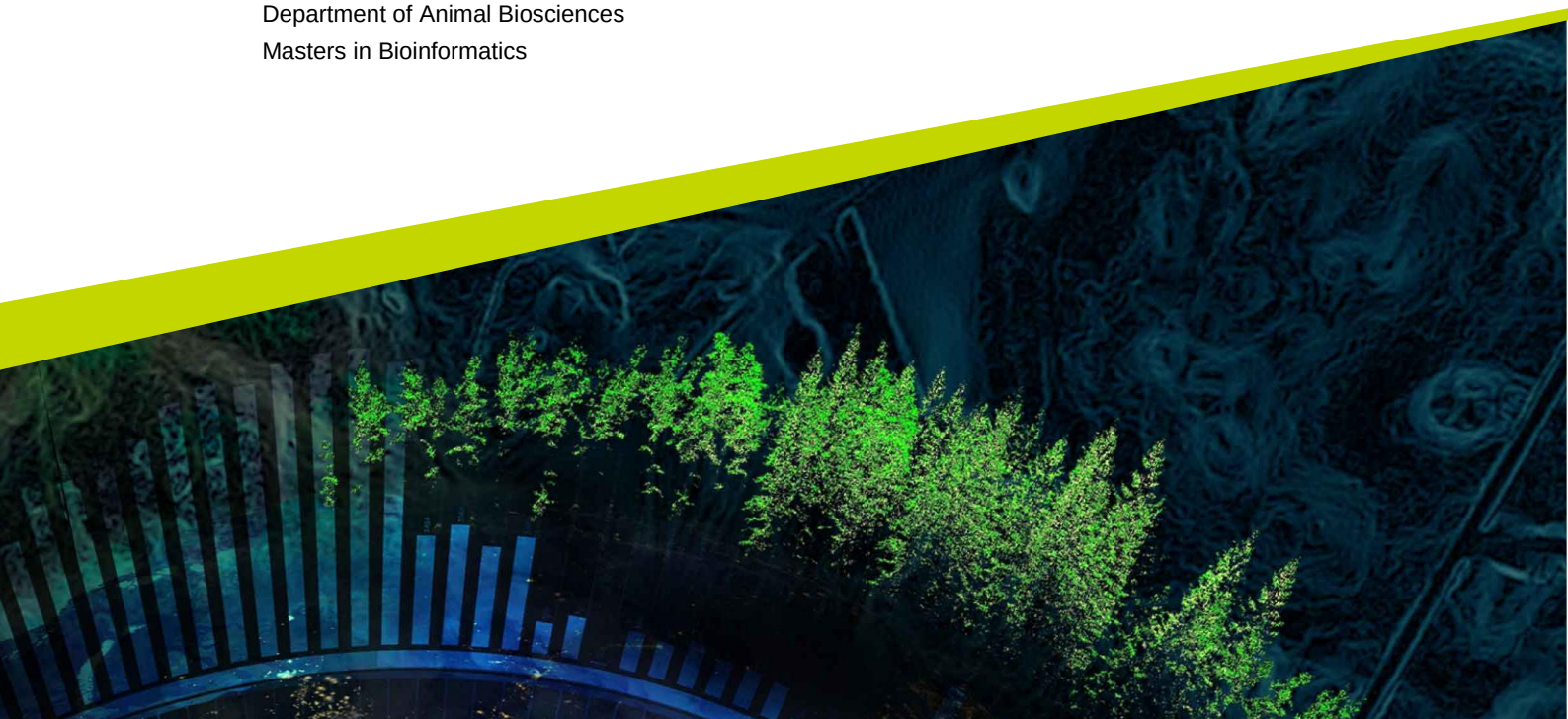
Development of Stable CRISPRi-dCas9-KRAB Cell Lines for Gene Suppression in Canine Cancer Models:

A Functional Study of SRGN

Ruwini Madhushika Herath Herath Mudiyanselage

Degree project/Independent project • 30 credits
Swedish University of Agricultural Sciences, SLU
Faculty of Veterinary Medicine and Animal Science.
Department of Animal Biosciences
Masters in Bioinformatics



Development of Stable CRISPRi-dCas9-KRAB Cell Lines for Gene Suppression in Canine Cancer Models: A Functional Study of SRGN

Ruwini Madhushika Herath Herath Mudiyanselage

Supervisor:	**Magnus Åbrink, Swedish University of Agricultural Sciences, Department of Animal Biosciences**
Assistant supervisor:	SofiaTengstrand, Swedish University of Agricultural Sciences, Department of Animal Biosciences
Assistant supervisor:	Amanda Liljenström, Swedish university of Agriculture, Department of Animal Biosciences
Examiner:	Göran Andersson, Swedish University of Agricultural Sciences, Department of Animal Biosciences

Credits:	30 credits
Level:	A2E
Course title:	Independent project in Bioinformatics
Course code:	EX1002
Programme/education:	Masters in Bioinformatics
Course coordinating dept:	Department of Animal Biosciences
Place of publication:	Uppsala
Year of publication:	2025

Keywords:	CRISPRi, CRISPR interference, dCas9-KRAB, Comparative oncology, Canine Mammary tumour (CMT)

Swedish University of Agricultural Sciences
Faculty of Veterinary Medicine and Animal Science.
Department of Animal Biosciences.

Abstract

Background: The discipline of comparative oncology has advanced by recognising the importance of pet dog spontaneous tumours as models for human cancer. Canine mammary tumours (CMTs) are pathologically and molecularly similar to humans, with conserved oncogenic mechanisms.

Objective: This study aimed to analyse publicly available human breast cancer data sets and identify up and downregulated genes, identify differently expressed gene in a mouse mammary tumour serglycin (SRGN)-deficient environment, and to design and verify functioning stable CRISPR interference (CRISPRi) systems in the CMT-U27 cell line, using two different plasmids containing dCas9-KRAB constructs, as the KRAB domain is a transcriptional repressor known to downregulate wide range of genes, including transposable elements (TEs) and protein-coding genes through epigenetic silencing. Then evaluate the interference of expression of SRGN, which is known to mediate inflammation, remodelling of the extracellular matrix, and tumour progression.

Methods: A public transcriptomic dataset from mouse mammary tumours and four datasets from human breast cancer samples were examined to study the expression profile of SRGN in the context of TP53-regulated genes. Experimentally, CMT-U27 cells were transfected with dCas9-KRAB constructs and G418 was used for selection. Subsequently, in the resulting G418-resistant cells qPCR verification was done to confirm construct expression and, after addition of guide RNAs targeting exon 1 and exon 2, evaluating SRGN repression.

Results: Bioinformatics analyses uncovered that SRGN expression is variable in human datasets of breast cancer. In a mouse knockout model, SRGN deficiency resulted in an extreme downward spiral of SRGN and inflammatory extracellular matrix remodelling associated genes. CMT-U27 cells stably transfected with CRISPRi constructs were shown to integrate the plasmid and exhibited variable SRGN repression. Most significant downregulation was seen in clones CRISPI-G and CRISPI-G-2 with 0.70 and 0.39 fold changes, respectively. Non-targeting controls like CRISPI C showed SRGN elevation, which may be explained by clonal heterogeneity after G418 selection.

Conclusion: The system for CRISPRi-dCas9-KRAB developed in this study should allow for efficient and precise transcriptional control of genes of interest (GOI) in canine mammary cancer cells, extending beyond SRGN for future gene suppression investigations. This system allows controlled exploration of manipulated genes, desired therapeutic targets, and tumour biology in comparative oncology.

Table of Contents

List of tables

List of figures

Abbreviations

Abbreviation	Description
A2M	Alpha-2-Macroglobulin
ABCA8	ATP Binding Cassette Subfamily A Member 8
ABCA10	ATP Binding Cassette Subfamily A Member 10
ADH1A	Alcohol Dehydrogenase 1A
ADH1B	Alcohol Dehydrogenase 1B
ADH1C	Alcohol Dehydrogenase 1C
ADIPOQ	Adiponectin, C1Q and Collagen Domain Containing
AKR1C2	Aldo-Keto Reductase Family 1 Member C2
AQP7P2	Aquaporin 7 Pseudogene 2
ATP4A	ATPase H+/K+ Transporting Alpha Subunit
BLNK	B-cell Linker Protein
BPIFA2	BPI Fold Containing Family A Member 2
BUB1	Budding Uninhibited by Benzimidazoles 1
C2orf40	Chromosome 2 Open Reading Frame 40 (ECRG4)
CCL2	C-C motif ligand 2 (CCL2), also known as monocytic chemotactic protein 1 (MCP-1)
CCNB2	Cyclin B2
CCNE2	Cyclin E2
CDC20	Cell Division Cycle 20
CDK1	Cyclin Dependent Kinase 1
cDNA	complementary DNA
CEACAM6	Carcinoembryonic Antigen-Related Cell Adhesion Molecule 6
CHRD-L1	Chordin-Like 1
CIDEC	Cell Death-Inducing DFFA-Like Effector C
CKAP2L	Cytoskeleton Associated Protein 2 Like
CMV	Cytomegalovirus (promoter)
CMT	Canine Mammary Tumour
CO_2	Carbon Dioxide
COL10A1	Collagen Type X Alpha 1 Chain
COL11A1	Collagen Type XI Alpha 1 Chain
COM	Catechol-O-methyltransferase
COMP	Cartilage Oligomeric Matrix Protein
CREB1	cAMP Response Element-Binding Protein 1
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRISPRi	CRISPR interference
CST1	Cystatin SN
CXCL10	C-X-C Motif Chemokine Ligand 10
CXCL9	C-X-C Motif Chemokine Ligand 9
CYP4B1	Cytochrome P450 Family 4 Subfamily B Member 1
Ct	Cycle threshold
dCas9	deactivated Clustered Regularly Interspaced Short Palindromic Repeats-associated protein 9
DCIS	Ductal Carcinoma In Situ
DCPP2	Demilune Cell and Parotid Protein 2
DEFB132	Defensin Beta 132
DEG	Differentially Expressed Gene
DMSO	Dimethyl Sulfoxide
DNA	Dioxyribonucleic acid

DSBs	Double Strand Breaks
DSCAM-AS1	DSCAM Antisense RNA 1
ECM	Extracellular Matrix
EGFR	Epidermal Growth Factor
EGFRP2	Epidermal Growth Factor Receptor Pseudogene 2
EIF3D	Eukaryotic Translation Initiation Factor 3 Subunit D
EMT	Epithelial-Mesenchymal Transition
ER	Estrogen Receptor
ESRP1	Epithelial Splicing Regulatory Protein 1
EZH2	Enhancer of Zeste Homolog 2
FAK	Focal adhesion kinase
FABP4	Fatty Acid Binding Protein 4
FBS	Fetal Bovine Serum
FC	Fold Change
FDA	Food and Drug Administration
FGF	Fibroblast Growth Factor
FOXA1	Forkhead Box A1, also known as Hepatocyte nuclear factor 3-alpha
GALNT6	Polypeptide N-Acetylgalactosaminyltransferase 6
GC-RMA	Gene Chip Robust Multi-array Average
GEO	Gene Expression Omnibus
GJB2	Gap Junction Protein Beta 2
GLYAT	Glycine-N-Acyltransferase
GO	Gene Ontology
GPD1	Glycerol-3-Phosphate Dehydrogenase 1
GPIHBP1	Glycosylphosphatidylinositol Anchored High Density Lipoprotein Binding Protein 1
GWAS	Genome-Wide Association Studies
H3K9me3	Trimethylation of Lysine 9 on Histone H3
HA	Hemagglutinin (tag)
HBC	Human Breast Cancer
HBD	Hemoglobin Subunit Delta
HBG2	Hemoglobin Subunit Gamma 2
HBM	Hemoglobin Subunit Mu
HEK293	Human embryonic kidney cell line
HET	Heterozygous
HIST1H2BB	Histone Cluster 1 H2B Family Member B
HLA-DRA	Major Histocompatibility Complex, Class II, DR Alpha
HP1β	Heterochromatin Protein 1 Beta
IDS	Invasive Ductal Carcinoma
IFI44L	Interferon Induced Protein 44 Like
IL6	Interleukin-6
KAP1	KRAB-Associated Protein 1
KIAA0101	Gene encoding PAF15 (PCNA-associated factor)
KIF11	Kinesin Family Member 11
KLB	Klotho Beta
KLF4	Krüppel- Like Factor 4
KLK1	Kallikrein-1
KO	Knockout
KRAB	Krüppel-associated box
KRAB-ZFP	Krüppel-Associated Box Zinc Finger Proteins
LIMMA	Linear Models for Microarray Data
MED10	Mediator Complex Subunit 10
MED14	Mediator Complex Subunit 14
MEM Na Pyruvate	Minimum Essential Medium with Sodium Pyruvate

MME	Membrane Metallo-Endopeptidase
MMP11	Matrix Metallopeptidase 11
MX1	MX Dynamin Like GTPase 1
NF-κB	Nuclear Factor Kappa-Light-Chain-Enhancer of Activated B Cells
NT	Normal Tissue
NSAIDs	Non-Steroidal Anti-Inflammatory Drugs
OAS1	2'-5'-Oligoadenylate Synthetase 1
OLFML2A	Olfactomedin-Like 2A
PBK	PDZ Binding Kinase
PCK1	Phosphoenolpyruvate Carboxykinase 1
PCR	Polymerase Chain Reaction
PIK3CA	phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha
PEST	Penicillin, Streptomycin, and Amphotericin B
POLR2D	RNA polymerase II subunit D
PPAPDC1A	Phosphatidic Acid Phosphatase Type 2 Domain Containing 1A
PPBP	Pro-Platelet Basic Protein (also called CXCL7)
PROL1	Proline Rich, Lacrimal 1
qPCR	quantitative Polymerase Chain Reaction
RAB25	RAB25, Member RAS Oncogene Family
RAM	Random-Access Memory
RMA	Robust Multi-array Average
RNA	Ribonucleic Acid
RNA Pol II	RNA Polymerase II
RNA-seq	RNA sequencing
RPMI 1640	Roswell Park Memorial Institute Medium 1640
RPS14	Ribosomal Protein S14
S100A7	S100 Calcium Binding Protein A7 (Psoriasin)
S100A12	S100 Calcium Binding Protein A12
S100P	S100 Calcium Binding Protein P
SBSN	Suprabasin
SCARA5	Scavenger Receptor Class A Member 5
SCGB3A1	Secretoglobin Family 3A Member 1
SDC1	Syndecan-1
SGCG	Sarcoglycan Gamma
SLC19A3	Solute Carrier Family 19 Member 3
SLU	Swedish University of Agricultural Sciences
SPC25	Spindle Component 25
SPTI	Serine Protease Inhibitor Type I
SRGN	Serglycin
SVA	SINE-VNTR-Alu retrotransposon
SYNPO4	Synaptopodin 4
TEs	Transposable elements
TGF-β2	Transforming Growth Factor Beta 2
TNF-α	Tumor Necrosis Factor alpha
TOP2A	DNA Topoisomerase II Alpha
TPX2	TPX2 Microtubule Nucleation Factor
TSS	Transcription Start Site
USA	United State America
ZNF91	Zinc Finger Protein 91
ZNF93	Zinc Finger Protein 93

Experimental samples-Abbrivations

CRISPI-C	CMT-U27 cells transfected with Addgene plasmid #73499 (no gRNA)
CRISPI-G	CMT-U27 cells transfected with Addgene plasmid #73499 and gRNA for SRGN exon 1
CRISPI-G-2	CMT-U27 cells transfected with Addgene plasmid #73499 and gRNA for SRGN exon 1
KRAB-C	CMT-U27 cells transfected with Addgene plasmid #112195 (no gRNA)
KRAB-G	CMT-U27 cells transfected with Addgene plasmid #112195 and gRNA for SRGN exon 1
KRAB-G-2	CMT-U27 cells transfected with Addgene plasmid #112195 and gRNA for SRGN exon 2

1. Introduction

According to WHO breast cancer caused about 670 000 deaths globally from 2.3 million women who were diagnosed in 2022. Understanding the complex genetic mechanisms is necessary to battle this situation. Bioinformatics analysis of transcriptomic data of human breast cancer would help to understand up- and down-regulated and co-regulated genes, which may shed light on further studies.

As it is unethical to subject human patients to the early stages of clinical trials, naturally occurring cancers in other species, such as dogs and cats, can be used as a tool to understand human tumour pathology. This gives more accurate results than laboratory mouse models, in which tumours are created artificially by genetic modifications or by injection of tumour cell lines (Sultan & Ganaie, 2018). Canines and humans are more genetically related than to mouse, and thereby canine cancers show close resemblance in histopathologically and molecular features to their human counterparts (Mestrinho & Santos, 2021). Above similarities allow hereditary risk studies of cancer development as well as gene interactions (GxE) and mutations existing in both species (Oh & Cho, 2023).

According to Mestrinho & Santos, (2021) and Oh & Cho, (2023) dog is an ideal organism for identifying therapeutic targets, and to evaluate in clinical trials and gene therapies, due to their genetic similarity to humans. Presenting similar counterpart in cancer types such as sarcomas, melanomas, etc. increase the usability of canine models for preclinical studies (Schiffman & Breen, 2015).

Canine mammary tumour (CMT) which shows a high prevalence in older female dogs is histologically and molecularly very similar to human breast cancer (HBC) (Oliveira-Lopes et al., 2024; Kim et al., 2020). Prognosis factors not limited to such as lymphatic metastasis (Abdelmegeed & Mohammed, 2018; Queiroga & Lopes, 2011) but also molecular pathways such as PI3K-Akt signaling pathway and the steroidogenic PIK3CA gene, along with alterations in the receptor EGFR and expression of mutated p53 genes (Kim et al., 2020) shows remarkable similarities. Due to this similarity studying CMT can shed insight into HBC and may assist in determining therapeutic intervention.

Cancer cell lines derived from both human and dog are robust experimental tools for studying the effects of individual genes on cancer development and treatment responses. Traditional knockout models which involve homologous recombination or CRISPR/Cas9 that induce double-strand breaks (DSBs) (Jinek *et al.*, 2012) can be used for above studies, but these models may cause irreversible genetic damages, which may mask the findings. However, the CRISPR interference (CRISPRi) system, that avoids DNA double-strand breaks, can provide a more precise and non-invasive approach for targeted gene silencing (Qi et al., 2013; Larson et al., 2013).

CRISPRi uses guide RNAs and a deactivated dCas9 protein to bind and prevent the transcription initiation or elongation, and the methodology has successfully been used to silence genes across several species, like bacteria, vertebrates such as zebrafish, and *Caenorhabditis elegans* (Larson et al., 2013; Qi et al., 2013; Long et al., 2015; Zhang et al., 2021). Furthermore, as reported by Goetzl & Alpert (2024), one of the latest advancements includes FDA approval of CASGEVY™ for sickle cell disease treatment, which has proven to increase foetal haemoglobin production. Together, this illustrates that the technique is easily applicable across different species without the uncontrolled effects and risks associated with permanent gene knockouts. CRISPRi capacity to simultaneously silence multiple genes enable increased flexibility in functional genomics and synthetic biology research, including cancer research (Zhang et al., 2021).

Serglycin (SRGN) is a proteoglycan involved in inflammation by storing and releasing cytokines and chemokines within the extracellular matrix (Kolset & Pejler, 2011) and it is expressed in hematopoietic and some cancer cells and acts as a scaffold for pro-inflammatory mediators, including TNF-α (Tumor Necrosis Factor alpha), IL-6 (Interleukin-6), and CCL2 (monocyte chemoattractant protein-1). Furthermore, SRGN has been associated with metastasis of some cancers, including, colorectal, lung, and triple negative breast cancer. SRGN-dependent metastasis seems to occur via CD44-dependent pathways via biochemical interaction & downstream signalling. When SRGN is highly glycosylated, it promotes binding CD44 which is a cell surface receptor in cancer cells and activates Src signalling. It leads to paxillin phosphorylation and disassembly of the FAK/paxillin complex, which accelerates focal adhesion turnover (Guo et al., 2017) Metastasis also could facilitate through TGF-β2 and CREB1 signalling interactions (REFs). SRGN also plays a role in immune evasion through the CD47–Sirpα pathway (Guo et al., 2017; Zhang et al., 2017; Xu et al., 2018; Chao et al., 2012). However, considering its significance in human oncology, SRGN has been relatively understudied in veterinary models, particularly in spontaneously occurring cancers.

The first aim of this thesis was to identify differently expressed genes in human breast cancers and evaluate how expression patterns change in a SRGN-deficient environment. The second aim was to establish two novel CRISPRi-expressing canine mammary tumour cell line (CMT-U27), and to conduct a functional repression study using guide RNAs targeting exon 1 or 2 of the SRGN gene.

2 Literature Review

2.1 Data sets used for bioinformatics analysis of SRGN expression

2.1.1 GSE29044

The GSE29044 dataset consist of 73 tumour samples and 36 normal tissue samples with age stratified metadata. The dataset was produced using the **Affymetrix** Human Genome U133 Plus 2.0 arrays and was normalised using the GC-RMA method (Gene Chip Robust Multi-array Average method combines multiple probe signals targeting same gene into a single expression value using a statistical model. It provided expression profiles of ductal carcinoma *in situ* (DCIS) and invasive ductal carcinoma (IDC). GSE29044 serves as an important resource for discovering age-associated transcriptional signatures of tumour invasiveness (Colak et al., 2013).

2.1.2 GSE42568

The GSE42568 dataset contains age-associated, ranging from 31 to 89 years, transcriptional profiles of 104 primary breast cancer biopsies alongside 17 normal breast tissue samples. Tumour characteristics included histological subtypes (ductal, lobular, and special types), grades (1–3), tumour sizes (T1–T3), along with lymph node involvement and estrogen receptor (ER) status. The dataset also contains clinical metadata with long-term follow-up data, some extending to 3026 days, making GSE42568 ideal for co-expression network analyses and survival transcriptional profiling. The expression data obtained from the **Affymetrix** Human Genome U133 Plus 2.0 arrays and normalized using GC robust multichip average method and using the ‘affy’ Bioconductor package. This dataset highlights the significance of transcriptional signatures and breast cancer outcomes (Clarke et al., 2013).

2.1.3 GSE89116

The GSE89116 dataset examines expression changes on a genome scale for early-onset (≤40 years) and late-onset (≥55 years) breast cancers to look for biologically relevant differences in transcription that occur with age. Data was produced using the Illumina HumanWG-6 and HumanHT-12 beadchip platforms and data normalization was conducted using the LIMMA package

in R. The two age group-tumour and adjacent normal tissues (as controls) provided the opportunity to study genes that are regulated in a patient age-dependent manner. The dataset enhances understanding of age-stratified oncogenic signatures and tumour onset among Indian women (Malvia et al., 2019).

2.1.4 GSE109169

The GSE109169 dataset produced on **Affymetrix** Human Exon 1.0 ST array and it gives gene expression on 25 paired tumour and adjoining normal breast tissue samples in early-onset (≤40 years) and late-onset breast cancer, specifically analysing genes related to metastasis and p53 signalling pathways (Chang et al., 2018).

2.1.5 GSE67806

This data set contains gene expression patterns from tumour samples of 6 mouse of the MMTV-PyMT tumour model with 3 heterozygous (SG+/−) and 3 serglycin knockout (SG−/−) tumour tissues. The data was produced on the Affymetrix Mouse Gene 2.0 ST array and normalised using the robust multi-array average (RMA) method. Notably, in the SRGN-deficient tumours, there was a striking repression of 666 genes and only six upregulated genes. The massive reduction of gene expression seem to confirm the previously suggested role of SRGN in moderating metastasis, inflammation, epithelial-mesenchymal transition (EMT), and extravasation. The dataset supports the idea that SRGN and its downstream effectors could be potential therapeutic targets in metastatic canine mammary tumours and human breast cancer. (Roy et.al., 2016)

2.2 CRISPRi and dCas9-based Gene Silencing

The use of Catalytically Inactive Cas9 (dCas9) and a single guide RNA (sgRNA) is referred to as CRISPR Interference (CRISPRi), and it is a powerful tool for gene suppression (Figure 1). According to Ghavami and Panddi (2021), CRISPRi is less risky and more accurate than “knockouts” as it does not produce DNA double-strand brakes. In addition to bacteria and human cells, CRISPRi has been applied in *C. elegans* and zebrafish, proving its versatility (Larson et al., 2013).

Thus, the transient CRISPRi technology provides a different use as compared to permanent gene knockout models and can avoid ambiguous and irreversible gene silencing (Qi et al., 2013). In synthetic biology and gene network studies, it is particularly convenient to have one gene simultaneously repressed by different elements, which is made possible by multiplex CRISPRi systems (Zalatan et al., 2015). The use of CRISPRi in cancer biology allows the exploration of essential genes, regulatory pathways, and therapeutic interventions (Gilbert et al.,2013).

Figure 1: CRISPRi based suppression. (Source: Li, Z., Xiong, X. and Li, J.-F., 2019. The working dead: repurposing inactive CRISPR-associated nucleases as programmable transcriptional regulators in plants. aBIOTECH. https://doi.org/10.1007/s42994-019-00003-z)

2.3 Role of KRAB Domains in Epigenetic Repression

Many zinc finger proteins involved in DNA binding contain potent transcriptional repressors called KRAB (Krüppel-associated box) domains (Margolin et al., 1994). The co-repressor KAP1, which

causes long-range transcriptional silencing through heterochromatin spreading, is recruited by KRAB-ZFPs (Lehner et al., 2013; Groner et al., 2010). Reduced histone acetylation and RNA Pol II recruitment are the outcomes of this repression mechanism, which involves the propagation of H3K9me3 and HP1β marks from the KRAB binding site to the target promoter. Interestingly, KRAB-ZFPs have co-evolved with a family of transposable elements (TEs) and may contribute to their suppression. According to recent data, TEs and KRAB-ZFPs work together to create species and cell-specific regulatory networks, influencing various aspects of development and physiology (Ecco et al., 2017).

In human genome over 400 of KRAB-ZFPs are precent (Imbeault *et al.*, 2017) and they regulate gene slicing during embryonic development, differentiation and homeostasis. One example is ZNF91 and ZNF93, which silence retrotransposons like SVA and LINE-1 elements in human pluripotent stem cells and protect genomic integrity. When it comes to gene editing, KRAB domain has fused to dCas9 and utilize it natural epigenetic suppression ability in CRISPR interference. As KRAB domain can used to repress TEs and any protein coding gene CRISPRi technology can be used to suppress many protein coding gene (Gilbert et al., 2013). Targeted genes can be varying such as oncogenes, cytokines or tumour suppressors, depending on the guide RNA used.

2.4 Supercoiled vs. Linearized Plasmids in Transfection

Previous studies have shown that the conformation of the plasmid affects their transfection efficiency. Supercoiled plasmids perform better than linear forms with delivery via lipids, chitosan microspheres, or electroporation (Lehner et al., 2013, Akbuğa et al., 2004; Chancham & Hughes, 2001; Escoffre et al., 2012). The higher efficiency of supercoiled DNA toward transfection is attributable to its round shape, which makes it easier for the cells to take up and become stable in the cell fluid (Lehner et al., 2013; Chancham & Hughes, 2001). Some research suggests that the plasmid size does not influence gene delivery efficiency (Akbuğa et al., 2004). Instead, as pointed out by Lehner et al. (2013) the arrangement of DNA transfection complexes plays a crucial role in effective gene transfer.

2.5 Lipofectamine Transfection.

Lipofection is considered as a gold standard for the safe delivery of exogenous DNA or RNA into cells (Cardarelli et al., 2016). Lipofectamine is a cationic liposome-based transfection reagent that can be used to deliver plasmid DNA into mammalian cells. It facilitates cytoplasmic entry of DNA. Nuclear membrane penetration is facilitated by combining Lipofectamine with nuclear targeting peptides like M9 (Byrnes et al., 2002).

2.6 Canine cancers

Cancer in dogs poses a significant and increasing health threat, with about 6 million new diagnoses each year in the US (Printz, 2011). This high number indicates the necessity of understanding what causes canine cancer and improving how to detect and treat the cancer. Some dog breeds have a higher probability of getting certain types of cancer, which shows that the genetic makeup play a key part in cancer risk (Dobson, 2013). For instance, breeds like Golden Retrievers and Boxers are likelier to get specific cancers, such as lymphoma and hemangiosarcoma, indicating that inherited traits are strongly linked to cancer development. Furthermore, hormones affect tumour development, *e.g.* female dogs develop mammary gland cancer much more often than male dogs according to the Animal Tumour Registry of Genoa (Merlo et al., 2008). This shows that gender can affect cancer risks, *i.e.* breast cancer is more common in women than in men.

In addition, comparing various cancer diseases across species has revealed striking similarities between dog and human cancers, with shared genetic changes and similar ways of cellular communication in both species (Schiffman & Breen, 2015). This suggests that comparative studies could lead to better treatments for both humans and dogs. Here, research efforts like the Canine

Comparative Oncology and Genomics Consortium, which aim to facilitate large-scale studies by gathering dog tumour samples to examine their genomic profiles, play a key role in spotting genetic factors that drive cancer and creating targeted treatments (Printz, 2011). Furthermore, purebred dogs exhibit less genetic variance, which makes them ideal for genome-wide association studies (GWAS), enabling scientists to identify the complex network of genetic factors that influence cancer development. Thus, comparative oncology has excellent potential to improve our understanding of cancer.

2.7 Canine mammary tumour

Mammary tumours in dogs (CMTs) are highly prevalent among female dogs. They occur in 25-50% of females who are not spayed, with increased risk in the aged dog, particularly after six years of age (Queiroga & Lopes, 2002; Ferreira et al., 2024). Again hormones play a role, since spaying before the first heat reduce the risk of developing mammary cancer (Queiroga & Lopes, 2002). In almost half of the cases the mammary cancers are malignant, and surgery is the primary treatment for most of the tumours, with exclusion for inflammatory carcinomas or when the cancer is advanced (Queiroga & Lopes, 2002; Sleeckx et al., 2011; Benavente et al., 2016).

2.8 CMT-U27 as a Model for Breast Cancer

CMTs represent a valuable translational model for human breast cancer research because they share similarities in histopathological features, clinical presentations, and molecular signatures (Graim et al., 2020; Oliveira-Lopes et al., 2024). CMTs also mirror human breast cancer subtypes sharing an analogous genomic landscape, microRNA expression changes, and metabolomic profile changes (Gherman et al., 2024). The canine model presents distinct benefits, such as having multiple spontaneously arising tumours in a single subject to provide the opportunity for investigation of tumour development from normal tissue to the cancerous state (Graim et al., 2020). This aspect allows the researcher to define transcriptional signatures and biological pathways specific to malignant tumours that may serve as prognostic markers in patients with breast cancer (Graim et al., 2020). Besides, dogs' shorter lifespan and higher incidence of cancer development allow for a faster assessment of novel therapeutic strategies compared to human studies (Oliveira-Lopes et al., 2024).

The CMT-U27 canine mammary carcinoma cell line has been extensively studied for its potential in cancer research. This cell line exhibits high growth rates, anti-apoptotic potential, and shorter cell cycles compared to other canine mammary tumour lines (Król et al., 2009). CMT-U27 cells are sensitive to antiprogestins, which reduce cell viability and may affect progesterone receptor expression (Guil-Luna et al., 2014). The CMT-U27 cells also shows responsiveness to nonsteroidal anti-inflammatory drugs (NSAIDs), where combined treatments with piroxicam and deracoxib cause significant cytotoxic effects and induction of apoptosis (Alkan et al., 2012)

2.9 Serglycin (SRGN) and its role in Tumour Biology

The serglycin (SRGN) proteoglycan plays an essential part in tumour biology. For example, in non-small cell lung cancer high SRGN expression levels increase binding to CD44 on the cancer cell surface. CD44 is a transmembrane glycoprotein cell surface receptor for hyaluronic acid and other ECM components (Guo et al., 2017). This process enhances cell migration, invasion and metastasis, *i.e.* tumour aggressiveness, leading to worsened patient outcomes (Korpetinou et al., 2014; Guo et al., 2017; Guo et al., 2020). In hematologic cancers such as multiple myeloma and leukaemia, CD44 helps to homing and adhesion to the bone-marrow niche (Asosingh *et al.*, 2001) immune evasion, and drug resistance (Purushothaman *et al.*, 2014; Guo *et al.*, 2020).

