



Master Project

# Engineering SNAP-tag based antibody fusion protein for Improved Cervical Cancer Therapy: Tackling Heterogeneity and Mesothelin Shedding

**Author:** Viacheslav Bolotnikov

**Host Supervisor:** Prof. Dr. Dr. Stefan Barth

**Host-Co-Supervisor:** Dr. Nkhasi Lekana

**EPFL Supervisor:** Prof. Dr. Li Tang

July 3, 2025

## Contents

<b>Plagiarism Declaration</b>	I
<b>Acknowledgements</b>	I
<b>List of Publications</b>	II
<b>List of Figures</b>	II
<b>List of Tables</b>	IV
<b>List of Abbreviations</b>	V
<b>Abstract</b>	VII
<b>1 Introduction</b>	1
1.1 Histological Subtypes	1
1.2 Current state of research	2
1.2.1 Focus on Squamous Cell Carcinoma	2
1.2.2 Etiology and Pathogenesis of Cervical Cancer	3

1.2.3	Conventional Cervical Cancer Therapies and Prevailing Challenges . . . . .	3
1.2.4	Targeted Therapy: Promise and Limitations of Conventional Antibody-Drug Conjugates (ADCs) . . . . .	4
1.2.5	SNAP-Tag Fusion Protein Technology for Antibody Targeting . . . . .	4
1.2.6	Mesothelin as a Target in Cervical Cancer . . . . .	5
1.2.7	Epidermal Growth Factor Receptor (EGFR) as a Co-Target for Bispecific Therapy . . . . .	8
1.2.8	Therapeutic Payloads: Auristatin F . . . . .	9
1.3	Additional study: Chondroitin sulfate proteoglycan 4 (CSPG4) . . . . .	10
1.4	Perspectives . . . . .	10
1.5	Aims and objectives . . . . .	11
<b>2</b>	<b>Materials and Methods . . . . .</b>	<b>12</b>
2.1	Bioinformatic Analysis for Secondary Target Discovery . . . . .	12
2.2	Cell culture . . . . .	13
2.3	In silico design of mammalian expression vectors . . . . .	13
2.4	Molecular cloning of SNAP-tag based fusion proteins . . . . .	14
2.4.1	Transformation of <i>E. coli</i> with plasmid DNA . . . . .	14
2.4.2	Isolation of plasmid DNA from <i>E. coli</i> . . . . .	15
2.4.3	Restriction enzyme digest . . . . .	15
2.4.4	Agarose gel electrophoresis . . . . .	15
2.4.5	Recovery of DNA fragments from agarose gels . . . . .	16
2.4.6	Ligation reactions and transformation of recombinant plasmids in <i>E. coli</i> . . . . .	16
2.4.7	Small-scale DNA isolation, restriction mapping and DNA sequencing . . . . .	17
2.4.8	PCR amplification of plasmid fragment . . . . .	17
2.4.9	Expression of the fusion proteins in HEK293T cells . . . . .	18
2.4.10	Protein purification using Immobilized Metal Affinity Chromatography (IMAC) . . . . .	19
2.4.11	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and western blot analysis of recombinant protein fractions . . . . .	19
<b>3</b>	<b>Results . . . . .</b>	<b>21</b>
3.1	Discovery and First Characterisation of a New Cervical Cancer Subtype: A Bioinformatic Analysis . . . . .	21
3.1.1	Dataset cleaning and first observation . . . . .	21
3.1.2	Subtypes identification . . . . .	23
3.1.3	Proteomics analysis and identification of potential targets . . . . .	25
3.1.4	EGFR as a target in MSLN-low tumors . . . . .	27
3.1.5	MSLN-low subtype characterisation . . . . .	27
3.2	Bispecific SNAP-tag Antibody Cloning and Production	29

