# Identification of novel therapeutic targets associated with aberrant glycosylation in stem cells of triple-negative breast cancer

Ricard Bonilla Amadeo

# Identification of novel therapeutic targets associated with aberrant glycosylation in stem cells of triple-negative breast cancer

Thesis report presented by

**Ricard Bonilla Amadeo**

to obtain the degree of

**Doctor of Philosophy (PhD) for the University of Barcelona (UB)**

under the Biomedicine Doctoral Programme

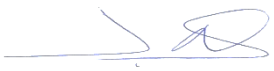
This thesis was conducted at the Translational Oncology Lab within the **Pathology and Experimental Therapy Department**, Faculty of Medicine in the University of Barcelona, under the scientific supervision of **Dr. Ruth Rodríguez Barrueco** and the academic supervision of **Dr. Josep Maria de Anta Vinyals**

Ruth Rodríguez Barrueco, PhD

*Director*

Josep Maria de Anta Vinyals, PhD

*Tutor*



Ricard Bonilla Amadeo, Doctoral student

*This thesis was made possible through the primary financial support of the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) and the Government of Catalonia, under the FI-SDUR grant (Code: 2020-FISDUR 00286). Additional funding from the projects 'Breast Cancer Stem-like Cells Specific Glycosylation as a Therapeutic Target' (Code: RYC-2016-19671) by the Spanish Ministry of Economy and Industry, and 'New Targeted Therapies Based on Aberrant Glycosylation of Breast Cancer Cells' (Code: CNS2022-135451) by the Spanish Ministry of Science and Innovation, provided crucial support for initiating and completing this research.*

# Abstract

Despite the generally high cure rates for breast cancer, a subset of patients still experiences metastasis and relapse, with triple-negative breast tumors (TNBC) being the most aggressive and associated with a poorer prognosis. A small population of cells within tumors, known as cancer stem-like cells (CSCs), are poorly differentiated and highly plastic, driving processes like migration, metastasis, and recurrence. Targeting these cells remains challenging due to their similarity to normal stem cells. However, critical distinctions between normal breast stem cells and breast cancer stem cells (BCSCs) have been identified. One notable difference is the presence of posttranslational modifications in BCSCs, particularly aberrant glycosylation, which often occurs on stemness markers.

Using our novel GlycoCRISPR library, which targets protein glycosylation genes, we performed the first comprehensive interrogation of the genes essential for maintaining stemness in TNBC cell lines. MDA-MB-231 cells (malignant) and MCF10A cells (non-transformed) were used for these screenings. Bioinformatic analysis revealed that ten glycosylation-related genes are specifically essential for maintaining the stem phenotype in cancer cells. Based on public patient data and extensive literature review, three key genes—EXT1, ST3GAL1, and DHDDS—were selected for further study. Clinical analysis using the METABRIC database showed that overexpression of these genes is significantly associated with worse prognosis, including lower overall survival and relapse-free survival rates.

*In vitro* experiments demonstrated that EXT1, ST3GAL1, and DHDDS are overexpressed in TNBC cells cultured in suspension, a condition used to enrich the stem-like cell population. Knockdown studies of EXT1 and DHDDS in triple negative breast cancer cell lines validated their role in stemness. A tumorsphere assay confirmed a significant reduction in sphere formation capacity following gene silencing in three different TNBC cell lines. Although no significant differences were observed in the $CD44^{+}/CD24^{-/low}$ population or stem-related gene expression between knockdown and control cells, the ALDH+ population was notably reduced when EXT1 and DHDDS were silenced, supporting their role in maintaining stemness.

In mouse tumor formation experiments, tumors with EXT1 and DHDDS knockdown showed significantly reduced growth compared to controls, with smaller tumor masses observed upon extraction. This underscores the role of these genes in sustaining the stem-like phenotype in TNBC.

Finally, by enzymatically degrading heparan sulfate in 231 cells using heparinase III and inhibiting N-glycosylation with tunicamycin to mimic the effects of EXT1 and DHDDS downregulation, respectively, we demonstrate that abnormal glycosylation can influence and sustain stem cell potential.

# Table of contents

**List of figures** ..... 9
**List of tables** ..... 16
**Abbreviations and acronyms** ..... 20
**Introduction** ..... 28
1. The breast: from the tissue to the functional units ..... 30
1.1. Breast tissue: physiology ..... 30
1.2. Breast cell types and microenvironmental cells ..... 31
2. Breast cancer ..... 33
2.1. Pathology ..... 33
2.2. Classification ..... 36
2.3. Triple-negative breast cancer (TNBC) ..... 38
2.4. Treatment for TNBC ..... 39
3. Metastasis, plasticity and stemness ..... 40
3.1. Metastasis process ..... 40
3.2. Phenotypic plasticity ..... 45
3.3. Stemness and its role in breast cancer ..... 46
4. Post-translational modification: Protein Glycosylation ..... 50
4.1. Types of glycosylation ..... 50
4.2. Biosynthetic process ..... 51
4.3. Physiological roles of protein glycosylation ..... 54
4.4. Glycosylation and cancer ..... 54
4.5. Therapeutic approaches against aberrant protein glycosylation ..... 56
4.6. Role of protein glycosylation in cell plasticity and stemness? ..... 58
5. CRISPR screenings: the edition for the elucidation ..... 59
5.1. CRISPR/Cas system ..... 59
5.2. Modified Cas nucleases ..... 60
5.3. CRISPR libraries ..... 61
5.4. Pooled CRISPR screenings ..... 61
**Hypothesis and objectives** ..... 67
**Materials and methods** ..... 71
1. Cell culture ..... 73

1.1 Cell types and characteristics ..... 73
1.2 Maintenance and counting ..... 73
1.3 Cryopreservation and thawing ..... 74
2. Engineering of cell lines by lentiviral transduction ..... 74
2.1 Transfection procedure ..... 74
2.2 Transduction of cell lines ..... 75
2.3 Plasmids for cell line genetic modification ..... 76
3. GlycoCRISPR library and screening experiments ..... 79
3.1 GlycoCRISPR library design ..... 79
3.2 Glycolibrary screening pipeline ..... 80
3.3 Glycolibrary screening set up ..... 82
3.4 Sample processing and NGS ..... 84
3.5 NGS data analysis ..... 85
4. Patient data analysis validation ..... 86
5. Bacterial procedures ..... 86
5.1 Bacterial transformation ..... 86
5.2 Colony formation from bacterial clones ..... 87
5.3 Plasmid amplification ..... 87
6. Gene expression analysis ..... 87
6.1 RNA extraction ..... 87
6.2 Reverse transcription PCR (RT-PCR) ..... 88
6.3 Real-Time quantitative PCR (RT-qPCR) ..... 88
7. Protein expression analysis ..... 89
7.1 Cell lysis and protein quantification ..... 89
7.2 Western Blot (WB) ..... 90
7.3 Cell surface protein detection by flow cytometry ..... 91
7.4 ALDH protein activity by ALDERED assay protocol using FACS ..... 91
8. Functional assays ..... 92
8.1 Cell viability assay by Thyazolyl Blue Tetrazolium Bromide (MTT) ..... 92
8.2 Tumorsphere formation assay ..... 92
9. *In vivo* assays ..... 93
10. Histology ..... 94
10.1 Paraffin embedding ..... 94

10.2 Hematoxylin/eosin staining ..... 94
11. Compounds ..... 94
12. Statistical analysis ..... 95
**Results** ..... 97
1. CRISPR/Cas9 screening in triple-negative breast cancer cell lines ..... 99
1.1 Defining the workflow for our screenings ..... 99
1.2 CRISPR screening in MDA-MB-231 cells ..... 101
1.3 CRISPR screening in MCF10A cells ..... 102
1.4 Bioinformatic analyses for essential hits in stemness conditions ..... 104
1.5 Hit selection ..... 107
1.6 Stemness gene sets and patient associations with our hits ..... 114
2. Functional validation in MDA MB 231 cells ..... 115
2.1 Hits expression in tumorsphere condition ..... 116
2.2 Hits silencing using miR.E system ..... 117
2.3 Effects of target silencing in tumorsphere formation capacity ..... 118
2.4 Analyses of stem properties after knocking-down our targets ..... 119
3. Validation of EXT1 and DHDDS in TNBC cell lines ..... 122
3.1 Gene expression in transformed and non-transformed cell lines ..... 122
3.2 Knockdown of EXT1 and DHDDS in HS578T and HCC cell lines ..... 123
3.3 Role in stemness of EXT1 and DHDDS in HS578T and HCC70 cell lines ..... 123
3.4 Analysis of $ALDH^+$ population in EXT1- and DHDDS-silenced HCC70 cell lines ..... 125
4. *In vivo* validation of EXT1 and DHDDS role in tumor progression ..... 126
5. Unveiling the mechanistic effects of underlying glycosylation in stemness ..... 129
5.1 Role of Heparin sulfate in cell stemness ..... 129
5.2 Impact of N-glycosylation in cell plasticity ..... 131
**Discussion** ..... 135
**Conclusions** ..... 147
**Bibliography** ..... 151
**Annexes** ..... 177

# List of figures

**Figure 1: Architecture of mature female breast ..... 30**

**Figure 2: Reproductive and metabolic hormones implicated in the different hormone-dependent stages of female life. ..... 31**

**Figure 3: The three different models of cellular lineage origin. ..... 32**

**Figure 4: Statistic data of incidence and mortality of BC.. ..... 34**

**Figure 5: Representation of the incidence rates' increment by age of diagnose in United States. 35**

**Figure 6: Different classification of breast cancer ..... 37**

**Figure 7: Schematic representation of the three steps carried out in the metastatic process ..... 40**

**Figure 8: Summary of all the target organs in BC.. ..... 44**

**Figure 9: Types of glycans and glycoconjugates found in humans.. ..... 51**

**Figure 10: Biological pathway of protein glycosylation.. ..... 52**

**Figure 11: All human glycosylation pathways and enzymes involved in. ..... 53**

**Figure 12: Role of glycans in different cancer processes ..... 55**

**Figure 13: Mechanism of CRISPR/Cas9 action ..... 60**

**Figure 14: Design to perform a pooled CRISPR screening ..... 62**

**Figure 15: Workflow of pooled CRISPR screening ..... 64**

**Figure 16: Pipeline of virus production and transduction process.. ..... 75**

**Figure 17: Map of the miR.E plasmid ..... 76**

**Figure 18: Map of the LentiNeoLuc plasmid. ..... 79**

**Figure 19: Graphical distribution of all sgRNA guides of our GlycoCRISPR library ..... 80**

**Figure 20: Schematic workflow of our screening. ..... 82**

**Figure 21: Poisson distribution that follows MOI, regarding the number of viral particles that can enter to 1 cell. ..... 83**

**Figure 22: Analysis of CD44 and CD24 surface protein markers by FACS in TNBC cell lines ..... 99**

**Figure 23: Analysis of ALDH+ subset measured by FACS in TNBC cell lines ..... 100**

**Figure 24: Validation of tumorsphere assay as a stem-like cell enriching method ..... 100**

**Figure 25: Pipeline of our CRISPR screening ..... 101**

**Figure 26: Representative images of MDA-MB-231 cell line through screening..... 102**

**Figure 27: Representative images of MCF10A cell line through the screening ..... 103**

**Figure 28: Changes in the representation of negative control sgRNA guides across conditions in the screenings ..... 104**

**Figure 29: Changes in the representation of positive control sgRNA guides across conditions in the screenings ..... 105**

**Figure 30: Volcano plots of sgRNA guides lost or enriched in MCF10A or 231 cell lines..... 106**

**Figure 31: Volcano plot of sgRNA guides lost and enriched in 3rd gen for MDA-MB-231 cell line. ..... 107**

**Figure 32: Venn's diagram of genes lost in the 2D and 3D conditions in both MDA-MB-231 and MCF10A. ..... 107**

**Figure 33: Representative scheme of heparan sulfate chain synthesis and elongation ..... 109**

**Figure 34: Correlation of EXT1 expression with survival and relapse in patients from METABRIC database..... 110**

**Figure 35: Differential expression across molecular subtypes and signatures associated with EXT1 expression. . ..... 110**

**Figure 36: Function of DHDDS, together with NgBR in the complex DDS, into the dolichol phosphate formation pathway ..... 111**

**Figure 37: Clinical parameters from the METABRIC dataset related to DHDDS expression ..... 111**

**Figure 38: Differential expression across molecular subtypes and signatures associated with DHDDS expression using METABRIC data ..... 112**

**Figure 39: Mucin type O-glycan formation ..... 112**

**Figure 40: Clinical parameters from the METABRIC dataset related to ST3GAL1 expression..... 113**

**Figure 41: Differential expression of DHDDS across molecular subtypes and associated signatures, as analyzed using METABRIC data..... 114**

**Figure 42: Stemness gene expression sets' correlation with each target expression ..... 115**

**Figure 43: Expression of our targets in MDA-MB-231 cells under suspension and in attachment conditions..... 116**

**Figure 44: Expression of targets in MDA-MB-231 cells after performing the silencing using the miR.E system ..... 117**

**Figure 45: Abiliity for tumorsphere formation of the EXT1 and DHDDS-silenced MDA-MB-231 cells ..... 118**

**Figure 46: Changes in the ability for tumorsphere formation in the ST3GAL1-silenced MDA-MB-231 cells ..... 119**

**Figure 47: Analysis of CD44+/CD24-/low population in EXT1- and DHDDS-silenced MDA-MB-231 cells under 2D condition..... 120**

**Figure 48: Analysis of CD44+/CD24-/low population in EXT1- and DHDDS-silenced MDA-MB-231 cells in 3D condition ..... 121**

**Figure 49: Changes in RNA expression in EXT1- (left) and DHDDS-silenced cells (right) when cultured in suspension ..... 122**

**Figure 50: RNA expression of different basal cell lines in 2D and 3D ..... 122**

**Figure 51: Knockdown of EXT1 and DHDDS conducted in HS578T and HCC70 ..... 123**

**Figure 52: Ability for tumorsphere formation in the EXT1 and DHDDS-silenced HS578T cells ... 124**

**Figure 53: Ability for tumorsphere formation in the EXT1 and DHDDS-silenced HCC70 cells..... 124**

**Figure 54: Analysis of ALDH+ population by FACS in EXT1-silenced HCC70 cell lines ..... 125**

**Figure 55: Analysis of ALDH+ subset by FACS in DHDDS-silenced HCC70 cell lines ..... 125**

**Figure 56: In vivo experiments in immunocompromised NSG mice using EXT1-silenced MDA-MB-231 cells ..... 126**

**Figure 57: Hematoxylin/eosin staining of scramble and shEXT1 tumors at 4X and 10X magnification ..... 127**

**Figure 58: In vivo experiments in immunocompromised NSG mice using DHDDS-silenced MDA-MB-231 cells ..... 128**

**Figure 59: Hematoxylin/eosin staining of scramble and shDHDDS tumors at 4X and 10X magnification. ..... 129**

**Figure 60: Heparan sulfate detection by FACS ..... 130**

**Figure 61: Heparinase treatment and its effect in stemness ..... 131**

**Figure 62: The initial step of N-glycosylation ..... 131**

**Figure 63: Impact of tunicamycin treatment in MDA-MB-231 cells ..... 132**

**Figure 64: Changes in the ability for tumorsphere formation in tunicamycin-treated MDA-MB-231 cells ..... 132**

# List of tables

**Table I:** Different glyco-based therapies against cancer pathology. ..... 57

**Table II:** Commercial cell lines used and their molecular features. ..... 73

**Table III:** Media used for each cell type. ..... 73

**Table IV**: All antisense sequence to knock-down our targets. ..... 76

**Table V**: Primers overhangs with the restriction enzymes sites. ..... 77

**Table VI**: Components for PCR (left) and thermocycler protocol to carry out the PCR (right). ..... 78

**Table VII**: Components to perform the digestion of the template. ..... 78

**Table VIII**: All reagents for ligation procedure. ..... 78

**Table IX**: List of primers used for sequencing all the samples. The specific barcode for each primer is highlighted in yellow. ..... 84

**Table X**: Products to carry out the RT-PCR. ..... 88

**Table XI**: SYBR® Green mix for qPCR. ..... 88

**Table XII**: Primers designed for qPCR. ..... 89

**Table XIII**: All antibodies used in the thesis, indicating the dilution factor, host of the antibody and their references ..... 90

**Table XIV**: Recipe to produce a 50mL aliquot of tumorsphere media. ..... 92

**Table XV**: Summary of the parameters changed in each different experiment to set up the screening conditions in MDA-MB-231 cells. ..... 101

**Table XVI**: Summary of the parameters modified in each different experiment to set up the screening in MCF10A.. ..... 103

**Table XVII**: Summary of all the information compiled for all the targets found ..... 108

# Abbreviations and acronyms

| First letter | Abbreviation | Meaning |
|---|---|---|
| A | ABC | ATP-binding cassette |
| | ALDH | Aldehyde Dehydrogenase |
| | APC | Fllophycocyanin |
| | APS | Ammonium persulfate |
| | ATCC | American Type Culture Collection |
| | ATM | Ataxia Telangiectasia Mutated Kinase |
| B | BBB | Blood-brain barrier |
| | BCSC | Breast cancer stem cell |
| | BL1 | Basal-like 1 |
| | BL2 | Basal-like 2 |
| | BLIA | Basal-like immune activated |
| | BLIS | Basal-like immune suppressed |
| | bp | Base pair |
| | BRCA1 | Breast cancer 1 |
| | BRCA2 | Breast cancer 2 |
| | BSA | Bovine Serum Albumin |
| | BTIC | Breast tumor-initiating cell |
| C | CAR-T | Chimeric antigen receptor T |
| | CDP | Cytidine diphosphate |
| | CHEK2 | Checkpoint Kinase 2 |
| | CMP | Cytidine monophosphate |
| | CRISPR | Clustered regulatory interspaced short palindromic repeats |
| | CRISPRa | Clustered regulatory interspaced short palindromic repeats of activation |
| | CRISPRi | Clustered regulatory interspaced short palindromic repeats of inhibition |
| | CSC | Cancer stem cell |
| | Ct | Cycle threshold |
| | CTC | Circulating tumor cell |

| | | |
|---|---|---|
| **D** | DBS | Double-strand break |
| | DCIS | Ductal carcinoma *in situ* |
| | DDS | Dehydrodolichyl diphosphate synthase |
| | DEAB | Diethylaminobenzaldehyde |
| | DEAE | Diethylethanolamine |
| | DMEM | Dulbecco's Modified Eagle Medium |
| | DMSO | Dimethyl sulfoxide |
| | DNA | Deoxyribonucleic acid |
| | dNTP | Deoxynucleoside triphosphates |
| | Dol-P | Dolichol phosphate |
| | DTC | Disseminated tumor cell |
| | DTT | Dithiothreitol |
| **E** | ECM | Extracellular matrix |
| | EDTA | Ethylenediaminetetraacetic acid |
| | EGFR | Epidermal growth factor receptor |
| | EMP | Epithelial-mesenchymal plasticity |
| | EMT | Epithelial-to-mesenchymal transition |
| | EpCAM | Epithelial Cell Adhesion Molecule |
| | ER | Estrogen receptor |
| | Er | Endoplasmic reticulum |
| **F** | FACS | Fluorescent-activation cell sorting |
| | FBS | Fetal Bovine Serum |
| | FC | Flow cytometry |
| | FDR | False discovery rate |
| | FOXA1 | Forkhead Box A1 |
| | FW | Forward |
| **G** | GAPDH | Glyceraldehyde 3-phosphate dehydrogenase |
| | GATA3 | GATA Binding Protein 3 |
| | GDP | Guanosine diphosphate |
| | GFP | Green fluorescent protein |
| | GH | Growth hormone |

| | | |
|---|---|---|
| | GO | Gene Ontology |
| | GPI | Glycosylphosphatidylinositol |
| **H** | hbFGF | Human basic fibroblast growth factor |
| | hCIT | Human cis-prenyltransferase |
| | HDR | Homologous directed repair |
| | hEGF | Human epidermal growth factor |
| | HEMA | 2-hydroxyethyl methacrylate |
| | HER2 | Human epidermal growth factor receptor 2 |
| | HRP | Horseradish peroxidase |
| | HS | Heparan sulfate |
| | HSPG | Heparan sulfate proteoglycan |
| **I** | IDC | Invasive ductal carcinoma |
| | IGF | Insulin growth factor |
| | IGFR | Insulin growth factor receptor |
| | ILC | invasive lobular carcinoma |
| | IM | Immunomodulatory |
| **K** | KI | Knock in |
| | KO | Knock out |
| **L** | LAR | Luminal androgen receptor |
| | LCIS | Lobular carcinoma *in situ* |
| | logFC | Logarithmic fold change |
| **M** | M | Mesenchymal breast TNBC subtype |
| | MACS | Magnetic cell sorting |
| | MaSC | Mammary stem cell |
| | MET | Mesenchymal-to-epithelial transition |
| | METABRIC | Molecular Taxonomy of Breast Cancer International Consortium |
| | MHC | Major histocompatibility complex |
| | MIC | Metastasis-initiating cells |
| | MOI | Multiplicity of infection |
| | mRNA | Messenger RNA |

| | | |
|---|---|---|
| | MSL | Mesenchymal stem-like |
| | MTT | Thyazolyl Blue Tetrazolium Bromide |
| | MUC | Mucin |
| **N** | NADPH | Nicotinamide adenine dinucleotide phosphate |
| | NgBR | Nogo-B receptor |
| | NGS | Next Generation Sequencing |
| | NHEJ | Non-homologous end joining |
| | NSG | NOD scid gamma |
| | NST | No special type |
| **O** | OS | Overall survival |
| **P** | PAM | Protospacer Adjacent motifs |
| | PAM50 | Predictor Analysis of Microarray 50 |
| | PARP | Poly(ADP-Ribose) Polymerase |
| | PBS | Phosphate-buffered saline |
| | PCR | Polymerase chain reaction |
| | PD-1 | Programmed Cell Death 1 |
| | PD-L1 | Programmed Cell Death receptor 1 |
| | PE | Phycoerythrin |
| | PR | Progesterone receptor |
| | PTEN | Phosphatase And Tensin Homolog |
| **R** | RNA | Ribonucleic acid |
| | RPMI | Roswell Park Memorial Institute |
| | RT | Room temperature |
| | RT-qPCR | Real time quantitative PCR |
| | RV | Reverse |
| **S** | SDS | Sodium dodecyl sulfate |
| | sgRNA | single-guide RNA |
| | SMO | Smoothened, Frizzled Class Receptor |
| | SOX | SRY-Box Transcription Factor |
| | SPF | Specific Pathogen-Free |
| | SSB | single-strand break |

| | | |
|---|---|---|
| **T** | TACA | tumor-associated carbohydrate antigens |
| | TBS | Tris-buffered saline |
| | TEMED | Tetramethylethylenediamine |
| | TF | Transcription factors |
| | TGF-β | Transforming growth factor beta |
| | TIC | Tumor-initiating cell |
| | TME | Tumor microenvironment |
| | Tn | Thomsen-noveau |
| | TNBC | Triple-negative breast cancer |
| | TP53 | Tumor protein 53 |
| **U** | UDP | Uridine diphosphate glucose |
| | UNS | Unspecified group TNBC subtype |
| **V** | VEGF | Vascular Endothelial Growth Factor |
| **W** | WB | Western Blot |

# Introduction

# 1. The breast: from the tissue to the functional units

## 1.1. Breast tissue: physiology

The mature female breast comprises two primary structural components: the glandular and the stromal portion. The glandular portion, known as the **mammary gland**, is organized into lobes. Within these lobes, there are **lobules** consisting on clusters of **alveoli** responsible for milk secretion. Milk is ejected through milk ducts connecting the lobules to the nipple. The stromal portion is formed by collagen extracellular matrix (ECM) and connective tissue. The later contains vascular, lymphatic, immune elements but is mainly composed of fatty tissue, which is responsible of the overall architecture of the breast (Fig. 1) (Hannan et al., 2023).

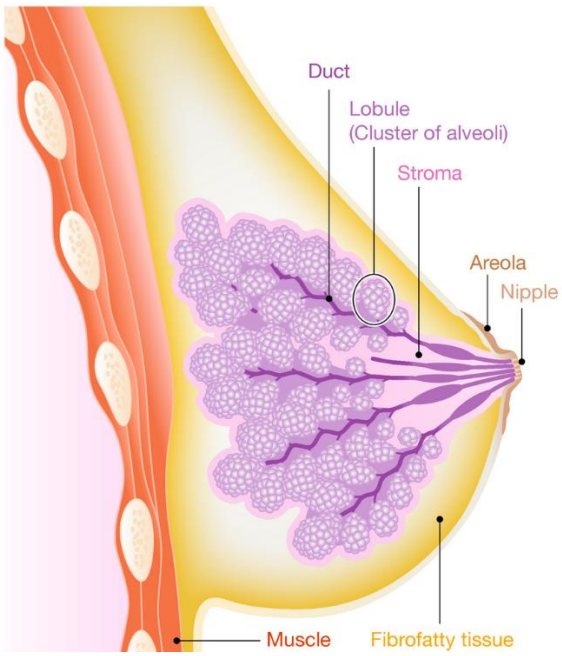



**Figure 1 Architecture of mature female breast** in which the mammary gland and the surrounding stroma can be observed (adapted from Pellacani et al., 2019).

The main function of the mammary gland is the milk ejection for breastfeeding. During the onset of puberty, the mammary gland suffers changes in preparation for branching morphogenesis. These changes are orchestrated mainly by female sex hormones, progesterone and oestrogen, but also by metabolic hormones such as growth hormone (GH) and insulin-like growth factor (IGF). During pregnancy, additional hormones play important roles in the mammary gland expansion. For instance, prolactin, along with progesterone, enables cell proliferation, ductal branching and alveologenesis (Biswas et al., 2022; Hannan et al., 2023). Insulin also influences the secretory differentiation in the mammary gland during this process (Neville et al., 2013). With childbirth, feminine sex hormones drop their influence in the mammary gland. For instance, prolactin, along with oxytocin, helps milk secretion (Biswas et al., 2022; Hannan et al., 2023) (Fig. 2). Additionally, the mammary gland is and endocrine organ, releasing parathyroid hormone-related peptide which stimulates the circulation of maternal calcium for milk synthesis (Hannan et al., 2023).

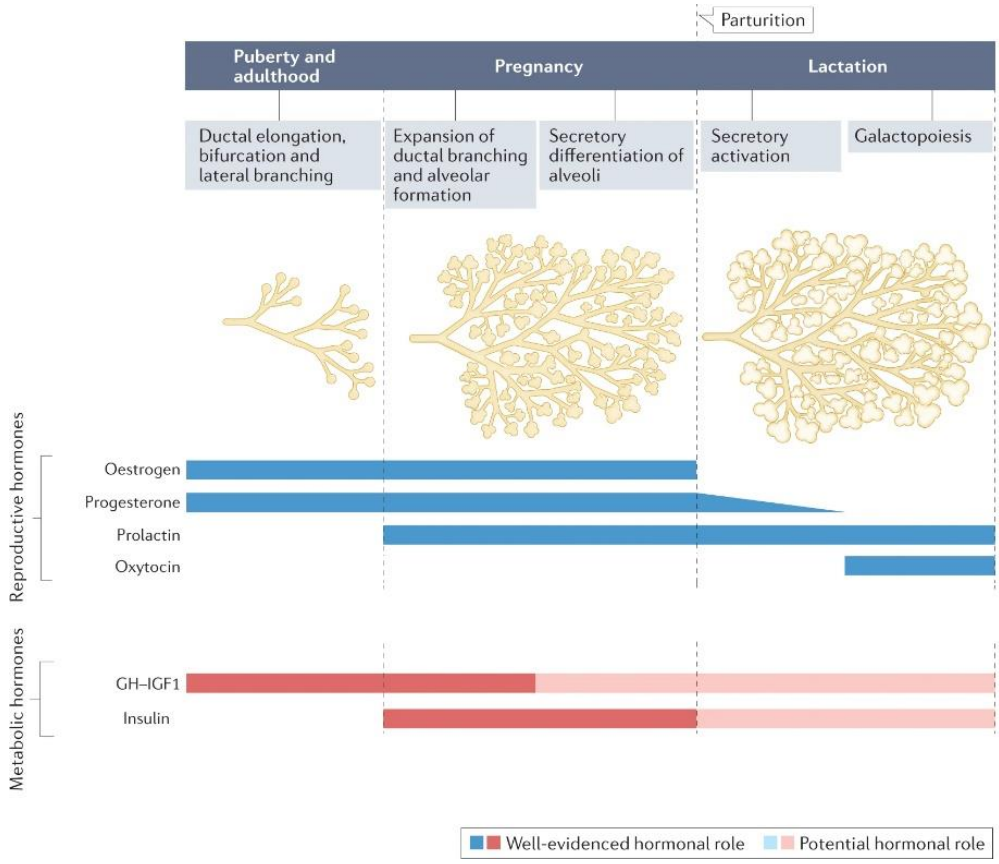



**Figure 2: Reproductive and metabolic hormones implicated in the different hormone-dependent stages of female life.** (adapted from Hannan et al., 2023)

## 1.2. Breast cell types and microenvironmental cells

The hotspot of the mammary gland is its functional unit, the **alveolus**. This structure, together with the ducts connecting to the nipple, comprises two main cell epithelial lineages (Lloyd-Lewis et al., 2017; Watson & Khaled, 2020). **Luminal cells**, which encircle the ducts and alveoli's lumen, are situated in the inner layer. The outer layer, directly connected to the basement membrane, consists of **myoepithelial cells** or **basal cells** (Cristea & Polyak, 2018; Seldin et al., 2017). All the mammary gland is embedded in a collagen-rich stroma mainly comprised by adipocytes, but also, by immune cells, blood and lymph vessels (Pellacani et al., 2019).

The origin of both basal and luminal cell lineages is not clear and three models have been proposed. In the first one, the **model of multipotent stem cells**, there is a niche of mammary stem cells (MaSCs) at the moment of birth, which gives rise to the committed luminal and myoepithelial progenitors and, subsequently, to the completely differentiated cell types during the adult life. Conversely, the second hypothesis consists on the **model of restricted progenitors** in which there is no MaSCs in the adult human being, only unipotent progenitors in the adulthood that derive to the mature cell types (Fu et al., 2020; Slepicka et al., 2021; Watson & Khaled, 2020). Nevertheless, single-cell RNA sequencing (scRNAseq) has led to a third hypothesis suggesting an interconnection among

multipotent stem cells and commited progenitors of each lineage and even between both lineages, defining the **continuous progression model** (Fig. 3) (Anstine & Keri, 2019; Wicker & Wagner, 2023).

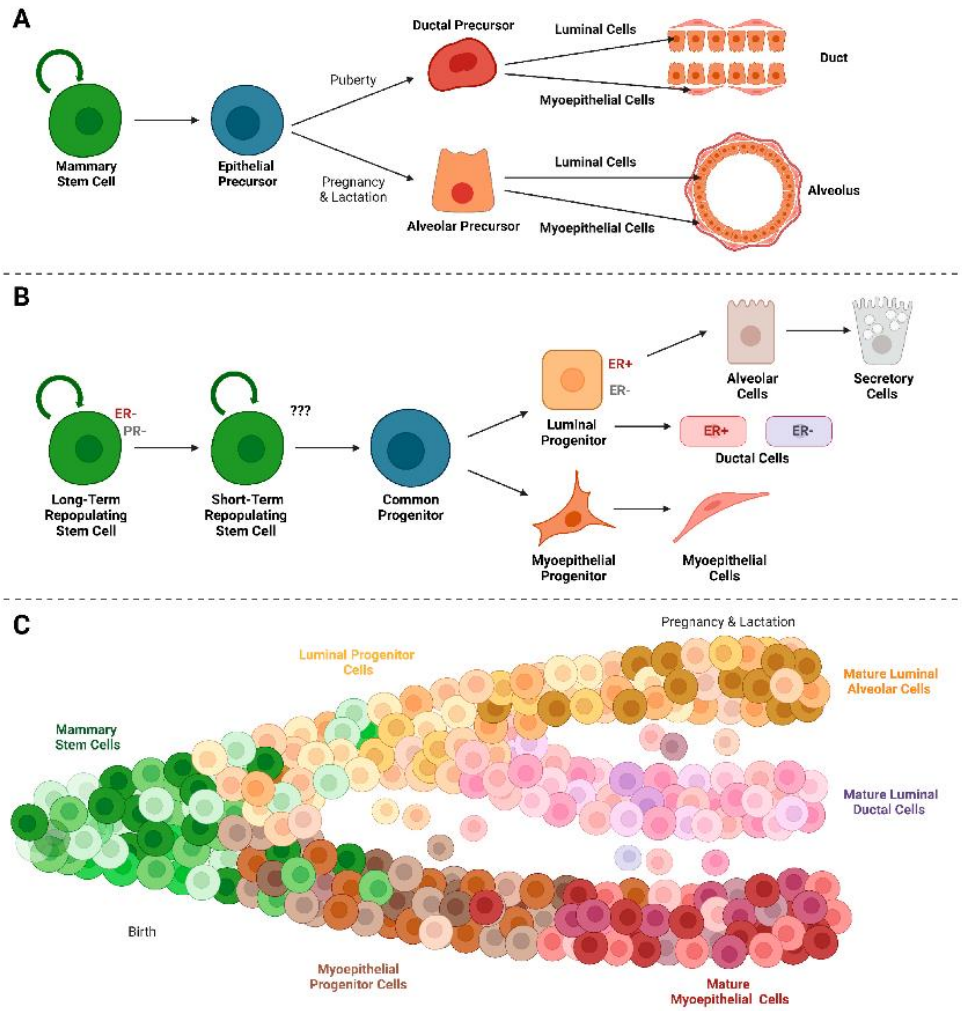



**Figure 3: The three different models of cellular lineage origin. A)** Multipotent stem cell model, **B)** restricted progenitors' model and **C)** continuous progression model (Wicker & Wagner, 2023)**.**

Within the luminal cell lineage, two distinct cell types can be identified based on the expression of two specific molecules: the estrogen receptor (ER) and the progesterone receptor (PR). Cells are classified as either ER/PR-positive if they express these receptors or ER/PR-negative if they do not (Taurin & Alkhalifa, 2020; Visvader & Lindeman, 2006; Watson & Khaled, 2020). During puberty and pregnancy, both receptors play a critical role with the rise of both hormones. As sensing cells, $ER^+/PR^+$ ductal luminal cells send paracrine signals to hormone-responding progenitor cells to induce proliferation. In late pregnancy, $ER^-/PR^-$ alveolar progenitors derive into secretory-specialized cells, the responsible of synthesis and secretion of milk (Fu et al., 2020; Hannan et al., 2023; Rodilla & Fre, 2018).

The term myoepithelial comes from the fact that the differentiated cells express smooth muscle protein to allow the contraction of the alveoli and ducts for milk secretion (Gieniec & Davis, 2022;

Samocha et al., 2019). The hierarchy of basal cells is yet controversial due to the fact that there are several markers that define different population of basal progenitors. In addition, there are basal early progenitors which derive to mature luminal specialized cells suggesting a plasticity in epithelial mammary cells (Rodilla & Fre, 2018; Samocha et al., 2019; Wicker & Wagner, 2023).

As previously mentioned, the whole mammary gland is embebbed in the **stroma** which conforms the main structure of the breast. One of the largest populations in breast are the **adipocytes**, whose main function is to store lipids within them. They provide a scaffold for epithelial but also for other stromal, immune, lymphatic and vascular cells. In addition, they are also responsible of secreting factors to support the branching of mammary gland, as well as the formation of new vasculature. Finally, they promote the paracrine activation of several pathways including ERα, IGF1 and HGF within the stromal compartment (Biswas et al., 2022; Inman et al., 2015).

The most relevant cell type in the breast are **fibroblasts**, which have the function to produce a lot of components of ECM such as collagen 1, fibronectin, laminin and elastin. Furthermore, they release to ECM cytokines, chemokines and growth factors which allow the bidirectional communication with epithelial cells of the mammary gland during its morphogenesis. Moreover, they secrete proteolytic enzymes that can modify ECM composition and influence cellular and tissue function (Biswas et al., 2022; Ingthorsson et al., 2022; Inman et al., 2015).

Finally, the **immune population** has an important role, promoting the elongation of mammary gland ducts. In addition, they phagocyte dead epithelial cells and are required for adipocyte repopulation in involution process. Interestingly, immune cells can maintain the epithelial progenitor cell niche. Additionally, eosinophils can secrete growth factors and cytokines for vascular formation, which is closely associated with lymphatic vessel organization (Biswas et al., 2022; Inman et al., 2015).

# 2. Breast cancer

## 2.1. Pathology

Breast cancer is a complex disease characterized by having multiples subtypes, sharing common features across diverse biological entities. The pathology of these subtypes is marked by their heterogeneity, exhibiting particular genomic alterations, gene expression patterns and distinct tumor microenvironment features that collectively influence patient's outcome and response to treatment.

### 2.1.1. Incidence and mortality

Breast cancer is a problem globally, being the cancer with the highest incidence and the fifth in cancer mortality worldwide in 2020 (Fig. 4A). Although in general breast cancer patients have a higher likelihood of survival and improved prognoses, the incidence of the disease continues to rise worldwide, creating a significant global health challenge (Sung et al., 2021). Interestingly, recent investigations from 2022 highlight breast cancer as the one with highest incidence and the most

mortal cancer among females in an age-world-standardized rate (ASW) of 12.6 over 100000 (Fig. 4B) (Ferlay J, 2022).

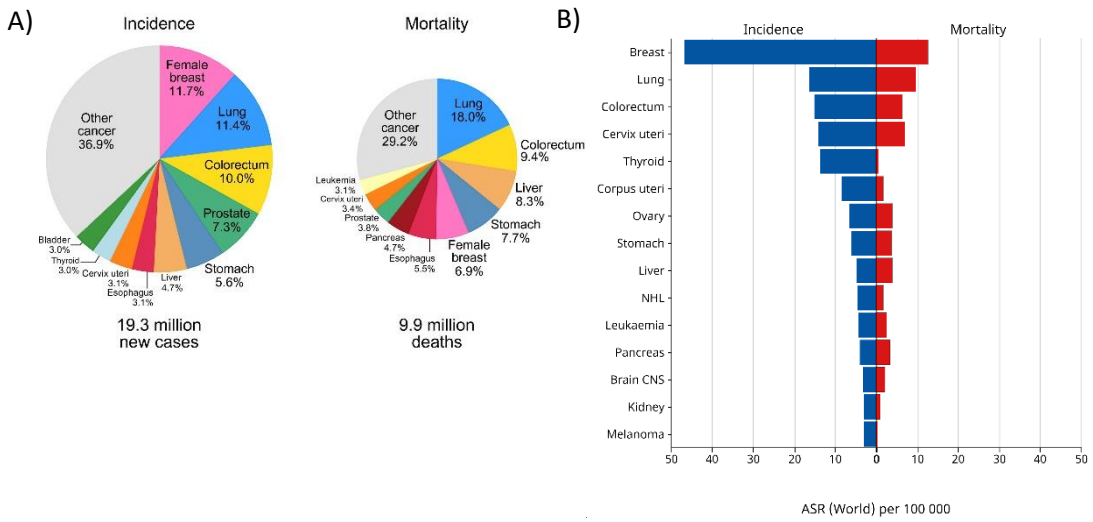



**Figure 4: Statistic data of incidence and mortality of BC. A)** Statistics of incidence and mortality of cancer disease in both sexes (Sung et al., 2021). **B)** New data from 2022 elucidating the incidence and mortality of cancer in women (Ferlay J, 2022).

Focusing on Spain, in 2020 breast cancer was the most frequent cancer in females and the third leading cause of cancer-related deaths (Ferlay J, 2020; Sung et al., 2021).

Although it is almost exclusively a woman issue, it also occurs in men representing only 1% of cases, being mainly associated with obesity and longer lifespan (Smolarz et al., 2022).

### 2.1.2. Risk factors

As a major problem for women's health, it is important to elucidate the primary risk factors that could lead to the breast cancer formation. While some of these factors are thoroughly researched and evident, there is a need for a deeper comprehensive investigation into others. Raising awareness about potential modifiable factors in women's life is crucial for minimizing the risk of experiencing the disease.

In first places, **sex** and **age** are the most important risk factors to develop breast cancer. As previously mentioned, almost all patients are **women** although male breast cancer represents 1% of the cases. In addition, increasing age is associated with higher incidence rates (Smolarz et al., 2022; Sun et al., 2017). As such, this trend is depicted in Fig. 5, illustrating the rising incidence rates depending on the age of diagnosis.

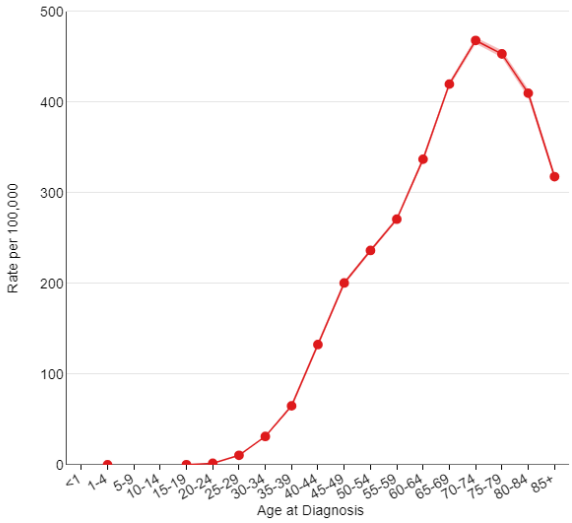



**Figure 5: Representation of the incidence rates' increment by age of diagnose in United States** (Surveillance Research Program, 2023).

About only 5 to 10% of all cases of breast cancer are inherited due to **hereditary gene mutations**. This susceptibility occurs because of well-studied mutations in, among others, both genes BRCA1 and BRCA2 (Houghton & Hankinson, 2021; Sarhangi et al., 2022). Due to various genetic mutations in molecular pathways related to breast cancer, it is estimated that approximately 10-30% of cases are attributed to hereditary factors, such as alterations in *TP53*, *PTEN*, *CHEK2* or *ATM* (Houghton & Hankinson, 2021; Sarhangi et al., 2022). However, having a first-degree relative with breast cancer, there is a 2-fold higher probability to develop it, and this likelihood increase if it is contracted in younger ages. (Smolarz et al., 2022).

Regarding **reproductive factors**, younger age of menarche, later menopause occurrence or late age of first pregnancy are associated with higher risk of breast cancer. These facts are correlated to the time of exposure of **estrogen hormone** and other sexual hormones, such as progesterone and prolactin. Therefore, variations in endogenous levels of hormones may modify the risk to suffer breast cancer. Additionally, exogenous exposure to sexual hormones is a crucial source of risk increase. Hence, oral contraceptives can raise the risk to develop breast cancer even for up to 10 years after stopping usage. Besides, the same occurs with hormone replacement therapy for menopausal or postmenopausal women (Houghton & Hankinson, 2021; Smolarz et al., 2022; Sun et al., 2017).

Concerning **lifestyle**, there are several components to take into account. First, **obesity** has a great impact in the breast cancer risk, particularly in postmenopausal women, who are more likely to develop hormone receptor-positive breast cancer. Furthermore, the combination of overweight with **sedentarism** raise the likelihood of being affected of breast cancer. Hence, maintaining a regular physical activity has been proven to be a protective factor. Moreover, fat tissue provides substrates for estrogen production, having an impact in breast cancer incidence. Adiposity can be linked to **insulin resistance** and the synthesis of insulin-like growth factor 1 (IGF-1), which is associated to a poorer overall survival, as well as higher possibilities of metastasis. The effect of **diet** on breast

cancer incidence needs to be further studied to be established as a protective or risk factor for breast cancer. However, an increased consumption of processed and red meat, animal fat and less ingestion of vegetables and fruit have been validated to be potential breast cancer risk factors. Finally, **alcohol intake** and smoking habit increase the probability of suffering breast cancer in a dose-dependent manner (Coughlin, 2019; Łukasiewicz et al., 2021; Smolarz et al., 2022).

## 2.2. Classification

Human breast tumors can be stratified according to several characteristics, primarily falling into two main categorizations: histological and molecular classification (Tsang & Tse, 2019).

### 2.2.1. Histologic subtypes

Mainly, histological groups have been established according to pathologic growth structure. Depending on cell's source, they can be carcinomas (from epithelial cells of mammary gland) or sarcomas (from connective tissue) (Zubair et al., 2021). At the moment of the diagnose, carcinomas are subset as *in situ* or invasive, regarding the penetration of the tumor to the surrounding tissues. The most typical form of pre-invasive carcinoma is the ductal carcinoma *in situ* (DCIS), which accounts for approximately 80% of pre-invasive tumors. The remainder are lobular carcinoma *in situ* (LCIS) (Nolan, Lindeman, et al., 2023). In case of invasive carcinomas, they are subdivided regarding cell morphology. Hence, the most common is the invasive ductal carcinoma (IDC), which shows no special type (NST) of cell histology and comprises 70-80% of invasive breast tumor cases, followed by invasive lobular carcinoma (ILC) representing 10-15% of cases. The rest of cases are less common histologic types, such as mucinous, papillary or metaplastic. Apart from cell source, this classification is made according to architectural properties and immunohistochemical profile. Nevertheless, as the NST categorization does not display concrete morphologic properties, many of them are classified as NST, not reflecting the vast heterogeneity of breast cancer (Fig. 6) (Nolan, Lindeman, et al., 2023; Tsang & Tse, 2019; Zubair et al., 2021).

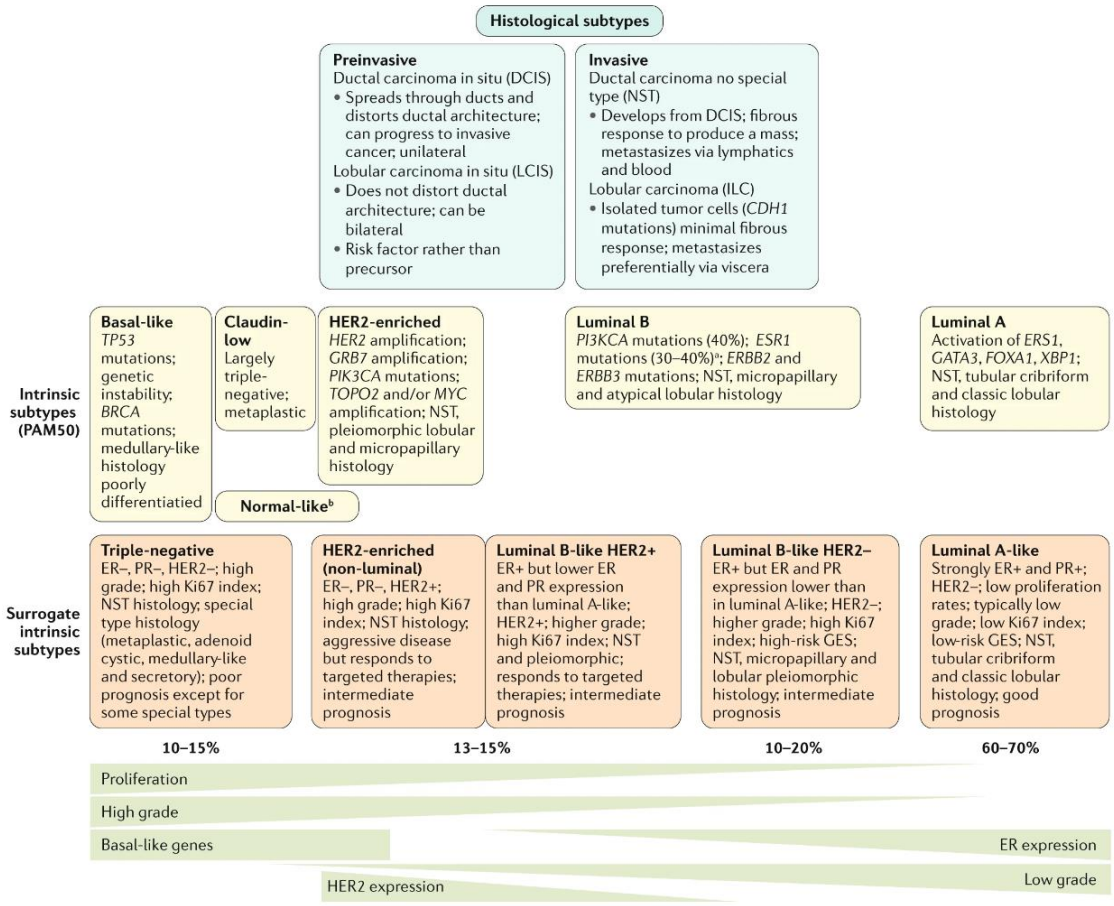



**Figure 6: Different classification of breast cancer.** In the upper part, there is the histological classification, pointing preinvasive and invasive forms. In the middle, both intrinsic and surrogate intrinsic classification are shown with the specific %of frequency of them. Finally, in the lower part, different features of the surrogated intrinsic subset are depicted (adapted from Harbeck et al., 2019).

### 2.2.2. Molecular subsets

Apart from the histologic classification, microarray expression profiling research deciphered six main molecular subsets, based on the signature of gene differential expression: **Luminal A**, **Luminal B**, **HER2-enriched**, **Basal-like**, **Normal-like** and **Claudin-low** (Fig. 6) (Nolan, Lindeman, et al., 2023). This classification, also named as intrinsic subtypes, is made according to Predictor Analysis of Microarray 50 (PAM50) test study, which is one of the most used, although there are other molecular profiling assays to classify breast tumors (Sarhangi et al., 2022).

**Luminal A** tumors constitute 50% of breast cancer, being the most prevalent subtype. They present a high expression of ER and/or PR, along with the manifestation of luminal gene markers like *GATA3* and *FOXA1* which are similarly observed in luminal normal mammary epithelium. These tumors, characterized by a lower presence of cell proliferation genes, tend to have a slower growth rate. Consequently, they have a favorable prognosis in terms of overall survival and relapse-free survival.

In contrast, **Luminal B** tumors, comprising 20% of all breast tumors, display a more aggressive phenotype and have a poorer prognosis compared to Luminal A tumors. Despite a high expression of ER, luminal B tumors show lower PR expression. The luminal signature expression is lower than in Luminal A tumors. However, they exhibit higher expression of proliferative genes and, occasionally, they can show HER2 expression.

**HER2-enriched** subtype depicts 15% of breast cancer cases and is characterized by the amplification of *ERBB2* gene, which codifies the HER2 protein, The overexpression of HER2 and the kinases of HER2 pathway and the absence of ER/PR. These tumors are more aggressive than luminal, but patients can take advantage of anti-HER2 therapies. It is important to remark that not all clinically $HER2^+$ tumors become part of the HER2-enriched subtype, due to the fact that a subset of Luminal B cancers also express HER2 as previously mentioned.

**Basal-like** tumors, which include approximately 15% of breast cancer patients, are characterized by the absence of ER, PR and HER2 expression. They present gene expression characteristic of normal mammary basal/myoepithelial cells, overexpression of cell proliferative-related genes, such as *EGFR*, and are frequently associated with germline *BRCA1* gene mutations. Due to these facts, they are the most aggressive tumors with poorer prognosis, and they are capable to invade surrounding tissue and form distant metastases easily.

**Normal-like** subtype was initially identified through the gene expression of normal breast epithelium with minimal changes. Subsequent observations revealed that this phenomenon might be attributed to a contamination of normal cells and the low cellularity of tumors. Therefore, the categorization of Normal-like subtype is currently questioned.

Finally, **Claudin-low** tumors are uncommon, exhibiting no expression of ER, PR and HER2 protein and low expression of proliferative genes and adhesion proteins including different types of claudins and E-cadherin. Moreover, this subset is distinguished by the enrichment of mesenchymal and stem cell-like signature pattern and the increment of immune cell infiltration. In spite of the fact that they are extensively considered as a subtype, it was later studied and redefined as a breast cancer phenotype (Harbeck et al., 2019; Nolan, Lindeman, et al., 2023; Sarhangi et al., 2022; Smolarz et al., 2022; Tsang & Tse, 2019; X. Xu et al., 2020; Zubair et al., 2021).

Along with this classification, breast cancer is typically and clinically stratified in a surrogate five groups based on immunohistologic and molecular features. These are **Luminal A-like**, **Luminal B-like HER2-**, **Luminal B-like HER2+**, **HER2-enriched (non-luminal)** and **Triple-negative breast cancer** (Fig. 6). This categorization also helps clinicians to establish which patients can benefit from different treatments (Harbeck et al., 2019; Łukasiewicz et al., 2021).

## 2.3. Triple-negative breast cancer (TNBC)

TNBC is not considered strictly the same as basal-like tumors although sometimes the terms are used interchangeably. Basal-like is the intrinsic molecular subtype which is included among TNBC surrogate subtype. Into the TNBC surrogate subtype is included the claudin low and normal-like subsets, apart from the basal-like subtype (Bou Zerdan et al., 2022). TNBC is the most common

subtype that arises in women premenopausal below 40 years, being highly aggressive and metastatic. Therefore, a clear effort to improve not only the classification but the subsequent establishment of the treatment has been made over the years. Hence, using gene expression profiling by tumor patient microarrays, seven clusters of TNBC emerged: **basal-like 1** (BL1), **basal-like 2** (BL2), **immunomodulatory** (IM), **mesenchymal** (M), **mesenchymal stem-like** (MSL)**, luminal androgen receptor** (LAR) and **unspecified group** (UNS) (Garrido-Castro et al., 2019; Lu et al., 2023a; Zubair et al., 2021). This classification was made according to different molecular alteration patterns, including transcriptome, somatic mutations and copy-number variations, as well as different expression of genes implicated in several pathways, in order to take them as models for therapy in the clinics (Garrido-Castro et al., 2019; Lu et al., 2023a).

BL1 and BL2 subsets exhibit transcriptionally increased levels of DNA damage-response genes. Additionally, BL1 contains elevated nuclear levels of Ki-67 and alterations in cell-cycle regulation proteins, whereas in BL2, the growth factor signaling and metabolic pathways are aberrantly more activated. Higher expression of genes related to antigen processing and presentation, as well as increment in immune-related cell and cytokine signaling pathways are characteristic of the IM subgroup. There is a similar protein expression pattern of cell migration, differentiation and epithelial-to-mesenchymal transition (EMT) among M and MSL subtypes. Nevertheless, MSL differs from M subset due to the higher expression of angiogenesis and stem-related genes and low expression of cell proliferation genes. Finally, LAR is negative for ER expression but shows expression profiles of mRNA and proteins of androgen hormone pathway. In addition, LAR category displays gene expression from luminal intrinsic subtypes (Garrido-Castro et al., 2019; Lu et al., 2023a; Tsang & Tse, 2019).

These subgroups were lately analyzed again, finding out that there were infiltrating lymphocytes and tumor-related mesenchymal cells in tumor samples of IM and MSL subtypes, redefining the categorization only in BL1, BL2, M and LAR. Apart from this classification, there is another which use a whole-genome sequencing analyzing the gene-expression profiling of tumor samples, determining 4 different clusters: LAR, mesenchymal (MES), basal-like immune suppressed (BLIS) and basal-like immune activated (BLIA).

All these attempts to define molecularly TNBC and the lack of a universally accepted gold-standard diagnosis on an international level clearly highlight the significant heterogeneity within this breast cancer subtype (Garrido-Castro et al., 2019; Lu et al., 2023a; Tsang & Tse, 2019).

## 2.4. Treatment for TNBC

The primary approach to remove TNBC is typically through **surgery** whenever feasible. Nevertheless, as it has tendency of recurrence and metastasis, patients have limited treatment alternatives, with **chemotherapy** being the predominant choice due to its favourable response. Regarding the different subtypes of TNBC, different combination can be given to achieve the best treatment outcome. There are different drugs commonly utilized in chemotherapy including taxanes (such as paclitaxel), anthracyclines (like adriamycin), cyclophosphamides, and cisplatin. TNBC patients in early stages can benefit from **neoadjuvant treatment**, showing a better prognosis in comparison with other breast

cancer subtypes helping later treatment decisions. Adjuvant chemotherapy is employed to eliminate any potential residual cancer lesions. However, as previously mentioned, there is still a risk of relapse and the emergence of metastases following this treatment.

Concerning the huge heterogeneity of TNBC, **targeted therapy is still challenging but can be beneficial in almost 70% of TNBC patients who do not respond to chemotherapy**. Some examples of them are PARP inhibitors which affect tumors with BRCA1/2 mutations or antiandrogens which are used in LAR subtypes specifically. In comparison with other breast cancer subtypes, some TNBCs exhibit an increase in infiltrating lymphocytes, together with a rise in PD-L1 expression, leading to a genomic instability. Therefore, they are treated with immunotherapy, using especially **PD-1 and PD-L1 inhibitors**, such as pembrolizumab, which is used against PD-1.

In spite of all the efforts, there are a percentage of TNBC patients who suffer recurrence and metastasis, making them a focal group for ongoing research into novel therapeutic approaches (Bou Zerdan et al., 2022; Lu et al., 2023a; Łukasiewicz et al., 2021).

# 3. Metastasis, plasticity and stemness

## 3.1. Metastasis process

The metastatic process is the combination of dynamic mechanisms used by tumor cells to spread from the primary tumor towards distal organs, forming secondary tumors called **metastasis**. Over 90% of cancer-related deaths are due to these secondary tumors (Mittal, 2017). The metastatic cascade involves several cellular pathways which enable tumor cells to invade the stroma and reach blood vessels. Cells are able to survive in suspension in systemic bloodstream, evade immune detection and modulate the microenvironment to create a niche for their colonization. Hence, in general terms, 3 distinct phases that overlap in time have been defined: **dissemination, dormancy** and **colonization**. time (Fig. 7) (Gerstberger et al., 2023; Suhail et al., 2019).

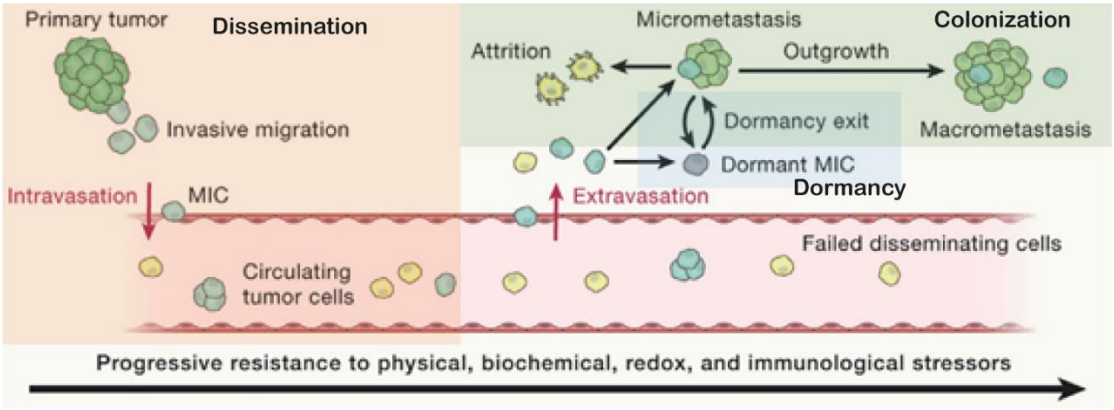



**Figure 7: Schematic representation of the three steps carried out in the metastatic process.** The first stage includes migration, invasion, survival in circulatory system and extravasation (in orange). The second implies dormancy in the target organ (in blue) and the last involves colonization, forming micro- and macrometastasis (in green). MIC: metastasis-initiating cells (Adapted from Gerstberger et al., 2023).

### 3.1.1. Dissemination step

For the dissemination step (Fig. 7), malignant cells modify the basement membrane, acquiring an invasive phenotype deepening into tissues until they reach the circulatory system (Gerstberger et al., 2023). Therefore, cancer epithelial cells suffer a switch becoming more mesenchymal, in a process named **epithelial-to-mesenchymal transition (EMT)**. This conversion leads to the loss of polarity and cell-cell junctions, enabling cancer cells to adopt migratory and stem-like capacities (Castaneda et al., 2022). The EMT process is governed by several transcription factors in breast cancer like Snail, Slug and Twist (J. Xu et al., 2024). This migratory phenotype is not only dependent on changes within cancer cells themselves. **Tumor microenvironment (TME)**, such as fibroblast, endothelial cells or immune cells, as well as the **extracellular matrix** itself, promote the transition into the process (Massagué & Ganesh, 2021; Suhail et al., 2019).

Additionally, the primary tumor sends **angiogenic signals** to stimulate the growth of new blood vessels, known as neovasculature, to obtain nutrients and oxygen. They are hyperplastic, disrupted and leaky new-formed vessels which facilitate cells to be released from the tumor into bloodstream. Specific well-studied proteins, including Integrins and Notch, help tumor cells intravasate into the circulatory system (Liang et al., 2020).

Despite most tumor cells die from stress or are eliminated by the immune system upon reaching the vessel lumen, a few manage to survive these stressors. This survival is often facilitated by the EMT process, which involves cytoskeletal reorganization, or by immune evasion through soluble factors and platelet protection surrounding the tumor cells. These surviving cells are known as circulating tumor cells (CTCs) or disseminated tumor cells (DTCs).

Once in the circulatory system, CTCs can get stuck in capillaries. Some organs are prone to present metastasis, such as liver or bone which have sinusoid capillaries. Moreover, CTCs attach to endothelial cells through adhesion molecules, facilitating the motility among endothelial junctions towards the tissue's depth, which is a process called paracellular migration (Castaneda et al., 2022; Massagué & Ganesh, 2021). Once cells arrive, the tumor cells finally reverse the mesenchymal phenotype through **the mesenchymal-epithelial transition (MET)**, the opposite process of EMT, to become more epithelial for establishing into the target organ (Huang et al., 2022; M. Park et al., 2022).

### 3.1.2. Dormancy

In the dormancy step (Fig. 7), not all the extravasated CTC might form detectable metastasis. Most of them enter a cell-cycle arrest and quiescent state, becoming less proliferative and resistant to antimitotic drugs. Thus, patients whose primary tumors were resected might relapse after months, or even years. The little number of experimental models makes even harder the extensive research of this field (Gerstberger et al., 2023; Massagué & Ganesh, 2021; Suhail et al., 2019).

According to literature there are three types of dormancy states. One of them is **cellular dormancy** which is depicted as the lack of both growth and apoptotic signals. In the case of the **angiogenic dormancy**, there is a deficiency in angiogenic cues due to the fact that dormant cells could not

induce the vessel formation in the metastatic niche, avoiding the arrival of respiratory and nutritious substances and the spreading of the metastatic cells. This stressful situation also provokes the induction of the latency state. Finally, in the **immunologic dormancy**, dormant mass cells may reach an equilibrium among the suppression of the metastasis outgrowth by immune reactivity and the evasion of the innate and adaptative immunity by the quiescent cells (Castaneda et al., 2022; Fares et al., 2020; Massagué & Ganesh, 2021).

Hence, dormancy phenotype is not only due to an intrinsic phenomenon in dormant cancer cells. The epigenetic regulation, together with an acquisition of adaptative metabolism, among other cellular paths facilitate the dormant cell phenotype but they are not the only essential factors. The paracrine signals from the TME contribute to the latency condition.

At the end, dormant cells must *awake* to form the secondary tumor in the metastatic niche. They have the reversible capacity to reactivate cell cycle, becoming an aggressive cell type. Little is known about the factors that initiate this process. However, aging is one of the intrinsic characteristics of patients that can ignite it. Additionally, inflammation, as well as immune surveillance removal, could be triggers for the exit of latency state (Gerstberger et al., 2023; Massagué & Ganesh, 2021; Suhail et al., 2019).

Therefore, genetics, the TME and timing are crucial factors for dormancy (Suhail et al., 2019).

#### 3.1.3. Colonization

Finally, upon dormant tumor cells emergence from their latency, colonization step initiates (Fig. 7). The restart of growth cycle and thus, the formation of macrometastasis in their niche, is the consequence of their concrete and complex tissue microenvironment interaction, as well as with the physical organ-specific barriers.

Therefore, it has been postulated the **"seed and soil" assumption** which explains that tumor cells spread to all points of the body (seed) but only those organs which accomplish the suitable conditions (soil) might harbour them (Castaneda et al., 2022; Liang et al., 2020; Mi Young Kim, 2021). Thus, the primary tumor releases soluble signals to support the immune escape and proliferative growth in metastatic niche, such as TGF-β or VEGF. Exosomes are particles secreted by primary tumor cells which are supposed to assist tumor cells to establish them in a specific organ, priming the resident cells to create an inflammatory environment and metabolic active niche. However, the role of target organ's stroma is also crucial for this process and the drug resistance (Castaneda et al., 2022; Massagué & Ganesh, 2021; Suhail et al., 2019), which is reprogrammed into a permissive microenvironment niche (Nolan, Kang, et al., 2023). In breast cancer the bi-directional interchange of signals is evident, resulting in a tendency of tumor cell to metastasize to particular organs, which is called organotropism (J. Xu et al., 2024).

#### 3.1.4. Organotropism of breast cancer

Breast cancer cells preferentially metastasize to four different organs which are bones, liver, brain and lung, and each of them have different context and molecules for homing tumor cells (Fig. 8) (Liang et al., 2020).

The **bone** is the primary location for breast cancer metastasis, affecting around 70% of patients with metastatic disease. There are two types of metastatic bone lesions: osteoblastic and osteolytic, being this last the predominant form of secondary tumors. In this context, bone stroma plays an essential role forming tumor cell niche, being critical osteoblasts, osteoclasts and immune cells. The colonization of CTCs in the bone niche is heavily supported by cell adhesion molecules which allow the harbour of cells to extracellular matrix (J. Xu et al., 2024).

After bone metastasis, **lung metastasis** is the second most common form of breast secondary tumors, being TNBC one of the most susceptible subtypes to develop it. They account for 15-20% of breast cancer patients, having poor survival rates and reaching 37% in 3-year overall survival (OS). Lung metastasis shows an intricated connection among tumor cells and lung-stromal cells. Upon their arrival, breast CTCs influence lung cells, especially immune cells, fibroblasts, and alveolar cells, to create a suitable niche for them. Particularly immune cells, macrophages and neutrophils evoke a pro-tumoral environment to enable metastasis progression (Nolan, Kang, et al., 2023; R. Wang et al., 2019; J. Xu et al., 2024) .

Following lung, there is the liver, which is notably lethal, as the 3-year overall survival rate does not reach 40%, while for bone metastases, it is over 50%. The tendency for breast cancer cells to colonize this organ is likely attributed to two factors: the particular tissue vascular structure, featured by fenestrated endothelium without subendothelial basement membrane, and the crosstalk of tumor cells with local resident cells, such as stellate, Kupffer and sinusoidal endothelial cells which is essential for the survival and the correct growth in the niche. In addition, influenced by stromal cells, resident immune cells might potentiate the homing of CTCs in this organ (Nolan, Kang, et al., 2023; R. Wang et al., 2019; J. Xu et al., 2024).

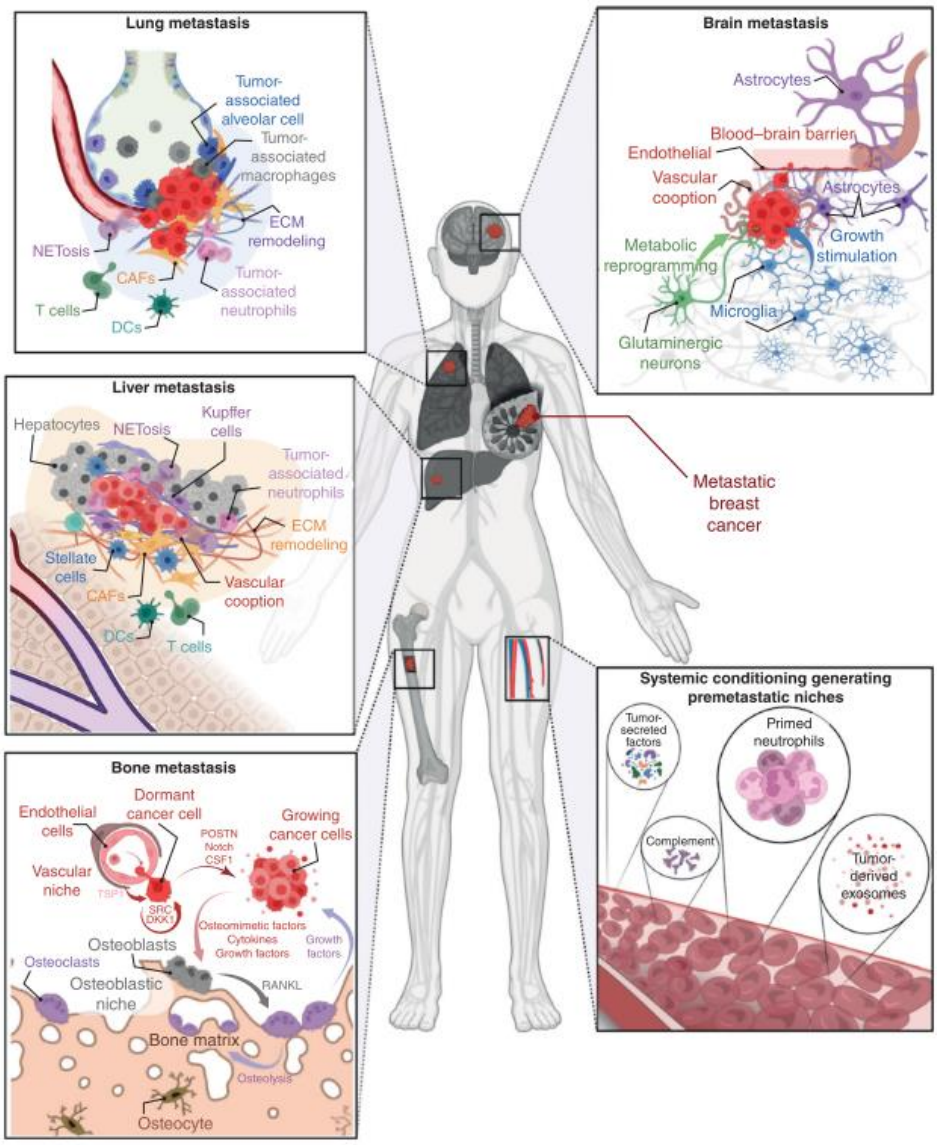



**Figure 8: Summary of all the target organs in BC.** It is also exhibited the mechanisms controlling the organotropism, observing a crosstalk among resident-niche cells and tumor cells (Nolan, Kang, et al., 2023).

Finally, focusing on the **brain metastasis**, it is a severe form of metastasis, with almost 15% of breast cancer patients developing it. The survival outcome is poor with a probability of 20% of one-year survival. HER2-positive breast cancer is the most frequent subtype which present brain metastasis. First, breast CTCs must go across the blood-brain barrier (BBB), which is a semi-permeable membrane that protects the brain. Tumor cells acquire the ability to permeabilize the BBB though specific cytokines. Furthermore, the concrete microenvironment, characterised by specialized resident cells as astrocytes and microglia, enhances the progression and the immune surveillance scape (Liang et al., 2020; Mi Young Kim, 2021; J. Xu et al., 2024).

## 3.2. Phenotypic plasticity

To survive through metastatic process, breast cancer cells need to adapt to different contexts during all the steps to detach from the primary tumor, survive in circulation, extravasate to distant organ, overcome the immune system and grow properly in the target organ. Thus, these cells must switch their phenotype to enable their survival in each situation. As such, **phenotypic plasticity** has been depicted as the capacity of cells to dynamically fluctuate their cellular state by changing their gene and protein expression patterns. This switch can result in differentiation, dedifferentiation or transdifferentiation and can be triggered by cell-intrinsic factors o external stimuli, mainly by TME. Additionally, plasticity is involved in specific processes, playing critical role in them and permitting breast cancer progression, and contributing to the intratumoral heterogeneity of these tumors (Fares et al., 2020; Jehanno et al., 2022; Massagué & Ganesh, 2021).

### 3.2.1. Plasticity in EMT

As explained in previous sections, EMT program is activated in the metastatic process to enable cells spreading from the primary tumor to bloodstream and, eventually, to distant organs. In these organs, the reverse program MET is triggered to establish tumor cells in the metastatic niche. Nonetheless, it is not a binary state, only epithelial or mesenchymal, but there are also transient phenotypes, defined as hybrid EMT, partial EMT or epithelial-mesenchymal plasticity (EMP). In the EMP, tumor cells can change through the EMT spectrum, adapting them to the microenvironment and the specific context of the patients (Castaneda et al., 2022; Kvokačková et al., 2021; Lüönd et al., 2021).

Those cells in EMP co-exhibit both epithelial and mesenchymal markers at the same time, such as EpCAM or E-cadherins to have adhesion properties, but weaker, as well as vimentin, which facilitates migrative capacities. Remarkably, tumor cells in this hybrid state are reported to be the most aggressive, with highest tumor initiation and metastatic potential, as well as enhance resistance to therapy. Furthermore, cells within EMP are expected to impact the surrounding microenvironment to become increasingly immunosuppressive. Consequently, patients exhibiting these cellular characteristics tend to experience poor outcomes (Castaneda et al., 2022; Jehanno et al., 2022).

### 3.2.2. Plasticity in dormancy

Dormant cells must return to a proliferative condition for the metastasis outgrowth. This procedure is in a constant dynamic flux, moving from one state to the other. The fluctuation between both phenotypes (dormant and cycling) is depending on the external influences (such as tobacco exposure), as well as stromal microenvironment signalling (Dalla et al., 2023; Jehanno et al., 2022).

In breast cancer, microenvironment of primary tumor site may send signals to target organ to stimulate the dormancy onset in there. Moreover, organ-site microenvironment promotes the dormancy state or the exit from it, using signals, such as hypoxia or inflammation, or other components of the stroma, as ECM (Dalla et al., 2023).

### 3.2.3. Metabolic plasticity

In order to survive to changes, CTCs adapt their metabolism to each particular context (in intra- and extravasation, in blood circulation and in invasion). Hence, the metabolic plasticity is characterised by the processing of a concrete metabolite in different pathways to assist tumor cells depending on the context requirements. The main metabolic pathways that are reprogrammed involve the pyruvate, glutamine and fatty acids metabolism (Jehanno et al., 2022).

There are several studies in disagreements of which type of pyruvate metabolism (glycolytic or oxidative phosphorylation) is preferred for breast plastic cancer cells. Likely, due to their high plasticity, cell fluctuate according to the specific context. Furthermore, depending on the particular organ of metastasis, CTCs enhance different metabolic pathways. For example, liver metastatic breast cancer cells tend to potentiate the glycolytic pathway, while reduce the glutamine pathways and oxidative phosphorylation. Nevertheless, brain metastatic breast cancer cells do not consume glucose but do enhance the glutamine metabolism (Gandhi & Das, 2019; Walsh et al., 2019).

### 3.2.4. Plasticity in stemness

The concept of "stemness" is referred as the ability of cells to differentiate to any cellular type into the tissue, as well as their self-renewal capacities. Thus, the **cancer stem cells (CSCs)** or **tumor-initiating cells (TICs)** are described as those cells with tumor-initiating and self-renewal competences, recapitulating a stem-like or a dedifferentiated state. The origin of CSCs and, hence, tumor initiation is illustrated by three models: in first place, the clonal evolution or stochastic model, by which cells acquire several mutations or epigenetic patterns that give advantage to the malignant over the normal cells, forming a tumor. In the deterministic model or CSC model, normal stem cells are transformed by somatic mutations, leading to the formation of cancer cells enable self-renewal and differentiation abilities. This creates a hierarchy of offspring cells committed to tumorigenic and/or metastatic capabilities. Nonetheless, the last model unified both which is the reversible plasticity model, stating that tumor cells move through different cell phenotypes, mainly between CSC and non-CSC state. Consequently, the stemness status can exhibit plasticity, fluctuating between these states. All these models may explain the existence of tumor heterogeneity (Castaneda et al., 2022; Warrier et al., 2023).