Apart from its pro-tumorigenic role in epithelial cancers, there is also evidence that SRGN is released from tumour-associated macrophages and myeloid-derived suppressor cells, which impact immune suppression and chronic inflammation in colonic and pancreatic tumours (Wang et al., 2019). Furthermore, some studies in breast cancer have shown that SRGN expression is controlled

through the NF-κB signalling pathway, connecting inflammation and cancer malignancy. This data indicates that SRGN may serve as a candidate biomarker and a molecular target for inhibition of malignant tumours.

New bioinformatics studies in both mouse models of mammary tumours and human breast cancer cohorts have now shown aggressive tumour subtypes to consistently overexpress SRGN, indicating a likely evolutionary conserved function that promotes tumour development. In dogs, while SRGN has not been studied in great detail, its role in immune responses and remodelling of the extracellular matrix renders it a compelling candidate for translational cancer research. Therefore, silencing of SRGN in the CMT-U27 cell line could shed light on standard inflammatory and metastatic processes of the mammary gland tissue in dogs and breast cancers in humans.

2.10 Rationale for Experimental Design of the Thesis Work

Due to its biological and molecular features, CMT-U27 cell line can be used as a model for human breast cancer. SRGN was chosen as the gene of interest for functional testing due to its involvement in inflammation, metastasis, and immune system interactions. This thesis work aimed at suppressing SRGN expression in the CMT-U27 cell line using the CRISPRi technology. Cell lines with stably integrated CRISPR components will allow for improved understanding of the role genes of interests (e.g. SRGN) in cancer-related mechanisms, thus advancing the comparative oncology field.

3 Materials and Methods

3.1 Data Analysis

Four publicly accessible gene expression datasets were analysed to identify up- and down-regulated gene, in correlation with SRGN expression, in human breast cancer. The datasets, GSE42568, GSE29044, GSE89116, and GSE109169 were acquired from the Gene Expression Omnibus (GEO) repository. Each dataset contained expression profiles from breast tumour samples and controls (normal tissue or context-specific control groups). The series matrix files were downloaded and separated into sample-wise and group-wise data frames. These data frames served as a baseline for assessing the SRGN regulation throughout tumour contexts as well as for finding differentially expressed genes (DEGs) related to extracellular matrix (ECM) modification, immune regulation, and cancer progression.

The fifth dataset contained mouse mammary cancer SRGN-knockout data, which allowed to understand transcriptomic shifts in an ablated SRGN expression environment. This provided insight into downstream genes regulated by SRGN and served as a comparative model to validate findings from human data. Finally, a sixth qPCR-based SRGN expression dataset from the CRISPR interference (CRISPRi) CMT-U27 canine mammary tumour cells was generated experimentally during this study.

3.1.1 The Computational Environment

All computational analyses were conducted in JupyterLab on a Windows 11 laptop equipped with an Intel Core i7 processor and 16 GB of RAM. The entire workflow was implemented using the Python programming language. Preprocessed expression data was imported from .txt files, which had already been normalised. Python libraries including pandas, numpy were used for data manipulation, while matplotlib and seaborn were employed for data visualization, including boxplots, volcano-style plots, and density plots. Custom scripts were written for parsing, comparing, and interpreting gene-level expression values across different experimental groups. All

code was deposited in the GitHub repository to ensure traceability and reproducibility of the project.

3.1.2 Analysis of Differential Expression

Expression matrices were log-transformed for each dataset, with focus on the genes and two conditions tumour versus control. Statistical significance was assessed using a Welch's t-test. Adjusted p-values (multiple testing) and log2FC values were used in the human datasets to identify the most significantly differentially expressed genes, In the mouse dataset non-adjusted p-values and log2FC were used. The top ten up- and down-regulated genes were selected.

All analyses were performed using **Python (version 3.12.5)** in **Jupyter Lab (4.4.3)**, using the following packages: pandas and numpy for data handling, scipy.stats for Welch's t-test, statsmodels for multiple testing correction.

3.1.3 Visualisation and Output of Results

All visualisations were created using Python (version 3.12.5) in the Jupyter Notebook (4.4.3) environment. Graphical outputs were created using the matplotlib package (REF), seaborn. Volcano plots were made which displayed DEGs. SRGN normalized expression levels were also compared via boxplots for tumour and control groups across datasets, which were generated using ggplot2. Moreover, within-group variation and SRGN expression consistency among biological replicates is shown with sample level expression plots.

3.1.4 Quality Control of experimental data

When selecting data points from the qPCR analysis, melting curves were manually analysed as a quality control measure to remove non-specific amplifications (Figure 2). Samples with multiple melting peaks, broad peaks, or irregular peak shapes, which indicated non-specific amplification, primer dimers, or some form of contamination, were removed from further processing. This step ensured that all quantification was based on reliable and target-specific amplifications with Ct values below 38 (Figure 3).

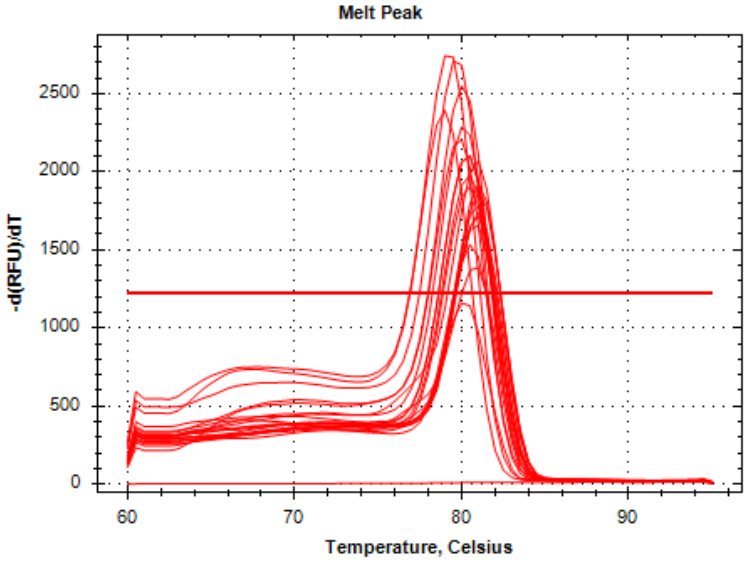


Figure 2: Melting curve peaks for RPS5 and SRGN qPCR products. The graph shows the derivative of fluorescence with respect to temperature, indicating single melting peaks around 80 °C and 81 °C, consistent with specific amplification. (Source: Author's own, generated during qPCR analysis using Bio-Rad CFX Maestro, 2025)

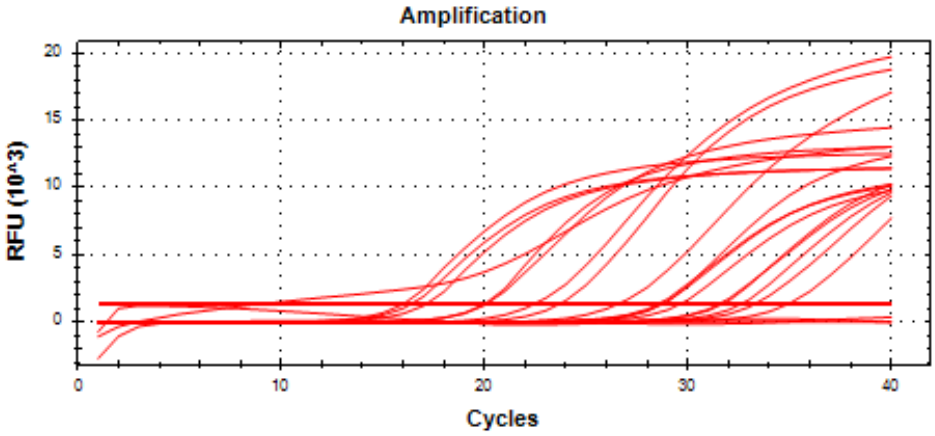


Figure 3; Amplification curve- RPS5 & SRGN. Fluorescence increases indicate successful and specific amplification of the target sequences. (Source: Author's own, generated during qPCR analysis using Bio-Rad CFX Maestro, 2025)

3.1.5 Data Management and Storage

All data management adhered to the Principles of FAIR data (Findable, Accessible, Interoperable, Reusable) for processed and raw data files. Data was stored on **personal encrypted storage** to ensure security and backup. Output tables, figures, intermediate data frames, and final DEG lists were archived for reproducibility and transparency in reporting.

3.2 Cell culture, Plasmid Preparation and CRISPRi Transfection

3.2.1 Cell Lines and Culture Conditions

CMT-U27 cells in freezing media (DMSO 10% + fetal bovine serum (FBS) 90%) were thawed, resuspended in 10 ml cell culture media (RPMI 1640 GlutaMAX™ medium +10% FBS +1% penicillin-streptomycin (PEST), +1% MEM Na pyruvate). Cells were centrifuged at 1100 rpm for 5 minutes, and the supernatant containing the diluted freezing media was discarded. Cells were resuspended in 5 mL of cell culture media and cultured in T-25 flasks at 37 °C and 5% CO_2. In every 2-3 days, culture media was changed, and cells were passaged at approximately 70-80% confluency. 1.5X trypsin was used for cell detachment. To achieve uniform cell seeding densities across experiments, an automated cell counter (VWR, Fluo) was used.

3.2.2 Selection of Plasmid

The CRISPR interference (CRISPRi) system was integrated to induce specific gene silencing in the CMT-U27 cell line using either one of two plasmids, *i.e.* pAAVS1-NC-CRISPRi (Gen3) plasmid (Addgene plasmid #73499, Watertown, Massachusetts, USA) and Addgene plasmid #112195 (Watertown, Massachusetts, USA). Plasmid structures are included in Appendix 2.

3.2.3 Plasmid Amplification and Linearization

The dCas9-KRAB and CRISPRi plasmids were delivered in bacterial swabs and cultured on agar plates containing ampicillin. Then, a few colonies were transferred into LB broth + ampicillin, and plasmids were extracted using the QIAprep Spin Miniprep Kit as per the manufacturer's protocols The QIAprep Spin Miniprep Kit (Qiagen, Cat. No. 27106) was purchased in 2020 from Qiagen GmbH, Hilden, Germany, and was available in the laboratory at the time of use. A NanoDrop 8000 Spectofotometer (Thermo scientific) was used to measure the concentration and purity of the plasmids.

To linearize the plasmid *Age*I and *Not*I restriction enzymes were used and then linearization was checked via agarose gel electrophoresis. These enzymes were available in laboratory at the time, but purchase history was unknown. Concentrations of the linearised plasmids were measured again via NanoDrop.

3.2.4 Transfection Procedures

Two forms of transfections were done: supercoiled plasmid transfection and linearized plasmid transfection. For the transfections, cells were cultured in 6-well plates (protocol – Appendix 3) at a confluency of approximately 50–70%. Plasmid was transfected using CRISPRMAX™ lipofectamine (Invitrogen, Thermo Fisher Scientific, USA) and the complete protocol has been included in Appendix 4.

After the cells were transfected, they were grown at 37°C for 1-2 days, depending on the plasmid DNA type. When transfecting the linearised plasmid, a longer incubation period before antibiotic selection was found to be beneficial (48 hours) for better plasmid uptake.

3.2.5 Antibiotic Selection and Stable Line Establishment

For the selection process, growth medium containing 300 µg/ml G418 (geneticin) (Invitrogen, Thermo Fisher Scientific, USA) was used 48 hours after transfection. After two weeks of selection, surviving cells were expanded into new T-25 flasks to generate stable populations. Cytotoxic effects of gentamicin selection were evaluated throughout the selection process with phase contrast microscopy.

3.2.6 gRNA Co-transfection

After developing the cell populations expressing dCas9-KRAB and establishing CRISPRi, cells were detached using 1.5X Trypsinase (protocol- Appendix 3). gRNA was transfected using CRISPRMAX™ (protocol – Appendix 4). After 48 hours of incubation, cells were harvested for downstream analyses. Used gRNAs are previously ordered and tested gRNA from Invitrogen, Thermo Fisher Scientific.

3.2.7 RNA extraction

Extraction of RNA was done with NucleoSpin® RNA Plus kit (Machery Nagel, GmbH & Co., Germany) according to the manufacturer's protocols. All steps were conducted in a RNase-free environment. The concentration of extracted RNA was measured using the NanoDrop 8000 before cDNA synthesis.

3.2.1 Freezing and Cryopreservation of Cells

Transfected CMT-U27 cells were also cryopreserved for long-term storage. The cells were harvested and resuspended in freezing medium, consisting of 90% FBS and 10% dimethyl sulfoxide (DMSO). The cells were aliquoted into cryogenic vials and placed in Mr, Frosty and placed in -80 °C. (all these reagents were previously prepared aliquots presented in laboratory at the time of conducting this step)

3.3 cDNA Synthesis and Quantitative PCR (qPCR)

One microgram of total RNA from CMT-U27 wild-type and CRISPRi-transfected cells were treated with DNase I to remove contaminating genomic DNA. First, reverse transcription with SuperScript IV (Invitrogen™, Thermo Fisher Scientific, USA) was done on the reactions containing oligo(dT) (Invitrogen™,Thermo Fisher Scientific, USA) and random hexamers primers (Thermo Scientific™). Master mixes were prepared in a clean room while reactions were conducted in a template room under RNase-free conditions. Specificity of amplification was confirmed with melt curve analysis following qPCR, and those with atypical melt curves were removed from further analysis. A complete step-by-step protocol is provided in Appendix 5 and 6.

4 Results

4.1 Statistical Analysis of Mouse Mammary Tumour Data

To identify key transcriptional changes in the SRGN "low" environment, knockout (KO) mice were compared to heterozygous (HET) mice. There were three KO (mus30, mus31, mus38) and three HET (mus32, mus37, mus39). Six genes were selected based on the largest absolute $\log_2$ fold change and statistical significance ($p < 0.05$) from 672 genes with a significantly altered expression level. Those are BPIFA2, KLK1, SPT1, BC023719, DCPP2, and A2M. However, there was no statistical significance ($p > 0.05$) in the Mann–Whitney U tests, and this is likely due to the limited sample size. BPIFA*2* demonstrated upregulation in the KO group, whereas DCPP2 and A2M were downregulated. This may indicate potential roles in the biological processes affected by gene knockout (Figure 4). As the microarray assay may produce a noise level reaching up to 4.5, values below 4.5 needs to be evaluated cautiously.

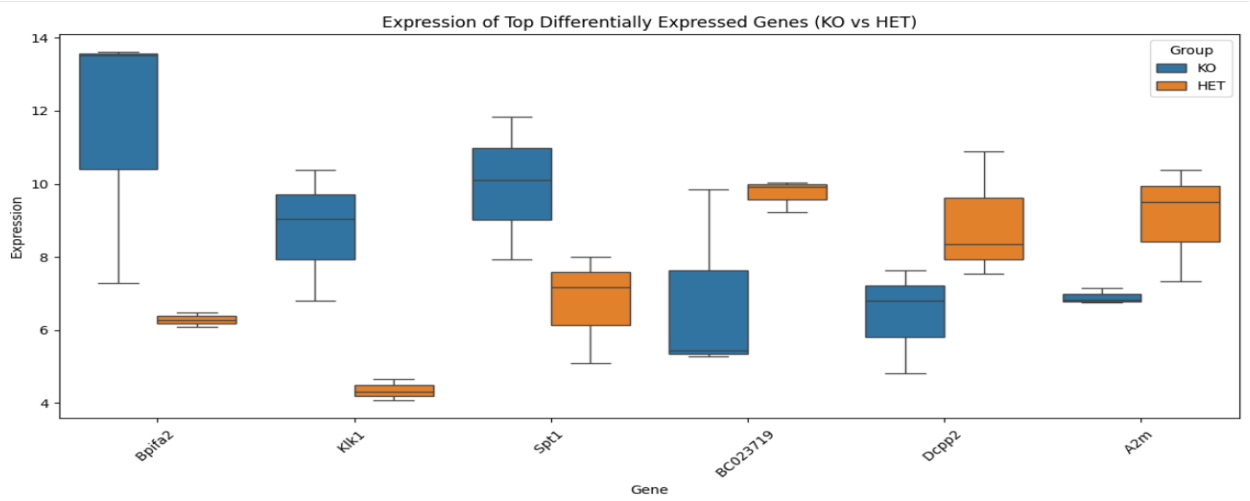


Figure 4; DEG in Knockout vs control mouse sample. Values below 4.5 are not significant as the noise level = 4.5. (Source: Author's own, generated during bioinformatic analysis using Python programming language in Jupiter Lab, 2025)

4.2 *SRGN* mRNA expression of Mouse Mammary Tumour

The microarrays contain seven different oligonucleotides for each gene, and two different assays were conducted. SRGN mRNA expression was compared in heterozygous (HET) and knockout (KO) mouse mammary tumour samples. In Assay 1 (all 7 values), SRGN expression was lower in KO mice (mean = 5.15) relative to HET (mean = 7.48). This difference was confirmed by an unpaired two-tailed t-test, which showed a significant difference ($p = 0.0016$), graph is attached to appendix 7. Assay 2 (with discarded values below noise level) also showed same results but the differences were even higher, with KO mice expressing a mean SRGN of 4.89 and HET mice averaging 10.52 (Figure 5). The difference was highly significant with $p = 2.16 \times 10^{-7}$. Moreover, all KO replicates displayed lower expression values, suggesting a consistent KO phenotype. These statistical assessments demonstrate that KO mice exhibit a robust reduction of SRGN expression compared to HET controls.

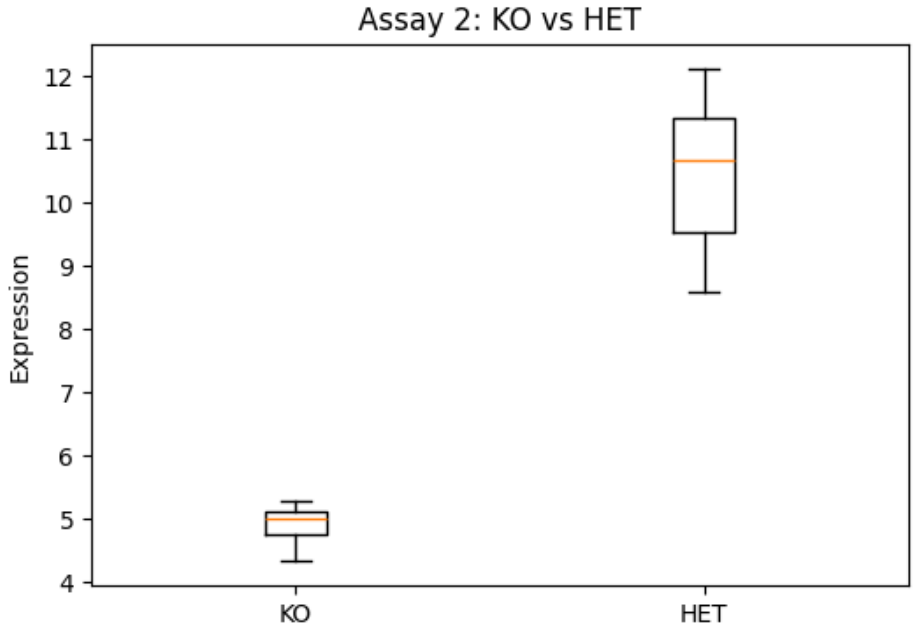


Figure 5; SRGN expression Knockout vs control. Which indicate knockout success. (Source: Author's own, generated during bioinformatic analysis using Python programming language in Jupiter Lab, 2025)

4.3 Statistical Analysis and Quality Control - Human breast cancer data sets

In GSE42568, the volcano plot suggested a balanced number of upregulated and downregulated genes, with neither side dominating the other. In GSE29044, however, the plot showed greater numbers of upregulated genes, revealing a stronger skew towards overexpression. In GSE89116, late-stage tumours showed greater dispersion in expression changes, and GSE109169 had significant gene expression changes in early onset tumours related to immune-related context.

In all four datasets, expression values remained uniform within sample groups, and no significant outliers or unconventional distributions appeared in the volcano plots or sample-level boxplots.

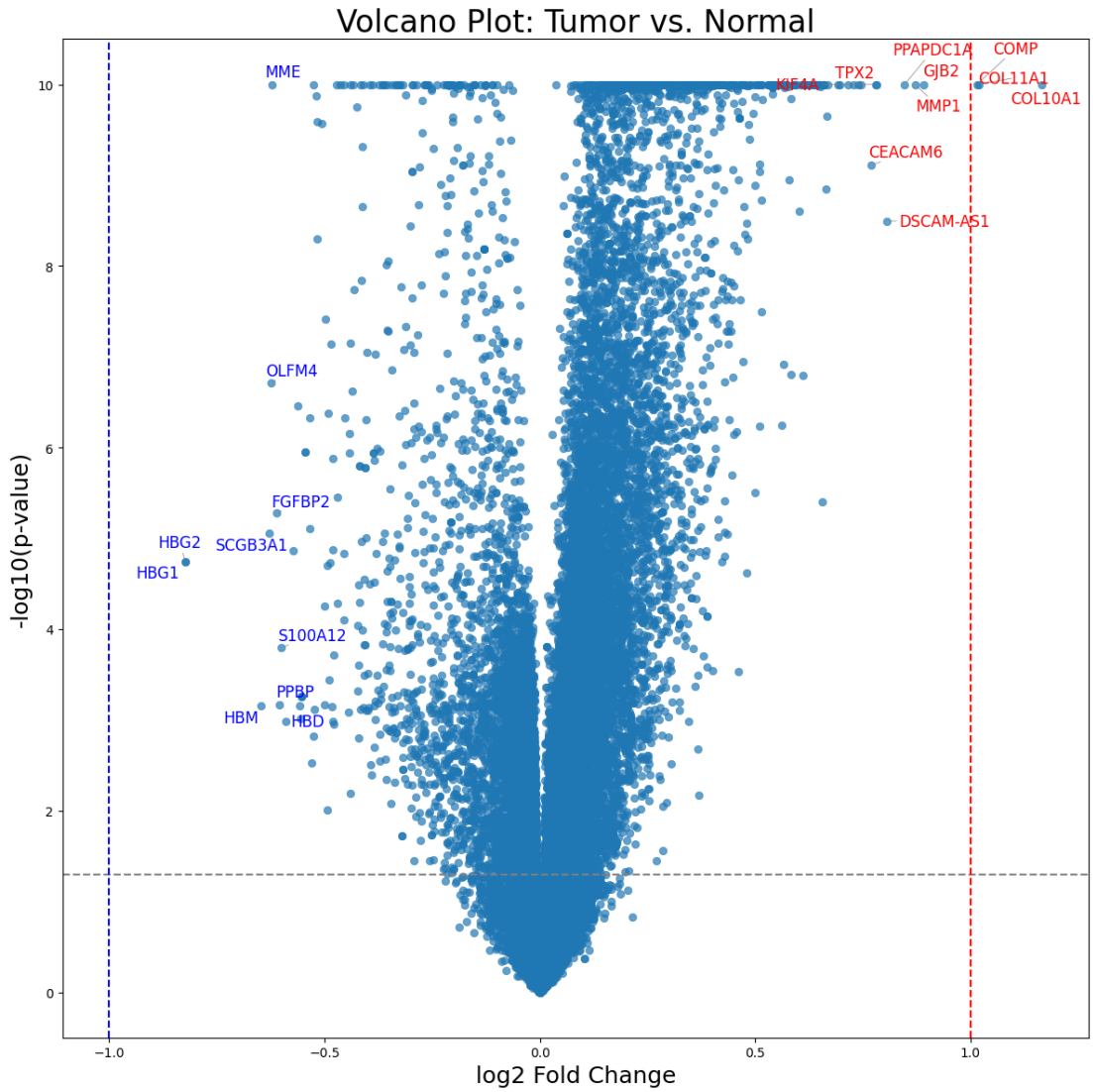


Figure 6; Volcano plot of differential gene expression between tumour and normal tissue samples (GSE29044). Genes with significant upregulation (log□ fold change p > 0.05) are marked in red, and significantly downregulated genes (log□ fold change p < -0.05) are marked in blue.
(Analysis and figure generated by the author using GSE29044 data from the Gene Expression Omnibus (GEO) using Python programming language in Jupiter Lab, 2025)

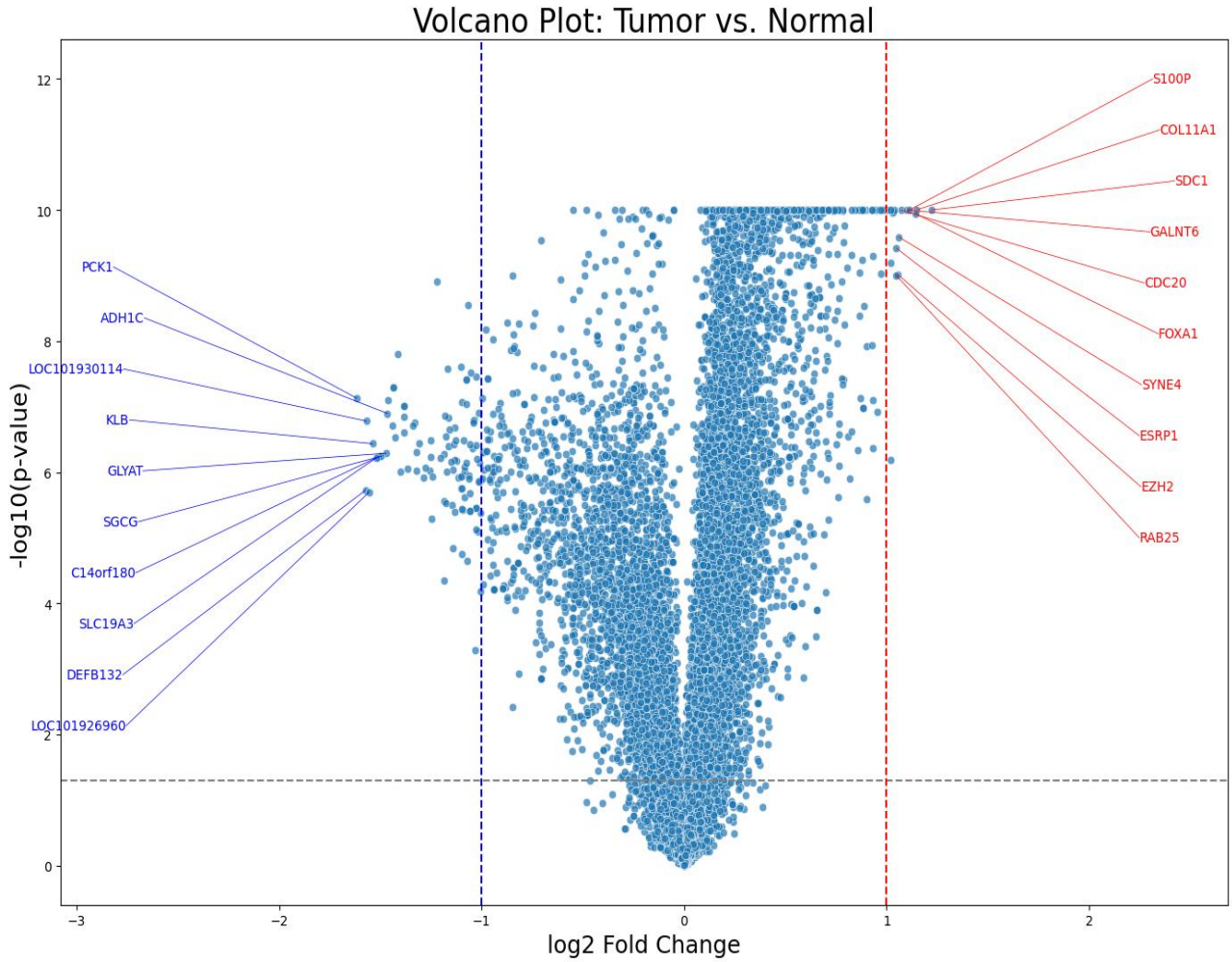


Figure 7: Volcano plot of differential gene expression between tumour and normal tissue samples (GSE42568). Genes with significant upregulation (log□ fold change p > 1) are marked in red, and significantly downregulated genes (log□ fold change, p < -1) are marked in blue.
(Analysis and figure generated by the author using GSE42568 data from the Gene Expression Omnibus (GEO) using Python programming language in Jupiter Lab, 2025)

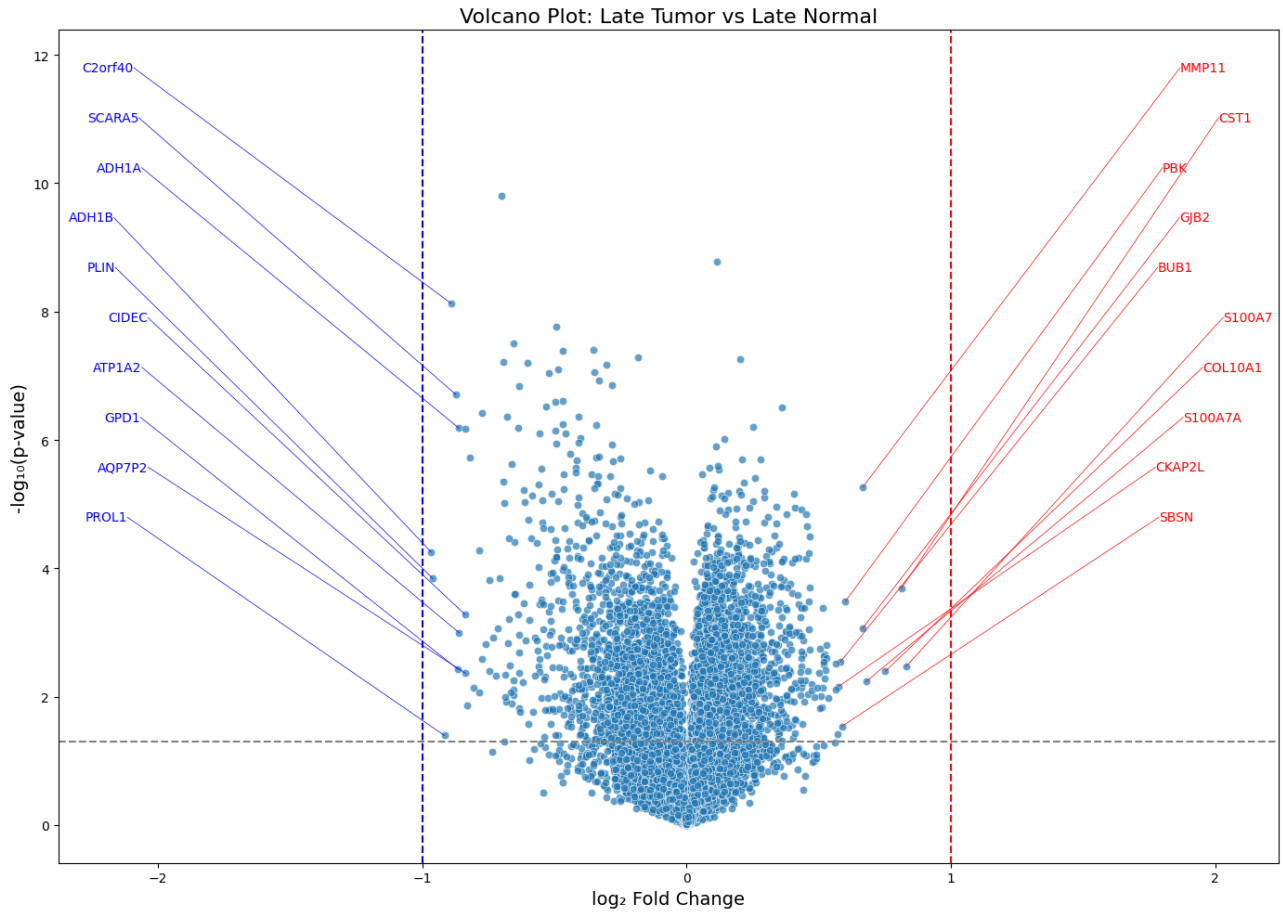


Figure 8: Volcano plot of differential gene expression between tumour and normal tissue samples (GSE 89116). Genes with significant upregulation (log□ fold change p > 0.05) are marked in red, and significantly downregulated genes (log□ fold change p < -0.05) are marked in blue.
(Analysis and figure generated by the author using GSE89116 data from the Gene Expression Omnibus (GEO) using Python programming language in Jupiter Lab, 2025)

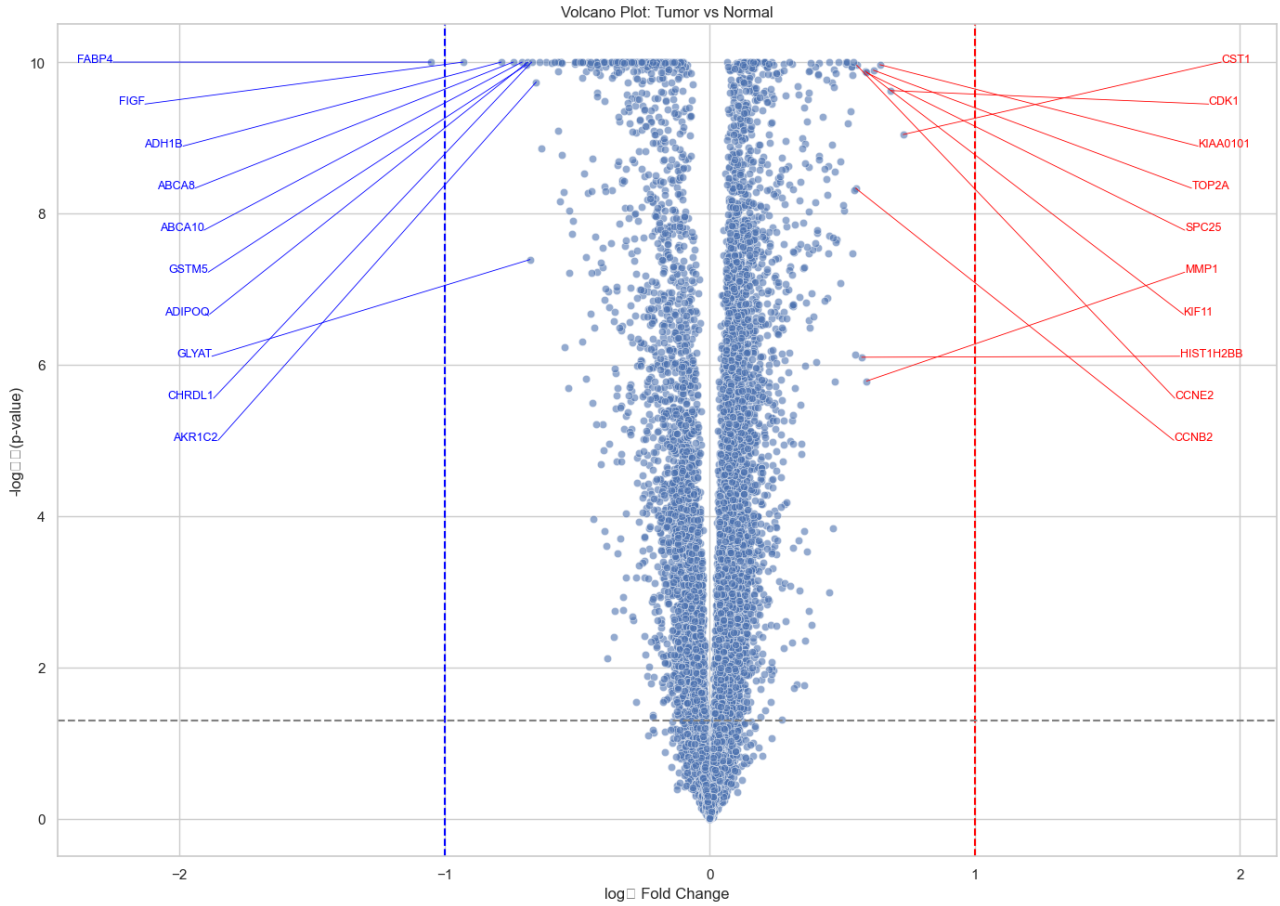


Figure 9: Volcano plot of differential gene expression between tumour and normal tissue samples (GSE109169). Genes with significant upregulation (log□ fold change p > 0.05) are marked in red, and significantly downregulated genes (log□ fold change p < -0.05) are marked in blue.
(Analysis and figure generated by the author using GSE109169 data from the Gene Expression Omnibus (GEO) using Python programming language in Jupiter Lab, 2025)

4.3.1 DEG Summary per Dataset

Differential gene expression analysis was conducted for each dataset using an adjusted p-value cutoff of 0.05 and a minimum $\log^2$ fold-change of ±1. The adjusted thresholds for significance as well as representative examples of the upregulated and downregulated genes are summarized below.

