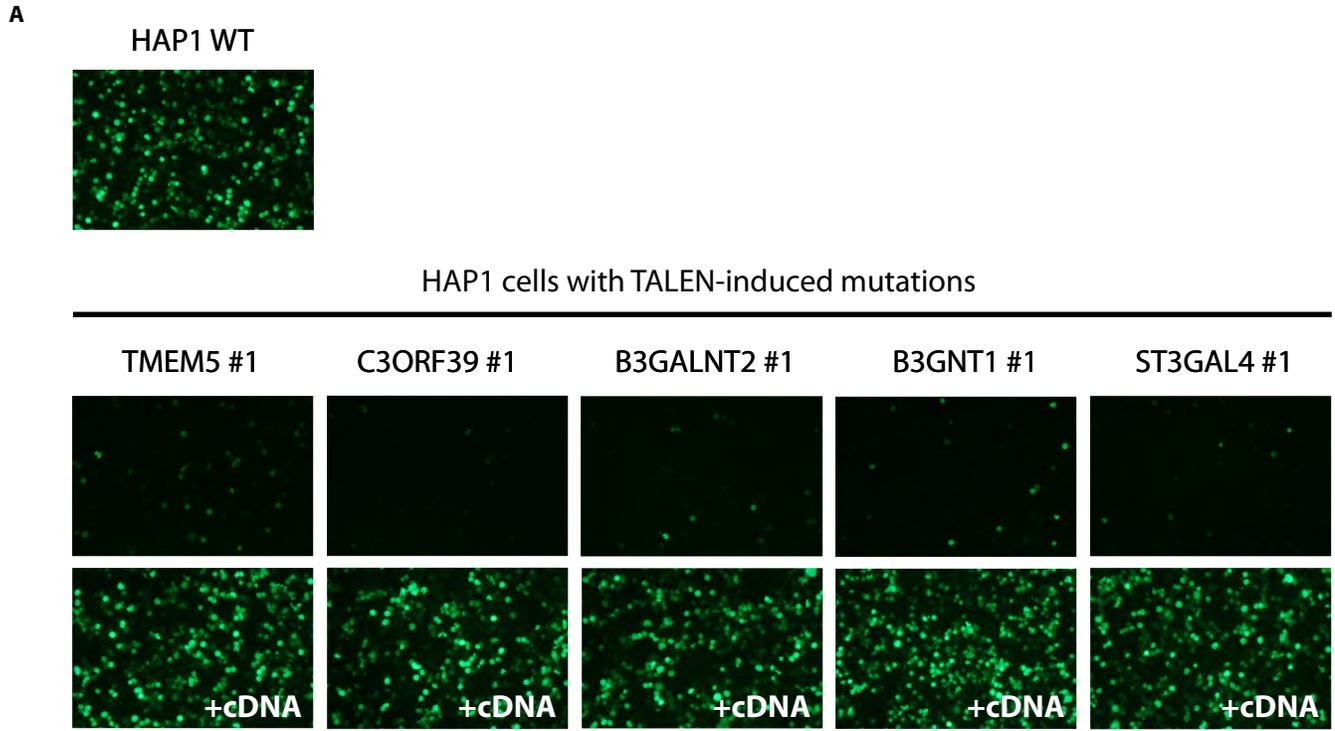
**B**



**C**



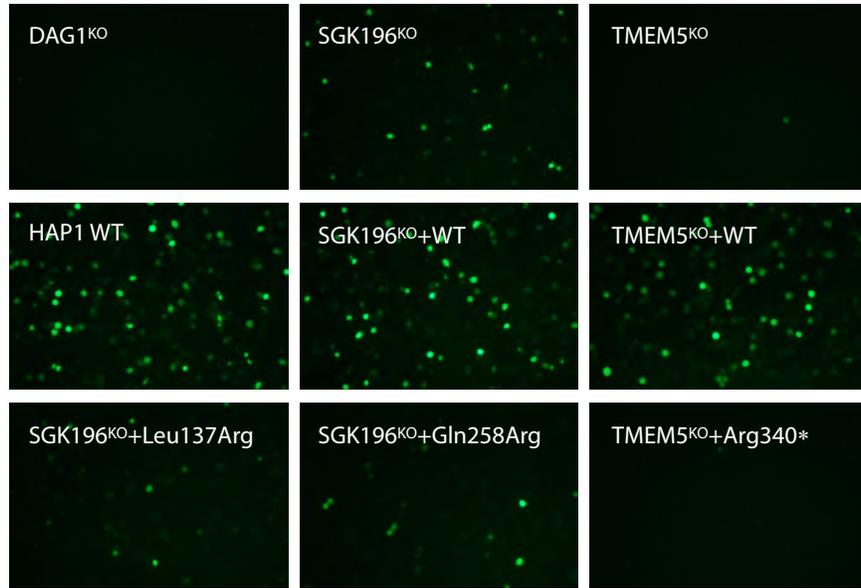
**Fig S11**



# Fig S12

Exostosin           saaaappakRktlffaGekrragksrsgairdklleekassdvqcee.....gneskqspenymellkssrFClvppGdtptsrrlfdallagCiPVi  
match           +++ ++ +R++l f G+ ++++   + +++l++ +++                   +e+++s +ny++ l +s +l+p G +   r+++a +g iPV+  
TMEM5           257-ASWSMLHDERPYLCNFLGTIYENSSR--QALMNILKKDGNDKLCW--VsarehwqPQETNESLKNYQDALLQSDLTLCPVGVNTECYRIYEACSYGSIPVV-353

**Fig S13**



## References and Notes

1. T. Yoshida-Moriguchi *et al.*, O-Mannosyl phosphorylation of alpha-dystroglycan is required for laminin binding. *Science* **327**, 88 (2010). doi:10.1126/science.1180512 Medline
2. C. Godfrey, A. R. Foley, E. Clement, F. Muntoni, Dystroglycanopathies: Coming into focus. *Curr. Opin. Genet. Dev.* **21**, 278 (2011). doi:10.1016/j.gde.2011.02.001 Medline
3. D. E. Michele *et al.*, Post-translational disruption of dystroglycan-ligand interactions in congenital muscular dystrophies. *Nature* **418**, 417 (2002). doi:10.1038/nature00837 Medline
4. A. Rambukkana *et al.*, Role of alpha-dystroglycan as a Schwann cell receptor for *Mycobacterium leprae*. *Science* **282**, 2076 (1998). doi:10.1126/science.282.5396.2076 Medline
5. W. Cao *et al.*, Identification of alpha-dystroglycan as a receptor for lymphocytic choriomeningitis virus and Lassa fever virus. *Science* **282**, 2079 (1998). doi:10.1126/science.282.5396.2079 Medline