Stemness is strongly linked to the EMT program, although they are not the same concept. The EMT program activates transcription machinery that not only modifies some characteristics of tumor cells but also increases their stem-like expression patterns. Subsequently, these tumor cells, apart from being more metastatic, would be more stem, with enhanced invasive potential. Thus, the transcription factors that govern EMT are crucial regulators of stemness (Huang et al., 2022; Jehanno et al., 2022).

## 3.3. Stemness and its role in breast cancer

The stemness has been a wide field of research in last years, in particular in breast cancer. This section explains the features of breast CSC (BCSC) or breast tumor-initiating cells (BTIC), their

biomarkers and cellular pathways involved, as well as different techniques to isolate them. Additionally, it is also depicted the role of stem cells in breast cancer and response to current therapies.

### 3.3.1. BCSC characteristics, biomarkers and isolation methods

BCSCs possess self-renewal abilities and exhibit stem-like properties, mirroring the expression patterns of normal mammary stem cells . Moreover, they are able to be more resistant to drug treatment, as well as demonstrating the ability to initiate tumor formation from a limited number of cells. Finally, they are able to survive upon loss of anchorage.

BCSCs are characterised by the expression of cell surface markers CD44, EpCAM, CD133, MUC1 or CK5, the low expression of CD24, among others specific markers, although the most frequently used are the positivity for CD44 ($CD44^{+}$) and low presence or absence of CD24 ($CD24^{-/low}$). In addition, BTIC population can be also detected by its increased activity of aldehyde dehydrogenase ($ALDH^{+}$) 1 protein. These stem markers are not mutually exclusive. As such, different subpopulations of BCSC can be defined. For instance, the ALDH+ BCSC subpopulation is defined as an epithelial-like stem cells, whereas the $CD44^{+}/CD24^{-/low}$ displays a mesenchymal-like phenotype. Nonetheless, a small subset of BCSC expresses both markers concurrently, showing a plasticity between epithelial and mesenchymal-like states. This pattern is more frequent in TNBC patients and enhances the tumor-initiating and metastatic potential of BCSC, resulting in a poorer outcome in patients. (D. Kong et al., 2020; Taurin & Alkhalifa, 2020).

Besides, BCSC show high expression of ATP-binding cassette (ABC) transporters enabling the release of cytotoxic drugs from the cytoplasm to the extracellular space. This feature makes them resistant to chemotherapy, leading to recurrence and relapse after long periods of time. The dormancy state also has an important role in this trait. BCSCs have a quiescent and immunoevasive phenotype, as dormant tumor cells have, indicating that an overlap in the characteristics of both populations may exist (Chiotaki et al., 2016).

Regarding the fact that breast cancer cells are stem-like, it might be obvious that they regain the expression of pluripotency transcription factors (TF), such as SOX2, SOX9 or SLUG. In particular, NANOG, OCT4 and SOX2 TFs have been found increased in BCSC of TNBC samples, maintaining the self-renewal capacity of cells. Additionally, as it is pointed in the previous section, EMT program differs from the stemness, but the activation of EMT-TFs induces the rise of the BCSC. Examples of these EMT-TFs are ZEB1, SNAIL, SLUG or TWIST. Nevertheless, both EMT and stemness are plastic processes. Hence, their link could be the EMP, which is likely the key player for tumor initiation, invasion and relapse related to EMT and stemness (Celià-Terrassa, 2018; Fultang et al., 2021; D. Kong et al., 2020; L. Zhang et al., 2023).

In order to isolate or identify BCSC, several approaches can be performed. Mainly, it is broadly used the fluorescent-activation cell sorting (FACS) or magnetic cell sorting (MACS) detecting the cell surface markers CD44 or EpCAM and the absence of CD24. The assessment of ALDH activity through ALDEFLUOR or ALDERED assay is another technique to detect BCSC. It is based on the ALDH activity, obtaining a bodipy-fluorescent product which is detected by FACS. The spheroid formation assay

enables the ability of BCSC to form tumorsphere, being one of the most used approach to assess the stemness of cells *in vitro* and enabling the obtention of cells with stem capacity in a functional way. Finally, a combination of these techniques can be also carried out to deepen in the different BCSC populations (Ray & Mukherjee, 2024).

### 3.3.2. Pathways involved in BCSC and their role in breast cancer

In mammary stem cells, there are specific cellular pathways which regulate their self-renewal, thereby they are altered in BTIC, some of them are described below.

Wnt/β-Catenin signalling support the renewal and proliferation of BCSC, showing a stabilisation in 50% of breast cancer patients. The overactivation of Wnt/β-Catenin pathways support the mammosphere formation and the chemoresistance.

Notch signalling is involved, not only in the renewal, determining the stem cell fate, but also in the resistance of treatment. This pathway is upregulated in BCSC during hypoxia.

Finally, Hedgehod pathway control cell fate in physiologic condition. However, alterations in this pathway lead to increase in the relapse probability of patients, potentiating BCSC presence (Ibragimova et al., 2022; Ray & Mukherjee, 2024; Zeng et al., 2021).

BCSC self-renewal and differentiation properties are essential for tumor initiation and recurrence, thus maintaining themselves in a stem-like state as well as committing their offspring to differentiated phenotypes. Furthermore, BTIC are able to mimic **vasculogenesis** through transdifferentiation. Hence, they differentiate into an endothelial-like state in presence of cytokines, such as the vascular endothelial growth factor (VEGF). $CD133^+$ BCSC have been found to exhibit this capacity. Noticeably, during the **metastatic process**, the majority of cells involved exhibit characteristics of stem-like phenotype. As such, they do not only activate the EMT-TFs, but also show the anoikis-resistant ability, enabling their survival in circulation and generating tumorspheres. The anoikis-resistant BCSCs display a higher expression of Wnt/β-Catenin pathway. Additionally, TGF-β signalling contribute both to the metastasis and BCSC maintenance. Ultimately, BTIC are **radio- and chemoresistant**, due to the increased expression of the cellular drug transporters, as well as high levels of $ALDH1^+$ protein which metabolizes anticancer treatments. Moreover, the potentiation of DNA damage repair systems, as well as antiapoptotic and antioxidant signalling, contribute to the BCSC resistance (Butti et al., 2019; L. Zhang et al., 2023).

### 3.3.3. Therapies to specifically target BCSC

As previously mentioned, BCSCs have mechanisms to avoid drug effects of chemotherapy through drug pumps or due to the intrinsic low-cycling capacity. Thus, different therapeutic approaches must be developed to target them and avoid the relapse.

Self-renewal pathways are altered in BCSC, being promising targets to develop treatments. Some **Notch-blocking drugs**, such as γ-secretase inhibitors, as well as **inhibitors of Hedgehog signalling**, such as the SMO inhibitor sonidegib or cyclopamine, have been tested in clinical trials for breast cancer in combination with chemotherapy to treat TNBC. Hedgehog pathway drugs can target the

ligands, receptors or the downstream TFs. **Inhibitors of Wnt family**, as vantictumab or sulforaphane (Bai et al., 2018), diminished the BTIC *in vitro* and *in vivo*, allowing their test in clinical trials. Nevertheless, more research is currently running to reduce side effects (D. Kong et al., 2020; Zeng et al., 2021).

There are several druggable signalling pathways implicated in the maintenance of EMP and/or BCSC conditions. **PI3K/AKT/mTOR, MAPK and STAT3 signalling pathways** are dysregulated in BCSC, promoting their cancer cell plasticity and the acquisition of stem properties. Therefore, several drugs and natural compounds have been found to target BCSC. For instance, inhibitors of AKT function can block the EMP and the formation of tumorspheres. Additionally, pharmacological inhibition of STAT3, as seen with napabucasin, reduces EMT and the stem-like population in breast cancer. Similarly, MAPK blockade, through agents like selumetinib that inhibit MEK, can decrease the $CD44^{+}/CD24^{-/low}$ population. However, compounds targeting these pathways require further investigation, as clinical trials have yet to demonstrate significant improvements (Hua et al., 2022; D. Kong et al., 2020; Zeng et al., 2021).

Other strategies can be used. High expression of DNA damage response proteins in BCSC can be targeted with their inhibitors, such as **PARP inhibitors olaparib and talazoparib**, specially for TNBC. They have been proved in clinical trials observing a benefit in locally advanced and metastatic breast cancers (Paul et al., 2022). Furthermore, targeting drug pumps could be a great opportunity to eliminate BCSC. As an example, Dofequidar, an ABCB1 inhibitor, was tested in combination with different chemotherapy drugs, observing a sensibilization of patients to chemotherapy (Saeki et al., 2007). Besides, since TME plays a critical role in BCSCs, drugs targeting chemokines or TME signaling molecules, such as TGF-β, are promising therapeutic options. Furthermore, enhancing the immune system's response against BCSCs is an attractive strategy, as it has shown effectiveness against non-BCSCs and it has been observed that BCSC evade the immune reaction. Hence, specific **immune checkpoint inhibitors** could exhibit good outcomes for eliminating BCSC, as well as the development of vaccines targeting antigens associated with BCSC markers. For instance, CD73 was proved to correlate with BCSC population (Yu et al., 2017). Moreover, another example of immunotherapy is the chimeric antigen receptor T cell (CAR-T cell) therapy to treat TNBC, which possesses the largest BCSC population among subtypes. An example is the CAR-T cell therapy against EpCAM surface protein in *in vitro* and *in vivo* models using TNBC cells, showing great antitumor results (B. L. Zhang et al., 2019) is clinical trial in development (NCT02915445). However, the research of these treatments needs greater efforts to improve the outcome of breast cancer patients (Hua et al., 2022; D. Kong et al., 2020; Zeng et al., 2021).

All these treatments have limitations. First, some markers are common among normal and tumorigenic stem cells. Hence, they have fatal adverse effects in normal tissue with off-target toxicities. In addition, there is a heterogeneity among BCSC, making difficult to specifically eliminate all of them. Finally, lots of drugs have limited effects *in vivo,* less solubility and instability (Ali et al., 2024; Khan et al., 2021; L. Zhang et al., 2023). Altogether, new strategies should be found to improve the specific targeting of BCSC improving patient's outcome.

# 4. Post-translational modification: Protein Glycosylation

Protein glycosylation is the addition of carbohydrates molecules to the residues of certain proteins. It is the most complex post-translational modification being a multistep process. This complexity arises from the number of enzymes involved, the specific position where the glycans are attached in the proteins and the glycan structures assembled. Additionally, protein glycosylation is different based on cell type and cellular needs (Schjoldager et al., 2020). The glycome of the cell is defined as the group of monosaccharides and glycans (referred as polysaccharides or complex oligosaccharides (Eichler, 2019) in cells) which are found free or assembled in glycoconjugates. These glycoconjugates are glycoproteins, glycosylphosphatidylinositol (GPI)-anchored proteins, proteoglycans and glycosphingolipids (Fig. 9). All the glycoconjugates at the cell surface conform the glycocalyx, which contributes to the stability of the protein and help to the interaction with microenvironment components (Schjoldager et al., 2020; Spiro, 2002). In this section, the glycosylation the process will be outlined, including the different types of glycosylation. The functional roles of glycosylation will also be discussed, along with its involvement in cancer. Finally, we will explore any potential connections between glycosylation and stemness.

## 4.1. Types of glycosylation

Glycosylation can be classified based on the molecule to which glycans are attached. N-glycosylation involves the attachment of sugars to the nitrogen atom of asparagine residues (or less commonly, arginine) in proteins via an amide bond. In contrast, O-glycosylation occurs when sugars are primarily added to serine and threonine residues in the oxygen atom, and less frequently to hydroxylysine and tyrosine residues by glycosidic linkage. Other types of glycosylation include C-glycosylation (or C-mannosylation), where a mannose sugar is attached to a tryptophan residue through a carbon-carbon bond, and glypiation, where sugars are linked to a GPI anchor that connects to protein backbones (Fig.9) (Eichler, 2019; Moremen et al., 2012).

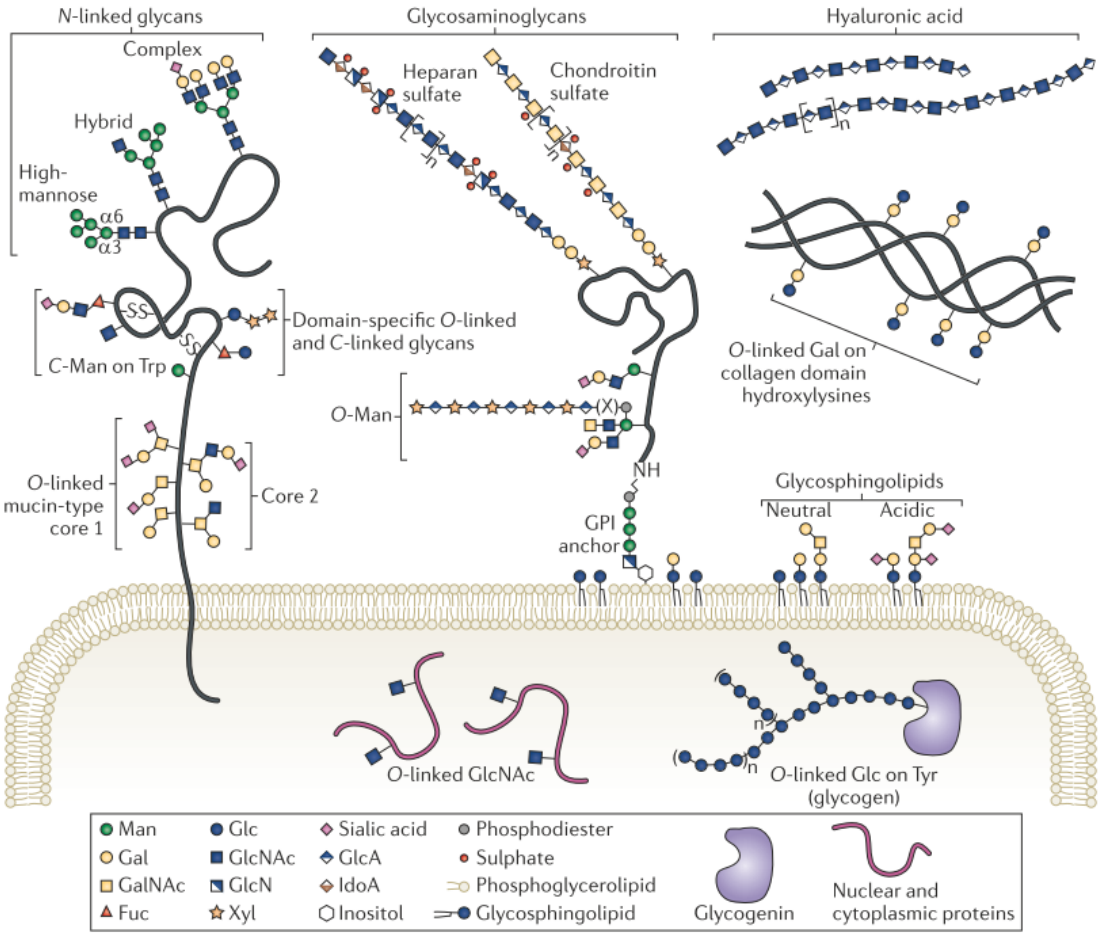



**Figure 9: Types of glycans and glycoconjugates found in humans.** Different glycosylation types are represented, including N-glycosylation and O-glycosylation. The formation of proteoglycans is also illustrated with the glycosaminoglycans attached to a polypeptide chain, as well as glycosphingolipids composition. There is little amount of proteins which undergo O-GlcNAcylation into the cytoplasm and nucleus (Moremen et al., 2012).

Moreover, various types of O-glycosylation are defined by the initial sugar added to the corresponding amino acids. In humans, the sugar more commonly attached is *N*-acetylgalactosamine (GalNAc), known as mucin-type O-glycan, and *N*-acetylglucosamine. O-glycosylation can also occur as O-fucosylation, O-xylosylation, O-mannosylation, or O-glucosylation (Dai et al., 2024; Reily et al., 2019). Finally, depending on the composition of glycans and their linkage, N-linked and O-linked glycans can be also classified through their composition and linkage position of branches (Bagdonaite et al., 2022).

## 4.2. Biosynthetic process

Unlike proteins synthesis, protein glycosylation takes place without a template. This implies that the cell glycome relies on the protein expression together with other factors such as the localisation of these proteins, the abundance of the enzyme substrates or the competition among different

acceptors and donors, resulting in various complexes glycan configurations in specific different glycosylation sites, commonly called **macroheterogeneity**, or in a concrete glycosylation site, referred to as **microheterogeneity.** (Bagdonaite et al., 2022; Moremen et al., 2012).

Protein glycosylation is a sequential intricate biosynthetic event mediated by over two hundred enzymes mainly by glycosyltransferases, but also by glycosidases. The process initiates in the secretory pathway, that is, the endoplasmic reticulum (Er) and Golgi and proceed along the secretory system performing the core extension, elongation, branching and capping. N-glycosylation and glypiation begin in Er, whereas O-glycosylation starts in the Golgi apparatus (Fig. 10). Nevertheless, one type of O-glycosylation, the O-GlcNAcylation takes place in cytoplasm and nucleus. (Reily et al., 2019; Schjoldager et al., 2020).

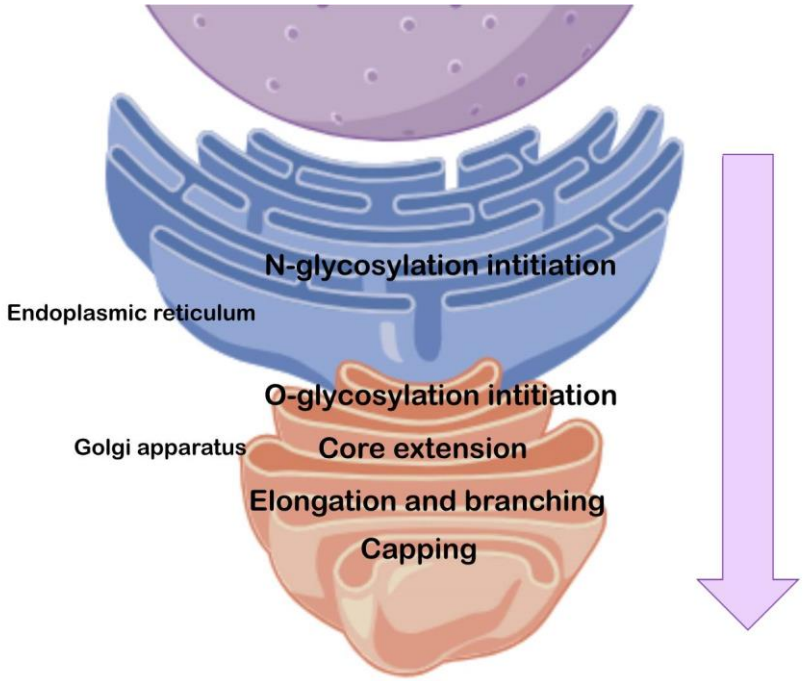



**Figure 10: Biological pathway of protein glycosylation.** N-glycosylation begins in the Er and progresses into the Golgi apparatus, where O-glycosylation also initiates and continues. Image created with BioRender.

Lastly, glycosyltransferases utilise ten different monosaccharides derived from highly energetic nucleotide sugar or dolichol-linked donors to form all the glycosydic structures (Fig 11). The ten monosaccharides are glucose, galactose, *N*-acetylglucosamine, *N*-acetylgalactosamine, fucose, glucuronic acid, mannose, sialic acid, xylose and ribose and are transferred using UDP, GDP, CMP and CDP to form activated sugar intermediates. Additionally, dolichol-linked donors are crucial for N-glycosylation. As such, all the enzymes implicated in the formation of dolichol are important for the protein glycosylation (Bangarh et al., 2023; Reily et al., 2019; Schjoldager et al., 2020).

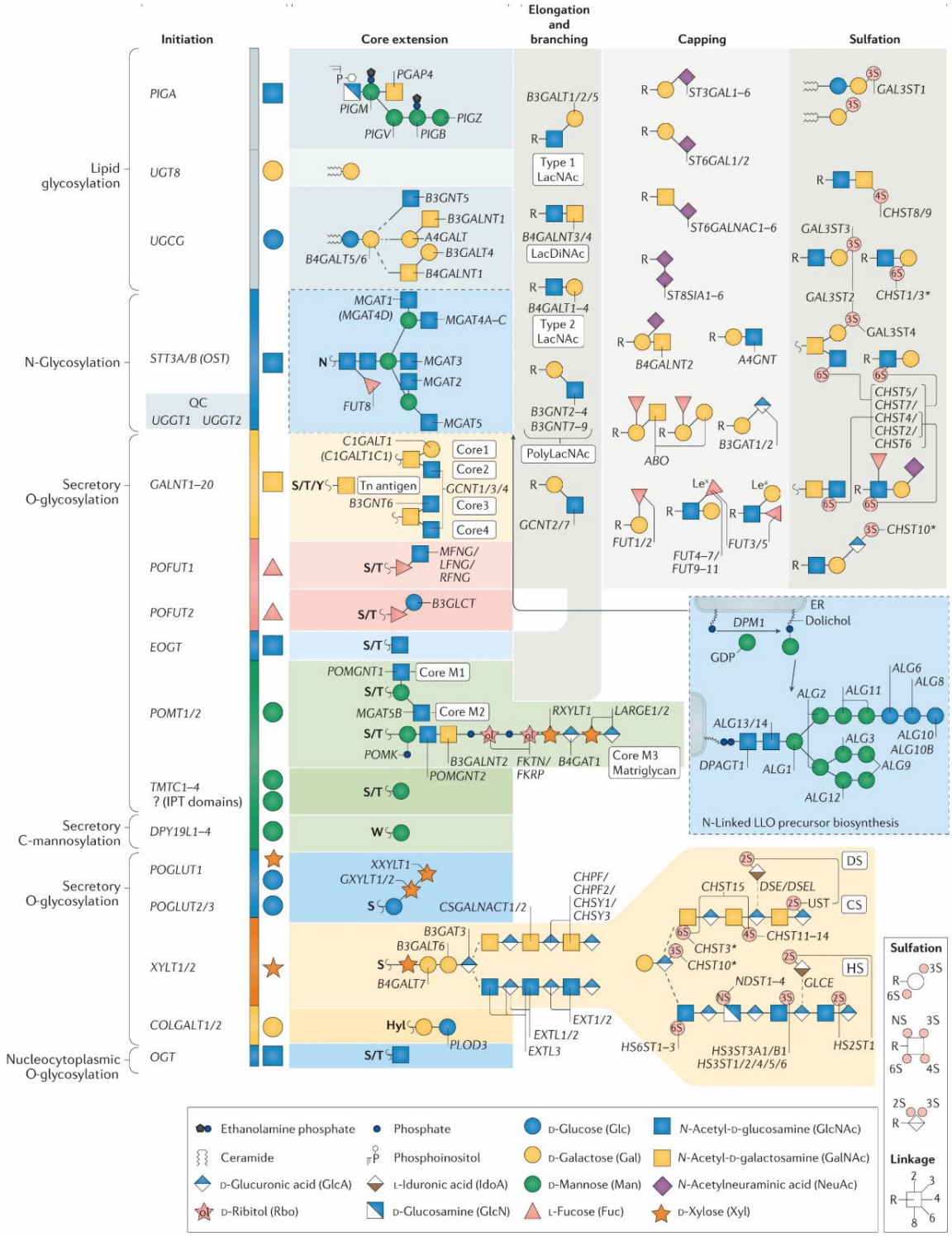



**Figure 11: All human glycosylation pathways and enzymes involved in.** The lipid glycosylation is also represented as indicated. The colored background illustrated the colors of the initial monosaccharide Adapted from Schjoldager et al., 2020.

## 4.3. Physiological roles of protein glycosylation

Protein glycosylation is involved in several physiological process. It mainly has a direct role in protein folding and the quality control. Furthermore, the addition of sugars stabilizes proteins, facilitates their transport across various cellular compartments, and is crucial for their proper function, even switching their function in some cases (Varki, 2017). Altered protein glycosylation can modify signaling pathways, this is the case of Notch, EGFR or IGFR protein, which require specific glycosylation for their activity (Dai et al., 2024; Varki, 2017).

The role of glycosylation modulating immune system is well established, involving the detection of different pathogens, as well as immune cell differentiation. The self-recognition and, thus, the regulatory modulation of immune system is also controlled by glycosylation (Reily et al., 2019; J. Y. Zhou et al., 2018). Furthermore, glycosylation is essential for cell-cell contacts, such as those between leukocytes and endothelial cells, as well as for cell-matrix interactions. Lastly, the glycocalyx provides protection to cells against molecular or cellular damage (Eichler, 2019; Varki, 2017; J. Y. Zhou et al., 2018).

## 4.4. Glycosylation and cancer

As previously explained, protein glycosylation plays several critical roles in maintaining cellular homeostasis and ensuring the proper functioning of biological systems. Consequently, aberrant glycosylation can disrupt this physiological balance, leading to abnormal cellular behaviour and the acquisition of tumorigenic properties. Two different processes are associated with tumor-related glycan alterations: incomplete synthesis, which is related to the early stages of cancer, and neo-synthesis, which is associated to the late stages of the disease. These alterations can be attributed to one or several factors: (I) shift in the Glycosyl transferases or glycosidase expression, (II) changes in the localisation of these enzymes within the secretory pathways, (III) alterations in chaperone activity that modify the peptide backbone conformation and, consequently, the formation of the glycan chain, and (IV) fluctuations in metabolism and the availability of acceptor and donor substrates (Mereiter et al., 2019; Pinho & Reis, 2015).

The most common glycosidic alterations observed in cancer include truncated O-glycans, such as short-chain mucin-type glycans, increased sialylation and fucosylation, and increased branched N-glycans (A. F. Costa et al., 2020; Vajaria & Patel, 2017).

Aberrant protein glycosylation is involved in several cancer cell functions, contributing to the cancer progression through the (de)regulation of cell-cell and cell-matrix interactions, cellular and proliferative signaling, and paracrine and distal communications among other processes. These abilities are involved in hallmarks of cancer, such as immune avoidance, metastasis, promotion of inflammation, deregulation of metabolism, or sustained proliferation, reviewed by Hanahan (Fig. 12). Hence, altered glycosylation state is postulated to be another hallmark in cancer properties (Hanahan, 2022; Munkley et al., 2016; Peixoto et al., 2019).

As examples, the role of ST6GAL1 in breast cancer has been studied, associating its increased activity with a reduced cell-cell contact and thus, increased metastatic potential. In addition, overexpression

of MGAT5 enzyme increases the N-glycan branching, mislocalising E-cadherin and promoting invasiveness. Heparan sulfate proteoglycan is also observed as a crucial mediator for the angiogenesis in ovarian cancer and hepatic cancer (Munkley et al., 2016; Peixoto et al., 2019). As such, glycosylation has a great impact in cancer cell biology.

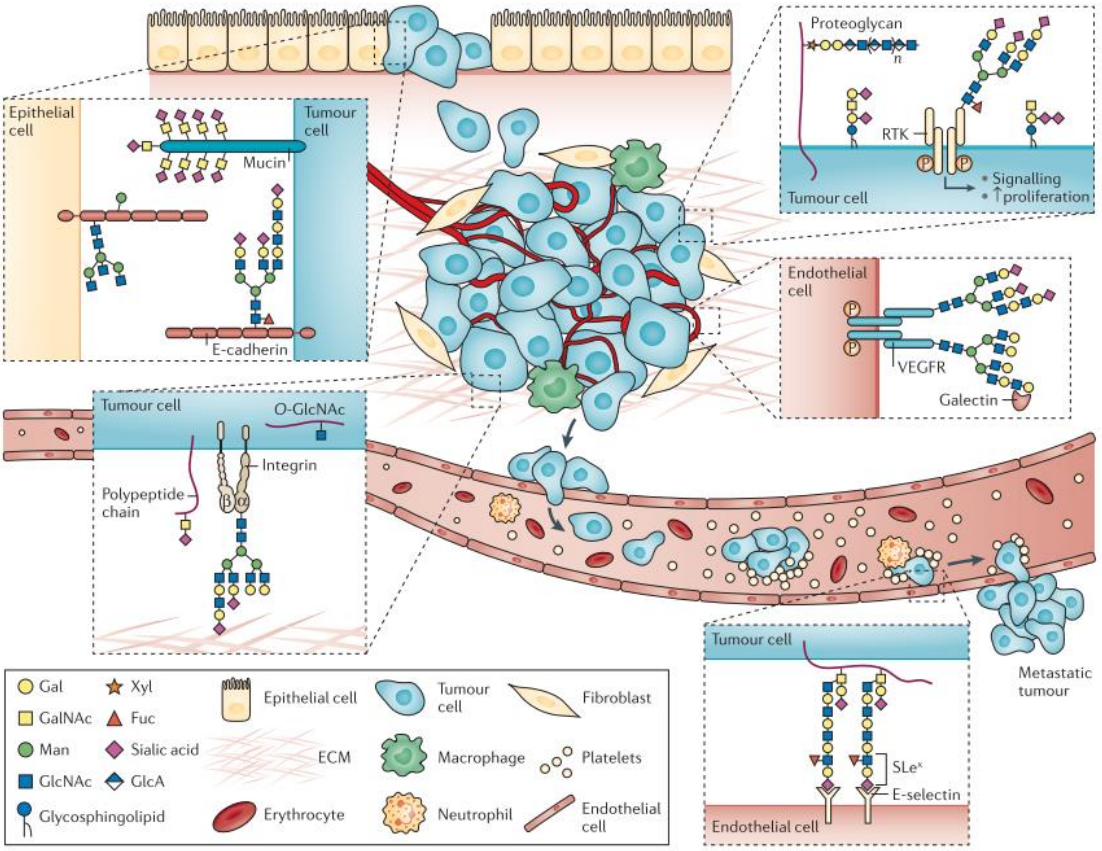



**Figure 12: Role of glycans in different cancer processes.** Aberrant protein glycosylation is implied in all the hallmarks of cancer. This image represents the role of it in cell-cell adhesion, tumor progression, angiogenesis, metastasis or immune evasion. Adapted from Pinho & Reis, 2015.

In the specific context of breast cancer, the role of aberrant protein glycosylation has been thoroughly investigated. There are some examples that represent this fact.

Over 90% of breast cancer cases exhibit expression of mucin-type glycans related to O-glycosylation. Aberrant O-glycosylation as a result of MUC1 protein increased expression enhances the cells' metastatic potential.

Concerning sialylation, sialic acid is commonly linked to N- and O-glycoproteins in breast cancer. These glycosidic pattern can be identified as “self” by the immune system and, thereby contribute to the evasion of the immune surveillance. The rise of sialyltransferase expression, such as ST3GAL1, has been examined in breast cancer tissue in comparison with the normal one, and it has been demonstrated that promotes tumorigenic capacity when it is overexpressed (Picco et al., 2010).

Finally, fucosylation is also detected as altered in breast cancer, resulting in an upregulation of fucosyltransferases which leads the cells to be more metastatic and participate in the EMT program (D. Liang et al., 2023; Peric et al., 2022; Scott & Drake, 2019).

## 4.5. Therapeutic approaches against aberrant protein glycosylation

The specific altered glycosidic patterns in cancer, also named as tumor-associated carbohydrate antigens (TACAs), are used as tumor biomarkers as well as therapeutic targets for cancer treatment (Matsumoto & Ju, 2023).

Several approaches have been already employed to target TACAs, such as the vaccinated-based immunity, the development of unconjugated and conjugated monoclonal antibodies or the application of CAR-T cells anti-TACAs (Berois et al., 2022).

The vaccines to target TACAs have emerged as a promising strategy, due to the specificity of TACAs in tumor cells. However, carbohydrates are challenging to be used as therapeutic agents for three reasons: they are molecules with low immunogenic response, especially those that contains less sugars and glycolipids (Matsumoto & Ju, 2023), they do not generate a robust T cell reaction and complex glycans are synthesis of glycans is challenging (Smith & Bertozzi, 2021). To overcome these facts, TACAs have been conjugated with carrier immunogenic proteins or peptides, such as diphtheria toxoid, oligonucleotides or nanoparticles that can be processed and trigger the long-term immunogenicity (Anderluh et al., 2022). For example, a vaccine against sialylated Thomsen-noveau (sTn) TACA was developed conjugated to an immunostimulant to raise a strong IgG antibody response (X. G. Yin et al., 2017).

Antibodies against glycans can be designed. (Mastrangeli et al., 2018; Smith & Bertozzi, 2021). They might be unconjugated to mediate an antibody-dependent cell-mediated cytotoxicity, or conjugated with toxins or drugs which exert the cytolytic activity (Berois et al., 2022). As an example, an antibody (trastuzumab) against $HER2^+$ cells conjugated with neuraminidase was also created to modify their glycocalyx and effectively increase the cytotoxicity of the antibody (Xiao et al., 2016), being this antibody under a phase I/II clinical trial (NCT05259696).

As antibodies can be developed against specific glycan structures, CAR-T cell therapy can also be engineered to produce T cells with artificially expressed glycan-based antigen receptors with the intracellular domains that facilitate the T-cell cytotoxicity (Anderluh et al., 2021). Additionally, CAR-T cell therapy do not require MHC presentation for T-cell activation (Singh & Mcguirk, 2020). One example of CAR-T cell therapy against the glycol form $Tn^+$ in MUC1 was engineered, demonstrating cytotoxicity and controlling tumor growth (Posey et al., 2016).

These examples are chosen from a wide range of studies and clinical trials that support the idea that aberrant protein glycosylation can be targeted with various treatment strategies (Table I).

*Table I: Different glyco-based therapies against cancer pathology.*

| Type of therapy | Name | Effect | Type of cancers | Currently clinical trials |
|---|---|---|---|---|
| **Vaccine-based therapy** | OBI 822(adagloxad simolenin)/OBI -821 | Reaction against Globo H glycosphingolipid antigen | TNBC | NCT03562637 |
| | STn/KLH (THERATOPE®) | Detection of sialyl-Tn antigen | Metastatic breast cancer | NCT00003638 |
| | Polysialic acid-KLH + QS21 (adjuvant) | Detection of polysialylation | Small-cell lung cancer | NCT00004249 |
| | Bivalent vaccine (GD2L/GD3L) + OPT-821 (adjuvant) | Reaction against gangliosides GD2 and GD3 | Neuroblastoma | NCT00911560 |
| **Antibodies against TACA** | Anti-TA-MUC1 (PankoMab-GEX™) | Antibody detecting a Tn epitope of MUC1 | Solid tumors | NCT01222624 |
| | hu mAb-5B1 (MVT-5873) | Antibody detecting Lewis A antigen | Tumors that express CA19-9 | NCT03801915 |
| | BMS-986012 | Antibody detecting fucosyl-monosialoganglioside (fucosyl-GM1) | Relapsed small-cell lung carcinoma | NCT02247349 |
| **CAR-T cell therapy** | huMNC2-CAR44 | Reactivity against extracellular domain of MUC1 | Metastatic breast cancer | NCT04020575 |
| | CART-TnMUC1 | Reactivity against Tn epitope of MUC1 | TNBC, pancreatic ductal adenocarcinoma, ovarian cancer | NCT04025216 |

| | | | | |
|---|---|---|---|---|
| | GD2-CART01 | Reactivity against disialylgangliosi de GD2 | Neuroblastoma and GD2 positive solid tumors | NCT03373097 |

Another strategy to affect aberrant glycosylation is target specifically those proteins that generate the altered glycosylation in the certain glycosydic pathway (N- or O-glycosylation) or target the biosynthetic pathway that produce the substrate for the aberrant glycosylation (Almahayni et al., 2022; Vasconcelos-dos-Santos et al., 2015).

## 4.6. Role of protein glycosylation in cell plasticity and stemness?

More than a decade ago, aberrant glycosylation was considered to be important for the acquisition of stem-like properties. The attempt to find the differences among normal stem cells and cancer stem-like cells has been and is still arduous, since they are nearly identical transcriptionally. Nevertheless, since most stem cell markers are glycoproteins or glycolipids and glycosylation is altered in tumor biology, aberrant glycosylation may play a key role in cancer stem cells (Karsten & Goletz, 2013).

Actually, the onco-fetal glycan structure Thomsen-Friedenreich (or CD176) was identified as a tumor-specific marker and was detected specifically in stem cell markers. Additionally, Oct4 and Sox2 are O-GlcNAc-modified in mouse embryonic stem cells and disappear when they differentiate (Barkeer et al., 2018; Karsten & Goletz, 2013). Glycosylation of CSC markers such as CD44, CD133 or EpCAM and its implication in some signalling pathways is well studied (Barkeer et al., 2018; Khan & Cabral, 2021).

Nonetheless, very few studies connect the aberrant glycosylation with the acquisition of stemness. As examples, Notch signalling and its regulation has been associated with stemness and metastasis in renal cancer. Also, the role of MGAT5 glycosyltransferase in controlling the Wnt/β-catenin signalling pathway has been determined (Barkeer et al., 2018; Khan & Cabral, 2021). It has been also investigated the highly expression of *N*-acetylgalactosamine and *N*-acetylglucosamine in the $CD133^{+}$ glioblastoma CSCs, permitting their identification through these glycosidic patterns (Tucker-Burden et al., 2012).

The relationship between aberrant protein glycosylation and the maintenance of stem-like properties in breast cancer is not yet fully understood. Therefore, it is necessary to investigate whether aberrant glycosylation promotes a stem-like phenotype and its persistence in breast cancer, as well as to identify the key factors involved. Understanding these mechanisms could lead to the development of new strategies and the identification of druggable targets to specifically treat and eliminate cells with stem-like properties.

# 5. CRISPR screenings: the edition for the elucidation

In the past years, CRISPR (clustered regulatory interspaced short palindromic repeats) editing has reached a relevant position as the gold standard approach for research in cancer field to be easier, more specific and cheaper than other methods. Besides, the use of loss- or gain-of-function libraries for gene or drug screening is extensively applied in the cancer disease investigation. In this section, CRISPR screenings will be explained from their bases, the different types of CRISPR systems and the strategies to apply in one of the most used CRISPR Cas nuclease, the knock out (KO).

## 5.1. CRISPR/Cas system

CRISPR nucleases are endonucleases used by prokaryotic organisms as an adaptative immune system to defend themselves from viral or exogenous DNA that can damage the host cell. The first discovered and the most frequently used is the Cas9 from *Streptococcus Pyogenes* (Katti et al., 2022; Sanjana, 2017).

The mechanism of CRISPR/Cas9 system is based on the endonuclease Cas9, which is guided by a 20-base pair (bp) single-guide RNA (sgRNA) to recognize and target specific DNA regions via complementary base pairing. The endonuclease comprises two lobes: one binds to the sgRNA while the other has nuclease activity. Once sgRNA binds to the target DNA sequence, Cas9 induces a double-strand break (DBS) into DNA. This cleavage requires the detection of the protospacer adjacent motifs (PAM), a specific sequence of 3 bp (5'-NGG-3'). Without the PAM sequence, Cas9 will not induce a DSB, even in the presence of the complementary sgRNA. Upon the DBS near the PAM sequence, cells induce DNA damage repair through non-homologous end joining (NHEJ), leading to insertions or deletions that cause a functional KO, or homologous directed repair (HDR), which uses a DNA template to introduce a specific gene sequence by homologous recombination, and thus, to generate a knock in (KI) (Fig. 13) (Ding et al., 2023; S. W. Wang et al., 2022; Y. Zhu, 2022).

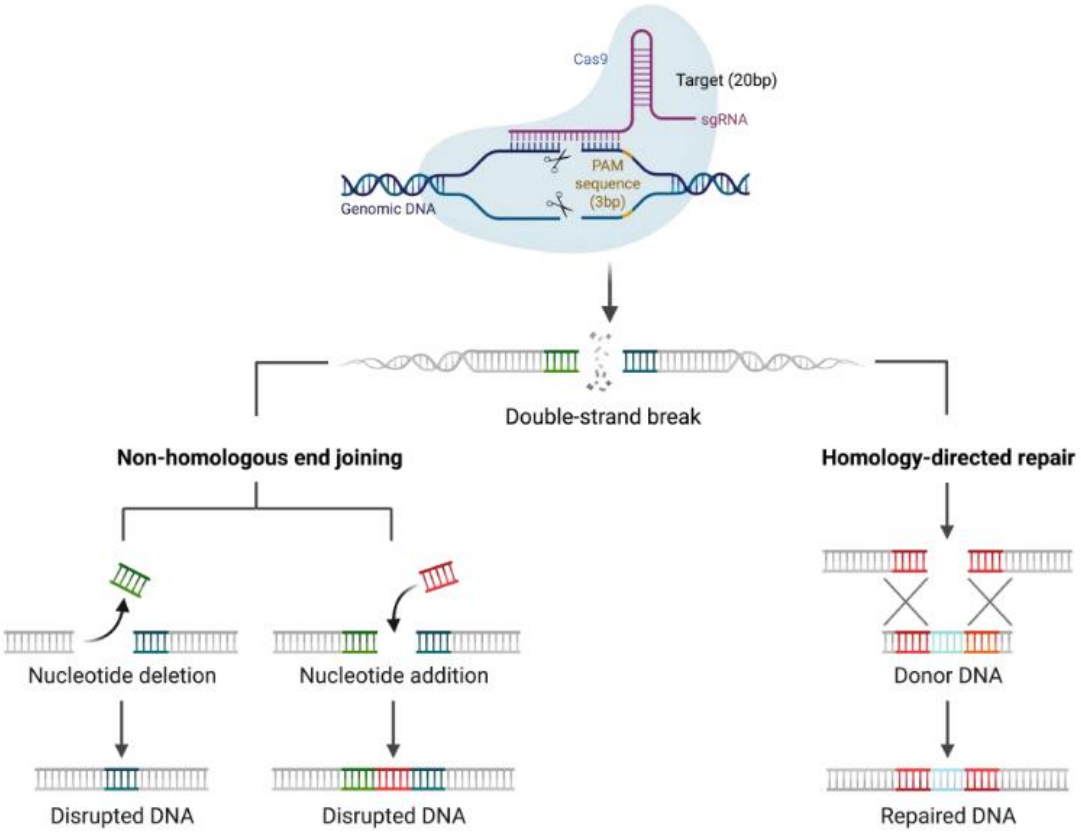



**Figure 13: Mechanism of CRISPR/Cas9 action.** Cas9 uses a single-guide RNA (sgRNA) to identify and bind to a specific DNA sequence, along with the presence of a PAM sequence, to induce a DNA break. To repair it, the NHEJ (left) and HDR (right) can be conducted to reach a KO or KI, respectively. Image from (Ding et al., 2023)

This system has evolved along the years, giving us powerful tools for gene editing, discussed in the following point.

Additionally to Cas9 nuclease, other nucleases from different species have been employed for gene editing. For instance, Cas12, which recognises different PAM sequences, requires a shorter sgRNA than Cas9 and induces staggered DBS that promotes the HDR repair system. Another example is Cas13 which target RNA instead of DNA (Hillary & Ceasar, 2023; Zhan et al., 2019).

## 5.2. Modified Cas nucleases

Cas9 can be engineered to improve the technique, its safety, or its purpose. Cas9 has been modified by inactivating one of the nuclease domains, generating a single-strand break (SSB). Hence, Cas9 turns as Cas9 nickases (Cas9n) to edit homozygous lethal genes. If they are fused with deaminases or DNA polymerase, they turn into base editors to transient modify a single nucleotide within the genome. These nucleases can be improved to prime editors, which apply a permanent base change through reverse transcription (Balon et al., 2022; Zhan et al., 2019; Y. Zhu, 2022).

When the Cas9 nuclease domain is altered to completely lack catalytic activity, it is named dead Cas9 (dCas9). This dCas9 can be fused with to various effector domains to regulate gene transcription.

When combined with activator transcriptional regulators, it is known as the CRISPR activation system (CRISPRa). Conversely, when fused with a repressor, it forms a CRISPR interference system (CRISPRi) to inhibit gene expression. Additionally, dCas9 can be fused with epigenetic modifiers to explore and investigate the epigenome (S. W. Wang et al., 2022; Zhan et al., 2019).

## 5.3. CRISPR libraries

CRISPR system has been broadly employed to reveal crucial genes in cancer or useful drugs for new therapies in large-scale genetic screens. CRISPR libraries have been a great tool for screenings, due to the fact that they can cover any genomic region, addressing various biological questions. A CRISPR library consist in a collection of sgRNAs targeting different genes. Libraries can be introduced into cells temporarily using adeno-associated viruses, or permanently using lentiviruses and retroviruses. Adeno-associated viruses target non-cycling cells, while lentiviruses and retroviruses target cycling cells. Other mechanisms can be applied such as lipid-based strategies (Balon et al., 2022; Joung et al., 2017; Shi et al., 2023; Zhan et al., 2019).

There are two main strategies to conduct a screening with CRISPR libraries: arrayed and pooled CRISPR screenings. In arrayed CRISPR screenings, each library component is screened separately, being easier for the researcher observe the resulting phenotype. High-content imaging is often used to measure the read-out, although arrayed screenings are time-consuming and more expensive. Conversely, in pooled CRISPR screens, all plasmids with each sgRNA are introduced in bulk into the cells, which are then perturbed by each sgRNA. This type of screening is less expensive, reduces the time required for its implementation and can be applicable at a large-scale, which might make it difficult to use FACs for its readout. (Joung et al., 2017; Shi et al., 2023).

## 5.4. Pooled CRISPR screenings

A common pipeline of a pooled CRISPR screenings is depicted in Fig. 14.

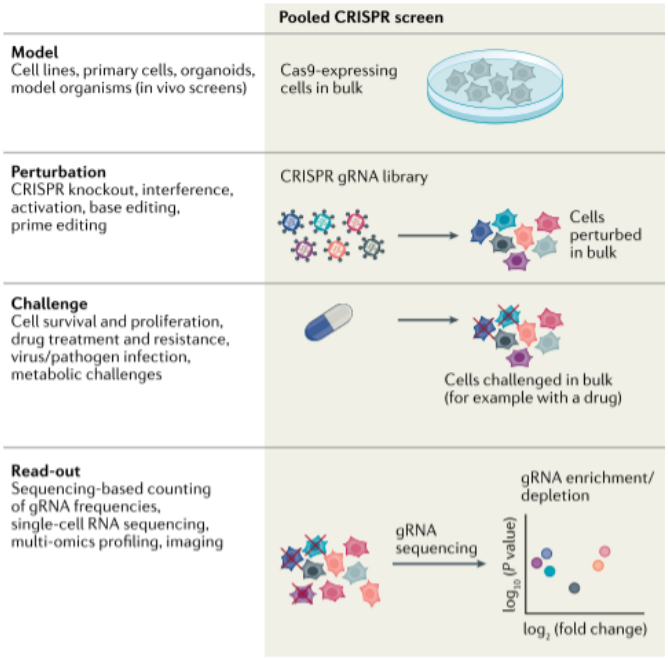



**Figure 14: Design to perform a pooled CRISPR screening.** There are four variables to consider. The model to use, which perturbation you are going to employ in your model, which will be the challenge after the perturbation and the specific read-out after phenotypic changes. Adapted from Bock et al., 2022

### 5.4.1. Models for CRISPR screenings

The chosen model is crucial for applying our pooled CRISPR library. The most commonly used model are cell lines, but primary cells and animal models for *in vivo* screenings are also employed. Their use benefits the screening by recreating the effects of the host microenvironment. However, depending on the cell type, they might not grow properly to conduct the screening, or it may be necessary to use immunocompromised animals to inject the cells. Other models as organoids are employed, which mimics the 3D structure of the specific organ. Nonetheless, the use of 3D matrices growth factors and induced-pluripotent stem cells or adult stem cells can be challenge for laboratories (Bock et al., 2022; Geurts & Clevers, 2023; Holen et al., 2017; Zhan et al., 2019).

### 5.4.2. Perturbations for CRISPR screenings

The type of perturbation that will alter our cells is crucial. Screenings can be classified according to the CRISPR system that is used. These include CRISPR KO, CRISPRi, CRISPRa, and base/prime editors, being CRISPR KO the most frequently used (Bock et al., 2022; Ding et al., 2023).

Pooled CRISPR screenings are often conducted using lentiviral vectors. Transduction is usually performed at a multiplicity of infection (MOI) smaller that 0.3 to ensure that only one sgRNA infects each cell. This fact allows the determination that the phenotypic effect will be caused by the particular perturbation (Sanjana, 2017; Zhou Y, 2022).

In this section, it is important to have into account the library type used. In the case of CRISPR KO screening, there are genome-wide libraries to cover all the genome, but different libraries can be

designed, targeting specific gene subsets. More than 4 sgRNA per target gene must be included to avoid the stochastic effect and perform a feasible screening, together with specific positive and negative controls. Furthermore, the library coverage, and consequently the amount of sgRNAs required, must be sufficient to maintain their representativity throughout the screening experiment, ensuring accurate representation within the cell population (Bock et al., 2022; Castells-Roca et al., 2021; Joung et al., 2017; Poirier, 2017).

### 5.4.3. Biological selective pressure in CRISPR screenings

After the perturbation, library-transduced cells will be selected by a biological pressure, competing for pool representativity at the final time point of the screening. Several approaches can be employed, such as cell viability, drug resistance or functional assays, as well as the selection using surface markers analysed by FACS (Castells-Roca et al., 2021; Joung et al., 2017; Katti et al., 2022; S. W. Wang et al., 2022).

Functional screenings can be categorized as **positive** or **negative**. In positive screenings, there is an enrichment of sgRNA at the final point, meaning that few cells survive to selective biological pressure, giving the mutations proliferative advantages over those which have not acquired the alteration. Therefore, the representation of selected sgRNAs increases significantly, enhancing their signals. For this reason, an initial coverage of 100-200 times is sufficient, making this approach ideal to elucidate genes related to drug or pathogen resistance. On the contrary, negative screens are focus on the depleted sgRNAs. In other words, most of the cells survive to the selective pressure and only those carrying the perturbation will be eliminated. Unlike positive screens, this type of experiments requires larger representativity, as the signal from depleted sgRNA is low. Negative screenings are widely used to discover essential genes (Bock et al., 2022; Castells-Roca et al., 2021; Joung et al., 2017; S. Sharma & Petsalaki, 2018).

The choice between positive and negative screens will depend on the type of perturbation selected and the biological question made.

### 5.4.4. Screening read-out

The sgRNA enrichment or depletion is determined in bulk using frequently next generation sequencing (NSG). Nevertheless, single-cell sequencing has expanded its influence in the CRISPR screening field, including spatial imaging. Notably, Perturb-seq, which integrates single-cell RNA-seq with CRISPR screening, allows researchers to investigate the effects of gene expression perturbation at the transcriptomic level (Bock et al., 2022; Meyers et al., 2023; Schraivogel et al., 2023).

Finally, a schematic process of pooled CRISPR screening is represented in Fig. 15.

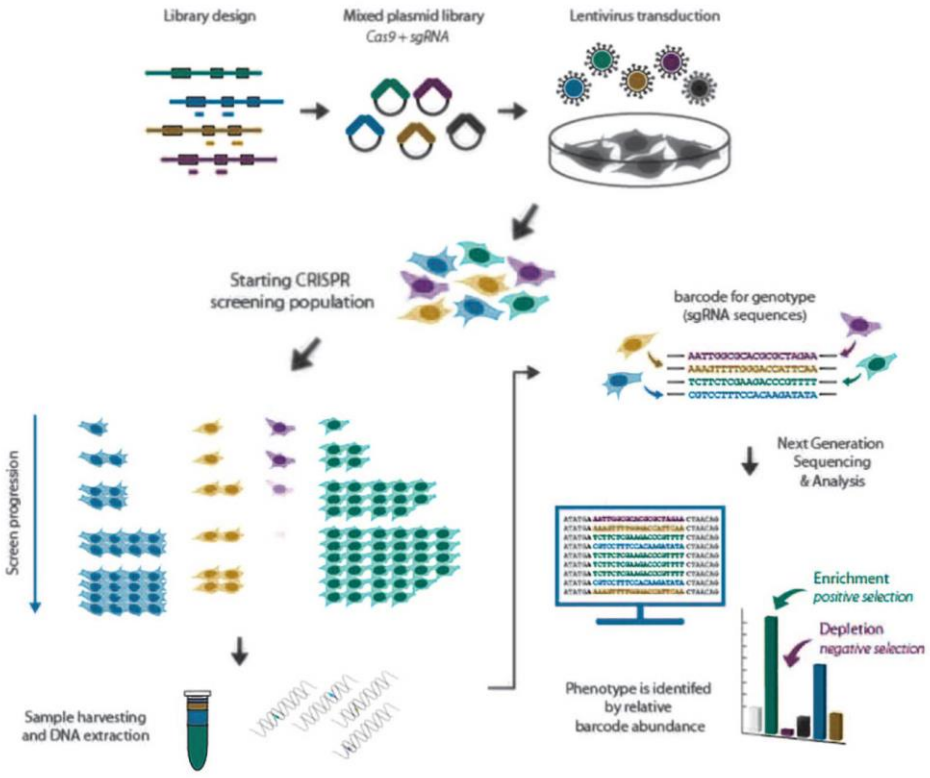



**Figure 15: Workflow of pooled CRISPR screening.** This workflow starts from the design of the library, the next cloning into the vector and the infection into target cells, the progression of the screening throughout the biological selection and the subsequent NGS and analysis. Adapted from le Sage et al., 2020**.**

### 5.4.5. Computational analysis of pooled CRISPR screening

Following NGS, bioinformatic analysis is necessary. This analysis relies on the screening results, typically the sgRNA counts obtained from amplicon sequencing. Generally, this involves a five-step pipeline.

Initially, **data processing** is necessary for the subsequent analysis. The reads are aligned against reference sgRNAs, resulting in matrices with counts of each experiment, condition and gene. Next, the **quality control** is crucial to rely on the analysis, based on the average reads or the percentage of missing sgRNAs. These metrics should be consistent across replicates and can be compared using visual tools like principal component analysis. Additionally, the non-essential genes can be evaluated to see their imperturbability in all the conditions and experiments.

Later, **gene ranking** is performed, reflecting the effects of perturbation and selective pressure. Depending on the type of selection, sgRNAs will be either enriched or depleted, and this is determined using statistical methods. Various software packages in programming languages like R or Python can be used to execute this pipeline, including tools such as MAGeCK (and its derivatives), CERES, or BAGEL.

The next step is the **hit analysis**. Once ranked, the most significant hits (either depleted or enriched) are assessed according to their relevance taking into account out biological question. Online resources can be used, as PubMed, Ensembl, The Human Protein Atlas or STRING. Finally, **visual interpretation** enables the ease in the result understanding, using volcano plots or sgRNA rank plots for representing all the analysis (Bock et al., 2022; X. Li et al., 2023; Zhao, Zhang, et al., 2022).

In summary, this thesis explores the use of a pooled CRISPR library, developed by our laboratory, to conduct negative screening for identifying genes that are essential in cell plasticity. Building on the premise that abnormal glycosylation might be pivotal in sustaining the stemness phenotype, this work hypothesizes that targeting these glycosylation processes could reveal key genetic dependencies in advanced breast cancer

.

# Hypothesis and objectives

# Hypothesis

Our hypothesis is that aberrant protein glycosylation plays a key role in the acquisition and maintenance of stem-like properties in tumor cells. Identifying glycosylation-related genes involved in stemness could enhance our understanding of how these cells retain their plasticity, driving tumor progression and metastasis. Targeted inhibition of such genes may reduce the stem-like characteristics of tumors, potentially leading to more effective and selective therapies that minimize side effects. This strategy could also support the discovery of new drugs or the repurposing of existing ones to specifically target this tumor cell subset.

# Objectives

The main objectives of this thesis are:

I. Establish the conditions to conduct CRISPR screenings across different breast tumor and non-transformed cell lines.

II. Identify protein glycosylation genes involved in the maintenance of breast cancer stem cells using the novel CRISPR library developed in our laboratory.

III. Validate the role of candidate genes in stemness through functional assays, assessing stemness and stem-related markers, including RNA expression and/or cell surface markers.

IV. Investigate the mechanism by which the glycosylation contributes to stemness, focusing on the selected targets.

# Materials and methods

# 1. Cell culture

## 1.1 Cell types and characteristics

Breast cancer cell lines have been used during this thesis to perform all the experiments. They have been selected according to their different features, detailed in the following table (Table II):

*Table II: Commercial cell lines used and their molecular features.*

| Type | Subtype | Cell line | Surface protein expression molecules | | |
|---|---|---|---|---|---|
| | | | HER2 | ER | PR |
| **Non-transformed** | Basal B | MCF10A | - | - | - |
| **Tumorigenic** | | MDA-MB-231 | - | - | - |
| | | HS578T | - | - | - |
| | Basal A | HCC70 | - | - | - |

All cell lines were purchased from American Type Culture Collection (ATCC) repository.

Not only breast cancer cell lines but also modified HEK293T cell line from embrionary kidney, namely Phoenix amphotropic cells, were used. These cells are engineered to express packaging and envelope proteins in a stable manner.

## 1.2 Maintenance and counting

To maintain the cell lines, they were grown in humidified atmosphere at 37ºC and 5% $CO_2$. In addition, they have been cultured in their specific media, as it is specified in the next table (Table III):

*Table III: Media used for each cell type.*

| Cell type | Base media | Supplementation |
|---|---|---|
| **MCF10A** | DMEM/F-12 | 5% Horse serum<br>1% Penicillin/Streptomicin<br>10 ng/mL hEGF<br>100 ng/mL Cholera toxin<br>10 µg/mL insulin<br>500 ng/mL Hydrocortisone |
| **MDA-MB-231**<br>**Phoenix (Φ)** | DMEM (high glucose) | Glutamax (4mM)<br>10% FBS<br>1% Penicillin/Streptomicin<br>1% Pyruvate sodium |
| **HS578T** | DMEM (high glucose) | Glutamax (4mM)<br>10% FBS<br>1% Penicillin/Streptomicin<br>1% Pyruvate sodium<br>10 µg/mL insulin |
| **HCC70** | RPMI 1640 | 10% FBS<br>1% Penicillin/Streptomicin |

After reaching subconfluence they were harvested by tryspinization using trypsin 0.05% EDTA (Gibco, Fisher Scientific, 11580626), previously washing them with PBS 1X. They were split at 70-80% confluence to avoid phenotypic, behavioural and metabolic alterations in the cells.

They were centrifuged for 5 minutes at 0.3 x g, at room temperature (RT). Afterwards, pellets were resuspended in a volume of 1 mL complete media, seeding them in the required dilution in order to culture them every 2-3 days.

Before each experiment, all cell lines were counted using the LUNA-II™ Cell Counter (Logos Biosystem, Aligned Genetics, Inc.), plating the specific number of cells stated in each experimental procedure.

All cells were analysed to check the absence of mycoplasma infection, a common affection of cell culture, carrying out monthly PCR with specific primers for Mycoplasma. Hence, all of them were mycoplasma-free cells.

### 1.3 Cryopreservation and thawing

Work with lower passage cells is the better way to carry out the experiments. For that, cells were maintained frozen in stock in -80ºC for short-term and in liquid nitrogen for long-term conservation.

Cell pellets were resuspended in 1mL of completed media supplemented with DMSO according to repository's instructions. Cryovials were placed in a cooler to gradually reduce the temperature when stored at -80ºC. One day later, the cryovials could be stored at -80ºC or at liquid nitrogen.

To facilitate the thawing of frozen cells while mitigating the toxic effects of DMSO, a rapid thawing method was employed. This involved the addition of warmed complete media to the frozen cell vial, aiming to dilute and remove DMSO after performing a centrifugation step. Once pelleted cells, they were plated in free-DMSO completed media to let them growth properly.

## 2. Engineering of cell lines by lentiviral transduction

Parental cell lines can be genetically modified to express or silence certain genes of our interest. The workflow for this process is detailed below, separating it by days for clarity (Fig. 16):

### 2.1 Transfection procedure

To initiate the process, Phoenix was utilized as our lentivirus packaging producer cells. To initiate transfection, they were plated at day 1 to reach 50-70% confluency within 24h. At day 2, transfection was performed using jetPEI® (Polyplus-transfection S.A, Illkirch, France) following the manufacturer instructions. Briefly, 2 eppendorfs were prepared. In the eppendorf A, the quantity in µg of the plasmid and helper plasmids psPAX2 (#12260, Addgene) and pMD2.G (#12259, Addgene) were added in a proportion 2:1:1 respectively, adding the NaCl solution given in the kit reaching a final volume of 250 µL. In the eppendorf B, JetPEI was mixed with NaCl solution, using a volume of JetPEI double that of the DNA amount used in Eppendorf A. To reach the final volume of 250 µL, NaCl solution was added. Thereafter, both eppendorfs were mixed in

a vortex for 20 seconds. Then, the content of eppendorf B was introduced in eppendorf A, vortexing the mixture again 20 seconds. The directionality of this step was compulsory. Finally, an incubation of this eppendorf at RT for 20 minutes was required to form the positively charged particles named complexes. Meantime, phoenix cells media was replaced for fresh medium. Once the incubation was completed, the content of the eppendorf was added to phoenix cells drop by drop through all the plate without prior mixing (to avoid the disruption of the complexes) and let incubate for 24h.

## 2.2 Transduction of cell lines

Change of media was performed 24 hours post-transfection with media of our target cells (day 3), and all virus media were collected 48 hours after transfection (day 4). Since the transduction efficiency had to be as high as possible, virus media collection was repeated 72 hours post-transfection to carry out a re-infection (day 5). Virus media from each day was first filtered with a 0.22 µm filters and concentrated using a 15 mL Amicon® (MilliporeSigma, Burlington, Merck Life Science, MA, USA, UFC910024) for 30 minutes, at 4ºC and 800 g. It is important that the lid of Amicon should not be fully closed to allow the virus media filtered. Once centrifuged, concentrated virus media could be used for infection or could be stored at -80ºC for long-term storage.

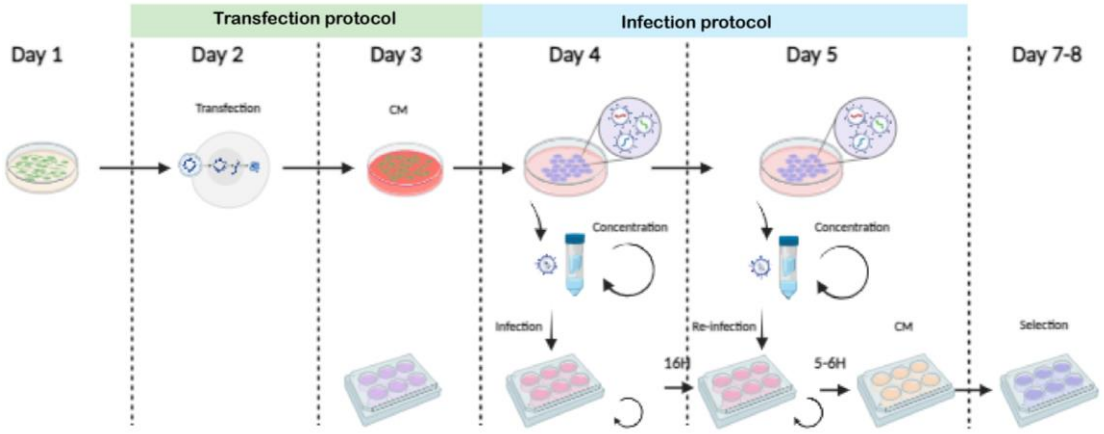



**Figure 16: Pipeline of virus production and transduction process.** The upper part concerns to virus media generation and the lower part regards the cell culturing and transduction. Upper: Phoenix cells were plated to reach a 50-70% confluency at day 2. On this day, transfection procedure was carried out, performing a change of media 24h later. At day 4, virus media was collected, filtrated, and concentrated. Fresh media was added again to collect virus media 72h posttransfection, repeating the same strategy as the previous day. Lower: target cells were plated in day 3, the next day, they were transduced for 16h as indicated. Subsequently, a reinfection was performed, changing media 5-6h later. After 48-72h, cells could be selected using antibiotic.

At the same time on day 3, target cells were plated in 6-well plate to achieve a 50% confluency the day of transduction. Cells were seeded for both a negative control (uninfected cells) and for subsequent infection. On day 4, cells were transduced adding completed fresh media, concentrated virus media and a polycation, either Polybrene (Millipore, Darmstadt, Germany, TR-1003-G) or DEAE-dextran (Sigma-Aldrich, 30461) to a final concentration optimised for each cell line. Additionally, to maximize the efficiency of infection, the 6-well plate was centrifuged at 400 g, RT for 90 minutes. Hence, cells were incubated overnight, performing a re-infection 16 hours later and repeating the protocol. After 5-6 hours of second infection, virus media was

replaced for fresh media, allowing the cells stabilize for 2-3 days. Following this, cells were maintained with fresh media containing antibiotic selection, until the uninfected cells died.

## 2.3 Plasmids for cell line genetic modification

Different vectors have been employed in this thesis for modify genetically our parental cells.

Firstly, miR.E plasmid was used for gene knocking-down. This system is a well-established method to silence our desired targets with more effectiveness and with less off-target effects. For this, the experimental miR-30 was used as a backbone to perform the cloning (Fellmann et al., 2013) (Fig. 17).

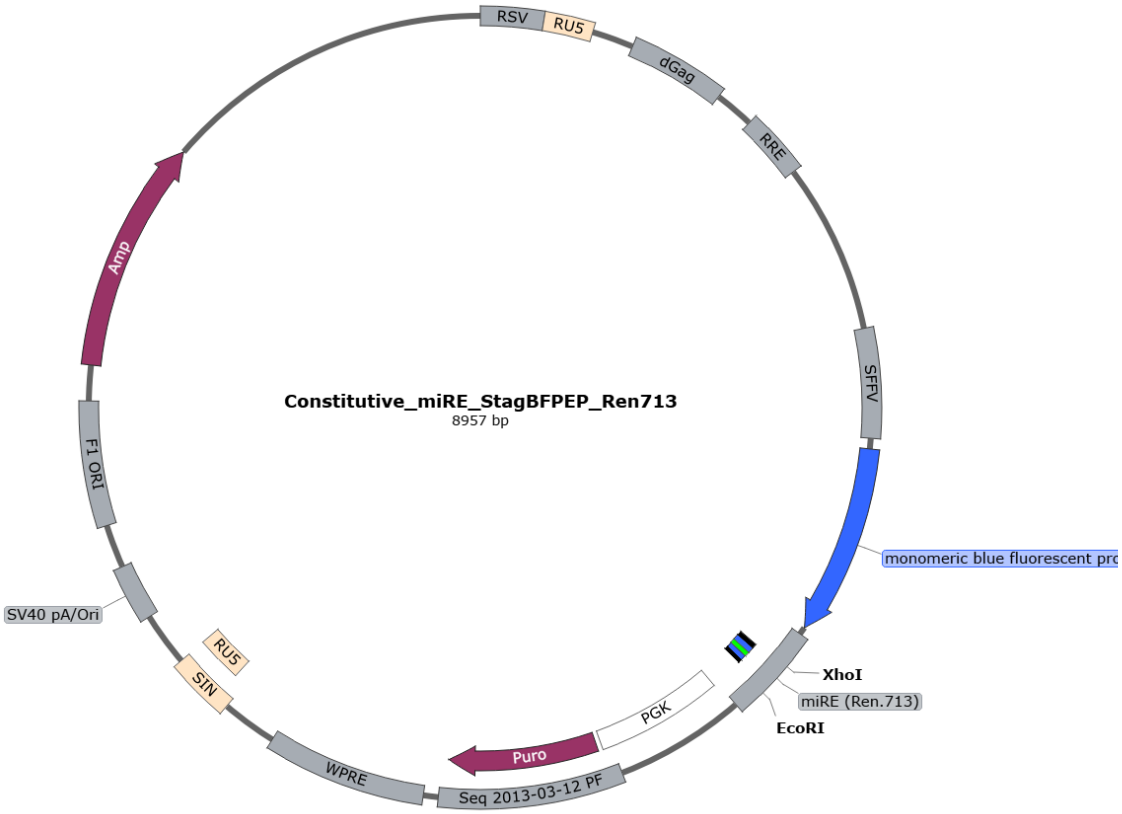



**Figure 17: Map of the miR.E plasmid.**

This system is based in a synthetic 96-mer template, consisting in a 21 base antisense sequence of our targeted genes designed using the splashRNA website (http://splashrna.mskcc.org/), and sequencing flanks containing EcoRI and XhoI recognition sequences. Three templates from three genes are described in Table IV.

*Table IV: All antisense sequence to knock-down our targets.*

| Gene template | Antisense sequence templates (from 5' to 3') |
|---|---|
| **ST3GAL1_1** | TGCTGTTGACAGTGAGCGCTCCCATTTTACTGATGAGAAATAGTGAAGCCACAGATGTAT<br>TTCTCATCAGTAAAATGGGAATGCCTACTGCCTCGGA |
| **ST3GAL1_2** | TGCTGTTGACAGTGAGCGCCCTGTCTCTACAAAAATAAAATAGTGAAGCCACAGATGTAT<br>TTTATTTTTGTAGAGACAGGATGCCTACTGCCTCGGA |

| | |
|---|---|
| **ST3GAL1_3** | TGCTGTTGACAGTGAGCGCGGGCTGGAAGAAAGTTCATAATAGTGAAGCCACAGATGTA TTATGAACTTTCTTCCAGCCCATGCCTACTGCCTCGGA |
| **EXT1_1** | TGCTGTTGACAGTGAGCGCCCTTACTACTATGCTAATTTATAGTGAAGCCACAGATGTATA AATTAGCATAGTAGTAAGGATGCCTACTGCCTCGGA |
| **EXT1_2** | TGCTGTTGACAGTGAGCGATAGGAATCATTTAATTTTTAATAGTGAAGCCACAGATGTATT AAAAATTAAATGATTCCTACTGCCTACTGCCTCGGA |
| **EXT1_3** | TGCTGTTGACAGTGAGCGCAGGTTGTGTACAGTTTAATTATAGTGAAGCCACAGATGTAT AATTAAACTGTACACAACCTATGCCTACTGCCTCGGA |
| **DHDDS_1** | TGCTGTTGACAGTGAGCGAACCTCCTTTCCTGATAATGAATAGTGAAGCCACAGATGTAT TCATTATCAGGAAAGGAGGTGTGCCTACTGCCTCGGA |
| **DHDDS_2** | TGCTGTTGACAGTGAGCGAGATCTGCTAGTAAATAACTAATAGTGAAGCCACAGATGTA TTAGTTATTTACTAGCAGATCCTGCCTACTGCCTCGGA |
| **DHDDS_3** | TGCTGTTGACAGTGAGCGACACCTGGGATTTGCTATTGAATAGTGAAGCCACAGATGTA TTCAATAGCAAATCCCAGGTGGTGCCTACTGCCTCGGA |

Briefly, a PCR was carried out using the template and forward and reverse primers with XhoI and EcoRI overhangs, respectively (Table V). Three reactions per template were run. Primer sequences and components for the PCR using High Fidelity (Roche Diagnostics, Mannheim, Germany, 04738292001) are detailed in the following tables (Table VI).

*Table V: Primers overhangs with the restriction enzymes sites.*

| Components | Sequences (from 5' to 3') |
|---|---|
| **miR.E-XhoI-Fw [10µM]** | TGAACTCGAGAAGGTATATTGCTGTTGACAGTGAGCG |
| **miR.E-EcoRI-Rev [10µM]** | TCTCGAATTCTAGCCCCTTGAAGTCCGAGGCAGTAGGC |

*Table VI: Components for PCR (left) and thermocycler protocol to carry out the PCR (right).*

| Components | µL (1X) |
|---|---|
| **$H_2O$** | 12.44 |
| **PCR 10X buffer** | 2 |
| **dNTPs [10mM]** | 0.4 |
| **FW primer [10µM]** | 2 |
| **RV primer [10µM]** | 2 |
| **Template [2µM]** | 1 |
| **High fidelity polymerase** | 0.16 |
| **Final volume** | 20 |

| Steps | Temperature (ºC) | Time |
|---|---|---|
| **1** | 95 | 5’ |
| **2** | 95 | 30’’ |
| **3** | 58 | 30’’ |
| **4 (GOTO 2→33X)** | 72 | 30’’ |
| **5** | 72 | 7’ |
| **6** | 4 | ∞ |

The PCR product was run in a 2% agarose gel and purified using the PCR clean up (Macherey-Nagel GmbH&Co, Düren, Germany, 740609.50). Finally, the template was digested with both EcoRI and XhoI restriction enzymes (New England BioLabs Inc., R3101S and R0146S) in a thermocycler incubation of 10 minutes at 25ºC, 120 minutes at 37ºC, an inactivation step of 65ºC for 20 minutes and a final infinite step of 4ºC. The specific components for digestion are reported in the table VII.

*Table VII: Components to perform the digestion of the template.*

| Reagents | Volume (µL) |
|---|---|
| **$H_2O$** | Up to 50 |
| **10X CutSmart Buffer** | 5 |
| **EcoRI HF** | 1 |
| **XhoI** | 1 |
| **Template extracted** | 25 |

Concurrently, 2-3 µg of miR.E vector were also digested and dephoshphorylated, adding 3 µL of FastAP enzyme (ThermoFisher Scientific, Vilnius, Lithuania, ED0651) to the mix detailed in table VII. This product was purified after running in a 1% agarose gel.

Finally, the digested template and the digested and dephosphorylated vector were ligated using the T4 ligase (New England BioLabs Inc. M0202S) as detailed in table VIII overnight at 16ºC.

*Table VIII: All reagents for ligation procedure.*

| Components | Volume(µL) |
|---|---|
| **T4 DNA ligase reaction buffer (2X)** | 5 |
| **Vector DNA (50ng)** | 1 |
| **Insert DNA (in a proportion of 1:3 with vector)** | 1.617 ng |
| **T4 DNA Ligase** | 1 |
| **$H_2O$** | Up to 10 |

The ligated plasmids were transformed into Sbl3 competent bacteria (Thermo Fisher Scientific, Carlsbad, CA, USA, 92008, C7373-03) to isolate colonies and extract DNA from them, which was

checked by sequencing. Only well-cloned vectors, confirmed through sequencing, were utilized for cell infection.

Another plasmid employed was a vector that expresses luciferase within a lentiviral vector, which includes neomycin resistance (#105621, Addgene) (Fig. 18).