Table 1; DEG analysis summary

Dataset	Total DEGs	Upregulated Genes	Downregulated Genes	Adjusted p-value Cutoff	Top Upregulated Genes	Top Downregulated Genes
GSE42568	1,412	689	723	FDR < 0.05	SDC1, COL11A1, S100P	PCK1, KLB, ADH1C
GSE29044	1,236	641	595	FDR < 0.05	COL10A1, COMP, GJB2	HBG2, SCGB3A1, MME
GSE89116	1,782	902	880	FDR < 0.05	PBK, COL11A1, MMP11, S100A7	CYP4B1, KLF4, GPIHBP1
GSE109169	2,214	1,074	1,140	FDR < 0.05	CXCL9, CXCL10, HLA-DRA	FABP4, FIGF,ADH1B, ABCA8

4.4 Functional Enrichment of Differentially Expressed Genes

An analysis was performed to investigate the differences in functional enrichment with regards to the significantly upregulated and downregulated genes in GSE29044 and GSE42568. Gene Ontology (GO) terms were created with p-value cutoff lower than 0.05 after correction for multiple testing.

As part of the enrichment analysis in GSE29044, upregulated genes were associated with functions related to extracellular matrix. Their most significant GO terms include “extracellular matrix structural constituent” (GO:0005201), “tendon development” (GO:0035989), “extracellular matrix organization” (GO:0030198), “collagen type X trimer” (GO:0005599), and “supramolecular polymer” (GO:0099081). Those hits suggest the enrichment of structural and matrix components in the tumour tissue.

In GSE42568, genes SDC1, COL11A1, FOXA1, S100P, and GALNT6, with adjusted p-values < 1e–10, were among the most upregulated genes. They are associated with cellular proliferation, matrix signaling, and epithelial plasticity. In contrast, PCK1 ($\log_2$FC = –1.62), KLB, ADH1C, and DEFB132 mostly active in metabolism and tissue-specific functions were downregulated

In GSE29044, the most significantly downregulated genes included MME (–0.62, p = 9.15e–23), SCGB3A1, OLFM4, and HBG2. These genes are associated with secretory and differentiation processes which indicate an epigenetic silencing of epithelial characteristics.

GSE89116 revealed that genes involved in cell proliferation (e.g., PBK, BUB1), ECM remodelling (e.g., MMP11, COL10A1) and metastasis or immune evasion (e.g., S100 family genes) are highly expressed in late-stage tumours in contrast to normal tissues.

4.5 Up- and Downregulated Genes in Human Breast Cancer

In GSE42568, tumour samples displayed GO-terms “ECM remodelling” and “invasion” with for example SDC1, COL11A1, FOXA1, and S100P as significantly upregulated and with PCK1, KLB, and ADH1C downregulated

Micrometastatic tumor dissemination was noted during the progression of breast cancer, describing an abundant form of tumoral dissemination in GSE29044. Tumours showed strong upregulation of the ECM associated genes COL10A1, COMP, and COL11A1 while HBG2, SCGB3A1, and MME were downregulated.

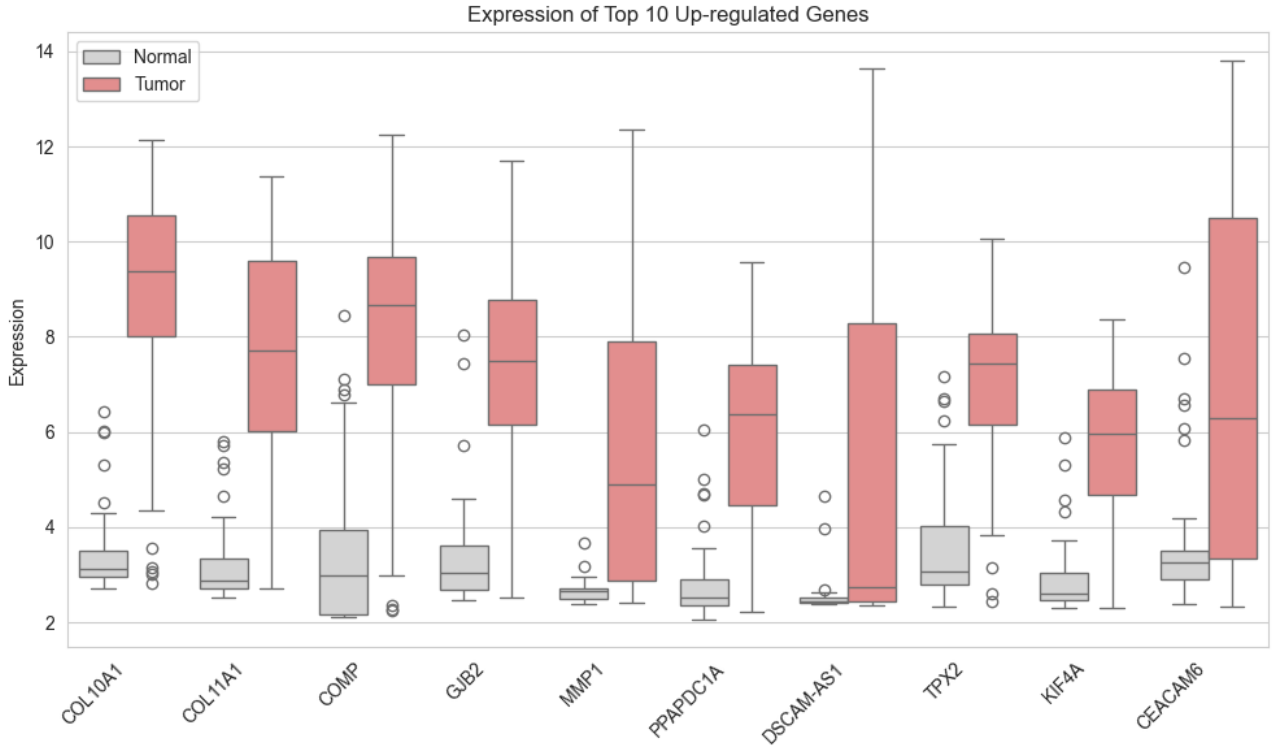


Figure 10; Boxplot showing gene upregulation in tumour samples, GSE29044. Tumour sample in (red) and normal samples (ash colour) Graph generated by the author using data from the Gene Expression Omnibus dataset GSE29044 using Python programming language in Jupiter Lab, 2025)

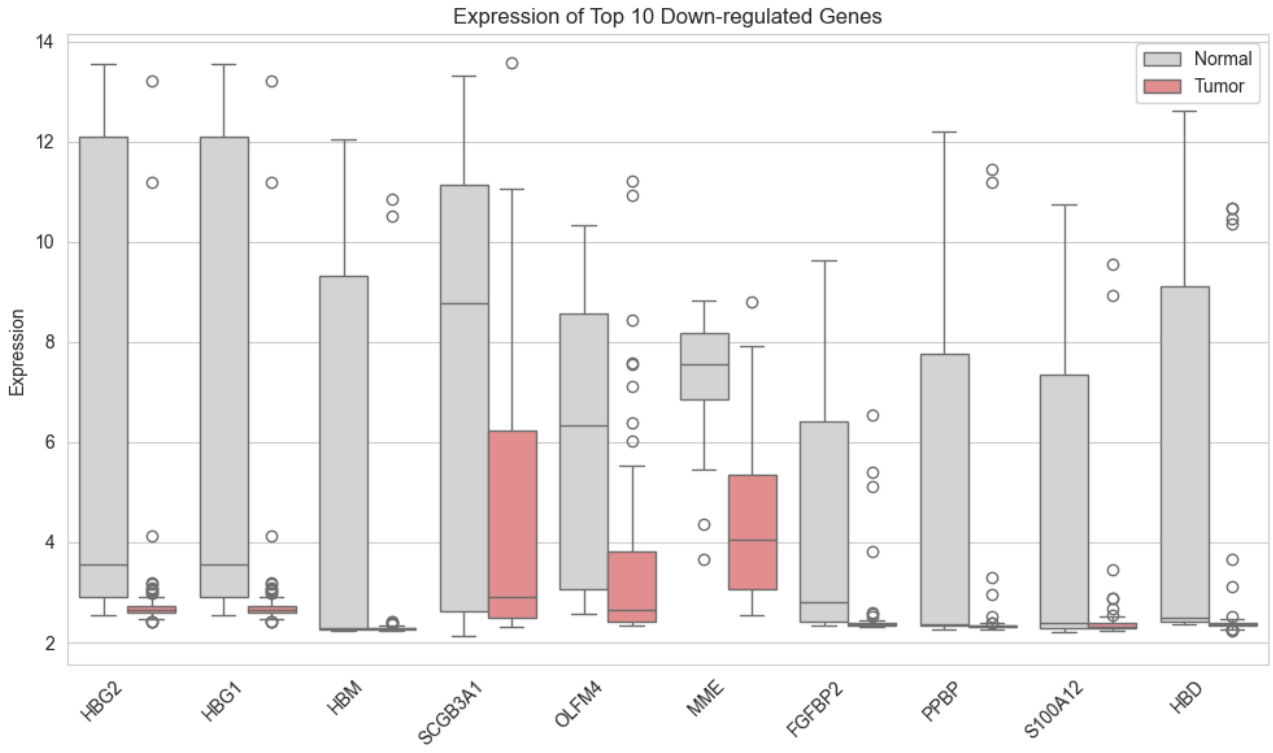


Figure 11; Boxplot showing gene down-regulation in tumour samples, GSE29044. Tumour sample in (red) and normal samples (ash colour). Graph generated by the author using data from the Gene Expression Omnibus dataset GSE29044 using Python programming language in Jupiter Lab, 2025)

Tumour samples taken from early onset have retained some normal tissue markers that were more strongly downregulated in samples from late onset tumours (Figure 12 & 13).

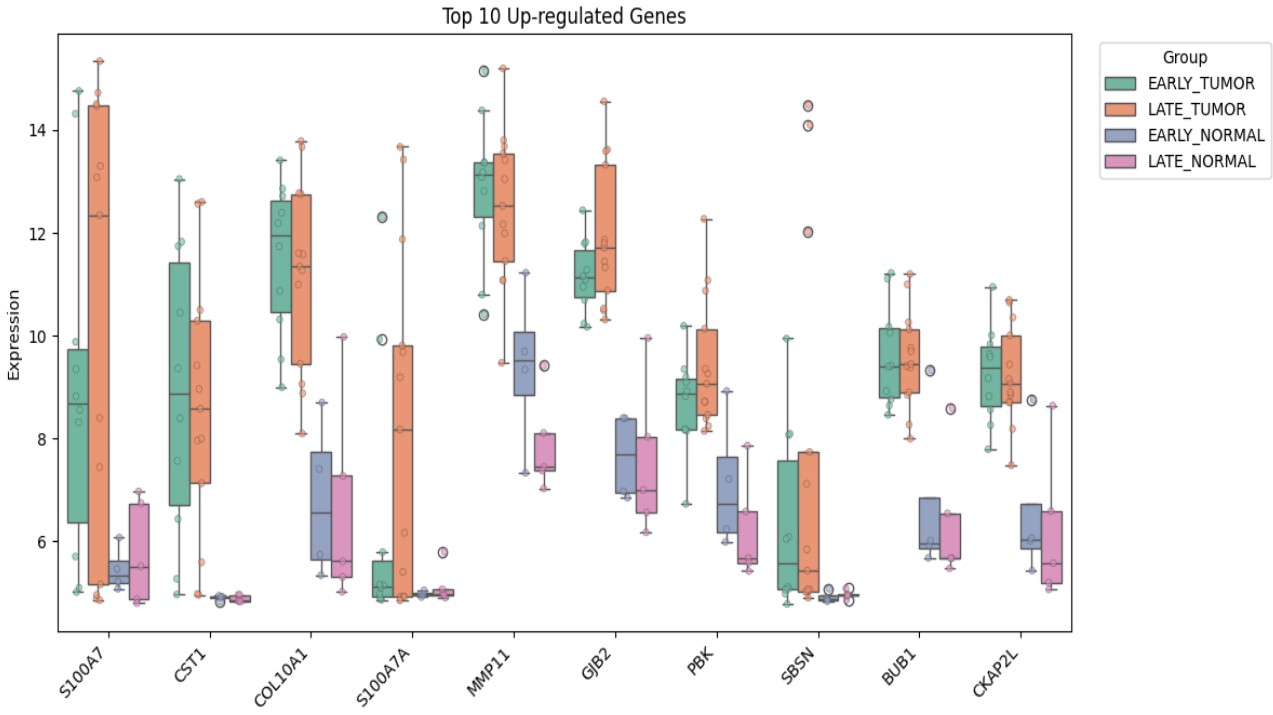


Figure 12; Boxplot showing gene upregulation in tumour samples early vs late in contrast to the same age normal tissue samples, GSE89116. Graph generated by the author using data from the Gene Expression Omnibus dataset GSE89116 using Python programming language in Jupiter Lab, 2025)

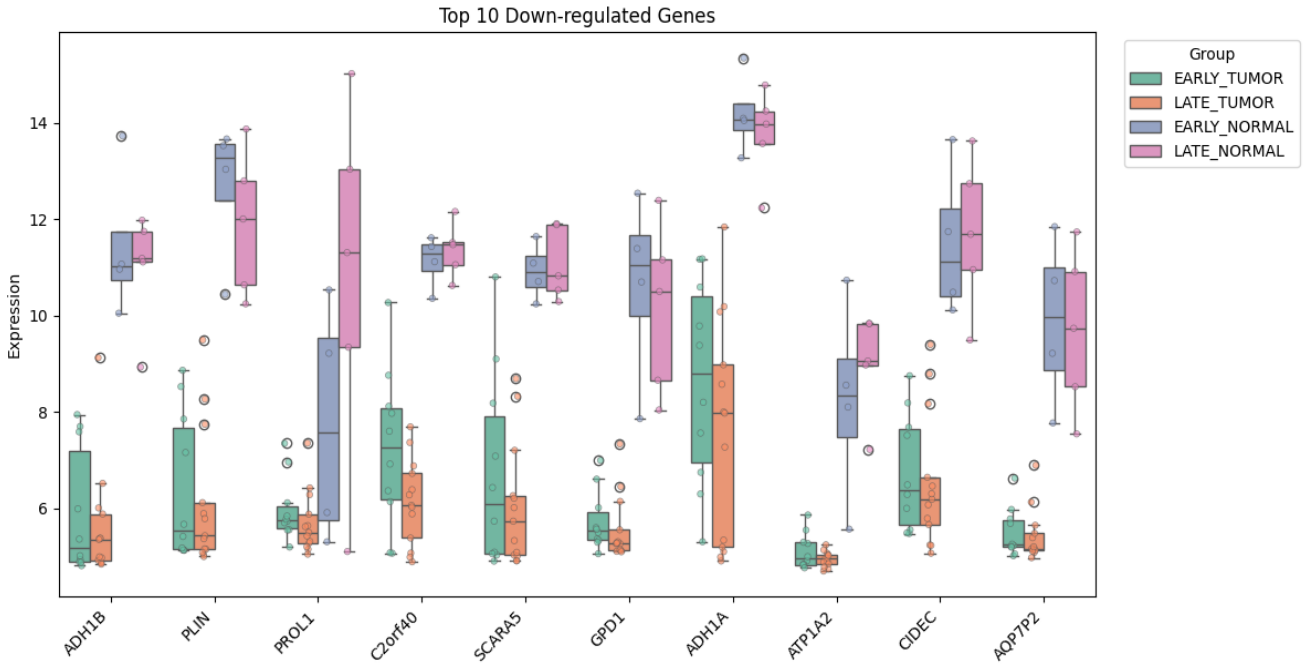


Figure 13; Boxplot showing downregulated genes in tumour samples early vs late in contrast to the same age normal tissue samples, GSE89116. Graph generated by the author using data from the Gene Expression Omnibus dataset GSE89116 using Python programming language in Jupiter Lab, 2025)

In GSE109169, tumours demonstrating elevated immune cell densities also exhibited upregulation of CXCL9, CXCL10, HLA-DRA, along with other genes of the immune system.

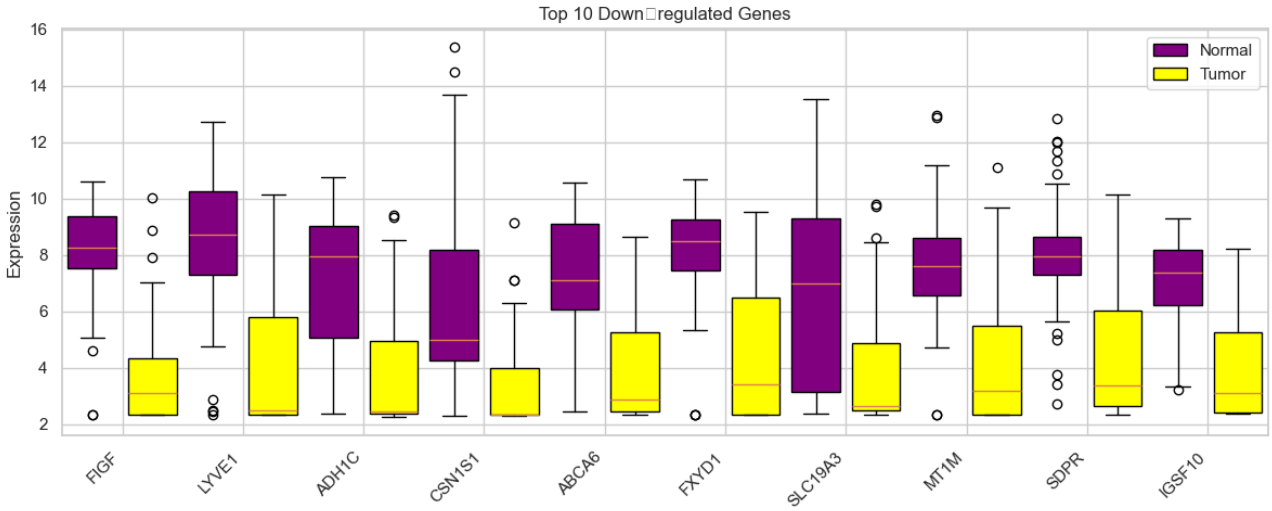


Figure 14; Boxplot showing top down-regulated genes in tumour vs normal sample cumulatively across all 4 data sets. Tumour sample in (yellow) and normal samples (purple). Graph generated by the author using data from the Gene Expression Omnibus datasets GSE29044, GSE42568, GSE89116 and GSE109169 using Python programming language in Jupiter Lab, 2025)

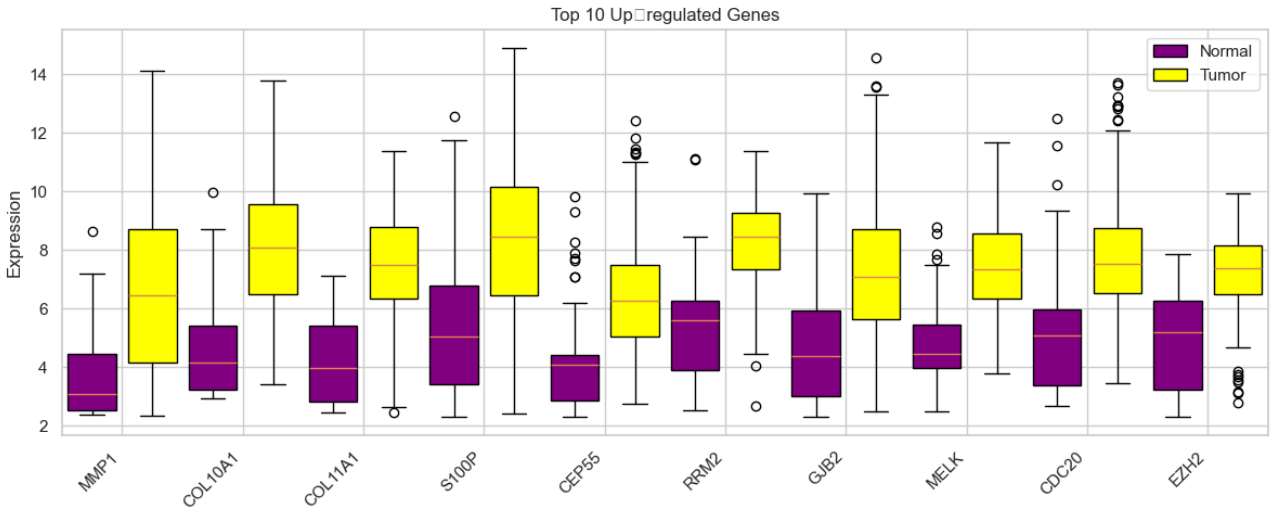


Figure 15; Boxplot showing top up-regulated genes in tumour vs normal sample cumulatively across all 4 data sets. Tumour sample in (yellow) and normal samples (purple). Graph generated by the author using data from the Gene Expression Omnibus datasets GSE29044, GSE42568, GSE89116 and GSE109169 using Python programming language in Jupiter Lab, 2025)

4.6 SRGN Regulation in Human Breast Cancer

The expression of SRGN exhibited lower level in tumour samples than in normal tissues. In the case of GSE42568, SRGN expression was lower in the tumour relative to normal tissue ($p = 5.786 \times 10^{-7}$). In GSE29044, tumour samples had a lower median SRGN expression compared to normal, though this difference did not reach statistical significance ($p = 0.0843$). In GSE89116, SRGN expression was upregulated in later stages of tumour progression compared to early stages and was consistently low in normal tissue. In GSE109169, immune-rich and immune-depleted tumours had comparable SRGN expression, with immune-rich tumours showing slightly elevated expression. Overall, context-dependent regulation was evident for SRGN as some datasets demonstrated lower tumour expression, while others showed greater dependence on stage or immune context.

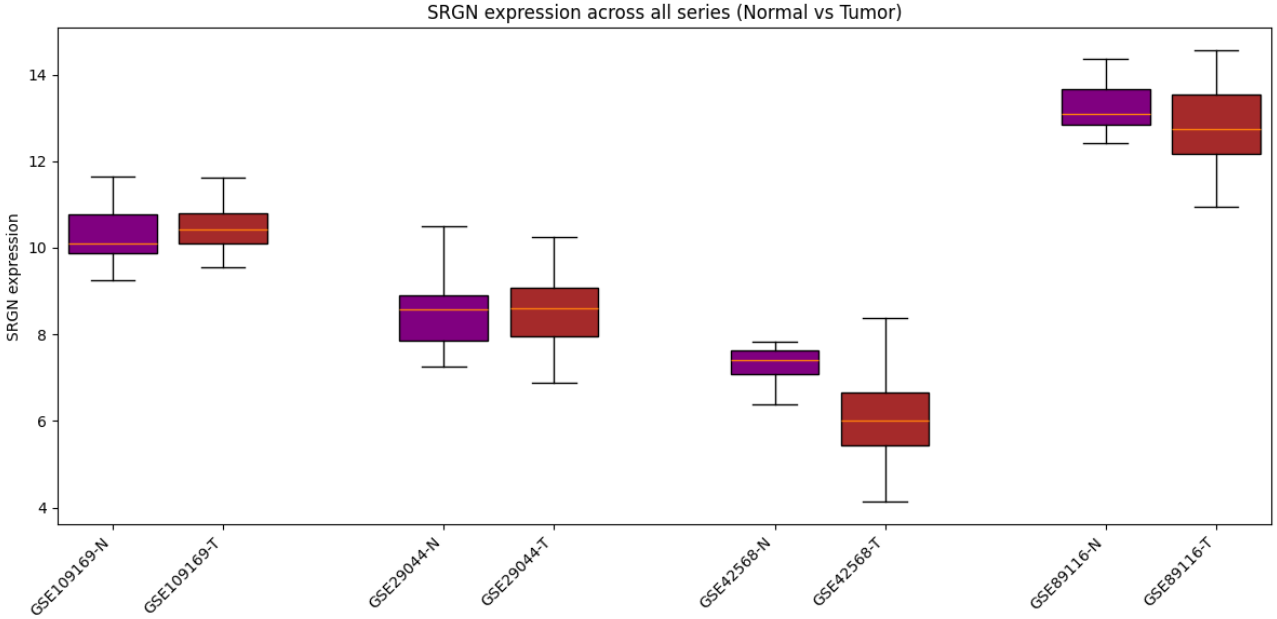


Figure 16; Boxplot showing SRGN expression in tumour vs normal samples across all 4 data sets. Tumour sample in (brown) and normal samples (purple). Graph generated by the author using data from the Gene Expression Omnibus datasets GSE29044, GSE42568, GSE89116 and GSE109169 using Python programming language in Jupiter Lab, 2025)

4.7 CMT-U27 Cell Line

4.7.1 Antibiotic Selection and Cell Survival Following CRISPRi Transfection

Following the transfection of CMT-U27 cells with CRISPRi-dCas9-KRAB plasmid, the cells were exposed to selection with G418. The morphological shape and density of cells were evaluated via phase-contrast microscopy twice a week to determine their tolerance to the antibiotic and survivability.

All 6 wells displayed a marked increase in cell death and detachment during tumour selection that lasted for 72 hours. This indicates effective G418 activity. In 7 days, surviving wells had a low amount of cell debris, with few colonies emerging from resistant cells. This shows successful uptake of the plasmid and expression of the resistance gene. The observed colonies were morphologically similar to the parental CMT-U27 cells, compact and more tightly adherent. Most wells exhibited recovery by day 14 post-selection, except control wells. The below phase contrast microscopy images show cells before and after antibiotic selection (Figure 17).