3.2.1	Cloning of pCB- $\alpha$ MSLN-SNAP-H22 (training phase with Thabo Matshoba) . . . . .	29
3.2.2	PCR Amplification and Restriction Digest of $\alpha$ MSLN (scFv) Insert . . . . .	31
3.2.3	Cloning and screening of pCB- $\alpha$ MSLN (scFv) -SNAP- $\alpha$ MSLN (scFv) (training phase with Thabo Matshoba) . . . . .	32
3.2.4	Cloning of pCB- $\alpha$ EGFR(scFv)-SNAP- $\alpha$ MSLN(scFv) . . . . .	34
3.2.5	Colony screening and validation of pCB- $\alpha$ EGFR(scFv)-SNAP- $\alpha$ MSLN(scFv) . . . . .	36
3.2.6	Sequencing confirmation of pCB- $\alpha$ EGFR(scFv) -SNAP- $\alpha$ MSLN (scFv) integrity . . . . .	37
3.2.7	Transfection into HEK293T cells for protein expression . . . . .	38
3.3	<b>Design of Shedding-resistant Anti-mesothelin Antibody</b> . . . . .	41
3.3.1	Cloning of pCB - $\alpha$ MSLN(m15B6)(scFv) - SNAP . . . . .	47
3.3.2	Colony screening and validation of pCB- $\alpha$ MSLN(m15B6)(scFv)-SNAP . . . . .	50
3.3.3	Transfection of pCB- $\alpha$ MSLN(m15B6)(scFv)-SNAP into HEK293T cells . . . . .	51
3.4	<b>Additional experiment: Purification and Characterization of <math>\alpha</math>CSPG4(scFv)-SNAP Fusion Protein</b> . . . . .	52
<b>4</b>	<b>Discussion</b> . . . . .	55
4.1	Unmet clinical need and the promise of targeted biologics . . . . .	55
4.2	Mesothelin and the rationale for bispecific strategies . . . . .	56
4.3	Production of bispecific SNAP-tag fusion proteins . . . . .	56
4.4	SNAP-tag scaffolds for next-generation ADCs and precision delivery . . . . .	57
4.5	Proteomic landscape of MSLN-low tumors supports pathway-guided targeting . . . . .	58
<b>5</b>	<b>Future Work and Perspectives</b> . . . . .	59
<b>6</b>	<b>Conclusion</b> . . . . .	60
<b>7</b>	<b>Appendix</b> . . . . .	61
7.1	Full DNA Sequence of m15B6 Construct . . . . .	61
7.2	Sequencing . . . . .	61
	<b>References</b>	63

# Plagiarism Declaration

I, **Viacheslav Bolotnikov**, hereby declare that the work contained in this thesis is my own original unaided work. It is being submitted for the degree of Master of Science MSc in Life Sciences Engineering at the Ecole Polytechnique Fédéral de Lausanne (EPFL). Neither has any part nor this whole thesis been submitted in the past, is being submitted, or will be submitted at any other university. I have used an author–year citation style, combining parenthetical author names with numeric references for clarity and consistency in scientific writing. Each significant contribution to, and quotation in, this thesis from the work or works of other authors and collaborators has been attributed, cited and referenced. I give the university the authority to reproduce the content, either partly or in whole for research purposes.

*Following the EPFL guidelines for the use of generative AI in academic work, I acknowledge that AI tools such as ChatGPT have been used in this thesis to improve language. This includes improving sentence structure, refining word choice, and improving overall readability. All substantive content, research findings, and conclusions remain my original work.*

Signature: 

Date: 03/07/2025

## Acknowledgements

The author would like to thank PhD students Thabo Matshoba and Marc Henry for their guidance and support throughout this project. Sincere thanks go to Professor Stefan Barth for his supervision, trust, and for providing the opportunity to carry out this research, as well as to Dr. Nkhasi Lekana for co-supervising this project. The author also acknowledges the valuable assistance of all members of the laboratory as well as those part of the Medical Biotechnology and Immunotherapy (MB&I) Research Unit, who offered helpful advice, technical support, and encouragement throughout the course of this work.