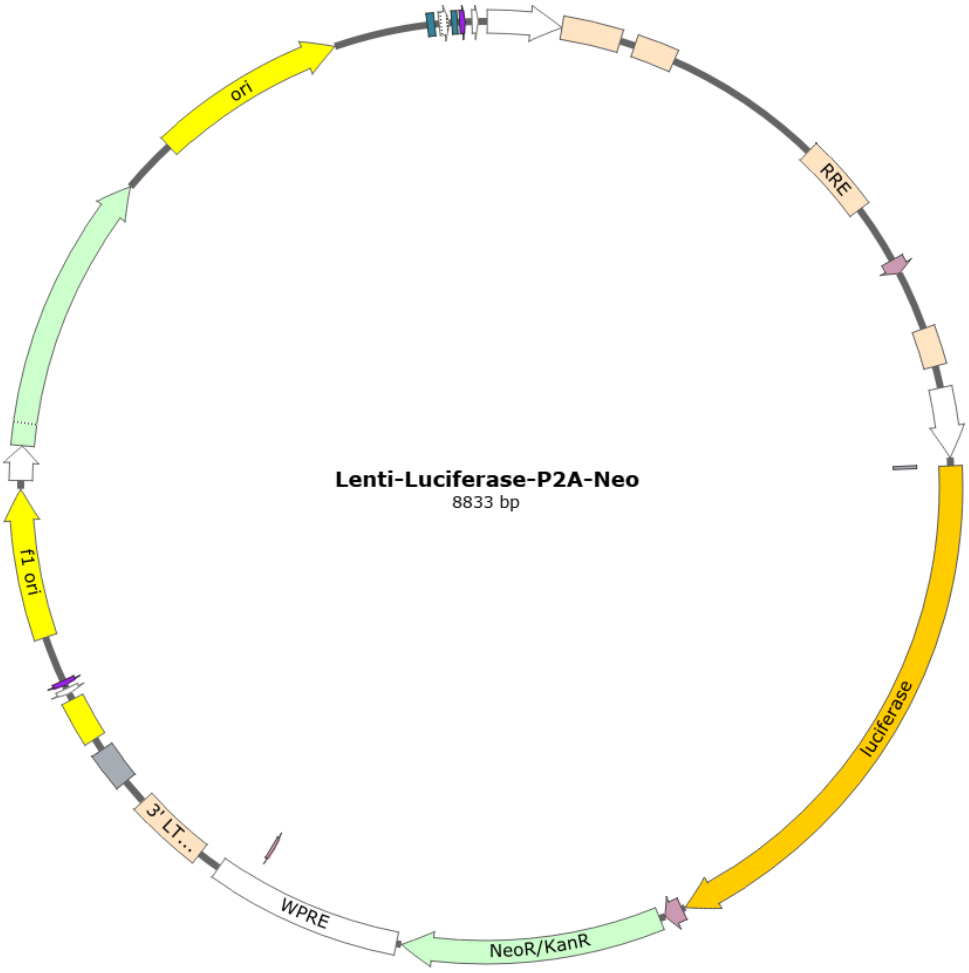



**Figure 18: Map of the LentiNeoLuc plasmid.**

# 3. GlycoCRISPR library and screening experiments

## 3.1 GlycoCRISPR library design

In order to elucidate the role of the protein glycosylation genes in stemness capacity in breast cancer, we have designed a novel CRISPR library targeting all the genes included in the Protein Glycosylation GO term (GO:0006486)

The backbone for our CRISPR library was the lentiCRISPR v2 plasmid (Addgene plasmid #52961). This plasmid encodes the Cas9 protein and a cassette for puromycin resistance. Furthermore, it was genetically modified by our laboratory to express turboGFP reporter protein and facilitate the tracking of our transduced cells.

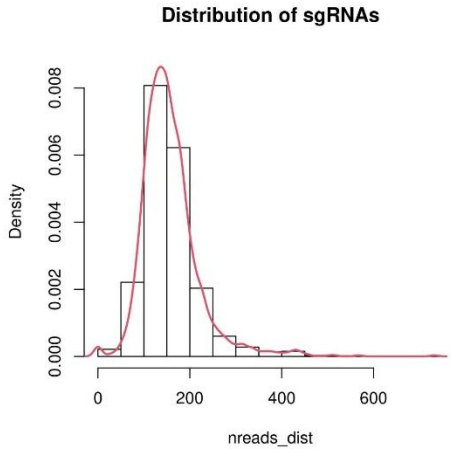



**Figure 19: Graphical distribution of all sgRNA guides of our GlycoCRISPR library.** On average, each guide was represented 157 times.

To specifically target genes linked to protein glycosylation, 10 unique guide sgRNAs were designed for each of the 284 genes identified within the Gene Ontology term GO:0006486. The purpose of creating this number of guides was to minimize off-target effects. Additionally, an equivalent number of guides was generated to target genes associated with stemness to assess changes in stem cell conditions. There are 12 different genes associated with stem-like phenotype. As a control measure, 13 guides were also designed to target non-coding sequences. All guides were synthesized by GenScript enterprise in an array microchip and, subsequently, all the guides were cloned in pool into the lentiCRISPR v2 vector with turboGFP. All genes and the corresponding guides are in the tables of the annexes 1 and 2.

The representativity of each guide in the GlycoCRISPR library was confirmed by NGS, quantifying an average representation of 157 copies for each guide (Fig. 19).

## 3.2 Glycolibrary screening pipeline

Depending on the ultimate goal of the CRISPR screening, there are two approaches that can be employed: positive screening or negative screening. In positive screenings, only the cells with the desired phenotype will be enriched after the selective challenge, indicating that these cells harbor the sgRNAs whose gene deletion will allow the cells to withstand the perturbation. Hence, this is the most suitable strategy to identify the key players in drug resistance. On the other hand, in the case of negative screening, cells which do not acquire the desired phenotype will not be affected by the selection, demonstrating that cells expressing the phenotype will be depleted after the perturbation, being then necessary for the selective condition. Thus, this is often used to detect essential genes and it was the chosen method for conducting our screening.

A carefully well-designed pipeline of our experiment is necessary to answer the biological question. Figure 20 schematically represents the design of the screenings that were performed in triplicate (Fig. 20).

As previously explained, Phoenix cells were our packing lentivirus producer cells. They were transfected using our GlycoCRISPR library and both PMD2.G and psPAX2 helper plasmids to potentiate the lentivirus production. The transfection procedure was performed using jetPEI® (Polyplus-transfection S.A, Illkirch, France) according to manufacturer's instructions. One day after transfection, media was replaced by media of our target cells, collecting all virus media 48h post-transfection. In cases where re-infection of our target cells was necessary, virus media was

once again collected 72 hours post-transfection (detailed protocol can be found in the previous section *2- Engineered cell lines production*). Subsequently, upon collecting virus media, our target cells could be transduced, ensuring optimal conditions, explained in the next section (see *3.3 Glycolibrary screening set up*).

Once the transduction conditions have been established, the next step is to define the selective pressure. Concerning our aim of elucidating the role of glycosylation genes in stemness, we need to identify a method to evaluate only the presence of stem-like cells after the library has been transduced and produced its effect. Therefore, the initial strategy involved evaluating the $CD44^{+}/CD24^{low/-}$ and $ALDH^{+}$ population through multiple passages. Nevertheless, since basal levels of $CD44^{+}/CD24^{low/-}$ and $ALDH^{+}$ cells vary among all breast cancer cell lines, it was not the appropriate method to appraise accurately the stem-like ability of them. Hence, tumorsphere formation assay is the functional tool enabling us to evaluate the stem capability of cell population.

Following the infection, cells were selected using puromycin antibiotic until the uninfected cells had died, taking the breast infected and selected cells as our initial time point (**T0**). Upon collection, cells were distributed for different purposes, ensuring the same representation with an equal cell count at each step (more details in the next section): **(I)** pelleted them for rapid freezing using liquid nitrogen, **(II)** frozen as stock and **(III)** plated them under two conditions. On one hand, they were seeded in monolayer or 2D for 15 doubling times to assess the essential glycosylation genes for their fitness, designating this as a final point (**Tf**). On the other hand, they were also plated in suspension or 3D, forming tumorspheres for three generations to analyse genes involved in stemness condition. At the end, cells from $3^{rd}$ generation were gathered, being this as a final point for 3D (**$3^{rd}$ gen**). When possible, tumorspheres from $1^{st}$ and $2^{nd}$ generation were collected as well. All final points were pelleted and frozen using liquid nitrogen.

Ultimately, DNA from all time points was extracted and gRNA amplified by PCR for sequencing by Next Generation Sequencing (NGS).

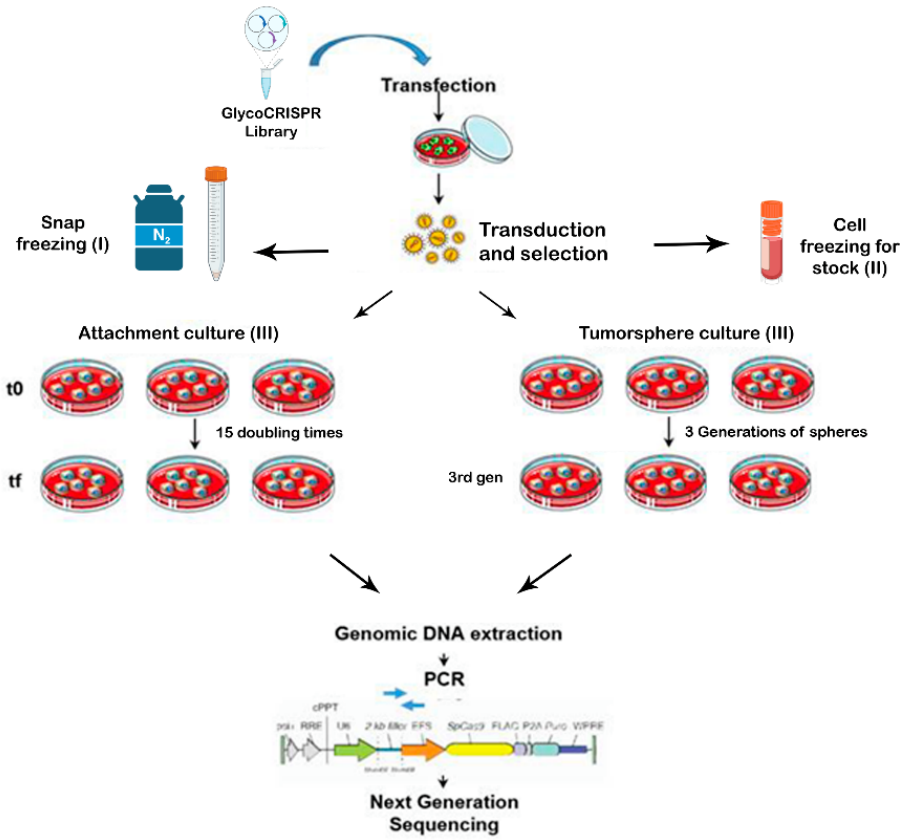



**Figure 20: Schematic workflow of our screening.** From up to down: our library was transfected into Phoenix packaging cells. Once collected virus media, it was used to transduce the Glycolibrary into our target breast cancer cell lines, culturing them, after selection, in both conditions: in 2D or monolayer for 15 doubling times and in 3D or tumorspheres for 3 generations. Pellets from T0, Tf and 3rd gen were collected for DNA extraction and amplicon sequencing.

## 3.3 Glycolibrary screening set up

Achieving the infection of a single viral particle per cell is crucial for deciphering the resultant phenotype. To set up the optimal conditions that allow the infection of target cells by a single viral particle, it is important to measure the multiplicity of infection (MOI) of the virus. MOI is the relation of the viral particle that infects each target cell. As an example, a MOI of 1 means that 1 of this viral particle can transduce 1 cell. Nevertheless, infecting cells using a MOI of 1 does not imply that all particles infect them. There is a **probability** of transduction in the MOI, which follows a Poisson Distribution (Fig. 21) (Target Discovery institute, 2024). Accordingly, we selected to use a MOI $\leq 0.3$ that results in a high probability of infection with a single viral particle while minimal infection with multiple guides.

Thus, concerning our pooled GlycoCRISPR library, it is important to maintain a MOI of 0.3 or lower to ensure that only 1 sgRNA guide is integrated per cell to see its effects.

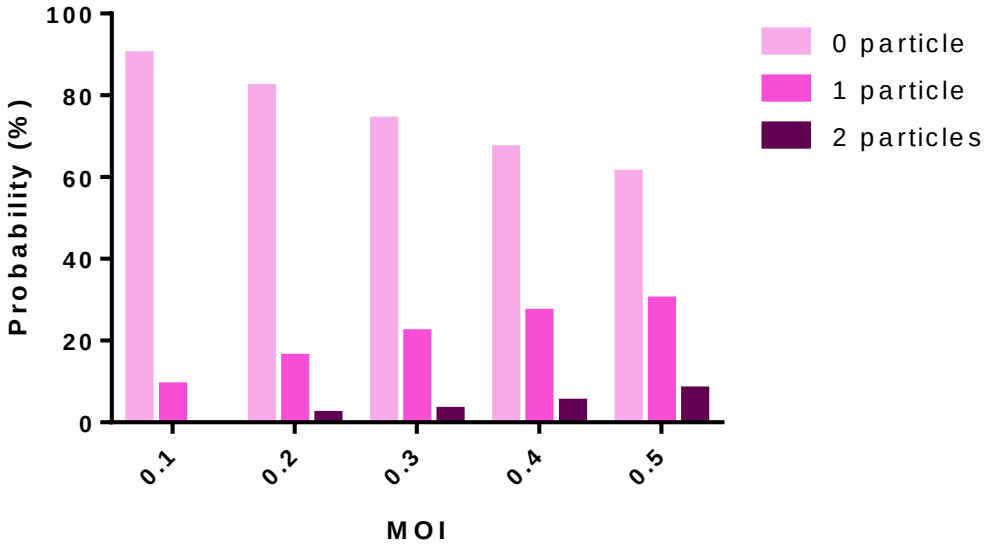



**Figure 21: Poisson distribution that follows MOI, regarding the number of viral particles that can enter to 1 cell.**

However, it is equally important to guarantee that a minimum number of cells are infected with each RNA guide to maintain the reliability of the screening process. It is accepted that maintaining a minimal representation of 100 times the quantity of sgRNAs is sufficient for conducting a reliable screen. In our case, the total representation has been 300 times (300X), maintaining this representativity in all steps of the screening, that is, passages, freezing cells and pellets.

The formula to calculate the final number of cells that need to be infected with our library is as follows:

$$nº\ infected\ cells = quantity\ of\ guides \times representativity$$

Hence, as our GlycoCRISPR library contains almost 3,000 different guides, to keep a representation of 300 times, the number of infected cells with our library must be minimum **900,000 cells**.

Despite this information, it is important to note that not all cells plated will undergo transduction when the MOI is 0.3 or lower. In other words, if the infection efficiency is 30% or less, only a fraction of the cells will be successfully infected. Therefore, to elucidate the initial number of plated cells, it was used this formula:

$$nº\ initial\ cells = nº\ infected\ cells \div MOI$$

Hence, the conditions to set a MOI of 0.3 or lower will depend on each breast cancer cell line properties. Different variants are influencing the efficiency of infection:

**Polycations used**: Polybrene and DEAE-dextran polycations were tested for the transduction.

**Concentrated virus**: This process rises the percentage of infected cells.

**Reinfection step**: As concentrating virus media, reinfecting cells increase the number of cells infected.

**Number of cells plated**: Seeding initially different quantities of cells may affect the number of resulted cells transduced. Depending on cell line, different confluence should be determined.

**Freeze-thaw cycles and spin step:** the effect of freezing the virus was examined as well since virus media change its title with freeze-thaw cycles.

**Centrifugation:** Furthermore, carrying out a spin may increase the efficiency of infection.

**Virus media volume**: the amount of virus media added to cells change the MOI in a likely dependent manner.

After establishing the appropriate conditions, the screening experiment was run.

## 3.4 Sample processing and NGS

All snap-frozen pellets were processed at the same time, and their DNA extracted using the NucleoSpin ® Blood Kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany) following the manufacturer's instructions.

For sequencing, primers were designed considering that NovaSeq 6000 system from Illumina was the chosen system, employing pair-end sequencing. These primers were designed to star at the common region and include the variable region of the GlycoCRISPR library, specifically the sgRNA guide-encoded sequences, resulting in a 314 bp amplicon. Hence, forward primer was the same for all samples. In contrast, reverse primers were designed with unique barcode enabling the distinction between samples at demultiplexing steps (Table IX).

*Table IX: List of primers used for sequencing all the samples. The specific barcode for each primer is highlighted in yellow.*

| Primer | Sequence 5' to 3' |
|---|---|
| **Illumina CPR Fw** | AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC GATCTATCTTGTGGAAAGGACGAAA |
| **Illumina CPR Rv 3** | CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCTGAGCCAATTCCCACTCCTTT |
| **Illumina CPR Rv 4** | CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCTGAGCCAATTCCCACTCCTT |
| **Illumina CPR Rv 5** | CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCTGAGCCAATTCCCACTCCTT |
| **Illumina CPR Rv 6** | CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCTGAGCCAATTCCCACTCCTT |
| **Illumina CPR Rv 7** | CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCTGAGCCAATTCCCACTCCTT |
| **Illumina CPR Rv 8** | CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCTGAGCCAATTCCCACTCCTT |

| | |
|---|---|
| **Illumina CPR Rv 9** | CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCAGACGTGT<br>GCTCTTCCGATCTGAGCCAATTCCCACTCCTT |
| **Illumina CPR Rv 10** | CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCAGACGTGT<br>GCTCTTCCGATCTGAGCCAATTCCCACTCCTT |
| **Illumina CPR Rv 11** | CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTCAGACGTGT<br>GCTCTTCCGATCTGAGCCAATTCCCACTCCTT |
| **Illumina CPR Rv 12** | CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTCAGACGTGT<br>GCTCTTCCGATCTGAGCCAATTCCCACTCCTT |
| **Illumina CPR Rv 13** | CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTCAGACGTGT<br>GCTCTTCCGATCTGAGCCAATTCCCACTCCTT |
| **Illumina CPR Rv 14** | CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCAGACGTGT<br>GCTCTTCCGATCTGAGCCAATTCCCACTCCTT |
| **Illumina CPR Rv 15** | CAAGCAGAAGACGGCATACGAGATTTGACTGTGACTGGAGTTCAGACGTGT<br>GCTCTTCCGATCTGAGCCAATTCCCACTCCTT |
| **Illumina CPR Rv 16** | CAAGCAGAAGACGGCATACGAGATGGAACTGTGACTGGAGTTCAGACGTGT<br>GCTCTTCCGATCTGAGCCAATTCCCACTCCTT |
| **Illumina CPR Rv 17** | CAAGCAGAAGACGGCATACGAGATTGACATGTGACTGGAGTTCAGACGTGT<br>GCTCTTCCGATCTGAGCCAATTCCCACTCCTT |

A PCR was performed to incorporate the primers and amplify the targeted amplicon, setting the conditions to conduct the minimum PCR cycles to prevent any alterations in the amplicon synthesis. Moreover, the same representation of the library had to be maintained. As detailed in the previous section, 900,000 cells had to be infected to have a 300X representation. Estimating that each cell harbour 6 pg of DNA, **5,4 µg of DNA** were initially loaded for each time point to carry out the PCR. Afterwards, PCR products were run in an electrophoresis on a 1% agarose gel for the extraction and purification of the 314 bp band.

Finally, all samples were sent in pool for sequencing using NovaSeq6000, 2x100bp and >8Gb per sample (40M pair-end reads) to Centre Nacional d'Anàlisi Genòmica (CNAG-CRG) in Parc Científic of Barcelona.

## 3.5 NGS data analysis

Once sequenced, FASTQ files were analysed, processing for quality control of the raw data which was performed with FasQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Principal components analysis was performed for clustering the samples based on sgRNAs read counts. Single guides were mapped to the library to generate the counts matrix, using the MAGeCK computational tool which enable the analysis of pooled screens. Negative controls were used for normalization of the expression.

Identification of essential genes was performed with CEDA package (CRISPR screen with Expression Data Analysis), which performs differential expression based on DESeq2 method (Zhao, Yu, et al., 2022). The normalization and differential expression analysis were fulfilled between each final point (Tf and 3rd) in comparison to the initial point (T0). Selection criteria were established based on the following conditions:

I) those genes in which at least 2 sgRNA that fulfilled >50 mean expression in T0,

II) the log2FC ≤-1.5 and

III) the adjusted p-value (FDR)<0.05.

Ultimately, only those genes lost in 3rd gen but not in Tf were chosen as hits. A total of 12 genes fulfilled these filters. The relative abundances of the sgRNAs were represented in a volcano plot.

# 4. Patient data analysis validation

Once the hits were identified, their significance was evaluated in different patient datasets, including datasets from initiatives such as the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC, n=1980) or The Cancer Genome Atlas (TCGA, n=1082). Gene expression and clinical data dataset were collected from the cBioPortal. Microarray data were downloaded in FPKM, then converted to log2 scale. Survival probabilities among different groups were evaluated by log-rank test which was used to compare the survival proportions and Kaplan-Meier plots were generated to illustrate the variations in overall survival (OS) and relapse-free survival (RFS) over time, not only in the overall breast cancer types but also by subtypes. Furthermore, correlations with stemness signatures were represented across subtypes and in overall. All comparisons were performed among patient groups with higher or lower protein expression of our genes of interest.

# 5. Bacterial procedures

Bacteria organisms are used for several purposes including the amplification of DNA from small quantities. Different strains exhibit different levels in their DNA replication reliability. We used *Escherichia Coli* Stbl3 bacteria strain for the experiments which is a strain with low replicative errors.

## 5.1 Bacterial transformation

The process of transformation consists in the introduction of external DNA into bacteria. This procedure was carried out to amplify miR.E plasmids. Briefly, a Stbl3 bacteria aliquot of 25 µL was thawn on ice. In sterile conditions, 100-200ng of the desired plasmid was added into the aliquot, standing on ice for 30 minutes. Later, a heat shock was performed, exposing the aliquot to 42ºC during exactly 30 seconds. This step allowed the entrance of the external vector into the bacteria. Immediately after this time, bacterial aliquot was put on ice again for 2 minutes. Then, 400 µL of S.O.C media (Invitrogen, CA, USA, 46-0700) was added to recover the bacterial aliquot and placed in a shaker 1h at 37ºC.

## 5.2 Colony formation from bacterial clones

As the cloning process outcomes several plasmid products, colony formation from clones was necessary to ensure that the vector was correctly cloned with the desired sequence.

For this, 200 µL of bacterial-recovered aliquot from the previous section was spread into an agar plate with antibiotic selection. It was important to extent all liquid throughout the plate and let it incubate at 37ºC for 16h in an inverted position.

If colonies grew, selected clones were picked and grown in LB low salt broth with antibiotic selection for amplification (explained in the following section). The cloning was confirmed by Sanger sequencing in Stab Vida enterprise, using specific sequencing primer.

## 5.3 Plasmid amplification

To acquire amounts of plasmid to work with, DNA amplification was necessary. Hence, two protocols were employed based on the desired amount of vector produced.

For the MINI protocol, utilizing the QIAprep Spin miniprep Kit (QIAGEN, Hilden, Germany, 27106), bacteria were grown in 5 mL of LB low salt broth with antibiotic selection for 16h at 37ºC in a shaker. Thus, DNA from bacteria was isolated according to manufacturer's protocol. This protocol enables the obtention of little quantity of vector.

In case of MAXI protocol, using the QIAGEN plasmid MAXI Kit (QIAGEN, Hilden, Germany, 12163), bacteria were cultured in an Erlenmeyer with 250 mL of LB low salt broth with antibiotic selection, allowing them growth for 16h at 37ºC in a shaker. To isolate the DNA, manufacturer's protocol was followed. Large quantities of vector are generated with this kit.

# 6. Gene expression analysis

The first step before producing the proteins is the **transcription** of the genome that assembles the DNA information into an RNA molecule.

The quantity of RNA molecules produced by our cells allows us to determine the expression of a specific gene, which can be assessed using quantitative PCR (qPCR).

## 6.1 RNA extraction

Isolation of the RNA from our cells of interest was the first step. Cells were seeded to achieve 70-80% after 48h. Before RNA extraction, cultured cells were washed with PBS 1X and maintained on ice to prevent RNA degradation and slow down the stress-induced transcription. Then, cells were scrapped using the lysis buffer from the Quick-Start RNeasy Mini Kit (QIAGEN, Hilden, Germany, 74104) and the resulting suspension was processed following the kit protocol. Once purified the RNA, Nanodrop (ThermoFisher, ND-ONEC-W) was used to quantify the concentration of RNA from the extracted samples. Finally, all samples were stored at -80ºC until the moment to use them.

## 6.2 Reverse transcription PCR (RT-PCR)

Reverse transcription is the step where single-strand complementary DNA (cDNA) is synthesized from RNA molecules, using the reverse transcriptase enzyme. The final product is analysed to assess gene expression. 400 ng of RNA was used to perform the reaction using the High-Capacity cDNA Reverse Transcription Kit (Invitrogen, MA, USA, 4368814). All components were mixed thoroughly, and the reverse transcriptase was added at the end of the process (table X).

*Table X: Products to carry out the RT-PCR.*

| Components | 1X (µL) |
|---|---|
| **RT 10X Buffer** | 2 |
| **dNTPs 25X** | 0.8 |
| **RT random primers 10X** | 2 |
| **RV transcriptase** | 1 |
| **Sample** | x |
| **Sterile $H_20$** | Up to 20 |
| **Final volume** | 20 |

Using thermocycler, samples were exposed to a first cycle of 25ºC for 10 minutes, a second step of 37ºC during 120 minutes and an inactivation step of 85ºC for 5 minutes. Subsequently, samples could be stored at 4ºC or -20ºC for long-term storage.

## 6.3 Real-Time quantitative PCR (RT-qPCR)

A PCR is a technique which amplifies DNA from a template using specific primers. In case of RT-qPCR, this approach enables the quantitative measurement of the PCR product in real time, facilitated by a fluorescence intercalating molecule. Therefore, once obtained the cDNA, qPCR amplifies a specific product with primers designed for qPCR.

To assess the amount of fluorescence, SYBR®Green (Applied Biosystems, Vilnius, Lithuania, A25742) served as the intercalant label, binding to the double-strand DNA molecules formed and releasing the fluorescence when DNA is annealed. This fluorescence is quantified at the end of every amplification cycle. Fluorescence intensity is proportional to the amount quantity of the PCR product.

Every component for the qPCR mix for each gene is detailed in the following table (table XI). This mix was prepared before adding the sample, 1 µL per sample, previously diluted 1:5. Every sample was loaded per triplicate.

*Table XI: SYBR® Green mix for qPCR.*

| Reagents | 1X (µL) |
|---|---|
| **Sterile $H_20$** | 2 |
| **Primer Fw (10uM)** | 1 |
| **Primer Rv (10uM)** | 1 |
| **Sybr®Green** | 5 |
| **Final volume** | 9 |

The results of the qPCR were shown as Ct (or Cq) values. These values correspond to the cycle where the accumulated fluorescence signal reaches a set threshold, and it is inversely proportional to the amount of gene expression. To normalize the Cq values, GAPDH was used as a housekeeping gene, and $2^{-\Delta Ct}$ was employed for the normalisation procedure.

All the designed gene primers for qPCR were specified in the following table (table XII).

*Table XII: Primers designed for qPCR.*

| Gene primer | Sequence (from 5' → 3') |
|---|---|
| **CD44 Fw** | CCGCTTTGCAGGTGTATTCC |
| **CD44 Rv** | TCTCCATCTGGGCCATTGTG |
| **DHDDS Fw** | AGAGCTGTCACTTTGGGAGC |
| **DHDDS Rv** | CACCGCAGAGTCTCAGCTAG |
| **EXT1 Fw** | GGGGAAGAGGTACCTGACAG |
| **EXT1 Rv** | CATTGTGCAGCATTTCCCGA |
| **GAPDH Fw** | TGACCTCAACTACATGGTCTACA |
| **GAPDH Rv** | CTTCCCATTCTCGGCCTTG |
| **KLF4 Fw** | ACCCACACAGGTGAGAAACC |
| **KLF4 Rv** | ATGTGTAAGGCGAGGTGGTC |
| **Myc Fw** | TCCCTCCACTCGGAAGGAC |
| **Myc Rv** | CTGGTGCATTTTCGGTTGTTG |
| **Nanog Fw** | AGAACTCTCCAACATCCTGAACCT |
| **Nanog Rv** | TGCCACCTCTTAGATTTCATTCTCT |
| **OCT3/4 Fw** | CTTGCTGCAGAAGTGGGTGGAGGAA |
| **OCT3/4 Rv** | CTGCAGTGTGGGTTTCGGGCA |
| **SOX2 Fw** | CGGAAAACCAAGACGCTCAT |
| **SOX2 Rv** | TTCATGTGCGCGTAACTGTC |
| **ST3GAL1 Fw** | TCATGCCCAAATCCCGGAAA |
| **ST3GAL1 Rv** | AGGTTGTCTGTCATCGGCTG |

# 7. Protein expression analysis

After transcription, the mRNA is translated into proteins, which carry out their functions within the cells. Therefore, it is also important to evaluate the protein expression and quantify its levels to elucidate the state and behaviour of cells.

## 7.1 Cell lysis and protein quantification

First, cells were seeded to achieve a 70-80% confluence after 48h. In this moment, cells were washed using PBS 1X to remove any debris and cell waste. Laemmli buffer with DTT (Fisher Scientific, 2440 Geer – Belgium, BP172-5) was used as lysis buffer, scrapping cells with a scrapper or pippete tip. Laemmli buffer formulation is 60 mM Tris ph 7, 10% Glycerol and 2% SDS. Cell lysates were boiled at 95ºC for 5 minutes to denaturalised proteins, performing a short spin to pull down all drops. Subsequently, cell extracts were quantified using Bradford (ThermoScientific, Rockforf, IL, USA, 1863028) according to manufacturer's instructions. Protein absorbance was measured at 595 nm wavelength in a spectrophotometer, comparing with a calibration curve made by Bovine Serum Albumin (BSA) (Fisher Scientific, 2440 Geer – Belgium, BP9703-100).

## 7.2 Western Blot (WB)

This technique allows the separation of protein according to their molecular weight, permitting the recognition of specific proteins in a cellular extract.

To prepare the samples for WB, protein extracts were prepared to reach a final concentration of 10 µg of protein, adding 4X Nupage LDS Sample Buffer (Invitrogen, CA, USA, NP0007) with DTT and diluted with EZ buffer to reach the specific loading volume.

Subsequently, samples were run on acrylamide gels. These gels consist in 2 parts: the stacking and the resolving. Stacking gels are composed by a low percentage of acrylamide to pack the proteins in the same front. On the other hand, the resolving gel can be prepared with different percentages of acrylamide depending on the size of the protein studied. In our case, 8-10% was sufficient. The reagents necessaries to polymerise an acrylamide gel were: distilled $H_2O$, Tris (Fisher Scientific, 2440 Geer – Belgium, BP152-5) pH=8.8 at a final concentration of 375 mM for the resolving and Tris pH=6.8 at a final concentration of 125 mM for the stacking, SDS at a final concentration of 0.1%, Acrylamide/Bis solution, 29:1 (Fisher Scientific, 2440 Geer – Belgium, BP1408-1) at a final concentration of 10%, APS (Fisher Scientific, 2440 Geer – Belgium, BP179-100) at a final concentration of 0.1% and TEMED (AppliChem, Darmstadt, Germany, A1148,0250).

Once polymerised the gel and loaded the samples, the electrophoresis was run at constant voltage of 100V. Afterwards, proteins were transferred to a 45 µm nitrocellulose blotting membrane (Amersham, Germany, 10600002) in a wet transfer method for 90 minutes at 400 mA. Then, transferred nitrocellulose membrane was blocked in 5% non-fat milk solution with TBS-0.1% Tween (VWR, Fontenay-sous-Bois, France, 663684B) (TBS-T) for 1h in RT and incubated with specific antibodies (see the table XIII) diluted in 3% of BSA and 0.02% sodium azide overnight at 4ºC in motion. The membrane was washed 3 times for 10 minutes each with TBS-T at RT and incubated for 1h at RT with the corresponding secondary antibody anti-mouse (Fisher Scientific, 11572122) or anti-rabbit (Fisher Scientific, 15217664) conjugated with horseradish peroxidase (HRP). The signal was developed using Immobilon Forte Western HRP substrate (Merck, Darmstadt, Germany, WBLUF0500) and Chemidoc (Amersham Imager 680, Amersham), according to manufacturer's instructions.

*Table XIII: All antibodies used in the thesis, indicating the dilution factor, host of the antibody and their references.*

| Antibodies used | Dilution factor | Host | Reference |
|---|---|---|---|
| **Anti-β-Actin** | 1:50000 | Mouse | Sigma-Aldrich (A5441-100UL) |
| **Anti-β-Tubulin** | 1:10000 | Mouse | Sigma-Aldrich (T4026-.2ML) |
| **Anti-DHDDS** | 1:500 | Rabbit | Atlas Antibodies (HPA026721-100UL) |
| **Anti-EXT1** | 1:500 | Rabbit | GeneTex (GNT54045) |
| **Anti-ST3GAL1** | 1:1000 | Mouse | Atlas Antibodies (HPA040466-25ul) |

## 7.3 Cell surface protein detection by flow cytometry

Flow Cytometry (FC) is a well-established technique that allows the detection of cell surface proteins and molecules in live cells, as well as intracellular proteins in fixed cells, through the use of antibodies bound to fluorescent dyes. This approach not only provides information about the presence of the protein but also the amount of it, in a single-cell level within a population. Furthermore, more than one protein can be evaluated at the same time, becoming a very useful tool to appraise the protein expression in cells.

In this case, two molecules were assessed at the same time, CD44 and CD24. Antibodies conjugated with two different fluorochromes were used. The human CD44 antibody was conjugated with the allophycocyanin (APC) fluorophore (BioLegend, San Diego, CA, USA, 338806), while the human CD24 antibody was conjugated with the phycoerythrin (PE) fluorophore (BioLegend, San Diego, CA, USA, 311106).

To detect these proteins, cells were seeded the previous day of analysis to achieve a 70-80% confluence. To harvest the cells, they were first washed with PBS 1X and then incubated for 5-7 minutes with Cell dissociation Buffer Enzyme-free PBS-based (Gaithersburg, MD, USA, 13151-014) to avoid the disruption of CD44/24 proteins. Once collected and centrifuged as normally did with completed media, pellets were resuspended and incubated for 10 minutes at 37ºC with 100 µL of Fluorescent Activated Cell Sorter solution (FACS solution) which consists in PBS 1X supplemented with 5% FBS (Fisher Scientific, 10270-106). Later, 1 µL of each antibody was added, previously diluted 1:50 in case of CD44 and 1:10 in case of CD24, for 10 minutes and resuspending by pippeting. Once working with fluorophore-conjugated antibodies, dark conditions are necessary to avoid photobleaching. After incubation time, marked cells were centrifuged and resuspended with FACS solution to analyse them in a Gallios cytometer from Beckman Coulter. All cytometry and statistics were performed using Kaluza software (Beckman Coulter, Brea, CA, USA).

For heparan sulfate analysis, cells were seeded and harvested as mentioned for CD44/24 protocol. When pelleted, cells were resuspended in 40 µL of FACS solution to add heparan sulfate monoclonal antibody (USBiological, H1890) at a final concentration of 250 ng/mL, and let them incubate on ice for 40 minutes. Later, PBS1X was added and cells were centrifuged 5 minutes at 0.3 g. Pellets were resuspended in 100 µL of FACS solution to add Alexa Fluor 488 anti-mouse secondary antibody (Invitrogen, A11001) at a final concentration of 2 µg/mL for 30 minutes on ice. After this time, PBS 1X was added and cells were pelleted to resuspend again in 400 µL of FACS solution to analyse through Gallios cytometer.

## 7.4 ALDH protein activity by ALDERED assay protocol using FACS

The ALDERED assay is based on the capacity of the cells to convert an ALDH substrate conjugated with a fluorophore to the corresponding acid. Therefore, the fluorescence emitted and detected by FACS is directly proportional to the ALDH activity within the lived cells. If the cells have not the ability to transform this substrate, they will release it, not being toxic for them. As a negative control, diethylaminobenzaldehyde (DEAB) is used as a control to let the substrate out of cells and have a background.

The protocol followed was the provided by the manufacturer but adapted for the lab. Briefly, cells were cultured in the previous day of the assay to reach a confluence of 70-80%. To harvest

cells, they were trypsinized as regularly did for their maintenance. Once pelleted, cells were resuspended in 500 µL of AldeRed Assay Buffer. One eppendorf with 2,5 µL of DEAB was prepared for each condition. Hence, 2,5 µL of AldeRed 588-A was added to the cell resuspension and immediately, 250 µL of them were moved to the Eppendorf with DEAB. Cells with and without DEAB were incubated at 37ºC for 40 minutes in dark conditions. After that, cells were centrifuged as normally and resuspended in 400 µL of AldeRed Assay Buffer and maintained on ice to avoid the efflux of the AldeRed and in dark conditions until analyse them in a Gallios cytometer from Beckman Coulter.

# 8. Functional assays

Functional assays enable the implication of proteins under study in a particular biological process through *in vitro* experiments.

## 8.1 Cell viability assay by Thyazolyl Blue Tetrazolium Bromide (MTT)

The MTT assay is a colorimetric experiment that provides insights into the metabolic cellular state, and, ultimately, assesses cell viability. Additionally, it is also used for drug cytotoxicity screening. The chemical principle of this experiment is the reduction of the MTT (yellow) to insoluble purple formazan product through the NADPH cellular oxidoreductase enzymatic system. Dissolving the formazan into dimethyl sulfoxide (DMSO), the solution turned purpled enabling its measurement by a spectrophotometer at 550 nm wavelength.

MTT from Sigma-Aldrich (St. Louis, MO, USA, M2128-1G) was used. First, 2000 cells/well were plated in a 96-well plate. Once adding the specific treatment to cells and waiting for the desired time, a final concentration of 0,5 mg/mL of MTT PBS 1X-based solution was added to each well with a subsequent incubation of 2h at 37ºC and 5% $CO_2$ in dark conditions. Later, all media was removed and 100 µL of DMSO (Thermo Scientific, Rockford, IL, USA, 20688) were added, allowing the dissolution of formazan for 10 minutes in motion in the dark. Then, a spectrophotometer was used to read the plate at 550 nm.

## 8.2 Tumorsphere formation assay

One of the abilities of stem-like cells is to survive in suspension. For this reason, low-attachment plates were employed in these experiments to enrich the stem-like population from our bulk of cells. This subpopulation was grown in the free-sera specific media DMEM/F-12 with concrete supplements. To prevent degradation, aliquots of 50 mL were prepared at the time of seeding the experiment. The recipe of tumorsphere media was detailed in the following table (table XIV):

*Table XIV: Recipe to produce a 50mL aliquot of tumorsphere media.*

| Reagent | From stock | Final concentration |
|---|---|---|
| **B27 (Gibco, Grand Island, NY, USA, 12587-010)** | 1 mL | 1X |
| **hbFGF (Sigma-Aldrich, St. Louis, MO, USA, F0291-25UG)** | 50 µL | 20 ng/µL |
| **hEGF (Sigma-Aldrich, St- Louis, MO, USA, E9644-5MG)** | 50 µL | 20 ng/µL |
| **DMEM/F-12 (serum-free+P/S)** | 49 mL | - |

These aliquots can be stored at 4ºC for a week, and it should be protected from light, due to the photosensitivity of the reagents.

To generate tumorspheres, a low number of cells were seeded to avoid as much as possible interactions and potential signalling pathways that could promote cell survival without exhibiting stem-like characteristics. Hence, 20,000 cells/well in a 6-well ultra-low attachment plate were seeded. Before being cultured, they were passed through a 40 µm cell-strainer to prevent cell aggregates. Additionally, they were grown for 1 week, supplementing every 2-3 days with fresh media at a quarter of the initial volume. Subsequently, the 1$^{st}$ generation of tumorspheres was created.

If a more enriched population was desired, 2$^{nd}$ and 3$^{rd}$ generation of spheres were formed. For that, the 1$^{st}$ generation of tumorspheres was collected in a 15mL falcon using glass pippetes to avoid adhesion in the plastic tips. Once centrifuged as normally did, pellet was well resuspended and incubated for 2' at 37ºC in 500 µL of trypsin to disaggregate the spheres, inactivating them with fresh completed media. After centrifuged them again, cells were resuspended with fresh media and counted using cell counter and trypan blue solution (Sigma-Aldrich, Darmstadt, Germany, T8154-20ML), in a dilution 1:1. Only viable cells were taken into account for culturing again as previously did in the 1$^{st}$ generation and the same quantity of cells.

# 9. *In vivo* assays

Since it is very important to unravel the behaviour of tumor cells and modified cells within a live system, human tumour cells can be injected into immunosuppressive mice. This live system enables us to appraise not only the growth of the tumor but also the potential to metastasise or invade distant organs or tissues.

Therefore, female NSG immunosuppressed mice were purchased and maintained in Specific Pathogen-Free (SPF) conditions. All procedures were conducted according to the guidelines and the approval of ethic committee from Bellvitge Biomedical Research Institute (IDIBELL) and its Animal Facilities head. Food and water were provided ad-libitum, and they were monitored to prevent any pathogens or other problems related to the procedures.

To investigate the role of our genes of interest, MDA-MB-231 cell lines with silenced targets and those with a scramble sequence were employed. Four animals were used for each cell line with every antisense sequence knockdown. Each female mouse was inoculated with 500,000 cells in both flanks in the fourth mammary gland. The cells were mixed in a 1:1 ratio with Cultrex ® (R&Dystems, MN, USA, 3632-005-02) product, a substance that remains in a liquid state at temperatures below 4ºC but begins to gelify at higher temperatures. Thus, it allows the cells to create a matrix for their establishment into the mammary gland.

Briefly, the procedure was as follows: at the age of 6 weeks, female mice were anesthetized using isofluorane 5% with $O_2$ 2%. Subsequently, they were operated, opening an incision next to the 4$^{th}$ nipple to properly visualise the 4$^{th}$ mammary gland and to inject 500,000 MDA-MB-231 cells with Cultrex®, either silencing our targets or the control using a 25G syringe. Upon finishing, the incision was carefully sutured. After that, they were monitored weekly to perform the measurements of the tumor size and the weight of the mice in the indicated days. Additionally, given that human tumor cells were modified to genetically express luciferase reporter gene, 15 mg/mL of Luciferine reagent (Biosynth, Bratislava, FL08608) were injected intraperitoneally per mice and tumor cells visualized using the IVIS Lumina XR and its tracking to prove their invasiveness or migratory capacity.

When tumors reached a specific size, mice were euthanized at the same time to remove the tumors. $CO_2$ was the final point procedure for mice. Tumors were cut in pieces for protein and RNA expression analysis and the biggest piece was fixed in formalin to perform immunohistochemistry. Not only the tumors were fixed but also lungs, which appeared to contain micrometastases in them.

# 10. Histology

## 10.1 Paraffin embedding

Upon the fixation of mice tumors, they were washed using PBS 1X pH 7.4. Briefly, for paraffin embedding, the tumors were dehydrated through a series of alcohol washes: three washes in 70% ethanol for 1 hour each, followed by an overnight wash in 96% ethanol, and finally, three washes in 100% ethanol for 1 hour each. Later, tissues were rinsed with xylol for 1h and then, all of them were embedded in paraffin overnight. Finally, the tissues were again embedded in paraffin to perform the paraffin blocks and carry out sections of µm.

## 10.2 Hematoxylin/eosin staining

Once the sections were made and placed on slides, they were heated for 30 minutes at 60ºC to deparaffinize before starting the staining process. To resume, the slides were washed three times with xylene for 5 minutes each, followed by two washes with 100% ethanol for 5 minutes each, then two washes with 96% ethanol for 5 minutes each, and finally one wash with 70% ethanol for 5 minutes. The final step was to rehydrate the slides with distilled water through three washes of 5 minutes each.

For staining, hematoxylin was applied for 3 minutes, followed by a 1-minute wash with water. Next, a 1% HCl solution was used for a quick wash of 3 seconds, followed by a 1-minute water wash. This was followed by a wash with 1% $NH_3$, another 1-minute water wash, and finally, a 1-minute wash with distilled water. The last step was the staining with eosin for 3 minutes. For mounting, previously the sections needed to be dehydrate washing 3 times with EtOH 96% for 1 minute each, 2 washes of EtOH for 1 and 2 minutes and finally, 3 washes of xylol for 5 minutes. Then, all sections could be mounting using mounting medium. Photos can be taken from this point.

# 11. Compounds

Different substances were used in this thesis. To treat cells and test the effects in glycosylation, tunicamycin drug (Fischer Scientific, 11446412) was employed to mimic the effects of DHDDS silencing, due to the fact that tunicamycin inhibits the GlcNAc phosphotransferase which catalyses the transfer of n-Acetylglucosamine 1 phosphate to dolichol phosphate in the first step of protein N-glycosylation. Additionally, heparinase III (Biotechne, 6145-GH-010) was used to mimic the impact of the knockdown of EXT1.

Ultra-low attachment plates were prepared using poly(2-hydroxyethyl methacrylate) or polyHEMA which was polymerised in cell culture dishes. Briefly, stock solution of 12% polyHEMA was prepared, diluting 2,4 g of polyHEMA (Sigma-Aldrich, St. Louis, MO, USA, P3932-10G) in 20 mL 95% EtOH for 2 complete days, in motion and at 37ºC. Later, the solution was centrifuged to remove any insoluble particles and supernatant was filtered through a 0,22 µm filter and syringe.

This solution should be protected from light. A 1:10 diluted solution with 95% ethanol was made, and 1 mL of the diluted polyHEMA solution was added to a 6-well plate. All prepared plates were dried in a 32ºC heater for 2 days in the dark. Before use, a wash with PBS 1X was required to remove any polyHEMA remainder.

For bacteria resistance treatment, carbenicillin (Fisher Scientific, 2440 Geer – Belgium, BP2648-5) at a final concentration of 100 µg/mL was used to kill all bacteria which did not integrate the constructs with the resistance cassette.

For cell selection after vector transduction, puromycin (Fisher Scientific, 2440 Geel – Belgium, BP2956-100) was employed as selective antibiotic.

# 12. Statistical analysis

All the experiments of the thesis were performed at least three times in independent experiments, if it is not specified. Data was represented and graphed as the mean with the standard deviation (SD). The comparisons between groups were statistically contrasted using Student t-test. All calculations were carried out using Graphpad Prism software (GraphPad for Science Inc., San Diego, CA, USA) and different levels of significance were set as p-value<0.05 (*), <0.01(**), <0.001(***) or <0.0001(****).

# Results

# 1. CRISPR/Cas9 screening in triple-negative breast cancer cell lines

## 1.1 Defining the workflow for our screenings

To elucidate the essential protein glycosylation-related genes involved in maintaining stemness, we have conducted a negative CRISPR screening using our novel CRISPR KO library. As described in figure 25, the lost guides indicated the genes that are essential for stemness under selective pressure. Different TNBC cell lines were first analysed to check whether there might be a unified method to specifically detect the stem-like subset in our cells. These were MDA-MB-231, HS578T and HCC70 cell lines.

In breast cancer, there are two main phenotypes for plastic stem-like cells: $CD44^{+}/CD24^{-/low}$ and $ALDH^{+}$ populations, that are not mutually exclusive. Hence, the presence of both papulations was investigated in the TNBC cell lines above mentioned. As shown in the Fig. 22A and B, the three cell lines had different proportion of the $CD44^{+}/CD24^{-/low}$ cells, having both MDA-MB-231 and HS578T the higher proportion of them. The presence of $ALDH^{+}$ cells was higher in HCC70 in comparison with 231 and HS578T (Fig. 23A and B). The heterogeneity found when analysing these markers indicated that these approaches could not be used as a read out in our screening as we would have to use different criteria among cell lines.

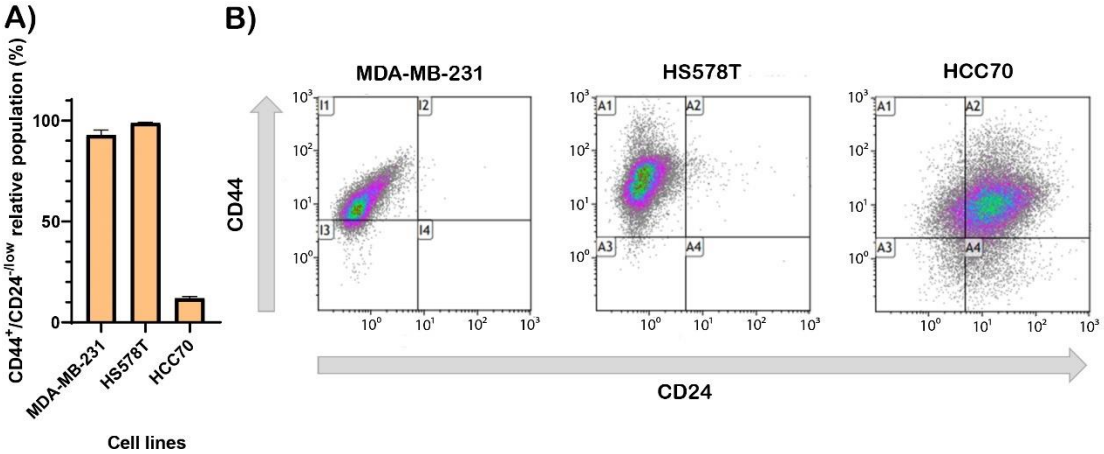



**Figure 22: Analysis of CD44 and CD24 surface protein markers by FACS in TNBC cell lines. A)** Bar plot represents the percentage of $CD44^{+}/CD24^{-/low}$ population in each cell line. The average and SD of two technical replicates are plotted. **B)** Representative images of CD44 and CD24 immunostaining dot plots from FACS analysis in each cell line.

As mentioned in the introduction, growth in suspension is a characteristic of cells with stem properties. By forcing the cells to grow unattached we could select the population with stem properties in all the three cell lines used. To verify that tumorsphere growth (3D) resulted in an enrichment in cells with stem characteristics across the three cell lines, RNA expression of stem-related genes was quantified in suspension. The heatmap represented in Fig. 24A illustrates how these genes increased their expression when cells were grown in 3D condition. The basal non-transformed epithelial cell line MCF10A was also analysed.

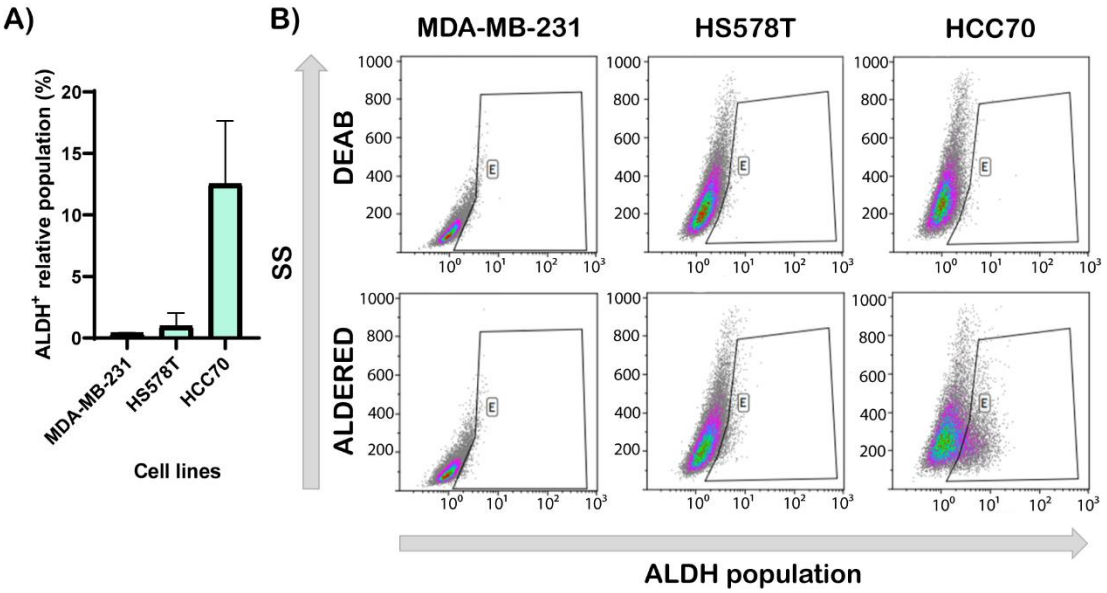



**Figure 23: Analysis of ALDH+ subset measured by FACS in TNBC cell lines. A)** Bar plot represents the ALDH$^+$ population in the different cell lines. **B)** Representative images of ALDH$^+$-stained cell dot plots for FACS in each cell line. Incubation with the ALDH inhibitor DEAB was used to establish the baseline of the fluorescence.

To finally confirm that tumorsphere assay enrich in cells with stem properties, MDA-MB-231 cells were cultured in suspension for 3 generations and CD44$^+$/CD24$^{-/low}$ and ALDH$^+$ phenotype analysed. According to Fig. 24B, CD44$^+$/CD24$^{-/low}$ population in 3D increased significantly in comparison with cells in 2D in all generations. Additionally, ALDH$^+$ subset slightly but significantly rised when cells were grown in 1$^{st}$ generation of spheres.

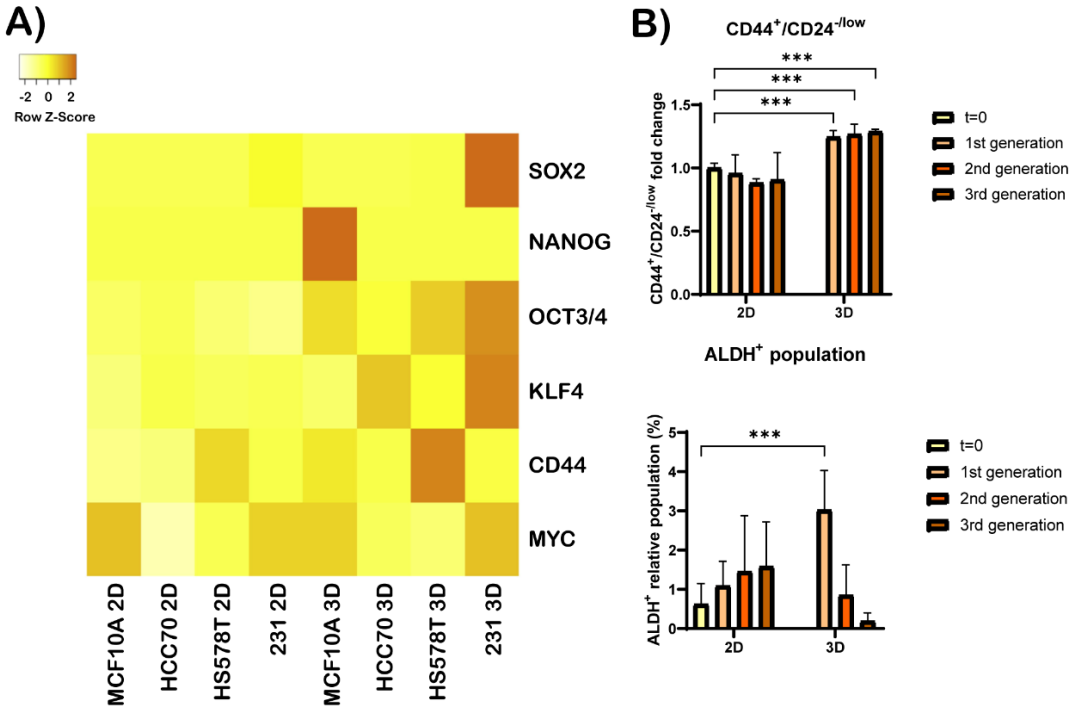



**Figure 24: Validation of tumorsphere assay as a stem-like cell enriching method. A)** Heatmap of RNA expression measured by RT-qPCR in different basal cell lines, cultured under attachment condition and in suspension. **B)** Bar plots of CD44$^+$/CD24$^{-/low}$ (upper panel) and ALDH$^+$ (lower panel) populations in 231 cells. Different days were analysed, regarding the three generations of spheres. Data represent means ± SD, n=6, ***p-value<0.001.

Thus, the pipeline for the screening was established as shown in Figure 25. Briefly and as detailed in the materials and methods section, the pipeline of the screening started with the transfection

of the packaging Phoenix cells with GlycoCRISPR library and helpers. Upon collecting the lentivirus media, breast cell lines were infected with the optimal conditions to ensure that 1 sgRNA integrated into each cell DNA. Following selection, which marked our initial time point (T0), the cells were cultured under two different conditions: in adhesion until reaching fifteen doublings to obtain the final 2D time point (Tf), and in suspension or tumorsphere cultures until the third generation, representing the final 3D time point (3rd gen). When possible, the intermediate generations were also collected for their sequencing. All this procedure was fulfilled three times. Cells from all time points were then processed for sequencing via Next Generation Sequencing (NGS). The detailed methodology was specified in the section 3.2 of the Materials and Methods chapter of this thesis.

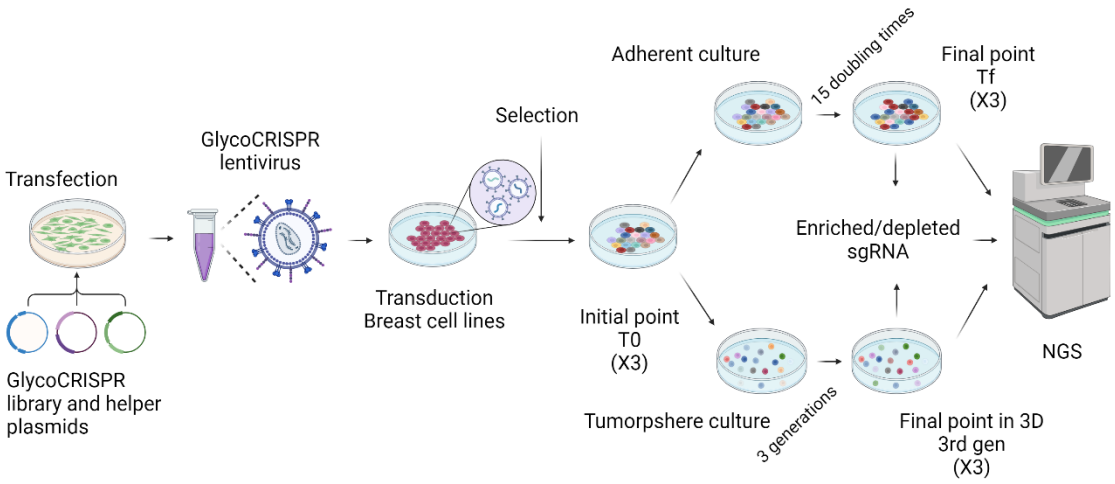



**Figure 25: Pipeline of our CRISPR screening.** Once obtained the pooled GlycoCRISPR library lentivirus media, breast cells lines could be infected according to the optimal conditions of MOI. All this process was executed three times and DNA was extracted from the specified time points (T0, Tf, and 3rd gen) for sequencing.

## 1.2 CRISPR screening in MDA-MB-231 cells

As explained in the materials and methods section, the infection of the glycoCRISPR library in the MDA-MB-231 cells aimed to achieve a MOI of 0.3 or an efficiency of infection of 30% to ensure that only one sgRNA guide entered into each cell. Hence, transduction conditions were optimized to accomplish the established MOI by adjusting several parameters. Table XV summarizes the conditions that were modified to identify the best option.

***Table XV: Summary of the parameters changed in each different experiment to set up the screening conditions in MDA-MB-231 cells.*** *Spin step referred to the fact that the plate could be centrifuged (S) or not (WS) to facilitate the interaction among virus and cells attached. Fresh media (FS) or frozen media (FZ) used are indicated. The percentage of virus added indicates the amount of virus media added considering the final volume used to culture cells.*

| Experiments | Spin (S)/without S (WS) | Fresh (FS) or frozen (FZ) virus | % of virus media added | Number cells obtained | % efficiency of infection |
|---|---|---|---|---|---|
| **1** | S | FS | 25 | $8.8x10^5$ | 17.7 |
| | | FS | 50 | $1.76x10^6$ | 16.7 |
| **2** | S | FS | 25 | $5.55x10^5$ | 13.75 |
| | | FS | 50 | $4.55x10^5$ | 11.4 |
| **3** | S | FS | 50 | $7.9x10^5$ | 17 |
| | | FZ | | $7.35x10^5$ | 16 |
| **4** | WS | FZ | 5 | $1.15x10^5$ | 5 |

| | | | | | |
|---|---|---|---|---|---|
| | WS | FZ | 12.5 | $3.85x10^5$ | 15 |
| | WS | FZ | 25 | $6.45x10^5$ | 23 |
| | WS | FZ | 50 | $1.10x10^6$ | 51 |
| **5 (screening)** | WS | FZ | 25 | $14.25x10^6$ | 20 |

Final conditions were 250,000 cells plated per well in a 6-well plate using DEAE-dextran polycation. This experiment was conducted without performing a spin, using frozen virus, and adding 25% virus media to the cells. All these conditions were determined to facilitate the experimental procedure and saving resources.

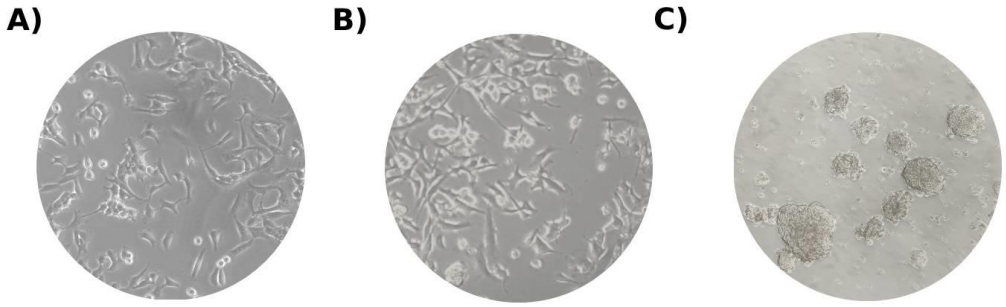



**Figure 26: Representative images of MDA-MB-231 cell line through screening. A)** MDA-MB-231 parental cells. **B)** MDA-MB-231-transduced cells with the glycoCRISPR library. **C)** MDA-MB-231-infected cells grown in ultra low-attachment plates. 10X magnification was used to take A and B images. For image C, a 4X magnification was employed.

MDA-MB-231 cells did not change either their phenotype nor their behaviour after the infection with the glycoCRISPR library, as shown in figure 26A and B, and the transduced cells were cultured until they reach the 3$^{rd}$ generation (Fig. 26C).

## 1.3 CRISPR screening in MCF10A cells

Running the screening in the non-transformed immortalised epithelial breast cells MCF10A allow us to identify essential genes for the mammary normal cells. For this reason, the transduction conditions to obtain a MOI of 0.3 were determined and recapitulated in the table XVI.

***Table XVI: Summary of the parameters modified in each different experiment to set up the screening in MCF10A.*** *The conditions described are the same as in Table I.*

| Experiments | Plated cells | Polycation used | Spin (S)/without S (WS) | Reinfection | Fresh (FS) or frozen (FZ) virus | % of virus media added | Number cells obtained | % MOI or efficiency of infection |
|---|---|---|---|---|---|---|---|---|
| **1** | $5x10^5$ | PB | S | No | FZ | 25 | $3.5x10^4$ | 1.2 |
| | | | | | FZ | 50 | $1.95x10^5$ | 7.8 |
| **2** | $2.5x10^5$ | | | | FZ | 25 | $2x10^4$ | 1 |
| | | | | | FZ | 50 | $1.55x10^5$ | 7.9 |
| **3** | $5x10^5$ | DEAE | | Yes | FZ | 25 | $1.15x10^5$ | 4.4 |
| | | | | | FZ | 50 | $1.5x10^5$ | 6 |
| **4** | $2.5x10^5$ | PB | | No | FS | 25 | $3.5x10^4$ | 4.1 |
| | | | | | FS | 50 | $2x10^4$ | 2 |
| **5** | | | | Yes | FS | 25 | $3.5x10^5$ | 19.6 |
| | | | | | FS | 50 | $1.15x10^5$ | 8 |
| **6** | | | | | FZ | 25 | $1.5x10^4$ | 1.3 |
| | | | | | FZ | 50 | $3x10^4$ | 3 |
| **7** | | | WS | | FZ | 25 | 0 | 0 |
| | | | | | FZ | 50 | 0 | 0 |
| **8** | $2x10^5$ | | S | | FS | 25 | $8x10^4$ | 2.7 |
| | | | | | FS | 50 | $6.6x10^5$ | 32 |
| | | | | | FS | 100 | $1.6x10^5$ | 13 |
| **9 (screening)** | | | | Yes | FS | 50 | $14x10^6$ | 17 |

In this case, 200,000 cells per well were plated to perform the screening. Polybrene was used as the polycation, conducting a spin to the plate with a reinfection and using fresh virus. The amount of virus added was the 50% to the cells.

Similarly to MDA-MB-231 cells, MCF10A-transduced cells did not behave differently to the parental cells (Fig. 27A and B). Nevertheless, since it is not a cancerous cell line, its capacity to form mammospheres is reduced and only one generation of sphere was harvested for sequencing (Fig. 27C).

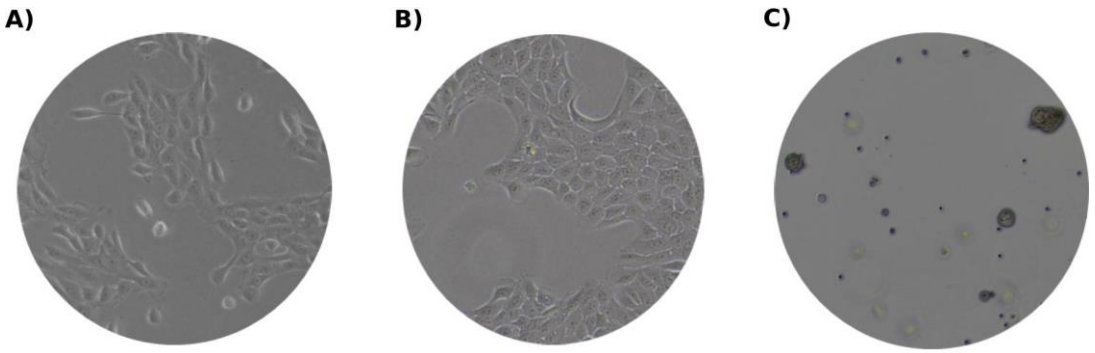



**Figure 27: Representative images of MCF10A cell line through the screening. A)** MCF10A parental cells. **B)** MCF10A cells transduced with the glycoCRISPR library**. C)** MCF10A-infected cells grown in ultra low-attachment plates. 10X magnification was used to take A and B images. For image C, a 4X magnification was employed**.**

In conclusion, the set up for the MCF10A screening was also established, with additional parameters adjusted for the screening. Due to the limited stem capacity of these cell lines, only one generation of spheres could be obtained, resulting in three different time points: T0, Tf and the 1st generation of spheres (1st gen).

## 1.4 Bioinformatic analyses for essential hits in stemness conditions

To perform the sequencing, DNA extraction from all time points was fulfilled and processed to include primers suitable for the sequencing and later analyses. These primers allowed the amplification of specific common genomic regions that contained the sgRNAs of our library. Once sequenced, bioinformatic analyses were conducted. Initially, negative controls were examined to check whether they varied across conditions. These negative controls were guides that did not target any region of human genome, and therefore, they should remain unchanged across conditions. As illustrated in Figure 28, the negative controls remained consistent across different conditions and experiments, both in the MDA-MB-231 screening (Fig. 28A) and in the 10A screening (Fig. 28B).

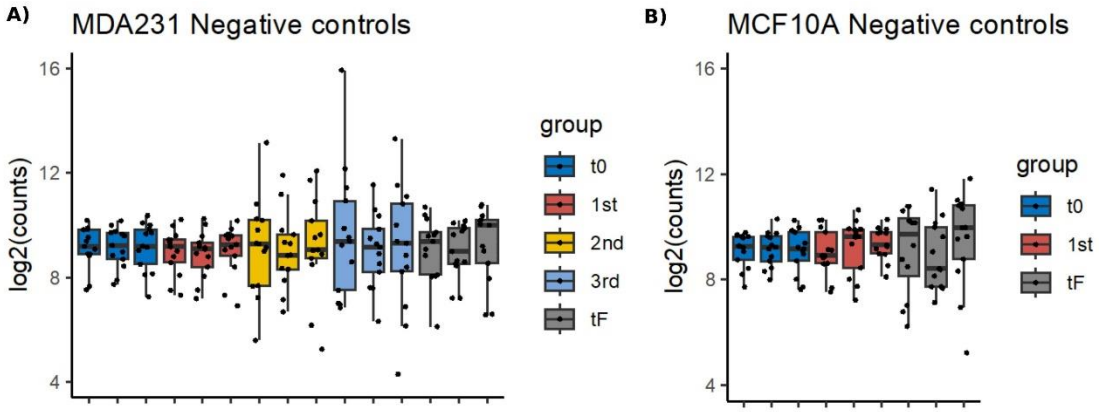



**Figure 28: Changes in the representation of negative control sgRNA guides across conditions in the screenings.** **A)** Negative controls in MDA-MB-231 cell screening. **B)** Negative controls in MCF10A cell screening. Data was normalized with the read counts. Each condition has triplicates.

Moreover, the representative loss of positive gene sgRNA guides in the 3rd and 1st generations (3D condition) of the MDA-MB-231 and MCF10A screenings, respectively, was studied. Our positive controls were guides that targeted genes related to stem-like characteristics, and thus their loss indicated that the screening was successful. sgRNA rank plots in Figure 29A showed the disappearance of guides when compared to the initial point, both in MDA-MB-231 and MCF10A screenings. The loss of selected guides against positive controls are depicted in Figure 29B, revealing a decrease in their representation under suspension conditions across experiments in MDA-MB-231 screening (Fig. 29B) and MCF10A experiment (Fig. 29C). This confirmed the successful execution of the screening.

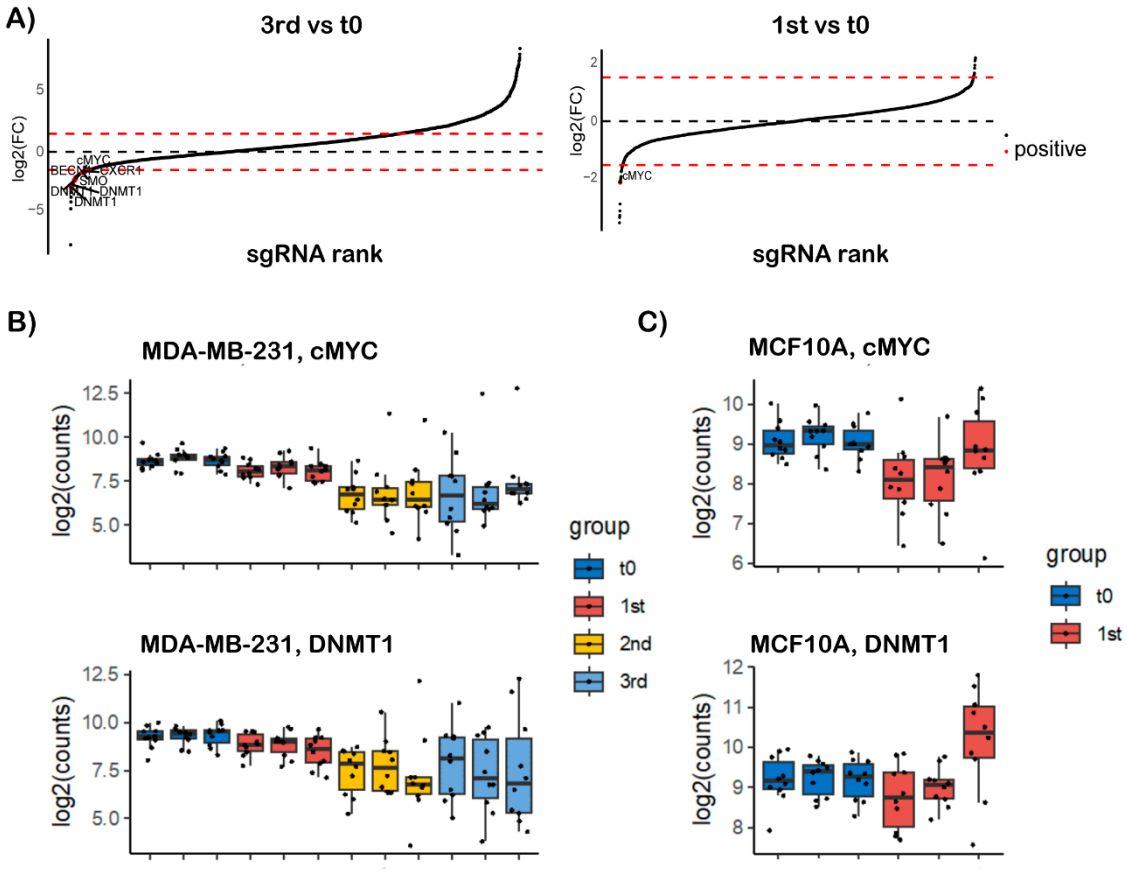



**Figure 29: Changes in the representation of positive control sgRNA guides across conditions in the screenings. A)** sgRNA rank plot where each point represents each sgRNA guide lost. In red, some disappeared positive controls sgRNA guides with their names. **B and C)** Bar plots of positive controls cMYC and DNMT1 in 231 **(B)** and MCF10A **(C)** screenings. Each condition has triplicates.

To select our hits, we set the following criteria: genes must have at least 50 mean reads in the initial point, and the log of fold change (logFC) and false discovery rate (FDR) must be lower than -1.5 and 0.05, respectively. The requirement of a minimum of 50 mean control reads ensured initial representation and avoided stochastic effects. A lower logFC indicated a greater decrease in gene representation. The FDR, an adjusted p-value, helped us confirm that the differences were statistically significant.

The results after applying these criteria were represented in volcano plots for each time point comparison and cell line. The lost and enriched sgRNA guides were highlighted in the volcano plots for MCF10A (Fig. 30A) and 231 cell lines (Fig. 30B).

Therefore, to elucidate the essential genes only for the stemness phenotype, comparisons among 3D and 2D conditions in both cell lines were carried out. The drop of sgRNA representation in 2D allowed us to discard genes essentials for cell fitness. Therefore, only the guides which reduced their representation in 3D conditions would be essential for the stem-like condition.

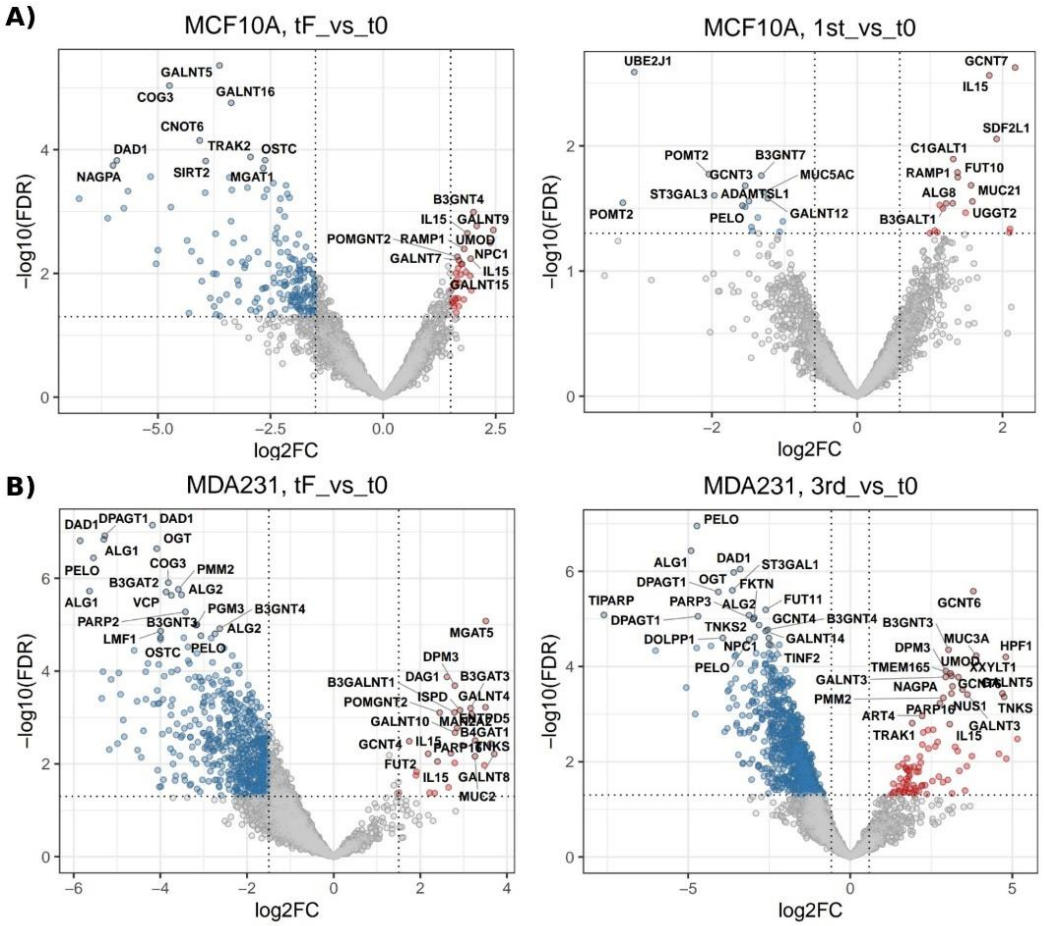



**Figure 30: Volcano plots of sgRNA guides lost or enriched in MCF10A (A) or 231 (B) cell lines.** The name of genes that fulfilled the criteria of more than 50 mean control reads in the initial point, |logFC|>1.5 and FDR<0.05.