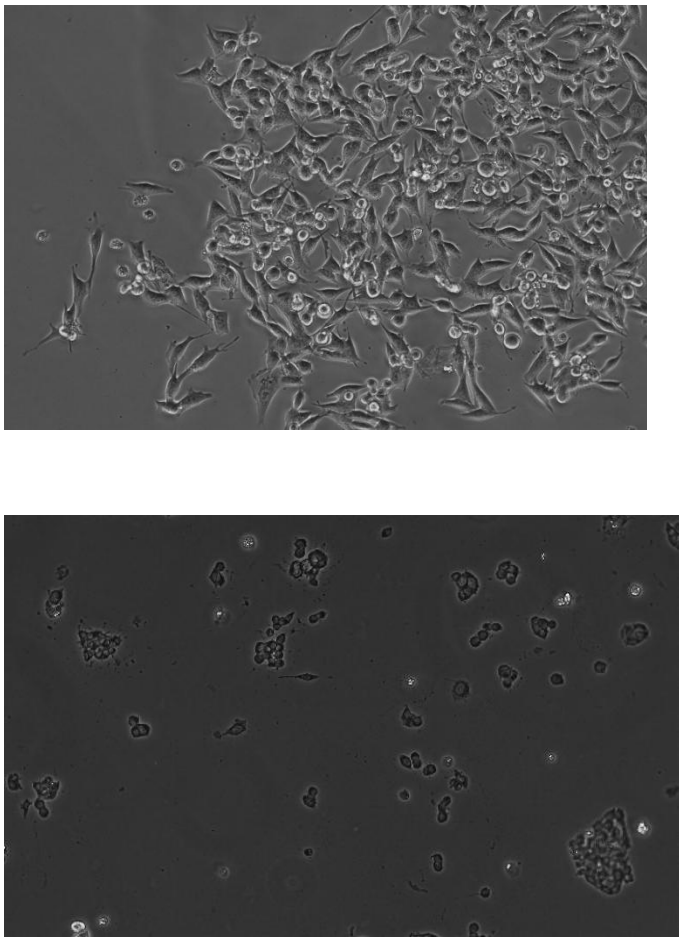


Figure 17; Phase-contrast microscopy images showing CMT-U27 wild type cells before (top) and after (bottom) G418 antibiotic selection. The upper image displays a dense population of untreated cells, while the lower image shows cellular debris from dead cells following selection.
(Images captured by the author using phase-contrast microscopy, 2025)

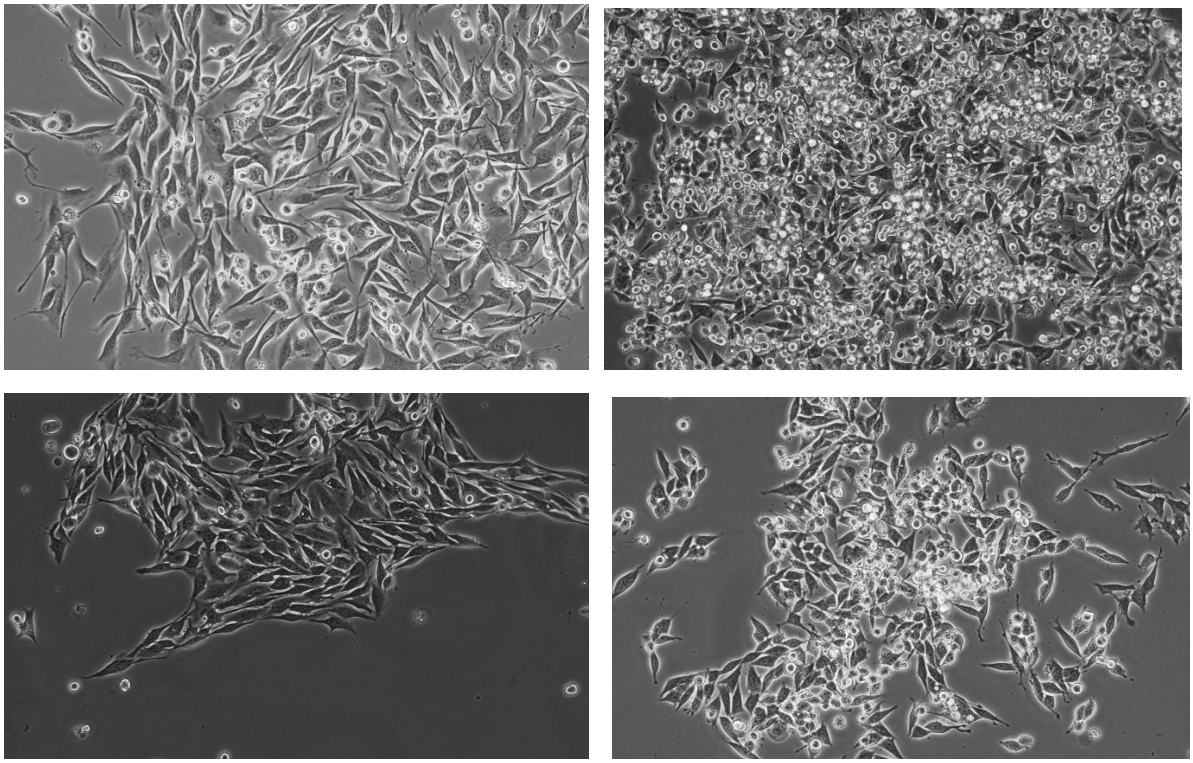


Figure 18; Surviving transfected populations 5–7 days post-selection with G418, showing high viability and dense adherence in multiple colonies, indicating successful transfection. (Images captured by the author using phase-contrast microscopy, 2025)

4.8 qPCR Analysis of Plasmid Delivery

Plasmid delivery and stable expression of the CRISPRi system was further verified via qPCR using primer pairs for plasmids #112195 (KRAB) and #73499 (CRISPI) with RPS5 serving as a reference gene. For plasmid #112195, only two groups, the CMT, KRAB-C (only containing plasmid) and KRAB-G (containing qRNA for SRGN exon i) groups, had adequate replicates (n = 2 each). Levene's test showed variance inequality ($W = \infty$, $p = 0.000$) and the Kruskal-Wallis test produced $H = 2.400$, $p = 0.121$, indicating no significant difference between groups in delta Cq values at the 0.05 alpha level.

Plasmid 73499_2 (plasmid detection with primer pair 2) had at least two biological replicates per sample for CMT, CRISPI-C (contain only plasmid 73499), CRISPI-G (containing gRNA for SRGN exon 1) and CRISPI-G-2 (containing gRNA for SRGN exon 2). All groups failed to meet normality due to low sample count (Shapiro–Wilk $p > 0.05$).

4.9 Fold change of plasmid expression

For qPCR, one primer pair was used for #112195 and plasmid and two different primer pairs were used for plasmid #73499, which was named as 73499_1(contain primer pair 1 for #73499 plasmid recognition) and 73499_2 (containing primer pair 2 for #73499 plasmid recognition). When compared to the non-transfected control (CMT), cells with KRAB-C (containing plasmid #112195) and KRAB-G (contain plasmid #112195 and gRNA for SRGN exon 1) transfections displayed robust stimulation of #112195 plasmid with fold (FC) changes of 21.63 and 21.98, respectively. The 73499 _2 target gave even greater signals in CRISPI-C (FC = 123.70), CRISPI-G (containing gRNA for SRGN exon 1) (103.97), and CRISPI-G-2 (149 .30).

```
Fold Change (2^-ΔΔCq):
 Biological Set Name  CMT    CRISPI-C    CRISPI-G  CRISPI-G-2
Target
112195                1.0         NaN         NaN         NaN
73499_1               1.0    8.945887    1.963265    9.597145
73499_2               1.0  123.702125  103.965244  149.297318
RPS5                  1.0    1.000000    1.000000    1.000000
```

Figure 19; Fold change in plasmid expression levels of targets 112195, 73499_1, and 73499_2 CMT-U27 cells transfected with different CRISPRi control, gRNA containing SRGN exon 1 and 2. In contrast to wild type. RPS5 was used as the reference gene. Calculated using the 2^-ΔΔCq method using Python programming language in Jupiter Lab, 2025)

4.10 SRGN Expression in CRISPRi-Treated Cell Lines

Statistical analysis was confined to groups with biological replicates of three or greater: CMT and CRISPI-C (n = 4 each). Normality was confirmed using the Shapiro-Wilk test (CMT: W=0.858, p=0.255; CRISPI-C: W=0.933, p=0.612}). Levene's test for equality of variances yielded results W=1.523, p=0.282. A one-way ANOVA with CMT, CRISPI-C, CRISPI-G, and CRISPI-G-2 resulted in F=0.954, p=0.459, demonstrating no statistically significant differences in SRGN expression between groups at the $p < 0.05$ level. No comparisons were performed due to the low replicate numbers in CRISPI-G and CRISPI-G-2 (n = 2 each)

4.11 Fold change

SRGN expression was evaluated by qPCR in CRISPRi-transfected canine cancer cell lines, each compared to its corresponding control. For the CRISPRi-guided constructs, CRISPI-G and CRISPI-G-2 were compared to CRISPI-C. Both conditions showed effective SRGN repression, with fold changes of 0.70 and 0.39, respectively. In contrast, SRGN expression in CMT (non-transfected) was lower than in CRISPI-C, with a fold change of 0.27, indicating that CRISPI-C itself may elevate SRGN levels or that the "subcloning" via G418 of CRISPRi transfected cells by chance selected CMT U27 cells with a higher expression of SRGN. For the KRAB-dCas9 conditions, KRAB-C and KRAB-G were compared to KRAB-C as the internal reference. KRAB-G showed an increased SRGN expression (fold change = 3.58), while CMT had a slightly elevated expression relative to KRAB-C (fold change = 1.65). These results indicate that SRGN suppression was successful in CRISPI-G and CRISPI-G-2, while KRAB-only conditions yielded variable expression patterns depending on the gRNA used.

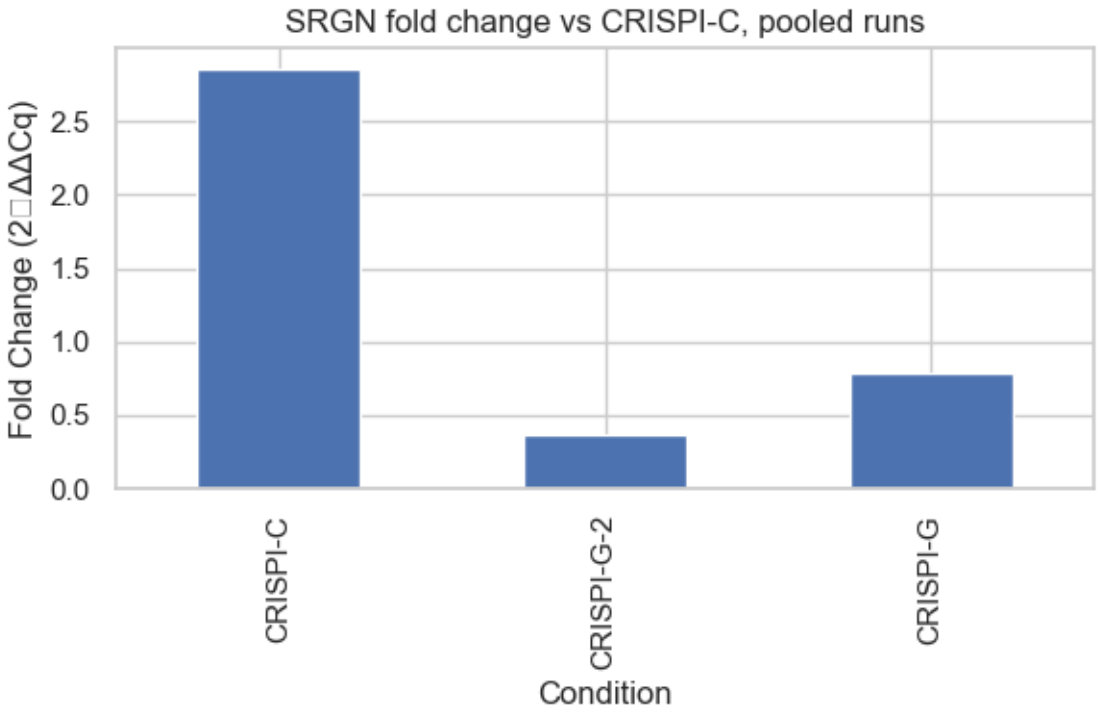


Figure 20: Fold change in SRGN expression relative to the control (CRISPI-C) in CMT-U27 cells transfected with plasmid #73499. Expression was measured by qPCR and calculated using the 2^−ΔΔCq method. Both CRISPI-G and CRISPI-G-2 constructs showed strong suppression of SRGN expression compared to the control. Graph generated by the author using Python programming language in Jupiter Lab, 2025)

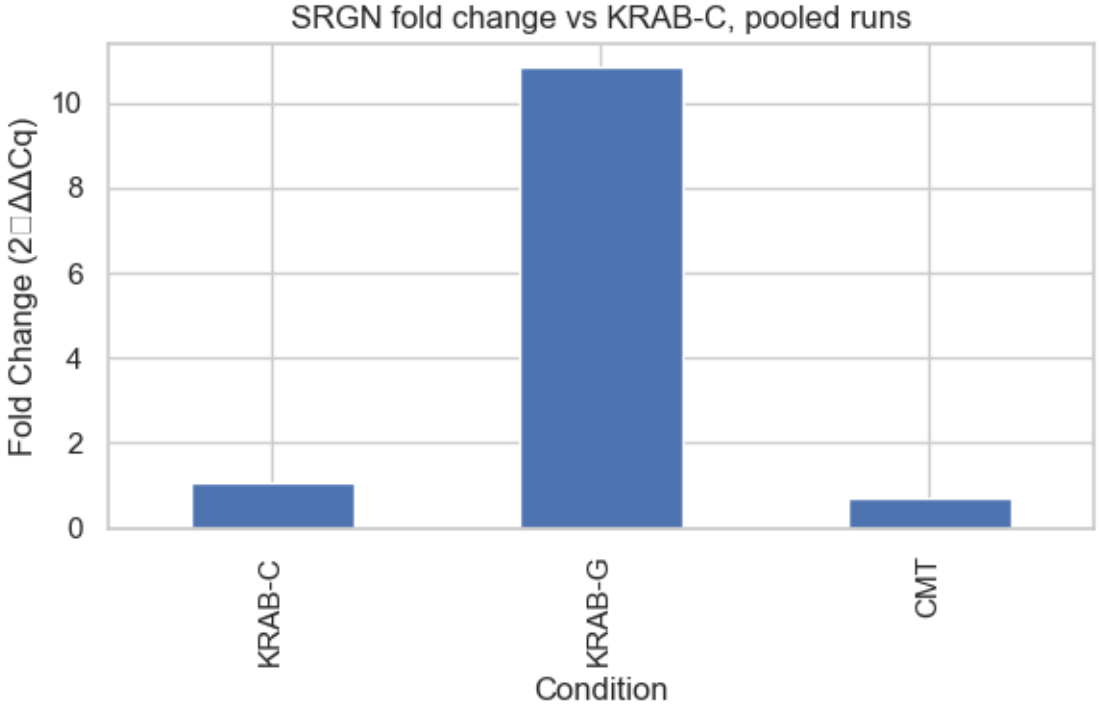


Figure 21: Fold change in SRGN expression relative to the control (KRAB-C) in CMT-U27 cells transfected with plasmid #112195. Expression was measured by qPCR and calculated using the 2^−ΔΔCq method. Graph generated by the author using Python programming language in Jupiter Lab, 2025)

5 Discussion

5.1 Overview of Findings

This thesis analysed changes in the regulated transcriptome of mammary tumour tissue with special focus on the potential importance of the SRGN gene that encodes serglycin. In addition, the

CRISPRi system was used to repress SRGN gene expression in the canine mammary tumour cell line CMT U27.

Four human datasets indicated that the expression level of SRGN changed with tumour stage and immune infiltration. It was either downregulated (as in GSE42568 or GSE29044) or upregulated in late-stage and immune-rich tumours (GSE89116 and GSE109169 respectively). Functional enrichment of differentially expressed genes indicated the persistent involvement of the extracellular matrix (ECM) remodelling, inflammation and cell division related pathways. In the mouse model, Serglycin (SRGN) knockout caused notable transcriptional changes, including the downregulation of immune and ECM-remodelling associated genes.

CRISPRi plasmids were delivered to CMT U27 cell line and with two different guide RNAs targeting exon 1 and exon 2 of the SRGN gene significant downregulation of the SRGN expression was demonstrated. These results demonstrate that transcriptional regulation in the CMT U27 is possible when using the CRISPRi system, and thus, provides a novel tool for evaluation of context-dependent mechanisms in canine mammary cancer progression.

5.2 Statistically significant and Quality control in Human data sets

When it comes to deferential expressed genes (DEGs) mRNAs from some genes which were highly regulated (1> log2FC >-1) were present at low levels in all datasets. It could be due to heterogeneity of the tumour (consisting of a mixture of cancer, stromal, and immune cells). Such diversity may dilute the intensity of gene expression signals in bulk samples, and changes would be more prominent in homogeneous cell populations. On the other hand, normal tissues may have a more uniform cell composition, which makes their gene expression levels more stable and easier to interpret. Therefore, as only a few genes showing extreme fold changes this does not imply a lack of cancer-related dysregulation but highlights the limitations of bulk analysis. Approaches like single-cell RNA-seq or spatial transcriptomics could help reveal these masked expression patterns more clearly.

5.3 Functional Implications of Shared Differentially Expressed Genes

In all four human breast cancer datasets, a subset of genes showed a consistent pattern of differential expression in tumour tissues relative to normal controls. Some of the most frequently upregulated genes included COL10A1, COL11A1, TPX2, COMP, and SDC1. These genes have important roles in extracellular matrix remodelling but also in cell growth and the activation of stromal tissues.

COL10A1, a collagen isoform, is one of the most frequently overexpressed genes in many mammary tumour samples. It is linked to cell proliferation, migration and invasion of the tumour cells. The presence of COL10A1 in multiple datasets in this study suggests its possible use as a marker of ECM remodelling and breast cancer invasiveness. COMP and SDC1 also serve as ECM-associated proteins and are involved in metastases and angiogenic processes.

S100 family genes are highly expressed in late-onset tumours, in contrast to normal tissues, indicating metastasis and immune cell infiltration. On the other hand, the downregulated genes PCK1, SCGB3A1, and KLB are related to a tissue’s specific metabolism, as well as epithelial and secretory functions. These genes showed downregulation across several human tumour datasets, indicating that tumour progression not only activates invasive growth and advancement pathways, but also downregulates genes that maintain normal tissue identity and metabolic homeostasis.

The persistent recurrence of these genes along with their associated biological processes across various datasets strongly suggests the existence of a conserved core set of transcriptional alterations

in breast tumour progression. This work provides a foundational molecular framework that helps understand the control of SRGN expression, which seems to be involved with other ECM and immune-related genes to modulate the mammary/breast cancer microenvironment.

5.4 SRGN Regulation Across Human Breast Cancer Datasets

The regulation of SRGN across human breast cancer datasets showed a context-dependent expression pattern according to age and onset, which may reflect the gene's dual roles in immune modulation and extracellular matrix (ECM) dynamics. In some tumour samples, SRGN appeared to be downregulated, especially when comparing tumour samples to normal tissue (NT) in paired samples. However, when compared to older onset tumours or in immune cell-dominated regions, SRGN was increased. This suggests SRGN expression in cancer is highly dependent on the surrounding microenvironment and immune cells. SRGN is known to control the immune & inflammatory mediators, but epithelial tumour cells might downregulate SRGN as part of a dedifferentiation or inactivated secretory pathways. Thus, the expression trends across datasets can be attributed to the interplay of these contrasting influences.

Additionally, the more advanced stage of the tumour and associated increase in SRGN expression suggest it is involved in tumour progression rather than initiation. However, the variability across datasets suggests that SRGN is not a biomarker for a fixed state but a dynamic regulator, with its expression reflecting the changing interaction between tumour cells and their surrounding microenvironment. These results underscore the need to discriminate and stratify breast cancer by immune phenotype in relation to evaluating the dual purpose of SRGN as a possible therapeutic target or biomarker.

5.5 Mouse SRGN Knockout: Systemic Gene Expression Changes

The SRGN knockout model described here, deepened our understanding regarding the transcriptional changes that accompany SRGN loss within mammary tumour tissues. Here the systemic changes in gene expression in the knockout animals pointed to wider regulatory networks that were likely associated with SRGN’s role within the tumour microenvironment.

The pattern and magnitude of the transcriptional changes imply that the silencing of SRGN may interfere with the crosstalk between tumour and stroma and perhaps blunts critical pathways that promote tumour progression. This interpretation is consistent with earlier observations where SRGN controlled the storage of chemokines and inflammatory mediators, aided in the migration of leukocytes toward tissues, and participated in the remodelling of the extracellular matrix during inflammation and cancer (REF). The co-regulation of certain genes with SRGN, especially those exhibiting robust positive or negative expression associations, reinforces the notion that SRGN may control the expression of some genes directly, as part of an upstream regulatory network.

Noteworthy is the fact that the SRGN involvement in the shedding and expression of some other genes in the knockout model correlates functionally with several gene sets revealed in the human dataset analysis, including ECM and immune system-related genes. This cross-species comparability enhances the biological significance of SRGN as a putative tumour microenvironment architect.

5.6 CRISPRi-Mediated Repression of SRGN

5.6.1 Influence of Plasmid Architecture on CRISPRi Functionality

CRISPRi-mediated repression is dependent on plasmid architecture. Plasmid #73499, containing a strong CAG promoter, showed higher expression of genes in the plasmid and suppression of SRGN compared to plasmid #112195 which had a CMV promoter that had relatively low expression of plasmid genes. Promoter cassette and dCas9-KRAB fusion in #73499 are likely contributing factors

to stable nuclear retention and persistent transcriptional activity. That could be the reason for the enhanced SRGN repression observed in the CRISPI-G-2 clone.

5.6.2 Guide RNA-Specific Repression Patterns and Interpretation

Although CRISPI-G1 is located closer to the transcription start site (TSS), the region generally associated with higher repression efficiency in CRISPRi systems, **CRISPI-G-2 surprisingly produced stronger knockdown**. This suggests that the gRNA targeting Exon 2 may have had more effective chromatin accessibility, reduced off-target effects, or better recruitment of the KRAB-dCas9 complex in the canine cancer context. Alternatively, differences in guide sequence efficiency or local epigenetic landscape may have contributed to this unexpected pattern.

In some conditions, especially those with non-targeting or poorly placed gRNA, SRGN expression was either un-repressed or even elevated. Notably, the CRISPI-C line without a guide RNA had greater SRGN expression relative to wild-type cells; however, this may be due to clonal heterogeneity within the pleiomorphic CMT-U27 population or selection bias during G418 screening, instead of the impact of unanchored CRISPRi elements.

5.6.3 Statistical and Biological Interpretation of Results

While qPCR data supported the successful expression and integration of dCas9-KRAB, statistical comparisons were constrained by the limited number of biological replicates in some groups.

The most significant SRGN repression was observed in CRISPI-G (0.70×) and CRISPI-G-2 (0.39×) in comparison to internal controls. Expression in CRISPI-C was higher than in wild-type giving an expression change of 1.65 which is suggestive of clonal variation or selection bias during stable integration rather than a direct effect of the CRISPRi system itself. KRAB-G, using plasmid #112195, also showed an increase (FC = 3.58), suggesting variable suppression or compensatory upregulation.

Statistical tests ANOVA and Kruskal–Wallis did not achieve any significance ($p > 0.05$) due to low replicate numbers ($n = 2–4$). These bounds, however, do not impact the functional validation of the CRISPRi framework how the results are achieved. Rather, they point to the need for a broader dataset, quantification at the protein level (e.g. western blot), and phenotypic measurement such as migration or cytokine release.

5.7 Limitations and Technical Challenges

At an experimental scale, G418 selection did enrich for transfected CMT-U27 cells, but some variability in SRGN expression across clones was noted. For instance, the CRISPI-C line, which does not contain a guide RNA, surprisingly exhibited higher SRGN expression than the wild-type. Clonal variability, integration plasmid off-target effects, or even stochastic biological noise could explain this. A broader array of gRNAs alongside clonal duplicates would enhance paradigm specificity and reliability of repression.

Another limitation of this study is that all qPCR measurements were conducted at 48 h post-transfection. This time point was chosen based on reported peaks of CRISPRi activity by Yu et al. (2016) in HEK293FT cells and Kosicki et al. (2017) for plasmid-based delivery. However, Lin et al. (2021) observed detectable effects as early as 2.5 h following RNP lipofection, and different cell types may exhibit distinct kinetics. Future studies should include a comprehensive time-course analysis (e.g., 24, 48, 72, and 96 h) to define the onset and durability of dCas9 KRAB/gRNA-mediated knockdown as cells proliferate and dilute the effector complex.

In addition, identification of dCas9-KRAB at the protein level was not achieved in this study. Although functional SRGN repression occurred in our CMT U27 CRISPRi cell line, suggesting that the construct is present, protein verification (e.g., Western or immuno-staining) step should be performed as a confirmatory test. Several phenotypic assays of SRGN repression at a downstream

level (for example, migration, invasion, cytokine secretion assays) should also be conducted. The time constraints to perform this study is the main reason why this was not performed, and the interpretation of the data relies exclusively on transcriptional levels.

As the level of CRISPR interference relies on the position of the guide RNA and the chromatin context, modifications to guide RNA design in subsequent experiments could be beneficial. The addition of multiple gRNA could enhance repression and lower clonal variability. Regardless of these setbacks, the key aims of the research were accomplished: transcriptional SRGN suppression was achieved in CMT-U27 cells, and the CRISPRi system was confirmed as a robust and operational framework for conducting gene interference experiments in dog cancer cell models.

6 Conclusion

In this thesis, differential gene expression analysis and the involvement of serglycin in cancer were examined using integrated transcriptomics, along with functional gene silencing in a dog cancer cell line model. This study, combining available datasets with experimental validation, identified SRGN as a gene of interest due to its association with aggressive tumours.

A significant milestone in the study was the successful establishment of the CRISPR interference (CRISPRi) based platform in CMT-U27 canine mammary tumour cell line. The system relies on a dCas9-KRAB fusion and permits stable transcriptional repression of SRGN, controlled by guide RNAs. It was functionally validated via stringent antibiotic selection, morphological stasis, and quantitative PCR, confirming reproducible knockdown across numerous clones.

The CRISPRi system characterised in this study is not limited to SRGN. It functions as a gene suppression apparatus that is reusable and expandable and offer a versatile platform for the investigation of other candidate genes in canine mammary tumours as well as other cancer models. The ability to stably, reversibly, and non-mutagenically shut down transcriptional activity makes this tool valuable for various functional studies involving, but not limited to, basic and applied research, prospective therapeutic target validation, and gene network elucidation.

6.1 Future Directions

This platform may be enhanced in multiple ways in future work. First, the phenotypic consequences of SRGN suppression should be evaluated with functional assays such as migration, invasion, and cytokine profiling. Then, transcriptomic analysis of CRISPRi-repressed clones can identify SRGN-associated gene networks, potentially suitable for SRGN-controlled intervention.

The CRISPRi-dCas9-KRAB system developed here can also be modified to span a variety of targets beyond SRGN in the CMT-U27 cell line or other spontaneous canine tumour models, where multiple guide RNAs can be added to enable multiplexed gene suppression and combinatorial gene regulation studies. As a research tool, this system provides the capability for reproducible and sustained gene regulation investigations in comparative oncology and SRGN-controlled oncogenesis, providing cross-species ramifications between veterinary and human oncology.

7 References.

1. Abdelmegeed, S., & Mohammed, S. (2018). Canine mammary tumors as a model for human disease (Review). *Oncology Letters*. https://doi.org/10.3892/ol.2018.8411
2. Akbuğa, J., Özbaş-Turan, S., & Erdoğan, N. (2004). Plasmid-DNA loaded chitosan microspheres for in vitro IL-2 expression. *European Journal of Pharmaceutics and Biopharmaceutics*, *58*(3), 501–507. https://doi.org/10.1016/j.ejpb.2004.04.015
3. Chancham, P., & Hughes, J. A. (2001). RELATIONSHIP BETWEEN PLASMID DNA TOPOLOGICAL FORMS AND IN VITRO TRANSFECTION. *Journal of Liposome Research*, *11*(2–3), 139–152. https://doi.org/10.1081/LPR-100108458
4. Chao, M. P., Weissman, I. L., & Majeti, R. (2012). The CD47–SIRPα pathway in cancer immune evasion and potential therapeutic implications. *Current Opinion in Immunology*, *24*(2), 225–232. https://doi.org/10.1016/j.coi.2012.01.010
5. Escoffre, J.-M., Nikolova, B., Mallet, L., Henri, J., Favard, C., Golzio, M., Teissie, J., Tsoneva, I., & Rols, M.-P. (2012). New Insights in the Gene Electrotransfer Process: Evidence for the Involvement of the Plasmid DNA Topology. *Current Gene Therapy*, *12*(5), 417–422. https://doi.org/10.2174/156652312802762554
6. Ferreira, T., Miranda, M., Pinto-Leite, R., Mano, J. F., Medeiros, R., Oliveira, P. A., & Gama, A. (2024). Integrated Study of Canine Mammary Tumors Histopathology, Immunohistochemistry, and Cytogenetic Findings. *Veterinary Sciences*, *11*(9), 409. https://doi.org/10.3390/vetsci11090409
7. Gherman, L.-M., Chiroi, P., Nuţu, A., Bica, C., & Berindan-Neagoe, I. (2024). Profiling canine mammary tumors: A potential model for studying human breast cancer. *The Veterinary Journal*, *303*, 106055. https://doi.org/10.1016/j.tvjl.2023.106055
8. Graim, K., Gorenshteyn, D., Robinson, D. G., Carriero, N. J., Cahill, J., Chakrabarti, R., Goldschmidt, M. H., Durham, A. C., Funk, J., Storey, J. D., Kristensen, V. N., Theesfeld, C. L., Sorenmo, K. U., & Troyanskaya, O. G. (2020). Abstract 2504: Modeling molecular development of breast cancer in canine mammary tumors. *Cancer Research*, *80*(16_Supplement), 2504–2504. https://doi.org/10.1158/1538-7445.AM2020-2504
9. Guo, J.-Y., Chiu, C.-H., Wang, M.-J., Li, F.-A., & Chen, J.-Y. (2020). Proteoglycan serglycin promotes non-small cell lung cancer cell migration through the interaction of its glycosaminoglycans with CD44. *Journal of Biomedical Science*, *27*(1), 2. https://doi.org/10.1186/s12929-019-0600-3

10. Guo, J.-Y., Hsu, H.-S., Tyan, S.-W., Li, F.-Y., Shew, J.-Y., Lee, W.-H., & Chen, J.-Y. (2017). Serglycin in tumor microenvironment promotes non-small cell lung cancer aggressiveness in a CD44-dependent manner. *Oncogene*, *36*(17), 2457–2471. https://doi.org/10.1038/onc.2016.404
11. Kim, T.-M., Yang, I. S., Seung, B.-J., Lee, S., Kim, D., Ha, Y.-J., Seo, M., Kim, K.-K., Kim, H. S., Cheong, J.-H., Sur, J.-H., Nam, H., & Kim, S. (2020). Cross-species oncogenic signatures of breast cancer in canine mammary tumors. *Nature Communications*, *11*(1), 3616. https://doi.org/10.1038/s41467-020-17458-0
12. Larson, M. H., Gilbert, L. A., Wang, X., Lim, W. A., Weissman, J. S., & Qi, L. S. (2013). CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nature Protocols*, *8*(11), 2180–2196. https://doi.org/10.1038/nprot.2013.132
13. Lehner, R., Wang, X., & Hunziker, P. (2013). Plasmid linearization changes shape and efficiency of transfection complexes. *European Journal of Nanomedicine*, *5*(4). https://doi.org/10.1515/ejnm-2013-0028
14. Long, L., Guo, H., Yao, D., Xiong, K., Li, Y., Liu, P., Zhu, Z., & Liu, D. (2015). Regulation of transcriptionally active genes via the catalytically inactive Cas9 in C. elegans and D. rerio. *Cell Research*, *25*(5), 638–641. https://doi.org/10.1038/cr.2015.35
15. Lupo, A., Cesaro, E., Montano, G., Zurlo, D., Izzo, P., & Costanzo, P. (2013). KRAB-Zinc Finger Proteins: A Repressor Family Displaying Multiple Biological Functions. *Current Genomics*, *14*(4), 268–278. https://doi.org/10.2174/13892029113149990002
16. Mandegar, M.A., Huebsch, N., Frolov, E.B., Shin, E., Truong, A., Olvera, M.P., Chan, A.H., Miyaoka, Y., Holmes, K., Spencer, C.I., Judge, L.M., Gordon, D.E., Eskildsen, T.V., Villalta, J.E., Horlbeck, M.A., Gilbert, L.A., Krogan, N.J., Sheikh, S.P., Weissman, J.S., Qi, L.S., So, P.L. & Conklin, B.R., 2016. CRISPR interference efficiently induces specific and reversible gene silencing in human iPSCs. *Cell Stem Cell*, 18(4), pp. 541–553. Available at: https://doi.org/10.1016/j.stem.2016.01.022.
17. Margolin, J. F., Friedman, J. R., Meyer, W. K., Vissing, H., Thiesen, H. J., & Rauscher, F. J. (1994). Krüppel-associated boxes are potent transcriptional repression domains. *Proceedings of the National Academy of Sciences*, *91*(10), 4509–4513. https://doi.org/10.1073/pnas.91.10.4509
18. Merlo, D. F., Rossi, L., Pellegrino, C., Ceppi, M., Cardellino, U., Capurro, C., Ratto, A., Sambucco, P. L., Sestito, V., Tanara, G., & Bocchini, V. (2008). Cancer Incidence in Pet Dogs: Findings of the Animal Tumor Registry of Genoa, Italy. *Journal of Veterinary Internal Medicine*, *22*(4), 976–984. https://doi.org/10.1111/j.1939-1676.2008.0133.x
19. Mestrinho, L. A., & Santos, R. R. (2021). Translational oncotargets for immunotherapy: From pet dogs to humans. *Advanced Drug Delivery Reviews*, *172*, 296–313. https://doi.org/10.1016/j.addr.2021.02.020
20. O'Geen, H., Ren, C., Nicolet, C.M., Perez, A.A., Halmai, J., Le, V.M., Mackay, J.P., Farnham, P.J. & Segal, D.J., 2017. dCas9-based epigenome editing suggests acquisition of histone methylation is not sufficient for target gene repression. *Nucleic Acids Research*, 45(17), pp. 9901–9916. Available at: https://doi.org/10.1093/nar/gkx578.
21. Oh, J. H., & Cho, J.-Y. (2023). Comparative oncology: Overcoming human cancer through companion animal studies. *Experimental & Molecular Medicine*, *55*(4), 725–734. https://doi.org/10.1038/s12276-023-00977-3
22. Oliveira-Lopes, A. F., Götze, M. M., Lopes-Neto, B. E., Guerreiro, D. D., Bustamante-Filho, I. C., & Moura, A. A. (2024). Molecular and Pathobiology of Canine Mammary Tumour: Defining a Translational Model for Human Breast Cancer. *Veterinary and Comparative Oncology*, *22*(3), 340–358. https://doi.org/10.1111/vco.12996
23. Pawlik, T. M., Paulino, A. F., Mcginn, C. J., Baker, L. H., Cohen, D. S., Morris, J. S., Rees, R., & Sondak, V. K. (2003). Cutaneous angiosarcoma of the scalp: A multidisciplinary approach. *Cancer*, *98*(8), 1716–1726. https://doi.org/10.1002/cncr.11667
24. Printz, C. (2011). Pet animals with cancer help advance human cancer research: Similarities help to explore future human treatments. *Cancer*, *117*(21), 4807–4808. https://doi.org/10.1002/cncr.26604
25. Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. (2013). Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell*, *152*(5), 1173–1183. https://doi.org/10.1016/j.cell.2013.02.022
26. Queiroga, F. and Lopes, C. (2002). Tumores mamários caninos – novas perspectivas [Canine mammary tumors – new perspectives]. In: *Congresso de Ciências Veterinárias [Proceedings of the Veterinary Sciences Congress]*, SPCV, Oeiras, 10–12 October 2002, pp.183–190.