The author gratefully acknowledges the École Polytechnique Fédérale de Lausanne (EPFL), and in particular the School of Life Sciences, for offering the opportunity to undertake this Master's thesis abroad. Special thanks are extended to Professor Dr. Li Tang for his support and supervision.

This research was supported by the Centre for Developing Cancer Therapeutics for Africa (CDCTA) as part of the University of Cape Town's Grand Challenges Programme (Vision 2030), and by the Faculty of Health Sciences at the University of Cape Town. The project also made use of core resources and services provided by the Flow Cytometry Facility and other shared infrastructure within MB&I.

# List of Publications

No publications have resulted from this work at the time of submission.

# List of Figures

1	Histological and anatomical organization of the human cervix. . . . .	2
2	Mesothelin maturation and shedding into membrane-bound and soluble fragments. . . . .	7
3	Shed mesothelin impairs CAR-T cell targeting and function. . . . .	8
4	Chromatogram exemplifying the applied imidazole concentrations and the expected elution peaks. . . . .	20
5	Distribution of MSLN mRNA expression in the dataset. . . . .	22
6	Clinical characteristics of MSLN-high and MSLN-low tumor groups. . . . .	23
7	Volcano plot of differentially expressed genes (DEGs). . . . .	24
8	Dimensionality reduction of tumor transcriptomic profiles. . . . .	25
9	Correlation between MSLN mRNA and protein expression in multiple tumor types. . . . .	26
10	Restriction digestion of pCB-H22-SNAP-H22 and pCB- $\alpha$ MSLN (scFv) -SNAP using agarose gel electrophoresis. . . . .	30
11	Growth of E. coli cells transformed with potential recombinant pCB- $\alpha$ MSLN(scFv)-SNAP-H22 plasmid DNA. . . . .	30
12	Colony screening by <i>PvuII</i> digestion to confirm insertion of MSLN and SNAP in pCB- $\alpha$ MSLN (scFv) -SNAP-H22. . . . .	31
13	PCR amplification of $\alpha$ MSLN (scFv) insert. . . . .	32
14	Growth of E. coli cells transformed with potential recombinant pCB- $\alpha$ MSLN(scFv)-SNAP- $\alpha$ MSLN(scFv) plasmid DNA. . . . .	33
15	Screening of pCB- $\alpha$ MSLN(scFv)-SNAP- $\alpha$ MSLN(scFv) by <i>NaeI</i> digestion. . . . .	34
16	Restriction digestion of pCB- $\alpha$ EGFR(scFv)-SNAP and pCB- $\alpha$ MSLN(scFv)-SNAP- $\alpha$ MSLN(scFv) using agarose gel electrophoresis. . . . .	35
17	Growth of E. coli cells transformed with potential recombinant pCB- $\alpha$ EGFR(scFv)-SNAP- $\alpha$ MSLN(scFv) plasmid DNA. . . . .	36