Additionally, to select our target genes, a second filter was established. This requirement was that at least two sgRNA against the gene fulfilled the first criteria applied. Those sgRNA guides that accomplished all the conditions and were lost exclusively in the 3rd gen point of 231 are represented in the volcano plot of Figure 31.

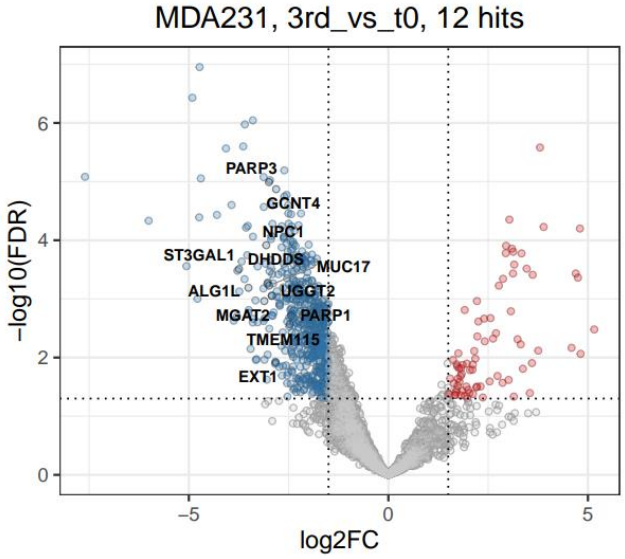


**Figure 31: Volcano plot of sgRNA guides lost (indicated) and enriched in 3rd gen for MDA-MB-231 cell line.** The names of the genes shown are those that had a minimum of two sgRNAs lost, though each is represented only once in the volcano plot.

Overlap of genes included in each condition is represented in the Venn's diagram shown in Figure 32. After this analysis, 10 genes were established as essential to grow in suspension in MDA-MB-231 breast cancer cells. Only one gene (POMT2) was essential for stemness in MCF10A.

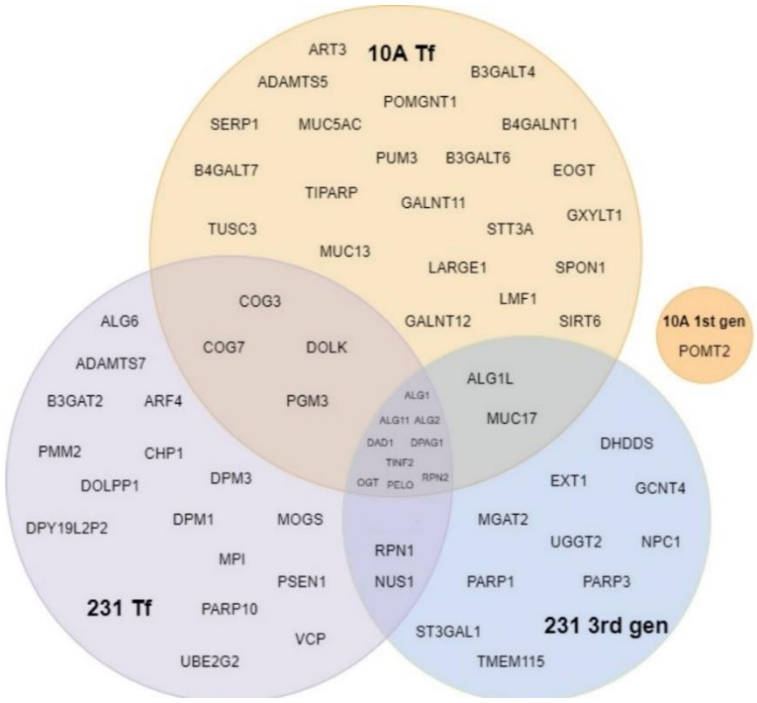



**Figure 32: Venn's diagram of genes lost in the 2D and 3D conditions in both MDA-MB-231 and MCF10A.** Some genes are common for the conditions but only the condition of 3rd generation of MDA-MB-231 sphere are our focus. Number of genes for each condition: 231 Tf (2D)=31, 231 3rd gen (3D)=23, MCF10A Tf (2D)=37, MCF10A 1st gen (3D)=1.

In summary, these data identified a group of 10 genes essential for the stemness condition in MDA-MB-231 cells, distinct from genes crucial for the fitness of MDA-MB-231 cells or the normal non-transformed MCF10A cell line.

## 1.5 Hit selection

To select the targets for extensive study, analyses of their correlation with clinical parameters was carried out. Additionally, bibliographic research was conducted to explore the relationship between the hits obtained and fields such as cancer or stemness. The bibliographic investigation also tried to identify potential drug inhibitors that could specifically target these genes. Therefore, the following table tries to resume all information found analyzing METABRIC data, as well as other site resources such as UniProt, the human protein atlas platform or Pubmed to collect as much information as possible to select our hits.

At first, we questioned whether all genes are protein-coding. The analysis of patients' publicly available data provided correlation with clinical data, including survival rates and relapse-free periods using Kaplan-Meier plots, along with details on gene alterations, such as amplification or deletion, and the levels of mRNA transcripts observed. In addition, we conducted a literature

review to connect our findings to breast cancer and/or stemness. Finally, we also investigated the availability of inhibitors for each of our identified targets (table XVII).

***Table XVII: Summary of all the information compiled for all the targets found.*** **OS: Overall survival; RFS: Relapse-Free Survival. Data obtained from cBioPortal website. **There is only one paper which associated circular RNA of DHDDS with breast cancer, but not the protein or gene itself.*

| Gene | OS* | RFS* | Alterations | Papers related to breast cancer | Papers related to stemness/plasticity | Published inhibitors |
|---|---|---|---|---|---|---|
| ***NPC1*** | NS | NS | 4% mostly high transcripts | Yes | Yes (Neural stem cells) | Yes |
| ***PARP1*** | | | 26% mostly gene amplifications | | Yes (Colorectal cancer/Neuroblastoma) | |
| ***PARP3*** | | S | 4% mainly low transcripts | | Yes (BC) | |
| ***ST3GAL1*** | S | | 25% mostly gene amplifications | | Yes (Glioblastoma) | |
| ***EXT1*** | | | 28% mostly gene amplifications | | Yes (Breast and hepatocellular carcinoma) | No |
| ***GCNT4*** | NS | NS | <1% almost all cases gene amplifications | No | No | |
| ***UGGT2*** | | | 6% mainly high transcripts | | | Yes |
| ***DHDDS*** | S | S | 5% mostly high transcripts | Yes** | | No |
| ***MGAT2*** | NS | NS | 7% mostly gene amplifications | No | | Yes |
| ***TMEM115*** | | | 5% mainly low transcripts | Yes | | No |

Based on overall and relapse-free survival data, three of our identified genes—ST3GAL1, EXT1, and DHDDS—were found to be statistically significant when comparing altered versus non-altered groups. In contrast, PARP3 only showed significance in relation to relapse-free survival. Among these, EXT1 exhibited the highest alteration rate, primarily through gene amplifications, making it a particularly promising candidate for further study. ST3GAL1 is also notable because its translated protein can be targeted by specific inhibitors. DHDDS, while altered in only 5% of breast cancer patients—primarily through mRNA amplification—has limited existing research linking it to breast cancer, with just one paper discussing its association via circular RNA and none

correlating it with stem cell properties. Following this analysis, these three genes were selected for further bioinformatic analysis.

### 1.5.1 EXT1

*EXT1* gene encodes for the protein exostosin glycosyltransferase 1, EXT1. This protein forms an heterocomplex with EXT2 protein, and together lead the elongation of the heparan sulfate proteoglycan chain with the addition of glucuronic acid and N-acetylglucosamine monomeric sugars (Fig. 33) (Busse-Wicher et al., 2014; Mccormick et al., 2000). Heparan sulfate is a glycosaminoglycan that binds to protein core through a linkage region to form the heparan sulfate proteoglycan (HSPG). Finally, the elongation of the heparan sulfate itself is carried out by EXT1/2 complex. Increased levels of HSPGs correlate with tumor cell proliferation, adhesion or invasion (Faria-Ramos et al., 2021; Knelson et al., 2014).

Alterations in *EXT1* gene, particularly through missense and nonsense mutations, are linked to hereditary multiple exostoses, a condition marked by the formation of osteochondromas (Alvarez et al., 2006; Bukowska-Olech et al., 2021). Because of this, EXT1 was initially considered a tumor suppressor (Mccormick et al., 1998). Nevertheless, the role of EXT1 in various cancers, including breast cancer, has been studied, with some research associating it with increased aggressiveness (Solaimuthu et al., 2024). Despite this, its specific role in stemness remains largely unexplored.

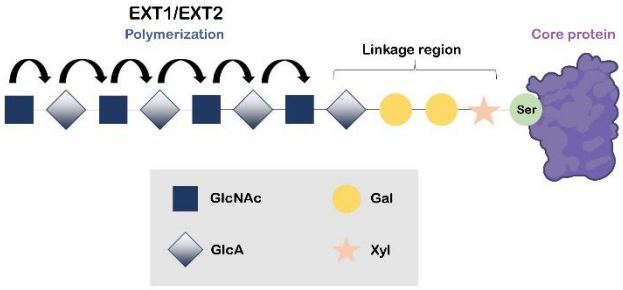



**Figure 33: Representative scheme of heparan sulfate chain synthesis and elongation.** GlcNAc, *N*-acetylglucosamine; Gal, galactose; GlcA, Glucuronic acid; Xyl, xylose. Adapted from (Nadanaka & Kitagawa, 2021).

We analysed in METABRIC database the relationship between EXT1 expression and clinical parameters. By stratifying the patients according to EXT1 expression best cut-off, Kaplan-Meier curves indicated a worse prognosis and a higher likelihood of relapse for patients with higher EXT1 levels. The hazard ratio analysis indicated that high levels of EXT1 increased the risk of death by 41% and the risk of relapse by 55% when compared with patients with low EXT1 expression. (Fig. 34A). The relapse can be plotted as recurrence of the disease, observing a significant higher recurrence in patients with EXT1 higher levels.(Fig. 34B).

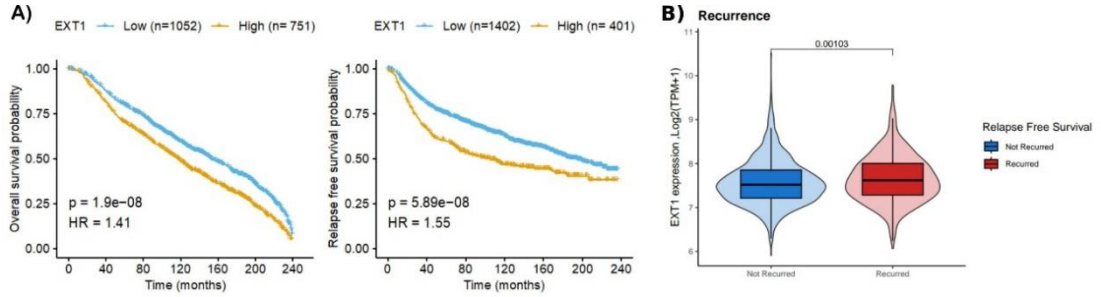



**Figure 34: Correlation of EXT1 expression with survival and relapse in patients from METABRIC database. A)** Kaplan-Meier curves depicting overall survival (left) and relapse-free survival (right), comparing patients with high versus low EXT1 expression. P-value and hazard ratio (HR) are showed. **B)** Box plot representing the recurrence based on the EXT1 expression. The Wilcoxon test p-value is indicated.

Analysis of molecular subtypes showed that basal subset had the highest EXT1 expression in comparison with the other subtypes (Fig. 35A). Furthermore, signatures of stemness and protein glycosylation were significantly associated with patients who had increased EXT1 expression (Fig. 35B). The stemness signature used is described in Ben-Porath, *et al*., specifically in the ES_exp1 signature (Ben-Porath et al., 2008). The protein glycosylation signature referred as the genes into the GO term: protein glycosylation (GO:0006486).

In conclusion, EXT1 has emerged as one of the most promising candidates for further investigation, as all the evidence indicates its significant involvement in stem cell properties.

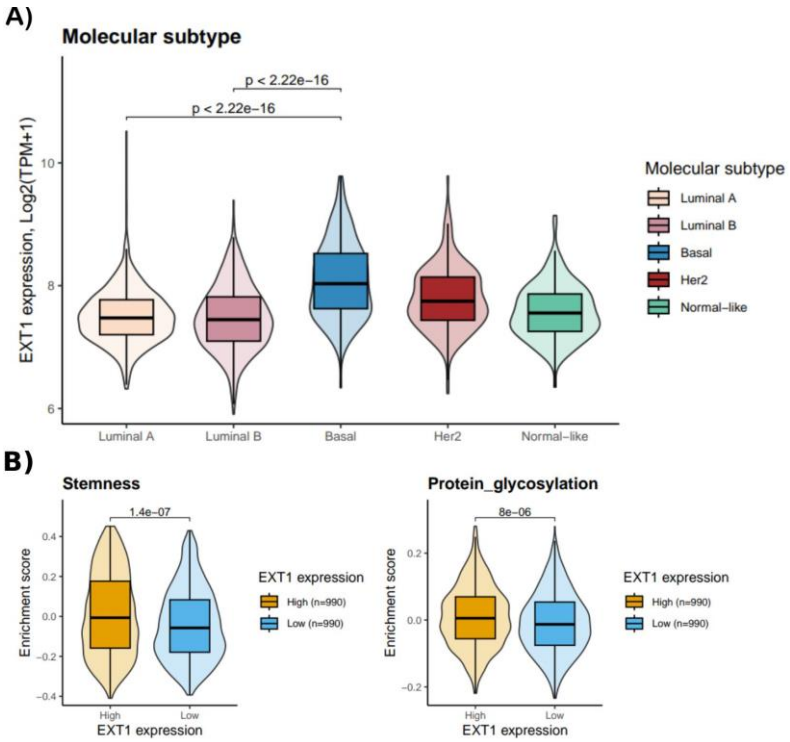



**Figure 35: Differential expression across molecular subtypes and signatures associated with EXT1 expression. A)** Box plot showing the EXT1 expression in every molecular subtype. Only the P-values statistically significant were shown, observed between basal and Luminal A and B subtype and tested by Wilcoxon test. **B)** Box plots representing the enrichment in stemness and protein glycosylation signatures according to EXT1 expression. The Wilcoxon test p-value is indicated.

### 1.5.2 DHDDS

DHDDS protein is the subunit of the human cis-prenyltransferase (hcis-PT, hCIT) or the dehydrodolichyl diphosphate synthase (DDS) complex, whose other subunit is the Nogo-B receptor (NgBR). The function of this complex is the cis-prenyl chain elongation to produce the polyprenyl backbone of dolichol phosphate (Dol-P), condensating various isopentenyl pyrophosphate with a farnesyl diphosphate (Fig. 36).

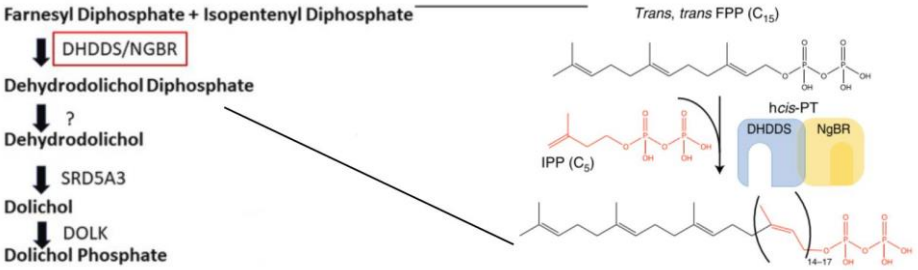



**Figure 36: Function of DHDDS, together with NgBR in the complex DDS, into the dolichol phosphate formation pathway.** Adapted from (Bar-El et al., 2020; Mousa et al., 2022).

The Dol-P is the lipidic glycosyl carrier for the N-type protein glycosylation that happens in the endoplasmic reticulum (Edani et al., 2020; Harrison et al., 2011). Alteration in DHDDS can lead to different pathologies, such as retinitis pigmentosa, epileptic encephalopathies, neurodevelopmental disorders and fatal congenital disorders of glycosylation (reviewed by Bar-El et al., 2020). Alterations in DHDDS have been documented in a few cancers, but not in breast cancer or in relation to stemness.

Applying the optimal cut-off on patients' data from METABRIC database, higher DHDDS expression was correlated with a poorer prognosis and significantly increase patient's probability of relapse. The increased risk of death and relapse are 23% and 22%, respectively, when patients had higher DHDDS expression (Fig. 37A).

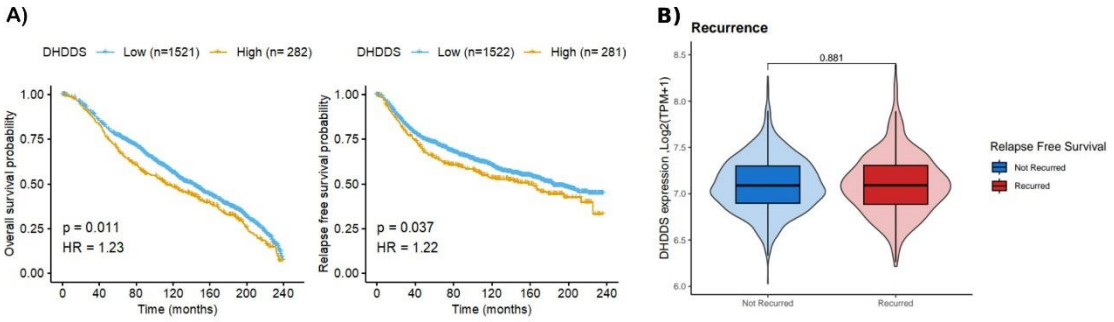



**Figure 37: Clinical parameters from the METABRIC dataset related to DHDDS expression. A)** Kaplan-Meier plots exhibiting worse overall and relapse-free survival in patients with higher DHDDS. Significance and HR are indicated in the figure. **B)** Box plot representing the recurrence based on the DHDDS expression. The Wilcoxon test p-value is indicated**.**

Unlike EXT1 gene, increased DHDDS expression was not related to a higher ratio of recurrence (Fig. 37B) or with the any breast cancer molecular subtype (Fig. 38A). The protein glycosylation signature enrichment was significantly higher in patients with elevated DHDDS expression in the same line as EXT1 higher expression patients. However, and against what we obtained in the screening results, patients with higher expression of DHDDS were negatively correlated with the stemness signature studied (Fig. 38B).

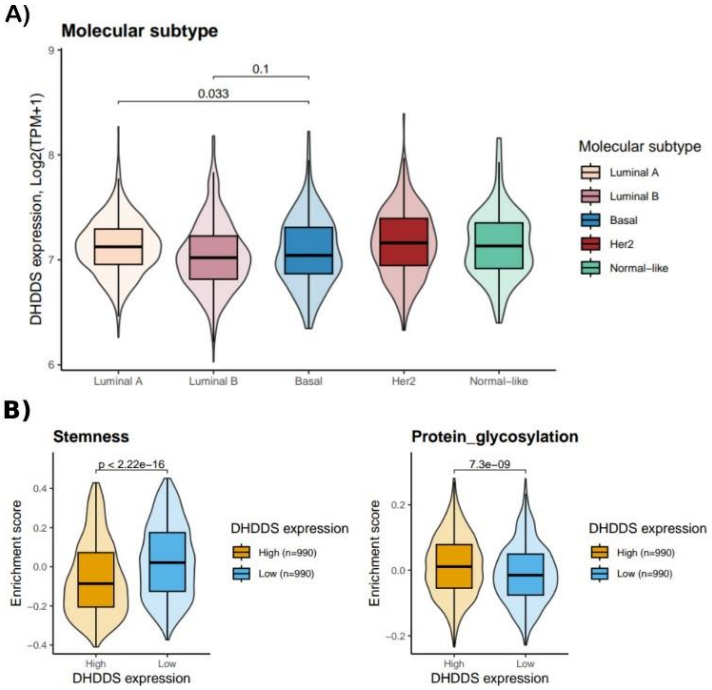



**Figure 38: Differential expression across molecular subtypes and signatures associated with DHDDS expression using METABRIC data. A)** Box plot showing the DHDDS expression based on the molecular subtypes. P-values above boxes indicate the statistical differences between basal and Luminal A and B subtype tested by Wilcoxon test. **B)** Box plots representing the stemness (left) and protein glycosylation (right) signatures based on the DHDDS expression. The Wilcoxon test p-value is indicated.

Lastly, DHDDS could be a promising target aspirant for validation. Despite being negatively correlated with the stemness signature studied, it appeared to be essential for stemness in our screening. Additionally, clinical data verified its association with poor prognosis and increased relapse, being compelling hit for further investigation.

### 1.5.3 ST3GAL1

*ST3GAL1* gene encodes the sialyltransferase ST3GAL1 protein, an enzyme responsible for adding sialic acids to a galactose-containing oligosaccharides. Specifically, this protein is the β-galactoside-α-2,3-sialyltransferase-1, transferring the sialic acid in an α2,3 linkage (Fig. 39) (Pietrobono et al., 2020; X. Wu et al., 2018). The alteration of this gene is strongly associated with several tumors (Dall'Olio et al., 2014). Nevertheless, the relationship between this protein and stemness has not been investigated yet.

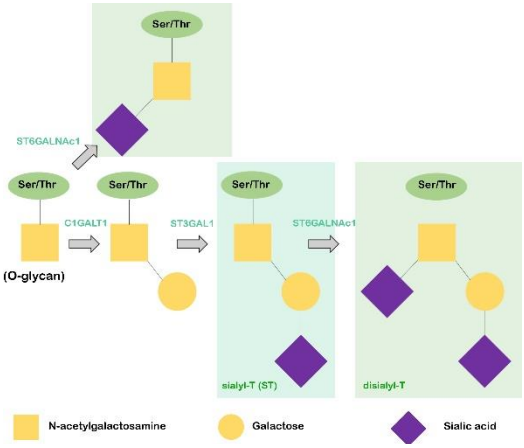



**Figure 39: Mucin type O-glycan formation.** Sialic acid is also named as N-acetylneuraminic acid. Adapted from (Zhou et al., 2023).

Analyzing the METABRIC patient database, individuals with higher ST3GAL1 expression showed poorer overall survival rates and higher probability of relapse. Specifically, the risk of death was 36% higher in patients with elevated ST3GAL1 expression compared to those with lower expression, while the risk of relapse was elevated by 43% (Fig. 40A). Additionally, the recurrence rate was significantly higher in those patients with elevated ST3GAL1 expression (Fig. 40B)

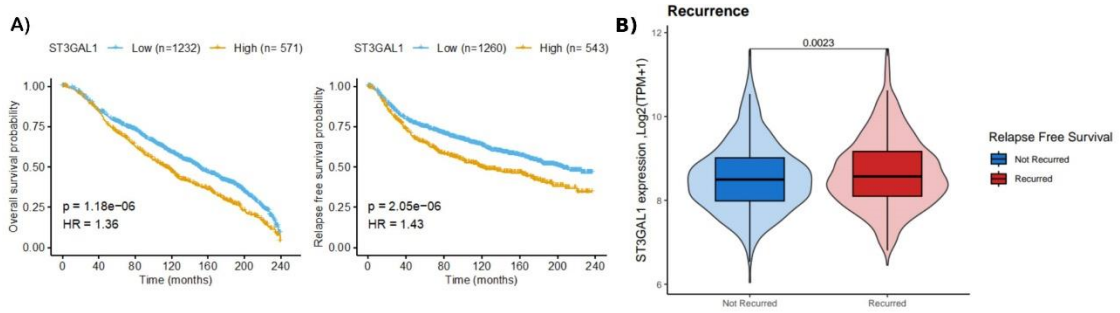



**Figure 40: Clinical parameters from the METABRIC dataset related to ST3GAL1 expression. A)** Kaplan-Meier curves representing the overall survival (left) and relapse-free survival (right) comparing high vs. low ST3GAL1 expression. P-value and hazard ratio (HR) are represented. **B)** Box plot representing the recurrence based on the ST3GAL1 expression. The Wilcoxon test p-value is indicated.

Increased ST3GAL1 expression was not related to any breast cancer molecular subtype. Basal tumors expressed less ST3GAL1 compared to other subtypes (Fig. 41A). In addition, unlike patients with high expression of EXT and DHDDS, those with higher ST3GAL1 expression did not show statistical differences in the stemness signature and had a significant negative correlation with patients who had lower ST3GAL1 expression (Fig. 41B).

Clinical data revealed that elevated ST3GAL1 levels were correlated with worse prognosis and higher relapse rates. Despite the absence of differences in stemness signatures among patients, and the negative association between protein glycosylation signatures and protein glycosylation status, ST3GAL1 was found to be essential for the stem-like phenotype in our screening. Therefore, ST3GAL1 stands out as a promising target candidate.

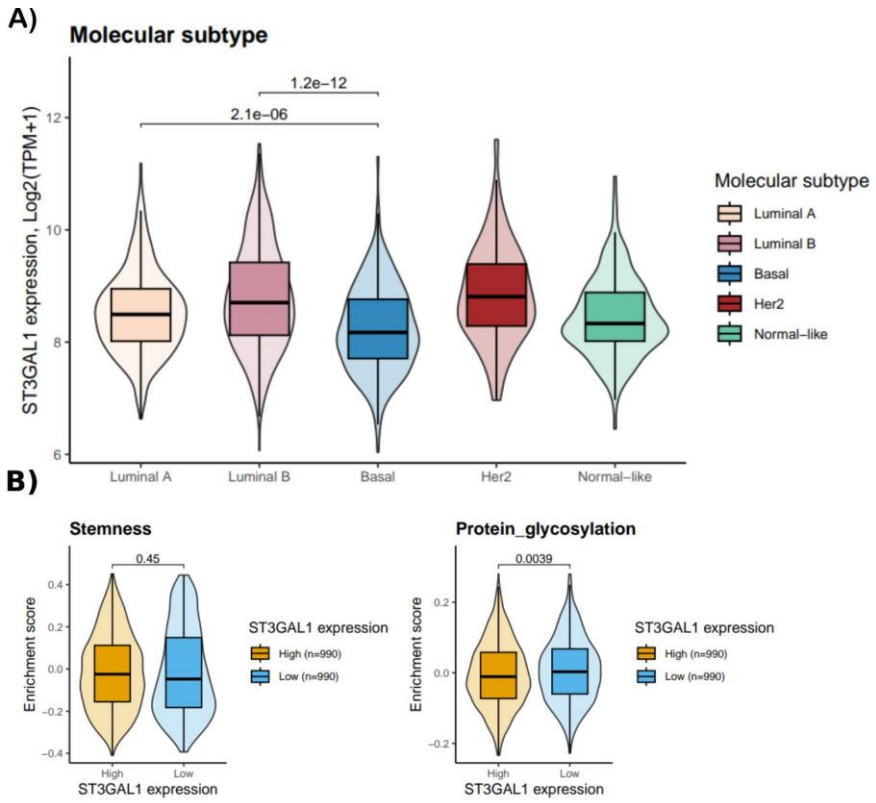



**Figure 41: Differential expression of DHDDS across molecular subtypes and associated signatures, as analyzed using METABRIC data. A)** Box plot showing the ST3GAL1 expression based on the molecular subtypes. P-values above boxes indicate the statistical differences between basal and Luminal A and B subtype tested by Wilcoxon test. **B)** Box plots representing the stemness (left) and protein glycosylation (right) signatures based on the DHDDS expression. The Wilcoxon test p-value is indicated.

## 1.6 Stemness gene sets and patient associations with our hits

As we wanted to investigate the relationship between our hits and the stemness phenotype, we aimed to explore the different gene expression signatures associated with stem-like status. According to Ben-Porath, *et al*., different sets of stem-like gene expression signatures can identify patterns resembling embryonic undifferentiated stem cells. While we utilized one signature from their study, we sought to evaluate the association of our targets with the remaining signatures. This study also linked these different gene expression patterns with breast tumor samples, strongly supporting the use of them for the linkage of our hits with stem-like phenotype in our research. We thus examined patients with elevated target gene expression to determine whether it correlated with the expression of all stemness gene sets (Fig. 42).

Unlike the other targets, patients with increased DHDDS expression were related to an overexpression of polycomb repressor signatures, as well as a significantly negative associated with undifferentiated state signatures, such as Sox2 and Oct4_targets or NOS_TF gene expression sets. Consequently, these patients are correlated to a differentiated state, although this gene appeared to be essential for the stemness in TNBC cell lines. Therefore, it might be interesting the implication of DHDDS in stemness.

Concerning EXT1, there was a significant positive correlation with some stemness-related gene expression sets, such as ES_exp1 and 2, and NOS signatures. Additionally, the Polycomb repressor signatures as PCR2, Suz12 and H3K27 are negatively correlated to increased EXT1

expression patients. These signatures are expressed in a cell-differentiated state. Hence, all these correlations showed an increased stem genes expression for these patients, making it one of the best candidates for further investigation.

In case of ST3GAL1, a significant positive correlation with the ES_exp2 signature and a significant negative correlation with polycomb signatures indicated that patients with elevated ST3GAL1 expression exhibit a less differentiated status. Even though some other signature sets were significantly underexpressed, ST3GAL1 could be a good candidate to research its implication in stemness (Fig. 42).

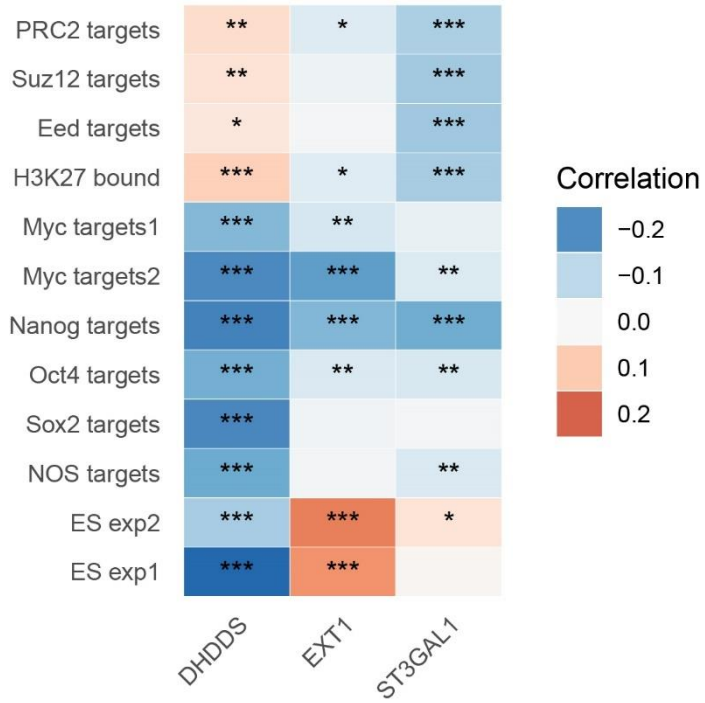



**Figure 42: Stemness gene expression sets' correlation with each target expression.** Positive correlations are represented in red whereas negative correlations are in blue. Significance is represented as: *p-value<0.05, **p-value<0.01, ***p-value<0.001.

Considering all these factors, and following extensive bioinformatic analyses of EXT1, ST3GAL1, and DHDDS, these three genes were chosen for validation of the screening results.

## 2. Functional validation in MDA MB 231 cells

As the screening was conducted in MDA-MB-231 TNBC cell line, the validations were initially carried out in this cellular model. The validation is conducted through RT-qPCR and WB to see the basal levels of our targets in 2D and 3D. Additionally, the gene expression was decreased using the miR.E system to later culture these cells in suspension. Furthermore, the $CD44^{+}/CD24^{-/low}$ population was studied in 2D and 3D, as well as the gene expression in 3D.

## 2.1 Hits expression in tumorsphere condition

To explore the potential role of all hits in stemness conditions, firstly the expression of EXT1, ST3GAL1 and DHDDS was studied when cells were cultured in suspension (3D). Three generations of spheres were analysed through RT-qPCR, comparing them with the attachment condition (Fig. 43A). The expression of the three targets was increased in the 1st generation in comparison with cells attached, suggesting a requirement of them to grow in suspension. Strikingly, the expression in 2nd and 3rd generation tended to decrease gradually along generations in all of our genes.

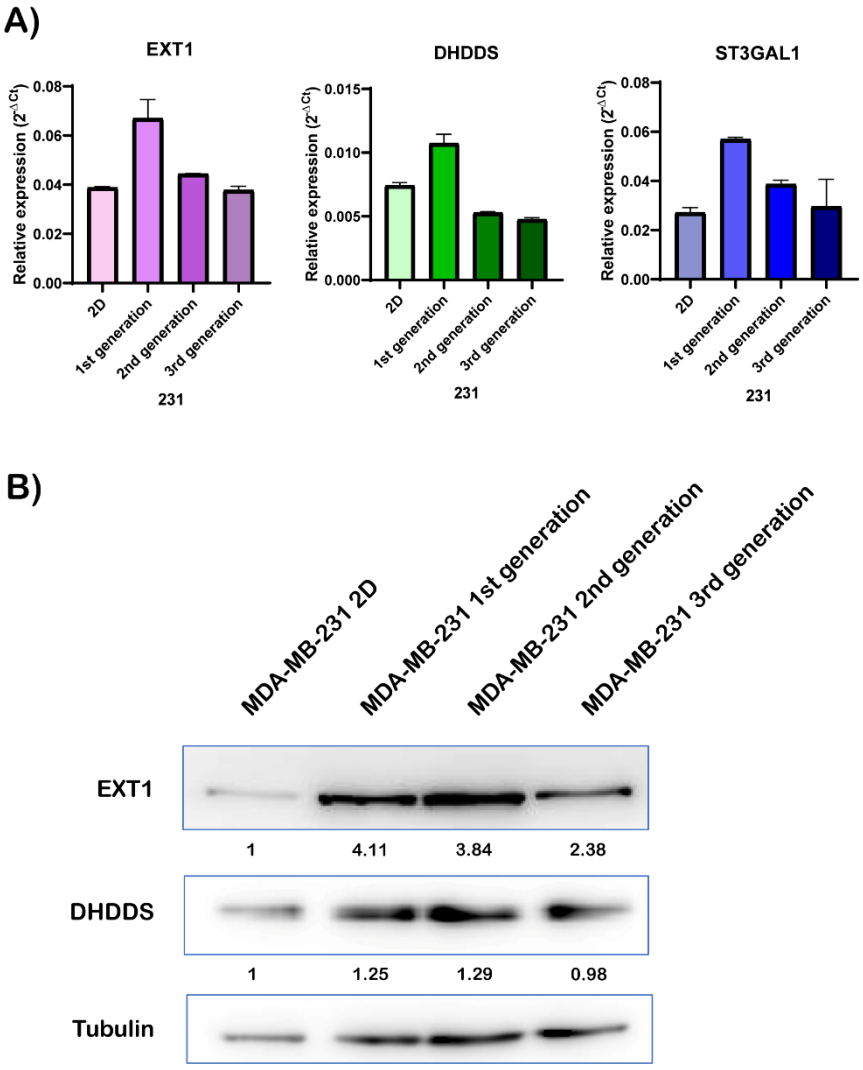



**Figure 43: Expression of our targets in MDA-MB-231 cells under suspension and in attachment conditions. A)** Gene expression analysed by RT-qPCR of all three genes in attached culture (2D) and in suspension (1st, 2nd and 3rd generations). Bar plots represent data as mean ± SD of three technical replicates. **B)** Protein expression of DHDDS and EXT1 measured by western blot. Tubulin was used as a loading control. Fold change quantification is indicated, under each western blot image.

Next step was to observe if the gene expression was in line with the protein expression levels. As observed in the Figure 43B, EXT1 protein expression was increased along all generations of tumorspheres in comparison with cells in 2D. DHDDS protein levels were increased in 1st and 2nd sphere generation, although there was no change in 3rd generation. The detection of ST3GAL1 protein was not possible in that moment.

These results suggest the active transcription and translation of these genes for, in particular, suspension conditions.

## 2.2 Hits silencing using miR.E system

To further study the involvement of our targets in stemness, we performed a knockdown using the miR.E technology. For each target, 3 different antisense sequences were designed to specifically silence our target genes, and a scramble sequence was also used as a control. For clarity, these antisense sequences are referred to as sh.

Therefore, MDA-MB-231 cells were infected using lentivirus carrying each sh of our targets or the scramble sequence to constitutively exert their effect in the cells. Unfortunately, one of the sh against ST3GAL1 gene was not successfully cloned into miR.E plasmid. For this reason, there are only two knocked-down cells for this gene.

The effectiveness of the silencing was initially determined by quantifying mRNA levels. As shown in Figure 44A, all antisense sequences reduced the RNA expression of each target by more than 50%, except for the ST3GAL1 silencing, which achieved a knockdown of only about 30%. Additionally, parental cells were also analysed to determine if the transduction process itself affected the expression of our targets, which it did not.

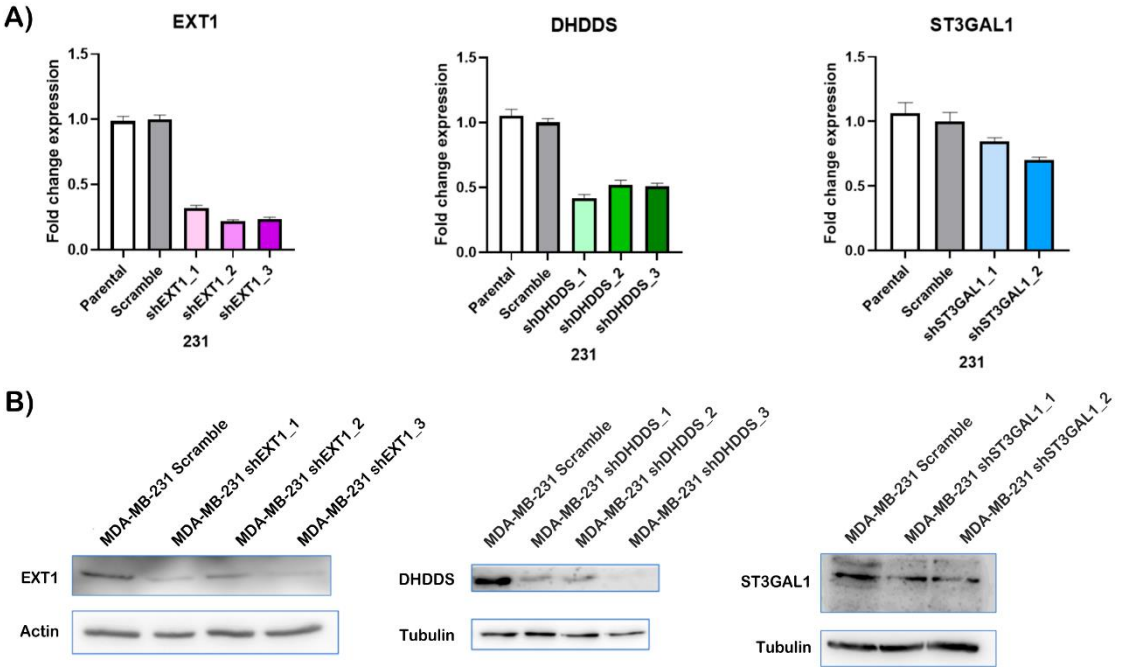



**Figure 44: Expression of targets in MDA-MB-231 cells after performing the silencing using the miR.E system. A)** Bar plots of the RNA expression levels for EXT1 (left panel), ST3GAL1 (middle panel) and DHDDS (right panel). Data from one experiment are represented as the mean of triplicate fold change respecting scramble ± SD. **B)** Western blots of each knock down cell line comparing with the scramble or parental cell lines. In case of EXT1, actin protein was used as a housekeeping. For ST3GAL1 and DHDDS, tubulin was used as the loading control.

Furthermore, the protein levels were also checked to confirm their decrease. As observed in the western blots shown in Figures 44B, the proteins for targets EXT1, DHDDS and ST3GAL1 were effectively knocked down.

In summary, MDA-MB-231 cells with the silencing of all target hits were obtained. The knockdown of both EXT1 and DHDDS targets was successfully achieved, as confirmed by both

RNA and protein levels analyses. Nonetheless, analyses in the ST3GAL1-silenced cells did not show a good decrease in either RNA or protein levels.

## 2.3 Effects of target silencing in tumorsphere formation capacity

After the silencing, we examined the role of our targets in stem-like phenotype performing tumorsphere formation assay. Hence, EXT1-, DHDDS- and ST3GAL1-silenced cells were cultured in ultra low-attachment plates to evaluate their capacity to form spheres. All photos were taken after 7 days, and the subsequent analysis was performed using Fiji software to quantify the tumospheres. We considered tumorspheres the cell aggregates larger than 60 µm.

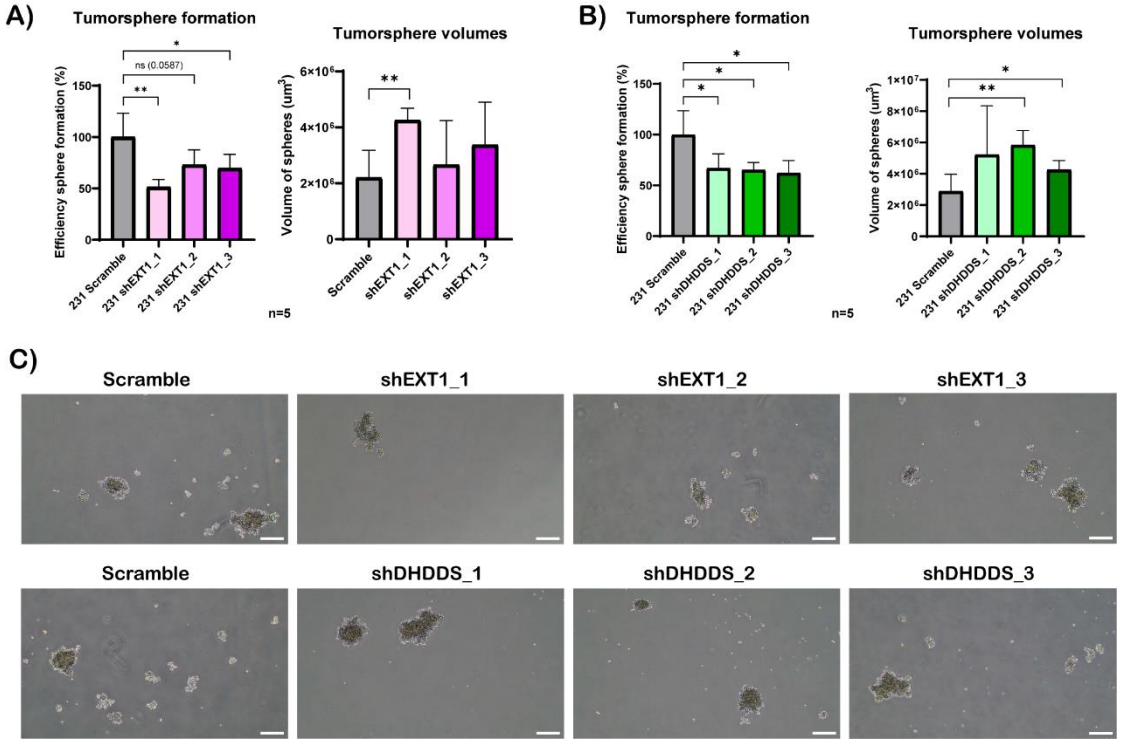



**Figure 45: Abiliity for tumorsphere formation of the EXT1 and DHDDS-silenced MDA-MB-231 cells. A and B)** Bar plots of the efficiency sphere formation (left panels) and the sphere volumes (right panels) for the EXT1 knocked-down cells (A) and DHDDS-silenced cells (B). Data are represented as the percentage of the mean ± SD. n=5. *p-value<0.05, **pvalue<0.01, ns=non-significant. **C)** Representative tumorsphere images taken from silenced cells as indicated, using 4X magnification. Scale bar in each image depicts 250 µm in length.

The tumorsphere formation efficiency was significantly reduced due to the knocking down of the EXT1, except for the second sh, which showed an almost significant reduction that could likely reach significance with an additional biological replicate. The volumes of spheres did not change significantly with the exception of the first sh (Fig. 45A and C, upper panels). The effect of the DHDDS silencing resulted in a significant decrease in the capacity to form tumorspheres in all the silenced cells. The volume of the spheres was significantly larger than the scramble control, even though the first did not show statistical differences (Fig. 45B and C, lower panels).

As previously mentioned, ST3GAL1 silencing was not as much as expected, despite considerable efforts to improve the knockdown were conducted. Then, there were not statistical differences in the sphere formation capacity between the scramble controls and the knockdown cells (Fig. 46A and B).

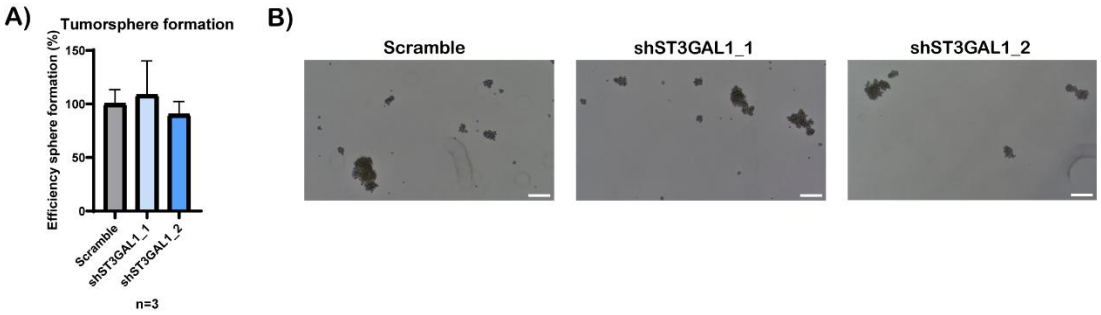



**Figure 46: Changes in the ability for tumorsphere formation in the ST3GAL1-silenced MDA-MB-231 cells. A)** Bar plots of the efficiency sphere formation for the ST3GAL1 knocked-down cells. Data are represented as the percentage of the mean ± SD. n=3. **B)** Representative tumorsphere images taken from silenced cells as indicated, using 4X magnification. Scale bar in each image depicts 250 µm in length.

These functional assays suggested an essential role of EXT1 in stemness, as expected after bioinformatic analyses and as predicted by the initial screening results. interestingly, DHDDS had also a critical importance in stem-like phenotype, as indicated by our screening, despite the fact that bioinformatic results did not show any connection between this gene and the stemness capacity. Unfortunately, the ineffective ST3GAL1 knockdown and the subsequent tumorsphere formation outcomes led us to eliminate this candidate from further consideration.

## 2.4 Analyses of stem properties after knocking-down our targets

Once confirmed the implication of EXT1 and DHDDS in the stem-like phenotype, we wonder how they influenced stem properties. Thus, $CD44^{+}/CD24^{-/low}$ population was first studied by analysing the surface markers by FACS.

This analysis revealed that $CD44^{+}/CD24^{-/low}$ population was not significatively reduced after the silencing of our targets, when cells were cultured previously in attachment (Fig. 47). Hence, given their involvement in the stemness characteristics, we assess this population when cells were plated under suspension conditions. Interestingly, neither EXT1 nor DHDDS knockdown significantly altered the percentage of $CD44^{+}/CD24^{-/low}$ population, except for the second DHDDS-silenced cell line (Fig. 48).

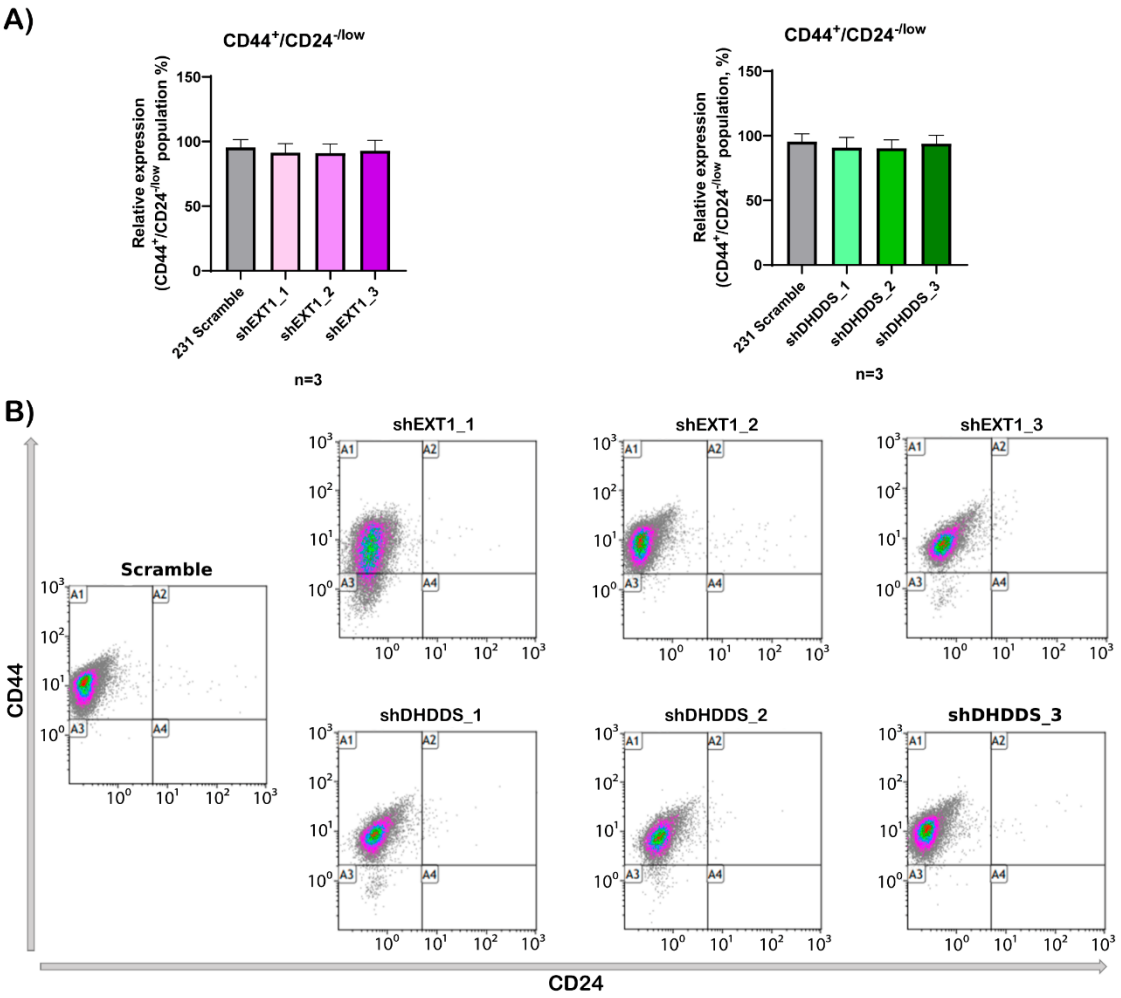



**Figure 47: Analysis of CD44+/CD24-/low population in EXT1- and DHDDS-silenced MDA-MB-231 cells under 2D condition. A)** Bar plots of the percentage of $CD44^{+}/CD24^{-/low}$ subset. Data are represented as the mean (in percentage) ± SD. n=3. **B)** Representative CD44 and CD24 immunostaining dot plots cells of knocked-down 231 cells.

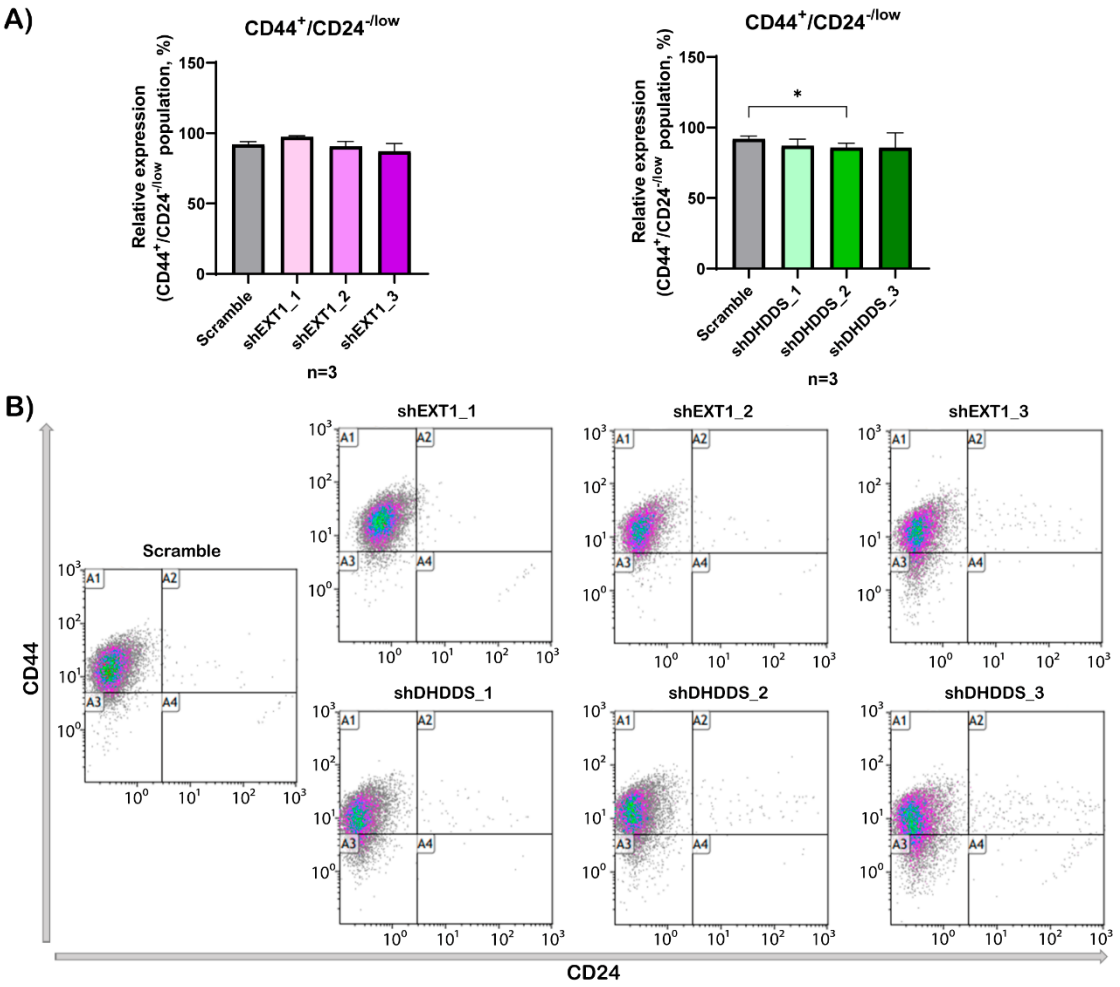



**Figure 48: Analysis of CD44+/CD24-/low population in EXT1- and DHDDS-silenced MDA-MB-231 cells in 3D condition. A)** Bar plots of the percentage of $CD44^{+}/CD24^{-/low}$ subset. Data are represented as the mean (in percentage) ± SD. n=3. **B)** Representative dot plots of double-marked immunostained cells with CD44 and CD24 antibodies.

Since there was no significant change in the stem-like $CD44^{+}/CD24^{-/low}$ population, we next aimed to study the expression of stemness-related genes in silenced cells grown in suspension to determine whether our targets might regulate these genes. The genes analyzed were CD44, SOX2, NANOG, OCT3/4, KLF4 and MYC.

As illustrated in Figure 49, the heatmap revealed that only few genes tended to show reduced expression of MYC when EXT1 was knockdown. KLF4 and CD44 also showed a decrease in the RNA levels in two of the EXT1-knocked down cells analyzed (Figure 49). A similar pattern was observed in cell lines where DHDDS was silenced, in which only CD44 tended to reduce its expression in two of the cell lines analyzed. Despite these apparent differences in gene expression among the cell lines, there were no statistically significant differences compared to the scramble control cell line.

To conclude, the silencing of our targets tended to slightly decrease the $CD44^{+}/CD24^{-/low}$ population but only when silenced-cells were cultured in 3D. Furthermore, the expression of stemness-related genes did not show significant changes in suspension in the EXT1- and DHDDS-knockdown cells, suggesting that there may be another mechanism through which our targets exert their role in stemness.

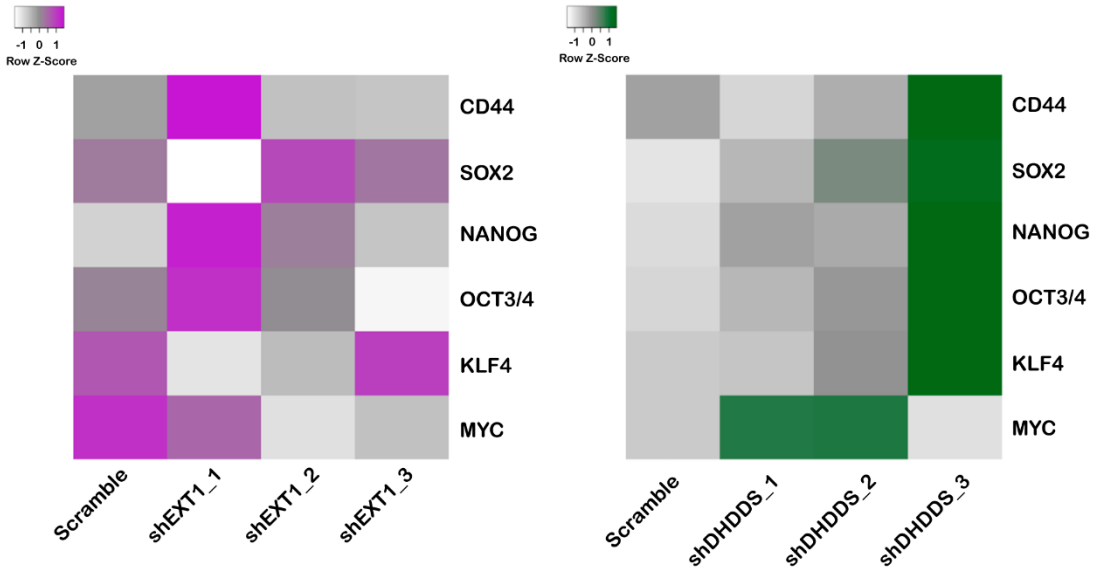



**Figure 49: Changes in RNA expression in EXT1- (left) and DHDDS-silenced cells (right) when cultured in suspension.** Colour scale in each heatmap and represents the mean of 3 independent experiments.

# 3. Validation of EXT1 and DHDDS in TNBC cell lines

## 3.1 Gene expression in transformed and non-transformed cell lines

Once established the implication of EXT1 and DHDDS targets in the stem-like characteristics, we aimed to explore their behaviour in other triple-negative breast cancer cell lines and a normal, non-transformed cell line.

We cultured HS578T, HCC70, and MCF10A in both 2D and 3D conditions to examine their gene expression patterns by RT-qPCR.

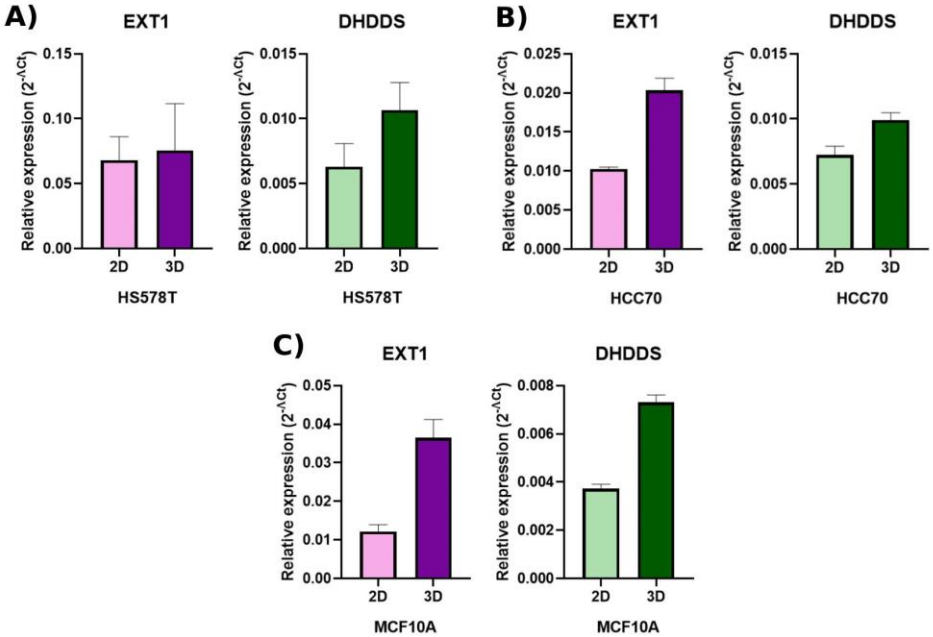



**Figure 50: RNA expression of different basal cell lines in 2D and 3D.** Bar plots of RNA expression in HS578T (A), HCC70 (B) and MCF10A (C). **A)** Data are represented as mean ± SD. n=2. **B and C)** Data of one experiment are represented as the mean of technical triplicates ± SD.

In the three cell lines, the expression of both EXT1 and DHDDS genes increased under suspension conditions (Fig. 50). Notably, DHDDS expression was higher in both transformed cell lines (Fig 50A and B, right panels) compared to MCF10A cells (Fig. 50C, right panel). A similar trend was observed for EXT1, except in HCC70.

Despite the increase in gene expression of both targets in 3D condition in normal non-transformed MCF10A cell line, screening did not indicate their essentiality for the stem cell capability of these cells. Therefore, we decided to study the role of both genes in HS578T and HCC70 to check their crucial role in stemness.

## 3.2 Knockdown of EXT1 and DHDDS in HS578T and HCC cell lines

Similar to the approach with MDA-MB-231 cells, EXT1 and DHDDS were constitutively silenced using the same three antisense sequences. This knockdown was verified by measuring RNA expression levels.

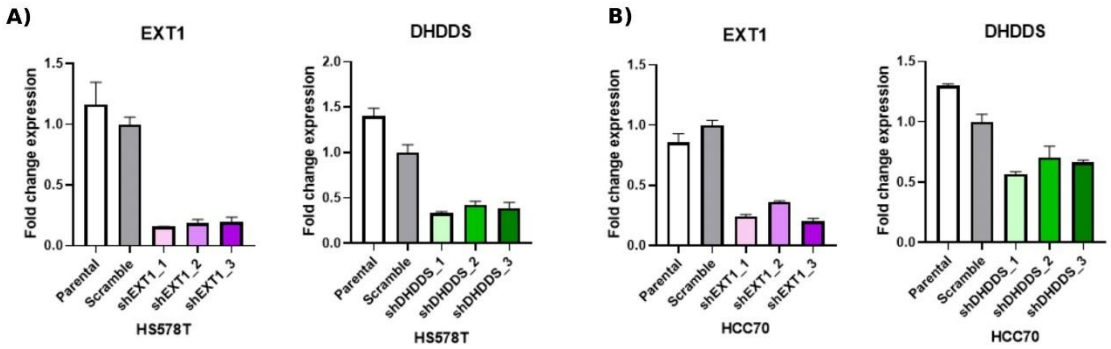



**Figure 51: Knockdown of EXT1 and DHDDS conducted in HS578T (A) and HCC70(B).** Bar plots data are represented as the mean of technical replicates in fold change ± SD.

Hence, the RNA levels in HS578T (Fig. 51A) and HCC70 (Fig. 51B) showed a reduction in gene expression of about 50% in HCC70 DHDDS-silenced cells and higher in all the others. As with the MDA-MB-231 cells, parental cells were also evaluated as controls for the transduction process.

## 3.3 Role in stemness of EXT1 and DHDDS in HS578T and HCC70 cell lines

To verify that both EXT1 and DHDDS play an important role in TNBC, all EXT1- and DHDDS-silenced cells of both cell lines were cultures in 3D.

Starting with HS578T cells, the EXT1 knockdown showed a significant decrease in sphere formation in the first antisense and a tendency in the rest of them. Regarding volume, there were no statistical differences among the silenced cells and the scramble control (Fig. 52A). The reduced expression of DHDDS led to a significant drop of the efficiency of tumorsphere capacity in all sh groups, although their volume remained unchanged with no significance (Fig. 52B).

For HCC70 cells, EXT1 silencing tended to reduce tumorsphere formation, with no significant change in volume (Fig. 53A). Similarly, DHDDS knockdown showed a tendency to decrease tumorsphere formation capacity, without altering the sphere volume (Fig. 53B).

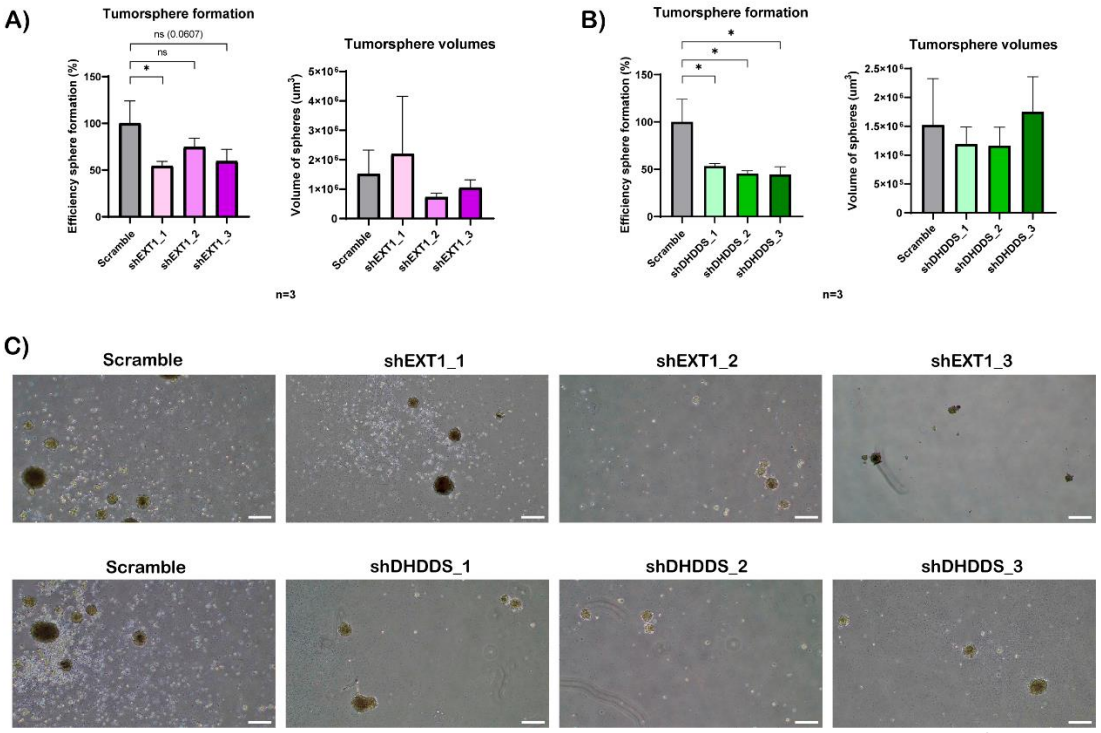



**Figure 52: Ability for tumorsphere formation in the EXT1 and DHDDS-silenced HS578T cells. A and B)** Bar plots of the efficiency sphere formation (left panels) and the sphere volumes (right panels) for the EXT1 knocked-down cells (A) and DHDDS-silenced cells (B). Data are represented as the mean (in percentage for the efficiency) ± SD. n=3. *p-value<0.05, ns=non-significant**. C)** Representative tumorsphere images taken from silenced cells as indicated, using 4X magnification. White scale bar in each image depicts 250 µm in length.

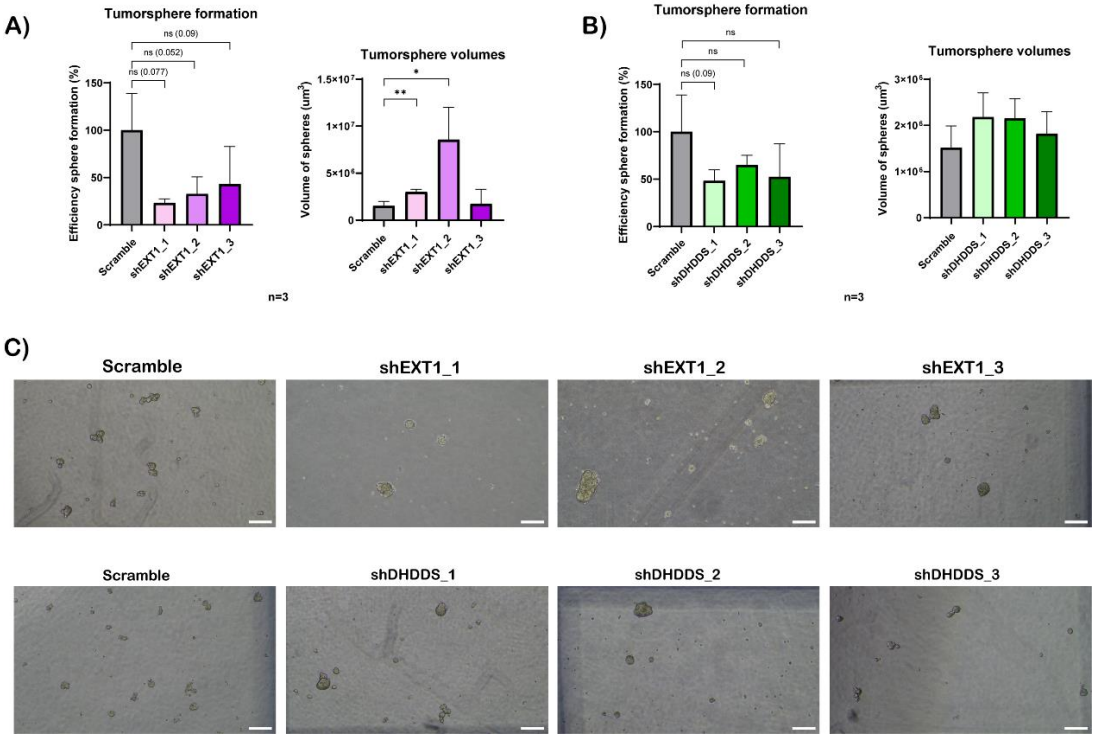



**Figure 53: Ability for tumorsphere formation in the EXT1 and DHDDS-silenced HCC70 cells. A and B**) Bar plots of the efficiency sphere formation (left panels) and the sphere volumes (right panels) for the EXT1 knocked-down cells (A) and DHDDS-silenced cells (B). Data are represented as the mean (in percentage for the efficiency) ± SD. n=3. *p-value<0.05, **p-value<0.01, ns=non-significant. **C)** Representative tumorsphere images taken from silenced cells as indicated, using 4X magnification. White scale bar in each image depicts 250 µm in length.

## 3.4 Analysis of ALDH⁺ population in EXT1- and DHDDS-silenced HCC70 cell lines

To verify the observed reduction in the cancer stem-like characteristics in HCC70 using the tumorsphere assay, we sought to determine if the ALDH+ population changed when our targets were knockdown. This approach was carried out only in this cell line, as it had the highest ALDH+ population,

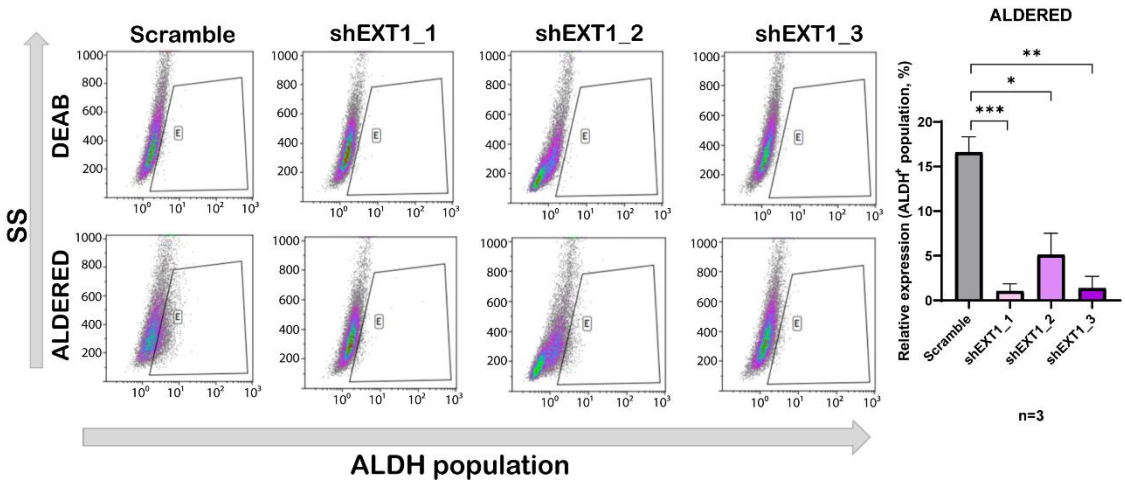



**Figure 54: Analysis of ALDH+ population by FACS in EXT1-silenced HCC70 cell lines.** Representative images of ALDH⁺ cells dot plots for FACS in each cell line. The ALDH inhibitor DEAB was used to establish the baseline of the fluorescence. The bar plot at the right represents the percentage of ALDH⁺ population in the different control and EXT1-silenced cell lines. Data are depicted as mean ± SD. n=3. *p-value<0.05, **p-value<0.01, ***p-value<0.001.

For the knockdown of EXT1, all silenced HCC70 cells significantly decreased the ALDH⁺ population (Fig. 54). Similarly, when DHDDS was silenced in HCC70, the ALDH⁺ subset was also significantly reduced, showing almost a complete remission of this population (Fig. 55).