27. Schiffman, J. D., & Breen, M. (2015). Comparative oncology: What dogs and other species can teach us about humans with cancer. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1673), 20140231. https://doi.org/10.1098/rstb.2014.0231
28. Sleeckx, N., De Rooster, H., Veldhuis Kroeze, E., Van Ginneken, C., & Van Brantegem, L. (2011). Canine Mammary Tumours, an Overview. *Reproduction in Domestic Animals*, *46*(6), 1112–1131. https://doi.org/10.1111/j.1439-0531.2011.01816.x
29. Sultan, F., & Ganaie, B. A. (2018). Comparative oncology: Integrating human and veterinary medicine. *Open Veterinary Journal*, *8*(1), 25. https://doi.org/10.4314/ovj.v8i1.5
30. Tinsley, A. (2020). *Canine Hemangiosarcoma: A Certainly Less Than Ideal, Very Ugly Cancer*. https://doi.org/10.20944/preprints202008.0528.v1
31. Xu, Y., Xu, J., Yang, Y., Zhu, L., Li, X., & Zhao, W. (2018). SRGN Promotes Colorectal Cancer Metastasis as a Critical Downstream Target of HIF-1α. *Cellular Physiology and Biochemistry*, *48*(6), 2429–2440. https://doi.org/10.1159/000492657
32. Zhang, R., Xu, W., Shao, S., & Wang, Q. (2021). Gene Silencing Through CRISPR Interference in Bacteria: Current Advances and Future Prospects. *Frontiers in Microbiology*, *12*, 635227. https://doi.org/10.3389/fmicb.2021.635227
33. Zhang, Z., Deng, Y., Zheng, G., Jia, X., Xiong, Y., Luo, K., Qiu, Q., Qiu, N., Yin, J., Lu, M., Liu, H., Gu, Y., & He, Z. (2017). SRGN-TGFβ2 regulatory loop confers invasion and metastasis in triple-negative breast cancer. *Oncogenesis*, *6*(7), e360–e360. https://doi.org/10.1038/oncsis.2017.53
34. Korpetinou, A., Skandalis, S.S., Labropoulou, V.T., Smirlaki, G., Noulas, A., Karamanos, N.K. and Theocharis, A.D., 2014. Serglycin: at the crossroad of inflammation and malignancy. *Frontiers in Oncology*, 3, p.327. https://doi.org/10.3389/fonc.2013.00327
35. Chang JW, Kuo WH, Lin CM, Chen WL et al. Wild-type p53 upregulates an early onset breast cancer-associated gene GAS7 to suppress metastasis via GAS7-CYFIP1-mediated signaling pathway. *Oncogene* 2018 Jul;37(30):4137-4150. PMID: 29706651
36. Malvia S, Bagadi SAR, Pradhan D, Chintamani C et al. Study of Gene Expression Profiles of Breast Cancers in Indian Women. *Sci Rep* 2019 Jul 10;9(1):10018. PMID: 31292488
37. Colak D, Nofal A, Albakheet A, Nirmal M et al. Age-specific gene expression signatures for breast tumors and cross-species conserved potential cancer progression markers in young women. *PLoS One* 2013;8(5):e63204. PMID: 23704896
38. Clarke C, Madden SF, Doolan P, Aherne ST et al. Correlating transcriptional networks to breast cancer survival: a large-scale coexpression analysis. Carcinogenesis 2013 Oct;34(10):2300-8. PMID: 23740839
39. Krol, M., Pawlowski, K.M., Skierski, J., Rao, N.A.S., Hellmen, E., Mol, J.A. and Motyl, T., 2009. Transcriptomic profile of two canine mammary cancer cell lines with different proliferative and anti-apoptotic potential. Journal of Physiology and Pharmacology, 60(Suppl 1), pp.95–106. Available at: http://www.jpp.krakow.pl/journal/archive/01_09_s1/pdf/95_01_09_s1_article.pdf
40. Guil-Luna, S., Hellmén, E., Sánchez-Céspedes, R., Millán, Y. and Martín de las Mulas, J., 2014. The antiprogestins mifepristone and onapristone reduce cell proliferation in the canine mammary carcinoma cell line CMT-U27. Histology and Histopathology, 29, pp.949–955. https://doi.org/10.14670/HH-29.949
41. Ustun Alkan, F., Ustuner, O., Bakirel, T., Cinar, S., Erten, G. and Deniz, G., 2012. The effects of piroxicam and deracoxib on canine mammary tumour cell line. The Scientific World Journal, 2012, Article ID 976740, 8 pages. https://doi.org/10.1100/2012/976740
42. Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., Lim, W.A., Weissman, J.S. and Qi, L.S., 2013. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, 154(2), pp.442–451. https://doi.org/10.1016/j.cell.2013.06.044
43. Li, Z., Xiong, X. and Li, J.-F., 2019. The working dead: repurposing inactive CRISPR-associated nucleases as programmable transcriptional regulators in plants. aBIOTECH, Available at: https://doi.org/10.1007/s42994-019-00003-z
44. Roy, A., Femel, J., Huijbers, E.J.M., Spillmann, D., Larsson, E., Ringvall, M., Olsson, A.-K. and Åbrink, M., 2016. Targeting Serglycin Prevents Metastasis in Murine Mammary Carcinoma. *PLOS ONE*, 11(5), p.e0156151. https://doi.org/10.1371/journal.pone.0156151
45. Cardarelli, F., Digiacomo, L., Marchini, C., Amici, A., Salomone, F., Fiume, G., Rossetta, A., Gratton, E., Pozzi, D. and Caracciolo, G., 2016. *The intracellular trafficking mechanism of Lipofectamine-based transfection reagents and its implication for gene delivery*. Scientific Reports, 6, p.25879. https://doi.org/10.1038/srep25879

46. Yu, X., Liang, X., Xie, H., Kumar, S., Ravinder, N., Potter, J., de Mollerat du Jeu, X. & Chesnut, J.D., 2016. Improved delivery of Cas9 protein/gRNA complexes using lipofectamine CRISPRMAX. *Biotechnology Letters*, 38(6), pp.919–929. doi:10.1007/s10529-016-2064-9.
47. Kosicki, M., Rajan, S.S., Lorenzetti, F.C., Wandall, H.H., Narimatsu, Y., Metzakopian, E. & Bennett, E.P. (2017) 'Dynamics of indel profiles induced by various CRISPR/Cas9 delivery methods', Progress in Molecular Biology and Translational Science, 152, pp. 49–67. doi: 10.1016/bs.pmbts.2017.09.003.
48. Lin, Q., Le, Q.A., Takebayashi, K., Thongkittidilok, C., Wittayarat, M., Hirata, M., Tanihara, F. & Otoi, T. (2021) 'Timing and duration of lipofection-mediated CRISPR/Cas9 delivery into porcine zygotes affect gene-editing events', BMC Research Notes, 14, 389. doi: 10.1186/s13104-021-05800-8.
49. Linder, C. (2022) Serglycin as a potential diagnostic biomarker for hemangiosarcoma in dogs. Second cycle, A2E. Uppsala: Swedish University of Agricultural Sciences, Department of Clinical Sciences. Epsilon Archive for Student Projects. Available at: https://stud.epsilon.slu.se/17749/
50. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E., 2012. A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337(6096), pp.816–821. https://doi.org/10.1126/science.1225829
51. Kolset, S.O. and Pejler, G., 2011. Serglycin: a structural and functional chameleon with wide impact on immune cells. *Journal of Immunology*, 187(10), pp.4927–4933. https://doi.org/10.4049/jimmunol.1100806
52. Kluyver, T., Ragan-Kelley, B., Pérez, F., et al., 2016. Jupyter Notebooks – a publishing format for reproducible computational workflows. In: Loizides, F., and Scmidt, B. (eds.) Positioning and Power in Academic Publishing: Players, Agents and Agendas. IOS Press, pp. 87–90.
53. Asosingh, K., Günthert, U., De Raeve, H., Van Riet, I., Van Camp, B. and Vanderkerken, K., 2001. A unique pathway in the homing of murine multiple myeloma cells: CD44v10 mediates binding to bone-marrow endothelium. *Cancer Research*, 61(7), pp.2862–2865.
54. Imbeault, M., Helleboid, P.Y. and Trono, D., 2017. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature*, 543(7646), pp.550–554. https://doi.org/10.1038/nature21683

Popular science summary

Dogs and humans share many similarities when it comes to cancer. In this project, I used a genetic tool called CRISPRi to "switch off" a specific gene, *Serglycin (SRGN)*, in cancer cells from dogs. This gene is linked to inflammation and cancer spread. Before doing the lab work, I analysed a large public dataset from mice and an unpublished dataset from humans to understand how this gene behaves in different cancers. Then, I created dog cancer cells with a system that can block SRGN without permanently changing the DNA. This combination of computer-based analysis and lab experiments helps us understand how specific genes drive cancer. My goal is to support future cancer research and treatments, both for animals and people, through this shared knowledge.

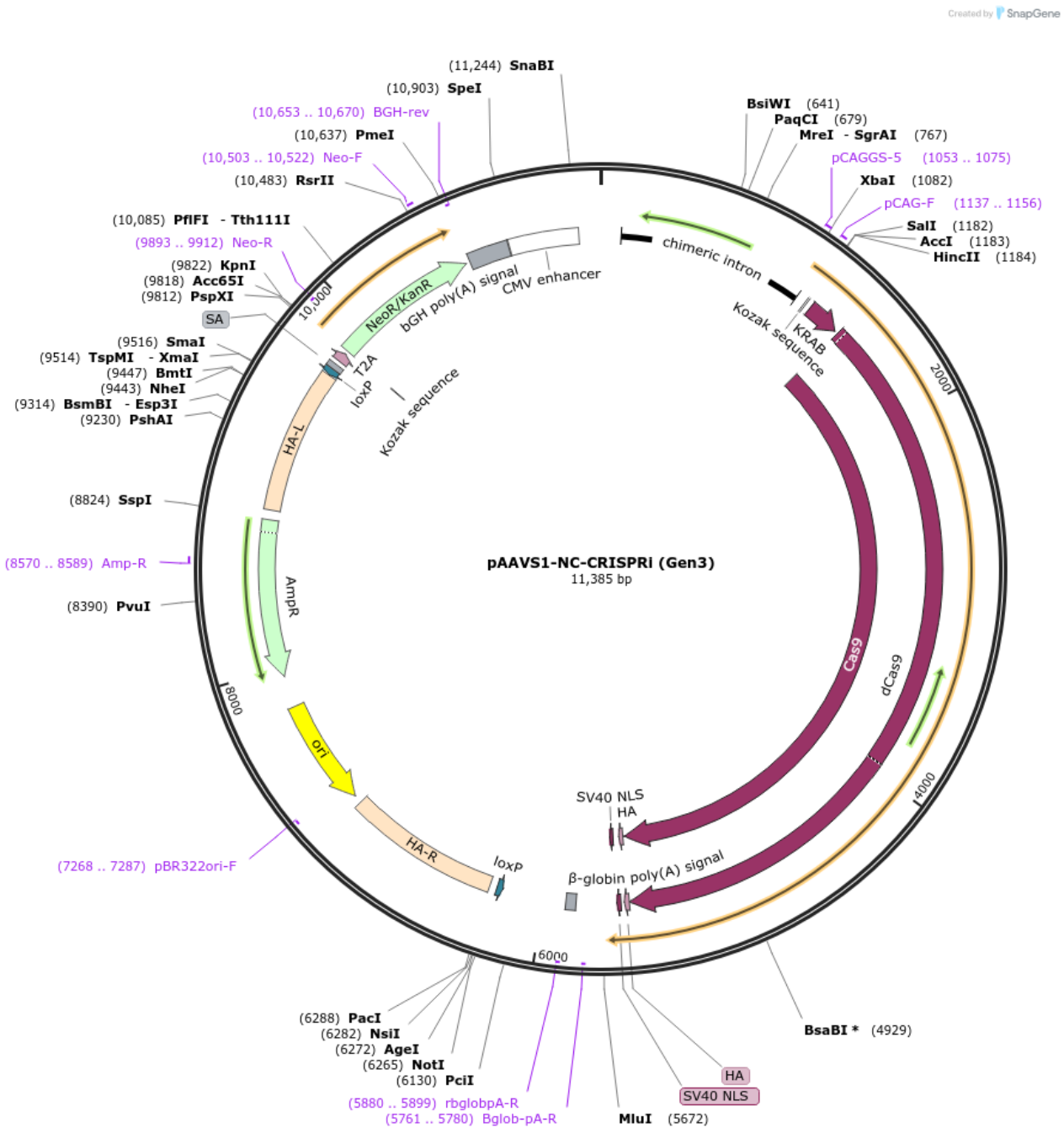
Appendix 1

code used for data analysis

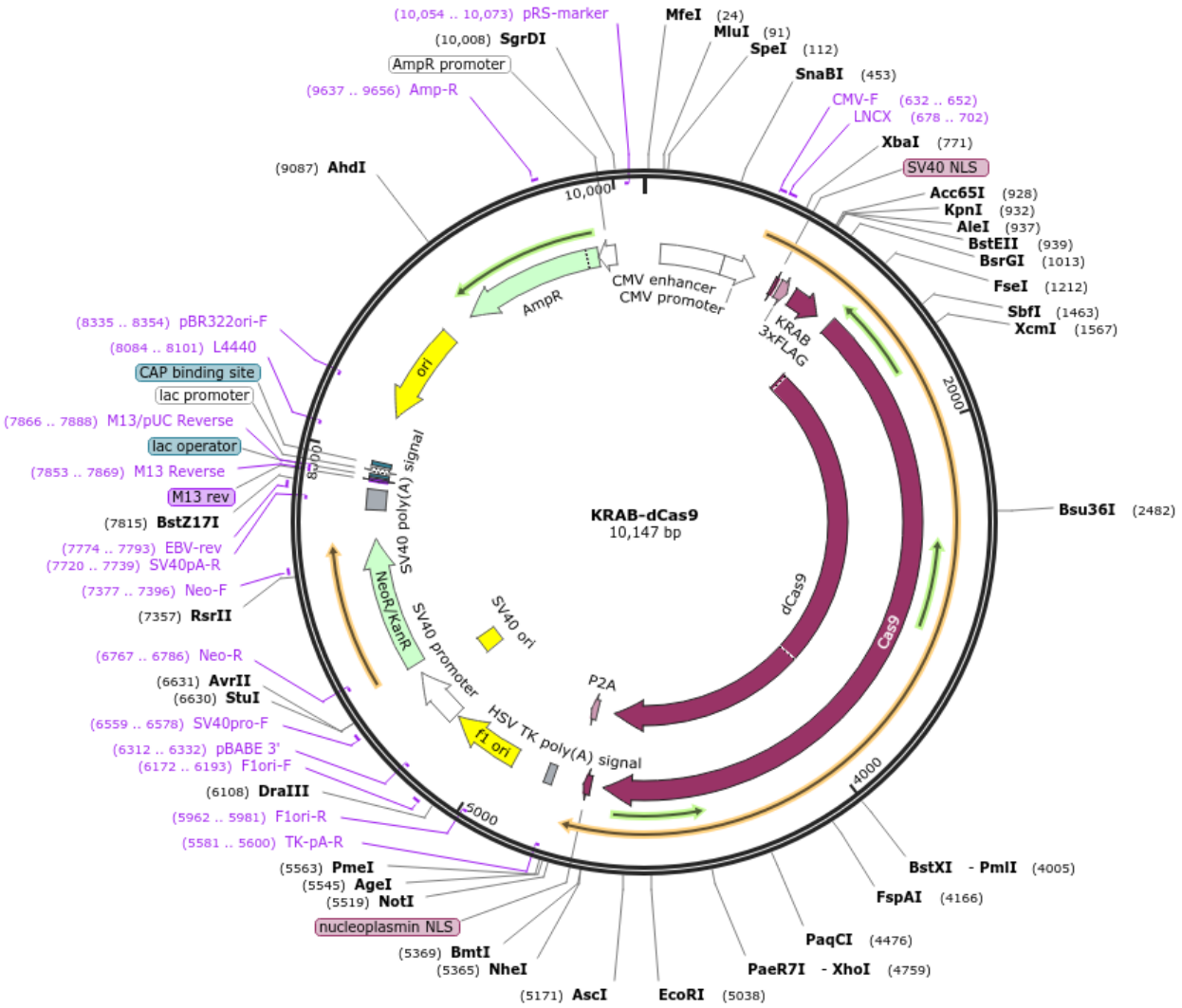
All codes used for data analysis have been stored on github suppository and can be accessed via following link:

https://github.com/ruwinimadhu/Bioinformatics_thesis

Appendix 2 -Plasmid #73499



Plasmid #112195



Appendix 3 – Cell transferring Protocol

1. remove the medium in the T25 flask.
2. Put 5 ml of PBS. tilt it carefully and discard it.
3. repeat step 2.
4. Add 500 mcl of trypsinise enzyme into the T-25 flask. Tilt carefully and incubate it at 37C for 5 min.
5. Observe under microscope. - These steps are conducted to detach cells from the surface. (T-25 flask) if the majority of cells are not detached from the surface these steps can be repeated.
6. Give a gentle tap to detach cells.
7. Add 5 ml of growth medium into T- 25 flack. mix them with up and down pipette movements.
8. take 5ml of solution and add into a 15ml conical tube. Centrifuge it at 11000 for 5 min.
9. in the same time, add 5ml of growth medium into to T - 25 flask and let cells grow again. (incubate in 37C and 5% CO2)
10. Discard the supernatant in the conical tube after centrifuge. add 12 ml of growth medium to the pellet and dissolve it using up and down pipette movements.
11. take 6 wells plate and add 2ml of cell-containing media into each well.

Appendix 4 - CRISPRMAX™ Transfection Protocol

Step	Action	Component	Volume (6-well)
Day 0	**1. Seed cells (30–70% confluency at transfection)**	Adherent cells	2.5–4.5 × 10^5 cells
Day 1	**2. Prepare Tube 1: Cas9 + gRNA mix (Mix well)**	Opti-MEM™ I Medium (Gibco)	125 µL
		Plasmid (dCas9-KRAB)	6250 ng
		gRNA (synthetic, previously ested inhouse gRNA used by Sofia Tengstrand)	1200 ng
		Cas9 Plus™ Reagent (**add last**)	12.5 µL
	3. Prepare Tube 2: Diluted CRISPRMAX™ (Mix well) **(Do not let sit >3 min)**	Opti-MEM™ I Medium	125 µL
		CRISPRMAX™ Reagent (Invitrogen, Thermo Fisher Scientific, USA)	7.5 µL
	4. Combine Tube 1 → Tube 2 (mix well immediately)	Final mix: Cas9/gRNA/transfection mix	
	5. Incubate at room temp (Do not exceed 30 min)		5–10 minutes
	6. Add complex to cells	Transfection complex	250 µL

Transfection of the plasmid is conducted without gRNA.

Transfection of gRNA is conducted without plasmid.

Appendix 5: cDNA Methodology (Linder, C. (2022))

Step A1 - Preparing DNase treatment mix for each RNA sample

Agent	Volume (µL)
10× DNase I Buffer	2.2
Nuclease free water	V
DNase I (Invitrogen AMPD1, don't vortex)	2.2
Total Volume	22

V = Amount to be added to give 1 µg RNA concentration in next step

Step A2 – Preparing cDNA master mix 1

Agent	Volume (µL)
dNTPs mix 10 µM	1.2
Random Hexamer	1
Oligo DT diluted 1:10	0.2
Total	2.4

Step A3 – Preparing cDNA master mix 2

Agent	RT+ (µL)
5× SSIV buffer	4
0.1 M DTT	1
Superscript 4 RT (200 U/µL)	1
Nuclease free water	1
Total	7

Combining cDNA and RNA

Step B1 – DNase treatment mix for each sample

Agent	Volume (µL)
RNA	V*
DNase treatment mix	22 – V

Total	22 µL

V = Calculated correct amount to give desired RNA concentration and amount.

- Mix gently by pipetting up and down, incubate at room temperature for 15 min.
- Add 2.2 µL DNase stop (EDTA).
- Now the total treated RNA volume will be 24.2 µL.
- Incubate at 70°C for 10 min, chill on ice for 5 min, and then spin down.

Step B2 – cDNA master mix 1 for each RNA

Agent	Volume (µL)
RNA treated	24.2
Master mix	2.4
Total	26.6

- Incubate at 59°C for 5 min and put on ice for 1 min.

Step B3 – cDNA master mix 2 for each reaction

Agent	RT+ (µl)
Master mix 2	7
Template RNA treated	26.6
Total	33.6

- Incubate at 25°C or room temperature for 15 min
- Incubate at 55°C for 60 min
- Inactivate reaction at 70°C for 15 min
- Dilute samples 1:5 with nuclease-free water
- Finished product can be stored in -20°C freezer

Appendix 6: qPCR Methodology (Linder, C. (2022))

- Volume per well: 25 µL total
- Vessel: Bio-Rad white plate
- Seal: Bio-Rad Seal
- Samples: synthesized cDNA 1:5 dilution with RNAse free water

Step 1: Master Mix Preparation

Agent	**Volume per reaction (µL)**	**Volume for duplicates +2 (µL)**
SsoAdvanced™ Universal SYBR® Green Supermix (Bio-Rad, USA)	12.5	25
Primer forward	1	2
Primer reverse	1	2
Nuclease free water	8.5	17
Total	23	46

Step 2: Sample and Master Mix Combining

- Add 2 µL sample per well
- Add 23 µL master mix per well

Step 3: Perform qPCR

- Centrifuge down sample plate for 1000 XG for 2 min
- qPCR was conducted using BioRed CFX Maestero machine.
- Thermocycle

Steps	Cycle	Temperature	Time
Denature	1	95 C	5 min
Amplification	39	95 C	15 sec
		60 C	30 sec
		72 C	30 sec Collect SYBR green
Melt Curve	60 C, 0.5 C steps to 95 C, 10 sec dwell time		

Appendix 6.1: qPCR primers for Plasmids recognition

Plasmid #112195

	Sequence (5'->3')	Length	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	AAGCCAGATGTGATCCTCCG	20	59.53	55.00	4.00	2.00
Reverse primer	GAGTACTTCTTGTCCATTGGCC	22	58.99	50.00	6.00	5.00

Plasmid #73499 -Primer pair 1

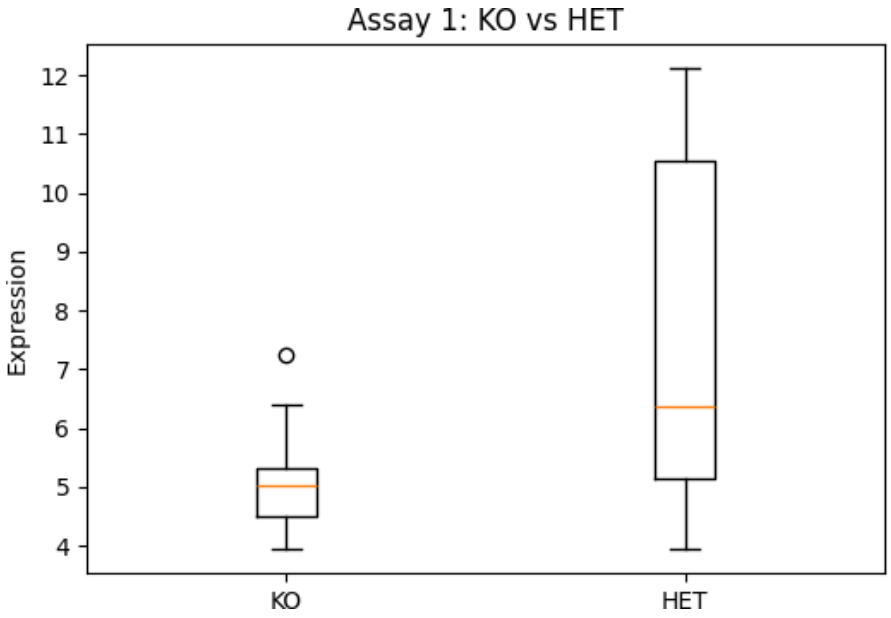
	Sequence (5'->3')	Length	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	ACTAAGCCAGATGTGATCCTCC	22	59.30	50.00	4.00	0.00
Reverse primer	TGCTGTACTTCTTGTCCATCGA	22	59.44	45.45	4.00	4.00

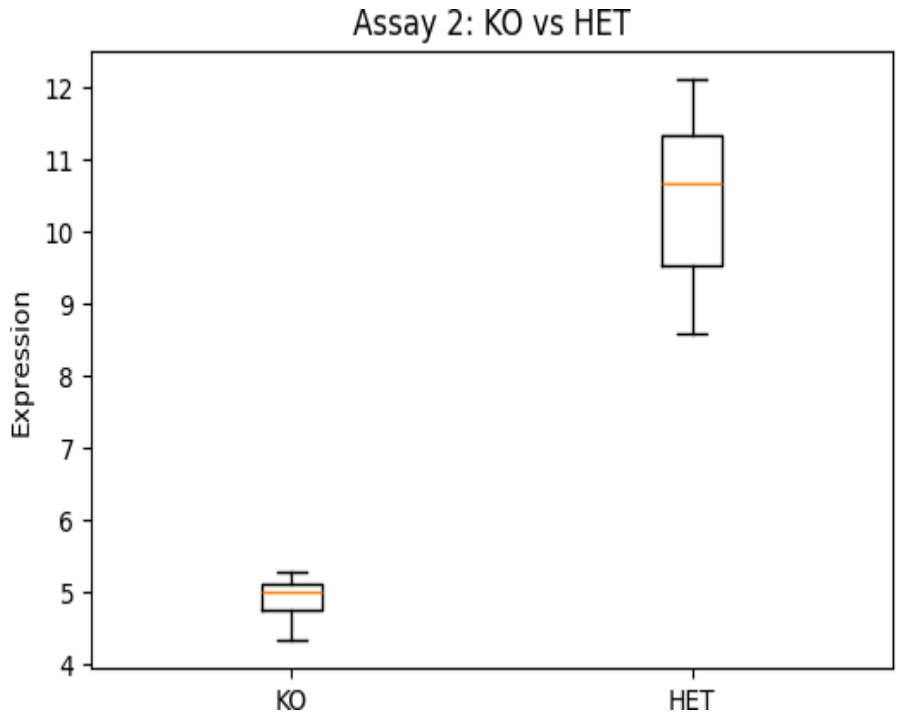
Primer pair 2

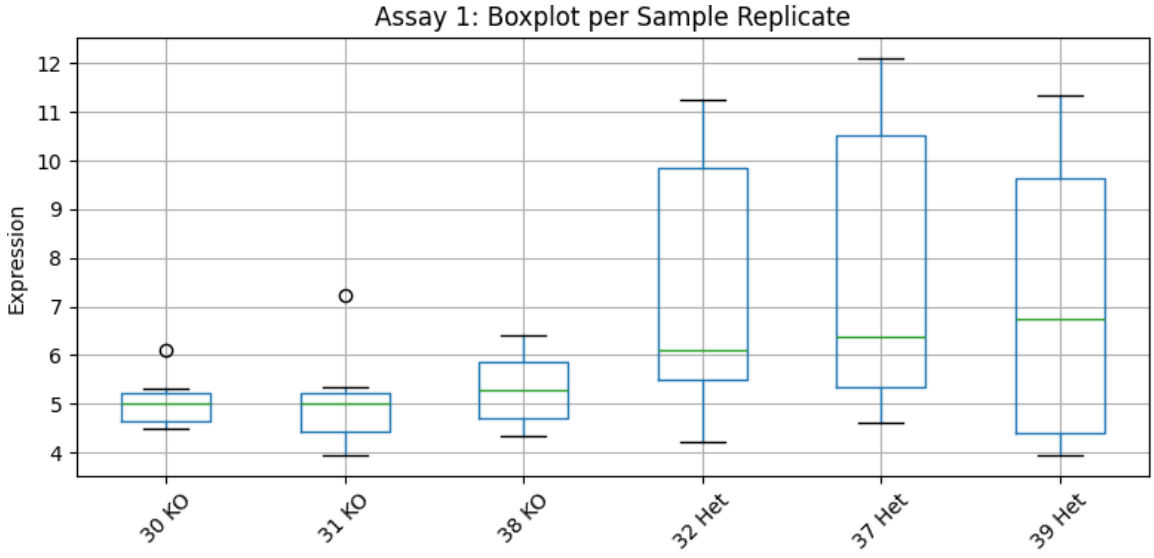
	Sequence (5'->3')	Length	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	ACTAAGCCAGATGTGATCCTCC	22	59.30	50.00	4.00	0.00
Reverse primer	GTTCTTCTTGATGCTGTGCCG	21	60.14	52.38	2.00	2.00

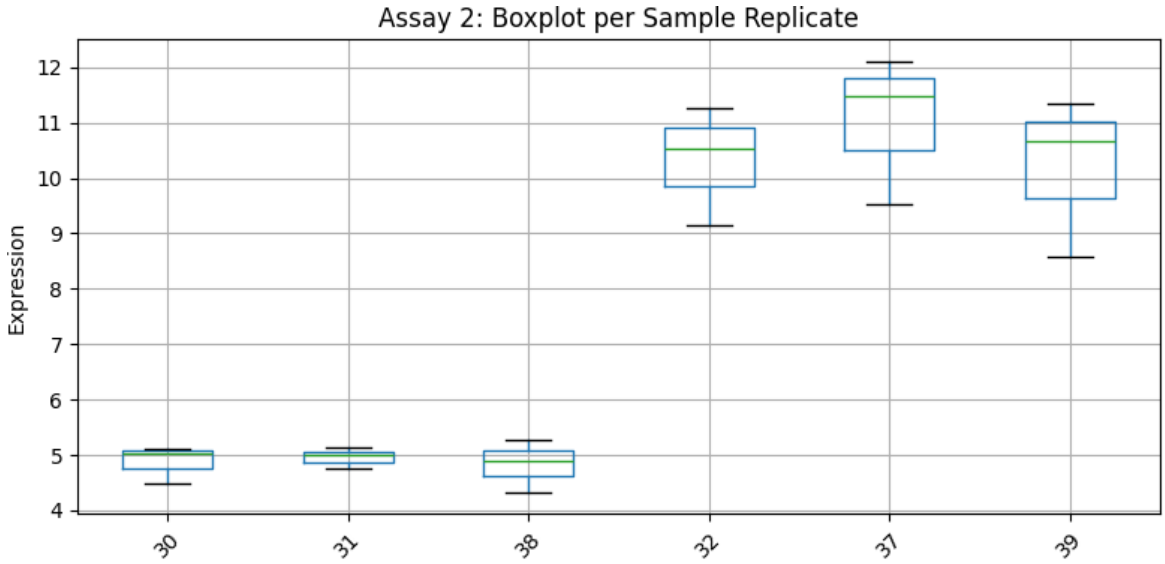
Appendix 7: GSE67806

```
SRGN Expression Summary:
        Assay  KO mean  HET mean        t-stat      p-value
        Assay 1 5.145195  7.483795  -3.582040 1.567880e-03
        Assay 2 4.891293 10.523342 -13.748012 2.155009e-0
```