6. S. Kunz *et al.*, Posttranslational modification of alpha-dystroglycan, the cellular receptor for arenaviruses, by the glycosyltransferase LARGE is critical for virus binding. *J. Virol.* **79**, 14282 (2005). doi:10.1128/JVI.79.22.14282-14296.2005 Medline
7. M. C. Manzini *et al.*, Exome sequencing and functional validation in zebrafish identify GTDC2 mutations as a cause of Walker-Warburg syndrome. *Am. J. Hum. Genet.* **91**, 541 (2012). doi:10.1016/j.ajhg.2012.07.009 Medline
8. T. Roscioli *et al.*, Mutations in ISPD cause Walker-Warburg syndrome and defective glycosylation of  $\alpha$ -dystroglycan. *Nat. Genet.* **44**, 581 (2012). doi:10.1038/ng.2253 Medline
9. T. Willer *et al.*, ISPD loss-of-function mutations disrupt dystroglycan O-mannosylation and cause Walker-Warburg syndrome. *Nat. Genet.* **44**, 575 (2012). doi:10.1038/ng.2252 Medline
10. J. E. Carette *et al.*, Haploid genetic screens in human cells identify host factors used by pathogens. *Science* **326**, 1231 (2009). doi:10.1126/science.1178955 Medline
11. J. E. Carette *et al.*, Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature* **477**, 340 (2011). doi:10.1038/nature10348 Medline
12. J. E. Carette *et al.*, Global gene disruption in human cells to assign genes to phenotypes by deep sequencing. *Nat. Biotechnol.* **29**, 542 (2011). doi:10.1038/nbt.1857 Medline
13. Materials and methods are available as supplementary material on *Science* Online.
14. D. J. Lefeber *et al.*, Deficiency of Dol-P-Man synthase subunit DPM3 bridges the congenital disorders of glycosylation with the dystroglycanopathies. *Am. J. Hum. Genet.* **85**, 76 (2009). doi:10.1016/j.ajhg.2009.06.006 Medline