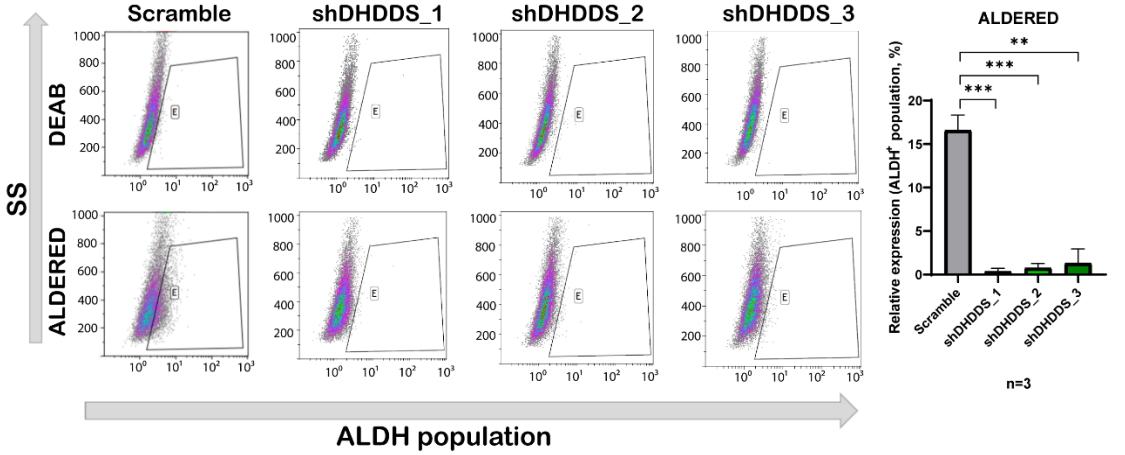



**Figure 55: Analysis of ALDH+ subset by FACS in DHDDS-silenced HCC70 cell lines.** Representative images of ALDH⁺-stained cell dot plots for FACS in each cell line. The ALDH inhibitor DEAB was used to establish the baseline of the fluorescence. At the right, bar plot represents the ALDH⁺ population in the different cell lines. Data are depicted as mean ± SD. n=3. **p-value<0.01, ***p-value<0.001.

All these results demonstrated that EXT1 and DHDDS play an important role in multiple TNBC cell lines, confirming that these proteins are essential for the stem-like phenotype in TNBC *in vitro*.

# 4. *In vivo* validation of EXT1 and DHDDS role in tumor progression

Once confirmed the role of EXT1 and DHDDS *in vitro*, the next step was to validate their function in mice, as one of the stem cell capacities is the tumor initiation in mice model. For this, MDA-MB-231 cell line was used. Two EXT1- and DHDDS-silenced MDA-MB-231 cells were transduced with luciferase vector to enable *in vivo* cell tracking in mice.

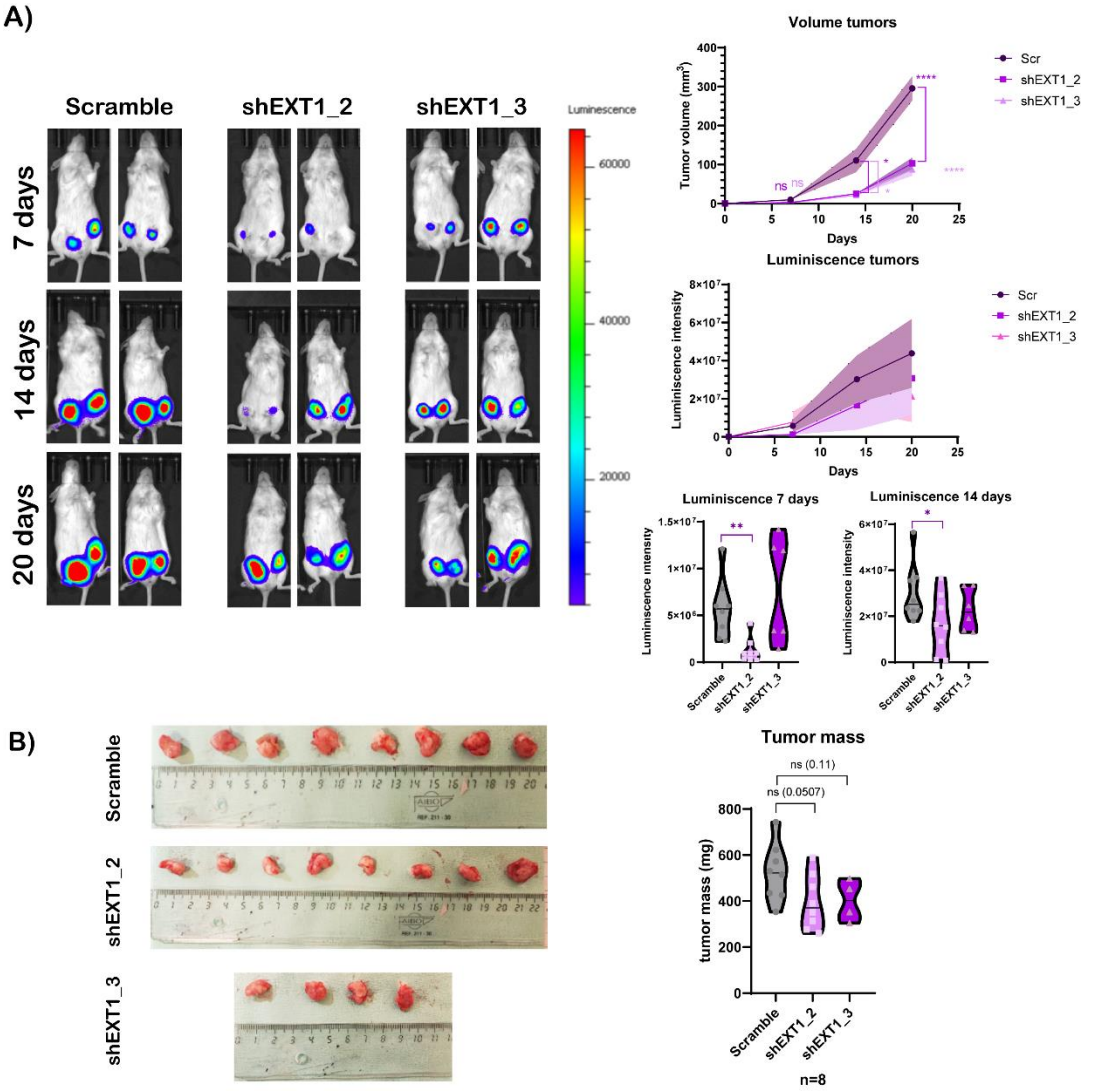



**Figure 56: In vivo experiments in immunocompromised NSG mice using EXT1-silenced MDA-MB-231 cells. A)** Representative pictures of IVIS imaging system for each group (left). At the right, curves of tumor volumes (top) and luminescence intensity emitted by tumor cells (middle). Luminescence at 7 and 14 days is represented using violin plots (down). Data are represented as mean ± SEM. n=8. **B)** Images of excised tumors from mice, showing their measures. Violin plot shows the mass of excised tumors. Data are indicated as mean of the tumor mass. Ns=non-significant, *p-value<0.05, **p-value<0.01, ****p-value<0.0001.

Cells were injected into the mammary fat pad of female immunodeficient NSG mice. These mice, also named as NOD scid gamma mice, are immunocompromised models that were selected to study human engraftment in mice to lack a functional immune system (Shultz et al., 2007). This choice was made to prevent the rejection that might occur with immunocompetent mice, given that the experiments involved the injection of a human cell line.

Four animals were used in each group in accordance with the 3R reduction principle. A total of 500,000 cells were injected into each of the fourth mammary fat pads, leading to a sample size of n=8. Mice were revised weekly after cell injection. Tumor volume was measured using a caliper, and mouse weight was monitored to ensure there was no significant weight loss due to the tumor, in line with the refinement principle to minimize animal suffering. Furthermore, the luminescence of the tumor cells was also quantified using IVIS Lumina XR optical imaging system, not only to follow up the tumor growth but also to detect the metastasis appearance.

Tumors in mice injected with EXT1-silenced cells grew significantly less than those in the control scramble group. The luminescence measurement shows a trend which did not result in a significant difference among the groups (Fig. 56A). Nevertheless, there was statistical difference at days 7 and 14 in the second antisense sequence. Resected tumors from the knockdown groups tended to have lower tumor mass, as it was also showed in the size observed in the images taken (Fig. 56B).

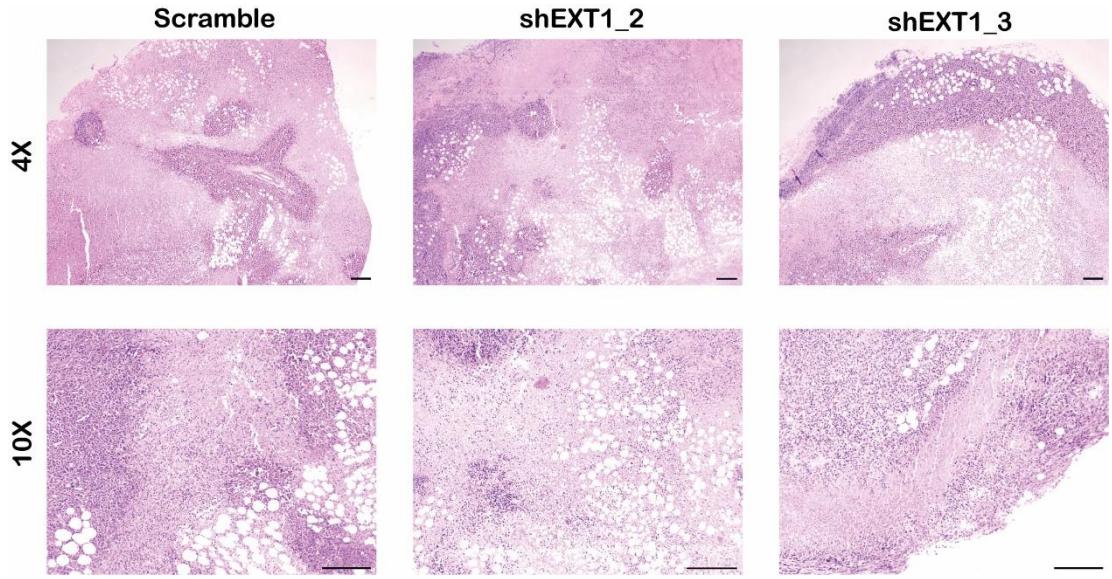



**Figure 57: Hematoxylin/eosin staining of scramble and shEXT1 tumors at 4X and 10X magnification.** The scaled bar in each image represents 2 mm

Hematoxylin/eosin staining of the scramble and shEXT1 tumor groups revealed no visible differences in either composition or structure (Fig. 57).

Regarding the depletion of DHDDS, tumor volume was significantly decreased when DHDDS-silenced cells were compared with scramble control in mice, even though the luminescence intensity again did not show any statistically significant difference (Fig. 58A), only in the antisense sequence 1 at day 7. In addition, tumor mass was generally lower in the DHDDS knockdown groups, with a significant reduction in the shDHDDS_3 group, which was also evident in the images of the excised tumors (Fig. 58B). One mouse from each DHDDS-silenced group was dead before the day of tumor removal.

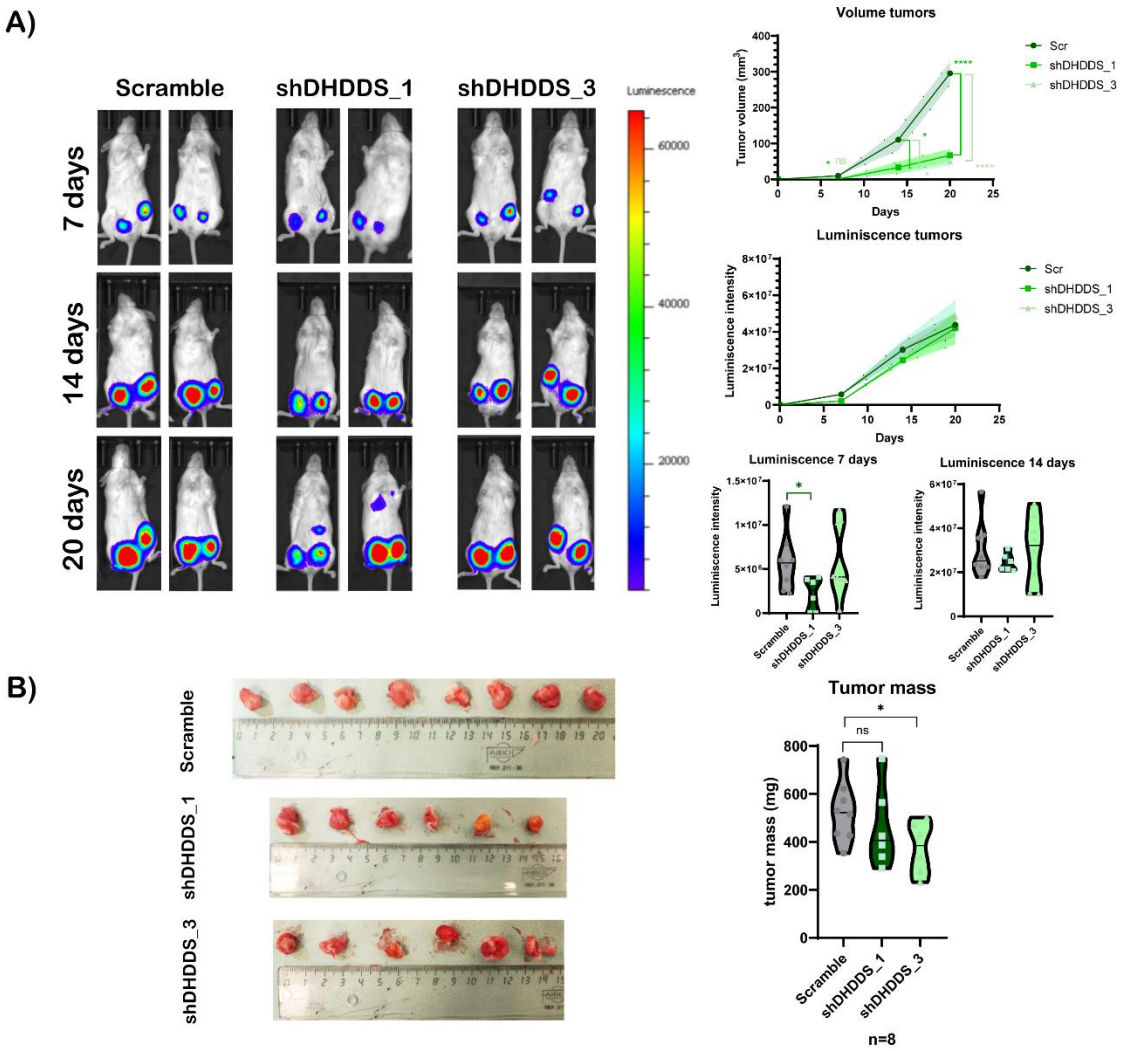



**Figure 58: In vivo experiments in immunocompromised NSG mice using DHDDS-silenced MDA-MB-231 cells. A)** Representative pictures of IVIS imaging system for each group (left). At the right, tumor volume curves (top) and luminescence intensity emitted by tumor cells (middle). Luminescence at 7 and 14 days is represented using violin plots (down). Data are represented as mean ± SEM. n=8. **B**) Images of excised tumors from mice, showing their measures. Violin plot shows the mass of excised tumors. Data are presented as mean of the tumor mass. Ns=non-significant, *p-value<0.05, ****p-value<0.0001.

After hematoxylin and eosin staining was conducted in mice tumors, no visible differences were observed in the tumors across all groups (Fig. 59).

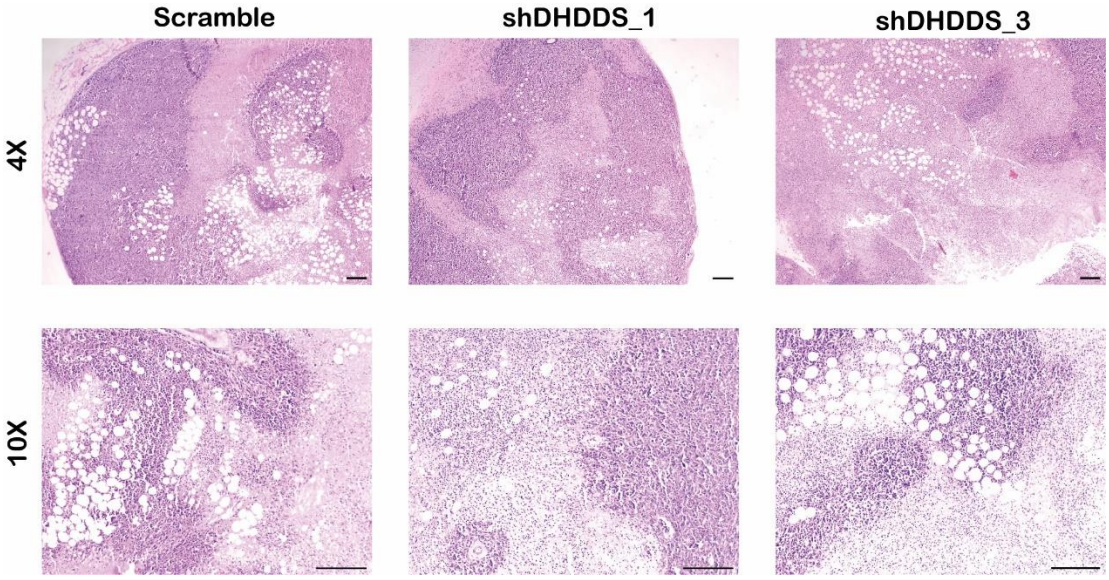



**Figure 59: Hematoxylin/eosin staining of scramble and shDHDDS tumors at 4X and 10X magnification.** The scaled bar in each image represents 2 mm.

These experiments confirmed that both EXT1 and DHDDS have and implication in the tumor progression properties of the malignant cells *in vivo*. Therefore, all this research demonstrated that EXT1 and DHDDS play an important role not only in stemness in TNBC, as shown by *in vitro* experiments but also in tumor progression through *in vivo* studies, pointing at them as promising drug targets.

# 5. Unveiling the mechanistic effects of underlying glycosylation in stemness

To confirm that abnormal glycosylation caused by EXT1 and DHDDS supports the stemness of tumor cells, this glycosylation was deliberately modified to assess its impact on stem cell properties.

## 5.1 Role of Heparin sulfate in cell stemness

Firstly, the amount of heparan sulfate was quantified by FACS using a specific antibody in MDA-MB-231 control and EXT1-knockdown cells. As shown in Fig. 60A and B, there was a significant drop of heparan sulfate-positive population up to almost 60% in the EXT1-silenced cells. Additionally, no changes between parental and scramble cell line were observed.

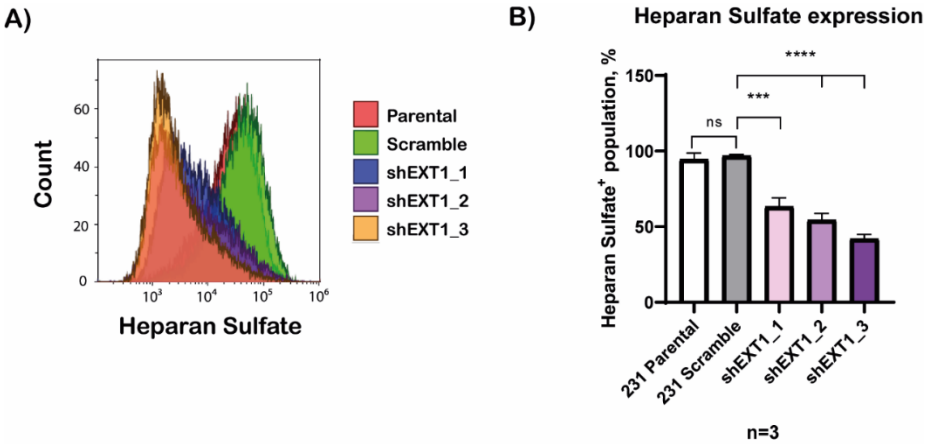



**Figure 60: Heparan sulfate detection by FACS. A)** Representative image of overlay histogram plot obtained when heparan sulfate was determined in MDA-MB-231 cell lines. **B)** Bar plots of heparan sulfate-positive population assessed by FACS. Data are represented as mean ± SD. n=3. n.s. non-significant, ***p-value<0.001, ****p-value<0.0001.

To achieve the objective of altering cellular glycosylation to mimic the effects of EXT1 silencing, we first targeted the removal of sugars forming heparan sulfate. For this, 231 cells were treated with heparinase III. Two different doses were tested, resulting in a dose-dependent reduction in heparan sulfate detection (Fig. 61A and B). Consistent with the observed effects of EXT1 downregulation, the 100 ng/mL dose effectively replicated the impact of EXT1 silencing.

Subsequently, 231 cells were pretreated with heparinase to remove heparan sulfate glycosylation before being plated for tumorsphere formation, with the treatment dose maintained during culturing. As shown in Fig. 61C, there was a trend towards reduced tumorsphere formation efficiency, suggesting a potential causal link between aberrant glycosylation and the maintenance of stemness capacity.

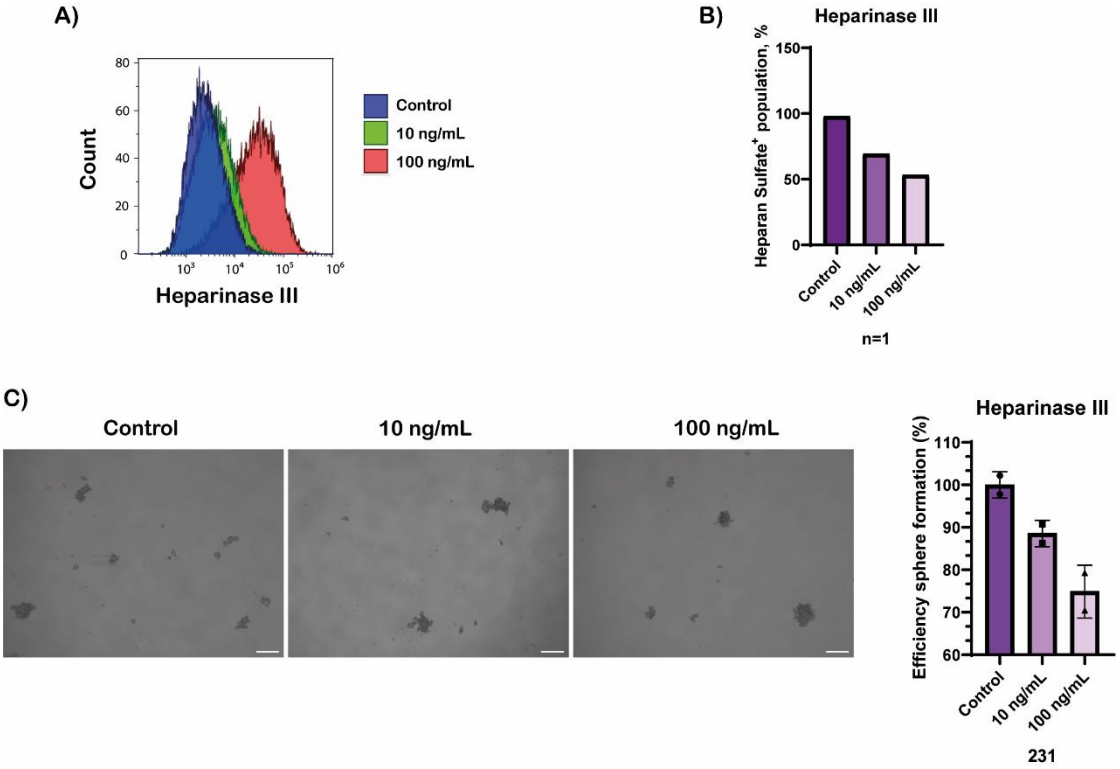



**Figure 61: Heparinase treatment and its effect in stemness. A)** Image of overlay histogram plot obtained when heparinase III was used to treat MDA-MB-231 cell line. **B)** Bar plots of heparan sulfate-positive population assessed by FACS and their modulation using heparinase III. N=1. **C)** Bar plot of the sphere formation efficiency using heparinase III in MDA-MB-231 cell line (right panel). Representative tumorsphere images taken from silenced cells as indicated, using 4X magnification. Scale bar in each image depicts 250 µm in length. N=2.

## 5.2 Impact of N-glycosylation in cell plasticity

Since DHDDS functions in the dolichol phosphate (Dol-P) synthesis which is essential for N-glycosylation and no inhibitory drug specifically targets DHDDS, we aimed to inhibit N-glycosylation downstream of Dol-P using Tunicamycin. This drug inhibits GlcNAc phosphotransferase, which catalyzes the transfer of N-acetylglucosamine-1-phosphate to dolichol phosphate in the first step of protein N-glycosylation (Fig. 62).

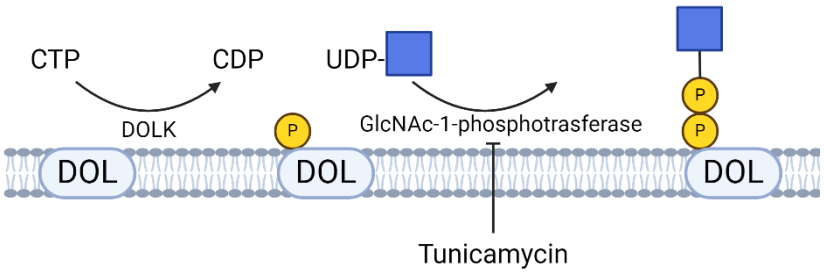



**Figure 62: The initial step of N-glycosylation.** Dolichol is first phosphorylated by dolichol kinase (DOLK), and GlcNAc (blue square) is then added by GlcNAc-1-phosphotransferase. This enzyme can be pharmacologically inhibited by tunicamycin resulting in the global inhibition of N-Glycosylation. Image created using BioRender.

First we assessed the effect of Tunicamycin on cell viability using MDA-MB-231 parental, scramble and DHDDS-silenced cells. As shown in Figure 63A, neither the untreated cells nor those treated with 10 ng/mL displayed notable differences in growth across the cell lines. However, at a 50 ng/mL, both parental and control cells exhibited slower growth, with a more

pronounced effect in DHDDS-knockdown cells. Western blot analysis revealed that DHDDS expression increased at the highest tunicamycin dose (Fig. 63B).

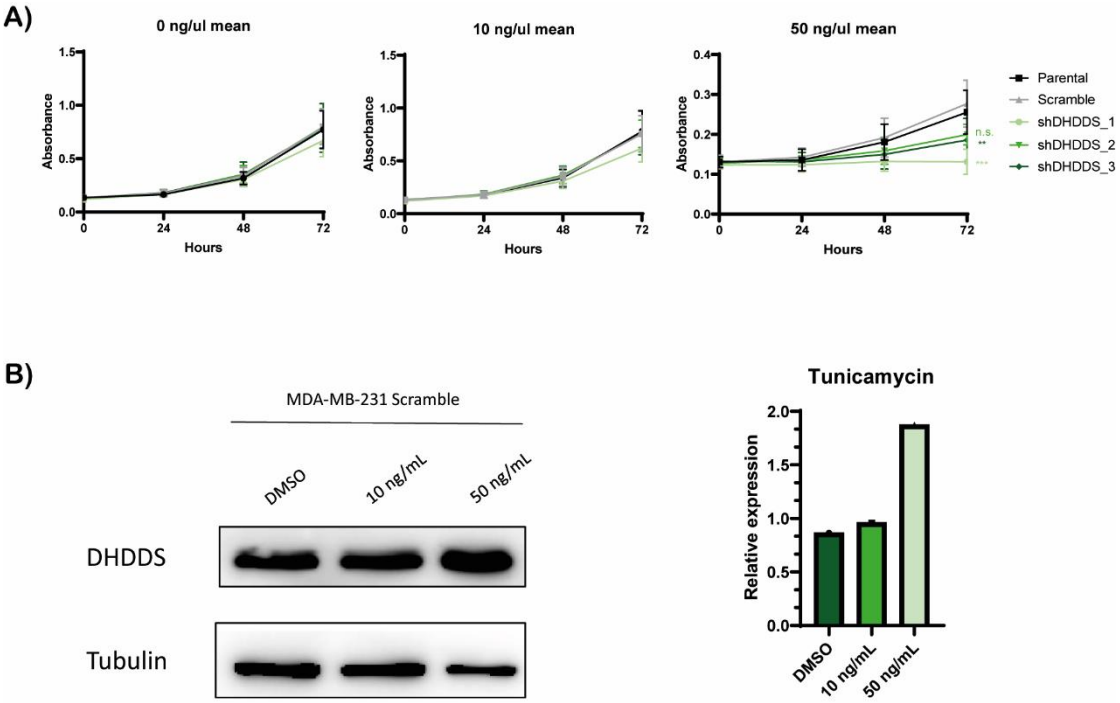



**Figure 63: Impact of tunicamycin treatment in MDA-MB-231 cells. A)** MTT assays showing the viability of cells after tunicamycin treatment. Untreated and 10ng/mL doses showed no significant differences among groups. In 50 ng/mL concentration, statistical differences are depicted in the graphs. n=3. n.s. non-significant, **p-value<0.01, ***p-value<0.001. **B)** Protein expression of DHDDS measured by western blot. Tubulin was used as a loading control. Bar plot represents the quantification of WB performed by Fiji software.

To analyze the effect of N-glycosylation on stem cell properties, the efficiency of sphere formation was analyzed on MDA-MB-231 cells treated with Tunicamycin. A 24-hour pre-treatment with tunicamycin was administered at both doses, 10 and 50 ng/mL in parental MDA-MB-231 cell line, before culturing the cells in suspension. Following this, the cells were seeded in 3D culture for 7 days while maintaining the respective tunicamycin doses.

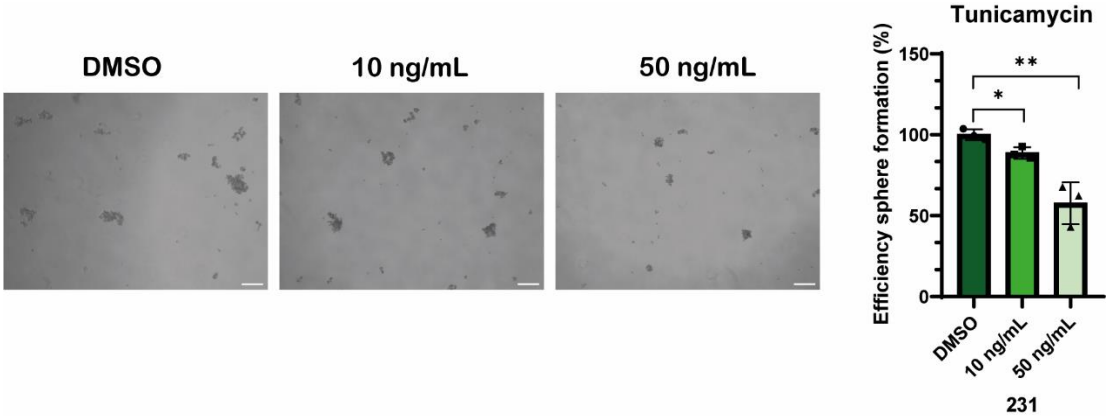



**Figure 64: Changes in the ability for tumorsphere formation in tunicamycin-treated MDA-MB-231 cells.** Representative tumorsphere images taken from treated cells as indicated, using 4X magnification. Scale bar in each image depicts 250 µm in length. Data are represented as the percentage of the mean ± SD. n=3. *p-value<0.05, **p-value<0.01.

As shown in Fig. 64, treatment with tunicamycin resulted in a significant reduction in tumorsphere formation, with the effect being more pronounced at the higher tunicamycin dose.

This indicates that tunicamycin, and therefore the N-glycosylation inhibition, reduces the stem cell potential of MDA-MB-231 cells in a dose-dependent manner.

These results suggest that treatment with tunicamycin, an inhibitor of N-glycosylation, triggers a negative feedback loop in DHDDS, leading to an increase in its expression. This, in turn, impairs the stemness capacity of the treated 231 cells, confirming the involvement of N-glycosylation in maintaining the stemness capacity.

# Discussion

According to recent studies, BC is the one with higher incidence and mortality in women in 2022 (Bray et al., 2024), making it a significant focus of current research. Among BC subtypes, TNBC is clinically the most aggressive subset and the most common in invasive cases. It exhibits a higher proliferation rate and earlier onset, being one with the worse prognosis, since is highly metastatic with significant likelihood of lethality, in spite of accounting for only a small percentage of all BC cases in comparison with other types of BC (Derakhshan & Reis-Filho, 2022; Karim et al., 2023; Leon-Ferre & Goetz, 2023).

Additionally, TNBC is characterised by the high risk of recurrence, with 50% of patients experimenting relapse in early stage within the 3 years after diagnose and 75% of death rates in the 3 months after relapse (R. L. B. Costa & Gradishar, 2017; Lu et al., 2023b). Different researches have focused on the reasons of this recurrence, trying to elucidate its mechanism or new therapeutic approaches (Stewart et al., 2019). This recurrence is widely attributed to a small subpopulation of cells within the tumor named **cancer stem cells (CSC) or tumor-initiating cells (TIC)** (Conde et al., 2022; Kumar et al., 2023) and this idea is supported by the fact that TNBC cells have a greater proportion of this subpopulation (Honeth et al., 2008; Ricardo et al., 2011).

BC tumors, and especially TNBC subset, are known for its significant heterogeneity. This is also attributed to the presence of breast CSC (BCSC) which fluctuate between dedifferentiated and differentiated state and give rise to a heterogeneous group of tumor cells (Z. Guo & Han, 2023; Mahmoud et al., 2022). These cells were firstly described as cancer cells with the ability to form tumors at low density in mice. Moreover, they express certain well-known surface markers such as the $CD44^{+}/CD24^{-/low}$, EpCAM or PROCR or the presence of $ALDH^{+}$ population, among others cell markers (Al-Hajj et al., 2003; Dittmer, 2018; Ginestier et al., 2007; Zheng et al., 2021). They are able to self-renew (Nalla et al., 2019), metastasize (H. Liu et al., 2010), even resist to drug treatments (L. He et al., 2021), features that are inextricably linked to the relapse of patients since chemotherapy cannot target this specific subset of cells and the recurrence may appear in distant organs (metastasis).

There is not a specific type of BCSC that possesses all the markers; instead, there are various subsets that can be identified by different markers, which often overlap (Al-Hajj et al., 2003; Ginestier et al., 2007; Wright et al., 2008). Together with the idea that they are not in a “static” state but instead shift between epithelial and mesenchymal phenotypes, exhibiting high plasticity, makes it difficult to study them extensively and, therefore, to identify a druggable target for their specific elimination (S. Liu et al., 2014). Nevertheless, different groups have studied different approaches to target BCSC, such as through CD44 (McClements et al., 2013), the classical signalling pathways as NOTCH or STAT3 pathways or designing different immunotherapeutic strategies as chimeric antigen receptor (CAR)-T cell or vaccines (A. Sharma et al., 2012; R. Zhang et al., 2022). In spite of the great efforts, further investigations are required to deepen in the BCSC biology and their therapeutic potential.

Hence, this thesis proposes the protein glycosylation as the biological process that may be a source of stemness and plasticity in BC tumor cells.

Protein glycosylation play an important role in breast tumor biology. Aberrant protein glycosylation promotes cell proliferation (H. B. Guo et al., 2010), cell-cell interaction and migratory capacity (S. Lin et al., 2002), invasiveness (Tu et al., 2017) or immune evasion, such as through the glycosylation and stabilization of PD-L1 (C. W. Li et al., 2016). Indeed, protein glycosylation is involved in all hallmarks of cancer not only in BC but in other tumor types

(Munkley et al., 2016). Thus, it is reasonable that this altered process has an impact in the tumor plasticity and stemness.

Most of CSC markers are glycoproteins that can undergo modifications in their glycosylated patterns, which may influence in CSC functions (Barkeer et al., 2018). Some studies show a loss of CD133 in colon CSC upon cell differentiation (Kemper et al., 2010) or highly sialylated CD44, which was also used to target CSC (T. Khan et al., 2020). Nonetheless, the role of aberrant glycosylation in promoting a stem-like phenotype has not been thoroughly investigated. Furthermore, CSC heterogeneity at both inter- and intratumoral levels adds significant complexity to target this little subset of cells, with the additional fact that CSC fluctuate among different states (T. Khan & Cabral, 2021). Aberrant protein glycosylation may give a logical explanation to the acquisition of stem properties, as it is a reversible biological process that could describe the plastic nature of CSC and is influenced by several factors. Therefore, it seems crucial to identify and understand the essential protein glycosylation players for the BCSC, encouraging an approach to unravel them.

In order to discover which is the particular gene responsible for a specific phenotype, forward genetics has been the main strategy to detect this genotype in a concrete biological context. The use of RNA libraries or cDNA plasmids, in loss- or gain-of-function investigations respectively, enables the study of the whole human genomic DNA in a large scale (Ali Khan et al., 2021; Heynen-Genel et al., 2012). For example, the use of a shRNA library to analyse the whole genome made it possible to identify genes in the mevalonate pathway that make lung and breast cancer cells more susceptible to statin-induced apoptosis (Pandyra et al., 2015). However, since CRISPR technology was adapted for genome editing in 2013 (Cong et al., 2013), this approach has been widely used for screenings. Comparisons between both strategies have shown that CRISPR provides more robust phenotypes, greater sensitivity, fewer off-target effects, and more significant results than shRNA libraries (Deans et al., 2016; Morgens et al., 2016). Consequently, CRISPR screening may be now considered as the gold standard technique for identifying essential genes.

The interrogation in a whole-scale was made throughout CRISPR libraries which have been developed and improved along all these years. The first one was GeCKO library for a dropout screening (Sanjana et al., 2014), and more others followed it as Toronto KnockOut (Hart et al., 2015) or Brunello libraries (Doench et al., 2016). Other libraries were also developed according to the wanted perturbation, as SAM library for CRISPRa screening (Joung et al., 2017) or Dolcetto library for CRISPRi screening (Sanson et al., 2018). In fact, CRISPRi screens are based on the same idea as shRNA screens. Hence, CRISPRi can overcome the results from shRNA and the difference between CRISPRi and CRISPR KO may be minimal. Indeed, the study from Sanson, *et al.* showed no significant differences among the screenings applied either with CRISPRi and CRISPR KO libraries (Sanson et al., 2018).

In spite of the significance of examining the whole-genome for a particular biological question, different libraries targeting specific subset of genes have been designed, such as those targeting genes related to DNA damage response and DNA repair (D. Su et al., 2020) or epigenetic genes (Williams et al., 2020). Thus, this thesis is based on the creation by our group of a **novel CRISPR library** which targets specifically genes related to protein glycosylation according to Gene Onthology term “protein glycosylation”. In 2021, Zhu, et al. published a library similar to the one described here (Y. Zhu et al., 2021). Nonetheless, while they share similarities, there are also notable differences. Both libraries focus on genes related to glycosylation, include negative controls, and contain approximately 10 sgRNAs per gene. However, our GlycoCRISPR library

specifically targets protein glycosylation genes, whereas Zhu's GlycoLibrary targets all genes involved in the biological glycosylation process, including glycolipids and GPI-anchored carbohydrates. This makes our library more focused on a specific pathway and simpler to analyse due to lower number of sgRNAs. Additionally, our library includes positive controls related to stemness, allowing us to study the involvement of protein glycosylation-related genes in a stem-like phenotype (Y. Zhu et al., 2021).

The presence of $CD44^{+}/CD24^{-/low}$ and $ALDH^{+}$ populations in breast cancer cell lines used in this study is well-studied (R. J. Kim et al., 2013; Ricardo et al., 2011; Sheridan et al., 2006; Y. Su et al., 2016; Vikram et al., 2020) and confirmed here to establish if the assessment of these markers may be a good a strategy to evaluate the stem-like population across cell lines. As they presented heterogeneity in their basal expression, other methodology should be found to evaluate the stem-like population. Hence, tumorsphere formation assay was the most suitable approach for the readout of our screening. This strategy enriched BCSC in our BC cell lines as proved here and also used in other studies (Ambrose et al., 2022; J. W. Park et al., 2022). In fact, it was determined as a good technique to screen anti-cancer compounds or CSC properties (Lee et al., 2015; Weiswald et al., 2015) and used to investigate the effect of some drugs as paclitaxel or doxorubicin in tumorsphere from breast tumor cells (Kessel & Chan, 2020), although some research did not see an enrichment in cancer stem-like population when some cells were cultured in suspension (Calvet et al., 2014).

The set up of screening conditions is a labor-intensive that consumes a significant amount of time and materials. It is always necessary to perform different proves to finally obtain the optimal settings to carry out the screening. The main objective was to achieve a multiplicity of infection (MOI) of 0.3 or less to ensure that only one guide enter to each cell (Shang et al., 2017). This approach was followed by different groups in their CRISPR screenings (C. Wang et al., 2022; T. Wu et al., 2021), even infecting MCF10A (L. Yin et al., 2024).

Initially, we faced some troubles with the MDA-MB-231 transduction. Those problems were overcomed with the use of PM2.G and psPAX2 helper plasmids as other group employed in its study (H. Liu et al., 2021) and the application of DEAE-dextran polycation instead of polybrene which was normally utilised by our lab and others. DEAE-dextran demonstrated a higher virus infection efficiency (Denning et al., 2013). In both 231 and MCF10A screenings, we modified different parameters to finally obtain the optimal conditions, such as the polycation or the amount of virus media employed. These conditions were also tested by Y. Kim et al. to increase the efficiency in immune cells (Y. Kim et al., 2023). The viability of virus particles through freeze-thaw cycles was also tested as a logistical issue, checked also by Kumru et al., 2018. These cycles affect viral supernatant, observing a decrease in the efficiency of infection, as it happened in all of our tries. Performing a spin step is supposed to improve the transduction efficiency as it did for infecting CAR-T cells (Rajabzadeh et al., 2021). In contrast, Lo Prestin et al. observed that the spin process, which they named as spinoculation, did not improved the infection efficiency as it was also observed in the set up proves of 231 screening (Lo Presti et al., 2021). Thus, different parameters should be established to perform CRISPR screenings in different cell lines.

The coverage of the library in the screening is also a critical point. The size of the library has to be considered, as well as the type of screening (positive or negative). Small customized libraries provide strength data and more sensitive when the coverage is higher (K. Lin et al., 2022). As reviewed, positive selection screenings require a representation of 100-200X per gene and the negatives are proned to need 500-1000X of coverage (Bock et al., 2022). As used in genome-wide CRISPR KO library screening in a study to research the chemoresistance in cancer (Zhong et

al., 2024), we also applied a coverage of our library of 300X. However, as our library contains 10 sgRNA per gene, a total representation per gene was 3000X, which give us a great tool to genetically screen the biological question with powerful outcomes. Other studies fluctuate the coverage between the 100X (Kiessling et al., 2016) and the 1000X (Mathiowetz et al., 2023).

Once runned the screenings and sequenced, all the results were analysed by bioinformatic tools. In our case, we selected the CEDA package to conduct the analyses, as it provides more control in false positive rates, has more sensitivity and offer a better recognition of genes with moderate fold change, taking into account the gene expression data (Zhao, Yu, et al., 2022). Firstly, the negative controls were checked, observing little variation across conditions and replicates. This strategy was also applied by Covarrubias et al., who displayed the invariability of negative control sgRNAs through a scatter plot and their correlation between time points (Covarrubias et al., 2020). Positive controls were analysed in CRISPR screenings to assess the correct performing of the CRISPR screening. Hart et al. supplies a list of essential genes that can be employed to this purpose (Hart et al., 2017). In fact, Schmierer et al. used ribosomal proteins to validate the loss of them after conducting CRISPR screening (Schmierer et al., 2017). In our case, positive controls, which are genes related to stemness phenotype, were disappearing across different generations.

In order to apply a selection criteria, groups performing genome-wide CRISPR screenings have a large sets of genes which vary among conditions. Therefore, the top enriched or the bottom depleted candidates, regarding log2 fold change, can be considered, as it was in an investigation to elucidate combinatorial drug targets against pancreatic cancer (Szlachta et al., 2018). Nevertheless, other study established a criteria similar to ours, in which they apply a log2 fold change of 0.5 with at least 2 sgRNA per gene (Takashi Ishio et al., 2022). We also added that FDR was 0.05 to select hits with higher confidence, as explained by Mathiowetz et al. (Mathiowetz et al., 2023). Establishing a lower FDR determines a robust statistical power.

After obtaining the hits, selection was primarily conducted through *in silico* analysis using public databases as cBioPortal for Cancer Genomics as an initial filter, followed by a review of related published literature and available inhibitors targeting them. Later on, extensive bioinformatic research and clinical data was evaluated for the final choice. Therefore, from 10 initial candidates we selected 3 of them. Although the amount of hits were not as much as other screenings, their validation may take too much time for this thesis, but can be considered for further investigations, even if the selected will not obtain a profitable results.

The *EXT1* protein coding gene that translate the exostosin glycosyltransferase 1 is responsible for the heparan sulfate chain elongation and appeared to be essential for stemness in 231 cells. It is involved in the development of osteochondromas (Bovée, 2008) and the etiology of hereditary multiple exostoses (Pacifici, 2017) with its alteration. In this regard, although it has been considered as a tumor suppressor (Mccormick et al., 2000), the role of EXT1 as tumor driver was studied in different cancer models, including breast (W. Kong et al., 2021; Solaimuthu et al., 2024). Furthermore, bioinformatic analyses and clinical data demonstrated to be the best hit, observing the negative clinical outcome concerning overall and relapse-free survival in patients with higher expression of EXT1 and a higher recurrence rate. Moreover, a positive correlation between stemness and protein glycosylation signature was observed, suggesting that EXT1 is probably linked to a stem-like phenotype and a higher cell protein-glycosylated status. Additionally, due to the fact that the screening was performed in TNBC cell line, it helped to determine the higher EXT1 expression in basal tumor subtypes.

The subunit of the dehydrodolichyl diphosphate synthase, DHDDS, was also shown to be required for stemness maintenance in 231 cells. All complex is necessary for the formation of dolichol phosphate, a lipid glycocarrier for N-glycosylation. Even though it was correlated with fatal congenital disorders of glycosylations or retinitis pigmentosa, little was known about this gene in cancer. Only two studies have noted its relevance in ovarian cancer and TNBC, where it is implicated as circular RNA (S. Cui et al., 2022; N. Li et al., 2020). Hence, the lack of available research of it made it attractive for the investigation. The clinical parameters of overall and relapse-free survival studied in this thesis support its further study. Strikingly, the recurrence ratio was not significant, probably because of the use of median distribution, whereas in the Kaplan-Meier plot is distributed with the best cutoff. Interestingly, the stemness signature was negatively and significantly correlated with the higher DHDDS expression, contrary with what we obtained in the screening. In addition, DHDDS are not preferentially expressed in any molecular subtype, which could explain the similar expression of this gene across all subsets.

ST3GAL1 protein is encoded by *ST3GAL1* gene which adds sialic acids to protein substrates. As it was shown to be essential for stem-like phenotype in 231 cells in our screening, and the strong relation of this protein with several cancer types such as ovarian (X. Wu et al., 2018), prostate (Garnham et al., 2024) or breast (H. X. Cui et al., 2016), it could be a good candidate. In fact, different research connected ST3GAL1 with resistance to anti-cancer drugs in ovarian cancer (X. Wang et al., 2017) or with the invasion and adhesion capacity in breast cancer (H. X. Cui et al., 2016), common properties of CSC. The overall and relapse-free survival was worse in patients with higher ST3GAL1 expression, observing a higher significant recurrence ratio with higher ST3GAL1 expression. Nevertheless and surprisingly, stemness signature was not statistically significant, perhaps because the genes included in the signature were not modified by ST3GAL1. The protein glycosylation signature was significantly lower in patients with higher ST3GAL1 expression suggesting a negative feedback regulation. What is more, the ST3GAL1 expression was not correlated to basal subtype, but could be associated with HER2 subtype, even with HER2 expression, as Luminal B subtype contains also the tumors with increased expression (Fig. 41).

The association of our targets with stemness signatures were analysed using the study of Ben-Porath et al. It gathered 13 different overlapping signatures from meta-analysis and found that undifferentiated tumors at histological level overexpressed genes typically found in embryonic stem cells and this signature was correlated with high-grade ER-negative tumors and basal-like subset in breast cancer which had worse clinical outcomes (Ben-Porath et al., 2008). As the study focused on breast tumors and included a comprehensive set of genes, it was particularly suitable for our analysis compared to other studies that, for example, examined stem-related genes from mice (Wong et al., 2008). Other recent investigations also used these signatures from Ben-Porath paper for their profitable research (Chokshi et al., 2024; Maciejewski et al., 2024). In fact, the study elegantly determined the stemness condition by assessing not only the activation of stem-related genes but also the repression of genes associated with differentiation, such as Polycomb sets. For this, this approach helped us to state the correlation of patients with higher expression of our targets with these signatures, identifying that higher EXT1 expression had the most related status to stem-like phenotype. Strikingly and according to what we observed in the previous bioinformatic analysis, DHDDS is strongly associated with a differentiated state (with the activation of differentiation gene sets and the repression of stem-like genes). Similarly, ST3GAL1 did not express undifferentiated gene sets, only some of them. However, as both of them showed an essentiality for the stem characteristics in 231 cells, we decided to further validate them.

The first validation performed was the study of targets expression by RNA and protein levels when cultured 231 cells in suspension. Using RT-qPCR, it was observed an increased expression of all of them in the first generation of tumorsphere and a subsequent drop of it across generations. These results were also confirmed by western blot at protein levels. This pattern indicates that these genes might initially influence cells by modifying them post-translationally, enabling their conversion into BCSC. The effects of these modifications may persist in future generations, making the continued expression of these genes unnecessary in later generations. This idea made us hypothesise that there could be genes more essential than others. The increase of a glycosyltransferase in CSC was also described by Guo, et al. where they observed an enhanced expression of Gnt-V (MGAT) in the compartment of $ALDH^+$ population (H. Guo et al., 2014). Additionally, the supposed state that there could be genes more essential than other might be supported by the study of Schultz et al. where ST6GAL1 protein expression rised along days in suspension because it was necessary but not at early stages in CSC of pancreatic and ovarian cancers (Schultz et al., 2016).

To confirm the implication of our hits in stemness, the down-regulation of them was performed using miR.E technology, a more potent approach than short hairpin RNAs (Fellmann et al., 2013). This technology was also employed in different studies but in its inducible form using doxorubicin (Goto et al., 2024; Ward et al., 2024). Since we did not want to study the temporal effect of our targets, we use the constitutive silencing form. Three different antisense sequences were used to demonstrate that the observed functional effects of target silencing were not stochastic. The result knockdowns were confirmed by mRNA and protein levels in 231 which allowed us to perform the functional validation through mainly tumorsphere formation ability. This approach was also employed by Ring et al. where after the silencing of EP300, an epigenetic modulator, they observed a drop in the tumorsphere formation capacity (Ring et al., 2020).

In this regard, the silencing of EXT1 and DHDDS provoked a significant diminish in the tumorsphere formation efficiency in almost all antisense sequencing, confirming the essential role of these genes in maintaining stemness in 231 cell line. One critical step is the selection criterion for tumorspheres, which was defined as having a diameter greater than 60 µm. This criterion is consistent with those used in other studies (Harbuzariu et al., 2017; Tian et al., 2017) and in a protocol of StemCell enterprise to produce tumorspheres from breast cell lines (*Tumorsphere Culture of Human Breast Cancer Cell Lines*, n.d.). Nevertheless, other studies have used different criterion for tumorsphere selection, with sizes ranging from 50 µm (Bailey et al., 2018) to 100 µm (Ospina-Muñoz & Vernot, 2020; Z. W. Zhu et al., 2018), highlighting the need to standardize these criterion. Unfortunately, due to ineffective silencing of ST3GAL1, the sphere-forming efficiency remained unaffected, leading to the exclusion of this target from further consideration. The volume of the spheres was also analyzed, revealing no significant differences except for one antisense sequence of EXT1 and two of DHDDS, although a general increased trend was observed. These collective results may suggest that in the suspension context, knockdown fails to maintain stemness properties. The stem-like phenotype is typically characterized by slow cycling (Francescangeli et al., 2023); however, silencing these genes might cause a loss of these stem traits, leading the cells to enter a proliferative state according to the tumorsphere volume results.

To further validate the implication of our target with a stem-like phenotype, silenced-231 cells were assessed using surface markers CD44 and CD24. The reduction of $CD44^+/CD24^{-/low}$ population might give us another verification that EXT1 and DHDDS are key players in stemness, as several studies did to check the implication of their proteins in stemness properties (L. Guo et

al., 2023; Ibrahim et al., 2017; S. L. Kim et al., 2023). However, our results showed that there was not a significant impact on the total $CD44^{+}/CD24^{-/low}$ subset when our targets were silenced in 231, neither in 2D nor in 3D. These results suggested that the involvement of our proteins did not influence in the amount of this population but it could do in the glycosylated state of specially CD44, which is a glycoprotein. One interesting experiment would be the detection of glycosylation pattern in CD44 or other proteins after the knocking down.

Next, the gene expression of some stem-related genes in 3D cultures was examined. While the differences were not statistically significant, there was a tendency for EXT1-silenced cells to show a decrease in some stemness genes, whereas DHDDS knockdown cells exhibited the opposite trend. Several factors should be considered. The tumorsphere assay showed considerable variability, leading to high deviation between experiments and making it difficult to detect significant differences. To better observe these differences, it may be necessary to increase the number of independent experiments. Additionally, while these genes are among the most studied, they are not the only ones involved in maintaining a stem-like state. Further experiments assessing other markers, such as CD133 or EpCAM, could provide additional insights into differences in stem-like phenotypes.

To confirm the crucial role of EXT1 and DHDDS in maintaining the stemness population not only in one cell line but across TNBC, two additional TNBC cell lines, HS578T and HCC70, were selected for validation. It is a common routine to corroborate the finding in different cell lines of a specific subtype, as done it in independent studies (Nam et al., 2017; Sato et al., 2022).

To the same extent as 231, EXT1 and DHDDS increased their mRNA in suspension. Hence, the knockdown of EXT1 and DHDDS was carried out in both cell lines, achieving a good silencing. The verification of the silencing through protein levels is necessary to perform for the complete validation.

Again, the validation for stemness properties was performed through tumorsphere forming assay. In HS578T cells, there was a notable decrease in tumorsphere formation with the first antisense sequence of EXT1, and an almost significant trend with other sequences. Silencing DHDDS significantly reduced tumorsphere formation across all antisense sequences. Similarly, in HCC70 cells, although none of the antisense sequences for both genes showed significant differences, there was a clear trend towards reduced tumorsphere forming ability. Tumorsphere volume did not change significantly for EXT1 or DHDDS in either cell line, except for an increase in volume with the first and second antisense sequences of EXT1 in HCC70. Thus, more experiments are required to see significant differences in these cell lines to finally verify the implication of both genes in the maintenance of stemness.

Regarding other validations in these TNBC cell lines, as 231 were similar as HS578T in $CD44^{+}/CD24^{-/low}$ population (Fig. 22), the assay evaluating this subset of cells was not conducted under the assumption that both cell lines would behave similarly. Nonetheless, it would be necessary to confirm this fact. Hence, analysis of $ALDH^{+}$ population was fulfilled in HCC70 has it contains the largest positive ALDH population. After the knockdown, this subset was depleted in all antisense sequences of both EXT1 and DHDDS genes. As such, this confirmed that, even though the tumorspheres forming capacity exhibited a reduced trend in HCC70 cells, EXT1 and DHDDS exerted an important role in stemness maintenance in this cell line. An additional experiment would be the assessment of $CD44^{+}/CD24^{-/low}$ cell subset, although the percentage was lower than 231 and HS578T. In this regard and concerning the existence of two subtypes of BCSC (S. Liu et al., 2014), the downregulation of our genes would likely affect only the

mesenchymal-like BCSC in case of HCC70, and not the epithelial-like BCSC compartment as in 231.

Once performed the validation *in vitro*, we wondered if the silencing of EXT1 and DHDDS could affect the tumor initiating ability in mice as one of the CSC capacities (S. Liu et al., 2014). Therefore, 231-silenced cells were our model to inject in immunodeficient mice. Two antisense sequences were selected for injecting them in mice, specifically those that behaved similarly in tumorsphere formation capability to see homogeneous results. Additionally, testing two of them might reduce firstly the number of animals and latter the possible stochastic effect observed only by one antisense sequence. Thus, after 20 days of experiment, there was a significant difference of tumor volume measured among scramble and silencing of EXT1 and DHDDS. The difference in volume was also noticed when tumors were removed from mice after their euthanasia, as well as through their difference in mass weight, which was significant between scramble and one of the antisense sequences of DHDDS. However, this significant difference was not observed through the luminescence intensity, due to the high variability of the data obtained per group. This experiment finally verified that both genes are then essential for the maintenance of stem-like phenotype in TNBC.

This thesis demonstrated that one of the genes implicated in maintaining stemness in TNBC is EXT1. This research is supported by several studies. In one of them, they found the implication of EXT1 in the stemness of doxorubicin-resistant breast cancer cells from luminal subtype (Manandhar et al., 2017). Additionally, EXT1 was also elucidated to regulate EMT and plasticity in TNBC directly by the effect of the heparan sulfate (HS) (Solaimuthu et al., 2024). Unlike this thesis, that research used an EXT1-KO model to perform the experiments, revealing that EXT1-silenced 231 cells could not form tumors in mice. HS was confirmed to be involved in the epidermal growth factor receptor (EGFR) signalling and in the resistance promotion of EGFR inhibitors in glioblastoma (Ohkawa et al., 2021). Moreover, it has been also examined that EXT1 mediated the role of transglutaminase 2 in stemness of hepatocellular carcinoma (Qin et al., 2023). On the contrary observed, other studies have found that the silencing and methylation of EXT1 induce chemoresistance, a stem-related feature (W. Kong et al., 2021; Pfeifer et al., 2023). Concerning all these studies, the work performed in this thesis enlighten the role of EXT1 in TNBC for the stemness maintenance.

This work also confirmed that DHDDS is directly involved in stemness preservation in TNBC. In other study investigated DHDDS as a transcription gene for HOXB4, which requires of DHDDS for tumor progression and invasion in ovarian cancer (N. Li et al., 2020), but it was not correlated to stem-like phenotype. Furthermore, circular RNA of DHDDS was associated with tumorigenic and invasive capabilities in TNBC (S. Cui et al., 2022). Nonetheless, they did not investigate the expression of the protein itself, which has been done in this thesis.

## Future perspectives

In summary, this thesis proved the significant role of glycosylation genes in stem-related phenotype in TNBC. In this line, a study explored the relevance of the N-glycosylation of EpCAM in stemness maintenance under hypoxic conditions in breast cancer (D. Zhang et al., 2020). Additionally, protein glycosylation can affect different stem-related pathways. As an example, the glycosylation by GALNT1 is necessary to maintain the CSC of bladder cancer through Hedgehog pathway (C. Li et al., 2016). Therefore, our GlycoCRISPR library can give important insights of stem properties related to protein glycosylation and should be published for further use by other groups. In fact, since several malignancies contains aberrant glycosylations as recently reviewed

(M. He et al., 2024), the relation with stemness can be further examine across different types of cancer.

Due to the variability in tumorsphere assay, it could be improved and optimised. The tumorspheres of 231 were tight spheroids as was described by Okuyama et al. (Okuyama et al., 2023). Then, using a matrix like methylcellulose for their growth could be an interesting approach to standardize results (Z. W. Zhu et al., 2018).

This work investigated the role of genes causing specific aberrant glycosylation in the context of stemness. However, due to time constraints, it was not able to examine specific glycosylations on individual proteins in this research. Future work should consider performing a glycoproteome analysis comparing scramble and knockdown cells to identify differentially glycosylated proteins.

To gain a deeper understanding of the connection between aberrant glycosylation and stemness properties, it is crucial to elucidate the mechanistic downstream pathways involved in our target genes. Therefore, experiments should be conducted to remove heparan sulfate in the case of EXT1 or N-glycosylation for DHDDS, followed by comprehensive detection of changes in key stem-related pathways. These experiments could utilize inhibitors for specific pathways or proteins, competitive molecules that prevent certain protein glycosylations, or enzymes that remove glycosidic complexes. For instance, downstream pathways involved in EXT1 signaling might include Wnt signaling, as demonstrated in a lung cancer study by Kong et al. (W. Kong et al., 2021) or the JAK/STAT axis as proved by Solaimuthu et al. (Solaimuthu et al., 2024). Nevertheless, further investigation is required to confirm their role in glycosylation-mediated stemness.

# Conclusions

The main conclusions of the present thesis are:

1. Negative screenings were performed using a novel CRISPR library developed in our lab, targeting protein glycosylation genes in both MDA-MB-231 transformed cells and MCF10A non-transformed cell lines. This approach identified genes crucial for maintaining stem-like properties not only in the triple-negative breast cancer cells but also in other cell types.

2. Bioinformatic analysis of the screenings identified 10 genes crucial for sustaining the stem-like phenotype in malignant cells. Among these, higher expression levels of EXT1, DHDDS, and ST3GAL1 were associated with poorer clinical outcomes, such as reduced overall survival and relapse-free survival, indicating their role in maintaining cancer cell plasticity and stemness.

3. EXT1 and DHDDS are crucial for maintaining stem/plastic properties in triple-negative breast cancer cell lines, particularly by supporting sphere formation. This highlights their potential for developing new targeted therapies.

4. Downregulation of EXT1 and DHDDS did not result in significant changes in the $CD44^{+}/CD24^{-/low}$ population or stem-related gene expression in MDA-MB-231 cells, indicating these genes may regulate a distinct subset of stem-like cells.

5. Silencing EXT1 and DHDDS resulted in a decrease on ALDH+ population in the HCC70 cell line, suggesting that these genes are specifically linked to the $ALDH^{+}$ population.

6. *In vivo* experiments demonstrated that the knockdown of EXT1 and DHDDS in MDA-MB-231 cells significantly reduced tumor growth when injected orthotopically, suggesting their role in tumor initiation and progression properties.

7. Treatment with heparinase III and tunicamycin, which alter glycosylation status of MDA-MB-231 cells, replicated the functional effects of EXT1 and DHDDS downregulation. This confirmed the key role of protein glycosylation in maintaining stem cell-like characteristics. These results suggest that modulating glycosylation could be a novel strategy to regulate stem cell capacity.

# Bibliography

Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J., & Clarke, M. F. (2003). *Prospective identification of tumorigenic breast cancer cells*. www.pnas.orgcgidoi10.1073pnas.0530291100

Ali Khan, A., Raess, M., & de Angelis, M. H. (2021). Moving forward with forward genetics: A summary of the INFRAFRONTIER Forward Genetics Panel Discussion. *F1000Research*, *10*. https://doi.org/10.12688/f1000research.25369.1

Ali, K., Nabeel, M., Mohsin, F., Iqtedar, M., Islam, M., Rasool, M. F., Hashmi, F. K., Hussain, S. A., & Saeed, H. (2024). Recent developments in targeting breast cancer stem cells (BCSCs): a descriptive review of therapeutic strategies and emerging therapies. In *Medical Oncology* (Vol. 41, Issue 5). Springer. https://doi.org/10.1007/s12032-024-02347-z

Almahayni, K., Spiekermann, M., Fiore, A., Yu, G., Pedram, K., & Möckl, L. (2022). Small molecule inhibitors of mammalian glycosylation. *Matrix Biology Plus*, *16*. https://doi.org/10.1016/j.mbplus.2022.100108

Alvarez, C., Tredwell, S., De Vera, M., & Hayden, M. (2006). The genotype-phenotype correlation of hereditary multiple exostoses. *Clinical Genetics*, *70*(2), 122–130. https://doi.org/10.1111/j.1399-0004.2006.00653.x

Ambrose, J. M., Veeraraghavan, V. P., Vennila, R., Rupert, S., Sathyanesan, J., Meenakshisundaram, R., Selvaraj, S., Malayaperumal, S., Kullappan, M., Dorairaj, S., Gujarathi, J. R., Gandhamaneni, S. H., & Surapaneni, K. M. (2022). Comparison of mammosphere formation from stem-like cells of normal breast, malignant primary breast tumors, and MCF-7 cell line. *Journal of the Egyptian National Cancer Institute*, *34*(1). https://doi.org/10.1186/s43046-022-00152-1

Anderluh, M., Berti, F., Bzducha-Wróbel, A., Chiodo, F., Colombo, C., Compostella, F., Durlik, K., Ferhati, X., Holmdahl, R., Jovanovic, D., Kaca, W., Lay, L., Marinovic-Cincovic, M., Marradi, M., Ozil, M., Polito, L., Reina, J. J., Reis, C. A., Sackstein, R., … van Vliet, S. J. (2022). Recent advances on smart glycoconjugate vaccines in infections and cancer. In *FEBS Journal* (Vol. 289, Issue 14, pp. 4251–4303). John Wiley and Sons Inc. https://doi.org/10.1111/febs.15909

Anderluh, M., Berti, F., Bzducha-Wróbel, A., Chiodo, F., Colombo, C., Compostella, F., Durlik, K., Ferhati, X., Holmdahl, R., Jovanovic, D., Kaca, W., Lay, L., Marinovic-Cincovic, M., Marradi, M., Ozil, M., Polito, L., Reina-Martin, J. J., Reis, C. A., Sackstein, R., … van Vliet, S. J. (2021). Emerging glyco-based strategies to steer immune responses. *FEBS Journal*, *288*(16), 4746–4772. https://doi.org/10.1111/febs.15830

Anstine, L. J., & Keri, R. (2019). A new view of the mammary epithelial hierarchy and its implications for breast cancer initiation and metastasis. In *Journal of Cancer Metastasis and Treatment* (Vol. 5). OAE Publishing Inc. https://doi.org/10.20517/2394-4722.2019.24

Bagdonaite, I., Malaker, S. A., Polasky, D. A., Riley, N. M., Schjoldager, K., Vakhrushev, S. Y., Halim, A., Aoki-Kinoshita, K. F., Nesvizhskii, A. I., Bertozzi, C. R., Wandall, H. H., Parker, B. L., Thaysen-Andersen, M., & Scott, N. E. (2022). Glycoproteomics. *Nature Reviews Methods Primers*, *2*(1). https://doi.org/10.1038/s43586-022-00128-4

Bailey, P. C., Lee, R. M., Vitolo, M. I., Pratt, S. J. P., Ory, E., Chakrabarti, K., Lee, C. J., Thompson, K. N., & Martin, S. S. (2018). Single-Cell Tracking of Breast Cancer Cells Enables Prediction of

Sphere Formation from Early Cell Divisions. *IScience*, *8*, 29–39. https://doi.org/10.1016/j.isci.2018.08.015

Bai, X., Ni, J., Beretov, J., Graham, P., & Li, Y. (2018). Cancer stem cell in breast cancer therapeutic resistance. In *Cancer Treatment Reviews* (Vol. 69, pp. 152–163). W.B. Saunders Ltd. https://doi.org/10.1016/j.ctrv.2018.07.004

Balon, K., Sheriff, A., Jacków, J., & Łaczmański, Ł. (2022). Targeting Cancer with CRISPR/Cas9-Based Therapy. In *International Journal of Molecular Sciences* (Vol. 23, Issue 1). MDPI. https://doi.org/10.3390/ijms23010573

Bangarh, R., Khatana, C., Kaur, S., Sharma, A., Kaushal, A., Siwal, S. S., Tuli, H. S., Dhama, K., Thakur, V. K., Saini, R. V., & Saini, A. K. (2023). Aberrant protein glycosylation: Implications on diagnosis and Immunotherapy. In *Biotechnology Advances* (Vol. 66). Elsevier Inc. https://doi.org/10.1016/j.biotechadv.2023.108149

Bar-El, M. L., Vaňková, P., Yeheskel, A., Simhaev, L., Engel, H., Man, P., Haitin, Y., & Giladi, M. (2020). Structural basis of heterotetrameric assembly and disease mutations in the human cis-prenyltransferase complex. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-18970-z

Barkeer, S., Chugh, S., Batra, S. K., & Ponnusamy, M. P. (2018). Glycosylation of Cancer Stem Cells: Function in Stemness, Tumorigenesis, and Metastasis. In *Neoplasia (United States)* (Vol. 20, Issue 8, pp. 813–825). Neoplasia Press, Inc. https://doi.org/10.1016/j.neo.2018.06.001

Ben-Porath, I., Thomson, M. W., Carey, V. J., Ge, R., Bell, G. W., Regev, A., & Weinberg, R. A. (2008). An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nature Genetics*, *40*(5), 499–507. https://doi.org/10.1038/ng.127

Berois, N., Pittini, A., & Osinaga, E. (2022). Targeting Tumor Glycans for Cancer Therapy: Successes, Limitations, and Perspectives. In *Cancers* (Vol. 14, Issue 3). MDPI. https://doi.org/10.3390/cancers14030645

Biswas, S. K., Banerjee, S., Baker, G. W., Kuo, C. Y., & Chowdhury, I. (2022). The Mammary Gland: Basic Structure and Molecular Signaling during Development. In *International Journal of Molecular Sciences* (Vol. 23, Issue 7). MDPI. https://doi.org/10.3390/ijms23073883

Bock, C., Datlinger, P., Chardon, F., Coelho, M. A., Dong, M. B., Lawson, K. A., Lu, T., Maroc, L., Norman, T. M., Song, B., Stanley, G., Chen, S., Garnett, M., Li, W., Moffat, J., Qi, L. S., Shapiro, R. S., Shendure, J., Weissman, J. S., & Zhuang, X. (2022). High-content CRISPR screening. In *Nature Reviews Methods Primers* (Vol. 2, Issue 1). Springer Nature. https://doi.org/10.1038/s43586-021-00093-4

Bou Zerdan, M., Ghorayeb, T., Saliba, F., Allam, S., Bou Zerdan, M., Yaghi, M., Bilani, N., Jaafar, R., & Nahleh, Z. (2022). Triple Negative Breast Cancer: Updates on Classification and Treatment in 2021. *Cancers*, *14*(5). https://doi.org/10.3390/cancers14051253

Bovée, J. V. M. G. (2008). Multiple osteochondromas. *Orphanet Journal of Rare Diseases*, *3*(1). https://doi.org/10.1186/1750-1172-3-3

Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide

for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *74*(3), 229–263. https://doi.org/10.3322/caac.21834

Bukowska-Olech, E., Trzebiatowska, W., Czech, W., Drzymała, O., Frąk, P., Klarowski, F., Kłusek, P., Szwajkowska, A., & Jamsheer, A. (2021). Hereditary Multiple Exostoses—A Review of the Molecular Background, Diagnostics, and Potential Therapeutic Strategies. In *Frontiers in Genetics* (Vol. 12). Frontiers Media S.A. https://doi.org/10.3389/fgene.2021.759129

Busse-Wicher, M., Wicher, K. B., & Kusche-Gullberg, M. (2014). The extostosin family: Proteins with many functions. In *Matrix Biology* (Vol. 35, pp. 25–33). Elsevier. https://doi.org/10.1016/j.matbio.2013.10.001

Butti, R., Gunasekaran, V. P., Kumar, T. V. S., Banerjee, P., & Kundu, G. C. (2019). Breast cancer stem cells: Biology and therapeutic implications. *International Journal of Biochemistry and Cell Biology*, *107*, 38–52. https://doi.org/10.1016/j.biocel.2018.12.001

Calvet, C. Y., André, F. M., & Mir, L. M. (2014). The culture of cancer cell lines as tumorspheres does not systematically result in cancer stem cell enrichment. *PLoS ONE*, *9*(2). https://doi.org/10.1371/journal.pone.0089644

Castaneda, M., den Hollander, P., Kuburich, N. A., Rosen, J. M., & Mani, S. A. (2022). Mechanisms of cancer metastasis. In *Seminars in Cancer Biology* (Vol. 87, pp. 17–31). Academic Press. https://doi.org/10.1016/j.semcancer.2022.10.006

Castells-Roca, L., Tejero, E., Rodríguez-Santiago, B., & Surrallés, J. (2021). *cancers CRISPR Screens in Synthetic Lethality and Combinatorial Therapies for Cancer*. https://doi.org/10.3390/cancers

Celià-Terrassa, T. (2018). Mammary stem cells and breast cancer stem cells: Molecular connections and clinical implications. In *Biomedicines* (Vol. 6, Issue 2). MDPI AG. https://doi.org/10.3390/biomedicines6020050

Chiotaki, R., Polioudaki, H., & Theodoropoulos, P. A. (2016). Stem cell technology in breast cancer: Current status and potential applications. In *Stem Cells and Cloning: Advances and Applications* (Vol. 9, pp. 17–29). Dove Medical Press Ltd. https://doi.org/10.2147/SCCAA.S72836

Chokshi, C. R., Shaikh, M. V., Brakel, B., Rossotti, M. A., Tieu, D., Maich, W., Anand, A., Chafe, S. C., Zhai, K., Suk, Y., Kieliszek, A. M., Miletic, P., Mikolajewicz, N., Chen, D., McNicol, J. D., Chan, K., Tong, A. H. Y., Kuhlmann, L., Liu, L., … Singh, S. K. (2024). Targeting axonal guidance dependencies in glioblastoma with ROBO1 CAR T cells. *Nature Medicine*. https://doi.org/10.1038/s41591-024-03138-9

Conde, I., Ribeiro, A. S., & Paredes, J. (2022). Breast Cancer Stem Cell Membrane Biomarkers: Therapy Targeting and Clinical Implications. In *Cells* (Vol. 11, Issue 6). MDPI. https://doi.org/10.3390/cells11060934

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *New Series*, *339*(6121), 819–823. https://doi.org/10.1126/science

Costa, A. F., Campos, D., Reis, C. A., & Gomes, C. (2020). Targeting Glycosylation: A New Road for Cancer Drug Discovery. In *Trends in Cancer* (Vol. 6, Issue 9, pp. 757–766). Cell Press. https://doi.org/10.1016/j.trecan.2020.04.002

Costa, R. L. B., & Gradishar, W. J. (2017). *ASSOCIATED CONTENT Triple-Negative Breast Cancer: Current Practice and Future Directions*. https://doi.org/10.1200/JOP

Coughlin, S. S. (2019). Epidemiology of Breast Cancer in Women. In A. Ahmad (Ed.), *Breast Cancer Metastasis and Drug Resistance: Challenges and Progress* (pp. 9–29). Springer International Publishing. https://doi.org/10.1007/978-3-030-20301-6_2

Covarrubias, S., Vollmers, A. C., Capili, A., Boettcher, M., Shulkin, A., Correa, M. R., Halasz, H., Robinson, E. K., O'Briain, L., Vollmers, C., Blau, J., Katzman, S., McManus, M. T., & Carpenter, S. (2020). High-Throughput CRISPR Screening Identifies Genes Involved in Macrophage Viability and Inflammatory Pathways. *Cell Reports*, *33*(13). https://doi.org/10.1016/j.celrep.2020.108541

Cristea, S., & Polyak, K. (2018). Dissecting the mammary gland one cell at a time. In *Nature Communications* (Vol. 9, Issue 1). Nature Publishing Group. https://doi.org/10.1038/s41467-018-04905-2

Cui, H. X., Wang, H., Wang, Y., Song, J., Tian, H., Xia, C., & Shen, Y. (2016). ST3Gal III modulates breast cancer cell adhesion and invasion by altering the expression of invasion-related molecules. *Oncology Reports*, *36*(6), 3317–3324. https://doi.org/10.3892/or.2016.5180

Cui, S., Zhang, Y., Xing, L., Li, R., Piao, Y., & Liu, H. (2022). Circular RNA dehydrodolichyl diphosphate synthase facilitated triple-negative breast cancer progression via miR-362-3p/DDX5 axis. *Environmental Toxicology*, *37*(6), 1483–1494. https://doi.org/10.1002/tox.23500

Dai, X. F., Yang, Y. X., & Yang, B. Z. (2024). Glycosylation editing: an innovative therapeutic opportunity in precision oncology. *Molecular and Cellular Biochemistry*. https://doi.org/10.1007/s11010-024-05033-w

Dalla, E., Sreekumar, A., Aguirre-Ghiso, J. A., & Chodosh, L. A. (2023). Dormancy in Breast Cancer. *Cold Spring Harbor Perspectives in Medicine*, *13*(11). https://doi.org/10.1101/cshperspect.a041331