18	Colony screening by <i>SacI</i> digestion to identify correct ligation of pCB- $\alpha$ EGFR(scFv)-SNAP- $\alpha$ MSLN(scFv). . . . .	37
19	Visualization of transient GFP expression at 48 and 72 hours post-transfection. . . . .	38
20	Gating strategy for quantifying transfection efficiency. . . . .	39
21	GFP-based quantification of transfection efficiency. . . . .	39
22	Histogram analysis of GFP expression. . . . .	40
23	Stable expression of GFP following one month of zeocin selection in HEK293T cells. . . . .	41
24	Variable region sequences of m15B6. . . . .	42
25	Full amino acid sequence of the bispecific m15B6-based antibody. . . . .	42
26	BLAST alignment of m15B6 antibody sequences. . . . .	42
27	IgBlast analysis of the m15B6 light chain (VL). . . . .	44
28	IgBlast analysis of the m15B6 heavy chain (VH). . . . .	44
29	<i>SnapGene</i> map of the m15B6 antibody cassette. . . . .	45
30	pUC57 - $\alpha$ MSLN (m15B6) (scFv). . . . .	45
31	pCB - $\alpha$ CD14(2F9)(scFv) - SNAP-Tag construct. . . . .	46
32	pCB - $\alpha$ MSLN(m15B6)(scFv) - SNAP-Tag construct. . . . .	47
33	Transformation controls for pUC57- $\alpha$ MSLN(m15B6)(scFv). . . . .	48
34	Restriction digestion of pCB- $\alpha$ CD14(2F9)(scFv)-SNAP and pUC57- $\alpha$ MSLN(m15B6)(scFv) using agarose gel electrophoresis. . . . .	49
35	Growth of E. coli cells transformed with potential recombinant pCB- $\alpha$ MSLN(m15B6)(scFv)-SNAP plasmid DNA. . . . .	50
36	Colony screening by <i>NheI</i> and <i>BamHI</i> digestion to identify cor- rect ligation of pCB- $\alpha$ MSLN(m15B6)(scFv)-SNAP. . . . .	51
37	Visualization of transient GFP expression in HEK293T cells trans- fected with pCB- $\alpha$ MSLN(m15B6)(scFv)-SNAP. . . . .	52
38	Chromatogram of $\alpha$ CSPG4(scFv)-SNAP protein purification. . . . .	53
39	Coomassie staining and western blot analysis of purified $\alpha$ CSPG4(scFv)- SNAP fractions. . . . .	54
40	Generation of a BSA standard curve for the determination of $\alpha$ CSPG4(scFv)-SNAP protein concentration. . . . .	55
1A	Sanger sequencing alignment of pCB- $\alpha$ EGFR (scFv) -SNAP- $\alpha$ MSLN (scFv) to the reference construct. . . . .	61

# List of Tables

1	Important components of the mammalian expression plasmid and their function . . . . .	13
2	Setting up the restriction enzyme digest reaction mixture . . . . .	15
3	Setting up ligation reactions of digested insert with vector back- bone . . . . .	17
4	Top 10 Differentially Expressed Genes (DEGs) between MSLN- High and MSLN-Low Groups . . . . .	24
5	Top RPPA Proteins Upregulated in MSLN-low Tumors (selected targets highlighted) . . . . .	27
6	Selected significantly enriched pathways in MSLN-Low tumors. . . . .	28
7	Transfection efficiencies of HEK293T cells with bispecific SNAP- tag antibody constructs, measured by GFP expression . . . . .	40
1A	Summary of alignment results per sequencing primer, before and after manual correction. . . . .	62