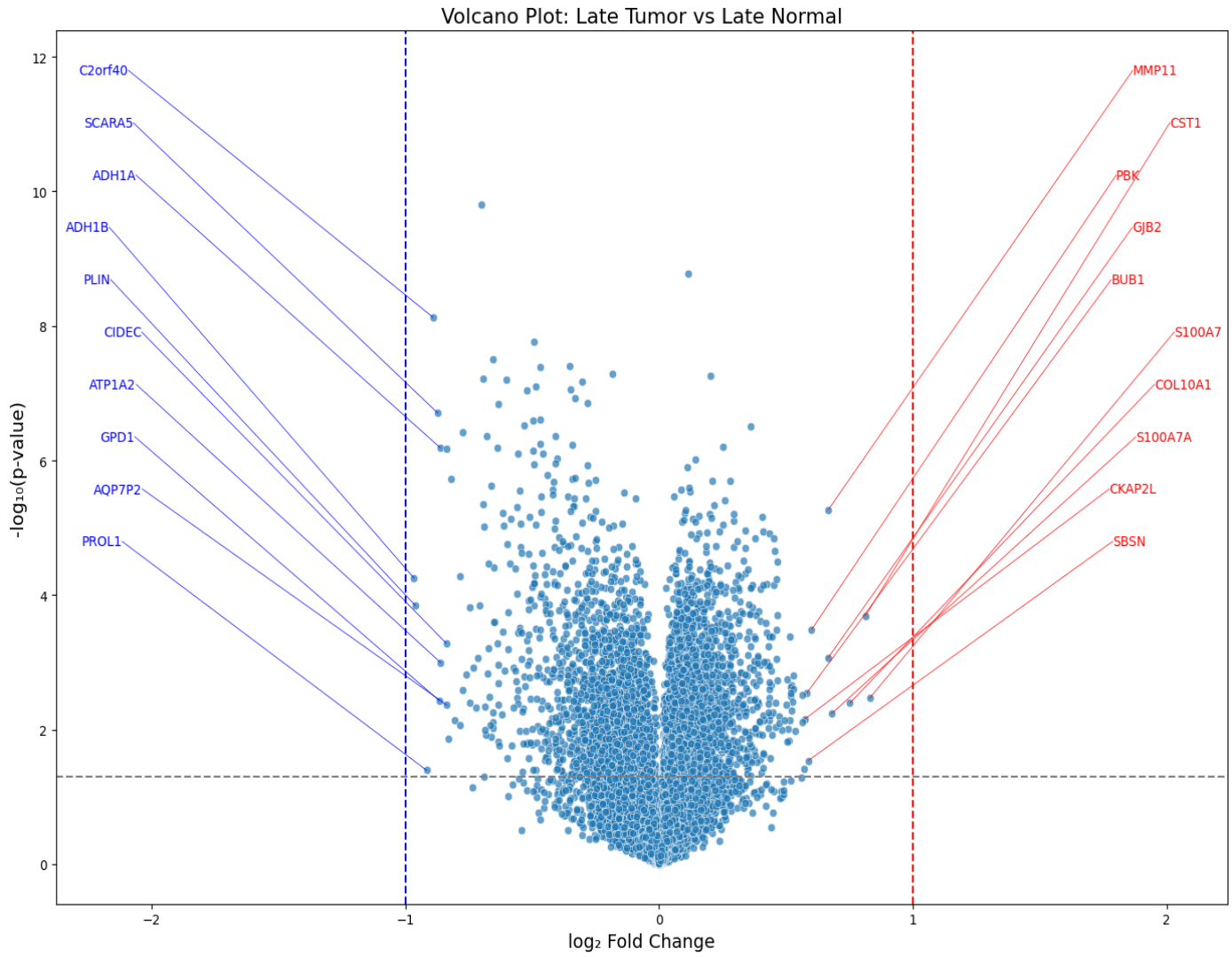


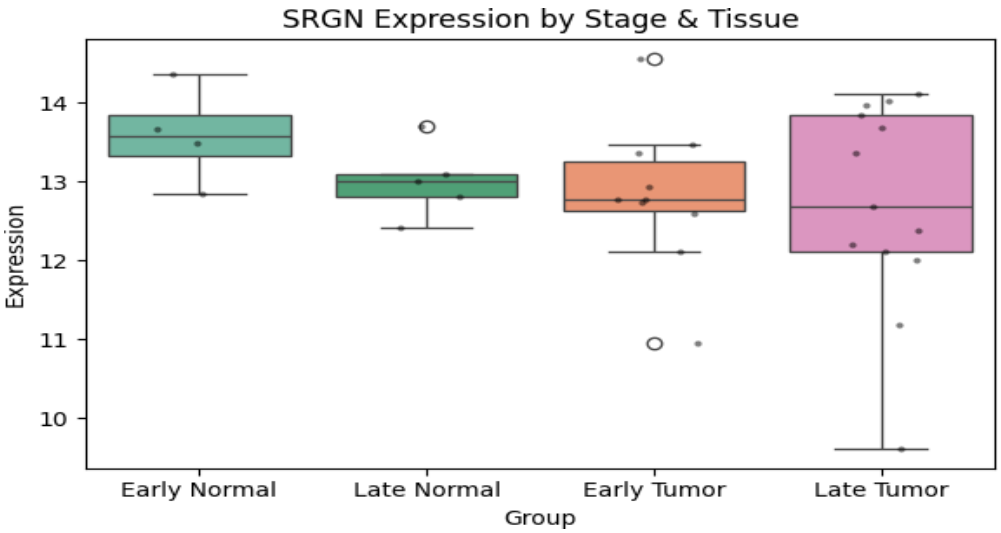

Assay 1: Boxplot per Sample Replicate
Expression
12
11
10
9
8
7
6
5
4
30 KO
31 KO
38 KO
32 Het
37 Het
39 Het

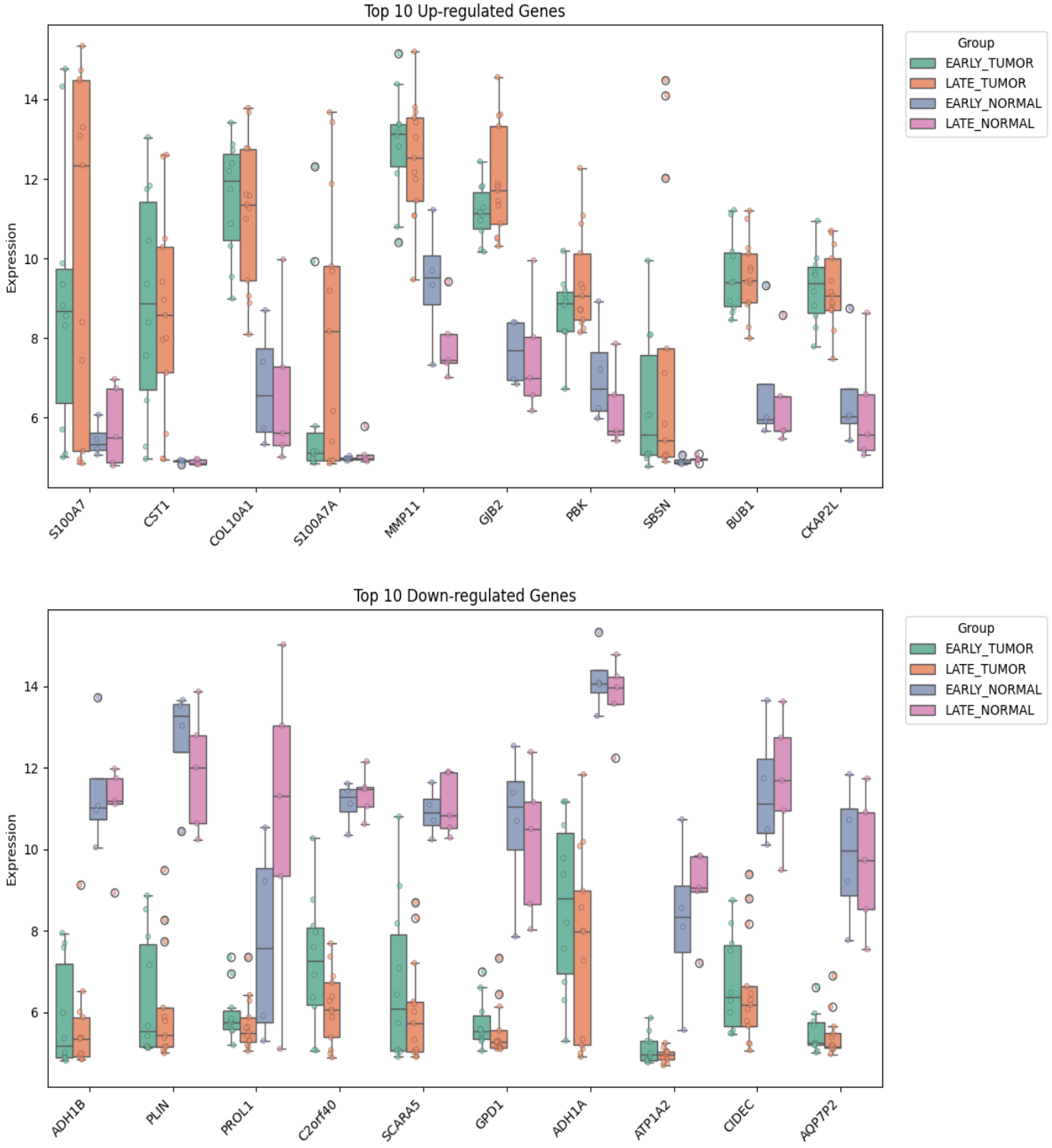

Assay 2: Boxplot per Sample Replicate
Expression
12
11
10
9
8
7
6
5
4
30
31
38
32
37
39

Appendix 8: GSE89116

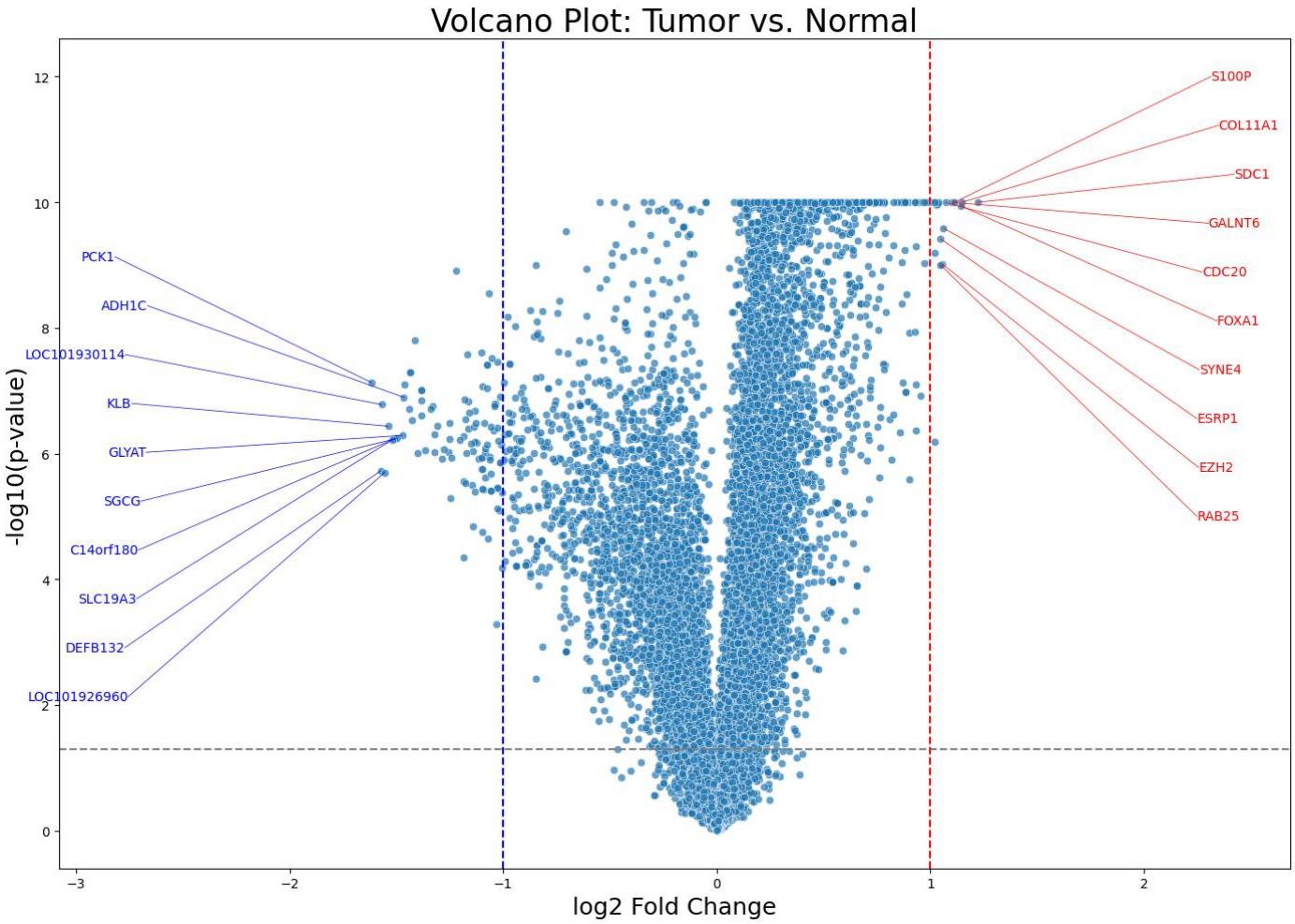
```
Early tumour: 10 Late tumour: 13 Early normal: 4 Late normal: 5
```

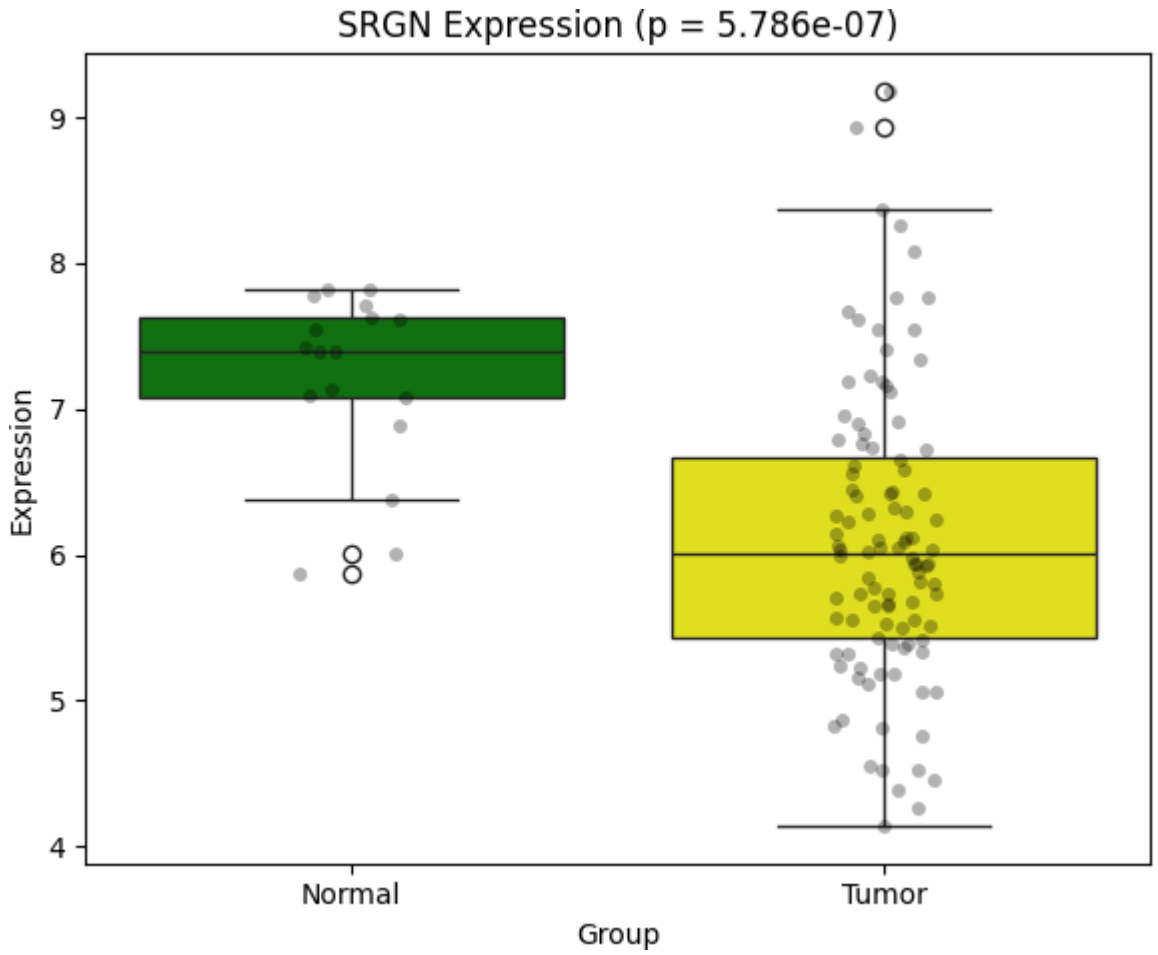


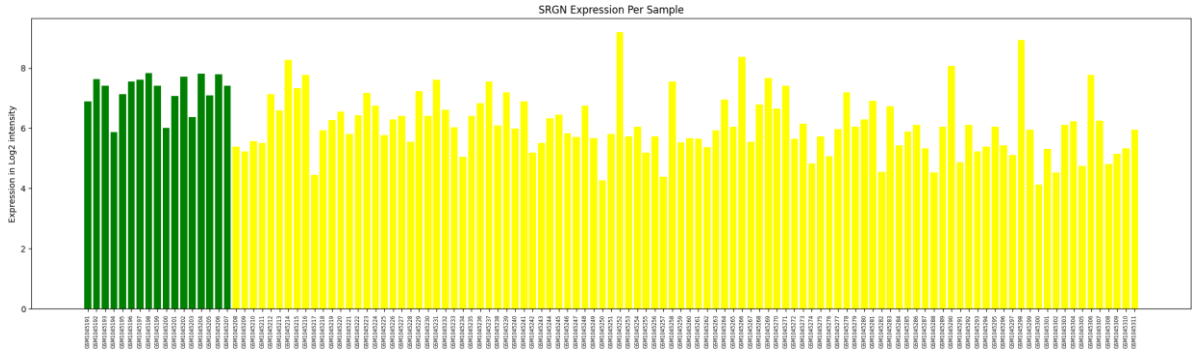



Top 10 Up-regulated Genes
Expression
Group
EARLY_TUMOR
LATE_TUMOR
EARLY_NORMAL
LATE_NORMAL
S100A7
CST1
COL10A1
S100A7A
MMP11
GJB2
PBK
SBSN
BUB1
CKAP2L
Top 10 Down-regulated Genes
Expression
Group
EARLY_TUMOR
LATE_TUMOR
EARLY_NORMAL
LATE_NORMAL
ADH1B
PLIN
PROL1
C2orf40
SCARA5
GPD1
ADH1A
ATP1A2
CIDEC
AQP7P2

Appendix 9: GSE42568




SRGN Expression (p = 5.786e-07)
Expression
9
8
7
6
5
4
Normal
Tumor
Group


SRGN Expression Per Sample
Expression in Log2 intensity
8
6
4
2
0

Top 10 Upregulated Genes:

Gene	log2FC	p_value
SDC1	1.222234	9.648563e-16
COL11A1	1.148206	8.838002e-17
FOXA1	1.141988	1.324632e-11
S100P	1.113816	1.235952e-19
GALNT6	1.101452	1.507053e-14
CDC20	1.071860	1.026745e-12
SYNE4	1.059639	1.568445e-10
EZH2	1.056040	8.615546e-10
ESRP1	1.048060	2.816567e-10
RAB25	1.047515	8.928803e-10

Top 10 Downregulated Genes:

Gene	log2FC	p_value
PCK1	-1.617203	7.384604e-08
DEFB132	-1.571821	1.873079e-06
LOC101930114	-1.568322	1.660221e-07
LOC101926960	-1.555296	2.016997e-06
KLB	-1.538618	3.668836e-07
SLC19A3	-1.517362	5.984505e-07
C14orf180	-1.508267	5.798532e-07
SGCG	-1.495885	5.721884e-07
GLYAT	-1.471931	5.136715e-07
ADH1C	-1.464445	1.270418e-07

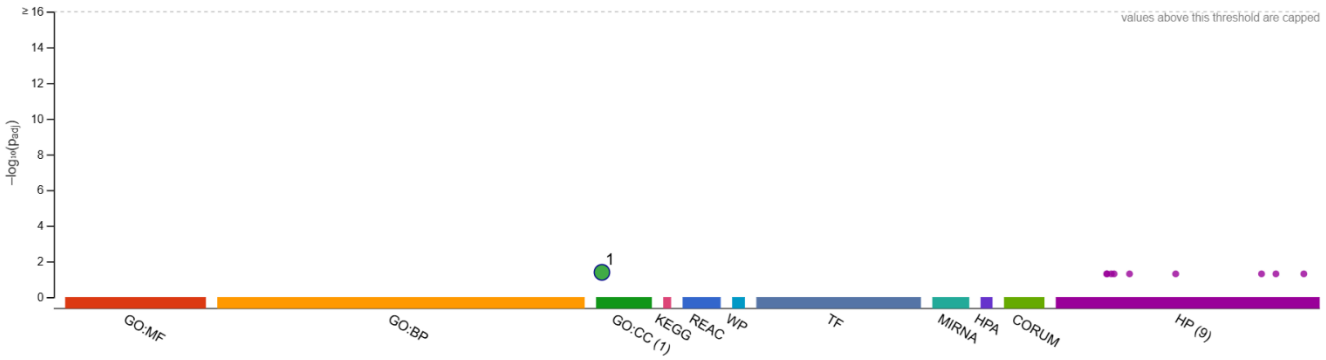
Gene Enrichment

GO:CC		stats		SDC1	COL1A1	FOXA1	S100P	GALNT6
Term name	Term ID	p_{adj}	$-\log_{10}(p_a...$ 0 ≤16					
microvillus	GO:0005902	4.104×10^{-2}						

1 to 1 of 1 Page 1 of 1

HP		stats		SDC1	COL1A1	FOXA1	S100P	GALNT6
Term name	Term ID	p_{adj}	$-\log_{10}(p_a...$ 0 ≤16					
Widely patent coronal suture	HP:0005442	4.996×10^{-2}						
Widely patent sagittal suture	HP:0005476	4.996×10^{-2}						
Wide tufts of distal phalanges	HP:0006095	4.996×10^{-2}						
Irregular proximal tibial epiphyses	HP:0006456	4.996×10^{-2}						
Posterior vertebral hypoplasia	HP:0008451	4.996×10^{-2}						
Small distal femoral epiphysis	HP:0012283	4.996×10^{-2}						
Lumbar hypolordosis	HP:0034770	4.996×10^{-2}						
Meningeal calcification	HP:0100250	4.996×10^{-2}						
Tympanic membrane hypermobility	HP:6000015	4.996×10^{-2}						

1 to 9 of 9 Page 1 of 1



ID	Source	Term ID	Term Name	p_{adj} (query...
1	GO:CC	GO:0005902	microvillus	4.104×10^{-2}

version	e112_eg59_p19_25aa4782
date	5/30/2025, 8:16:16 PM
organism	hsapiens

g:Profiler

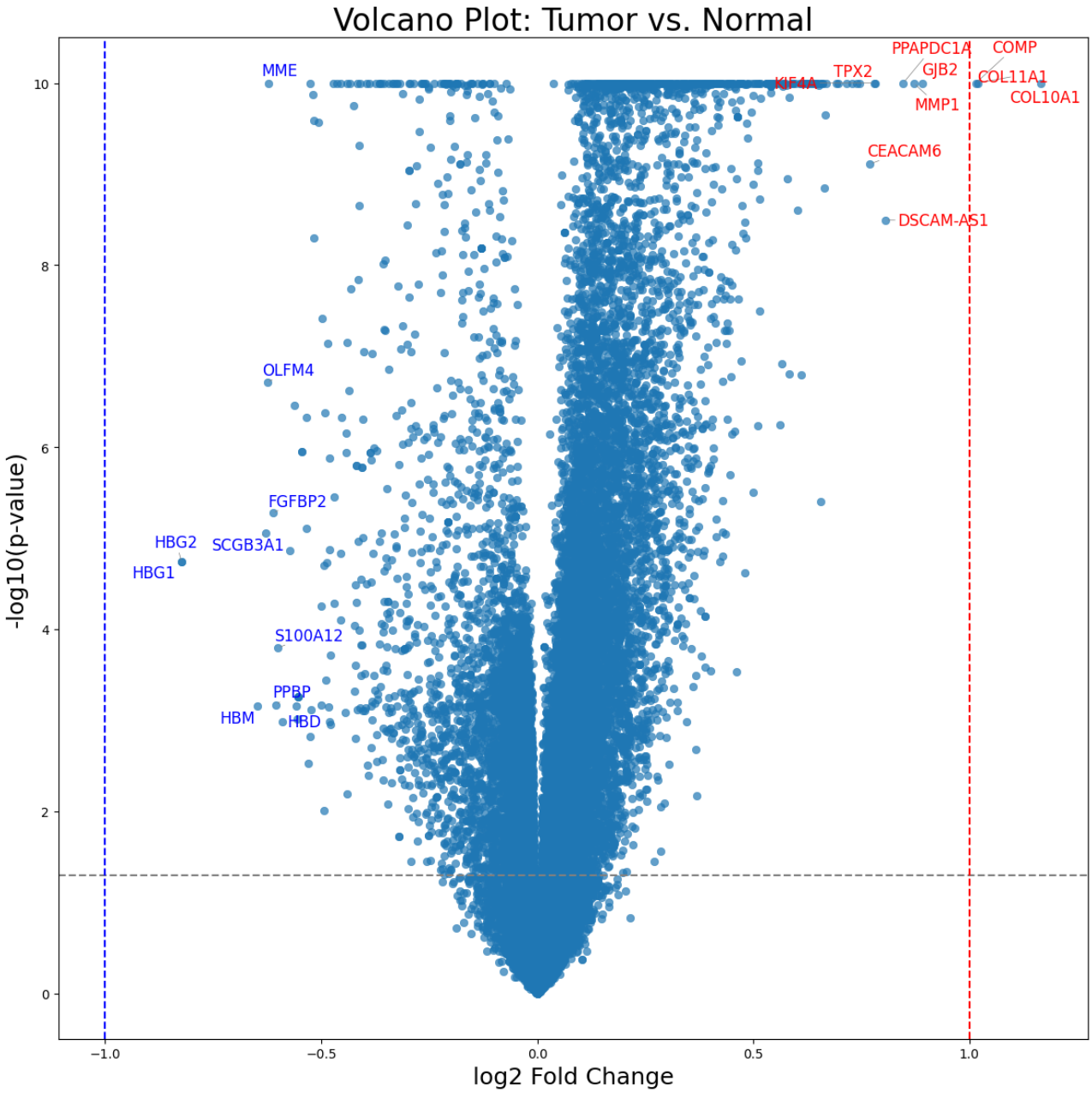
Appendix 10: GSE29044

△ Top 10 Upregulated Genes:

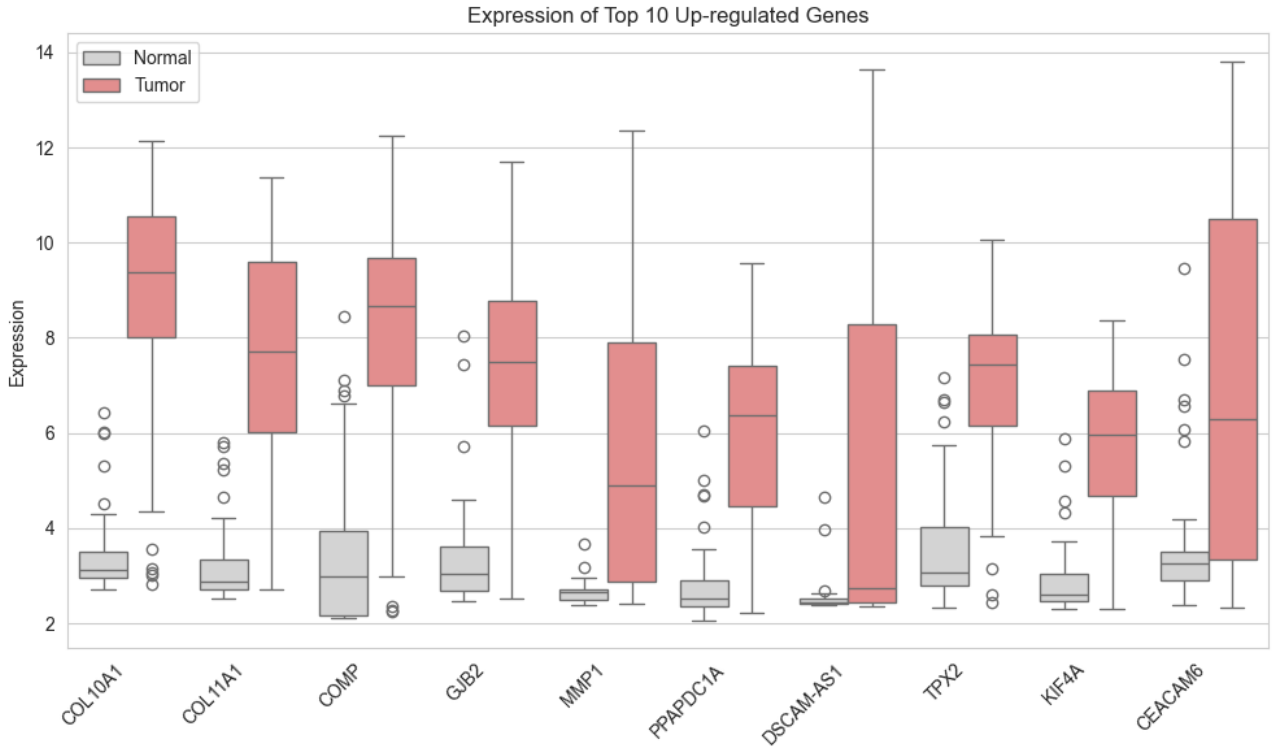
	Gene_Symbol	log2FC	p_value	-log10(p)
Gene_Symbol				
4108	COL10A1	1.166195	9.490338e-36	35.022718
4109	COL11A1	1.019259	7.001858e-26	25.154787
4173	COMP	1.015101	6.581372e-26	25.181684
7483	GJB2	0.891259	1.736958e-23	22.760211
13884	MMP1	0.871690	5.271276e-13	12.278084
16568	PPAPDC1A	0.847015	4.866692e-22	21.312766
5530	DSCAM-AS1	0.806208	3.110709e-09	8.507141
22324	TPX2	0.781553	5.221493e-23	22.282205
9634	KIF4A	0.779931	1.709364e-23	22.767166
3554	CEACAM6	0.770136	6.736910e-10	9.171539