15. K. Buysse *et al.*, Missense mutations in  $\beta$ -1,3-N-acetylglucosaminyltransferase 1 (B3GNT1) cause Walker-Warburg syndrome. *Hum. Mol. Genet.* (2013). doi:10.1093/hmg/ddt021 Medline
16. A. C. Combs, J. M. Ervasti, Enhanced laminin binding by alpha-dystroglycan after enzymatic deglycosylation. *Biochem. J.* **390**, 303 (2005). doi:10.1042/BJ20050375 Medline
17. S. K. Patnaik, P. Stanley, Lectin-resistant CHO glycosylation mutants. *Methods Enzymol.* **416**, 159 (2006). doi:10.1016/S0076-6879(06)16011-5 Medline
18. M. Eckhardt, M. Mühlenhoff, A. Bethe, R. Gerardy-Schahn, Expression cloning of the Golgi CMP-sialic acid transporter. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 7572 (1996). doi:10.1073/pnas.93.15.7572 Medline
19. R. J. L. Hari, G. Garg, Charles A. Hales, *Chemistry and Biology of Heparin and Heparan Sulfate* (Elsevier, Kidlington, Oxford, UK, 2005).
20. K. Inamori *et al.*, Dystroglycan function requires xylosyl- and glucuronyltransferase activities of LARGE. *Science* **335**, 93 (2012). doi:10.1126/science.1214115 Medline
21. F. Foulquier *et al.*, TMEM165 deficiency causes a congenital disorder of glycosylation. *Am. J. Hum. Genet.* **91**, 15 (2012). Medline
22. N. E. Sanjana *et al.*, A transcription activator-like effector toolbox for genome engineering. *Nat. Protoc.* **7**, 171 (2012). doi:10.1038/nprot.2011.431 Medline
23. P. Vogel *et al.*, Congenital hydrocephalus in genetically engineered mice. *Vet. Pathol.* **49**, 166 (2012). doi:10.1177/0300985811415708 Medline
24. S. Vuillaumier-Barrot *et al.*, Identification of mutations in TMEM5 and ISPD as a cause of severe cobblestone lissencephaly. *Am. J. Hum. Genet.* **91**, 1135 (2012). doi:10.1016/j.ajhg.2012.10.009 Medline
25. P. C. Sabeti *et al.*; International HapMap Consortium, Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913 (2007). doi:10.1038/nature06250 Medline
26. S. P. Whelan, L. A. Ball, J. N. Barr, G. T. Wertz, Efficient recovery of infectious vesicular stomatitis virus entirely from cDNA clones. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 8388 (1995). doi:10.1073/pnas.92.18.8388 Medline
27. A. C. Wong, R. G. Sandesara, N. Mulherkar, S. P. Whelan, K. Chandran, A forward genetic strategy reveals destabilizing mutations in the Ebolavirus glycoprotein that alter its protease dependence during cell entry. *J. Virol.* **84**, 163 (2010). doi:10.1128/JVI.01832-09 Medline
28. D. J. Lefeber *et al.*, Autosomal recessive dilated cardiomyopathy due to DOLK mutations results from abnormal dystroglycan O-mannosylation. *PLoS Genet.* **7**, e1002427 (2011). doi:10.1371/journal.pgen.1002427 Medline

29. A. Banerjee, M. C. Sammarco, S. Ditch, J. Wang, E. Grabczyk, A novel tandem reporter quantifies RNA polymerase II termination in mammalian cells. *PLoS ONE* **4**, e6193 (2009). doi:10.1371/journal.pone.0006193 Medline
30. L. S. Collier, C. M. Carlson, S. Ravimohan, A. J. Dupuy, D. A. Largaespada, Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. *Nature* **436**, 272 (2005). doi:10.1038/nature03681 Medline