# List of Abbreviations

**ADC** Antibody–Drug Conjugate

**AKT** Protein Kinase B

**BSA** Bovine Serum Albumin

**CAR-T** Chimeric Antigen Receptor T cells

**CDR** Complementarity-Determining Region

**CHO** Chinese Hamster Ovary

**CMV** Cytomegalovirus

**CIN** Cervical Intraepithelial Neoplasia

**CRT** Chemoradiation Therapy

**DC** Dendritic Cell

**DEG** Differentially Expressed Gene

**DMEM** Dulbecco's Modified Eagle Medium

**DNA** Deoxyribonucleic Acid

**EGFR** Epidermal Growth Factor Receptor

**eGFP** Enhanced Green Fluorescent Protein

**FACS** Fluorescence-Activated Cell Sorting

**FBS** Fetal Bovine Serum

**FN1** Fibronectin 1

**GO** Gene Ontology

**GPI** Glycosylphosphatidylinositol

**GSEA** Gene Set Enrichment Analysis

**GFP** Green Fluorescent Protein

**HER2** Human Epidermal Growth Factor Receptor 2

**HIV** Human Immunodeficiency Virus

**HPA** Human Protein Atlas

**HPV** Human Papillomavirus

**HRP** Horseradish Peroxidase

**IMAC** Immobilized Metal Affinity Chromatography

**KEGG** Kyoto Encyclopedia of Genes and Genomes

**mRNA** Messenger RNA

**MPF** Megakaryocyte Potentiating Factor

**MSLN** Mesothelin

**mTOR** Mechanistic Target of Rapamycin

**NCL** Nucleolin

**NK** Natural Killer

**PBMC** Peripheral Blood Mononuclear Cell

**PCR** Polymerase Chain Reaction

**pRb** Retinoblastoma Protein

**PI3K** Phosphoinositide 3-Kinase

**PVDF** Polyvinylidene Difluoride

**RNA** Ribonucleic Acid

**ROS** Reactive Oxygen Species

**RPPA** Reverse Phase Protein Array

**SCC** Squamous Cell Carcinoma

**SDS-PAGE** Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

**scFv** Single-Chain Variable Fragment

**TCGA** The Cancer Genome Atlas

# Abstract

Cervical cancer remains a significant global health concern, ranking as the fourth most common cancer among women worldwide. This disease disproportionately affects women in low- and middle-income countries, where approximately 90% of cervical cancer deaths occur due to limited access to preventive measures, early detection, and adequate health-care (WHO 2022) [1]. In South Africa, cervical cancer poses a particularly urgent public health challenge, ranking as the third most frequent cancer among women and associated with a notably higher incidence and mortality rate compared to high-income countries (IARC 2020) [2].

This Master's thesis presents the rational design and initial characterization of a bispecific SNAP-tag-based antibody fusion protein targeting mesothelin (MSLN) and epidermal growth factor receptor (EGFR) for the selective detection and elimination of cervical cancer cells. Mesothelin, frequently overexpressed in cervical cancer and minimally present in healthy tissues, represents a promising primary therapeutic target. To overcome antigen heterogeneity and the challenge of mesothelin shedding, EGFR was identified and validated as a complementary co-target through integrative transcriptomic and proteomic analyses. In addition, a shedding-resistant anti-mesothelin antibody variant was designed to improve surface retention and therapeutic stability. The bispecific antibody constructs were successfully cloned, sequence-verified, and transiently expressed, and the resulting protein material was prepared for downstream applications including purification, site-specific conjugation to cytotoxic payloads (auristatin F), and functional validation. Continuation of these experiments has been delegated to further students to ensure sustained progress. Collectively, this research establishes a modular bispecific SNAP-tag ADC platform tailored for heterogeneous cervical tumors and provides a foundation for future preclinical evaluation, particularly in low-resource settings.