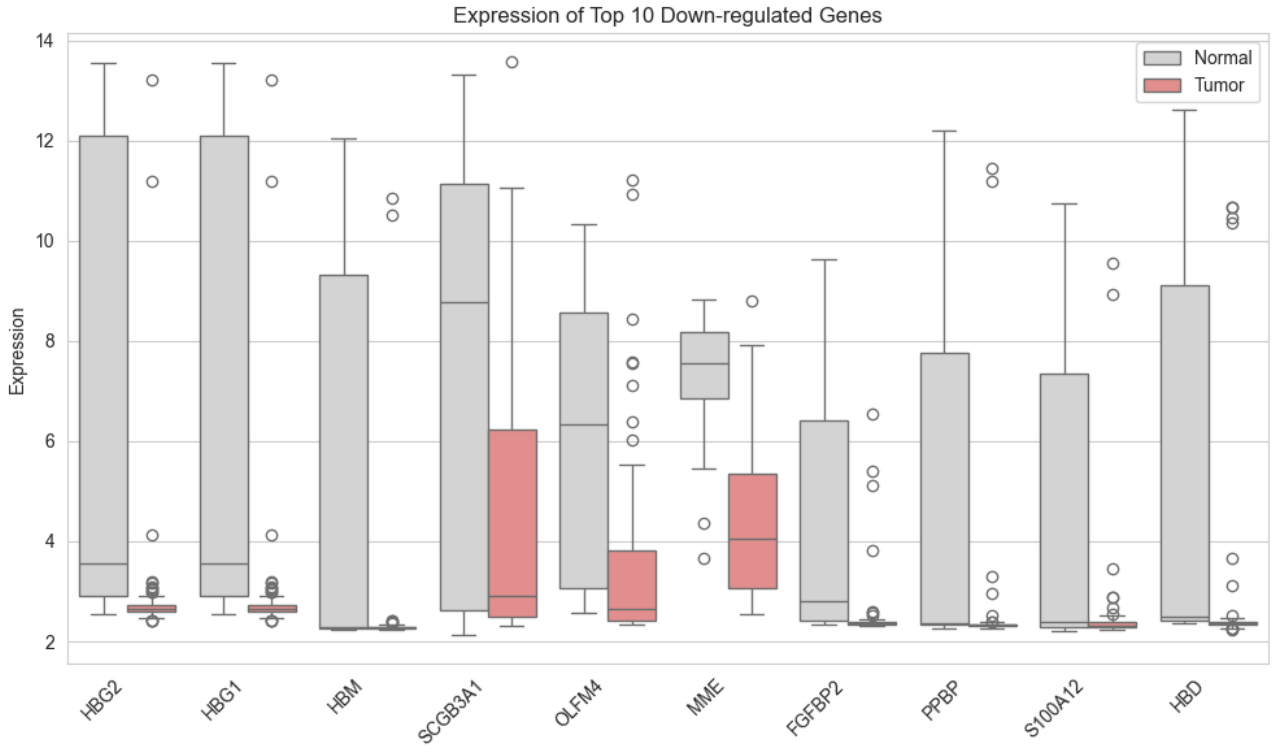
▽ Top 10 Downregulated Genes:

	Gene_Symbol	log2FC	p_value	-log10(p)
Gene_Symbol				
8122	HBG2	-0.823185	1.828512e-05	4.737902
8121	HBG1	-0.823185	1.828512e-05	4.737902
8123	HBM	-0.647471	6.995509e-04	3.155181
19299	SCGB3A1	-0.629635	8.764842e-06	5.057256
15268	OLFM4	-0.624777	1.957334e-07	6.708335
13881	MME	-0.622633	9.157710e-23	22.038213
6796	FGFBP2	-0.612412	5.298021e-06	5.275886
16578	PPBP	-0.604822	6.817885e-04	3.166350
19156	S100A12	-0.600276	1.600821e-04	3.795657
8118	HBD	-0.590297	1.045215e-03	2.980794

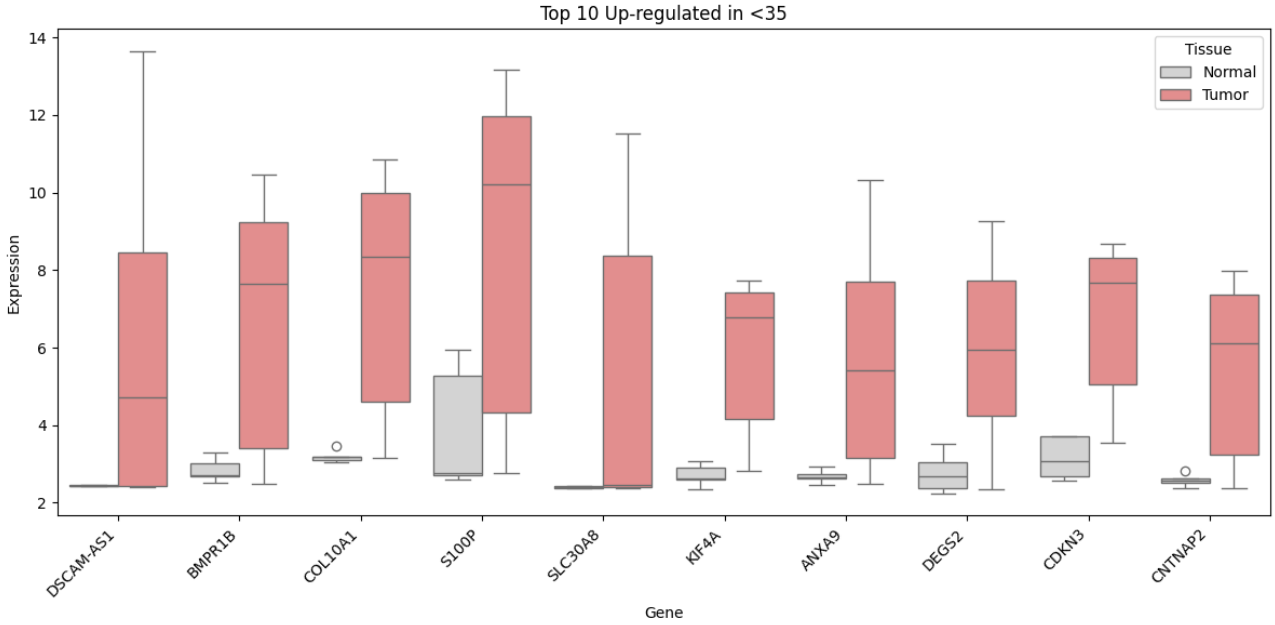


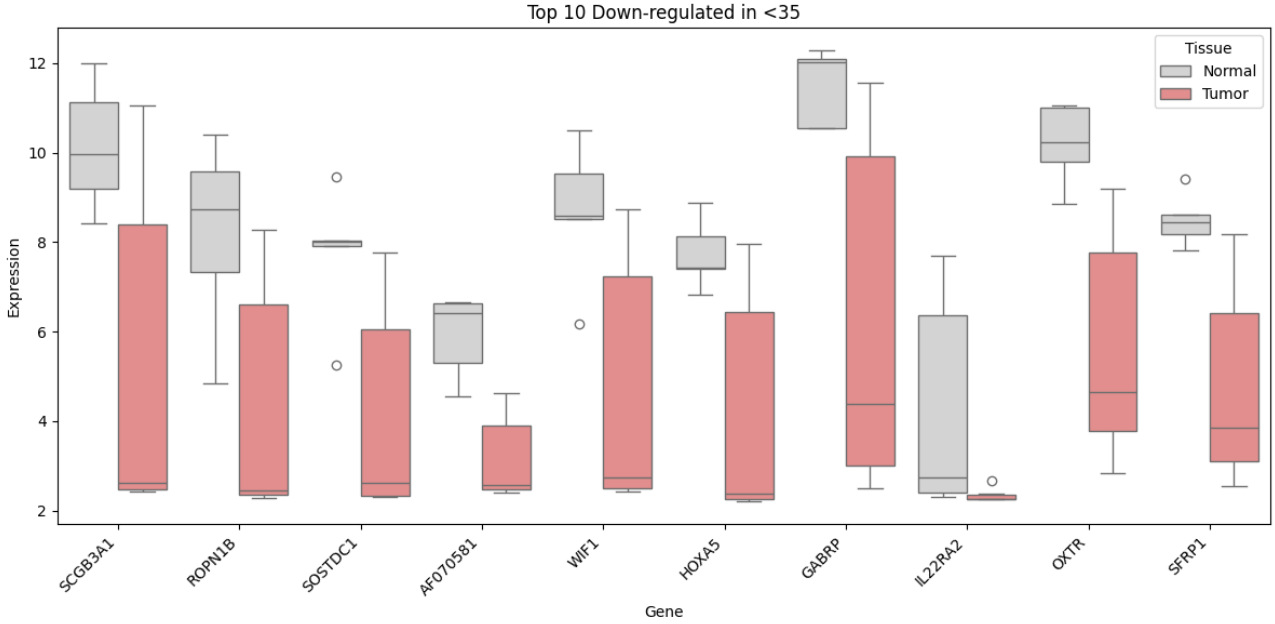
Top Up and down regulated genes – in complete data set

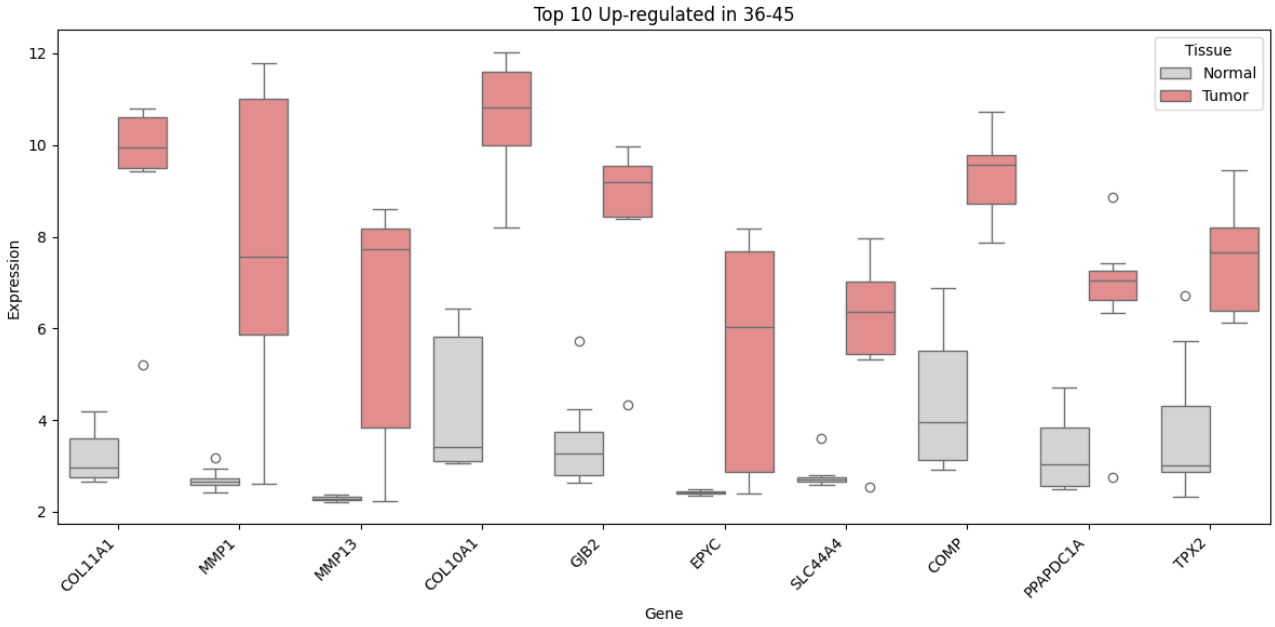


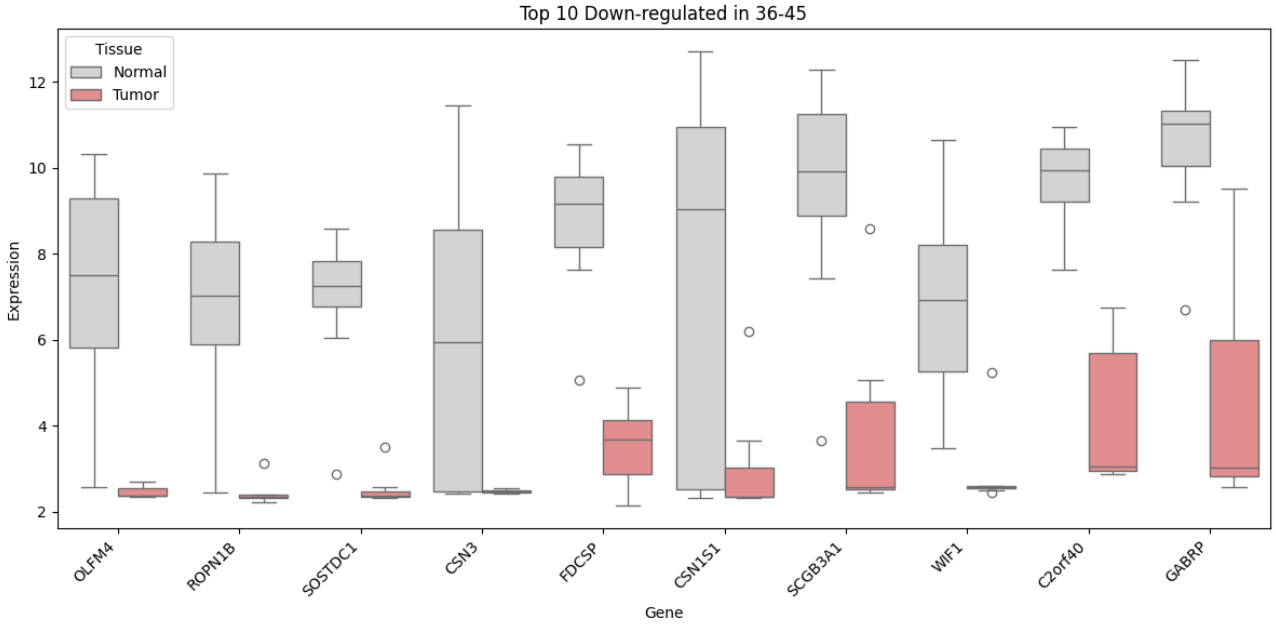


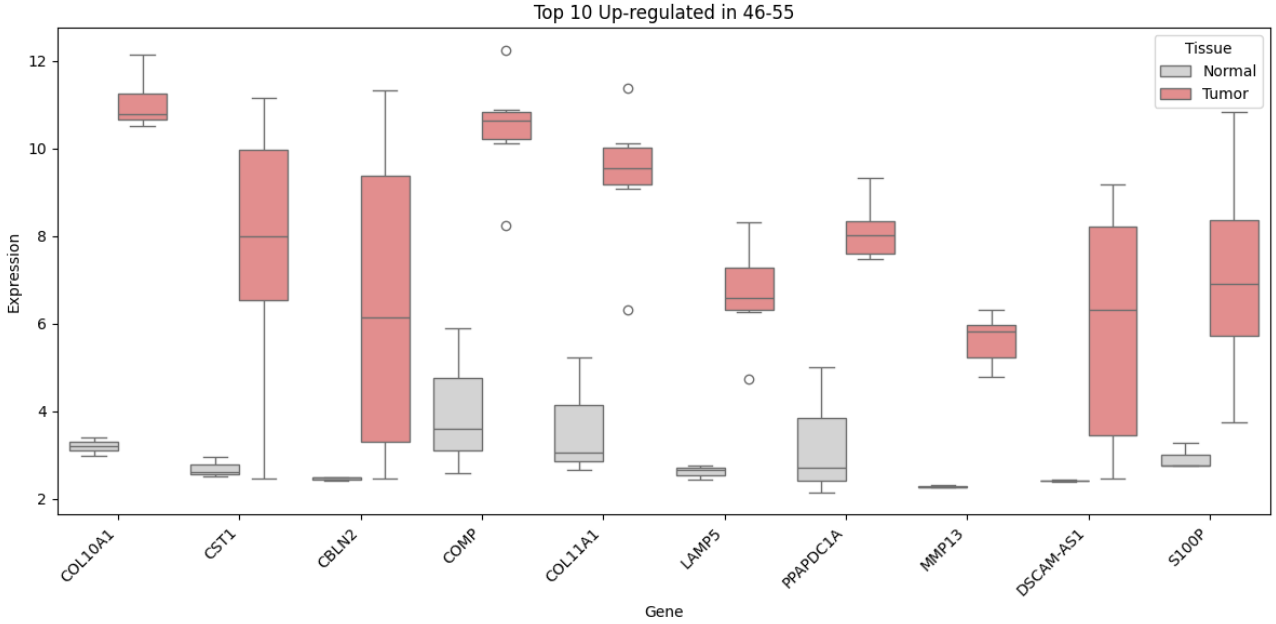
Top Up and down regulated genes – according to Age groups

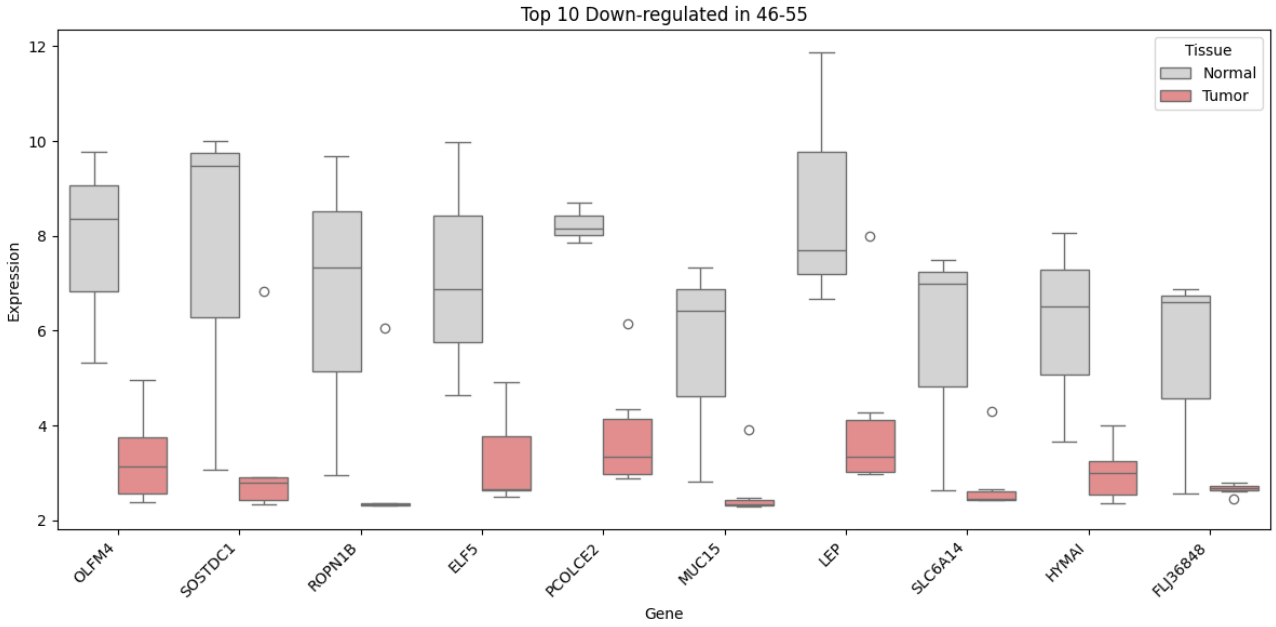

Top 10 Up-regulated in <35
Tissue
Normal
Tumor
Expression
Gene
DSCAM-AS1
BMPR1B
COL10A1
S100P
SLC30A8
KIF4A
ANXA9
DEGS2
CDKN3
CNTNAP2

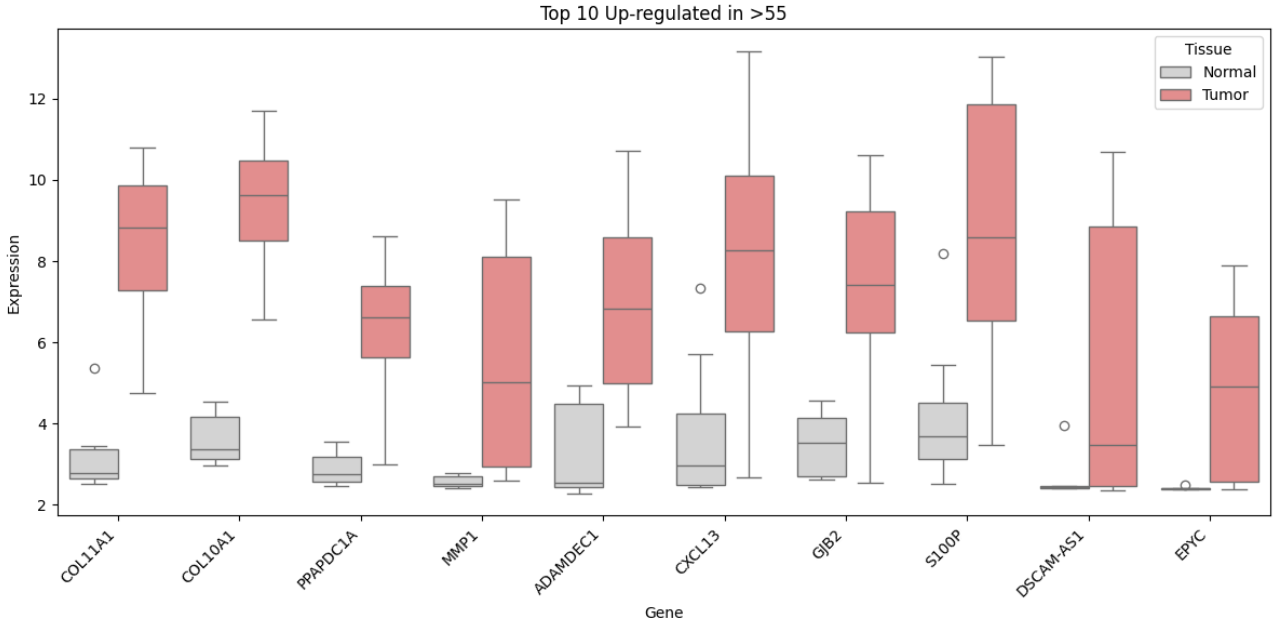

Top 10 Down-regulated in <35
Tissue
Normal
Tumor
Expression
Gene
SCGB3A1
ROPN1B
SOSTDC1
AF070581
WIF1
HOXA5
GABRP
IL22RA2
OXTR
SFRP1

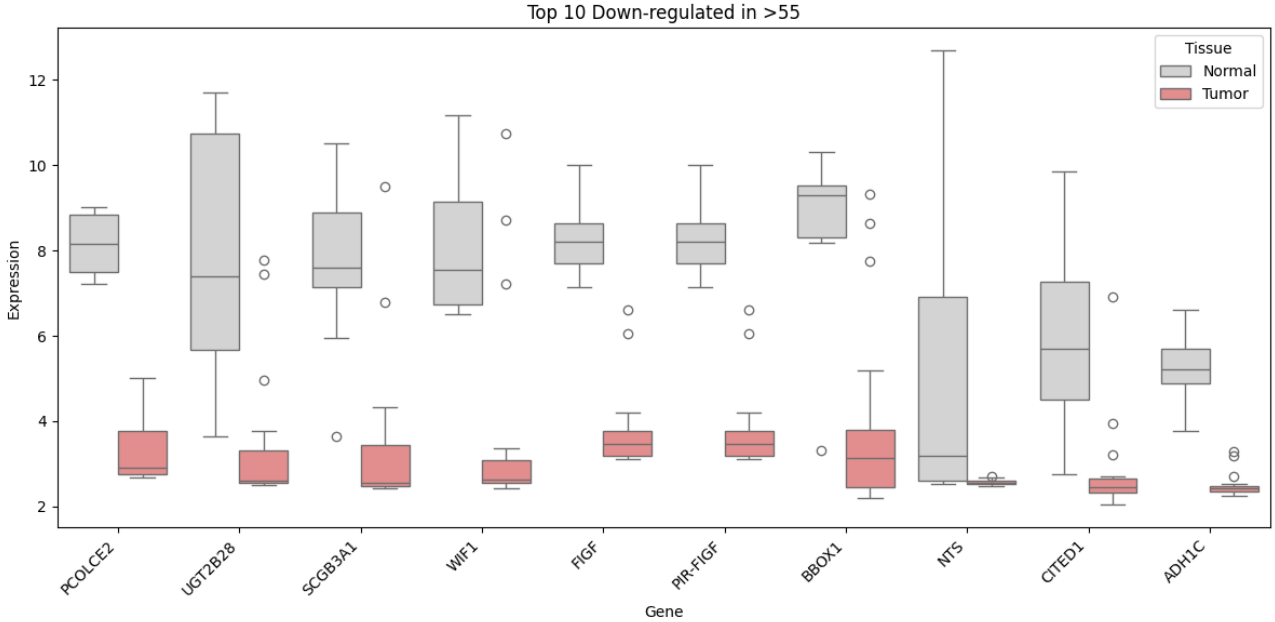

Top 10 Up-regulated in 36-45
Tissue
Normal
Tumor
Expression
Gene
COL11A1
MMP1
MMP13
COL10A1
GJB2
EPYC
SLC44A4
COMP
PPAPDC1A
TPX2


Top 10 Down-regulated in 36-45
Tissue
Normal
Tumor
Expression
Gene
OLFM4
ROPN1B
SOSTDC1
CSN3
FDCSP
CSN1S1
SCGB3A1
WIF1
C2orf40
GABRP

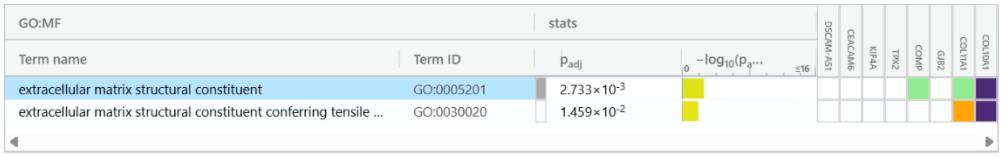

Top 10 Up-regulated in 46-55
Tissue
Normal
Tumor
Expression
Gene
COL10A1
CST1
CBLN2
COMP
COL11A1
LAMP5
PPAPDC1A
MMP13
DSCAM-AS1
S100P


Top 10 Down-regulated in 46-55
Tissue
Normal
Tumor
Expression
Gene
OLFM4
SOSTDC1
ROPN1B
ELF5
PCOLCE2
MUC15
LEP
SLC6A14
HYMAI
FLJ36848





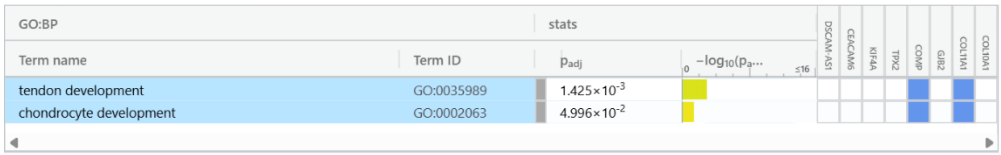
Gene Enrichment



GO:MF

Term name	Term ID	stats: p_{adj}	stats: $-\log_{10}(p_a\cdots$ (0 … ≤16)	DSCAM-AS1	CEACAM6	KIF4A	TPX2	COMP	GJB2	COL11A1	COL8A1
extracellular matrix structural constituent	GO:0005201	2.733×10^{-3}									
extracellular matrix structural constituent conferring tensile ...	GO:0030020	1.459×10^{-2}									

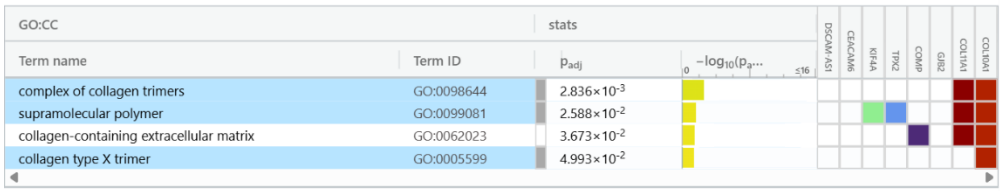
1 to 2 of 2 Page 1 of 1



GO:BP

Term name	Term ID	stats: p_{adj}	stats: $-\log_{10}(p_a\cdots$ (0 … ≤16)	DSCAM-AS1	CEACAM6	KIF4A	TPX2	COMP	GJB2	COL11A1	COL8A1
tendon development	GO:0035989	1.425×10^{-3}									
chondrocyte development	GO:0002063	4.996×10^{-2}									

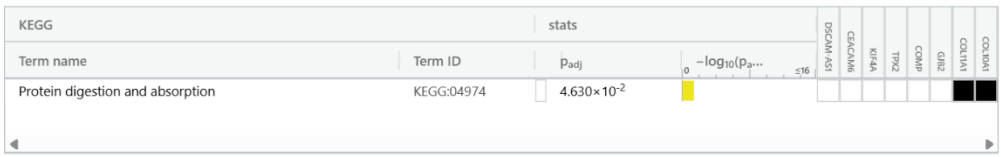
1 to 2 of 2 Page 1 of 1



GO:CC

Term name	Term ID	stats: p_{adj}	stats: $-\log_{10}(p_a\cdots$ (0 … ≤16)	DSCAM-AS1	CEACAM6	KIF4A	TPX2	COMP	GJB2	COL11A1	COL8A1
complex of collagen trimers	GO:0098644	2.836×10^{-3}									
supramolecular polymer	GO:0099081	2.588×10^{-2}									
collagen-containing extracellular matrix	GO:0062023	3.673×10^{-2}									
collagen type X trimer	GO:0005599	4.993×10^{-2}									

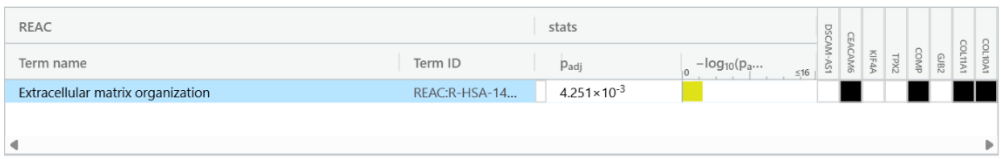
1 to 4 of 4 Page 1 of 1



KEGG

Term name	Term ID	stats: p_{adj}	stats: $-\log_{10}(p_a\cdots$ (0 … ≤16)	DSCAM-AS1	CEACAM6	KIF4A	TPX2	COMP	GJB2	COL11A1	COL8A1
Protein digestion and absorption	KEGG:04974	4.630×10^{-2}									

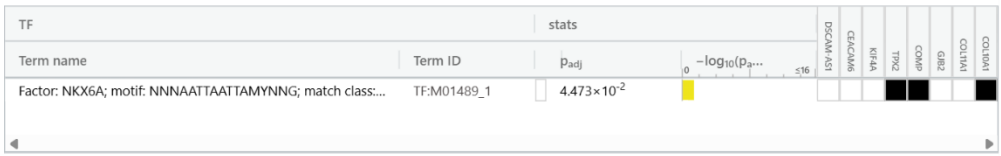
1 to 1 of 1 Page 1 of 1



REAC

Term name	Term ID	stats: p_{adj}	stats: $-\log_{10}(p_a\cdots$ (0 … ≤16)	DSCAM-AS1	CEACAM6	KIF4A	TPX2	COMP	GJB2	COL11A1	COL8A1
Extracellular matrix organization	REAC:R-HSA-14...	4.251×10^{-3}									

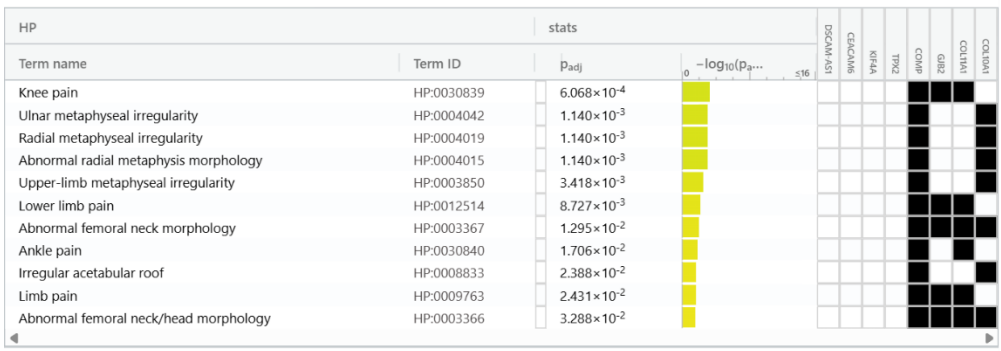
1 to 1 of 1 Page 1 of 1



TF

Term name	Term ID	stats: p_{adj}	stats: $-\log_{10}(p_a\cdots$ (0 … ≤16)	DSCAM-AS1	CEACAM6	KIF4A	TPX2	COMP	GJB2	COL11A1	COL8A1
Factor: NKX6A; motif: NNNAATTAATTAMYNNG; match class:...	TF:M01489_1	4.473×10^{-2}									