Dall'Olio, F., Malagolini, N., Trinchera, M., & Chiricolo, M. (2014). Sialosignaling: Sialyltransferases as engines of self-fueling loops in cancer progression. In *Biochimica et Biophysica Acta - General Subjects* (Vol. 1840, Issue 9, pp. 2752–2764). Elsevier. https://doi.org/10.1016/j.bbagen.2014.06.006

Deans, R. M., Morgens, D. W., Ökesli, A., Pillay, S., Horlbeck, M. A., Kampmann, M., Gilbert, L. A., Li, A., Mateo, R., Smith, M., Glenn, J. S., Carette, J. E., Khosla, C., & Bassik, M. C. (2016). Parallel shRNA and CRISPR-Cas9 screens enable antiviral drug target identification. *Nature Chemical Biology*, *12*(5), 361–366. https://doi.org/10.1038/nchembio.2050

Denning, W., Das, S., Guo, S., Xu, J., Kappes, J. C., & Hel, Z. (2013). Optimization of the transductional efficiency of lentiviral vectors: Effect of sera and polycations. *Molecular Biotechnology*, *53*(3), 308–314. https://doi.org/10.1007/s12033-012-9528-5

Derakhshan, F., & Reis-Filho, J. S. (2022). Pathogenesis of Triple-Negative Breast Cancer. *Annu. Rev. Pathol. Mech. Dis. 2022*, *17*, 181–204. https://doi.org/10.1146/annurev-pathol-042420

Ding, S., Liu, J., Han, X., & Tang, M. (2023). CRISPR/Cas9-Mediated Genome Editing in Cancer Therapy. In *International Journal of Molecular Sciences* (Vol. 24, Issue 22). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/ijms242216325

Dittmer, J. (2018). Breast cancer stem cells: Features, key drivers and treatment options. In *Seminars in Cancer Biology* (Vol. 53, pp. 59–74). Academic Press. https://doi.org/10.1016/j.semcancer.2018.07.007

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H. W., Listgarten, J., & Root, D. E. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology*, *34*(2), 184–191. https://doi.org/10.1038/nbt.3437

Edani, B. H., Grabi Nska, K. A., Zhang, R., Park, J., Siciliano, B., Surmacz, L., Ha, Y., & Sessa, W. C. (2020). Structural elucidation of the cis-prenyltransferase NgBR/DHDDS complex reveals insights in regulation of protein glycosylation. *Proc Natl Acad Sci U S A*. https://doi.org/10.1073/pnas.2008381117/-/DCSupplemental

Eichler, J. (2019). Protein glycosylation. In *Current Biology* (Vol. 29, Issue 7, pp. R229–R231). Cell Press. https://doi.org/10.1016/j.cub.2019.01.003

Fares, J., Fares, M. Y., Khachfe, H. H., Salhab, H. A., & Fares, Y. (2020). Molecular principles of metastasis: a hallmark of cancer revisited. In *Signal Transduction and Targeted Therapy* (Vol. 5, Issue 1). Springer Nature. https://doi.org/10.1038/s41392-020-0134-x

Faria-Ramos, I., Poças, J., Marques, C., Santos-Antunes, J., Macedo, G., Reis, C. A., & Magalhães, A. (2021). Heparan sulfate glycosaminoglycans: (un)expected allies in cancer clinical management. In *Biomolecules* (Vol. 11, Issue 2, pp. 1–28). MDPI AG. https://doi.org/10.3390/biom11020136

Fellmann, C., Hoffmann, T., Sridhar, V., Hopfgartner, B., Muhar, M., Roth, M., Lai, D. Y., Barbosa, I. A. M., Kwon, J. S., Guan, Y., Sinha, N., & Zuber, J. (2013). An optimized microRNA backbone for effective single-copy RNAi. *Cell Reports*, *5*(6), 1704–1713. https://doi.org/10.1016/j.celrep.2013.11.020

Ferlay J, E. M. L. F. L. M. C. M. M. L. P. M. Z. A. S. I. B. F. (2020, February 7). *Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer*. https://gco.iarc.who.int/media/globocan/factsheets/populations/724-spain-fact-sheet.pdf

Ferlay J, E. M. L. F. L. M. C. M. M. L. P. M. Z. A. S. I. B. F. (2022, February 7). *CANCER TODAY | IARC*. https://gco.iarc.fr/today/en/dataviz/bars?mode=cancer&key=asr&cancers=20&group_populations=1&sexes=2&types=0_1&sort_by=value1

Francescangeli, F., De Angelis, M. L., Rossi, R., Cuccu, A., Giuliani, A., De Maria, R., & Zeuner, A. (2023). Dormancy, stemness, and therapy resistance: interconnected players in cancer evolution. In *Cancer and Metastasis Reviews* (Vol. 42, Issue 1, pp. 197–215). Springer. https://doi.org/10.1007/s10555-023-10092-4

Fultang, N., Chakraborty, M., & Peethambaran, B. (2021). Regulation of cancer stem cells in triple negative breast cancer. In *Cancer Drug Resistance* (Vol. 4, Issue 2, pp. 321–342). OAE Publishing Inc. https://doi.org/10.20517/cdr.2020.106

Fu, N. Y., Nolan, E., Lindeman, G. J., & Visvader, J. E. (2020). Stem cells and the differentiation hierarchy in mammary gland development. *Physiological Reviews*, *100*(2), 489–523. https://doi.org/10.1152/physrev.00040.2018

Gandhi, N., & Das, G. M. (2019). Metabolic reprogramming in breast cancer and its therapeutic implications. In *Cells* (Vol. 8, Issue 2). MDPI. https://doi.org/10.3390/cells8020089

Garnham, R., Geh, D., Nelson, R., Ramon-Gil, E., Wilson, L., Schmidt, E. N., Walker, L., Adamson, B., Buskin, A., Hepburn, A. C., Hodgson, K., Kendall, H., Frame, F. M., Maitland, N., Coffey, K., Strand, D. W., Robson, C. N., Elliott, D. J., Heer, R., … Scott, E. (2024). ST3 beta-galactoside alpha-2,3-sialyltransferase 1 (ST3Gal1) synthesis of Siglec ligands mediates anti-tumour immunity in prostate cancer. *Communications Biology*, *7*(1). https://doi.org/10.1038/s42003-024-05924-0

Garrido-Castro, A. C., Lin, N. U., & Polyak, K. (2019). Insights into molecular classifications of triple-negative breast cancer: Improving patient selection for treatment. *Cancer Discovery*, *9*(2), 176–198. https://doi.org/10.1158/2159-8290.CD-18-1177

Gerstberger, S., Jiang, Q., & Ganesh, K. (2023). Metastasis. In *Cell* (Vol. 186, Issue 8, pp. 1564–1579). Elsevier B.V. https://doi.org/10.1016/j.cell.2023.03.003

Geurts, M. H., & Clevers, H. (2023). CRISPR engineering in organoids for gene repair and disease modelling. *Nature Reviews Bioengineering*, *1*(1), 32–45. https://doi.org/10.1038/s44222-022-00013-5

Gieniec, K. A., & Davis, F. M. (2022). Mammary basal cells: Stars of the show. In *Biochimica et Biophysica Acta - Molecular Cell Research* (Vol. 1869, Issue 1). Elsevier B.V. https://doi.org/10.1016/j.bbamcr.2021.119159

Ginestier, C., Hur, M. H., Charafe-Jauffret, E., Monville, F., Dutcher, J., Brown, M., Jacquemier, J., Viens, P., Kleer, C. G., Liu, S., Schott, A., Hayes, D., Birnbaum, D., Wicha, M. S., & Dontu, G. (2007). ALDH1 Is a Marker of Normal and Malignant Human Mammary Stem Cells and a Predictor of Poor Clinical Outcome. *Cell Stem Cell*, *1*(5), 555–567. https://doi.org/10.1016/j.stem.2007.08.014

Goto, N., Westcott, P. M. K., Goto, S., Imada, S., Taylor, M. S., Eng, G., Braverman, J., Deshpande, V., Jacks, T., Agudo, J., & Yilmaz, Ö. H. (2024). SOX17 enables immune evasion of early colorectal adenomas and cancers. *Nature*, *627*(8004), 636–645. https://doi.org/10.1038/s41586-024-07135-3

Guo, H. B., Johnson, H., Randolph, M., Nagy, T., Blalock, R., & Pierce, M. (2010). Specific posttranslational modification regulates early events in mammary carcinoma formation. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(49), 21116–21121. https://doi.org/10.1073/pnas.1013405107

Guo, H., Nagy, T., & Pierce, M. (2014). Post-translational glycoprotein modifications regulate colon cancer stem cells and colon adenoma progression in Apcmin/+ mice through altered Wnt receptor signaling. *Journal of Biological Chemistry*, *289*(45), 31534–31549. https://doi.org/10.1074/jbc.M114.602680

Guo, L., Li, F., Liu, H., Kong, D., Chen, C., & Sun, S. (2023). SIX1 amplification modulates stemness and tumorigenesis in breast cancer. *Journal of Translational Medicine*, *21*(1). https://doi.org/10.1186/s12967-023-04679-2

Guo, Z., & Han, S. (2023). Targeting cancer stem cell plasticity in triple-negative breast cancer. In *Exploration of Targeted Anti-tumor Therapy* (Vol. 4, Issue 6, pp. 1165–1181). Open Exploration Publishing Inc. https://doi.org/10.37349/ETAT.2023.00190

Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. In *Cancer Discovery* (Vol. 12, Issue 1, pp. 31–46). American Association for Cancer Research Inc. https://doi.org/10.1158/2159-8290.CD-21-1059

Hannan, F. M., Elajnaf, T., Vandenberg, L. N., Kennedy, S. H., & Thakker, R. V. (2023). Hormonal regulation of mammary gland development and lactation. In *Nature Reviews Endocrinology* (Vol. 19, Issue 1, pp. 46–61). Nature Research. https://doi.org/10.1038/s41574-022-00742-y

Harbeck, N., Penault-Llorca, F., Cortes, J., Gnant, M., Houssami, N., Poortmans, P., Ruddy, K., Tsang, J., & Cardoso, F. (2019). Breast cancer. *Nature Reviews Disease Primers*, *5*(1). https://doi.org/10.1038/s41572-019-0111-2

Harbuzariu, A., Rampoldi, A., Daley-Brown, D. S., Candelaria, P., Harmon, T. L., Lipsey, C. C., Beech, D. J., Quarshie, A., Ilies, G. O., & Gonzalez-Perez, R. R. (2017). Leptin-Notch signaling axis is involved in pancreatic cancer progression. In *Oncotarget* (Vol. 8, Issue 5). www.impactjournals.com/oncotarget/

Harrison, K. D., Park, E. J., Gao, N., Kuo, A., Rush, J. S., Waechter, C. J., Lehrman, M. A., & Sessa, W. C. (2011). Nogo-B receptor is necessary for cellular dolichol biosynthesis and protein N-glycosylation. *EMBO Journal*, *30*(12), 2490–2500. https://doi.org/10.1038/emboj.2011.147

Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, K. R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S., Mero, P., Dirks, P., Sidhu, S., Roth, F. P., Rissland, O. S., Durocher, D., Angers, S., & Moffat, J. (2015). High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell*, *163*(6), 1515–1526. https://doi.org/10.1016/j.cell.2015.11.015

Hart, T., Tong, A. H. Y., Chan, K., Van Leeuwen, J., Seetharaman, A., Aregger, M., Chandrashekhar, M., Hustedt, N., Seth, S., Noonan, A., Habsid, A., Sizova, O., Nedyalkova, L., Climie, R., Tworzyanski, L., Lawson, K., Sartori, M. A., Alibeh, S., Tieu, D., … Moffat, J. (2017). Evaluation and design of genome-wide CRISPR/SpCas9 knockout screens. *G3: Genes, Genomes, Genetics*, *7*(8), 2719–2727. https://doi.org/10.1534/g3.117.041277

He, L., Wick, N., Germans, S. K., & Peng, Y. (2021). The role of breast cancer stem cells in chemoresistance and metastasis in triple-negative breast cancer. In *Cancers* (Vol. 13, Issue 24). MDPI. https://doi.org/10.3390/cancers13246209

He, M., Zhou, X., & Wang, X. (2024). Glycosylation: mechanisms, biological functions and clinical implications. *Signal Transduction and Targeted Therapy*, *9*(1), 194. https://doi.org/10.1038/s41392-024-01886-1

Heynen-Genel, S., Pache, L., Chanda, S. K., & Rosen, J. (2012). Functional genomic and high-content screening for target discovery and deconvolution. In *Expert Opinion on Drug Discovery* (Vol. 7, Issue 10, pp. 955–968). https://doi.org/10.1517/17460441.2012.711311

Hillary, V. E., & Ceasar, S. A. (2023). A Review on the Mechanism and Applications of CRISPR/Cas9/Cas12/Cas13/Cas14 Proteins Utilized for Genome Engineering. In *Molecular Biotechnology* (Vol. 65, Issue 3, pp. 311–325). Springer. https://doi.org/10.1007/s12033-022-00567-0

Holen, I., Speirs, V., Morrissey, B., & Blyth, K. (2017). In vivo models in breast cancer research: Progress, challenges and future directions. In *DMM Disease Models and Mechanisms* (Vol. 10, Issue 4, pp. 359–371). Company of Biologists Ltd. https://doi.org/10.1242/dmm.028274

Honeth, G., Bendahl, P. O., Ringnér, M., Saal, L. H., Gruvberger-Saal, S. K., Lövgren, K., Grabau, D., Fernö, M., Borg, Å., & Hegardt, C. (2008). The CD44+/CD24-phenotype is enriched in basal-like breast tumors. *Breast Cancer Research*, *10*(3). https://doi.org/10.1186/bcr2108

Houghton, S. C., & Hankinson, S. E. (2021). Cancer progress and priorities: Breast cancer. *Cancer Epidemiology Biomarkers and Prevention*, *30*(5), 822–844. https://doi.org/10.1158/1055-9965.EPI-20-1193

Huang, Y., Hong, W., & Wei, X. (2022). The molecular mechanisms and therapeutic strategies of EMT in tumor progression and metastasis. In *Journal of Hematology and Oncology* (Vol. 15, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s13045-022-01347-8

Hua, Z., White, J., & Zhou, J. (2022). Cancer stem cells in TNBC. In *Seminars in Cancer Biology* (Vol. 82, pp. 26–34). Academic Press. https://doi.org/10.1016/j.semcancer.2021.06.015

Ibragimova, M., Tsyganov, M., & Litviakov, N. (2022). Tumour Stem Cells in Breast Cancer. In *International Journal of Molecular Sciences* (Vol. 23, Issue 9). MDPI. https://doi.org/10.3390/ijms23095058

Ibrahim, S. A., Gadalla, R., El-Ghonaimy, E. A., Samir, O., Mohamed, H. T., Hassan, H., Greve, B., El-Shinawi, M., Mohamed, M. M., & Götte, M. (2017). Syndecan-1 is a novel molecular marker for triple negative inflammatory breast cancer and modulates the cancer stem cell phenotype via the IL-6/STAT3, Notch and EGFR signaling pathways. *Molecular Cancer*, *16*(1). https://doi.org/10.1186/s12943-017-0621-z

Ingthorsson, S., Traustadottir, G. A., & Gudjonsson, T. (2022). Cellular Plasticity and Heterotypic Interactions during Breast Morphogenesis and Cancer Initiation. In *Cancers* (Vol. 14, Issue 21). MDPI. https://doi.org/10.3390/cancers14215209

Inman, J. L., Robertson, C., Mott, J. D., & Bissell, M. J. (2015). Mammary gland development: Cell fate specification, stem cells and the microenvironment. In *Development (Cambridge)* (Vol. 142, Issue 6, pp. 1028–1042). Company of Biologists Ltd. https://doi.org/10.1242/dev.087643

Jehanno, C., Vulin, M., Richina, V., Richina, F., & Bentires-Alj, M. (2022). Phenotypic plasticity during metastatic colonization. In *Trends in Cell Biology* (Vol. 32, Issue 10, pp. 854–867). Elsevier Ltd. https://doi.org/10.1016/j.tcb.2022.03.007

Joung, J., Konermann, S., Gootenberg, J. S., Abudayyeh, O. O., Platt, R. J., Brigham, M. D., Sanjana, N. E., & Zhang, F. (2017). Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nature Protocols*, *12*(4), 828–863. https://doi.org/10.1038/nprot.2017.016

Karim, A. M., Eun Kwon, J., Ali, T., Jang, J., Ullah, I., Lee, Y. G., Park, D. W., Park, J., Jeang, J. W., & Kang, S. C. (2023). Triple-negative breast cancer: epidemiology, molecular mechanisms, and modern vaccine-based treatment strategies. In *Biochemical Pharmacology* (Vol. 212). Elsevier Inc. https://doi.org/10.1016/j.bcp.2023.115545

Karsten, U., & Goletz, S. (2013). What makes cancer stem cell markers different? *SpringerPlus*, *2*(1), 1–8. https://doi.org/10.1186/2193-1801-2-301

Katti, A., Diaz, B. J., Caragine, C. M., Sanjana, N. E., & Dow, L. E. (2022). CRISPR in cancer biology and therapy. In *Nature Reviews Cancer* (Vol. 22, Issue 5, pp. 259–279). Nature Research. https://doi.org/10.1038/s41568-022-00441-w

Kemper, K., Sprick, M. R., De Bree, M., Scopelliti, A., Vermeulen, L., Hoek, M., Zeilstra, J., Pals, S. T., Mehmet, H., Stassi, G., & Medema, J. P. (2010). The AC133 epitope, but not the CD133 protein, is lost upon cancer stem cell differentiation. *Cancer Research*, *70*(2), 719–729. https://doi.org/10.1158/0008-5472.CAN-09-1820

Kessel, S. L., & Chan, L. L. Y. (2020). A High-Throughput Image Cytometry Method for the Formation, Morphometric, and Viability Analysis of Drug-Treated Mammospheres. *SLAS Discovery*, *25*(7), 723–733. https://doi.org/10.1177/2472555220922817

Khan, S., Suryavanshi, M., Kaur, J., Nayak, D., Khurana, A., Manchanda, R. K., Tandon, C., & Tandon, S. (2021). Stem cell therapy: A paradigm shift in breast cancer treatment. *World Journal of Stem Cells*, *13*(7), 841–860. https://doi.org/10.4252/wjsc.v13.i7.841

Khan, T., & Cabral, H. (2021). Abnormal Glycosylation of Cancer Stem Cells and Targeting Strategies. In *Frontiers in Oncology* (Vol. 11). Frontiers Media S.A. https://doi.org/10.3389/fonc.2021.649338

Khan, T., Igarashi, K., Tanabe, A., Miyazawa, T., Fukushima, S., Miura, Y., Matsumoto, Y., Yamasoba, T., Matsumoto, A., Cabral, H., & Kataoka, K. (2020). Structural Control of Boronic Acid Ligands Enhances Intratumoral Targeting of Sialic Acid to Eradicate Cancer Stem-like Cells. *ACS Applied Bio Materials*, *3*(8), 5030–5039. https://doi.org/10.1021/acsabm.0c00530

Kiessling, M. K., Schuierer, S., Stertz, S., Beibel, M., Bergling, S., Knehr, J., Carbone, W., de Vallière, C., Tchinda, J., Bouwmeester, T., Seuwen, K., Rogler, G., & Roma, G. (2016). Identification of oncogenic driver mutations by genome-wide CRISPR-Cas9 dropout screening. *BMC Genomics*, *17*(1). https://doi.org/10.1186/s12864-016-3042-2

Kim, R. J., Park, J. R., Roh, K. J., Choi, A. R., Kim, S. R., Kim, P. H., Yu, J. H., Lee, J. W., Ahn, S. H., Gong, G., Hwang, J. W., Kang, K. S., Kong, G., Sheen, Y. Y., & Nam, J. S. (2013). High aldehyde dehydrogenase activity enhances stem cell features in breast cancer cells by activating hypoxia-inducible factor-2α. *Cancer Letters*, *333*(1), 18–31. https://doi.org/10.1016/j.canlet.2012.11.026

Kim, S. L., Choi, H. S., & Lee, D. S. (2023). BRD4/nuclear PD-L1/RelB circuit is involved in the stemness of breast cancer cells. *Cell Communication and Signaling*, *21*(1). https://doi.org/10.1186/s12964-023-01319-6

Kim, Y., Lee, D. Y., Choi, J. U., Park, J. S., Lee, S. M., Kang, C. H., & Park, C. H. (2023). Optimized conditions for gene transduction into primary immune cells using viral vectors. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-39597-2

Knelson, E. H., Nee, J. C., & Blobe, G. C. (2014). Heparan sulfate signaling in cancer. In *Trends in Biochemical Sciences* (Vol. 39, Issue 6, pp. 277–288). Elsevier Ltd. https://doi.org/10.1016/j.tibs.2014.03.001

Kong, D., Hughes, C. J., & Ford, H. L. (2020). Cellular Plasticity in Breast Cancer Progression and Therapy. In *Frontiers in Molecular Biosciences* (Vol. 7). Frontiers Media S.A. https://doi.org/10.3389/fmolb.2020.00072

Kong, W., Chen, Y., Zhao, Z., Zhang, L., Lin, X., Luo, X., Wang, S., Song, Z., Lin, X., Lai, G., & Yu, Z. (2021). EXT1 methylation promotes proliferation and migration and predicts the clinical outcome of non-small cell lung carcinoma via WNT signalling pathway. *Journal of Cellular and Molecular Medicine*, *25*(5), 2609–2620. https://doi.org/10.1111/jcmm.16277

Kumar, H., Gupta, N. V., Jain, R., Madhunapantula, S. R. V., Babu, C. S., Kesharwani, S. S., Dey, S., & Jain, V. (2023). A review of biological targets and therapeutic approaches in the management of triple-negative breast cancer. In *Journal of Advanced Research* (Vol. 54, pp. 271–292). Elsevier B.V. https://doi.org/10.1016/j.jare.2023.02.005

Kumru, O. S., Wang, Y., Gombotz, C. W. R., Kelley-Clarke, B., Cieplak, W., Kim, T., Joshi, S. B., & Volkin, D. B. (2018). Physical Characterization and Stabilization of a Lentiviral Vector Against Adsorption and Freeze-Thaw. *Journal of Pharmaceutical Sciences*, *107*(11), 2764–2774. https://doi.org/10.1016/j.xphs.2018.07.010

Kvokačková, B., Remšík, J., Jolly, M. K., & Souček, K. (2021). Phenotypic heterogeneity of triple-negative breast cancer mediated by epithelial–mesenchymal plasticity. In *Cancers* (Vol. 13, Issue 9). MDPI. https://doi.org/10.3390/cancers13092188

Lee, C.-H., Yu, C.-C., Wang, B.-Y., & Chang, W.-W. (2015). *Tumorsphere as an effective in vitro platform for screening anti-cancer stem cell drugs* (Vol. 7, Issue 2). www.impactjournals.com/oncotarget

Leon-Ferre, R. A., & Goetz, M. P. (2023). Advances in systemic therapies for triple negative breast cancer. In *BMJ*. BMJ Publishing Group. https://doi.org/10.1136/bmj-2022-071674

le Sage, C., Lawo, S., & Cross, B. C. S. (2020). CRISPR: A Screener's Guide. *SLAS Discovery*, *25*(3), 233–240. https://doi.org/10.1177/2472555219883621

Liang, D., Gao, Q., Meng, Z., Li, W., Song, J., & Xue, K. (2023). Glycosylation in breast cancer progression and mammary development: Molecular connections and malignant transformations. In *Life Sciences* (Vol. 326). Elsevier Inc. https://doi.org/10.1016/j.lfs.2023.121781

Liang, Y., Zhang, H., Song, X., & Yang, Q. (2020). Metastatic heterogeneity of breast cancer: Molecular mechanism and potential therapeutic targets. In *Seminars in Cancer Biology* (Vol. 60, pp. 14–27). Academic Press. https://doi.org/10.1016/j.semcancer.2019.08.012

Li, C., Du, Y., Yang, Z., He, L., Wang, Y., Hao, L., Ding, M., Yan, R., Wang, J., & Fan, Z. (2016). GALNT1-mediated glycosylation and activation of Sonic hedgehog signaling maintains the self-renewal and tumor-initiating capacity of bladder cancer stem cells. *Cancer Research*, *76*(5), 1273–1283. https://doi.org/10.1158/0008-5472.CAN-15-2309

Li, C. W., Lim, S. O., Xia, W., Lee, H. H., Chan, L. C., Kuo, C. W., Khoo, K. H., Chang, S. S., Cha, J. H., Kim, T., Hsu, J. L., Wu, Y., Hsu, J. M., Yamaguchi, H., Ding, Q., Wang, Y., Yao, J., Lee, C. C., Wu,

H. J., … Hung, M. C. (2016). Glycosylation and stabilization of programmed death ligand-1 suppresses T-cell activity. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms12632

Li, N., Gou, J. H., Xiong, J., You, J. J., & Li, Z. Y. (2020). HOXB4 promotes the malignant progression of ovarian cancer via DHDDS. *BMC Cancer*, *20*(1). https://doi.org/10.1186/S12885-020-06725-4

Lin, K., Chang, Y. C., Marron Fernandez de Velasco, E., Wickman, K., Myers, C. L., & Bielinsky, A. K. (2022). Scalable CRISPR-Cas9 chemical genetic screens in non-transformed human cells. *STAR Protocols*, *3*(4). https://doi.org/10.1016/j.xpro.2022.101675

Lin, S., Kemmner, W., Grigull, S., & Schlag, P. M. (2002). Cell surface α2, 6-sialylation affects adhesion of breast carcinoma cells. *Experimental Cell Research*, *276*(1), 101–110. https://doi.org/10.1006/excr.2002.5521

Liu, H., Patel, M. R., Prescher, J. A., Patsialou, A., Qian, D., Lin, J., Wen, S., Chang, Y. F., Bachmann, M. H., Shimono, Y., Dalerba, P., Adorno, M., Lobo, N., Bueno, J., Dirbas, F. M., Goswami, S., Somlo, G., Condeelis, J., Contag, C. H., … Clarke, M. F. (2010). Cancer stem cells from human breast tumors are involved in spontaneous metastases in orthotopic mouse models. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(42), 18115–18120. https://doi.org/10.1073/pnas.1006732107

Liu, H., Zhou, Y., Qiu, H., Zhuang, R., Han, Y., Liu, X., Qiu, X., Wang, Z., Xu, L., Tan, R., Hong, W., & Wang, T. (2021). Rab26 suppresses migration and invasion of breast cancer cells through mediating autophagic degradation of phosphorylated Src. *Cell Death and Disease*, *12*(4). https://doi.org/10.1038/s41419-021-03561-7

Liu, S., Cong, Y., Wang, D., Sun, Y., Deng, L., Liu, Y., Martin-Trevino, R., Shang, L., McDermott, S. P., Landis, M. D., Hong, S., Adams, A., D'Angelo, R., Ginestier, C., Charafe-Jauffret, E., Clouthier, S. G., Birnbaum, D., Wong, S. T., Zhan, M., … Wicha, M. S. (2014). Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. *Stem Cell Reports*, *2*(1), 78–91. https://doi.org/10.1016/j.stemcr.2013.11.009

Li, X., Chen, Z., Ye, W., Yu, J., Zhang, X., Li, Y., Niu, Y., Ran, S., Wang, S., Luo, Z., Zhao, J., Hao, Y., Zong, J., Xia, C., Xia, J., & Wu, J. (2023). High-throughput CRISPR technology: a novel horizon for solid organ transplantation. In *Frontiers in Immunology* (Vol. 14). Frontiers Media SA. https://doi.org/10.3389/fimmu.2023.1295523

Lloyd-Lewis, B., Harris, O. B., Watson, C. J., & Davis, F. M. (2017). Mammary Stem Cells: Premise, Properties, and Perspectives. In *Trends in Cell Biology* (Vol. 27, Issue 8, pp. 556–567). Elsevier Ltd. https://doi.org/10.1016/j.tcb.2017.04.001

Lo Presti, V., Cornel, A. M., Plantinga, M., Dünnebach, E., Kuball, J., Boelens, J. J., Nierkens, S., & van Til, N. P. (2021). Efficient lentiviral transduction method to gene modify cord blood CD8+ T cells for cancer therapy applications. *Molecular Therapy Methods and Clinical Development*, *21*, 357–368. https://doi.org/10.1016/j.omtm.2021.03.015

Lu, B., Natarajan, E., Balaji Raghavendran, H. R., & Markandan, U. D. (2023a). Molecular Classification, Treatment, and Genetic Biomarkers in Triple-Negative Breast Cancer: A

Review. *Technology in Cancer Research and Treatment*, *22*. https://doi.org/10.1177/15330338221145246

Lu, B., Natarajan, E., Balaji Raghavendran, H. R., & Markandan, U. D. (2023b). Molecular Classification, Treatment, and Genetic Biomarkers in Triple-Negative Breast Cancer: A Review. In *Technology in Cancer Research and Treatment* (Vol. 22). SAGE Publications Inc. https://doi.org/10.1177/15330338221145246

Łukasiewicz, S., Czeczelewski, M., Forma, A., Baj, J., Sitarz, R., & Stanisławek, A. (2021). Breast cancer—epidemiology, risk factors, classification, prognostic markers, and current treatment strategies—An updated review. In *Cancers* (Vol. 13, Issue 17). MDPI. https://doi.org/10.3390/cancers13174287

Lüönd, F., Tiede, S., & Christofori, G. (2021). Breast cancer as an example of tumour heterogeneity and tumour cell plasticity during malignant progression. In *British Journal of Cancer* (Vol. 125, Issue 2, pp. 164–175). Springer Nature. https://doi.org/10.1038/s41416-021-01328-7

Maciejewski, K., Giers, M., Oleksiewicz, U., & Czerwinska, P. (2024). The Epigenetic Modifiers HDAC2 and HDAC7 Inversely Associate with Cancer Stemness and Immunity in Solid Tumors. *International Journal of Molecular Sciences*, *25*(14), 7841. https://doi.org/10.3390/ijms25147841

Mahmoud, R., Ordóñez-Morán, P., & Allegrucci, C. (2022). Challenges for Triple Negative Breast Cancer Treatment: Defeating Heterogeneity and Cancer Stemness. In *Cancers* (Vol. 14, Issue 17). MDPI. https://doi.org/10.3390/cancers14174280

Manandhar, S., Kim, C. G., Lee, S. H., Kang, S. H., Basnet, N., & Lee, Y. M. (2017). Exostosin 1 regulates cancer cell stemness in doxorubicin-resistant breast cancer cells. *Oncotarget*, *8*(41), 70521–70537. https://doi.org/10.18632/ONCOTARGET.19737

Massagué, J., & Ganesh, K. (2021). Metastasis-initiating cells and ecosystems. In *Cancer Discovery* (Vol. 11, Issue 4, pp. 971–994). American Association for Cancer Research Inc. https://doi.org/10.1158/2159-8290.CD-21-0010

Mastrangeli, R., Palinsky, W., & Bierau, H. (2018). Glycoengineered antibodies: Towards the next-generation of immunotherapeutics. In *Glycobiology* (Vol. 29, Issue 3, pp. 199–210). Oxford University Press. https://doi.org/10.1093/glycob/cwy092

Mathiowetz, A. J., Roberts, M. A., Morgens, D. W., Olzmann, J. A., & Li, Z. (2023). Protocol for performing pooled CRISPR-Cas9 loss-of-function screens. *STAR Protocols*, *4*(2). https://doi.org/10.1016/j.xpro.2023.102201

Matsumoto, Y., & Ju, T. (2023). Aberrant Glycosylation as Immune Therapeutic Targets for Solid Tumors. In *Cancers* (Vol. 15, Issue 14). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/cancers15143536

McClements, L., Yakkundi, A., Papaspyropoulos, A., Harrison, H., Ablett, M. P., Jithesh, P. V., McKeen, H. D., Bennett, R., Donley, C., Kissenpfennig, A., McIntosh, S., McCarthy, H. O., O'Neill, E., Clarke, R. B., & Robson, T. (2013). Targeting treatment-resistant breast cancer stem cells with FKBPL and Its peptide derivative, AD-01, via the CD44 pathway. *Clinical Cancer Research*, *19*(14), 3881–3893. https://doi.org/10.1158/1078-0432.CCR-13-0595

Mccormick, C., Duncan, G., Goutsos, K. T., & Tufaro, F. (2000). The putative tumor suppressors EXT1 and EXT2 form a stable complex that accumulates in the Golgi apparatus and catalyzes the synthesis of heparan sulfate. In *PNAS* (Vol. 97, Issue 2). www.pnas.org

Mccormick, C., Leduc, Y., Martindale, D., Mattison, K., Esford, L. E., Dyer, A. P., & Tufaro, F. (1998). *The putative tumour suppressor EXT1 alters the expression of cell-surface heparan sulfate*. http://genetics.nature.com

Mereiter, S., Balmaña, M., Campos, D., Gomes, J., & Reis, C. A. (2019). Glycosylation in the Era of Cancer-Targeted Therapy: Where Are We Heading? *Cancer Cell*, *36*(1), 6–16. https://doi.org/10.1016/J.CCELL.2019.06.006

Meyers, S., Demeyer, S., & Cools, J. (2023). CRISPR screening in hematology research: from bulk to single-cell level. In *Journal of Hematology and Oncology* (Vol. 16, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s13045-023-01495-5

Mittal, V. (2017). *Epithelial Mesenchymal Transition in Tumor Metastasis*. https://doi.org/10.1146/annurev-pathol-020117

Mi Young Kim. (2021). Breast Cancer Metastasis. In *Advances in Experimental Medicine and Biology*. http://www.springer.com/series/5584

Moremen, K. W., Tiemeyer, M., & Nairn, A. V. (2012). Vertebrate protein glycosylation: Diversity, synthesis and function. In *Nature Reviews Molecular Cell Biology* (Vol. 13, Issue 7, pp. 448–462). https://doi.org/10.1038/nrm3383

Morgens, D. W., Deans, R. M., Li, A., & Bassik, M. C. (2016). Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. *Nature Biotechnology*, *34*(6), 634–636. https://doi.org/10.1038/nbt.3567

Mousa, J., Veres, L., Mohamed, A., De Graef, D., & Morava, E. (2022). Acetazolamide treatment in late onset CDG type 1 due to biallelic pathogenic DHDDS variants. *Molecular Genetics and Metabolism Reports*, *32*. https://doi.org/10.1016/j.ymgmr.2022.100901

Munkley, J., Elliott, D. J., Munkley, J., & Elliott, D. J. (2016). Hallmarks of glycosylation in cancer. *Oncotarget*, *7*(23), 35478–35489. https://doi.org/10.18632/ONCOTARGET.8155

Nadanaka, S., & Kitagawa, H. (2021). *Regulation of growth factor activity by heparin/ heparan sulfate proteoglycan*. https://www.ncbi.nlm.nih.gov/books/

Nalla, L. V., Kalia, K., & Khairnar, A. (2019). Self-renewal signaling pathways in breast cancer stem cells. *International Journal of Biochemistry and Cell Biology*, *107*, 140–153. https://doi.org/10.1016/j.biocel.2018.12.017

Nam, K., Son, S.-H., Oh, S., Jeon, D., Kim, H., Noh, D.-Y., Kim, S., & Shin, I. (2017). Oncotarget 35804 www.impactjournals.com/oncotarget Binding of galectin-1 to integrin β1 potentiates drug resistance by promoting survivin expression in breast cancer cells. In *Oncotarget* (Vol. 8, Issue 22). www.impactjournals.com/oncotarget/

Neville, M. C., Webb, P., Ramanathan, P., Mannino, M. P., Pecorini, C., Monks, J., Anderson, S. M., & MacLean, P. (2013). *The insulin receptor plays an important role in secretory differentiation in the mammary gland*. https://doi.org/10.1152/ajpendo.00337.2013.-Insulin

Nolan, E., Kang, Y., & Malanchi, I. (2023). Mechanisms of Organ-Specific Metastasis of Breast Cancer. *Cold Spring Harbor Perspectives in Medicine*, *13*(11). https://doi.org/10.1101/cshperspect.a041326

Nolan, E., Lindeman, G. J., & Visvader, J. E. (2023). Deciphering breast cancer: from biology to the clinic. *Cell*. https://doi.org/10.1016/j.cell.2023.01.040

Ohkawa, Y., Wade, A., Lindberg, O. R., Chen, K. Y., Tran, V. M., Brown, S. J., Kumar, A., Kalita, M., James, C. D., & Phillips, J. J. (2021). Heparan sulfate synthesized by ext1 regulates receptor tyrosine kinase signaling and promotes resistance to EGFR inhibitors in GBM. *Molecular Cancer Research*, *19*(1), 150–160. https://doi.org/10.1158/1541-7786.MCR-20-0420

Okuyama, N. C. M., Ribeiro, D. L., da Rocha, C. Q., Pereira, É. R., Cólus, I. M. de S., & Serpeloni, J. M. (2023). Three-dimensional cell cultures as preclinical models to assess the biological activity of phytochemicals in breast cancer. In *Toxicology and Applied Pharmacology* (Vol. 460). Academic Press Inc. https://doi.org/10.1016/j.taap.2023.116376

Ospina-Muñoz, N., & Vernot, J. P. (2020). Partial acquisition of stemness properties in tumorspheres obtained from interleukin-8-treated MCF-7 cells. *Tumor Biology*, *42*(12). https://doi.org/10.1177/1010428320979438

Pacifici, M. (2017). Hereditary Multiple Exostoses: New Insights into Pathogenesis, Clinical Complications, and Potential Treatments. In *Current Osteoporosis Reports* (Vol. 15, Issue 3, pp. 142–152). Current Medicine Group LLC 1. https://doi.org/10.1007/s11914-017-0355-2

Pandyra, A. A., Mullen, P. J., Goard, C. A., Ericson, E., Sharma, P., Kalkat, M., Yu, R., Pong, J. T., Brown, K. R., Hart, T., Gebbia, M., Lang, K. S., Giaever, G., Nislow, C., Moffat, J., & Penn, L. Z. (2015). Genome-wide RNAi analysis reveals that simultaneous inhibition of specific mevalonate pathway genes potentiates tumor cell death. In *Oncotarget* (Vol. 6, Issue 29). www.impactjournals.com/oncotarget/

Park, J. W., Kim, Y., Lee, S. been, Oh, C. W., Lee, E. J., Ko, J. Y., & Park, J. H. (2022). Autophagy inhibits cancer stemness in triple-negative breast cancer via miR-181a-mediated regulation of ATG5 and/or ATG2B. *Molecular Oncology*, *16*(9), 1857–1875. https://doi.org/10.1002/1878-0261.13180

Park, M., Kim, D., Ko, S., Kim, A., Mo, K., & Yoon, H. (2022). Breast Cancer Metastasis: Mechanisms and Therapeutic Implications. In *International Journal of Molecular Sciences* (Vol. 23, Issue 12). MDPI. https://doi.org/10.3390/ijms23126806

Paul, S., Sinha, S., & Kundu, C. N. (2022). Targeting cancer stem cells in the tumor microenvironment: An emerging role of PARP inhibitors. In *Pharmacological Research* (Vol. 184). Academic Press. https://doi.org/10.1016/j.phrs.2022.106425

Peixoto, A., Relvas-Santos, M., Azevedo, R., Lara Santos, L., & Ferreira, J. A. (2019). Protein glycosylation and tumor microenvironment alterations driving cancer hallmarks. *Frontiers in Oncology*, *9*(MAY). https://doi.org/10.3389/fonc.2019.00380

Pellacani, D., Tan, S., Lefort, S., & Eaves, C. J. (2019). Transcriptional regulation of normal human mammary cell heterogeneity and its perturbation in breast cancer. *The EMBO Journal*, *38*(14). https://doi.org/10.15252/embj.2018100330

Peric, L., Vukadin, S., Petrovic, A., Kuna, L., Puseljic, N., Sikora, R., Rozac, K., Vcev, A., & Smolic, M. (2022). Glycosylation Alterations in Cancer Cells, Prognostic Value of Glycan Biomarkers and Their Potential as Novel Therapeutic Targets in Breast Cancer. In *Biomedicines* (Vol. 10, Issue 12). MDPI. https://doi.org/10.3390/biomedicines10123265

Pfeifer, V., Weber, H., Wang, Y., Schlesinger, M., Gorzelanny, C., & Bendas, G. (2023). Exostosin 1 Knockdown Induces Chemoresistance in MV3 Melanoma Cells by Upregulating JNK and MEK/ERK Signaling. *International Journal of Molecular Sciences*, *24*(6). https://doi.org/10.3390/ijms24065452

Picco, G., Julien, S., Brockhausen, I., Beatson, R., Antonopoulos, A., Haslam, S., Mandel, U., Dell, A., Pinder, S., Taylor-Papadimitriou, J., & Burchell, J. (2010). Over-expression of ST3Gal-I promotes mammary tumorigenesis. *Glycobiology*, *20*(10), 1241–1250. https://doi.org/10.1093/glycob/cwq085

Pietrobono, S., Anichini, G., Sala, C., Manetti, F., Almada, L. L., Pepe, S., Carr, R. M., Paradise, B. D., Sarkaria, J. N., Davila, J. I., Tofani, L., Battisti, I., Arrigoni, G., Ying, L., Zhang, C., Li, H., Meves, A., Fernandez-Zapico, M. E., & Stecca, B. (2020). ST3GAL1 is a target of the SOX2-GLI1 transcriptional complex and promotes melanoma metastasis through AXL. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-19575-2

Pinho, S. S., & Reis, C. A. (2015). Glycosylation in cancer: Mechanisms and clinical implications. In *Nature Reviews Cancer* (Vol. 15, Issue 9, pp. 540–555). Nature Publishing Group. https://doi.org/10.1038/nrc3982

Poirier, J. T. (2017). CRISPR Libraries and Screening. In *Progress in Molecular Biology and Translational Science* (Vol. 152, pp. 69–82). Elsevier B.V. https://doi.org/10.1016/bs.pmbts.2017.10.002

Posey, A. D., Schwab, R. D., Boesteanu, A. C., Steentoft, C., Mandel, U., Engels, B., Stone, J. D., Madsen, T. D., Schreiber, K., Haines, K. M., Cogdill, A. P., Chen, T. J., Song, D., Scholler, J., Kranz, D. M., Feldman, M. D., Young, R., Keith, B., Schreiber, H., … June, C. H. (2016). Engineered CAR T Cells Targeting the Cancer-Associated Tn-Glycoform of the Membrane Mucin MUC1 Control Adenocarcinoma. *Immunity*, *44*(6), 1444–1454. https://doi.org/10.1016/j.immuni.2016.05.014

Qin, X. Y., Furutani, Y., Yonezawa, K., Shimizu, N., Kato-Murayama, M., Shirouzu, M., Xu, Y., Yamano, Y., Wada, A., Gailhouste, L., Shrestha, R., Takahashi, M., Keillor, J. W., Su, T., Yu, W., Fujii, S., Kagechika, H., Dohmae, N., Shirakami, Y., … Kojima, S. (2023). Targeting transglutaminase 2 mediated exostosin glycosyltransferase 1 signaling in liver cancer stem cells with acyclic retinoid. *Cell Death and Disease*, *14*(6). https://doi.org/10.1038/s41419-023-05847-4

Rajabzadeh, A., Hamidieh, A. A., & Rahbarizadeh, F. (2021). Spinoculation and retronectin highly enhance the gene transduction efficiency of Mucin-1-specific chimeric antigen receptor (CAR) in human primary T cells. *BMC Molecular and Cell Biology*, *22*(1). https://doi.org/10.1186/s12860-021-00397-z

Ray, S. K., & Mukherjee, S. (2024). Breast cancer stem cells as novel biomarkers. In *Clinica Chimica Acta* (Vol. 557). Elsevier B.V. https://doi.org/10.1016/j.cca.2024.117855

Reily, C., Stewart, T. J., Renfrow, M. B., & Novak, J. (2019). Glycosylation in health and disease. In *Nature Reviews Nephrology* (Vol. 15, Issue 6, pp. 346–366). Nature Publishing Group. https://doi.org/10.1038/s41581-019-0129-4

Ricardo, S., Vieira, A. F., Gerhard, R., Leitão, D., Pinto, R., Cameselle-Teijeiro, J. F., Milanezi, F., Schmitt, F., & Paredes, J. (2011). Breast cancer stem cell markers CD44, CD24 and ALDH1: Expression distribution within intrinsic molecular subtype. *Journal of Clinical Pathology*, *64*(11), 937–944. https://doi.org/10.1136/jcp.2011.090456

Ring, A., Kaur, P., & Lang, J. E. (2020). EP300 knockdown reduces cancer stem cell phenotype, tumor growth and metastasis in triple negative breast cancer. *BMC Cancer*, *20*(1). https://doi.org/10.1186/s12885-020-07573-y

Rodilla, V., & Fre, S. (2018). Cellular plasticity of mammary epithelial cells underlies heterogeneity of breast cancer. In *Biomedicines* (Vol. 6, Issue 4). MDPI AG. https://doi.org/10.3390/biomedicines6040103

Saeki, T., Nomizu, T., Toi, M., Ito, Y., Noguchi, S., Kobayashi, T., Asaga, T., Minami, H., Yamamoto, N., Aogi, K., Ikeda, T., Ohashi, Y., Sato, W., & Tsuruo, T. (2007). Dofequidar fumarate (MS-209) in combination with cyclophosphamide, doxorubicin, and fluorouracil for patients with advanced or recurrent breast cancer. *Journal of Clinical Oncology*, *25*(4), 411–417. https://doi.org/10.1200/JCO.2006.08.1646

Samocha, A., Doh, H., Kessenbrock, K., & Roose, J. P. (2019). Unraveling heterogeneity in epithelial cell fates of the mammary gland and breast cancer. In *Cancers* (Vol. 11, Issue 10). MDPI AG. https://doi.org/10.3390/cancers11101423

Sanjana, N. E. (2017). Genome-scale CRISPR pooled screens. *Analytical Biochemistry*, *532*, 95–99. https://doi.org/10.1016/j.ab.2016.05.014

Sanjana, N. E., Shalem, O., & Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. In *Nature Methods* (Vol. 11, Issue 8, pp. 783–784). Nature Publishing Group. https://doi.org/10.1038/nmeth.3047

Sanson, K. R., Hanna, R. E., Hegde, M., Donovan, K. F., Strand, C., Sullender, M. E., Vaimberg, E. W., Goodale, A., Root, D. E., Piccioni, F., & Doench, J. G. (2018). Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-07901-8

Sarhangi, N., Hajjari, S., Heydari, S. F., Ganjizadeh, M., Rouhollah, F., & Hasanzad, M. (2022). Breast cancer in the era of precision medicine. In *Molecular Biology Reports* (Vol. 49, Issue 10, pp. 10023–10037). Springer Science and Business Media B.V. https://doi.org/10.1007/s11033-022-07571-2

Sato, F., Sagara, A., Tajima, K., Miura, S., Inaba, K., Ando, Y., Oku, T., Murakami, T., Kato, Y., & Yumoto, T. (2022). COL8A1 facilitates the growth of triple-negative breast cancer via FAK/Src activation. *Breast Cancer Research and Treatment*, *194*(2), 243–256. https://doi.org/10.1007/s10549-022-06635-y

Schjoldager, K. T., Narimatsu, Y., Joshi, H. J., & Clausen, H. (2020). Global view of human protein glycosylation pathways and functions. In *Nature Reviews Molecular Cell Biology* (Vol. 21, Issue 12, pp. 729–749). Nature Research. https://doi.org/10.1038/s41580-020-00294-x

Schmierer, B., Botla, S. K., Zhang, J., Turunen, M., Kivioja, T., & Taipale, J. (2017). CRISPR/Cas9 screening using unique molecular identifiers. *Molecular Systems Biology*, *13*(10). https://doi.org/10.15252/msb.20177834

Schraivogel, D., Steinmetz, L. M., & Parts, L. (2023). *Pooled Genome-Scale CRISPR Screens in Single Cells*. https://doi.org/10.1146/annurev-genet-072920

Schultz, M. J., Holdbrooks, A. T., Chakraborty, A., Grizzle, W. E., Landen, C. N., Buchsbaum, D. J., Conner, M. G., Arend, R. C., Yoon, K. J., Klug, C. A., Bullard, D. C., Kesterson, R. A., Oliver, P. G., O'Connor, A. K., Yoder, B. K., & Bellis, S. L. (2016). The tumor-associated glycosyltransferase ST6Gal-I regulates stem cell transcription factors and confers a cancer stem cell phenotype. *Cancer Research*, *76*(13), 3978–3988. https://doi.org/10.1158/0008-5472.CAN-15-2834

Scott, D. A., & Drake, R. R. (2019). Glycosylation and its implications in breast cancer. In *Expert Review of Proteomics* (Vol. 16, Issue 8, pp. 665–680). Taylor and Francis Ltd. https://doi.org/10.1080/14789450.2019.1645604

Seldin, L., Le Guelte, A., & Macara, I. G. (2017). Epithelial plasticity in the mammary gland. In *Current Opinion in Cell Biology* (Vol. 49, pp. 59–63). Elsevier Ltd. https://doi.org/10.1016/j.ceb.2017.11.012

Shang, W., Wang, F., Fan, G., & Wang, H. (2017). Key elements for designing and performing a CRISPR/Cas9-based genetic screen. In *Journal of Genetics and Genomics* (Vol. 44, Issue 9, pp. 439–449). Institute of Genetics and Developmental Biology. https://doi.org/10.1016/j.jgg.2017.09.005

Sharma, A., Paranjape, A. N., Rangarajan, A., & Dighe, R. R. (2012). A monoclonal antibody against human notch1 ligand-binding domain depletes subpopulation of putative breast cancer stem-like cells. *Molecular Cancer Therapeutics*, *11*(1), 77–86. https://doi.org/10.1158/1535-7163.MCT-11-0508

Sharma, S., & Petsalaki, E. (2018). Application of CRISPR-Cas9 based genome-wide screening approaches to study cellular signalling mechanisms. In *International Journal of Molecular Sciences* (Vol. 19, Issue 4). MDPI AG. https://doi.org/10.3390/ijms19040933

Sheridan, C., Kishimoto, H., Fuchs, R. K., Mehrotra, S., Bhat-Nakshatri, P., Turner, C. H., Goulet, R., Badve, S., & Nakshatri, H. (2006). CD44+/CD24-Breast cancer cells exhibit enhanced invase properties: An early step necessary for metastasis. *Breast Cancer Research*, *8*(5). https://doi.org/10.1186/bcr1610

Shi, H., Doench, J. G., & Chi, H. (2023). CRISPR screens for functional interrogation of immunity. In *Nature Reviews Immunology* (Vol. 23, Issue 6, pp. 363–380). Nature Research. https://doi.org/10.1038/s41577-022-00802-4

Shultz, L. D., Ishikawa, F., & Greiner, D. L. (2007). Humanized mice in translational biomedical research. In *Nature Reviews Immunology* (Vol. 7, Issue 2, pp. 118–130). https://doi.org/10.1038/nri2017

Singh, A. K., & Mcguirk, J. P. (2020). CAR T cells: continuation in a revolution of immunotherapy. In *Historical Review Lancet Oncol* (Vol. 21). www.thelancet.com/oncology

Slepicka, P. F., Somasundara, A. V. H., & dos Santos, C. O. (2021). The molecular basis of mammary gland development and epithelial differentiation. In *Seminars in Cell and Developmental Biology* (Vol. 114, pp. 93–112). Elsevier Ltd. https://doi.org/10.1016/j.semcdb.2020.09.014

Smith, B. A. H., & Bertozzi, C. R. (2021). The clinical impact of glycobiology: targeting selectins, Siglecs and mammalian glycans. In *Nature Reviews Drug Discovery* (Vol. 20, Issue 3, pp. 217–243). Nature Research. https://doi.org/10.1038/s41573-020-00093-1

Smolarz, B., Zadrożna Nowak, A., & Romanowicz, H. (2022). Breast Cancer—Epidemiology, Classification, Pathogenesis and Treatment (Review of Literature). In *Cancers* (Vol. 14, Issue 10). MDPI. https://doi.org/10.3390/cancers14102569

Solaimuthu, B., Khatib, A., Tanna, M., Karmi, A., Hayashi, A., Rmaileh, A. A., Lichtenstein, M., Takoe, S., Jolly, M. K., & Shaul, Y. D. (2024). The exostosin glycosyltransferase 1/STAT3 axis is a driver of breast cancer aggressiveness. *Proceedings of the National Academy of Sciences of the United States of America*, *121*(3). https://doi.org/10.1073/pnas.2316733121

Spiro, R. G. (2002). MINI REVIEW Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. In *Glycobiology* (Vol. 12, Issue 4).

Stewart, R. L., Updike, K. L., Factor, R. E., Henry, N. L., Boucher, K. M., Bernard, P. S., & Varley, K. E. (2019). A multigene assay determines risk of recurrence in patients with triple-negative breast cancer. *Cancer Research*, *79*(13), 3466–3478. https://doi.org/10.1158/0008-5472.CAN-18-3014

Su, D., Feng, X., Colic, M., Wang, Y., Zhang, C., Wang, C., Tang, M., Hart, T., & Chen, J. (2020). CRISPR/CAS9-based DNA damage response screens reveal gene-drug interactions. *DNA Repair*, *87*. https://doi.org/10.1016/j.dnarep.2020.102803

Suhail, Y., Cain, M. P., Vanaja, K., Kurywchak, P. A., Levchenko, A., Kalluri, R., & Kshitiz. (2019). Systems Biology of Cancer Metastasis. In *Cell Systems* (Vol. 9, Issue 2, pp. 109–127). Cell Press. https://doi.org/10.1016/j.cels.2019.07.003

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249. https://doi.org/10.3322/caac.21660

Sun, Y. S., Zhao, Z., Yang, Z. N., Xu, F., Lu, H. J., Zhu, Z. Y., Shi, W., Jiang, J., Yao, P. P., & Zhu, H. P. (2017). Risk factors and preventions of breast cancer. In *International Journal of Biological Sciences* (Vol. 13, Issue 11, pp. 1387–1397). Ivyspring International Publisher. https://doi.org/10.7150/ijbs.21635

Surveillance Research Program, N. C. I. (2023, November 16). *SEER*Explorer: An interactive website for SEER cancer statistics [Internet]*. SEER*Explorer: An Interactive Website for SEER Cancer Statistics [Internet]. Surveillance Research Program, National Cancer Institute; 2023 Apr 19. Available from: Https://Seer.Cancer.Gov/Statistics-Network/Explorer/. Data Source(s): SEER Incidence Data. https://seer.cancer.gov/statistics-network/explorer/application.html?site=55&data_type=1&graph_type=3&compareBy=sex&chk_sex_3=3&rate_type=2&race=1&advopt_precision=1&advopt_show_ci=on&hdn_view=0&advopt_show_apc=on&advopt_display=2#resultsRegion0

Su, Y., Pogash, T. J., Nguyen, T. D., & Russo, J. (2016). Development and characterization of two human triple-negative breast cancer cell lines with highly tumorigenic and metastatic capabilities. *Cancer Medicine*, *5*(3), 558–573. https://doi.org/10.1002/cam4.616

Szlachta, K., Kuscu, C., Tufan, T., Adair, S. J., Shang, S., Michaels, A. D., Mullen, M. G., Fischer, N. L., Yang, J., Liu, L., Trivedi, P., Stelow, E. B., Stukenberg, P. T., Parsons, J. T., Bauer, T. W., & Adli, M. (2018). CRISPR knockout screening identifies combinatorial drug targets in pancreatic cancer and models cellular drug response. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-06676-2

Takashi Ishio, Sarvesh Kumar, Joji Shimono, Anusara Daenthanasanmak, Sigrid Dubois, Yuquan Lin, Bonita Bryant, Michael N Petrus, Emmanuel Bachy, Da Wei Huang, Yandan Yang, Patrick L Green, Hiroo Hasegawa, Michiyuki Maeda, Hideki Goto, Tomoyuki Endo, Takashi Yokota, Kanako C Hatanaka, Yutaka Hatanaka, … Masao Nakagawa. (2022). Genome-wide CRISPR screen identifies CDK6 as a therapeutic target in adult T-cell leukemia/lymphoma. *Blood*, *139*(10), 1541–1556. https://doi.org/10.1182/BLOOD.2019000962

Target Discovery institute. (2024, September 23). *CRISPR loss of function screening*. Https://Www.Tdi.Ox.Ac.Uk/Research/Research/Cellular-High-Throughput-Screening-Hts/Crispr-Pooled-Screening/Crispr-Loss-of-Function-Screening.

Taurin, S., & Alkhalifa, H. (2020). Breast cancers, mammary stem cells, and cancer stem cells, characteristics, and hypotheses. In *Neoplasia (United States)* (Vol. 22, Issue 12, pp. 663–678). Elsevier Inc. https://doi.org/10.1016/j.neo.2020.09.009

Tian, J., Hachim, M. Y., Hachim, I. Y., Dai, M., Lo, C., Raffa, F. Al, Ali, S., & Lebrun, J. J. (2017). Cyclooxygenase-2 regulates TGFβ-induced cancer stemness in triple-negative breast cancer. *Scientific Reports*, *7*. https://doi.org/10.1038/srep40258

Tsang, J. Y. S., & Tse, G. M. (2019). *Molecular Classification of Breast Cancer*. http://journals.lww.com/anatomicpathology

Tu, C. F., Wu, M. Y., Lin, Y. C., Kannagi, R., & Yang, R. B. (2017). FUT8 promotes breast cancer cell invasiveness by remodeling TGF-β receptor core fucosylation. *Breast Cancer Research*, *19*(1). https://doi.org/10.1186/s13058-017-0904-8

Tucker-Burden, C., Chappa, P., Krishnamoorthy, M., Gerwe, B. A., Scharer, C. D., Heimburg-Molinaro, J., Harris, W., Usta, S. N., Eilertson, C. D., Hadjipanayis, C. G., Stice, S. L., Brat, D. J., & Nash, R. J. (2012). Lectins identify glycan biomarkers on glioblastoma-derived cancer stem cells. *Stem Cells and Development*, *21*(13), 2374–2386. https://doi.org/10.1089/scd.2011.0369

*Tumorsphere Culture of Human Breast Cancer Cell Lines*. (n.d.). Retrieved August 5, 2024, from https://www.stemcell.com/tumorsphere-culture-human-breast-cancer-cell-lines-lp.html

Vajaria, B. N., & Patel, P. S. (2017). Glycosylation: a hallmark of cancer? In *Glycoconjugate Journal* (Vol. 34, Issue 2, pp. 147–156). Springer New York LLC. https://doi.org/10.1007/s10719-016-9755-2

Varki, A. (2017). Biological roles of glycans. *Glycobiology*, *27*(1), 3–49. https://doi.org/10.1093/glycob/cww086

Vasconcelos-dos-Santos, A., Oliveira, I. A., Lucena, M. C., Mantuano, N. R., Whelan, S. A., Dias, W. B., & Todeschini, A. R. (2015). Biosynthetic machinery involved in aberrant glycosylation: Promising targets for developing of drugs against cancer. In *Frontiers in Oncology* (Vol. 5, Issue JUN). Frontiers Media S.A. https://doi.org/10.3389/fonc.2015.00138

Vikram, R., Chou, W. C., Hung, S. C., & Shen, C. Y. (2020). Tumorigenic and metastatic role of cd44−/low/cd24−/low cells in luminal breast cancer. *Cancers*, *12*(5). https://doi.org/10.3390/cancers12051239

Visvader, J. E., & Lindeman, G. J. (2006). Mammary stem cells and mammopoiesis. In *Cancer Research* (Vol. 66, Issue 20, pp. 9798–9801). https://doi.org/10.1158/0008-5472.CAN-06-2254

Walsh, H. R., Cruickshank, B. M., Brown, J. M., & Marcato, P. (2019). The flick of a switch: Conferring survival advantage to breast cancer stem cells through metabolic plasticity. In *Frontiers in Oncology* (Vol. 9, Issue AUG). Frontiers Media S.A. https://doi.org/10.3389/fonc.2019.00753

Wang, C., Feng, X., Su, D., Chen, Z., Wang, S., Tang, M., Huang, M., Nie, L., Zhang, H., Li, S., Yin, L., Johnson, R. L., Hart, T., & Chen, J. (2022). *Integrated screens uncover a cell surface tumor suppressor gene KIRREL involved in Hippo pathway*. https://doi.org/10.1073/pnas

Wang, R., Zhu, Y., Liu, X., Liao, X., He, J., & Niu, L. (2019). The Clinicopathological features and survival outcomes of patients with different metastatic sites in stage IV breast cancer. *BMC Cancer*, *19*(1). https://doi.org/10.1186/s12885-019-6311-z

Wang, S. W., Gao, C., Zheng, Y. M., Yi, L., Lu, J. C., Huang, X. Y., Cai, J. Bin, Zhang, P. F., Cui, Y. H., & Ke, A. W. (2022). Current applications and future perspective of CRISPR/Cas9 gene editing in cancer. In *Molecular Cancer* (Vol. 21, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s12943-022-01518-8

Wang, X., Zhang, Y., Lin, H., Liu, Y., Tan, Y., Lin, J., Gao, F., & Lin, S. (2017). Alpha2,3-sialyltransferase III knockdown sensitized ovarian cancer cells to cisplatin-induced apoptosis. *Biochemical and Biophysical Research Communications*, *482*(4), 758–763. https://doi.org/10.1016/j.bbrc.2016.11.107

Ward, N. P., Yoon, S. J., Flynn, T., Sherwood, A. M., Olley, M. A., Madej, J., & DeNicola, G. M. (2024). Mitochondrial respiratory function is preserved under cysteine starvation via glutathione catabolism in NSCLC. *Nature Communications*, *15*(1). https://doi.org/10.1038/s41467-024-48695-2

Warrier, N. M., Kelkar, N., Johnson, C. T., Govindarajan, T., Prabhu, V., & Kumar, P. (2023). Understanding cancer stem cells and plasticity: Towards better therapeutics. *European Journal of Cell Biology*, *102*(2). https://doi.org/10.1016/j.ejcb.2023.151321

Watson, C. J., & Khaled, W. T. (2020). Mammary development in the embryo and adult: New insights into the journey of morphogenesis and commitment. In *Development (Cambridge)* (Vol. 147, Issue 22). Company of Biologists Ltd. https://doi.org/10.1242/dev.169862

Weiswald, L. B., Bellet, D., & Dangles-Marie, V. (2015). Spherical Cancer Models in Tumor Biology. In *Neoplasia (United States)* (Vol. 17, Issue 1, pp. 1–15). Neoplasia Press, Inc. https://doi.org/10.1016/j.neo.2014.12.004

Wicker, M. N., & Wagner, K. U. (2023). Cellular Plasticity in Mammary Gland Development and Breast Cancer. In *Cancers* (Vol. 15, Issue 23). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/cancers15235605

Williams, R. T., Guarecuco, R., Gates, L. A., Barrows, D., Passarelli, M. C., Carey, B., Baudrier, L., Jeewajee, S., La, K., Prizer, B., Malik, S., Garcia-Bermudez, J., Zhu, X. G., Cantor, J., Molina, H., Carroll, T., Roeder, R. G., Abdel-Wahab, O., Allis, C. D., & Birsoy, K. (2020). ZBTB1 Regulates Asparagine Synthesis and Leukemia Cell Response to L-Asparaginase. *Cell Metabolism*, *31*(4), 852-861.e6. https://doi.org/10.1016/j.cmet.2020.03.008

Wong, D. J., Liu, H., Ridky, T. W., Cassarino, D., Segal, E., & Chang, H. Y. (2008). Module Map of Stem Cell Genes Guides Creation of Epithelial Cancer Stem Cells. *Cell Stem Cell*, *2*(4), 333–344. https://doi.org/10.1016/j.stem.2008.02.009

Wright, M. H., Calcagno, A. M., Salcido, C. D., Carlson, M. D., Ambudkar, S. V., & Varticovski, L. (2008). Brca1 breast tumors contain distinct CD44+/CD24-and CD133+cells with cancer stem cell characteristics. *Breast Cancer Research*, *10*(1). https://doi.org/10.1186/bcr1855

Wu, T., Wang, Y., Xiao, T., Ai, Y., Li, J., Zeng, Y. A., & Yu, Q. C. (2021). Lentiviral CRISPR-guided RNA library screening identified Adam17 as an upstream negative regulator of Procr in mammary epithelium. *BMC Biotechnology*, *21*(1). https://doi.org/10.1186/s12896-021-00703-9

Wu, X., Zhao, J., Ruan, Y., Sun, L., Xu, C., & Jiang, H. (2018). Sialyltransferase ST3GAL1 promotes cell migration, invasion, and TGF-β1-induced EMT and confers paclitaxel resistance in ovarian cancer. *Cell Death and Disease*, *9*(11). https://doi.org/10.1038/s41419-018-1101-0

Xiao, H., Woods, E. C., Vukojicic, P., & Bertozzi, C. R. (2016). Precision glycocalyx editing as a strategy for cancer immunotherapy. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(37), 10304–10309. https://doi.org/10.1073/pnas.1608069113

Xu, J., Gao, F., Liu, W., & Guan, X. (2024). Cell-cell communication characteristics in breast cancer metastasis. In *Cell Communication and Signaling* (Vol. 22, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s12964-023-01418-4

Xu, X., Zhang, M., Xu, F., & Jiang, S. (2020). Wnt signaling in breast cancer: biological mechanisms, challenges and opportunities. In *Molecular Cancer* (Vol. 19, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s12943-020-01276-5

Yin, L., Hu, X., Pei, G., Tang, M., Zhou, Y., Zhang, H., Huang, M., Li, S., Zhang, J., Citu, C., Zhao, Z., Debeb, B. G., Feng, X., & Chen, J. (2024). Genome-wide CRISPR screen reveals the synthetic lethality between BCL2L1 inhibition and radiotherapy. *Life Science Alliance*, *7*(4). https://doi.org/10.26508/lsa.202302353

Yin, X. G., Chen, X. Z., Sun, W. M., Geng, X. S., Zhang, X. K., Wang, J., Ji, P. P., Zhou, Z. Y., Baek, D. J., Yang, G. F., Liu, Z., & Guo, J. (2017). IgG Antibody Response Elicited by a Fully Synthetic Two-Component Carbohydrate-Based Cancer Vaccine Candidate with α-Galactosylceramide as Built-in Adjuvant. *Organic Letters*, *19*(3), 456–459. https://doi.org/10.1021/acs.orglett.6b03591

Yu, J., Liao, X., Li, L., Lv, L., Zhi, X., Yu, J., & Zhou, P. (2017). A preliminary study of the role of extracellular -5'- nucleotidase in breast cancer stem cells and epithelial-mesenchymal

transition. *In Vitro Cellular and Developmental Biology - Animal*, *53*(2), 132–140. https://doi.org/10.1007/s11626-016-0089-y

Zeng, X., Liu, C., Yao, J., Wan, H., Wan, G., Li, Y., & Chen, N. (2021). Breast cancer stem cells, heterogeneity, targeting therapies and therapeutic implications. In *Pharmacological Research* (Vol. 163). Academic Press. https://doi.org/10.1016/j.phrs.2020.105320

Zhang, B. L., Li, D., Gong, Y. L., Huang, Y., Qin, D. Y., Jiang, L., Liang, X., Yang, X., Gou, H. F., Wang, Y. S., Wei, Y. Q., & Wang, W. (2019). Preclinical Evaluation of Chimeric Antigen Receptor-Modified T Cells Specific to Epithelial Cell Adhesion Molecule for Treating Colorectal Cancer. *Human Gene Therapy*, *30*(4), 402–412. https://doi.org/10.1089/hum.2018.229

Zhang, D., Yang, L., Liu, X., Gao, J., Liu, T., Yan, Q., & Yang, X. (2020). Hypoxia modulates stem cell properties and induces EMT through N-glycosylation of EpCAM in breast cancer cells. *Journal of Cellular Physiology*, *235*(4), 3626–3633. https://doi.org/10.1002/jcp.29252

Zhang, L., Chen, W., Liu, S., & Chen, C. (2023). Targeting Breast Cancer Stem Cells. In *International Journal of Biological Sciences* (Vol. 19, Issue 2, pp. 552–570). Ivyspring International Publisher. https://doi.org/10.7150/ijbs.76187

Zhang, R., Tu, J., & Liu, S. (2022). Novel molecular regulators of breast cancer stem cell plasticity and heterogeneity. In *Seminars in Cancer Biology* (Vol. 82, pp. 11–25). Academic Press. https://doi.org/10.1016/j.semcancer.2021.03.008

Zhan, T., Rindtorff, N., Betge, J., Ebert, M. P., & Boutros, M. (2019). CRISPR/Cas9 for cancer research and therapy. In *Seminars in Cancer Biology* (Vol. 55, pp. 106–119). Academic Press. https://doi.org/10.1016/j.semcancer.2018.04.001

Zhao, Y., Yu, L., Wu, X., Li, H., Coombes, K. R., Au, K. F., Cheng, L., & Li, L. (2022). CEDA: integrating gene expression data with CRISPR-pooled screen data identifies essential genes with higher expression. *Bioinformatics*, *38*(23), 5245–5252. https://doi.org/10.1093/bioinformatics/btac668

Zhao, Y., Zhang, M., & Yang, D. (2022). *Bioinformatics approaches to analyzing CRISPR screen data: from dropout screens to single-cell CRISPR screens*.