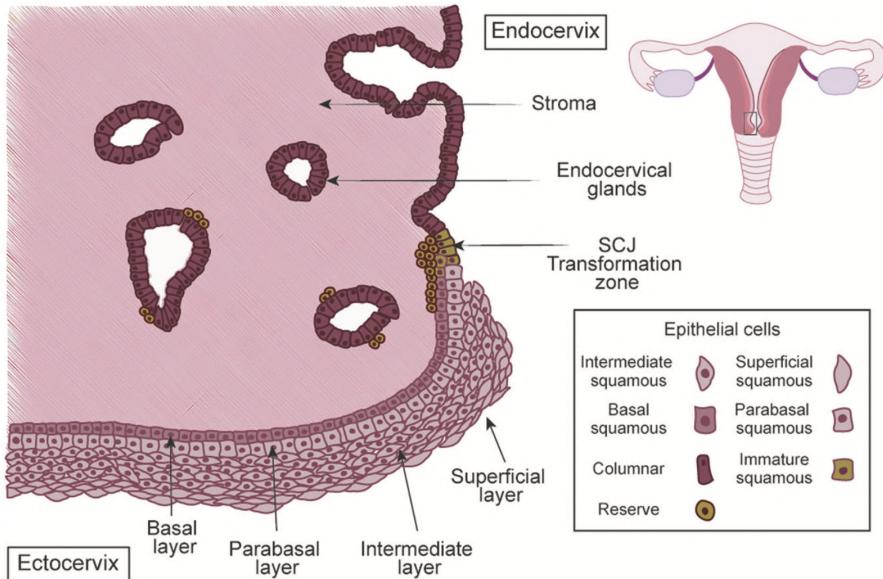
# 1 Introduction

Cervical cancer remains a major public health challenge worldwide, especially in low- and middle-income countries. Globally in 2020, an estimated 600,000+ new cervical cancer cases and over 340,000 deaths occurred, making it the fourth most common cancer and cause of cancer death in women (WHO 2022) [1]. The burden is disproportionately high in low-resource regions, which account for nearly 90% of cervical cancer deaths [1]. In sub-Saharan Africa, cervical cancer is often the leading cause of cancer-related death among women. For example, South Africa has one of the world's highest incidence rates, with an age-standardized rate over 30 per 100,000 and approximately 10,000 new cases annually (IARC 2020) [2]. Limited access to screening and HPV vaccination, coupled with a high prevalence of HIV co-infection, contributes to this elevated incidence in South Africa and other developing countries (*Cohen et al., 2019*) [3]. By contrast, high-income countries have seen declines in cervical cancer rates due to effective Pap smear screening and immunization programs. This stark disparity underscores the urgent need for improved preventive and therapeutic strategies tailored to high-burden settings.

## 1.1 Histological Subtypes

The cervix marks the transition between the vaginal epithelium and the uterine cavity, with two distinct epithelial types meeting at the squamocolumnar junction (SCJ-zone) as seen on Figure 1. The ectocervix and vagina are lined by stratified squamous non-keratinized epithelium, rich in glycogen to support the local microbiota. In contrast, the endocervical canal features a single-layered columnar epithelium that forms deep crypts responsible for mucus production, which varies with hormonal cycles. This abrupt epithelial transition is dynamic and prone to metaplasia—physiological cell-type replacement—especially during puberty, pregnancy, and menopause. The SCJ-zone is particularly vulnerable to oncogenic transformation, explaining why most cervical cancers arise here.

Cervical cancer is histologically classified primarily into squamous cell carcinoma (SCC) and adenocarcinoma, with rarer subtypes such as adenosquamous carcinoma and small-cell neuroendocrine carcinoma. These classifications and names are based on the type of cell from which the cancer originates.



**Figure 1: Histological and anatomical organization of the human cervix.** Schematic representation of the ectocervix and endocervix, highlighting the squamo-columnar junction (SCJ) and transformation zone. The stratified squamous epithelium of the ectocervix transitions into the columnar epithelium of the endocervix. Adapted from *Alzamil et al., 2020* [4].

## 1.2 Current state of research

### 1.2.1 Focus on Squamous Cell Carcinoma

Squamous cell carcinoma (SCC) of the cervix constitutes approximately 80-85% of cervical cancer cases. This claim is supported by evidence from multiple studies indicating the predominance of SCC in cervical malignancies, where it is recognized as the most prevalent form of cervical cancer. The remaining 15-20% of cervical cancers primarily comprise adenocarcinomas, which stem from glandular endocervical cells (*Wang et al., 2024*) [5].

The demographic distribution of these carcinoma types shows a notable trend: SCC is particularly prominent in low-resource settings and among older women. This is illustrated by the higher incidence rates observed in less affluent regions, likely due to variations in healthcare access, availability of screening programs, and public health initiatives (*Muntyanu et al., 2022*) [6]. Conversely, adenocarcinomas, while less common, have seen an increase in incidence in certain high-income populations. This trend can be attributed to advancements in cervical screening technologies such as Pap smears, which have historically been more efficient at identifying squamous dysplasia compared to glandular lesions (*Kang et al., 2020*) [7]. The differential detection capabilities of these screenings highlight the impact of healthcare systems on cancer epidemiology.