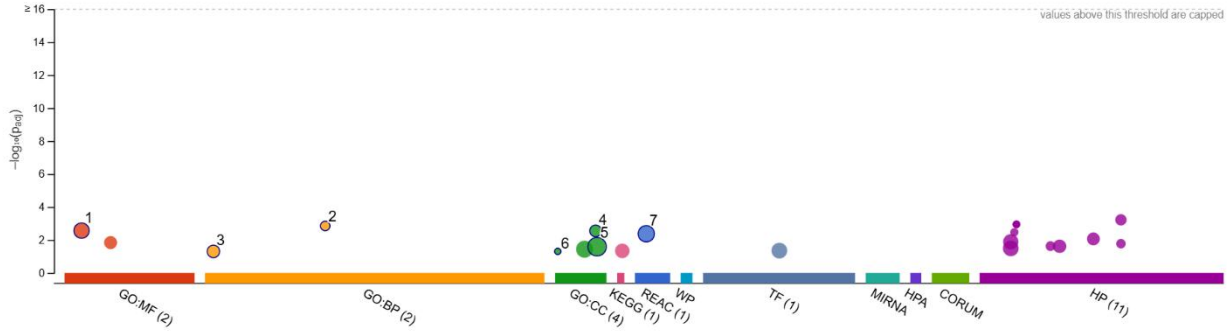
1 to 1 of 1 Page 1 of 1



HP

Term name	Term ID	stats: p_{adj}	stats: $-\log_{10}(p_a\cdots$ (0 … ≤16)	DSCAM-AS1	CEACAM6	KIF4A	TPX2	COMP	GJB2	COL11A1	COL8A1
Knee pain	HP:0030839	6.068×10^{-4}									
Ulnar metaphyseal irregularity	HP:0004042	1.140×10^{-3}									
Radial metaphyseal irregularity	HP:0004019	1.140×10^{-3}									
Abnormal radial metaphysis morphology	HP:0004015	1.140×10^{-3}									
Upper-limb metaphyseal irregularity	HP:0003850	3.418×10^{-3}									
Lower limb pain	HP:0012514	8.727×10^{-3}									
Abnormal femoral neck morphology	HP:0003367	1.295×10^{-2}									
Ankle pain	HP:0030840	1.706×10^{-2}									
Irregular acetabular roof	HP:0008833	2.388×10^{-2}									
Limb pain	HP:0009763	2.431×10^{-2}									
Abnormal femoral neck/head morphology	HP:0003366	3.288×10^{-2}									

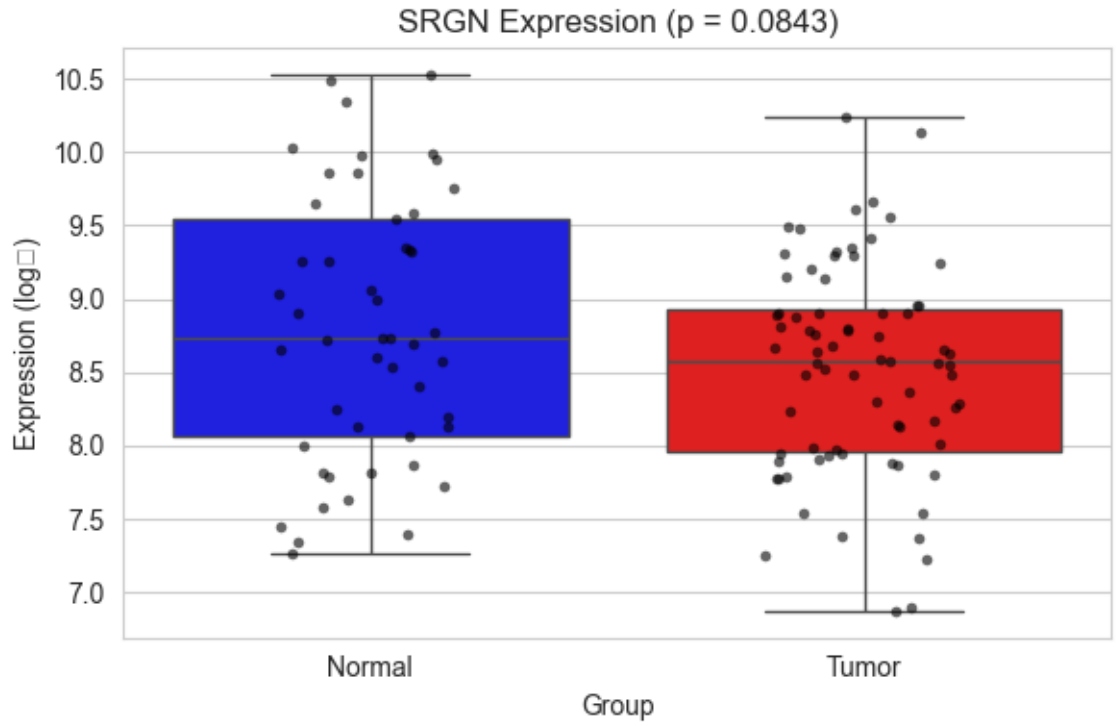
1 to 11 of 11 Page 1 of 1

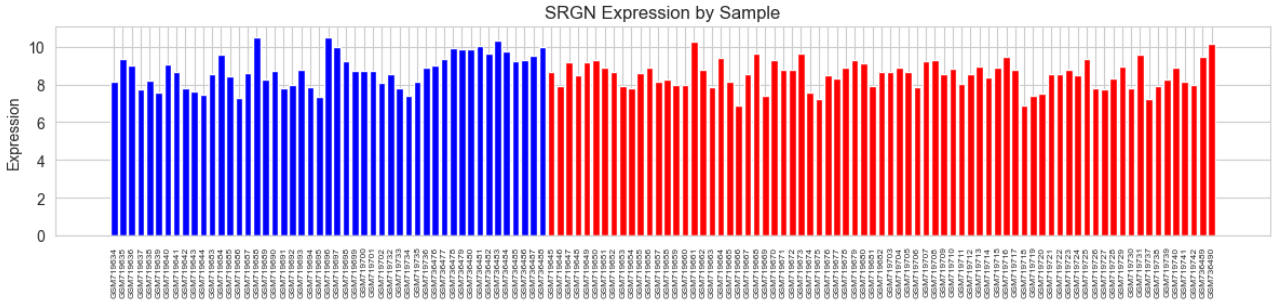


ID	Source	Term ID	Term Name	p_{adj} (query...
1	GO:MF	GO:0005201	extracellular matrix structural constituent	2.733×10^{-3}
2	GO:BP	GO:0035989	tendon development	1.425×10^{-3}
3	GO:BP	GO:0002063	chondrocyte development	4.996×10^{-2}
4	GO:CC	GO:0098644	complex of collagen trimers	2.836×10^{-3}
5	GO:CC	GO:0099081	supramolecular polymer	2.588×10^{-2}
6	GO:CC	GO:0005599	collagen type X trimer	4.993×10^{-2}
7	REAC	REAC:R-HSA-14...	Extracellular matrix organization	4.251×10^{-3}

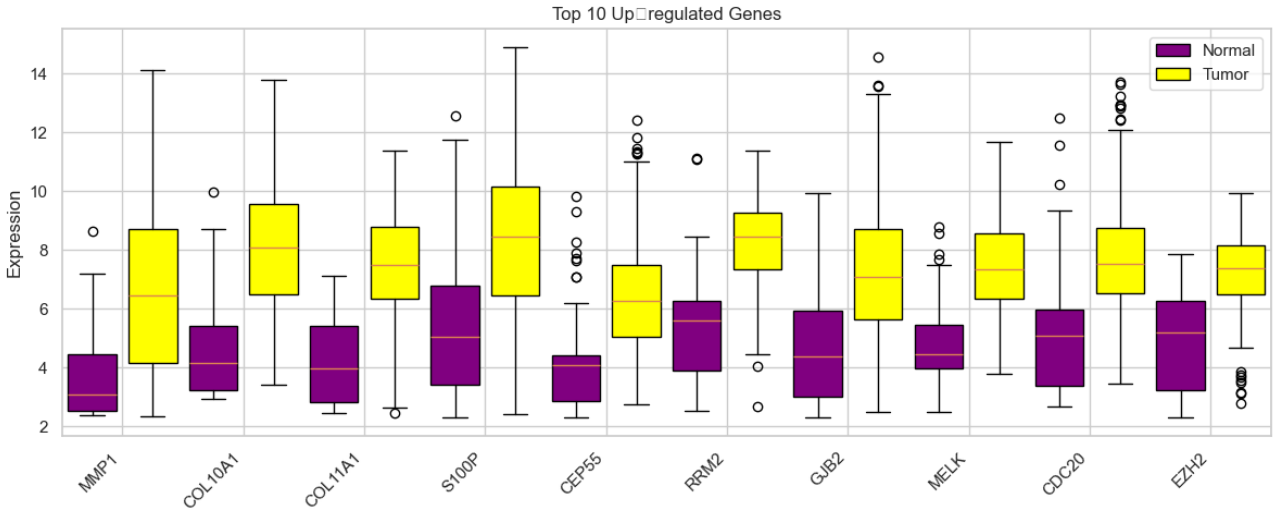
version	e112_eg59_p19_25aa4782
date	5/30/2025, 7:50:17 PM
organism	hsapiens

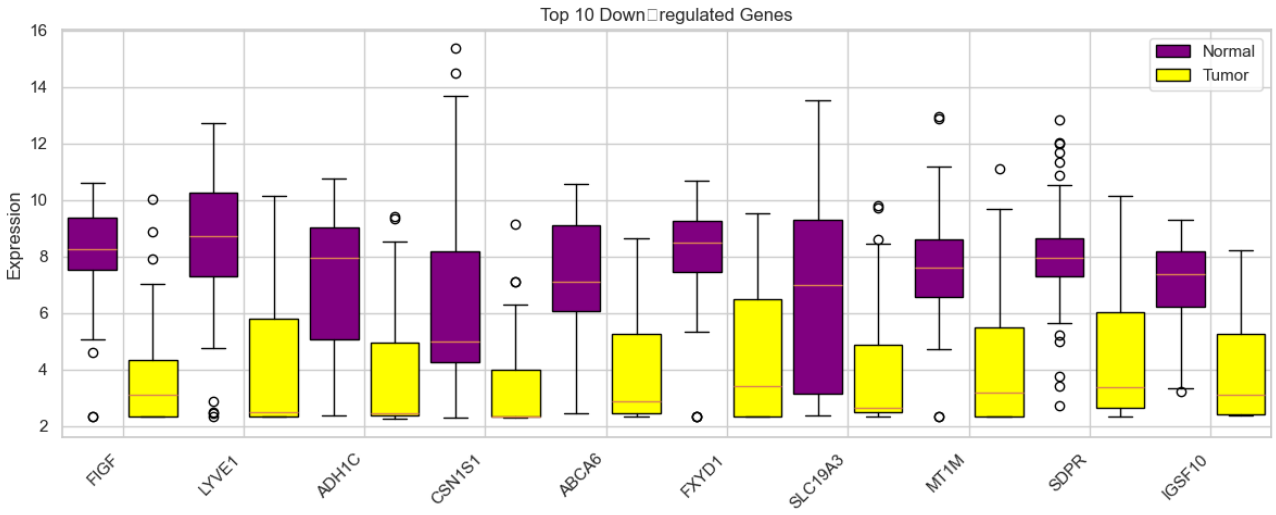
g:Profiler

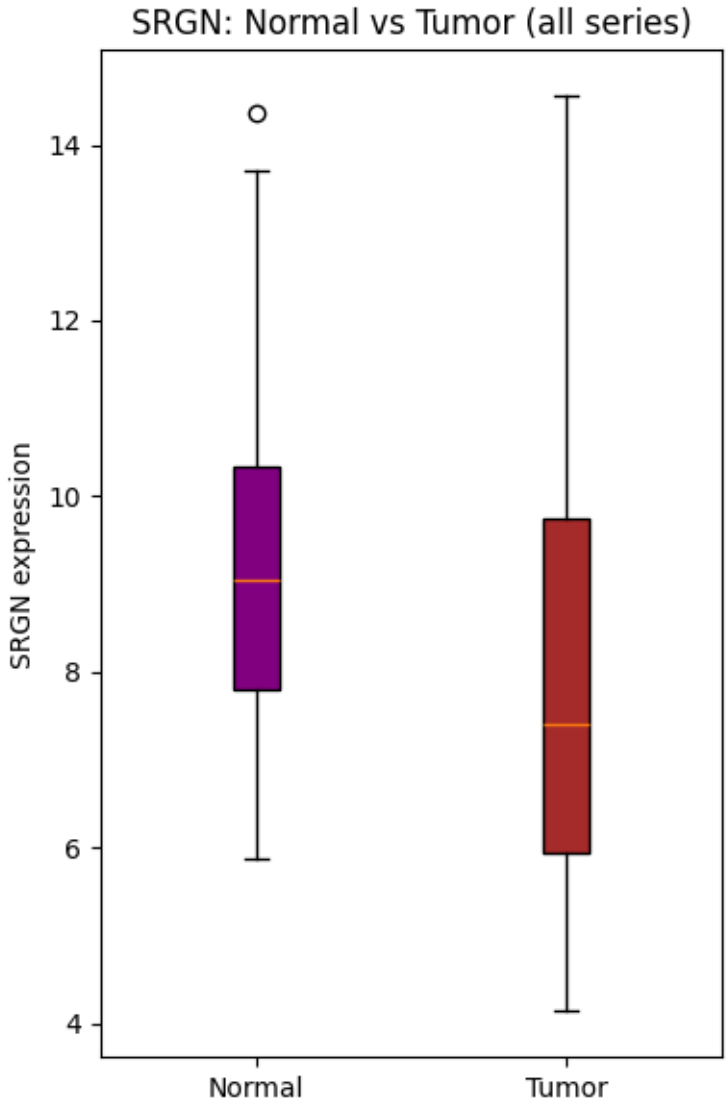


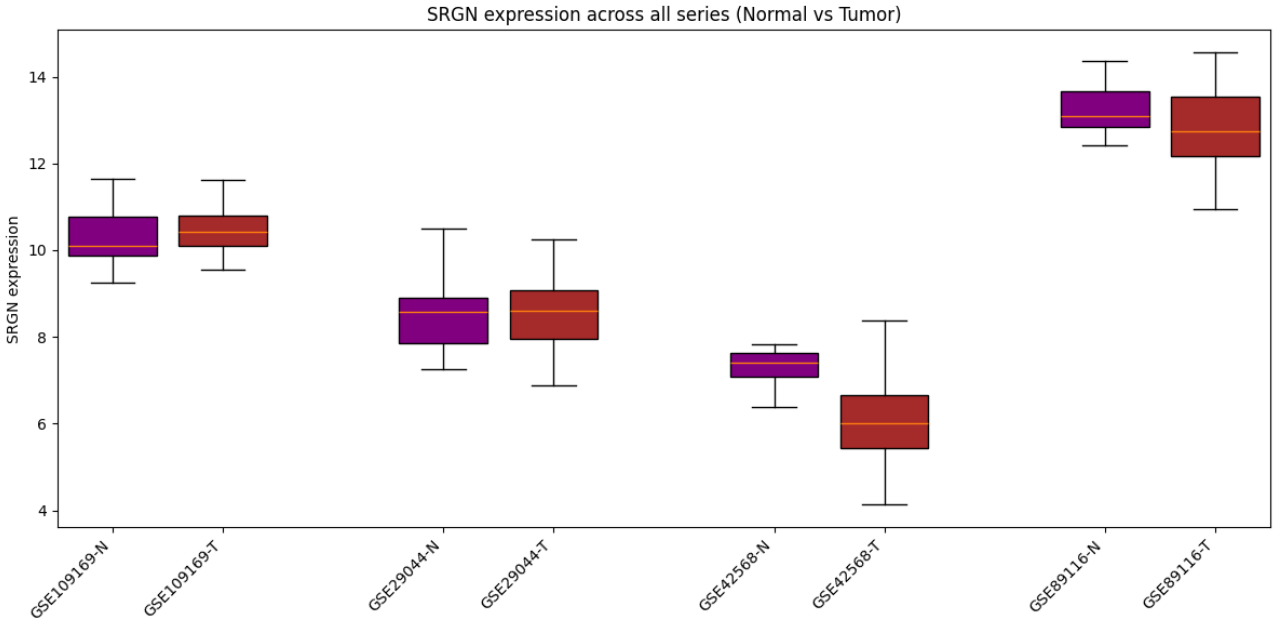


Appendix 11: expression details cumulative (all data set -Human)

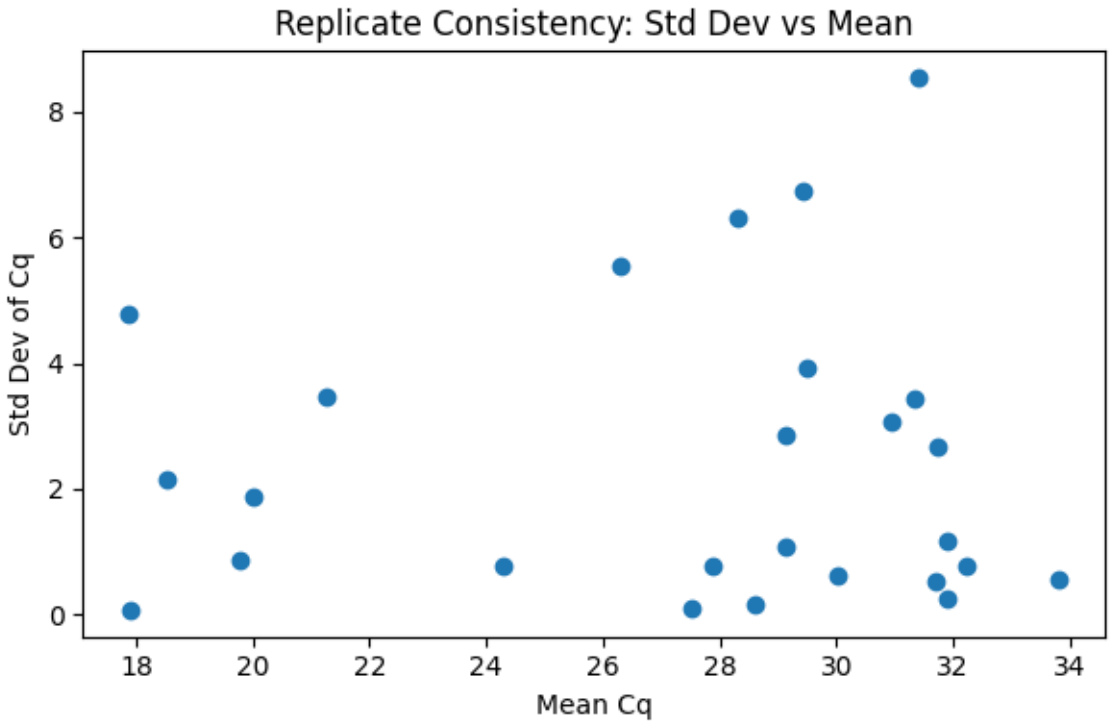


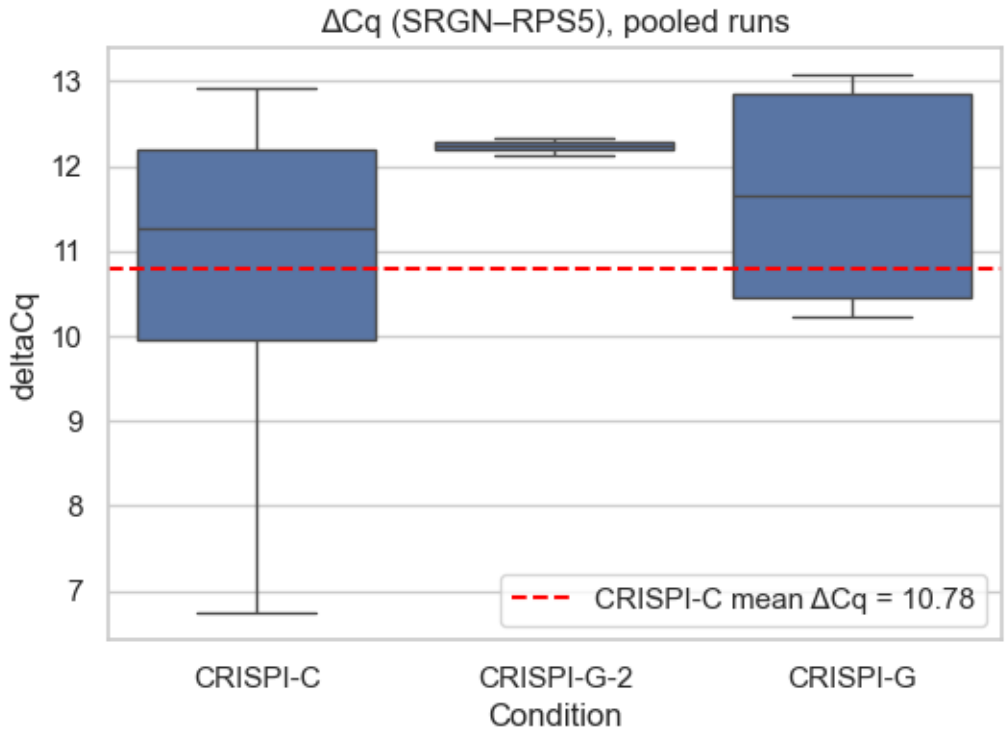


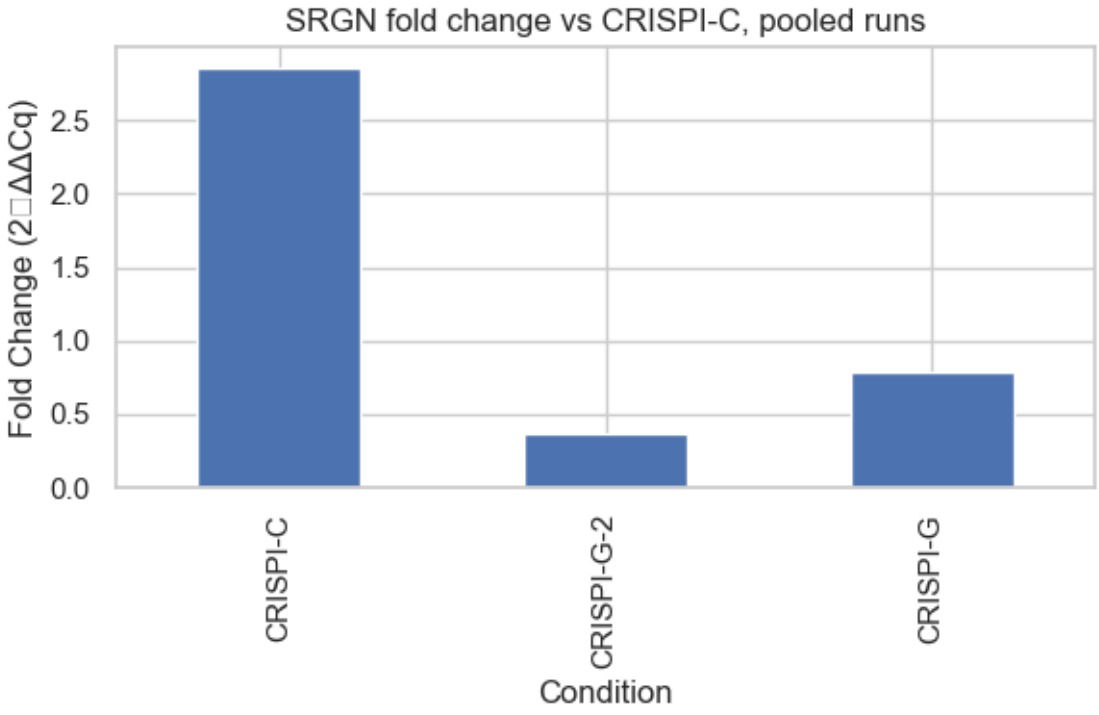

SRGN: Normal vs Tumor (all series)
SRGN expression
14
12
10
8
6
4
Normal
Tumor

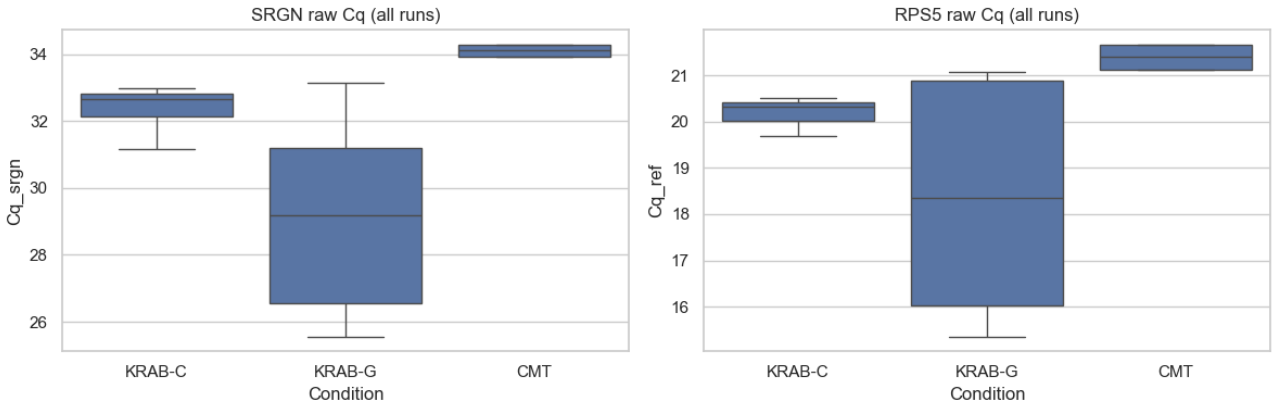

SRGN expression across all series (Normal vs Tumor)
SRGN expression
14
12
10
8
6
4
GSE109169-N
GSE109169-T
GSE29044-N
GSE29044-T
GSE42568-N
GSE42568-T
GSE89116-N
GSE89116-T

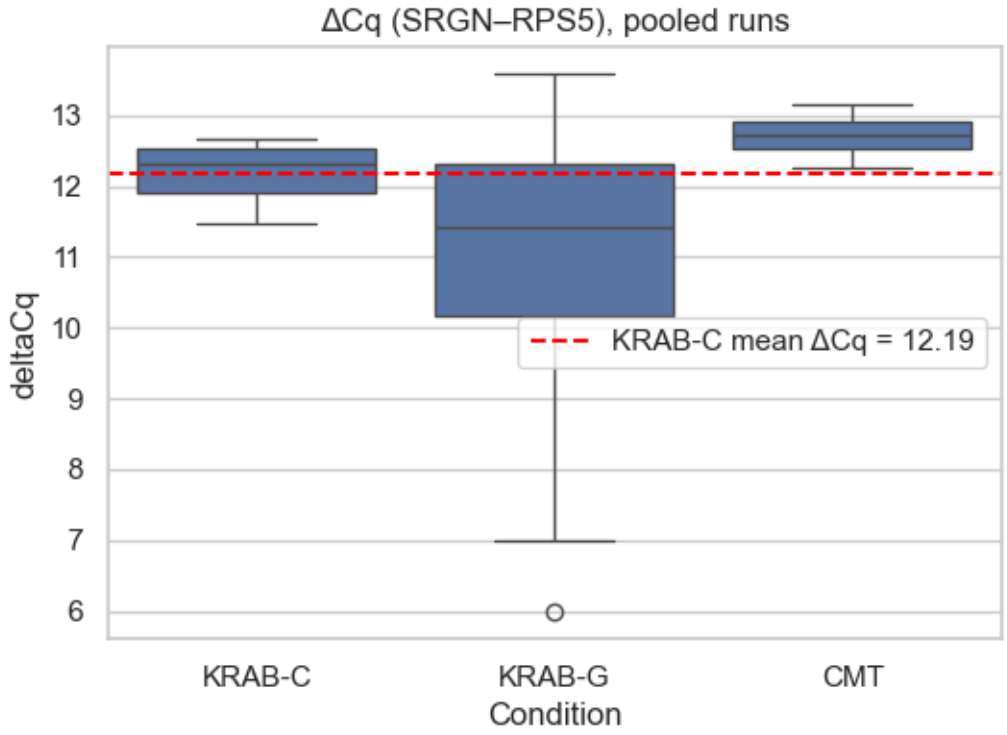
Appendix 12: Experimental Data Analysis (qPCR)

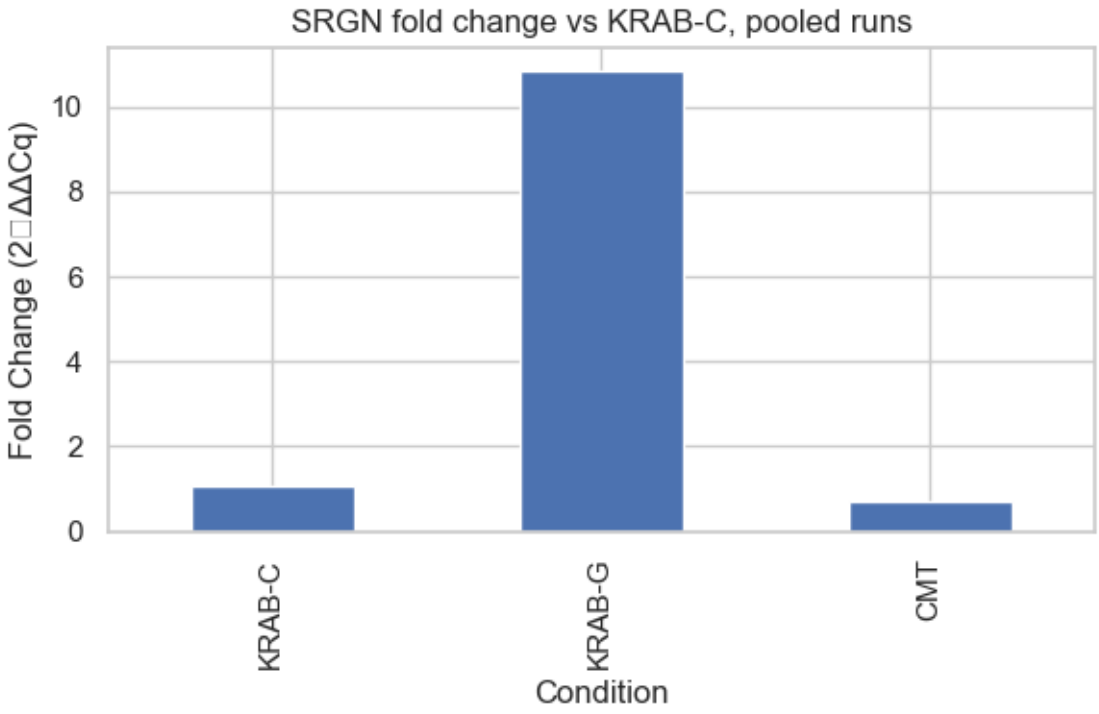



ΔCq (SRGN–RPS5), pooled runs
deltaCq
13
12
11
10
9
8
7
CRISPI-C mean ΔCq = 10.78
CRISPI-C
CRISPI-G-2
CRISPI-G
Condition


SRGN fold change vs CRISPI-C, pooled runs
Fold Change (2□ΔΔCq)
2.5
2.0
1.5
1.0
0.5
0.0
CRISPI-C
CRISPI-G-2
CRISPI-G
Condition


SRGN raw Cq (all runs)
Cq_srgn
34
32
30
28
26
KRAB-C
KRAB-G
CMT
Condition
RPS5 raw Cq (all runs)
Cq_ref
21
20
19
18
17
16
KRAB-C
KRAB-G
CMT
Condition


ΔCq (SRGN–RPS5), pooled runs
deltaCq
13
12
11
10
9
8
7
6
KRAB-C mean ΔCq = 12.19
KRAB-C
KRAB-G
CMT
Condition


SRGN fold change vs KRAB-C, pooled runs
Fold Change (2□ΔΔCq)
10
8
6
4
2
0
KRAB-C
KRAB-G
CMT
Condition

KRAB-G-2 did not have data on SRGN expression

Publishing and archiving

☒ YES, I, Ruwini Madhushika Herath Herath Mudiyanselage, have read and agree to the agreement for publication and the personal data processing that takes place in connection with this

☒ YES, I, Ruwini Madhushika Herath Herath Mudiyanselage have read and agree to the agreement for publication and the personal data processing that takes place in connection with this.