Zheng, Q., Zhang, M., Zhou, F., Zhang, L., & Meng, X. (2021). The Breast Cancer Stem Cells Traits and Drug Resistance. In *Frontiers in Pharmacology* (Vol. 11). Frontiers Media S.A. https://doi.org/10.3389/fphar.2020.599965

Zhong, C., Jiang, W. J., Yao, Y., Li, Z., Li, Y., Wang, S., Wang, X., Zhu, W., Wu, S., Wang, J., Fan, S., Ma, S., Liu, Y., Zhang, H., Zhao, W., Zhao, L., Feng, Y., Li, Z., Guo, R., … Fei, T. (2024). CRISPR screens reveal convergent targeting strategies against evolutionarily distinct chemoresistance in cancer. *Nature Communications*, *15*(1). https://doi.org/10.1038/s41467-024-49673-4

Zhou, J. Y., Oswald, D. M., Oliva, K. D., Kreisman, L. S. C., & Cobb, B. A. (2018). The Glycoscience of Immunity. In *Trends in Immunology* (Vol. 39, Issue 7, pp. 523–535). Elsevier Ltd. https://doi.org/10.1016/j.it.2018.04.004

Zhou, X., Chi, K., Zhang, C., Liu, Q., & Yang, G. (2023). Sialylation: A Cloak for Tumors to Trick the Immune System in the Microenvironment. In *Biology* (Vol. 12, Issue 6). MDPI. https://doi.org/10.3390/biology12060832

Zhou Y, F. Q. S. H. Z. G. (2022). CRISPR Guide RNA Library Screens in Human Induced Pluripotent Stem Cells. In *Methods Mol Biol* (pp. 233–257). http://www.springer.com/series/7651

Zhu, Y. (2022). Advances in CRISPR/Cas9. In *BioMed Research International* (Vol. 2022). Hindawi Limited. https://doi.org/10.1155/2022/9978571

Zhu, Y., Groth, T., Kelkar, A., Zhou, Y., & Neelamegham, S. (2021). A GlycoGene CRISPR-Cas9 lentiviral library to study lectin binding and human glycan biosynthesis pathways. *Glycobiology*, *31*(3), 173–180. https://doi.org/10.1093/glycob/cwaa074

Zhu, Z. W., Chen, L., Liu, J. X., Huang, J. W., Wu, G., Zheng, Y. F., & Yao, K. T. (2018). A novel three-dimensional tumorsphere culture system for the efficient and low-cost enrichment of cancer stem cells with natural polymers. *Experimental and Therapeutic Medicine*, *15*(1), 85–92. https://doi.org/10.3892/etm.2017.5419

Zubair, M., Wang, S., & Ali, N. (2021). Advanced Approaches to Breast Cancer Classification and Diagnosis. In *Frontiers in Pharmacology* (Vol. 11). Frontiers Media S.A. https://doi.org/10.3389/fphar.2020.632079

# Annexes

*Annex 1: Guide sequences of the positive and negative controls of the GlycoCRISPR library*

| Type of control | Gene name | Gene | Guide sequence | | | | |
|---|---|---|---|---|---|---|---|
| **Positive** | Octamer-binding transcription factor 4, | OCT4 | TTCTGTCGTTCACTGGCAGG | CCTGCCAGTGAACGACAGAA | ACCCTGGGGGTTCTATTTGG | AAATCTTCAGGAGGTAAGGG | CCCACCAAATAGAACCCCCA |
| | | | ACTGGCAGGTGGTCCGAGTG | TTGGCTGAATACCTTCCCTG | ACCCACCAAATAGAACCCCC | GTTTGGCTGAATACCTTCCC | GTGAGTGCCATGTCTCTCTG |
| | SRY-box 2 | SOX2 | GGGTTCGGTGGTCAAGTCCG | ATTATAAATACCGGCCCCGG | TTTTGTCGGAGACGGAGAAG | CGGATTATAAATACCGGCCC | ATGTCCCAGCACTACCAGAG |
| | | | TGAGCGTCTTGGTTTTCCGC | CATGTATCTCCCCGGCGCCG | GCTGTTTTTCTGGTTGCCGC | AAAGTTTCCACTCGGCGCCC | CGGCAATAGCATGGCGAGCG |
| | Krüppel-like factor 4 | KLF4 | TCGGTCATCAGCGTCAGCAA | CGGACTCCCTGCCATAGAGG | TGACCTTGGTAATGGAGCGG | ACTTGTGATTACGCGGGCTG | AATGAGGTAGGTGAGGCGCG |
| | | | GTGGTGGTGGCGCCCTACAA | AAGGATGGGTAATTGGGCCC | TCTCAGCAATGGCCACCGGC | CTGCACACGACTTCCCCCTG | CCGGTGGCCATTGCTGAGAG |
| | MYC proto-oncogene | cMYC | CTCTGAGACGAGCTTGGCGG | GGGGTCGATGCACTCTGAGG | CAGGCTGCGCGCAAAGACAG | GTCGATGCACTCTGAGGCGG | CTGGAGCGGGGCACACAAAG |
| | | | GGTGTGACCGCAACGTAGGA | TCCCTTCGGGGAGACAACGA | TGTTGGTGAAGCTAACGTTG | CAGATATCCTCGCTGGGCGC | GAGAGCAGAGAATCCGAGGA |
| | DNA methyltransferase 1 | DNMT1 | AGGAGAACGCCTTTAAGCGC | TTGAACGTGAAGGCCTCAGG | ACAGGCCGGTCAGTACGGCG | GTAGGAGATCTCCAGTGCCG | CAACTTGAACCGCTTCACAG |
| | | | AGATCAAGCTGCCCAAGCTG | GTCCAGGTGTACTCCATGGG | ATCGAGGTGCGGCTCTCAGA | AATGTATTCGGCAAATGCTG | CTGTGAAGCGGTTCAAGTTG |
| | Fascin actin-bundling protein 1 | FSCN1 | CTGCGGCAACAAGTACCTGA | CGCTCTGGGAGTACTAGGGC | GTGACCGGCGCATCACACTG | AGGCACTCACTTGCTGGAGG | CACCTTGAACCCGAACGCCT |
| | | | CATCGGCTGCCGCAAGGTCA | CGACCACGCAGGCGTCCTGA | GTGTGCCTTCCGTACCCACA | TACTGGACGCTGACGGCCAC | ATCGGCTGCCGCAAGGTCAC |
| | Notch receptor 1 | NOTCH1 | CATCGGGCACCTGAACGTGG | GAAGAACAGAAGCACAAAGG | ATCAGAGCGTGAGTAGCGGG | TCCTGCCAGAACACCCACGG | TCCTCGCCCTGCAAGAACGG |
| | | | CAGCCTCAACGGGTATGCGG | GCTGCACTTCATGTACGTGG | ACCAATACAACCCTCTGCGG | GGAAACAACTGCAAGAACGG | ACCTGCCACAACACCCACGG |
| | LDL receptor related protein 6 | LRP6 | GAGCGTGCCAACAAAACCAG | AGAGGACTTAGAGGAACCCC | GCAGTTTACTTGTTTCACGG | CTGAAGGCATGTATATCCAT | TGGCACGCTCAATGCTGCGT |
| | | | CTCGCGTTGGACCCTGCCGA | TTGTGGTAAACCCAGAGAAA | GTTGTTCATCAATCACCATG | CACAAGTCCAGTAGATGTAG | TGTTGCTTTATGCAAACAGA |
| | Smoothened, frizzled class receptor | SMO | CTCTGGTCGGGTAAGTGCGG | GAGCTCGTGCCGCTTAGAGA | AGGCACACGTTGTAGCGCAG | CACGGCAGACGATCTCTCGG | ACGGCAGACGATCTCTCGGC |
| | | | TGTCAGGCCAATGTGACCAT | CCACATTCGTGGCTGACTGG | GCAGTCGAGGAATGGTACTG | GGCGATGCCAGTTCCAAACA | ATCCGAGGTGAGTGAAGACC |
| | TNF receptor superfamily member 11a (RANK) | TNFRSF11A | TGGGACAGAGAAATCCGATG | CCAGCTAGAAAACCACCAAA | AGAAGAACTGCAAACCGCAT | CTACCTTATCTCCACTTAGG | AGCTATCCAAGTATTCATCC |
| | | | TGGCCGCCTAAGTGGAGATA | CAAGTATTCATCCGGGCCAC | GCACACATCCAACCCGTACA | TCCAACCCGTACACGGGTGC | GTAAACATGGGGTTCTGTAT |
| | C-X-C motif chemokine receptor 1 | CXCR1 | GAAATGACACAGCAAAATGG | TGTTTGGATGGTAAGCCTGG | TGACACAGCAAAATGGCGGA | GCCGATGAAGGCGTAGATGA | CAACAACATCGGCCGGGCCC |
| | | | CAGAACAGCATGACAAACAG | AGCGCCGCAACAACATCGGC | TGTGAGCGCCGCAACAACAT | TCTCAGTTTCTAGCATACAG | CAGCAGGAACACTAGGGCAT |
| | Beclin 1 | BECN1 | CGAGAGACACCATCCTGGCG | GGTTTCCGTAAGGAACAAGT | TCGCCTGCCCAGTGTTCCCG | AAACTCGTGTCCAGTTTCAG | GATTTTCTGCCACTATCTTG |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | GCCGAAGACTGAAGGCAAGT | TGAGTTCCTGGATGGTGACA | AACCCCCCAGAACAGTATAA | ATTTATTGAAACTCCTCGCC | GCGTTATGCCCAGACGCAGC |
| **Negative** | Addgene BRDN0001149383 | GCTCCCATCCATAGTAAAAA | | | | |
| | Addgene BRDN0001145482 | GTGAACTGCAATCTTATTAT | | | | |
| | Addgene BRDN0001149198 | GTATTACTGATATTGGTGGG | | | | |
| | HGLibA_65383 | GACAATCATGGTGAAAGCGG | | | | |
| | HGLibA_65382 | CTGAGTGAAAAATAAAAGTT | | | | |
| | HGLibA_65381 | TTTCCCATGATCATTTAGTG | | | | |
| | HGLibA_65380 | TAAACAAAAAGGAAATAGTT | | | | |
| | HGLibA_65379 | ATATTTTATGACATAAAAAT | | | | |
| | HGLibA_65378 | GACTGAAATCCAAGGACTGT | | | | |
| | HGLibA_65377 | AGAAGAAAAAAATGTCTACG | | | | |
| | HGLibA_65376 | GAGAAGTGGGGAGCCATTGG | | | | |
| | HGLibA_65375 | AAAGAAAGAGGAATAGTAGC | | | | |
| | HGLibA_65374 | GCCCCGCCGCCCTCCCCTCC | | | | |

*Annex 2: Guides of all protein glycosylation genes included in our GlycoCRISPR library*

| Gene name | Gene | Guide sequence | | | | |
|---|---|---|---|---|---|---|
| **Lactosylceramide 4-alpha-galactosyltransferase** | A4GALT | GGAGGCGCTACTATGAAGTG | ACGGAACAGGAGACCAAATG | TATGTGGAGTTTGAGGGCAT | GAGGCGCTACTATGAAGTGT | GGGCCTGATGGAGATGACAT |
| | | GCCACCGTATTTCCAGATGA | ACCCCATCTCCTATCGAGAG | TCCACTCTCGATAGGAGATG | GTAGCGCCTCCACTCTCGAT | GCTTCTCGGTACTCTAGTAA |
| **Histo-blood group ABO system transferase** | ABO | GAGCCATGGCCGAGGTGTTG | GCTTCCTGGCATTAGACTTC | GCAACGAGACGCGCTGCAGA | GGTCGGTGCAAGAGGTGCAG | AGGTGCGCGCCTACAAGCGC |
| | | AGGCCTTCACCTACGAGCGC | GCGATTTCTACTACCTGGGG | ATGATGGTCGACCAGGCCAA | TGATCAGTGACTTCTGCGAG | CCGGCTGCTTCCGTAGAAGC |
| **Alkaline ceramidase 2** | ACER2 | CCGAGTTCTACAACACGGTG | CGAGTTCTACAACACGGTGC | GACTCTCTGGACCCCCGTGC | CGCATGTTGTCACACCTGCA | CAAAGATCTTTCGGAATGAC |
| | | AATGTTTAGTGCAAGCTTAC | CATGTTGTCACACCTGCACG | TATCGCCGAGTTCTACAACA | TTTTGGTTGTAGTGGGTAAG | TACAACACGGTGCGGGGCGC |
| **A disintegrin and metalloproteinase with thrombospondin motifs 13** | ADAMTS13 | GAGAACTCTCAAGGACGCGG | TTTGTCTGCAGGTTTACAGG | ATGCGGATCTCCTCCAGGCG | CCTACTTCCAGCCTAAGCCA | GATCGGAGGGCGCTATGTCG |
| | | GGTTTACAGGCGGTATGGCG | GTACAGAGTGGCCCTCACCG | GAGCCACATTAATGAAGAGC | CGACACGACCATCCTCCAGG | GGTGGGCTCATTTGCAGGAG |
| **A disintegrin and metalloproteinase with thrombospondin motifs 5** | ADAMTS5 | GACCAACTCTACTCCGGCGG | CTGCACGTCTACACCCGCGA | ATGGCGCGGTTGTATGGCCG | ATCGACCAACTCTACTCCGG | GTAATTCCGTCTGCGTCCGG |
| | | GCCATGATTCAGGTGACCGA | TAGCACCACCACCTTCACCA | GTGGGCTACCTCGTCTACGC | GCTGTACAGCCTATTGGCGA | CACCTGCCCAGGATAAAGCC |
| **A disintegrin and metalloproteinase with thrombospondin motifs 7** | ADAMTS7 | TGGCCGGACAGGGTACATGG | GGAACCTGGTGATATACTGG | TCTACGAGCTACAATACCGC | TTCTACGAGCTACAATACCG | TGGGCTACAGTGAAGGCCAG |
| | | GGATGTGGGGCTGATCCCAG | GAGCTCGTGGGCTACAGTGA | GAGGCGTCTACACCAGCGGT | TGATATACTGGCGGCTGCAG | GATGAAAGGTCGTTTCCCAA |
| **ADAMTS-like protein 1** | ADAMTSL1 | TTTGTGATTAAGCTCATCGG | TGTTATGCAGGCCCATGCAG | GAATGCTCACGCACCTGCGG | CTGTATGTAGACCTTCCTGG | GTGGATCTTGAGATAGCCGA |
| | | ACGGAGATGGGTCCACCTGC | GATGATTGATTTACGTACCG | GTGTGTCTTACATTTGGTTG | ACCGTCAAGGAAGATAACTG | GTGAATGCTCACGCACCTGC |
| **ADAMTS-like protein 4** | ADAMTSL4 | ACTGTGTCCTGTACAAGGGG | TGCCGGCAGGAATTTGGGGG | GGATGAACGCAGCTGTGCCG | TCGGGGAACCTCACTGACCG | TCGGAGGAGCACTAGTGTAG |
| | | ACTGTGTCCTGTACAAGGGG | TGCCGGCAGGAATTTGGGGG | GGATGAACGCAGCTGTGCCG | TCGGGGAACCTCACTGACCG | TCGGAGGAGCACTAGTGTAG |
| **Activity-dependent neuroprotector homeobox protein** | ADNP | CTGGCCCGATGAGAGAGAAG | AATCTCTAACGATAACCCAG | GATTGGGCACACAAATGTAG | CTGGCCCGATGAGAGAGAAG | AATCTCTAACGATAACCCAG |
| | | GATTGGGCACACAAATGTAG | CTGGCCCGATGAGAGAGAAG | AATCTCTAACGATAACCCAG | ACCTGGTTTTTCGTAAGTGA | ACCTGGTTTTTCGTAAGTGA |
| **Chitobiosyldiphosphodolichol beta-mannosyltransferase** | ALG1 | TGCTATGCGAGAAGACCTGG | CAGGTCTTCTCGCATAGCAT | GGTACCATGGCCTGGGATGA | TGATGTGCCAGTTATCCGCC | GTCATTGACTGGCACAACTA |
| | | GCGCGTGGTACTGCATACGG | ACTCCGTACTGGAAAACTCG | TGCTATGCGAGAAGACCTGG | GGCCCGGCATGTAGTAGCGG | GCACCACCGCTACTACATGC |
| **Dol-P-Glc:Glc(2)Man(9)GlcNAc(2)-PP-Dol alpha-1,2-glucosyltransferase** | ALG10 | GCTGGAAGGTTACTATTTCT | CTCTCGCAACGCCCGGCTGA | GGCGTTGCGAGAGCCCTACA | AAACCTGCCATTTGGATCTT | TGTCATTGCACAAAAGTTAA |
| | | TAAAAAGGTACAGCTCAAGG | CAGGCGCAGCGCTACTGTGA | CCTCTTCTCCGCCTTCAGCC | GTGGAAGATCTCGTCCATGT | GTGGAATTGTTATTGGCGAT |
| **Putative Dol-P-Glc:Glc(2)Man(9)GlcNAc(2)-PP-Dol alpha-1,2-glucosyltransferase** | ALG10B | GTACCTGGTGTCAGTTGGAG | GGCGCTGCGAGAGCCCTACA | TAAAAAGGTACAGCTCAAGG | CAGGCGCAGCGCTACTGTGA | CCTCTTCTCCGCCTTCAGCC |
| | | GTGGAAGATCTCGTCCATGT | TGGAAGATCTCGTCCATGTA | AGCAGGAATGGCGCAGCTAG | GCAGGAATGGCGCAGCTAGA | GCTAGAGGGTTACTGTTTCT |
| | ALG11 | TATCCTTGCACACAATTCGG | ATCCTTGCACACAATTCGGG | CAATTGGTCTGCATACCATG | TTCTGTTGGCCAGTTTAGGC | TTATCCTTGCACACAATTCG |