Given that SCC is by far the most prevalent subtype globally, this thesis will primarily focus on squamous cervical carcinoma. Focusing on SCC is justified not only by its high prevalence but also by specific therapeutic challenges associated with this subtype, especially for late stages. Cervical cancer survival rates vary widely depending on stage

at diagnosis, tumor type, and treatment access. For instance, research has shown that the survival rate for early-stage cervical cancer patients can be as high as 90%, particularly when detected through screening techniques such as Pap smears (*AshingGiwa et al., 2010*) [8]. Conversely, for patients diagnosed with metastatic disease, the five-year survival rate drops dramatically to around 15% (*Lorin et al., 2015*) [9]. Chemotherapies, chemoradiation (CRT), or adjuvant/neoadjuvant chemotherapy have only shown small improvements for advanced cases (*Xiong et al., 2021*) [10].

### 1.2.2 Etiology and Pathogenesis of Cervical Cancer

Persistent infection with high-risk human papillomavirus (HPV), especially types 16 and 18, is the primary cause of both squamous cell carcinoma (SCC) and adenocarcinoma of the cervix, accounting for approximately 70% of all cases (*Kusakabe et al., 2023*) [11], (*Asmiati, 2018*) [12], (*Su et al., 2023*) [13]. The transformation zone (or squamocolumnar junction) of the cervix is particularly vulnerable to HPV infection due to active epithelial remodeling and metaplasia, creating a permissive environment for viral entry and replication (*Ntuli et al., 2022*) [14].

HPV-induced carcinogenesis is driven by two major viral oncoproteins: E6 and E7. E6 promotes the degradation of the tumor suppressor p53, impairing cell cycle arrest and apoptosis, while E7 inactivates the retinoblastoma protein (pRb), leading to uncontrolled proliferation and genomic instability [13], (*Broni et al., 2024*) [15], (*Zhang, 2016*) [16]. These disruptions facilitate the progression from cervical intraepithelial neoplasia (CIN) to invasive carcinoma (*Fan et al., 2021*) [17], (*Kaczmarek et al., 2022*) [18]. Additionally, high-risk HPV types can induce epigenetic changes that further promote malignant transformation in both SCC and adenocarcinoma (*Silva et al., 2021*) [19].

Importantly, not all HPV infections lead to cancer. Most are transient and cleared by the immune system, but persistent infection with high-risk types significantly increases the likelihood of progression to CIN and ultimately invasive cancer (*Ofoezie et al., 2025*) [20]. The mucosal immune response, along with the vaginal microbiota, plays a crucial role in determining whether the infection is eliminated or persists [14].

### 1.2.3 Conventional Cervical Cancer Therapies and Prevailing Challenges

Surgery is still standard for stage I, and cisplatin-based concurrent chemoradiotherapy with high-dose-rate brachytherapy remains the work-horse for stages IB2–IVA; metastatic relapse continues to rely on platinum–taxane doublets, occasionally augmented with bevacizumab, PD-1 inhibitors, or, since 2024, the antibody–drug conjugate tisotumab vedotin (*Coleman et al., 2021*) [21], none of which has yet become a universally adopted, stage-specific standard (*Cibula et al., 2018*) [22]. In high-income countries, near-universal HPV vaccination and organised cytology/HPV screening intercept most lesions early, so recent innovation has focused on refining dose delivery (e.g., IMRT boosts) rather than inventing new systemic drugs (*Lu et al., 2022*) [23]. By contrast, many middle- and low-income settings face late presentation, limited radiotherapy capacity, and prohibitive drug costs, forcing clinicians to rely on the same chemoradiotherapy templates developed decades

ago (*Beddoe et al., 2019*) [24]; five-year survival after radiotherapy still falls below 50 % in several cohorts (*Irwanto et al., 2022*) [25]. This prevention-treatment gap explains why no widely used, bespoke therapy for advanced cervical cancer has emerged.