| | | | | | | |
|---|---|---|---|---|---|---|
| **GDP-Man:Man(3)GlcNAc(2)-PP-Dol alpha-1,2-mannosyltransferase** | | AAGTTGTTGAGGTGAGCAGC | ACCGTTGACATTAACATCGC | GGAACCACAATGTCAAGCTT | TGGAACGAGCATTTTGGGAT | CTTCACTACAGAGAGCATGT |
| **Dol-P-Man:Man(7)GlcNAc(2)-PP-Dol alpha-1,6-mannosyltransferase** | ALG12 | CCTGAACACGATGATGGCGA | GCGAAAGCACGTAAACCGCG | CTGCAGGTGGACGTTGAAGG | AGCGAAAGCACGTAAACCGC | AGTCTTGGCTGTACAAAGCG |
| | | GAATGCCGCCTACTCAGCCA | AAAGTCTTGGCTGTACAAAG | GGTAGGATCCACACACAGCA | CGTTACAAAAGGAAGTGAGA | CAGCTACGTACCTACAGGCA |
| **Putative glycosyltransferase ALG1-like** | ALG1L | GCGGTCGGCCTTCATGGAGC | TGGTGATGCATCTCCGCGAG | AGCCCTGCTGGTCAGCAGCA | AGAACCTGAGGACGCAGCCA | GCCCTGCTGGTCAGCAGCAC |
| | | TGTGATAACAGGTACCGCCT | CGGCTCTTCATGAAGCTGGG | GGACGCAGCCATGGAGCGGT | CCAGCAGGGCTGGCCACTCG | GGGCCAGAGGCTAAAACCCC |
| **Putative glycosyltransferase ALG1L2** | ALG1L2 | CTGAGTCCTCAAAGACCAGG | CACAGAGCGGTCGGCCTTCA | GCCGGCATCTTTCTTTAAAG | ACGGAGCGGGATTCTGGGAG | GGACCCAGACACAGAGCGGT |
| | | TTTGAACAACTGACTCTTGA | ACAGGTCTGCATGACCACTG | AGGCCGACCGCTCTGTGTCT | TGGTGACGCGTCTCCACGAG | GACAGGTCTGCATGACCACT |
| **Alpha-1,3/1,6-mannosyltransferase ALG2** | ALG2 | TAGTGTGCGACCAGGTGAGG | TGAGCGGCTGGTGTTGGACG | CGCGGCTCTCGGCGAAACAG | CAGGAAAACCATGCGCACGT | CTGGACAGCGCACTACGACC |
| | | TGGATGACCTAGTCCCCAAG | TGGACAGCGCACTACGACCC | CGCCGACGAGGAGTTCGACG | CTCTCGGCGAAACAGTGGCC | CGACGTGGTAGTGTGCGACC |
| **Dol-P-Man:Man(5)GlcNAc(2)-PP-Dol alpha-1,3-mannosyltransferase** | ALG3 | GTATGCGGTGGTCAAGGCCA | ATTGATGACGCCTTCTACCT | AAGTATCTTGTCGCTGCTGA | GACGGCATGGCATATGTGCA | GGGTAGAAACGATCTGTATG |
| | | GTCCGGTGTCACCCTGCAGT | TAGAAACGATCTGTATGCGG | GATGTTCCCAGCGCAGGCGA | TTGTACTATGCCACCAGCCG | GTGCTCTTGGGGAAAGGCTG |
| **Dolichyl-phosphate beta-glucosyltransferase** | ALG5 | TCTCACCAGTACGAGGGCTG | CCCTGTTCTCACCAGTACGA | CATCTCTACACGTTGAACGA | AGCAGTTGTAAATGCAACGA | CAAAGTACGTGTGATAACCC |
| | | TCTACACGTTGAACGATGGT | ACCCATACATGAGAAGAGTA | TCCGTACTCTTCTCATGTAT | CCGAGTTTGCTCAAGCCTCC | TTACTCGAGAAGCAGCTTCA |
| **Dolichyl pyrophosphate Man9GlcNAc2 alpha-1,3-glucosyltransferase** | ALG6 | CTATGATAAGCTGTAAGAGG | AGACTCTTCCCGGTTGATCG | TCAAATAATCCACGATCAAC | GGATCCTCCTCAGAAACTAC | CAAATAATCCACGATCAACC |
| | | AACAAGTCCGGTAGTTTCTG | ACCGGACTTGTTTTCTGTAT | GCTGGTAAACCGCCTATGTT | TCATAATCACCAAACATAGG | CTTGTGACTGCGACCTCCTA |
| **Probable dolichyl pyrophosphate Glc1Man9GlcNAc2 alpha-1,3-glucosyltransferase** | ALG8 | GGCGGCGCTCACAATTGCCA | GCGGCGCTCACAATTGCCAC | GAGTACATCCATAAAGATGA | AGCGTGCGGTGCCGCAGCAA | CACATAGTAAGGGACTCGCA |
| | | CGTGCGAGTCCCTTACTATG | GAGTGATAGCAAGCCAGTTT | TCGTGTTATTTCCCTGGGAC | TTGCCACGGGTACTGGCAAT | CTCCATTGCACGATTATTTC |
| **Alpha-1,2-mannosyltransferase ALG9** | ALG9 | CGCTCTGTGGCACTGTTCAG | TGTGTTTCAACGATATCGCC | GATGAATATTTTGGCTCCCG | GTCCATATACCATCCAGTCA | CACCGGACCGAGTACGTGTG |
| | | TCATGTCAGTAGGAACAATC | TACTGTGACATCGAATTGGC | TTCAACTACTGGGAGCCAGT | AGAATTTAGGCCACCCGTAT | ATGTCACAGTATAGTGCTCC |
| **ADP-ribosylation factor 4** | ARF4 | CAAGGAACTGGTCTGTATGA | TCTGCTTCTTGCCAAATAGT | ACTTCCAGAATACCCAGGTA | CTGCTTCTTGCCAAATAGTC | TAAACTTTCTTACCAATGGT |
| | | TCAATTCATCTACCAGAAGC | CACAAGTGCCACTACCGCCA | ACATACAGTATCCCTTACCT | AAGCAGATGCGCATTTTGAT | ACAAGTGCCACTACCGCCAT |
| **Brefeldin A-inhibited guanine nucleotide-exchange protein 1** | ARFGEF1 | GCTGCGCAAAGCTTGCGAGG | CATATATCCCAAGAACACGA | CGCGTTGATACTCAAGACCA | CCTCCTCCCAGGCACTAACG | ACTGTTCCAGAAATGTATCG |
| | | ACTAAATGACGCTATACCTG | CTATGCGAGGACATTTGGAC | CAATATCTTCTCCAGAGCCC | GATACTCAGGTATAAATGGA | GAACTTGCATTCCAAACAAC |
| **GPI-linked NAD(P)(+)--arginine ADP-ribosyltransferase 1** | ART1 | CTTCCGCGATGAGCATGGGG | TCATCGCGGAAGCCCAGGGG | AGCGTACTGGTCATCAAAGG | ACTGCGAGTACATCAAAGGT | GGTCTCGTCGTGTGATGGGG |
| | | TGCTCATCGCGGAAGCCCAG | TCCTTTGATGACCAGTACGC | CCGCATCTACCTCCGAGCCC | TCTCCCGGATCTCAACCACA | GGGCTTCCGCGATGAGCATG |
| **Ecto-ADP-ribosyltransferase 3** | ART3 | CGTTCCCCAACTGCTAAAGG | ATACGTTCCCCAACTGCTAA | TAACCATGGAATAGCCCTGA | ACCAGCAATTAGATACTGTG | TTGGAGGGCTAAACCAAGCC |
| | | TGCCCGAGGATGCAGAAGGA | TTATCCATCTACACATGCCT | GTCTTACCAGGTATTTGGGT | CTACCTGGTAAGCAACTGAT | CCTGAAATGTACGGACAGGA |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Ecto-ADP-ribosyltransferase 4** | ART4 | CCCCTGTGTAGGCATTAAAG | ATCGACTTCGACTTCGCACC | CTCCTGCAACGATGAGAATC | TGGTTGCAGTTGAGGTCAAC | GAATTGGCCAAATCGAATGG |
| | | TATGGAGAAACTAACTCAAG | CAGATTCTCATCGTTGCAGG | TGAGGTGCATTATAGGACGA | AATGAACGTTCATACTGCTG | ACAGGGGCCACCATTCGATT |
| **Ecto-ADP-ribosyltransferase 5** | ART5 | ATTCTGGGCTTTGAAGCCAG | TGGAGGCAAACTGGCCCAAG | GGTGCCTCTTCGTCATATGA | TTGAATCAGGCCGTGCGGAC | GCGGCTTTGATGATCGCCCT |
| | | CCGTTCCCAGGGCTCCTAAG | GGCGCAGTTAAAATGGCTAC | ATGATCGCCCTCGGCAGCCT | GTTGAATCAGGCCGTGCGGA | ACCCACATAGGTATCGTCAA |
| **Asialoglycoprotein receptor 1** | ASGR1 | CAACTTCACAGCGAGCACGG | GGACGGGACGGACTACGAGA | AGTGATCTGGGGAGACCGGG | CTGACGCCGACAACTACTGC | GAGCCTGAGCTGTCAGATGG |
| | | TGTGCCCACTTCACCGACGA | AGCAGTTCGTGTCTGACCTG | TCCAGGGCCCGTTTTGGTCG | GACGGGACGGACTACGAGAC | TGGGCCTCCACGACCAAAAC |
| **Asialoglycoprotein receptor 2** | ASGR2 | CTCCTCGAGCACCCTGACGG | GCCTTTTCTCACACACCCAG | CTCAGCCAGATAATTGGCAC | GAGAAAAGGCGGAATGCCAC | ACTCAGCCAGATAATTGGCA |
| | | AGGCAAGTGTGCAGTTTCTG | GATTGCCTGGACCTCCGTCA | AGACTGTGTTGAAGTCCAGC | CACCTTGTCAGGTGGTAGTG | GACCTCCGTCAGGGTGCTCG |
| **UDP-GalNAc:beta-1,3-N-acetylgalactosaminyltransferase 1** | B3GALNT1 | TCTGGACTGTCCTTCCGAGT | ATGTGATAGAACGCGTGAAC | GATCTCAGTGACATCCTACT | TGAGCTGCTGTGGATGGCCT | CTTCATTACGTACTTGGCAT |
| | | AAGAAGGTGTTCATCCTCTA | TCAGTGACATCCTACTCGGA | TGAAAGTCTTGTCTGTAAAT | ACCCTTCAGATGTGAAAGCC | CGCGTTCTATCACATTGTAG |
| **UDP-GalNAc:beta-1,3-N-acetylgalactosaminyltransferase 2** | B3GALNT2 | GATTTCTTGACACAAGGCCG | CAAGACTGGTAATAACGATG | TTCTTGACACAAGGCCGTGG | ACGCTGACAACTCGATCCTC | CCAAGACTGGTAATAACGAT |
| | | TGGTAAGATGAATTGTTCCA | AGCATCCCACATTAAGTCAA | TCTGTCCGAAGACACTTCAT | GCGGCCAATGCGAAACTGGC | TAACGATGGGGTAGAGAACT |
| **Beta-1,3-galactosyltransferase 1** | B3GALT1 | TGGGCATATACCACTTACTG | ACGTCTCTCTGATTGCCTGA | TCTGATGCACAGTGATAACT | TTACTGCGGACATCCCGAAT | TAACTACCCACCTTTCTGTT |
| | | ACTGGCTATGTCATTAATGG | GCGATACCTACACAAACTGT | TTCGGGATGTCCGCAGTAAG | CAGTAAGTGGTATATGCCCA | AAGGTGGGTAGTTACTGTCT |
| **Beta-1,3-galactosyltransferase 2** | B3GALT2 | GAACACAAACTCATTGGGAG | GTACTCCAATAAGATGAGTG | AGACAAACCACAGCTCCCTT | AGGTATCCGCCGTTTGCACT | TGGGTGCATATCCTCGCATT |
| | | GAAGAGCTATTCGGCAAACT | TAGAAGCTAGAAGAGCTATT | CATACACTTTCCGAGGATTT | CATCTTCCAAGTGCAAACGG | GATAACGCTCACTTGGGTAG |
| **Beta-1,3-galactosyltransferase 4** | B3GALT4 | GTGGGGGTAAGTGCCCGACG | GCGGTAGGAGTCCTGGAAGG | CATGGTGCGGCGTACTCCGA | ACCGCGTATCAGAGGAGCAG | GACGATCACCAGCAGCAAAG |
| | | ATGGTGCGGCGTACTCCGAG | TCACCCTAAAGACCCTCAGC | CTTGAGGACGTATCGGGCCA | GGGCAGGCACCGCGTATCAG | CAACCCCGAAGGCCCGAAGA |
| **Beta-1,3-galactosyltransferase 6** | B3GALT6 | GTAGTCGCAGAGTTGCCAGG | AGAAGAAGCCCCAGTAGAGG | CTTCGAGTTCGTGCTCAAGG | ACGTGCTGCGGATCACGCTG | CTTCTCTGGCAGCACTGCGA |
| | | CTACTACCTGCCCTACGCGC | GGTATTCGGTGTCGAAGCGC | CGACGCCTACGAAAACCTCA | CGCGGACCAGTCGTACACGT | GCGCAGGTAGTGCACCAGGT |
| **Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 1** | B3GAT1 | GCTGCGCATCTACAAGGGGG | CTGGATTTGCCGTCAACCTG | GGGTGACAAGTTCTCGAAGG | CCCTTATGTACCGCGAGCAG | CTATTGCAAATGGCCGGTGG |
| | | GCACCGAGTACGTGTACACG | CTGGTGCTGCGCATCTACAA | GTCTGACCGCGACATCGTGG | TCTCCCGACTCCGCAGATGA | GCACCACCTCCACGATGTCG |
| **Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 2** | B3GAT2 | TGCTCGACGTGGACACGCGC | AAAGAAGCGGGTGAAAAGCG | CCAGACGGAGACCTTGCGGG | GTCAGCGAAGAAGAGCACGC | TGTGCCACACGAGAACCTGA |
| | | GCATAGATGGTGGGCAGCTG | CAACGAGCCAAAGTACCACC | TGTCATCATCATGCTCGACG | GCGCTACGAACGTCCGCTGG | GGCCAACACGTTCCGCCAGG |
| **Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 3** | B3GAT3 | GGGTCCAGCGCATCTAACGG | TTTGATTCCACCGCTCCCCG | GAGGGCCCTCAGGTACAGGA | AATACCTGAGGATAACGCAG | CAATTGCTGGGTCTGAGCCC |
| | | CGTTAGATGCGCTGGACCCG | TGTTCCGCTGCTCGACACCA | CGATGACAACACCTACAGCC | CACGGGTCCAGCGCATCTAA | CAGGGTTCGATCCATTCCCT |
| **Beta-1,3-glucosyltransferase** | B3GLCT | CTCCCTGGGTAAGTAGCGGG | CCTGCTCCCTGGGTAAGTAG | AGGTACGTAGCGATGGCTGG | CTCTACATCTGGGACAAAGG | ACTGGCGAGAAGTCTCCTGA |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | TGAGTTAAACAAACGTACTG | TCGGCCGGTGGATTACCCTA | CTGGAACATCGATCCAGTGA | TAGTCCTTAGGGTAATCCAC | CTTCTGTTACTCACACAAAG |
| **N-acetyllactosaminide beta-1,3-N-acetylglucosaminyltransferase 2** | B3GNT2 | CATGCTGACCAACCAGACGG | CAACGCAGGGAACCAAACGG | CTCTCGGTTCCAGTATGCCT | GAACGAGGCCGAGTTTCTGA | AAGCAACGCAGGGAACCAAA |
| | | GGAGGGGGGTTCCTCTACTC | GGAATGTGCCTTCAGAAACT | ACCTGACCTGAGGGTCACGT | TGCCAGAAGGCAAGCAATCC | TTGCCAGAAGGCAAGCAATC |
| **N-acetyllactosaminide beta-1,3-N-acetylglucosaminyltransferase 3** | B3GNT3 | TGAAGTATCTCCGGCACCGG | GTTGCCCCACGAAGAGGTGG | GGATCAGTTGCCCCACGAAG | GAGGAGGGTGAAAGCGCCGA | AGGAAGTTCTGAACGTGCTG |
| | | GCGCACATAGTTGCTAGGGG | CGGTACCCACCCTATTGTGG | GGATGAAGTATCTCCGGCAC | TGCTCCAAAAAGCCCGGATG | CGCCAGCTTCGTGCTCAACG |
| **N-acetyllactosaminide beta-1,3-N-acetylglucosaminyltransferase 4** | B3GNT4 | TCAAGACATTTGGAATCCGG | TTGTGGGTATGTGCCTGAGG | CATCTTCCATGATAGCCTGG | AGGTCAAGAAGAGACGGTGA | GTGCGGCTATCCGCAGCACG |
| | | TCAAGACATTTGGAATCCGG | TTGTGGGTATGTGCCTGAGG | CATCTTCCATGATAGCCTGG | AGGTCAAGAAGAGACGGTGA | GTGCGGCTATCCGCAGCACG |
| **Lactosylceramide 1,3-N-acetyl-beta-D-glucosaminyltransferase** | B3GNT5 | TTGGTAGCGAGGCCCCGCTG | AAAATAGGGATAGTACCGC | TCCGGAATTAGAAGGACGTG | GTCTCTTAAGCACACCTCAG | TCTCTTAAGCACACCTCAGC |
| | | ATCGACGTTCCGGAATTAGA | GGAGCTGCCTATGTAATCTC | GGAGATTACATAGGCAGCTC | TAGATAAACGCAGCCCTACA | GAAAACTATGATCGACGTTC |
| **Acetylgalactosaminyl-O-glycosyl-glycoprotein beta-1,3-N-acetylglucosaminyltransferase** | B3GNT6 | GGGTGCCCAATAGAAAGAGG | GCAACTCGCGGTACATGCAG | GCTTACCCGGTGTACTGCAG | GGTGCACTAGCAGCAACTCG | AGGAAGTCCTGGATGCGCGC |
| | | CGTCCCAAAGCAGCGGGAAG | GCATCTCGTAGGGCGCGAAG | GTTCTCCGGCCAGCTCATGG | CACGCGCGCTTTCTGCTCAG | CGAAGGGTCGGATGCCCTCG |
| **UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 7** | B3GNT7 | TATCCGCCGTATGCAGGCGG | TGTGCACCACGAGCATGGCG | CAACAAATACTACATCCCGG | TCTACGGCGACATCCTGCAG | AACCACCCGGAGAAGTGCAG |
| | | GCAGGATGTCGCCGTAGAGG | AGCTATCCGCCGTATGCAGG | AGCATGGCGCGGAAAAAGCA | AAGAGAAACTGCCGGAACTG | ACTGCAGGATGTCGCCGTAG |
| **UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 8** | B3GNT8 | TCGCAAGACAAAACTCACGG | ACTCTCATTGGGGAGGGCGG | ACTTGGGGCAGCGCATGACC | GATAGTCTGGCCCAGGCGGG | CGCAAGACAAAACTCACGGT |
| | | CAGTGTAGACGTCCTCAAAG | CAGTGGGGACAGCACTGAAA | CTGACTTGACGGCCAACAGC | GTAGCCACCCCCGCTTGCAT | TCGAGTGGACATCCGAGTCC |
| **UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 9** | B3GNT9 | GAAGGTGCTCAAATGCGGCG | TCAAAGGGTCCCGTAGGCGT | TAAAAACGAAGCGCACGTCG | GCTGGGGGATGCCAAAGGTG | AACCAGCCGCACAAGTGCCG |
| | | GCACGCCCAGCAAGAACACG | ATTTGCGCGCCAAGGACCAG | CGGTGGCAGAGGACTTCGAG | GTCGGGTAACTGGAACGCGC | CCAGCAAGCAGGTCTTGCGC |
| **Beta-1,4 N-acetylgalactosaminyltransferase 1** | B4GALNT1 | ACGGCGCAAGAGGTAGCCGG | TAACCGTTGGGTAGAAGCGG | CCCGTAGCCGATCATAACGG | TTTGCCGGAGGCAGTTCCCG | GCGATGACCACGGTAACCGT |
| | | ACCGGGATGTGTGCGTAGCG | GAAGAAGTTAACCACGCCGT | ACGACTTCGTCTTCACGGCG | CTAGCTGCACACCTTGCCGA | CCGGGCGTAAGTCTCTGCTC |
| **Beta-1,4 N-acetylgalactosaminyltransferase 2** | B4GALNT2 | CATCGGGTCCTTCAAACTGG | GCTTGGGTATGACAGGATGG | ACGTGACATACACCAGCACG | TGATCCCCGCCTGCAACGAG | AGTGAGCCACTCGTTGCAGG |
| | | ATGGTTCTTGAAGTAGTGGA | GATTCGGGATGACTTCGGGC | TCCGTGGACTGGGTACCCAA | TGTGCCGCATAAAGGTGTGA | CACGTGACATACACCAGCAC |
| **Beta-1,4-galactosyltransferase 1** | B4GALT1 | CCCGAGTCCTTACCAAGCAG | TGCTTGGTAAGGACTCGGGT | GTGGAGACTCCGACCAGTTG | TATATCTCGCCCAAATGCTG | GCCAAGGTGCAGAGCGCAGA |
| | | CCGAGTCCTTACCAAGCAGC | CATTCCATTCCGCAACCGGC | AACCCAAATGTGAAGATGGG | ACCCAAATCACAGTGGACAT | ACCGAGGTCAAGTTGCTAGC |
| **Beta-1,4-galactosyltransferase 2** | B4GALT2 | ACCGGTGAGTAAGCACGCGG | GAAGATGAAGCAGTCATAGG | CATTAGTGTCAGCCCCGAGG | ATCTACACCCCATCTTGAGG | GGGCGTAGACGTCAAAGTAG |
| | | TCAACCGGTGAGTAAGCACG | ATATCACAGTGGACATTGGG | ACCCCATCTTGAGGCGGCAG | CGGGCCAAGCTGCTTAACGT | AAGCTGACCATGAAGCGGGA |
| **Beta-1,4-galactosyltransferase 3** | B4GALT3 | AATCCCCGGGTAGAACCAGG | TCATAGCAACGGCAACATGG | ATAGTGTCCTACAGATGTGG | GGAGGAGTCAGTGTGAACCT | TTGCAGCATCTCTTGACGGA |
| | | GATTGTGGAGCGGAATCCCC | ATGATGTACCTGTCACTGGG | GTCAAGAGATGCTGCAACGC | CATTGTGCCTCATCGTGCCC | TGTTGCGGAGGCTGCTGGAG |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Beta-1,4-galactosyltransferase 4** | B4GALT4 | GTTCCGGTGGGGAACGAGGA | GTCATCCACCAGGTGAGCGT | GAACGCAGAACGGTAAGTCC | CGGGTACCAGGTCCACATCG | ATCCCAAAGTGTCCAGAGGC |
| | | GCAATGAGGTGAACGCAGAA | TCGAGCCAAACTCTTGAATG | TCGCCATCCTCGTTCCCCAC | CGTCATCCACCAGGTGAGCG | CGAGCCAAACTCTTGAATGT |
| **Beta-1,4-galactosyltransferase 5** | B4GALT5 | TGCTGAGGAAGTCAAAAGAA | CAAACTGCAAGCGCTGGCGC | ATGTGGCGCCCGGCATAGGT | TTCGGAGTGCTTATGCCAAG | TTGGGCCCTCTGAACAGAGA |
| | | CTTCTGATTGCATGCCTCGG | GTGCTCGTGGCGGTTCCGGA | TATTCTGTGAGCCGGCCAGA | TCATGATGTAGATCACATAC | TACTTCGTCTATGTGGCGCC |
| **Beta-1,4-galactosyltransferase 6** | B4GALT6 | GTGAAGTCCAGTTTTTAGGA | CAGGTAGATGATCCACATCG | AAGGAGCGTCAGTACATCGA | TGTTCATGGCGATTACGGAA | ATCTATGTGGCCCCAGGCAT |
| | | TCAAACAACAACGTATCTCC | GAAGATGTTCATGGCGATTA | GATGACGACTTACGCATATA | GAAAATGACCGGAACTATTA | CTCCACATCCGTAATAGTTC |
| **Beta-1,4-galactosyltransferase 7** | B4GALT7 | AACATCGTGCGGAGAGGCGG | AGACGGAACACTTCCGAGGG | CACTACAAGACCTATGTCGG | GCCCAGCTCCCTTAATGCGC | CATCTACGTGCTCAACCAGG |
| | | CGGGCAGCGCTCATCAACGT | GGGAACATCGTGCGGAGAGG | TCCGCGAGGGGAACATCGTG | GTTCCCCTCGCGGAGGAAAG | CTCCTCGAAGCGTTCGCGGA |
| **Beta-1,4-glucuronyltransferase 1** | B4GAT1 | TTTTGTTCATGGGCATGCGG | ACGGCTGCCTCGTAACGCGA | GCCCGGGATTAATTATGCGC | TGCCGGAAGAGAGCTTGCTG | GGCGATTACCGCGTCTACAG |
| | | GAGCCATTCTACGTGGCAGG | CTGGTAGAAGGCGCACCGGA | CATCGCCACCAGCATGAGCG | CTTCCGGCAGGTTGACCCAG | GTCCTGCCAAGGTACCACGT |
| **Glycoprotein-N-acetylgalactosamine 3-beta-galactosyltransferase 1** | C1GALT1 | ATTCTTTGCTGGGTTATGAC | TCTAAACTTACCTCTACAGG | ACAACACTTTGTTACAACGC | CAACACTTTGTTACAACGCT | TTTCTCTAGGTTTTGAGGGC |
| | | AGTATACGTTCAGGTAAGGT | CCACTACTTTTTTAGGGTCC | ATGTAGCCCTGCTTTACATA | ATTTCTAAACTTACCTCTAC | CAAACACGTCAAAGCTACTT |
| **C1GALT1-specific chaperone 1** | C1GALT1C1 | GTATGGGGTATACCGCCTTA | TGTATGGGGTATACCGCCTT | CAGCCTTCTACTACCTGGTT | TATGCCCAAATGCCCTAAGG | TCTGGAGACCTTGAATATGT |
| | | GTATGGGGTATACCGCCTTA | TGTATGGGGTATACCGCCTT | CAGCCTTCTACTACCTGGTT | TATGCCCAAATGCCCTAAGG | TCTGGAGACCTTGAATATGT |
| **Uncharacterized protein C20orf173** | C20orf173 | CGACATCGACCAATACCCCG | TCCATTCGTCGGCTGTGTGG | ACTTCCCGGAACTGAACCAA | AAATGGCTCACCGAGAGCCT | CACTTCCCGGAACTGAACCA |
| | | TGCTATCTTCCACTCCACTA | ATCAAGCCAGTTCCATTCGT | ATACCCCGTGGTTTTCAGGT | GCCACCACACAGCCGACGAA | GACTGCCCTTGGTTCAGTTC |
| **Coiled-coil domain-containing protein 126** | CCDC126 | ATTCTGATTTATTCCCCTAG | TGTCGAGAACGGTGCTTCTA | AACACAGTGGATGTCGAGAA | GGATGACATTTTGCAACGAT | GAGAACGGTGCTTCTATGGC |
| | | GGTGCTTCTATGGCAGGATA | ACACCACCAATGGTACTAGT | CAATGGTACTAGTGGGAATT | AATAAAAGAACGAATGTCTC | TCAGCAGCCAACACCACCAA |
| **Properdin** | CFP | GTAGGATTAGGTCCACAGGG | CGGCGGCAGCAACAATCGAG | AGTGGAGCCCCTGTATCCGA | GGTAGGATTAGGTCCACAGG | ACCTTCGACCATGGAAACGG |
| | | CATTCCAGCCCACACAGCGC | AGCGGCGGCAGCAACAATCG | TACTGGGTGAAGCAGAGCAC | TCATGTTCCGTCGGATACAG | ATCGATGTCCGTCAAACTTG |
| **Calcineurin B homologous protein 1** | CHP1 | TTTCATTATCCTCAATGGGG | TCGGGCCTCCACGTTACTGC | CCGGCTGTGAGTTCGGGTTG | GAGGAGATCAAGAAGGAGAC | CTCGGGCCTCCACGTTACTG |
| | | CAGGTGCTACGCATGATGGT | GTCTCCTTTAGTGAAGAAAT | AAGGAGACCGGCTGTGAGTT | CCATCAACCCACTGGGGGAC | TTCGTCCCGCAGTAACGTGG |
| **Carbohydrate sulfotransferase 4** | CHST4 | TTGACAGTCGCATTGTGATG | CTTGTTCAGATCTGACGTGG | CACAAGATGAAATCATCCCC | AGCCATGTGCAGCATCCAGG | GGTCCTCATAGCGCACAAGC |
| | | AGTCGAGAAACCTTTTCATA | GCATAACATCACCCGAGGCA | CGGGGGATGATTTCATCTTG | GATTGACAGTCGCATTGTGA | GCGTTCTCGGGAACGGAACA |
| **CCR4-NOT transcription complex subunit 6** | CNOT6 | TAATTGTATACATCCTCCGA | AGGAACCAGATGGAACAAGA | TACAAGAACCAGATAGGACA | TACATCCTCCGAGGGTCAGG | CTGGTAGAGCTGAAAGAACG |
| | | AGGTGGATGCCGTTGACTTG | ATGCCTAAGGTGTTCAGCTG | AATACGAGCCCCCTGACCCT | CCTTTCCTGCCCCAAGTCAA | TCAGTCCTAAGTCTAGAGCT |
| | COG3 | CTCGGTATTGGAGCTGAAGG | TGTTAGAAGATGTACAGGAG | GGATCGGAGACCGGACACGA | CGATAAGTTAGTCATGATGG | ACCAGTGTTACCTTGATCAG |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Conserved oligomeric Golgi complex subunit 3** | | TGTTTACTGCAGGTTCGTAG | CACAGGCCGATTTCAGATGA | CATGACAAAGGTATAGACCT | TCATGGAATGTGGTATCCTA | ACTCTTCCTTTAACAGAGCG |
| **Conserved oligomeric Golgi complex subunit 7** | COG7 | GAGAGGGGTGAGACTAAAGG | GTGGGTCCCCACAAGACAGG | GTAATTCACACTCCGCATGG | ATCAATGCGGCCTTCAGGGC | AGCCAGTCCACAGATTGTAG |
| | | CGGAATCGAAAGCCAGCTGG | GAACTTGGAGAAGTCCATGG | GCTGACTCGGCTTAACCAGC | CAAGCCCTTGGCGAAGTGGG | TGTAGTAGTAGGCCAGGAGC |
| **Cold shock domain-containing protein E1** | CSDE1 | TCTTTACCCTAAATCAGAAG | TGCACCAAACCCCTGTGGGG | GTCGAGTATAGCTTGTCCAA | CAATGGGTTTCCCAGTCCGT | TTGAAGTATCATCGGACCGA |
| | | GTGAATTCCACATCATCGCC | GCAAAGTAATTCGCCCCCTG | ACACCCCTGAAGATGTCGAA | GACTATTGCTTTTCAAGCCA | TGTATGCTACGAACGTAATG |
| **Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit DAD1** | DAD1 | CGGTAGTGTCTGTCATTTCG | GCTGTATATACTGCTGACCG | GTTCGGTTACTGTCTCCTCG | TGAAGGGGAAGGTCCCCACG | CGGGCTTCATCTCTTGTGTG |
| | | TGTGGGGAGTTTCATCCTAG | GATCAACCCACAGAACAAAG | TGGAAATCCGCTTTGTTCTG | CGCCGACATAACTGCACGCA | GCGTGCAGTTATGTCGGCGT |
| **Dystroglycan** | DAG1 | GACGCCCCATGCTGGGACAG | TGTCCCACTGTCCCAGCATG | GGCTTCCTGGGATCAGATGG | CGTGCTTTCGGACATCCGGG | GCTTTCGGACATCCGGGAGG |
| | | GGAGGCGGAAGCAGTGGTTG | GCGGCAGTGTCACAGCCTAG | CGGATTCCCAGACTCCTCTG | GAGCATAGTAAAGGTGAGAG | GCAAGGATGTGCTCCCAGCG |
| **Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit** | DDOST | CGATTGGGCTACCGTAGATG | GTGCTGGGTCCTTCGGCAGG | GAAGATGGAGCCCAGCACCG | TGTCCAGCAGCACTAAGGTG | TGCTGGACAACCTCAACGTG |
| | | CGGAGCCTGAAGGGTGAGAG | GGGTGAGAGCGGGGTCCGAG | GGTGAGAGCGGGGTCCGAGG | GTTTTTTCTTGTCCAGACCG | TGCCTGCTTTCAGATTTTGG |
| **Derlin-3** | DERL3 | TGAACCTTCTCAAGCACGCG | CGAGTTCCTGCAGGTGCCGG | TGCAGGTGCCGGCGGTGACG | AGTAGAGTTGAAAGGGGCTG | TGTTCCGGAAGTTCCAGGTG |
| | | GACGAGCCTCCAGACCTACG | AGGAGCCCTCTTCCAGCATG | TGAAGACGAAGTCGGCCGTG | CGTACACCAGCATGGCCATG | GTCTTCATGTTTCTCTTCGG |
| **Dehydrodolichyl diphosphate synthase complex subunit DHDDS** | DHDDS | GAGAGCTGTCACTTTGGGAG | GCTTGTTGAAGCCCTGTGAG | GGGCTTCAACAAGCTAGCTG | GAGGAGAGCGGTTGGTATAG | CGTATCAAGATGTCAGGATG |
| | | TGGTGTTCCAACCCGTTCTG | GGCCCGAGACATGTATGCAG | GAGACATGTATGCAGAGGAG | GACAGAGCAGCTGCTGCGAG | AGAGGGGCTCCAAGCCAGTG |
| **Dolichol kinase** | DOLK | GGGTCATATCTCTAGACCTG | AGTGCCCATCTCCGGCCCCG | GCCAGCACCGATCCACTCAG | GAGTGGATCGGTGCTGGCAG | TGGATCGGTGCTGGCAGAGG |
| | | GGCGGCAGTAGTGTTTGCAG | CGTAGAAGGCCTGCACTGCG | GGGCAATAGGCCACTGTTTG | GGCAGCCACTGGCATGGCAG | GAGGGCCACTGCCATGCCAG |
| **Dolichyldiphosphatase 1** | DOLPP1 | GTCTCCGGGTAAGATGGCAG | GCTCGCTCCCCGCTTCATGG | CGATGACAAATACAGGGCTG | TGACCCTCATCATATTTAAG | TGGGGGCCTGGCACTGAACG |
| | | GGGGCCTGGCACTGAACGAG | TGGTGCCCACTGCTGTGTGG | AGACTAGGAAGGCCACAGCG | GTCTCCTACAGCAGGTATGG | GTTCCTGCAAAACAGGATGG |
| **UDP-N-acetylglucosamine--dolichyl-phosphate N-acetylglucosaminephosphotransferase** | DPAGT1 | GCCCCTGACCGGTCACCATG | CGTCTCGCTGCTGGGATTTG | GGCCGGGATGAGGGTGACTG | AGAGGCGCGCAGCAATGAAG | TGTTGAGGTCCTGACCACAG |
| | | GCTGTCGGCTGGTTTTGTTG | GAATCCCAGGGAGTGATCAG | CGAACCCACTTACTTCATGG | TGAAGTAAGTGGGTTCGTGG | GACCATGAGGAGAGGTAGTG |
| **Dolichol-phosphate mannosyltransferase subunit 1** | DPM1 | GTAGTCCTCGCAGGTCTCGG | AGCTCCCGCCGAGACCTGCG | AGAACCTGCCGCTCATCGTG | TGGTGAAAAGCTTCTCCGAG | AGTCCCAACTTTTTCTCTCG |
| | | TGTTTTATTTGGCAGCCGTG | GACTTCCAAGGAGGCCATGG | GGCGGAACTGAGCCAGATGC | AGGACTACGACTGACTTCCA | GTCGTAGTCCTCGCAGGTCT |
| **Dolichol phosphate-mannose biosynthesis regulatory protein** | DPM2 | GGCCACGGGGACAGACCAGG | TGATCAGGCTAACGGCGACG | AGCAGGAGGCCTGCAGCCAG | GTGAAGGTCCCGCAGGGATG | TGGCTGAGCGCGCGGGGAAA |
| | | GCGCGGGGAAATGGTGAGAT | GTGAGATTGGCACCGTGTGC | CGAGGCCGAGTCCCACCACC | GGTGAAGATGATCAGGCTAA | TGTAGTAGGTGAAGATGATC |
| **Dolichol-phosphate mannosyltransferase subunit 3** | DPM3 | GCTGGTTTTAATGCTCTCCG | GTTTTAATGCTCTCCGTGGG | CGTGGTCAGGGCCACCCAGG | TGTCCTGCCAGGAAGTCCTG | GCCCAGGGCATAGCAGCCGG |
| | | GTCCTCGCAGTCATGAAAAG | CGAGCCGACTTAGCCCGCAG | GAGCCGACTTAGCCCGCAGG | GGCGGACACCAGCAAGTAGG | GCTGCAGAGCCAGATACAGG |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Probable C-mannosyltransferase DPY19L1** | DPY19L1 | AGGCGGCCGCCCTTCCATTG | GCGCGGGCGCCGAAGAGGTG | TGTGGGAGTGATGGGGCCGA | CGGCCCCATCACTCCCACAA | GGCGGCCGCCCTTCCATTGT |
| | | CGCGGGCGCCGAAGAGGTGT | GCCGAAGAGGTGTGGGTAGC | GAGGTGTGTTATATGGCTCC | TGCGGGTGCCATGCCCACGA | GTGTTAAGCTCTCTGCACTT |
| **Probable C-mannosyltransferase DPY19L2** | DPY19L2 | GCCGCCCCTTAGACTGGCTG | GCCGCAGCCAGTCTAAGGGG | AGCCAGTCTAAGGGGCGGCG | GCCGGAGGTAGAGGAGGAGA | GGAAACTGCCAAGGGGCTCC |
| | | GAGGACCTCCAGGAGCCCCT | GCTCCTGGAGGTCCTCCCCG | GCAGCTCCGGGAAAAGGTGC | GCCAGTCTAAGGGGCGGCGC | GGGCTCCTGGAGGTCCTCCC |
| **Putative C-mannosyltransferase DPY19L2P1** | DPY19L2P1 | GTTGACTAGCCGTTGACGAG | | | | |
| | | GCTCGTCAACGGCTAGTCAA | | | | |
| **Putative C-mannosyltransferase DPY19L2P2** | DPY19L2P2 | GTAGGGTGGGGCGCATGCGT | CGACTGTGAAATGTAGGGTG | GCGACTGTGAAATGTAGGGT | GGCGACTGTGAAATGTAGGG | GAAGTCTGCAGGCGCTGGGG |
| | | AGGCGCTGGGGAGGGTGTGG | GGCGCTGGGGAGGGTGTGGG | AGGGTGTGGGGGGCTGTGCG | GGGTGTGGGGGGCTGTGCGT | GCAGGCGCTGGGGAGGGTGT |
| **Probable C-mannosyltransferase DPY19L3** | DPY19L3 | GTTGTAATTACCTTGCACGA | TGATCTCTTCACAGAAGCGA | AGTTGTAATTACCTTGCACG | TAATTACCTTGCACGAGGGT | AGCGCCCAGTTCTCCCTCAG |
| | | GCCGGAGTCAAGCTGTGCAC | AAGACAGCATCTGCTACGAG | GACCTGCTGGACATTGCCAA | GCCTTTTTCCATTTTACCCA | ATAGTGCGGGTGGTTGGTTA |
| **Probable C-mannosyltransferase DPY19L4** | DPY19L4 | CTTCGCAGAAACGATGGCGG | CTCCGCCATCGTTTCTGCGA | GGAAGAAGGTGATTGCCGCG | TTTCATTGGCTGTCTTGCAG | CATCGTTTCTGCGAAGGCTA |
| | | GAGACCTATTATAGTAGGCA | GAAAAGTGTCATCCAAAGTT | ACTTATCAGCATACCATGAA | TGGCCACGTAAGTAACCATT | GTGTATGTTCTCCCGAACTT |
| **D(2) dopamine receptor** | DRD2 | GATGGAGGAGTAGACCACGA | GGGCAACTGTACTCACCCCG | TGACCAGTGTGGCGACGAGG | CAACGGGTCAGACGGGAAGG | CAGCTGTGTACCTGCAAGGG |
| | | CTGCGTTATTGAGTCCGAAG | AAGGCGCTGTACAGGACAGG | CACGTAGAAGGAGACGATGG | ATTCAGTGGATCCATCAGGG | CGTCTGACCCGTTGAAGGGC |
| **D(3) dopamine receptor** | DRD3 | CAGGCCATTGCCGAAGACGA | TGGTGCTGAAACAAAGGAGA | GCATGGCTGTGCTGAAGGAG | TCCCGAAGTGGCACTCCCCG | GGCATAGTAGGCATGTGGGC |
| | | CGTTCACTACCAGCATGGCA | CCAGACGGCCGTGATCATGA | CTGGCCATCGTCTTCGGCAA | CAAAAACATCACAGCAAATG | GTGCCATGCTGGTAGTGAAC |
| **D(4) dopamine receptor** | DRD4 | ACCTCGGAGTAGACGAAGAG | CTTCGTGGTGCACATCACGC | GTTGAAGACAGTGTAGATGA | AGGAACCCACCGACCACCAC | CTGCGCTACAACCGGCAGGG |
| | | ACGAGTAGACCACGTAGTCG | GTTCCGCAACGTCTTCCGCA | TCGTTGAGGCCGCACAGTAC | AGGCGGCGTGCCAAGATCAC | TGCGCTACAACCGGCAGGGT |
| **ER degradation-enhancing alpha-mannosidase-like protein 3** | EDEM3 | AGCGCGATGGAGACTAGTGG | ACAATTTGTACAGCTCGTGG | ATGGAGACTAGTGGCGGCGA | CTACCCTGCCAAAAAATGGG | GCGAGCGCGATGGAGACTAG |
| | | CAAGTCCGAGCATTCAGCAT | TCGACGCTGAAGATGGGTTG | CCACTGGTAGTGTTGAAAGC | ACACGGAGGTGGCCGACACC | GCATCAGAAAACCAGAAGCT |
| **Ectonucleoside triphosphate diphosphohydrolase 5** | ENTPD5 | GAAAGTGAACAACATAGAGA | TTTGGGTGGAGGCTCCCCCT | CTATACACATAGGTGAGGAC | TCTTTTCCCAAGATGTGGCT | TCTATACACATAGGTGAGGA |
| | | TCGTACCACCCTCAGCACTT | AACACCTCTCCCACCTTCGT | GGACTGATGGGCACACTTTC | ACTCTGGAACAAACTCCTAG | CTATGCCGAAGTGCTGAGGG |
| **EGF domain-specific O-linked N-acetylglucosamine transferase** | EOGT | ACAGCCCATACCTCATGCGG | TTCACTACATCACTTGGCGA | CTAAGGTAATGGATACACTG | TGCCAGCATCCGCTTGCCAG | TGTCAGCCTAAGGAAACGGT |
| | | GGACACTCTTGAGTCAGCTG | ACGTATTGCAACACCCAAAG | GCTCTGTCAGCCTAAGGAAA | GAGGCGTTCACTACATCACT | CTGACGTGTACATCGTGATG |
| **Exostosin-1** | EXT1 | TCTGGGATAACTCTAAGGAG | GACGACGACAGGCACAGCAG | AGCCGGAGAGAAGAACACAG | CGTATACCCACAGCAAAAAG | TGTTGTCGTAGGGCAGAAAA |
| | | CTTATATCACGTCCATAACG | CGTGGGGTTTGACATCGGCC | GAGGACGTGGGGTTTGACAT | CGGCATGTAGCCAAACCAGC | AGTGGTCAGGGTCAGCCCAA |
| **Exostosin-2** | EXT2 | TTTGAGGACTCGATAGAATG | GACAGGAGGAAGTATTGCCG | TCGGCTGGCAGCCTAACAAC | CGCTGCCCCACTTCTACAAG | CTGTGTCATTATGTGTGCGT |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | CCGACAGTCCCATCCCAGAG | GTCGTAGGTGAGGACTATGG | CGGGACCATGCCTCTCAAGG | GTTCTGGTTAAGCACATCGA | AGGGAACAAACAGACAGGCC |
| **Exostosin-like 1** | EXTL1 | TATCTGGTTGATGCAGTCGG | TAAGGTATTCGTGTACCCAG | CATCTCTGGCCACCGTCCCG | CCCTCTGAAGCTCATCCAGG | GTCGAGCCATTTCTGGGACG |
| | | GAAGCCCACCCGTTGCGAGG | TGTGGCAGAGCTTCCCAGAG | GGTCATCCATACCACTCTGG | CGCTGAATCTGCCCTCAGGG | CACTTGTGGAAAGACTGCTG |
| **Fukutin-related protein** | FKRP | ACTTGAGCTCCAGGAAGCGG | GCTGGACTTGACCTTCGCCG | TACTAGGGCCACAAACTCGG | CGAGACCGCCCGCTATGTGG | GCGCGGCTTCGTATGGGAGA |
| | | CGCAGGAAGCGCACGTCTGG | ATGGGATGATGTCCCCGTGG | CTGCGCCACGAAGCCGGCAA | CGGCTTCGTATGGGAGAAGG | CACGGTTCCGAAGCAGCGCG |
| **Fukutin** | FKTN | GTGTGCTATCAAATCCAATT | CACTCAATAAACCTAGAGTG | AGTGGCAACTACCTCTGGCA | AGATTGAGAGTAAAGATCCC | CCTCCCAATGTGCAACCCAA |
| | | AGTTGGCTTCAATGTATTCG | GGCCTTTCGGAAGAGTGCAA | AACGACCAAACTGTAACTTT | TGAGTCTATCCCGTCTAGCC | GTAAAGATCCCCGGCTAGAC |
| **Galactoside 2-alpha-L-fucosyltransferase 1** | FUT1 | CAGGGTGATGCGGAATACCG | CAGGCAGGATAAAGGCCCGG | GGGGTAGACAGTCCAGGTGC | GTTTTCGTGGTCACCAGCAA | CACTCTGCGCCTCTTCCCGA |
| | | GAAAACATCGACACCTCCCA | TCCACTTCTGGGGCCAGCAC | TAACCTGCAGATAGTCCCCA | GGACTGTCTACCCCAATGGC | GCTTCACGACTGGATGTCGG |
| **Alpha-(1,3)-fucosyltransferase 10** | FUT10 | TCTGACTGTACATACACCAG | CGAGCTGATGACTTACATCG | GGGACCACCAGAGCATAATG | CAGTCATGATGGGCTTTCCG | ACTGCAAAGGAACTAGGTAT |
| | | TCCAAGTATTGGGTAGTTAG | CGTGGCAGTGTAGTTGAACA | TGGGGGATCCGTAATATACA | GGGACTACCCCCAGTTTCAG | CGGGGACCACCAGAGCATAA |
| **Alpha-(1,3)-fucosyltransferase 11** | FUT11 | TGGAGGTAAGATTGAAGAGG | GCGGCTACAGGACACAGCCA | ACTGGGCAGTCCATGTGTGA | CGCCACATCCCGGTGAGTGA | GGATGCCGAGAAAGCCCACG |
| | | CGGAAAACCGCGCCATCCCA | TTCTTGCTGAGCCACGGCCC | CCGCGGTTCTCCCTCTGTGA | TTTCGTCTGTGACTACGAAC | CTTTGGCCCCACTCGGAAAA |
| **Galactoside 2-alpha-L-fucosyltransferase 2** | FUT2 | GTAGGGGTCCATGTTCGCCG | GGAATACCGCCACATCCCGG | TAGGGGTCCATGTTCGCCGA | CGACCGGCGATACCTACAGC | AGGGGTCCATGTTCGCCGAG |
| | | GCGATACCTACAGCAGGCCC | GTGAAGCGGACGTACTCCCC | GCTAGCACTGGTATCTGCAC | CTGCTGAACGTGAAATATAG | CTGCTGTAGGTATCGCCGGT |
| **Galactoside 3(4)-L-fucosyltransferase** | FUT3 | CCTTGCGGTCGGCAGTGATG | GGGTGGGAGTGGTGTCCTGT | ATAGCAGTGCGGCCAGACAG | ATGTCCGTAGCAGGATCAGG | GGATCCCCTAGGGCTCCCAG |
| | | CGATGCCACTGGATCCCCTA | CAGCGGCGCCATGGCCATTG | GTCCTGTCGGGAGGACCCAC | TCCTGATCCTGCTACGGACA | GCCATGTCCGTAGCAGGATC |
| **Alpha-(1,3)-fucosyltransferase 4** | FUT4 | GCGCGTGTTGGACTACGAGG | GAAGTAGCGGCGATAGACCG | TGATCACCTACGCTTGCTGG | TTTCCTCGACCGCAACCCCG | CGCCACTGTCCAGGAAACAG |
| | | TACGAGCGCTTTGTGCCCCG | TCCGGCGCCAGTGGAAGTAG | AAGTAGCGGCGATAGACCGC | AGTAGCGGCGATAGACCGCG | TCCCAGAATGCAAAGGCCGC |
| **Alpha-(1,3)-fucosyltransferase 5** | FUT5 | GTGGCCTTTTAACACACCCG | ACGTCCACAGCAGGATCAGT | TCGCCATGCTGTCCTGGCAG | CTGTGGGTACACACTGGAGT | GTCCAACTGGAAGCCGGACT |
| | | ATGGCAGTGGAACCTGTCAC | CTGTCACCGGGGCTCCCAAT | CGTCCACAGCAGGATCAGTA | CCATTGGGAGCCCCGGTGAC | TGTGGCGCCGCTGTCTGGCC |
| **Alpha-(1,3)-fucosyltransferase 6** | FUT6 | GGACCCATTAGGGTACACAG | GAATCCAGGTACCAGACACG | GGTTGTACATGACCTCTCGG | TGGCAGCTGAAAGCCATGGA | GTACCAGACACGCGGCATAG |
| | | GTACACGTCCACCTTGAGAT | CGTACACGTCCACCTTGAGA | GGCGGAGTTTGGCCCCCAGT | AGACACGCGGCATAGCGGCT | GACCCATTAGGGTACACAGT |
| **Alpha-(1,3)-fucosyltransferase 7** | FUT7 | AGTCGGTGAACAGTCGCACG | GAAGGCGTCAGCCGGCACGA | GTGGTGGAAGACCACGGCGT | GGTCCTCATAGACTTGGCTG | TGCGGGGTAGGTGTGGGTAG |
| | | GTAGAAGCGGTACTGGGCCA | ACATGCACGAAGGCGTCAGC | TGCGACTGTTCACCGACTGG | CACGCCAGGCAAAGAAGCGT | ACAGTGGCCGTCCATTGGCA |
| **Alpha-(1,6)-fucosyltransferase** | FUT8 | AATTGGCGCTATGCTACTGG | TGTCAGACGCACAGACAAAG | TTCTCTCGAACTTTGTAGGA | CAAGGGTAAATATGGAGGAC | CCGTCCTCCATATTTACCCT |
| | | ATCTGACAGAACTGGTTCAG | CTCGTACAAGTCGATCTGCG | TTCCAAGATGAGTGTTCGCT | GATCGACTTGTACGAGTGCA | GCTTCAAACATCCAGTTATT |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Alpha-(1,3)-fucosyltransferase 9** | FUT9 | CTTGGATATCTGAATCACGG | AAGGGTGGCCTAGCTTGCTG | AATCCACAAGGATTACATCA | GCACTTGGATATCTGAATCA | CACGGCGGTAAGTCAGAGTC |
| | | CTTACCGTCAAGAAGCCATA | AGTGCCTTATGGCTTCTTGA | TTTCCGTGATGTAATCCTTG | GGGATTTGTTGTACAGTGAA | CCATCACCGAGACATCAGTT |
| **Galactosylceramide sulfotransferase** | GAL3ST1 | CGGACCATCTGCATCGACGG | TCAAGAAGAGCATCGGGCAG | CATGCGCTTCCACTACGACG | CGGACCATCTGCATCGACGG | TCAAGAAGAGCATCGGGCAG |
| | | CATGCGCTTCCACTACGACG | GGAGGACTCGAACAAGCGGG | AAGAACACGATGTTGCGCCG | GGAGGACTCGAACAAGCGGG | AAGAACACGATGTTGCGCCG |
| **Polypeptide N-acetylgalactosaminyltransferase 1** | GALNT1 | GTTGGTTGCCTTTTAGGTGG | GGAGATATATCGTCAAGAGT | TTAGAGACTGCAATGGAAGT | AGAGTTGGTTGCCTTTTAGG | TCACTATTTCTCATTGGGAG |
| | | GTGAGCGATTAATGACACTA | CTGGCTCTAGAACTAAGAAT | GGTGATTGTTTTCCACAATG | GAATTTCAGGACACCTACCA | GGCTCTGATATGACCTATGG |
| **Polypeptide N-acetylgalactosaminyltransferase 10** | GALNT10 | TTTGGGCGCTGTACCGCGAG | CTTCGAACCAAGAAACGGGA | GCACACTGTGGACGGTGCGG | GCGAGCGATTGAGCACACTG | AGGTCTCCCGTGATGGCCGG |
| | | TGATAAGGACCCGAATGCTG | ATCTCGGCGACCAGCTCTGG | CTGGACTGCGACATCCCCAG | CGCTCAGCGCGTAGGTACGG | ATCATCCCCTTCCACAACGA |
| **Polypeptide N-acetylgalactosaminyltransferase 11** | GALNT11 | GGTGAGTACCCTCTAGGCGG | TCGAGGGAGAATGATTGGCG | CCACGGCTCATGAAATGCCA | GGCCAGATCTGGATGTGTGG | GGCGGTGAGTACCCTCTAGG |
| | | CTCGCTCATCAGACCCGCCA | TCTGAGCTAGGACGAGCGGA | CAATCGGCTATACCAGGTGT | GAGAAGTGATACCTGTCGCG | GTCAATCACTGGGCACACCA |
| **Polypeptide N-acetylgalactosaminyltransferase 12** | GALNT12 | TGACCTGATGACATCGACGG | AGATCGGCGGTTTCGACTGG | GATCCGCGCCAACAAGAGAG | AACTCACCAAGCGGGCACGG | GGCCGAGGAACTCACCAAGC |
| | | ACAGTGATAGAGGTGAGTCC | CGAATACCTGGGGAACTCCG | CAGTCGAAACCGCCGATCTG | AGTCGATCACATCAATCACC | CTGACCTGATGACATCGACG |
| **Polypeptide N-acetylgalactosaminyltransferase 13** | GALNT13 | TTTGGGGAACAGGATACCAG | AAGCAGCTGCTCCTCGAAGA | ATTAGGATGGAAGAACGCTC | GGAGCAAGTAACAATCTCCA | CTCTGAGAGTAGATAGTGTG |
| | | TTTAGACAACATGGGCCGCA | TTAGGATGGAAGAACGCTCT | AACATCCAAGCACAAGTCAT | TTAATACGTGCCCGTCTTCG | TCTAGCCACTTCGCTGATGT |
| **Polypeptide N-acetylgalactosaminyltransferase 14** | GALNT14 | TTCCGAGTGTGGATGTGCGG | TGTCTTACAGGACTACACGC | CGCTGCTCAGGACCATCCGC | GTGCAGACCCCTAAGGTAGG | TTGTCTTACAGGACTACACG |
| | | TGATGGCACCGAGAACGGCA | CGGTTTAATACACTGCAAGA | GATACAGATATGTTCGGTGA | AATGCTTGCGCAATAATGAA | TCTCCCCAGAGCAGAAGGCT |
| **Polypeptide N-acetylgalactosaminyltransferase 15** | GALNT15 | TCTCTTATGTCGCTGCGAGG | GGGATTCCCCAAAGTCCAGG | CTATCACCGGAGATACCACT | GGGTTCGCATTGCTGAGACC | TGGGTTGTCGGACATTCCAC |
| | | GCCACCCTTTATCTCACTGC | GACTACCCACCTGTTGCTGC | GGAGTTGACCCCGTTCAGCC | CACGTGGGTCCAGGCTGAAC | TACAGTGGGTGCCGCACCTC |
| **Polypeptide N-acetylgalactosaminyltransferase 16** | GALNT16 | TCCCCGGCATCATTAAGCAG | CCCCGGCATCATTAAGCAGG | ACTGCGAAGTGAACACCGAG | TGTGAGTACGCCCCGAGCGT | AGTGCCTGCGCAATGATCGG |
| | | CTTCCCGCCGATCATTGCGC | GGGGCGTACTCACAATGAGC | AGACGTTCTCCAGGTACCAG | GGAGGCCGTACCTGTAATGG | AATGATGGGACTCACCACGC |
| **Polypeptide N-acetylgalactosaminyltransferase 18** | GALNT18 | GAAGGCATGGAAGTCTACGG | TCGCCAGCGTGTATGTGCGG | TAGCTGCATTCGATGAGCCG | CCTTGGGGGGATTTAGGTAG | TGTAGTACACGTTCTGGGGG |
| | | ATTGCTTCAGGGCCACGCGC | CTGATTTGTGTCTTGACCAG | TCAAAGCCCTGGGCAGCCAG | GCATCAAGGAGAACCGGAAG | AATGTGGGCAATCCGTGAGC |
| **Polypeptide N-acetylgalactosaminyltransferase 2** | GALNT2 | GCATCTGCTGACTTGAAGGG | ACAGTCGCACGGCCAAGAGC | GTCGGCCCTACTCAGGACCG | GTACGGCTTGCAGGCACCCG | GACAGTCGCACGGCCAAGAG |
| | | TGCCACCACACTGCCACACG | CACCCCTACACGTTCCCGGG | TCATGCGCTCACGGGTTCGG | GGCCGATGCTGCCCAAGCCA | TTCCACTGCTGCGAAAGGGC |
| **Polypeptide N-acetylgalactosaminyltransferase 3** | GALNT3 | TATGGAAGTAACCATAACCG | GTGAAGAATAGAGCACACTG | ACTCTGCTCAACATGAAATT | CAGAATAGCTGAGAACTACA | CATTTGAATCCTGAGGTGGA |
| | | GAAAAGGTCTGATCACTGCT | GGGTTGCATGACACTAAACT | GAGTGTCTGGTCCAAGATCT | CTCACGTATCCAGATATAAC | TGTATACATGTCATGGACTT |
| | GALNT4 | TAGTGCTTATCGGACACCGG | GCAGGGGGCTTCTTATAAAG | CATGGGGCTATTCGCAGTAG | TTTCATGCCTCCGCAGGAGC | TTAGGACCAATAAGCGAGAG |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Polypeptide N-acetylgalactosaminyltransferase 4** | | AGTGCTTATCGGACACCGGA | CGAAAGTGGCCCCAATCAGA | ATTGGGGCCACTTTCGCCAC | ATTCCCGTGCACTTGGGGAG | TGGGCCATGTGTTCCCCAAG |
| **Polypeptide N-acetylgalactosaminyltransferase 5** | GALNT5 | GACCACCTCATCGACCAA | GAACTTGGTGCGGGTTGCCG | ACCGGATGAAGACAGTGGAG | CAGTGGAGCGGAACTTGGTG | GTTCTCCAACCAAAGTTCAT |
| | | CATCTAGCCCTTGGTCGATG | TGTTCTCCAACCAAAGTTCA | ATGTGGAATGTAACGTTGGT | GCCTTACGGCTCCTTTATCA | CTAGCCCTTGGTCGATGAGG |
| **Polypeptide N-acetylgalactosaminyltransferase 6** | GALNT6 | GTGGGTGAGAACAACCGCGG | TCAGCGATTCGAGCCAGGAG | ACTTACGGCACCATAGAAGG | AAGAGGCCACCAGCAAACGT | AGATCCCCGACGTTTGCTGG |
| | | GAGGCCACCAGCAAACGTCG | CTTTTCCTGGGTCTCCAGGG | CGAGGAATGGGAATTGGCCC | GCTTTGCGATGTTGTGGCG | CACTGTGAGTACCAGGGCAC |
| **N-acetylgalactosaminyltransferase 7** | GALNT7 | GAAGACCCAACATAAATGGG | GTCTTCCTATACAAAAGGGA | CTCGCCGCCGGAGGAAGATG | ATGAGGGTGAGTGACCCGCC | CTTACGCAGTTTGCTGGTGG |
| | | GGCGTAAGTCATTGACGCTG | TTGCCGCCTCGGTTGGGCAT | CAAAGTTACCGAGGATCCCT | CGATCTAAACACTTTCCTGA | TCTTCAGCCTCATCTTCCTC |
| **Probable polypeptide N-acetylgalactosaminyltransferase 8** | GALNT8 | TCATAAACAGAGATACCAAG | CTTGTGGCAGTGCATCGTAG | GATTTCCTTCAACAATCGAG | AAGCTCTGTCCATTATACAA | CCGACACGCGAGACTACAGG |
| | | CAGGAGACACAATCACAGTG | TTACAAAACCTGTTTACGGG | CAGGCAGCGATCACTAGCAC | TGTAGGAGTCCTTCAATCAT | TCCTCAGACTTACTTCACTG |
| **Polypeptide N-acetylgalactosaminyltransferase 9** | GALNT9 | AGATTGTCCGCAACAGCCGG | ATGTGGCGCGGCCAACACAG | CTACCTGGGAGGACATCCCG | GGATGATGTACATGCACCAG | CTCCCAGGTAGGAGTAGCAG |
| | | AGAGGTAGCACCTCCCCCGA | CGCCATACACCTCCATGCCG | TGGACCAGTACGTCAACAAG | CAACACCCTCACGTACGGAG | TGTTGCGGACAATCTTCACG |
| **Polypeptide N-acetylgalactosaminyltransferase-like 6** | GALNTL6 | CTGCTGCACAGAACCTGAAG | GTATAATTAACCGAACCCCA | ACCTTACCCCCTTACTGAAG | GCCTAAATACTACCCTCCAG | AGTATAATTAACCGAACCCC |
| | | TGGAGAGAAGATATTCGACC | AAGGTTGGACTTCACTCCTG | GTACATTTACCAGCGGCGGC | GACAGCAAGCATGGAGCCAC | GGAGTACAGGCATCTCTCCA |
| **Globoside alpha-1,3-N-acetylgalactosaminyltransferase 1** | GBGT1 | GGGAGGAGACATCCATGCGC | TAAGAGGGCTCACCGGGAGG | CTCCGGCGATGCATTGCTGG | TTGGCGCCCATCGTCTCCGA | TCATAGGGGAACTGCTGGCG |
| | | AGGGGACTTCTATTATGGTG | GCTCATAGGGGAACTGCTGG | TCTCCGGCGATGCATTGCTG | ATGGCATCATGGCTGCCTGG | CAGTCTCCTTACCATACCAC |
| **Beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase** | GCNT1 | TAGTCGTCAGGTGTCCACCG | ACCCCTTAGTAAAGAAGAGG | ACATGGACTCCATCGCAGGG | CCAAAGGATTCCTGAAGTCC | CATAAACCACACTCTCCAAT |
| | | TGAAATTTAAAAAGCGCCCT | GCGCACATGGACTCCATCGC | TCCGCCTCTTCTTTACTAAG | GATCATACTTATGGCTGGCA | TGCAGCTAAATAGGAATCCT |
| **N-acetyllactosaminide beta-1,6-N-acetylglucosaminyl-transferase** | GCNT2 | ATGGAACGAGAGCCTCACCG | GTAAATTGTGAGATTATGGG | CTTTGTTCTGCATGACCCAC | CCCCAGCTCATGCAATTGGA | TGGAACCCGTTGTCTATGGA |
| | | TTTCGCGATGCCTCAGTTCT | GGGGAGATCCAAGCTTCCAA | CGTCCAATTGCATGAGCTGG | GGAGATCCCTCCATAGACAA | GTACTTCCATGAGACCTCGA |
| **Beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase 3** | GCNT3 | CCCAGAAACTTTCAAAGAGG | CTGAGATGTCGTACTTGGGG | CTTCCTGAATACATGTGGGA | CATAAACGCAGATAGCCCGC | CTCAAAGTGATATTTCCAGC |
| | | CATAGATGGCCTTATAACGT | TACGTTATAAGGCCATCTAT | AAACATGTTGGACGAAATCT | AACATGTTGGACGAAATCTC | GCTGCATTACTTGTGGGCTC |
| **Beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase 4** | GCNT4 | TGGAGCAAATATGTTGGAGA | TCGGGTTCCAGGAATACCTG | CTGTTCGGGTATCTATGAAC | TTACCATCATGAACTTAGAC | CAGGTGCCTTACGATCATAA |
| | | CTTACCATCATGAACTTAGA | ATTCGGGTTCCAGGAATACC | GGCTGATTTAAATTGCTTGT | TCGTTCAAGACTTTTTTGCC | ACTTTTGGGCTACCTTGATT |
| **Beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase 6** | GCNT6 | CTGTATTCACGTTGATAAGG | TAGCCATTGAGTTACTCCAG | TGGATTTTGAGTTGACGTGG | TGGAAAATCCAGGCTCCAGG | TGTAGGTATCTTGAGACCGC |
| | | AAGTTACGTTGCCCTTACAA | AGACCGCTGGAGTAACTCAA | CAAGTTACGTTGCCCTTACA | CGCAGCCCTAAATGGGAAGA | CGTATCGAAATCTTGACTGA |
| **Beta-1,3-galactosyl-O-glycosyl-glycoprotein beta-1,6-N-acetylglucosaminyltransferase 7** | GCNT7 | TGGTTACCTGCTTAGAGCGG | GAAGGCGACTGAATGAGCCA | ACTGCTCCAGGATTTCTCGG | ACGGACCAGGAGACCTGCCG | AGAAAGCAAACACTCCCTAG |
| | | CTAAGCAGGTAACCACAAGC | AGTAGTGTTGCTCTGGGCTG | TCTAAGCAGGTAACCACAAG | GCGACTGAATGAGCCACGGC | CGCTTTATTTGAAGGGAAAG |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Inactive N-acetyllactosaminide alpha-1,3-galactosyltransferase** | GGTA1P | CTTCTCGGGCCTGCACGTGG | TGCCTTCTCGGGCCTGCACG | ACACCACCACCACCACGTGC | GCACGTGGTGGTGGTGGTGT | |
| | | CACCACGTGCAGGCCCGAGA | CTCGGGCCTGCACGTGGTGG | GGGCCTGCACGTGGTGGTGG | | |
| **Golgin subfamily A member 2** | GOLGA2 | TGCGACATGTGTCAATGGGG | GAGCCGGTACCAACAGCTAG | AGGTAAAGTGCACCGGGTCG | AGTCGGAGCAGCACGTCAAG | GATGACACCGTGTTACCTGG |
| | | TCGTGAGATGCAGAACCCCC | TCGTCAGCCCGGTAAAAAAA | TTCGTGAGATGCAGAACCCC | ATTCTTGCAGGTAATGGAGT | GTCGGTTAGACAACTACAAA |
| **Golgi reassembly-stacking protein 1** | GORASP1 | AGAGGGCCTAGATACTGGGA | GTAACTCCCAACGCAGCCTG | CTGAGCTTCAAGCAGCTCGG | TTGGCTATGGGTATCTACAC | TGGGAGTTACAGTCACCTCC |
| | | GACACGTCCAGGAAGCCTAG | TATGCGTTTCTCACAGTCCG | CGCTCATCGAGTCTCATGAG | TACGCTCATCGAGTCTCATG | TGTAACTCCCAACGCAGCCT |
| **Glucoside xylosyltransferase 1** | GXYLT1 | CACTAGGGGAAAAAACTTAC | CTGGAAGAAGGAACGGGCGG | TCCCTGGAAGAAGGAACGGG | CACACAGCACCACGACGCGC | ACAACTGTACGACTACAATG |
| | | ATGAGGAACCTCGAATAGGA | GGTACGGGCAGTTCCCGGTG | TGACAACTGTACGACTACAA | GTACGGGCAGTTCCCGGTGT | ATTCGAGGTTCCTCATGTTC |
| **Glucoside xylosyltransferase 2** | GXYLT2 | CATGAAGCTCCGCAGCAAGG | AGAAAGGATGCCTAGCAAAG | GGCCTGTGGCAATCGGCTGG | AATACATAGAGACACTCTGG | TCATAGAGTGCTCTGAACGT |
| | | AGTACAAGAATGCCATCACG | CGTACATGCAGTGATCGGGA | AGGTCTCAGAAAGAGGACAT | GATCGGGACGGTAGTTCCAC | TATCGTCATGGTAGACGCCT |
| **Histone PARylation factor 1** | HPF1 | TGCAGAATGGTCGGCGGTGG | GTACTGCCGGAAAGGCTGTG | AAGCGAGTACCCCAATTCTC | ATCTCTCTGTTTCATCTTCA | GCAGAATGGTCGGCGGTGGC |
| | | AGAGGGGCCGCAGGTACTGC | TACCTGCGGCCCCTCTCCGC | AGCCTTTCCGGCAGTACCTG | ATGAGTAAGGTGGCTTTGAT | TTCTGGAGCAACTCACCAGA |
| **Interleukin-15** | IL15 | ATTCTCATTACTCAAAGCCA | CAAGTTATTTCACTTGAGTC | ATCATGAATACTTGCATCTC | GCTTTGAGTAATGAGAATTT | ACTTATTACATTCACCCAGT |
| | | TATTGATGCTACTTTATATA | CAACAGTTTGTCTTCTAATG | GGCATTCATGTCTTCATTTT | TCCTAAAACAGAAGCCAACT | CAAAGAATGTGAGGAACTGG |
| **Isoprenoid synthase domain-containing protein** | ISPD | TGTAGGGTGTAGCTGATGAG | CTGGAGAGGTAATGCGGCGC | TGGAGAGGTAATGCGGCGCC | GGCAGGCAACACAGCTGCCA | CTACAAACGAGATCTCTATG |
| | | GGCGCCGCATTACCTCTCCA | GGATGGGGCAGAATTGCTTC | AACACAGCTGCCACGGCTTG | GCGGCTGAATCGATTATTAA | GCTCATCAGCTACACCCTAC |
| **Histone acetyltransferase KAT2B** | KAT2B | ACACTCGGCCAGGGATGAGG | AGGTGGATTGATAACTGCCC | TGTAGTATTCACTCTCAGGG | TGGGATGTGAGCTAAATCCA | ACCTGTGTGGTTTCGTACCG |
| | | CAGTTTCTTGGCCCGCGGAG | AGTTCTGCGACAGTCTACCT | AAAGATGGCCGTGTTATTGG | TCCGATGGAATTAATCAACG | GAAGTTATAAGGTTCCCCAT |
| **Potassium voltage-gated channel subfamily E member 1** | KCNE1 | TCTACATCGAGTCCGATGCC | CAAGGCCTATGTCCAGGCCC | TCTACATCGAGTCCGATGCC | CCTGTAGCTCTCCAGGACCC | CGAAGAATCCCAGTACCATG |
| | | CCTGTAGCTCTCCAGGACCC | ACAAGGCCTATGTCCAGGCC | TCTACATCGAGTCCGATGCC | CAAGGCCTATGTCCAGGCCC | CGAAGAATCCCAGTACCATG |
| **LARGE xylosyl- and glucuronyltransferase 1** | LARGE1 | ACTGCCCGGAGTACGACCGG | ATGACGGCAATCTTCTGAGG | CTTGGTAATGTCGAAGCTGG | GTAGAAGTCCACACGCACAG | TGTTACTTCTGGATAAGCTG |
| | | CCACTCAACCCGGTAAGGCG | CTTATGAGCCGCCACAACGT | TCAGGGCAGCAAAGCCGTAG | GTAGCGCAGTGTCTCGAACG | GCGCCGGATACAATGCCAGC |
| **LARGE xylosyl- and glucuronyltransferase 2** | LARGE2 | GTTATGCCCCGCACACACGA | CCCGTACCGTGTGCAATGGG | CCACTTGGTGACTGACGCCG | GGCCCTGGCACAACTGGACG | AGTACGATGGGAACCTGCTG |
| | | GGCCGGGGATTTAACACAGG | CAGATGGATGGTGAAGGCCT | CTGCTGCCGGAACTCAAAGC | CCACAGACTATGCCCGCTGG | GGACTTGTCCCGCCACCATG |
| **Protein ERGIC-53** | LMAN1 | TCTTGGATTCTGGAACGCGG | AGATGGCGGGATCCAGGCAA | CCCGTCACTATAGTGTAAGG | CGTTCCAGAATCCAAGATGG | TCACTCGGTCGCTTCGTCCG |
| | | TGACGGGGCTAGTCAAGCTT | GGACAGGATAGGGTTTGTTG | AGAGGTCGAATTGGAGCTGA | TTGTACTCGAAACGGCGATG | CATTCCACTGACCAGTCTGA |
| **Lipase maturation factor 1** | LMF1 | CCGGCACTCACAGTACACGA | CATCACCAAGGAGCGGGCGG | CGTGTACTGTGAGTGCCGGG | TTGTCCTGTGGGGCTTCCGG | TTTCTGGTGCACGTCTAGAG |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | GGGCCTGATCAAGATCCGGG | CAGCATCACCAAGGAGCGGG | TGCTGAGCGTCTCGAAGCGA | TCTCCGGCCCATCTCCACAC | GTGGTTCGCGGCCTTCCAGG |
| **Magnesium transporter protein 1** | MAGT1 | GTGGGGCTTTCACAAGGCGA | GAAAGAAGGAGGTGAGAACG | CTATAAGTGAAACTTTGCTC | GAGCGAACATGGCAGCGCGT | GCCCATAAGAATCCCCACAC |
| | | GGGCGTGAGAACAGGCAAAT | CAGTTCTGTCGGCGATCCAC | AGTTCTGTCGGCGATCCACC | TGCCCATAAGAATCCCCACA | AGCAGTGAACATGACGATAA |
| **Mannosyl-oligosaccharide 1,2-alpha-mannosidase IA** | MAN1A1 | CGAACTTCTCCGTCAGGCGG | GATCCGCGAAAACCACGAGC | TTGATCCGCAAGTCTAGCAG | GCACGAACTTCTCCGTCAGG | CTTCGTGCCCCCAATCGGGG |
| | | GCATCGCTCCCGCTGTCCAG | CGTGACAAGGCGCCGTTCAG | TTTCTCGCGGATGGCGGCGT | GGACTTCGTGCCCCCAATCG | CTTTGCCCTTTTCTCGCGGA |
| **Mannosyl-oligosaccharide 1,2-alpha-mannosidase IB** | MAN1A2 | GTGGTATCCTACGTCCAGAG | GATCGTGAGTGAATCGCCAT | TTAATTCCACATGTAGATGC | TAAAGGTAAAAGGTCATCAC | CTGGTCATTATTTAGAGCTA |
| | | GAATCGCCATAGGTACCAAT | AGTATTGCCGAGTTAATGGT | TCCATGGATCAGGAAGACTC | AGGTTTCAATTACTTCTGGA | CAGAAAACCCACCATTAACT |
| **Endoplasmic reticulum mannosyl-oligosaccharide 1,2-alpha-mannosidase** | MAN1B1 | ACAACAGCAAGAGTTGGCGG | GAGATGAAGTCCCGATGAGG | TTTGTGGGGGAGCTTGCCCA | GTTCAGATGCACTAGTAGGG | CGGTGGGTACATGACTACAG |
| | | TCTGCGGATCTCCTTCCGGG | AGAGCACGATCCGCATCCTG | AGCAACTGTCGAGATTGCAG | CTGGCTCTGGGCGTCTACCA | CGTCGATCAGTGTGAGACCG |
| **Mannosyl-oligosaccharide 1,2-alpha-mannosidase IC** | MAN1C1 | TGCTGAATGTCTCTCCCGGG | TTGCTGAATGTCTCTCCCGG | CTTGCTGAATGTCTCTCCCG | TAACTCCGGCAGAGAGGCCG | CTACATCCTCCGGCCAGAGG |
| | | GTGAACATCCGCTACATCGG | GGCTGACCTACATTGCCGAG | TGAAATCGCCGGCCATGCCC | CACGGGAATCCCAAAGGGCG | CATGATCGCCCTTGGCGCCG |
| **Alpha-mannosidase 2** | MAN2A1 | TGCCGAACACGGTGAACTGG | TGTGTCACTCTGACAAAGGG | CGAGATGATCATTACTGGAG | CATTTTCCTCGATAGACTCG | TACACTTTCAGTCCCAATGG |
| | | GGCACCACAAAGACTTGAAG | TCAATAGCCCAGCCGGACCG | TCACCTTGCAGCTCAAGGGT | ATTCCGTGTATTCACAGTAG | ACAGGAGTGAAACCTCGGTC |
| **Alpha-mannosidase 2x** | MAN2A2 | CGGTTCGCCGCTCTAGCAGG | CTATGAGCACATTCACCAGG | TTCAGTGACACGCAGCACGG | CGGGGCGTTCAGGTACATGG | AACCAGGCACGACTCACCCG |
| | | GTCCGAAGGCCAGTACCAGG | TCAGTGACACGCAGCACGGG | CGTGCGTGTCCTTTCGGAGG | AAAAGAGAGCGGCAGTCCGA | CATGGTGGAGCTGTATCCAA |
| **Multiple coagulation factor deficiency protein 2** | MCFD2 | TTCTCCCAACCCGGCAGCAT | CTTATCCAGGCCCATGCTGC | CATCACTCATGTCCATAAGG | CAAGAGTACGTATTCAGCCC | ACTCATGTCCATAAGGAGGT |
| | | CGTGCACTGTGTTCTTATCC | TTATCCAGGCCCATGCTGCC | TGTCCATAAGGAGGTAGGTC | CCGGGCTGAATACGTACTCT | GTCATCAATATCTGTGAGAT |
| **Alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase** | MGAT1 | TCTGGGACGACTGGATGCGG | TCATGGATGACCTTAAGTCG | ACCAGGATGGGAATCACCGC | TAGCCCTCCCACGTCAGTGG | GCCCAGGCCAGGGAAAAAGT |
| | | GTCTCGGCCTGGAATGACAA | AGATCGCGCGCCACTACCGC | TGGACCTGTCTTACCTGCAG | TGAGATCTCAGGGCGTATGC | CCATCGAGAGCGCTGACTGA |
| **Alpha-1,6-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase** | MGAT2 | GGTTCGGCGTCCAGCAACGG | CAGGGTCTCCAACGTGTCGG | GATGCACCCCAATTTCAAAG | GCTAATCCTGACGCTCGTGG | GATCAATCAGCTGATCGCCG |
| | | AGGCGGACAACCTGACGCTG | GTTCGGCGTCCAGCAACGGT | AACCGGACAGAAATTCACCC | CCGCAGGGTCTCCAACGTGT | TGAGGTTCCGCATCTACAAA |
| **Beta-1,4-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase** | MGAT3 | GCATGGTGTAGTACTGGCGG | CCAACCTGCCCACCAAGGAG | CGGGACCTGAACTACATCCG | CGGGACCTGAACTACATCCG | ATGGCGTTGATGACGCGGCG |
| | | CTTGTGGCGGATGTACTCGA | GGGTTTGAAGCAGACGCCGC | CTTTTTCTGGAACAATGCCC | TATTTCGTGCGCACCAAGGC | AGACATAGAGCACCTTGTGG |
| **Alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase A** | MGAT4A | GAAGTAGATACCTCCGCAGG | GGTCTACCAAGGGCATACGC | AGGTAAAATGTTTCAAGCGC | AATCCATGTAAACCCACCTG | ACGCTGGAGAAAACTTACAT |
| | | TCTACAATCAGAGTAAGATC | TAAAACAGCAATGGTGCCTG | TGGGCTATCACACCGATAGC | GGAGGTATCTACTTCCTTGA | AAGATGTAGTCTCCAGCTAT |
| **Alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase B** | MGAT4B | GAAGCGGATGAAGTCCCCCG | CATCGTGGTGCTGATCGCCG | GAAAAAGGCCGACTAAGCTG | ACGTTGTGGACGTTTACCAG | CGTTCCACGGCTTCAATCGG |
| | | AGGACCCCCGATTGAAGCCG | GGGAGTCCGTCTGGATCGAG | AAAAAGGCCGACTAAGCTGC | ATGCGGTGAGCAAGAGCTGG | ACCACCGACACTATGGGGGA |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase C** | MGAT4C | CATCATGGAGCCCTAGATGT | ACACTGTAGAACGTTTTCTC | AACGTTTTCCCCAACATCTA | ACGTTCTACAGTGTCATTCT | ACGTTATGCCTAGCAAACAA |
| | | CTGTGTAATATCCTGGACCA | ACGCTTTCTTTGTAAAGGTG | TGTGGCAGCTAGGTAGCGAT | TTTGCGCACCATATTATTGC | CTTTTAAGGCCATCTAGGAT |
| **Alpha-1,6-mannosylglycoprotein 6-beta-N-acetylglucosaminyltransferase A** | MGAT5 | ACGCTGAAGTTTTCATCGGG | CTGCTACGCAGACTATGGAG | CCATATGAATTTACGTGCGA | TCCGGCGAATGGCTGACGCA | ACATGAAACAGGAATACTC |
| | | ACCATGGTATCCTCAGTGGA | GCTTGACCTGTAGGGCGCTG | TGTGTATTGCCTCCTATGGA | AATCTTGTTGTCAATGGCAC | CAGTGAGGGTAGCCGTCCAT |
| **Alpha-1,6-mannosylglycoprotein 6-beta-N-acetylglucosaminyltransferase B** | MGAT5B | ATTGGCGATGGCCTCCAGGG | GAGAAGCGGCTCATCAAAGG | CTACAGTGTCAGAAGGCCGG | CCGTGTACTACGAGAGCCAG | TGGTTGAGGGAGCTGTGGGG |
| | | ACAAAGGCTGGCACCTCGGG | GCCCATGAAGGAGTTGTCGG | GTCCGAGGAGCTCAACGAGA | CAGGAGGGAACTTTACCGAA | GGGTTTGGCTTCCCCTACGA |
| **Protein O-GlcNAcase** | MGEA5 | TCAATTGGCAGCAAACGCTG | GATCCACAGAACTCATCCCA | TTGGCGAGAGATGTATTCAG | ACTGGTTCCGCAGTGTACAA | CTGTTGAGATCAGATCCAAG |
| | | GTCAAGCGACGTTGGAGGAG | TAGACCACCTCTAAAGGCCC | CTGTTTCTGGGCCCGTACAA | GAAGGAGAGTCAAGCGACGT | TCAAGCGACGTTGGAGGAGC |
| **Mannosyl-oligosaccharide glucosidase** | MOGS | TCGGCTAGAATGTCCAGCAG | TTGGTGGCAAAAGTGTAAAG | TGGGCTGGATGACTACCCCC | TATGGCTGGGAGTTCCACGA | TGTGTTGGCGCCCGAAGGAG |
| | | TGGGCTGGATGACTACCCCC | CAAGGGAGCGTAAACCAAAG | TACAGTGACCGCGATGGGCG | CACACCCTTCAGTAACCGAG | TCACCACTGAGTTCGTCAAG |
| **Mannose-6-phosphate isomerase** | MPI | CAATGGCTCACCTCGCGGAG | GTTCCAACAGCGAAGTGGCG | ACAAGGTAAAGGACAGAGTG | TTAGGGTCTTCTGTGAGATG | TTGGCATCGGGGTAGTGCTG |
| | | CGCTCCGCGAGGTGAGCCAT | CGCGAGGTGAGCCATTGGCT | AGCTCAACCTGTTGGTGAAG | GGCATACTGCTGCACCGCAC | CAGCGAAGTGGCGCGGCTGT |
| **Metallothionein-3** | MT3 | GTGGCGTCGCCCTCTCTAGG | CGCGGACTCCTGCAAGTGCG | GCGGACTCCTGCAAGTGCGA | AGGTCTCAGGGTCCATGTCG | ACGGAGGGGTGCCTTCTCAC |
| | | CAAAGGCGGAGAGGCAGCTG | GCACACTTCTCACACTCCGC | AAGGACTGTGTGTGCAAAGG | TGAGAAGGCACCCCTCCGTG | AGGTGGCTCCTGCACCTGCG |
| **Mucin-1** | MUC1 | GCTGCCCGTAGTTCTTTCGG | GATCGTCAGGTTATATCGAG | CTGACGATCTCAGACGTCAG | CTACGATCGGTACTGCTAGG | GCTGCCCGTAGTTCTTTCGG |
| | | CTACGATCGGTACTGCTAGG | GAACCCGTAACAACTGTTGC | CTGCAGCTCTTGGTAGTAGT | AGTGCCGCCGAAAGAACTAC | GCTACGATCGGTACTGCTAG |
| **Mucin-12** | MUC12 | ATTGCTGTGAGTATATTGGG | GAATCTACTACCTTCCACAG | ATGGGGACACTCACAGGCAG | TTCCCATCCACTCTCGGCAG | TCAAGAATCGGGTAAGACCA |
| | | GAATCAACAGTATCCCACAG | TGAAGTTGTGATGCTGCCAG | ACGATTCCCACTTCCAGCAA | GGTACTGCCGGGTAACGCTG | GTTGCAGCTGACATGTGCCC |
| **Mucin-13** | MUC13 | CTCACCTAATAGTCAGGGCG | GGTGCTGGTCTCCAATAAAG | ACTCACCTAATAGTCAGGGC | GACTCACCTAATAGTCAGGG | GCCATAATAATCACACCGAA |
| | | CATAACGAATTATCTGCACA | AACACTTACCAACGCAGAAA | CTTTCCTAAGGTCAGGATAA | CCAACGCAGAAAGGGCTCTG | ACTGCGGATGACTGCCTCAA |
| **Mucin-15** | MUC15 | TCTGCGATTAGACAATGCAC | TGTCGTCATAAAGTCGCCGA | TATACAGAAGTACGAAGTGG | AAACGGATTCATTTTCCCAT | TCACTTCTATCGGGGAGCCA |
| | | TTTGTGGTAAGCCATCCACT | GTCGTCATAAAGTCGCCGAT | ACCAAAGAAGCCTACAATGT | AAACTCACATCATAAGGTTC | TGCACGTGATGGCATTCCTA |
| **Mucin-16** | MUC16 | GAGGAGAACATGGGTCACCC | GCCGGTGGCTATAGTGAAGG | GCCCTGGACTGGACAGAGAG | GCAGACAGCATCCACTCTGG | GGCTTTTGGGGTCAGGACGA |
| | | GCTTTTGGGGTCAGGACGAT | GGTGCAGACAGCATCCACTC | GTGAGCAACTTTGATGTGGT | GAACAAGGGCTTGAGCTGTT | GTTCCAGCACAGCTGCCCAC |
| **Mucin-17** | MUC17 | GGGTCTATGTGTCTCAAGGA | GGGTGAACATCACAAAGCTA | TGTAGATGTTGTACCTGTGG | GTTTCACCTCTCTCTTGGAG | ATATGGAGACTACTTCGTAG |
| | | TTGCTCACCTCTACCAAAGC | TATAACTGTCAGAACAACAC | CCTACAACTGCTGAAGGTAC | TGGAGTAGACATACGCATAG | AGCAGTTGTAGGAGATGAAC |
| **Mucin-19** | MUC19 | GATCTATGACCACTGCACTG | GCAGTCAAACAGGTGAGCAT | GGATCTATGACCACTGCACT | GATAACAGTCACATTTACAA | GTTCAGATGGCAAAATGCAC |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | GACCATCTGAGGCACAGAGG | GTCGCTTCGCCAGGATTCAA | GTCCTGGCACCTTTGAATCC | GGCACCTCTTGGGTGGCGCC | GCATGTTATTTTATTAAAAC |
| **Mucin-2** | MUC2 | GTTGGGGTCAACCGTAGTGG | AGGCGAAGTTGTAGTCGCAG | GATGTGGTCAACTCAGCAAT | TTGGGGTCACCCGTTATTGG | GGGTCACCCGTTATTGGTGG |
| | | TTGGGGTCAACCGTAGTGGT | GTAGGGGTTGTCGTTGAGAA | GGGGTTGGGGTCAACCGTAG | GAATTGATTGGAGACGTCTG | ATGGGTGTCAGGGTTGTAC |
| **Mucin-20** | MUC20 | GTCAGCTGCACCTGATGCCA | TTGGTTGCGATGTCCATGGT | ATGGAGGTTTCCTCCTCCTG | TCAGATCCCGGGGACCATGA | CGGCTGCTGTTGGTTGTAGT |
| | | CTACAACCAACAGCAGCCGA | ACTACAACCAACAGCAGCCG | CTAAGGTGCTGTTCGTCCCT | GGGGTCCCAACCGTGGCATC | TGCGTGTCAGGAGAGGCTAA |
| **Mucin-21** | MUC21 | GCTCATCTCCATGGCTATCG | ACCAGTATCCTCGATAGCCA | CGATAGCCATGGAGATGAGC | CACCCTGGTCTCGGTTGTGG | TCACTGCAGTAGATGCACTA |
| | | GGAAGATTTCCCACGGCACC | TCCATGGCTATCGAGGATAC | GTGGTAGACAGCTGTGTTAA | TGATCCCACTGGAGGTTGTA | GAAGATTTCCCACGGCACCA |
| **Mucin-3A** | MUC3A | CCACAGGTAAGGGGGAGAGG | CGTCGTGGTCACACAGATGG | GTGTCACCAAATGCACGTCG | TTGAAGGCAAACTTGGACGT | AGGAAAGCCTGTGATAGAAG |
| | | GGTAAGGCCCCGCTCACCAT | GAAGGTCCAACCTGCAGTGG | CACCAAATGCACGTCGGGGG | AAGTCCTCAGGATGGTCGGG | GGCTGACGTCAGTACAGGAG |
| **Mucin-4** | MUC4 | TTTTGCCGCAGCCATGAAGG | CACTTACCTGGGACCACATG | GCGTGCCCCCTTCATGGCTG | ACTTTTGCCGCAGCCATGAA | GACTTTTGCCGCAGCCATGA |
| | | TACCTGGGACCACATGCGGA | GGCTGCGGCAAAAGTCCCCC | CAGCCATGAAGGGGGCACGC | GTCTGGTAAAACCTTCACCA | AGGCCTCGGGCCACTTACTG |
| **Mucin-5AC** | MUC5AC | ACATTTGCCTTACCAAGCGG | CCAGGAGTGCACGTGTGAGG | GATCCCGACCTACCAGGAGG | CGGGGTGATGACAAACGAGG | GGACGTCGGCAGCTACCTGG |
| | | CACTTTTCCCCACAGACCGG | TCGCAGCTGACCGATGCCGA | GTGTGCCTGCGTCTACAACG | CTCGCTCGAGGGCAACACGG | TCACGGTTCTGGTACGGGGG |
| **Mucin-5B** | MUC5B | GACTGGCGAGGTGCAACCGG | TCGTAGTCGTGACAGAGCGG | TTCAACGTCCAGCTACGCCG | AAAGAGCATAGAGTGCCGGG | TCTGGCAGGGCACCTCAAGG |
| | | CCTCTTCCACAACAACACCG | GTGGAACAAAGCTCACGCGC | GCAGGGCACACCATGGATGG | CCAGCGCTACGCCTACGTGG | CTCTTCCACAACAACACCGA |
| **Mucin-6** | MUC6 | CCTGGTACACGCACCTCTGC | GCCTGGTACACGCACCTCTG | GCAGCCACTGCATCAACGGG | CCTTACGCAGGCAACACCGG | TGTCCCGAGAGAAGATACCG |
| | | TTCGCTCACAGTGACGACGG | GGAGTCCACAAACAGCGAGG | TGAGTACTTCGACCACGAGG | AGCTATTGTGACCTCCCCGG | TTCCGGGCAGAAGCACCCGA |
| **Mucin-7** | MUC7 | CACACACCAGACTACTACTT | GCGTTTGTGCAGACATTTAT | TGCACTGAGTGCTTGCTTCT | TTGGGGGGGTTATTAGGTGA | CAGAATTATTGACGACATGG |
| | | AATGTCTGCACAAACGCTGT | TCTTGTGGAGCTGGGGAAGA | TGTGGTCAACCCTACCTTAG | AATTTGGGTTGTAGCCACTA | TTATAGGACTTTCTAATGAA |
| **Mucin-like protein 1** | MUCL1 | CAGGGACACACTCTACCATT | GCAGTTTTACCCAAATGGGT | TCAGCTGGAGCAGCTGTTGT | CAGTTTTACCCAAATGGGTT | CGGGAGATCCCCAACCCATT |
| | | ACGAGCAGTGGTAGAAGCAG | CTAGCTGGATACGTGTCAGC | TGGTGGTGACCTGAAGATCA | AGGGACACACTCTACCATTC | TGGGTTGGGGATCTCCCGAA |
| **N-acetylglucosamine-1-phosphodiester alpha-N-acetylglucosaminidase** | NAGPA | CTGCCGAGACAAGACCGGGG | GGTCCCGTTGAGCACAAAGG | TGCTGCGGCACAAAGCACCG | GGAACGTGGTGAGCGACGAG | CGGTCCATCCGGCATCACAG |
| | | CATAGTCCCCATGCAGGCGC | TTTCCCACAGGTTGATGCTG | TCCCCGTTCATCTCCTGCAG | GGGTTCGTGCACACACACCA | GACGGGCACTGCCAATGCAC |
| **Niemann-Pick C1 protein** | NPC1 | GCAGGCTTCTGGTAAGCCGG | TCCAGGACAAAGTACACAGG | CGCAGGCTTCTGGTAAGCCG | GGTGCAGCAGATATTTAACG | ATCGTCGATCCAGGACGAGG |
| | | TACCTGGACAGAAACTGTAG | TAGCCTCCCAACACAAGCCA | ATAATCGTCGATCCAGGACG | CTTGATCGTGCTGAGCTCGG | CAGGCTTCTGGTAAGCCGGG |
| **Uridine diphosphate glucose pyrophosphatase** | NUDT14 | AACGGTCACGCTGTGTACGG | GTCACGCTGTGTACGGGGGG | TAGCAGCTGTAGACCAGGAC | GCTCACGCTGCATTACCGCC | CTGTAGACCAGGACGGGCCT |
| | | CTAGCAGCTGTAGACCAGGA | AGAACGGTCACGCTGTGTAC | TGGCGGTAATGCAGCGTGAG | TCACGCTGTGTACGGGGGGA | AACTGTCACCCCCGCTGAGC |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Dehydrodolichyl diphosphate synthase complex subunit NUS1** | NUS1 | CGTCTACGACCACCAAGGTG | AGCTGAAAAGAAAACACGGG | ACGACCACCAAGGTGAGGCC | GAGCTGAAAAGAAAACACGG | CGCACCGGGCCTCACCTTGG |
| | | CGGTTCCTGCCGACTGCCGG | GCAGGAACCGCCGTCACCAC | CGGCGGTTCCTGCCGACTGC | GCGGTTCCTGCCGACTGCCG | CGCGCGGGTGCCGGTGGTGA |
| **2'-5'-oligoadenylate synthase 2** | OAS2 | CTTACTCAGAGCGTTGAAGG | GGACGGAAAACAGTCTTAAG | GTCTTAAGAGGCAACTCCGA | GAAGAGGACAAGGGTACCAT | GGCAGGGTGGCTCCTATGGA |
| | | TAAGACTGTTTTCCGTCCAT | ATGCTTACTCAGAGCGTTGA | CGTAGGGCTTCAGGTATTCC | CTGCACCAGGGGGAACTGTT | GACATTCTTCGTAGGGCTTC |
| **UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase 110 kDa subunit** | OGT | TGGCAATAGACACCTACAGG | AATAAGCTTCTGCCAGAAGG | CTCCAGATGGCGTCTTCCGT | GGCAATAGACACCTACAGGC | TCATACCTACCTTCAGGTGA |
| | | AAAGCGCACCACTCGTCCCA | AATACACAATGGAACTAGAG | AATACACAATGGAACTAGAG | TAACCCGTACTGAGAACGGG | GGGTGGTTACAATAATGGTA |
| **Olfactory receptor 13F1** | OR13F1 | TATTTCCTGTGAATCTACAG | TCCTGTGAATCTACAGCGGA | TTTATGGCTTTGGTGTATGC | AGTTCAGGTCATCATATTTG | TTACAGATGTCCAATTTGCC |
| | | GGTAGTTTTGTTCTATGGGA | GACAGGCTGTCTCACTGCCA | CCTGGTGCAGTTAATCATGC | TTTACAGATGTCCAATTTGC | GCCCTCCGCTGTAGATTCAC |
| **Olfactory receptor 56A1** | OR56A1 | CTAGTGTCTTCATTGTGGTG | CGGACCAAAGAGATAAAACA | GTCCGAACCCCATACACAAT | TCCGAACCCCATACACAATA | ACCCTATTGTGTATGGGGTT |
| | | GGTGATGAAGGACGTTCAGC | GGCTGAATCATGAGCTGAGT | ATAAACGTGCAGGACTCCAT | TAGATTCAAAGCAGAGGGGG | CAACACCACAACCAGCAGTA |
| **Olfactory receptor 56A4** | OR56A4 | CCGGGCTATAACAAAGACCA | TTCTAGACGTAGTAGAGTGT | GTGGGAACCACACGTGCTCA | ACTGGGTCTCCAGATTGATC | ACTGGTCAGTGATGATAGAC |
| | | GAAAGTTGTGCTTAGGATCA | ATTCTAGACGTAGTAGAGTG | ACTTTATTGGGACAGATACA | AACTGGTCAGTGATGATAGA | CTGTTCCCATGCTTTCTGCC |
| **Olfactory receptor 56A5** | OR56A5 | GCTACCAGTTTGTTATAGGT | GGGTCCAACCTATAACAAAC | TTGAGCAGGATGGGGACATC | TAGGGCTGCCATCTTTGTTG | TCACTGATCAATTTGTCGCT |
| | | ACCACAAGTACCTAGAGCTT | CACTGATCAATTTGTCGCTA | GTTATAGGTTGGACCCTGCT | GCCAAAGCTCTAGGTACTTG | TCTCGACTCAGATACTGTGC |
| **Olfactory receptor 5T1** | OR5T1 | GCTGGTTGTACCGATCATTG | AGCCAGAAATCCCCAATGAT | GGGCTGGTTGTACCGATCAT | GGCTGGTTGTACCGATCATT | TACCGATCATTGGGGATTTC |
| | | GTCATTGTCCGAAGTGTAGC | ACCAAGTTCCAGCTACACTT | TCACTCTAATAGGCAATTTA | TCTAATAGGCAATTTAGGGC | TTGTATAGATCCATATCTGA |
| **Oligosaccharyltransferase complex subunit OSTC** | OSTC | ATTGTTGAACCTCCAAGTGT | CTCATCACCGGAGGTAACTC | GTGTCTTACTTCCTCATCAC | CCTCATCACCGGAGGTAACT | CGGAGGTAACTCGGGCTGTC |
| | | GCCCTGGTTGCACATGCCGT | GCATACACAGTCATGGCCGA | GGACATTCGAGCACTAAGAA | GACGGCATGTGCAACCAGGG | GTCATAGAACCGACACTTGG |
| **Poly [ADP-ribose] polymerase 1** | PARP1 | AGAAACCAGCGCCTCCGTGG | CTGTCTGCATTGTTCAGGAG | ATCTTGGACCGAGTAGCTGA | AGCCTATGGACAAGACCCAG | AGTGCCTTCTGGGGAGTCGG |
| | | CATACCACGGGCGCTTCAGG | CGGTCAATCATGCCTAGCTG | GTGGGTAACCCCAAAGGTAA | AGTGGGTAACCCCAAAGGTA | TCCAACAGAAGTACGTGCAA |
| **Poly [ADP-ribose] polymerase 10** | PARP10 | GCGTCTGGCAGAGAACACCG | CTCCCCAGACACTTAACCGA | TCGGTTAAGTGTCTGGGGAG | GCCATTGCAACCCTGGGACG | CTGATGCATGGGGAAAAGCG |
| | | TTTTCCCCATGCATCAGTGG | CAACGCCGATGGCCATAAGG | ATGGTGCCCAGCTGAGCCTG | AGTTCTCGGACCTCCTCTGG | ATGGACTCACACGAACGACG |
| **Mono [ADP-ribose] polymerase PARP16** | PARP16 | GCATCTGCAAGCTGAAGCGA | AAAGTCCTTACAGTCGCCGC | CCGTTTGTGGTTGTCTCCGG | TGGAATGGAAGTTTTCTAGG | GAAGAGGCTGCACCGGAGGT |
| | | ACAAGCGCGACTCGGTGCTG | CTTGGACCTTACCTCTTGGG | CAAAGTCCTTACAGTCGCCG | CTCCGGAGACAACCACAAAC | ATGGTCAATGACCTCACACA |
| **Poly [ADP-ribose] polymerase 2** | PARP2 | TCAGCGTTCGAATTCCATGG | TCTACGAGTTTTCTTGGCAG | GTTCGAATTCCATGGCGGCG | AGGTTCGGAGCTCAATATCG | GGGGGCGCAAGGCACAATGT |
| | | TCCTGTATTCTTTAGGCGAG | GATTGCCTGAACAAGCCACC | GGTTCGGAGCTCAATATCGC | AAAGGTACCGCATACGGACC | AGGCCTTACCCAAAGTCATG |
| **Poly [ADP-ribose] polymerase 3** | PARP3 | ACTTATCGAAGTACAGGCAG | GAGCCCACTAGTGAGGATGG | GGTTCCAGCAGGTGAAGAAG | TTCGACAGTGTCATTGCCCG | AGAGAAGCGCATAATCCGCG |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | GAGCAAGCAACAGATTGCAC | GCTGGTAGTCTCGGTCCAGG | TGTCGGTTGCAGGGAGAGGT | CGATAAGTGTGTACTTGCCC | TCCGATTACCCAGTTTGGAG |
| **Poly [ADP-ribose] polymerase 4** | PARP4 | CGGGGGAAGATAGGAACCAA | TGAGACTCACCCGATACCGC | CGAATAACCTGGTGTGCGGG | TTGCCAGGGTGCTACGGTGG | GGTGGAGCTTCAGTGTTCGC |
| | | CAAAGCTCGGTATTTGGCCA | TGCTGACTGGATCCCACAGT | CTCATTTGTGACGTAGCCCT | AGGGAGGTCGACATCGAGAT | CACACCAGGTTATTCGCCTG |
| **Protein pelota homolog** | PELO | CAGCTTTAGTGTCTGAAAGG | GGAGCGGTTCTATGAACAGG | AGCCTGGAGCGCTGATGTGG | CATGACCCTCACTCGGGCCA | GTACAGACAGAGTCCTCCAC |
| | | ACTCTCAAGCCTGCCAGCTG | AACATCGAGAAGGACAATGC | AGCATGACCGGGCCTTGGAG | GGTACAGACAGAGTCCTCCA | AGGACTCTGTCTGTACCTTG |
| **Phosphoacetylglucosamine mutase** | PGM3 | GTGTATTGTCGAAACACGGG | CACGGGTGGCCGATATGGAA | TGGGATTATTAGCTGTCCTG | CATCTTCTGTACCAGAGGGC | GGGCTTGACTGTACAACAGT |
| | | ATGGTGTATTGTCGAAACAC | CTGTTTAATGATGGGTCCAA | GATCATGTCATGTTTCGCAT | GCATTACACGCCAAGCCCAA | GGCACTAAACCCATTACCTC |
| **Phosphomannomutase 1** | PMM1 | GAGGTACGTCCACCCTGCAG | TGCTCAGGCTGCCCAAGAAG | TGTGTTTGCCGAGAACGGGA | CAGCTTTGACGTCTTCCCCG | AGAGGACGCGCTCCTTCCTG |
| | | GGGACGGTGCAGTATAAGCA | ATGCTGAACATCTCGCCCAT | AAGTAGAGTGCAGATCGGTG | ATCGCTGAGCAGCTGGGTGA | CTACGAAGTAGAGTGCAGAT |
| **Phosphomannomutase 2** | PMM2 | ATCGGACTTTGAGAAAGTGC | TCCTAACGTGGGAGCGGGAG | CTGTTCTCCTAACGTGGGAG | AAACTCTTGTGTAGACAGGT | CGCTGTCACGGAGTAGCCCA |
| | | AATCGGAGTGGTAGGCGGAT | GATCTACGGAAAGAGTTTGC | ACACGTTTAACATCCCATTT | AAGCAAACCCACCTCTTCTT | AGCAAACCCACCTCTTCTTC |
| **Polypeptide N-acetylgalactosaminyltransferase** | GALNT4 | TAGTGCTTATCGGACACCGG | GCAGGGGGCTTCTTATAAAG | AGTGCTTATCGGACACCGGA | TTTCATGCCTCCGCAGGAGC | TTTCATGCCTCCGCAGGAGC |
| | | CGAAAGTGGCCCCAATCAGA | TGAAATAACGCTCCAGAACG | CATGGGGCTATTCGCAGTAG | TGGGCCATGTGTTCCCCAAG | ATTCCCGTGCACTTGGGGAG |
| **GDP-fucose protein O-fucosyltransferase 1** | POFUT1 | TCCATGCCGAAGAAAGAAGA | TAAAGTGGTCGGCTTGGCCG | CAGTAGAGCAGGTAACCGGC | GCACATAGTCATCGTGAGGG | TCAGTGGGAAGTACTCACGT |
| | | GATGAATGCCCACATAGGGC | AGCAGGAAAGACACGCTCAG | GTATGCCACCCGCTTCTCAG | AAGGAGGGACAGCCAAGGTA | ACAGTTGCCAATAAAGTGGT |
| **GDP-fucose protein O-fucosyltransferase 2** | POFUT2 | CGTCCTGCAAAGTTACGCAG | GGCCGATATTCTGTCGGGGG | CGGGTCTGTGGAAACGGCGA | TATGACGTCAACCCCCCGGA | CAAACACCTTGTCCAGCCGG |
| | | GGCGGCCGATATTCTGTCGG | CGGAAGGCTTCAACCTGCGC | CGGCGGCCGATATTCTGTCG | GCGTTGCGATTATTGACCAG | ACTGCTCATACTCGATGACG |
| **Protein O-glucosyltransferase 1** | POGLUT1 | GATCTAACTCCTTTCCGAGG | ATCCTACAGGTCTTGGACGG | GTCTGCCGGCGATGGAGTGG | TGCAGTAGGTCTGCCGGCGA | ACAGGAAGAGGTGTTTAAAC |
| | | TTTATCCTACAGGTCTTGGA | TCACTAAGAACAGACTGTAC | ATCCTCTCATTCTTCTGTCT | GAGGCGTAGCTGCAAGTTTC | GCCCACCTGACTCCTTCTGG |
| **Protein O-linked-mannose beta-1,2-N-acetylglucosaminyltransferase 1** | POMGNT1 | CGGATGCCTGAACAACGCCG | GAAAAGGTAACTGTCAGGGT | GGGGTATGAACACACGGCTG | GATGTGCGTGGCAACCATCG | TATGCAGTACCTGAGACACG |
| | | GCGTGTCTCAGGTACTGCAT | GCGGATGCCTGAACAACGCC | GTATAAACTGACAAACCAGC | ACACACGTTTTGCCATCACG | CATTCGAATAAAGGCCACGT |
| **Protein O-linked-mannose beta-1,4-N-acetylglucosaminyltransferase 2** | POMGNT2 | CGTATGGTTTGAATGAGGGA | GGATCTTGTTGAAGATGCAG | CATTAGGATGCACCTCTCGG | AGTAGCAGAGCCACTTGAAG | GCCGGTGTTCGTGCCAGACG |
| | | GTGGCTCTGCTACTCCAACG | CACAGGAGTCCCCCTAGGCG | AGATTCGGAAGAGCCACTCG | GTGTGCAACACGTAGCGAGC | TTCCGGCAACAGAGATGCCG |
| **Protein O-mannose kinase** | POMK | CGTTCTGGAGACCTACCAGA | CCCCTGGTGAACCACAGCTC | CGAGAGGTGCCGCCAGCTGT | GCGTGACAACATGTGTGCCT | TGACTTGGACGCCTTACCCC |
| | | CACATGACCCGTGTGCCCAC | GGGTCCACAGTGGATTGTCG | CAACAGCTGGCGGCACCTCT | GGTTACTCAAGGAACCTAGA | GTAGGTCTCCAGAACGTCCT |
| **Protein O-mannosyl-transferase 1** | POMT1 | TCGGAGCTGCAGGTGAGGAG | GTTGGACACGTGGCACGCGG | TGTGGCGACTCACCTACCCG | GAGACTGGTGAGTAAGGCTG | ACACTCACGTGTTCAGGGAG |
| | | GTAGTTCACTGCCCAGCCAC | ATACGGCGCGAGTGAGTCCG | CGTGTTGGACACGTGGCACG | TCGGCAACTGGAGATCGTCG | TTTGGTCGTGGGGCCCAGAG |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Protein O-mannosyl-transferase 2** | POMT2 | ATCACTCACCAGATGTGCGG | GATAGCCGATGGCCATCCGG | CCATCCCGTAAGCCAGAGGG | ACAGGGCCTACGCTTCTCAG | TTCTGTGGGACACCCTCCTG |
| | | GGGCCTGAAAATCAACAAGA | GGGACTCAAATGATTTCTGG | CTGTGGCGCGAAGCCCCAAA | CATTGACCCCTGAGAAGCGT | GCTCGTCCAAGCGGTGGAAG |
| **Presenilin-1** | PSEN1 | GGCAGGAGCACAACGACAGA | ACCTGCCGGGAGTTACCCTG | AGTCAGCTTTTATACCCGGA | AGGTCCACTTCGACTCCAGC | TATCTAATGGACGACCCCA |
| | | GAATATGGCAGAAGGAGACC | ACCCCAGGGTAACTCCCGGC | CACAACGACAGACGGAGCCT | ATTATCTAATGGACGACCCC | AGTCAGTCAGCTTTTATACC |
| **Pumilio homolog 3** | PUM3 | TGAAGCCATCCGCGAAGCGG | GCCACGTGAGGAAGATGCTG | CTCCCATTTGGTTTTACTGG | ACCAAAGAACAAATTCCAGC | ATTGTATGCGTACTCCACGA |
| | | GTAGACCACCGCTTCGCGGA | GATTGAAGCCATCCGCGAAG | GGAGCAGAGGAACATGCTGA | AAGATACAGAGGTCCGCAGA | CCAGGTAGACCACCGCTTCG |
| **Receptor activity-modifying protein 1** | RAMP1 | GCGCACTGAGGGCATTGTGT | AAGGGGTAGAGGATGCTGCC | GCCTGCCAGGAGGCTAACTA | CTAACTACGGTGCCCTCCTC | CAGTGCATGGCCGCTACTTC |
| | | TAACTACGGTGCCCTCCTCC | TGGGGCAGCTCCTGAAGTAG | CGACTGCACCTGGCACATGG | TGTCTACCTGGAACTGGGTG | GGACCACGATGAAGGGGTAG |
| **Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1** | RPN1 | TGTAGGCAACAATCACAGGG | ACGTCCTCATTGATCAGCGG | GGAGCTGGTGCTGAAGTCGG | GTAGCTCTCCACATTTCGAG | AGTGCAGCTCATCTGGGGCA |
| | | TTCACGTCCTCATTGATCAG | TGACGGTCTCGTCAAAGTGA | CCATACCTTAAAAGAACGGA | GGCTTTGGAGCCTGAGCTCG | TCTCGTCAAAGTGACGGTAA |
| **Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2** | RPN2 | GCGGGACCTGCTCGGAGGAA | GAGTCCCACGATGGAGTAGA | CATGCAGTTGCAGCTCTAAG | GAATGTATTGGACACCACGG | GATCAGGACTTACCCTACAG |
| | | CTACCTCACCAAGCATGACG | GTATTGGACACCACGGTGGG | TGTATTGGACACCACGGTGG | GTGAGTGGCTGTAATTAGCG | CTCATGAACAGGCTATCTTG |
| **Stromal cell-derived factor 2** | SDF2 | ACAGTTACTGGAGGATACGG | ACCTATCAGTGGGCAAAAAG | CGTATTGAGTAGCTTCACCA | AGCCGCAAGTAACGACACCC | CAGCCATCCTAACTGTATCG |
| | | GGTCAGGTAATGCTGGGGAC | GCAACAGTTACTGGAGGATA | AGCCTGGGTGTCGTTACTTG | AGTGGTGACTATGGAGGTTT | AACAGTTACTGGAGGATACG |
| **Stromal cell-derived factor 2-like protein 1** | SDF2L1 | GGGTGTCACTCCTCAGGCAG | ACGACGCCAATAGCTACTGG | CTCACCTGGTTGTTGGACAG | GCAATCGGTGACCGGCGTAG | TCAATACGCACCACCGCGTG |
| | | AATAGCTACTGGCGGATCCG | GTGCCAACACGCACAATACG | CGGACGACGCCAATAGCTAC | GCCCTCGCCGTCTTCCCCAA | CAACACGCACAATACGTGGA |
| **Stress-associated endoplasmic reticulum protein 1** | SERP1 | AGACCTCGGTAAGGAAAGAG | CGAAGATGGTCGCCAAGCAA | AGCAAGAACATCACCCAGCG | CGCCCGCTCTTTCCTTACCG | TCTTGGCGACGTTGCCGCGC |
| | | CGGCAACGTCGCCAAGACCT | ACGTCGCCAAGACCTCGGTA | TTGCTTGGCGACCATCTTCG | CGCCAAGCAAAGGATCCGTA | CCTACAGACGCCTTCTCTTC |
| **Stress-associated endoplasmic reticulum protein 2** | SERP2 | AAGCCATGGTGGCCAAACAG | TAGCCAAAACCCTGGTAAGG | ACGTAGCCAAAACCCTGGTA | GGCCAAACAGCGGATCCGGA | GATCCGCTGTTTGGCCACCA |
| | | AGCCAAAACCCTGGTAAGGC | TGGTGGCCAAACAGCGGATC | AGCCATCCGGATCCGCTGTT | ATTTCAGAGCGCACAAGCCA | GGGGAACGTAGCCAAAACCC |
| **NAD-dependent protein deacetylase sirtuin-2** | SIRT2 | TAGGTTGTCATAGAGGCCGG | ACTGTCTGCTTCTCCACCAG | GACTTTCGCTCTCCATCCAC | TCTGGGAGAATAAGTTCCGC | CTGAGGTGGAGACAGATGGA |
| | | ATACCCTGGAGCGAATAGCC | TGGACGAGCTGACCTTGGAA | AGCTTAGCGGGTATTCGTGC | GGGATGCCTGCGGCTAGGAA | CTCTAGGTTGTCATAGAGGC |
| **NAD-dependent protein lipoamidase sirtuin-4, mitochondrial** | SIRT4 | ATGCTTTGCACACCAAGGCG | TTTATGCCCGCACTGACCGC | TGTGGGGAACAGACTCCCCG | TGGCACAAATAACCCAATGG | TTCTCGCCCAGTACCGCTGG |
| | | CAGGAATCTCCACCGAATCG | GGGCGAGAAACTTCGTAGGC | AGCCGACTCCCTCTTGGTGG | TCAGGTCAGCAAAAGGCCGT | AACGCTCTTGCAGCACCCCC |
| **NAD-dependent protein deacetylase sirtuin-6** | SIRT6 | GATGTCGGTGAATTACGCGG | GCAGGTTGACGATGACCAGG | CGTTCTGGCTGACCAGGAAG | TGTCGGTGAATTACGCGGCG | GAGGATGTCGGTGAATTACG |
| | | ATCAACGGCTCTATCCCCGC | ACGTTGACGAGGTCATGACC | CTGTCGCCGTACGCGGACAA | TGGTCAGCCAGAACGTGGAC | TACGCGGACAAGGGCAAGTG |
| **Solute carrier family 35 member C2** | SLC35C2 | CGAGAATCCCGCCAAGGAAG | CTGGGGGCCTCGTTCATCGG | TGGTGTCGATGGGATTCTGG | TACCTTAAAAATGCCGGCAA | GCAGCCCTGTGTCCTGGAAA |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | GCATTTTTAAGGTACAGACT | ATCCTACCAGCGCGCGGCAC | GGCTGCTCCTGCGGGTACTT | GCGAATGCCACCGATGAACG | GTCCACACAGTTCAACGTGG |
| **Organic solute transporter subunit beta** | SLC51B | GGTAAGTGGTGAGAGATTGA | TGTGGTGGTCATTATAAGCA | GGGGTACCACAGTACCGGCT | GGGGTACCACAGTACCGGCT | ACAGTACCGGCTGGGTCTCC |
| | | CACAGTACCGGCTGGGTCTC | CTCCTGGGGTACCACAGTAC | CAGTTTCTGGTACATCCGGA | AGCCGGTACTGTGGTACCCC | CTCTCAGTTTCTGGTACATC |
| **Spondin-1** | SPON1 | GTTTTGGATAGCACCACCAG | GCGGCATGGGCATGAAGAAG | AGGGCATGCGAACCCGACAG | AGTGACTGCAGCGTGACCTG | GGTTCGCATGCCCTTCCCGC |
| | | CGAGTTCAGCCTCCGCGTGG | GTCCAGGGTCTCGTCGGAGA | CTCCGATGATCGCAGACCAG | TTGTAGGATGCGCCCATGGA | TTCCCGGCTTGTAGAAGTCG |
| **Polyprenol reductase** | SRD5A3 | TTTGCTCAGAATTCTCGGGG | GTTGTGGAACCCAAAGGTGA | TCTTGGGGACATCAAAGGCT | GGTTCCACAACTTAACTTGG | CTAGCCACCAAGTTAAGTTG |
| | | TGCCTTCTGAATTGTAGGAG | ACCGGGGCGCGGTTACCTCT | CACACTTGGTTTTCCCATAG | GCACGGTGGTTCCATATTCT | GCTATGGGAAAACCAAGTGT |
| **CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase 1** | ST3GAL1 | GGAGAACAACCCATCCGCGG | GATAAGGTGAGTGGAGGCCC | CGTCTTGCGAAAAGCCCCCG | TCTGCGGGGATGGGAAGACA | AAGCCGTACAAGTCCACCTG |
| | | ACAACTGGCTGCAAGGGCAC | CACGGGCGATACCCATCTAC | TCTCACCAGCCACCATCGGT | CATCGGGCAGCGCAAGCTCT | AGCCGTACAAGTCCACCTGT |
| **CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase 2** | ST3GAL2 | GGTGGACAGAGCATCACGGG | TCAAGTATATCCACGACAGG | GAAAAGCACCAGCATCCCCG | ACGGGCACAACTTCATCATG | TCACAATCGGATCTGCCCCG |
| | | CACCGGACGTCCAGAGGTGG | AAAGTGGCTGTCAAACCAGT | CAGGGAGCACTTCATGGTGC | TCTACCGGGGCAACTGAGCC | TCCTGCTCAAAGCCCACGGT |
| **CMP-N-acetylneuraminate-beta-1,4-galactoside alpha-2,3-sialyltransferase** | ST3GAL3 | AGGGATGTTCTAGAAGACGG | CATTGCCCACGATGATGCAG | TGCCCACGATGATGCAGCGG | TGACCTTTGATCCCAAAAGG | ACAATGGCCTCATGGGCCGG |
| | | ACGGTCTCATAGTAGTGCAG | CAAGTACGCAAACTTTTCAG | TAGGGATGTTCTAGAAGACG | GGTAGGGATGTTCTAGAAGA | AAAGTCCTGCCACTTGAAGC |
| **CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase 4** | ST3GAL4 | CGTTCCCCACGACCACACAG | CTACAAAAAGTACCAGGCGG | GATCACGCTCAAGTCCATGG | AGTCCCGTGAGTGTCATCCG | CCATCCTGAGTGATAAGAAG |
| | | AAGAACATCCAGAGGTAAGG | CCGCCTGGTACTTTTTGTAG | GATGGCCAACAGGCCCGTGG | AAGAGGCCCTGGCCATTAAG | CAGCATCCGCTTAATGGCCA |
| **Alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 6** | ST6GALNAC6 | TCTTGAGGTTGACAGGTCGG | CTCGCTGGATCACACGCACG | GTGTGTTCCGCGTGCTGAGG | GACCTGTCAACCTCAAGAAG | CGTGCGTGTGATCCAGCGAG |
| | | AGGTGTCGGCGTCCTGCAGG | AAGAAGTGGAGCATCACTGA | CCATGCCATAGACATGCACG | CAATTTGACGACCTCTTCCG | TCTTCCATTACGGCTCCCTG |
| **Alpha-N-acetylneuraminide alpha-2,8-sialyltransferase** | ST8SIA1 | CCAACTCTATACTAACCTGT | CAGCTAATCCCAGCATAATT | CCAATGCTACGCAGAAAGTT | TACCTTTGCCGAATTATGCT | CAATGCTACGCAGAAAGTTG |
| | | GTAAGACGTTGTCATAGTAG | TTGGAGAAATTCCTCGGGCA | GGAGTGAGGTATCTTCACAT | TCCAATGCTACGCAGAAAGT | CCAGAGTTGGAGAAATTCCT |
| **Alpha-2,8-sialyltransferase 8B** | ST8SIA2 | TCGGGGGTCTTGCTGAACAG | TGGCCAGGGTATACATCAAG | ACTTGGTCAATGCCACGTGG | AAGGCCCGCTGGATGACCGA | CACATCAAAAGACCCACCAC |
| | | GGGGCTTTGAAACTGACTGT | GACTGTCGGCCAGTGCGATG | CAGGGTATACATCAAGAGGC | TTTGGGACTTGTGCCATCGT | GTGCCATCGTGGGCAACTCG |
| **Sia-alpha-2,3-Gal-beta-1,4-GlcNAc-R:alpha 2,8-sialyltransferase** | ST8SIA3 | TTGCAATTTCGCCCCTACGG | GTGTTGGGGTCAAATCCAAA | CGAATGTACATGTTCCACGC | GCACTCACCGGAATCCCGCG | ACTCCCAAGTACGCCAGCCC |
| | | CTGTGCCTAAGAACTCCAAA | TTCGTAATGGGCACGAATGA | ACCAAGAATAGTGTTCGGAT | ATGTACATTCGGGGCGCCCC | TGGAAGATCTTCCCTTGTGT |
| **CMP-N-acetylneuraminate-poly-alpha-2,8-sialyltransferase** | ST8SIA4 | ATGGTGAATTGTCTTTGAGT | CACCAGGAGACGCAACTCAT | CATGAGAAATGTTTAGTGTC | ACCAGGAGACGCAACTCATC | ACCCGATGAGTTGCGTCTCC |
| | | AGATGCGCTCCATTAGGAAG | ACTATGTGCTTGACAGGCGC | GACTTATTCATGCTGTCAGA | TGCAGAACGAGATGTGTCAG | GATCGTCCACCTCTTCCTAA |
| **Alpha-2,8-sialyltransferase 8E** | ST8SIA5 | ACAAGCTGGAGAAGTGGCGG | TGTTGATCTCCCTCCCGCAG | CAACAAATCCCGGTTGGCCG | TCCACAAGCTGGAGAAGTGG | GATGATCCAGCAGTGCAGAG |
| | | AGTCAGAGCTGTTCGACAGG | CATCCGCGTCAAGTACGTGC | GGTGTTGCGCGTGTTGTAGA | TGTGCACGCGGAGGATGCCT | CTTGACGCGGATGGACACGT |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Alpha-2,8-sialyltransferase 8F** | ST8SIA6 | ACTTGTGATCATCAAGCCGG | GACTAATATGAGTTACGAGG | CGCGCCACTAACAGGTACAG | TGAGAAGTCGCTCCAACTGA | GTTTTGAACAGCATCACAGC |
| | | CTGTACCTGTTAGTGGCGCG | ACTGCATACCGCTTGTCCAC | TACAGAGTTGTCCATGGAAA | GTCTTACCTAAAAACGAAGT | TGGGACTAATATGAGTTACG |
| **Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3A** | STT3A | TAGAGGCCAACTACAAAAGA | GCTTCGTCATAATACTCCAG | TACATGGTAGATTGCAGCAG | GCTGCGCAGGTAATCCACAA | AAGGTGGTACGTGACGATGG |
| | | ATTCTACTTTCAGGATGCGA | CGGACACGGTCAAAGCCTGG | CGGGCAGATAGTACAATGGA | CAGGCAGTAAACAGTACAGT | GGTGCGTCTAATGCTAGTGT |
| **Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3B** | STT3B | TCATGGCCCTGGGAAACAGC | TCATCCACGAGTTCGACCCG | CGTGGATGATGCTTTCGAAG | CTATAGGTCTCTGCTTGGGG | AATCCAATGAATAAGGCCAG |
| | | CAGCGGTTATCATCAACCCT | TAGACTACCAAAAGGAAGCG | GATGTAAGGCCGCTAAAAGT | TACGTGTTCGGTCAAAACCT | ACGTGTTCGGTCAAAACCTG |
| **E3 ubiquitin-protein ligase synoviolin** | SYVN1 | ATACCTGGTTAGGATGACAG | AGGGGATACTCACAGTTGGG | ATGGGCCGGATGGCAAAGAG | TATCACAGCATCCTGACCCG | GTCATCCCGAAAAACGGTGA |
| | | TGCCGTGCGGAACATTGCCC | GGCCAGGGCAATGTTCCGCA | ACACAGCCTTGTTGTCCCAG | CTTACAGGGTGTTCATGTTG | GCTGAAGTCATCCCGAAAAA |
| **Methylcytosine dioxygenase TET1** | TET1 | GATTTGGCTACGACCAGTGG | GATAGGGGTTCGTCTGATGG | GGTAAAAGACAAGGGAGAGG | ACAAAGTTCATGCAACACGG | CCACCCCATCAAGAGATCGG |
| | | GAGGAATAACACCCAAAGAG | GAGGTACTGACCATTGGCAC | TTCCCGAAGGCATCGTACAG | AGCCGGTCGGCCATTGGAAG | TTGATGTGGGATAGCACCAA |
| **Methylcytosine dioxygenase TET2** | TET2 | CCACCAATCCATACATGAGA | TGAGGCCTTTCAGAAAGCAT | AAAGACGAGGGAGATCCTGG | TGACCCATGAGTTGGAGCCA | AGTTTGTCAGCCAGAGACAG |
| | | TTCCGCTTGGTGAAAACGAG | ACGGCACGCTCACCAATCGC | CCCATGGCAAAAAAGTGAAA | GTTCTATCATGGTTAAGAGC | GTTTGTCAGCCAGAGACAGC |
| **Methylcytosine dioxygenase TET3** | TET3 | CAAGTACAGCGGCAACGCGG | TCCTCATCGAGTGTGCCCGG | TGCAGATCTGGTGCGTGCGG | TTCCTGGACGAGAACATCGG | AGCCAAGATGAAGCAGCTGG |
| | | GACGGCTGCGAGGCAAACCG | TCGAGAAGGTCATCTACACG | GAGGGCCTCAATGGCAATGG | GCTGTTCATGCTGTAAGGGT | TCCGGGAACTCATGGAGGAG |
| **Thrombospondin-1** | THBS1 | CCAGGGTGTCGAACATGCCA | GTAGCTAGTACACTTCACGC | ACTTGTCATCAGGCACAGGG | CGTGGTGTCCAATGGCAAGG | AGTTGGCCAATGAGCTGAGG |
| | | ATGTTCGACACCCTGGCCGA | GCGTGGTCACAATGGTGCGC | TGTTGAGGCTATCGCAGGAG | CTGTACCTGTTTGTTGGCCG | GCAGGTCCGTGTCTGGACCG |
| **TERF1-interacting nuclear factor 2** | TINF2 | CCGGTAGCGAACCAAGCCAG | GCCGGTAGCGAACCAAGCCA | ATTCAGTACCGTACAGGTGG | ATTGGGACTGAACTCTTCGT | GTATTGGAGCAAGTAGGACC |
| | | CAGAATCTGGACCTATAGTG | TGGCCTTTAGGCCCATACAA | AGAATCTGGACCTATAGTGC | TACCTACCCCCTTCTGGCCA | CGCAGAAACTCCAGTACTCG |
| **TCDD-inducible poly [ADP-ribose] polymerase** | TIPARP | GAAGCAGTATAAAACAGGAG | ACACTGCCAGGATTGACTGG | GGTTCACCAGCTCAAACACG | CTGAATTTGACCAACTACGA | ATCAGAAACCCTCAGTGGGA |
| | | CAACTCTCGGGGTCTGAAAG | CACTGAAGCTCCAGAACGAG | TGATTGAAGAAGCCAACTCT | TACCACACTCACCAAGAGAA | AAATGTTGGGGACCAGATAC |
| **Transmembrane protein 115** | TMEM115 | GGTGAAGGCTCTGTGTGCGG | GTGGTCTGGAGTTACAGCGT | AAGCCATAGGAAGCCAGCGC | GGCACGTTGCATCTTCCTGG | GACGCCACCTAGGAAGCCCA |
| | | GGTAAAGATATGCCAGAAGA | GAATGGCCCCCAAGTGCTGG | GGCACTCAAGCAAACCATGG | GACGGTGAAGCGCTACGATG | GGTCTGGAGTTACAGCGTCG |
| **Transmembrane protein 165** | TMEM165 | GTGCCCCACAGTTCCACCCA | CAGATGAAGACCTTAGCCAC | GGACCCCTATGGTGTAGCCG | GACCCCTATGGTGTAGCCGT | TGGACCGGGAGATGTTGAAA |
| | | GTACCCGTTTCAACATCTCC | GTTCCTCTTGACCCTCATCA | GTTATAGCGCATTGCCATGA | AAGCATTGCACCAGCCAGCA | AACCGCCTGACCGTGCTGGC |
| **Transmembrane protein 258** | TMEM258 | AGGATACACGAAGAACCAGG | CTCACCATTTTGCCCCGCGA | GCCAATGGCCAAAAGCACCA | TTCCCTCCTCTAGGAGCTCG | TTAGCCTTCGCGGGGCAAAA |
| | | AACTGCTGTGCCCAGTTACG | CATCTACGTGTGAGCACCCA | GTAACTGGGCACAGCAGTTA | GGTAGAGGTGACCTCGTAAC | AAAAGCACCACGGTCAGATG |
| **Transmembrane protein 5** | TMEM5 | CTGCAGTTACTCAAGTCCAT | GATCTCACATTGTGCCCGGT | GAGTGATCTCACATTGTGCC | ATGGGAACGATAAGCTTTGT | ATTCTGTGTTTACTCCGACC |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | CATTCTGTGTTTACTCCGAC | ATCTATGAGGCTTGCTCCTA | TGGGAACGATAAGCTTTGTT | GGTCAATGCTGCATGATGAG | GTCATCACGTCTTCCACCAC |
| **Transmembrane protein 59** | TMEM59 | GGTAGGTCAACTGACAGGCC | GTGATACGGCGTCTTGCCAC | TGATACGGCGTCTTGCCACC | TTCGGCTGAAGCATTTGACT | AGAGGTTATGAAGCTCTGTG |
| | | GCGTACAACTCCTCTTCCTA | GTCAAATGCTTCAGCCGAAG | TCCGGACGGGCCCTACCTTA | CTCCGGACGGGCCCTACCTT | TGACTCGGTCTTGGGTGATA |
| **Tankyrase-1** | TNKS | CTTTCGGCTGCAACATGCAG | GACTCCTTTACATCTAGCAG | ATTGTAGAGACTTAGAGGGC | CAGGGGAGAAGACTTCCGGC | CTGTTTCAAAGATATCCCGA |
| | | CTGCAGGTGACCTGCTAGGG | CTCCGCCAAGCTCGATCCAG | GAGCGACATGCATACATAGG | GGTAAAGAGGCTGGTGGACG | CGCGGCAAACGTAAATGCAA |
| **Tankyrase-2** | TNKS2 | GTGCAGCTTGCAGCAACGAG | CACTTCGCCGCAGGTAACCG | AGATCACTACAATGAGAAGG | GTATAACAGAGTGTCCGTGG | GTAAATTGCCGCGATACCCA |
| | | CGTGGAACGAGTCAAGAGGC | ACTTTGAACACCTCTGGTAG | ACCCATAGCTCAGGAGTAGG | TCTCTGACACGTACCCTAGG | ATTTCCTGCCCGCCGTGTCG |
| **Trafficking kinesin-binding protein 1** | TRAK1 | TTGTCCGAATGAGCTTGCCG | TCAACCCCGTAAGTCACCAG | CTGCCGTAGAAGCTGGACCG | TGCCGTAGAAGCTGGACCGG | AGCTCAATAGTGGCATCCGG |
| | | TCGTGTGGCTGTTGAAGGAG | CCTCTGCCAGGACTCTGCCA | GGTGGGCTGCAGCTCAATAG | CACGAGGTTGAGGTTGGAGA | CGCTGGAAGGTGATCACGCG |
| **Trafficking kinesin-binding protein 2** | TRAK2 | TCTTTGCTGAAGAATGGCAG | AGGCTTTACCCAGTTGCCCG | AAGCTTCCAAAGATGCCCAA | TTGACAGAGCCCAACAGAAG | CGTTCTGCTCAGATAAGACA |
| | | TCAGGATTGCCAATGACACA | TCATTATCCTGTGGAAGTAG | TTGGACAAGCTCTCTTAAAG | CGGTAGCAGTTCATCCAACA | GCAGTTTATTAGGTGGACTA |
| **Thyroid receptor-interacting protein 11** | TRIP11 | GAAGGTAACAGCTGGAGCGA | TTTAGGGTCAGCCACAGGTG | TAAACGTTGAACCTAGGTAG | TTGTTTCATGACGATCAGGG | GCAACTGCAGGCTTATGCTA |
| | | GCCGAGCGAGGAGCCAAAAA | GTCACGTAAACGTTGAACCT | GCGATGTCGTCCTGGCTTGG | TAGCTGCTTGATTTCTGAAT | ATGAAGTGTTACGGTTAATG |
| **Tumor suppressor candidate 3** | TUSC3 | ATTACCTATAAGGATAGCCG | AAGGCGATGTTGGAAAAAGA | GCAGATGCCGCTATCACCAT | TCCTTCACGCCGTAGGCAAG | GAGACACTGCCCTGCCGCGA |
| | | TCCAACAAGCGACACTAACA | AAAGTGGATTGCTGACAGAA | AACCATCCCCATGGTGATAG | CCTTCACGCCGTAGGCAAGC | CTCCAACAAGCGACACTAAC |
| **Ubiquitin-conjugating enzyme E2 G2** | UBE2G2 | CATGGGCTACGAGAGCAGCG | ACGTGGATGCGTCCAAAATG | CGCGCTGCTCTCGTAGCCCA | GCTGTCGGTGGTGAGCATGC | TGACGAAAGTGGAGCTAACG |
| | | ACCCTTGCAGTCTACCCTGA | TCCCATCAGGGTAGACTGCA | CGCTGCTCTCGTAGCCCATG | CCGGTCATCGCGCCACATTT | TCATCTTTGGGGGACTTAAC |
| **Ubiquitin-conjugating enzyme E2 J1** | UBE2J1 | ATGGAGGAGTTTATCACGGG | TAACACTTACTACTCCACGA | CCATCAAAATCGGAGTCTGG | GTTGTAGCGGGTCTCCATGG | TTGATGGAGGAGTTTATCAC |
| | | CGTTTGCCAGATATATTCGT | TTTGATGGAGGAGTTTATCA | AATGGCACTTCACGGTTAGA | AGGTATTCTTAGCTACAGGT | TTCTTAGCTACAGGTTGGGT |
| **UDP-glucose:glycoprotein glucosyltransferase 1** | UGGT1 | GTAGTGGGTCAACAACCTGG | TTGTAGCTGAGTAAGAACGA | TTCAGTACAAATGGCCCCGG | GTCGCTTGCCCTGAAAGAAG | GTAGAGATGACCAGATGCAA |
| | | TTTCAGCTGAAAGTAGCCCT | GACAGCCGAAGAGAAATGGA | GTGATGACGCCTCTAAGAAA | AATGAAGCTCGGGTAATGGA | TTGTTCAGTACAAATGGCCC |
| **UDP-glucose:glycoprotein glucosyltransferase 2** | UGGT2 | GGCAGTCACCGACTTGGACG | AATGGAATCATAAACTGGAG | CTTCTGGAGTAATCATGTTG | GATATCCATCCATTTCCCTG | CGAAAGCCACGAACGTGGTG |
| | | CTTTCGCTGGCGCCATGGCA | CTGCCCACTTGGCCGCGAAG | GTTCTGGTTCCAGAACAAAA | GCCAGCGAAAGCCACGAACG | TCATATCGAGATGCACGCTT |
| **Uromodulin** | UMOD | GTTGTAGACGTAGTAGCCGC | CCTGACCATTGGCTGTAGGG | CTACTTGTGCGTATGCCCCG | ACTGAGGCTTTTCTCTACGT | TGTCATTGAAGCCCGAGCAC |
| | | ACGCATCCGTCCAGCGACGA | TCTGGACGGAAAATCGGCCC | CGAGGGCATCGTGAGCCGCA | TAGAGACTCAACTATCCAAG | TCATGTACCTGAGTGACAGC |
| **Transitional endoplasmic reticulum ATPase** | VCP | TGATAGGCTCCCCTTCGCAG | TCACCTTTTGAACTAGAAGG | GTGCTCCACAGGATACTAGG | GCGCAGGTTAGCCTTGAGGA | CCGTACTTCACATCAGGGCA |
| | | ACTGCGGGAAACCGTGGTAG | CTGTGTTTACCAACTATAGG | CCACAGCACGCATCCCACCA | GTTTGCCGTACTTCACATCA | GACACAGTGATCCACTGCGA |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Vascular endothelial growth factor B** | VEGFB | TCAGGTGCCGGAAGCTGCGA | AACGGAGGAAGCTGCGGCGT | CTAAGCCCCGCCCTTGGCAA | TGCTCGGGTACCGGATCATG | CCGGTACCCGAGCAGTCAGC |
| | | TGACACCACTGGGAGCAGAC | ATCTCCCCCAGCTGACTGCT | ACCTGCCCAGTACCTGCATC | CTTGGCAACGGAGGAAGCTG | CACTGGGAGCAGACCGGTAG |
| **Putative polypeptide N-acetylgalactosaminyltransferase-like protein 3** | WBSCR17 | CCTGGCATGGATGTATACGG | GCTCATGACCTACCAGCCAG | TGGTGGACAACTCCAAGAGT | TGAATGGTAAGGACGCACGC | TAATACCGTTGCTTACGGGG |
| | | CAACAGCGACGAAGGTACAG | TGGAGTTGTCCACCAGGCAG | ATGTCCACAAACGCTACCCC | AATTACCTCCCCGTAAGCAA | GGTACGAGAACTCGGCCCA |
| **Xyloside xylosyltransferase 1** | XXYLT1 | TCTGATGCATGGCGACCGAG | GGCCACGTCAAGATCTACCA | GCCGACAAGTACCACTTCCG | CCGTCAGCACAGCAACATCG | TCCAGCAGGCGGCTGTAGAG |
| | | GCCACGTCAAGATCTACCAC | CGAAGACGTCACTGTAGCCA | CAAATTCCTCAAACAACTCC | TGCTGGACTGTACCTGGAAC | CGCCATCATCGGCATAGCCC |
| **Alpha-1,4-N-acetylglucosaminyltransferase** | A4GNT | GGAGGCGCTACTATGAAGTG | ACGGAACAGGAGACCAAATG | GAGGCGCTACTATGAAGTGT | TATGTGGAGTTTGAGGGCAT | GGGCCTGATGGAGATGACAT |
| | | GCCACCGTATTTCCAGATGA | ACCCCATCTCCTATCGAGAG | TCCACTCTCGATAGGAGATG | GTAGCGCCTCCACTCTCGAT | GCTTCTCGGTACTCTAGTAA |