#### 1.2.4 Targeted Therapy: Promise and Limitations of Conventional Antibody-Drug Conjugates (ADCs)

First-generation antibody-drug conjugates (ADCs) built with random lysine or cysteine chemistry yield heterogeneous drug-to-antibody ratios (DAR 0–8) and thus erratic pharmacokinetics, efficacy, and toxicity (*Okojie et al., 2023*) [26]. A deeper hurdle, however, is biological: truly “ideal” surface biomarkers, highly and uniformly over-expressed across an entire patient population yet absent from vital healthy tissues are rare. Even canonical targets such as HER2 or Trop-2 are present in only a subset of cases or show variable intratumoural expression, which can drive primary resistance and on-target/off-tumour side-effects (*Qiu et al., 2025*) [27]. Reviews of clinical ADC pipelines consistently highlight the scarcity of robust, assayable biomarkers that meet all three criteria of abundance, homogeneity, and tumour specificity (*Katrini et al., 2024*) [28]; (*Hurwitz et al., 2023*) [29]. Consequently, current development strategies rely on narrower biomarker-guided patient selection, site-specific conjugation, or dual-payload formats to compensate, but these add cost and complexity and have yet to supplant conventional random-conjugate ADCs in routine oncology (*Muppa et al., 2025*) [30].

#### 1.2.5 SNAP-Tag Fusion Protein Technology for Antibody Targeting

Recent advances in protein engineering have enabled more precise, tailor-made antibody-based therapies. A key innovation utilized in this thesis is the SNAP-tag technology for site-specific conjugation to therapeutic payloads or imaging agents. The SNAP-tag is a 20 kDa engineered version of human O<sup>6</sup>-alkylguanine-DNA alkyltransferase that covalently and irreversibly binds benzylguanine (BG) derivatives. It was first described and utilized by the Prof. Dr. Kai Johnsson during his time at EPFL (*Keppler et al., 2003*) [31], (*Gautier et al., 2008*) [32]. By genetically fusing a SNAP-tag to an antibody fragment (such as a single-chain variable fragment, scFv), MB&I has created a unique antibody format for a one-step self-labeling of the antibody with any BG-modified molecule. This catalytic suicide reaction yields a 1:1 stoichiometric conjugation under physiological conditions, in contrast to traditional conjugation chemistries that randomly attach drugs to lysine or cysteine residues and produce heterogeneous mixtures [33].

The controlled site-specificity and stoichiometry of SNAP-tag conjugation offer multiple benefits: it produces a uniform antibody-drug conjugate (ADC) population with predictable pharmacokinetics, and it preserves antibody binding affinity by avoiding modification of the antigen-binding site. Furthermore, SNAP-tagged antibodies are extremely versatile – the same fusion protein can be conjugated with different functional moieties simply by swapping the BG-linked payload. For example, a SNAP-tag antibody can be coupled to fluorescent dyes for imaging, radionuclides for radio-immunotherapy, or cytotoxic drugs for targeted killing, all using the same conjugation platform. This “plug-and-play” flexibility is highly advantageous in a theranostic context.

In a seminal demonstration of this technology, (*Hussain et al., 2019*) [33] described a one-step procedure to produce scFv–SNAP fusion proteins and conjugate them with diverse effector molecules, including a photosensitizer for photodynamic therapy. They showed that the resulting conjugates retained specific tumor cell binding and functional activity, illustrating the potential of SNAP-tag ADCs as precision immunotheranostics. More recently, (*Huysamen et al., 2023*) [34] applied SNAP-tag conjugation to develop a targeted drug delivery system against EGFR-positive cancers. They generated an EGFR-specific scFv–SNAP fusion and attached a potent cytotoxin (auristatin F) via a click chemistry linker to a BG group, creating a well-defined ADC [34]. This approach yielded highly active conjugates that specifically killed EGFR-overexpressing tumor cells in preclinical models [34].

These studies highlight the feasibility and advantages of SNAP-tag-mediated antibody conjugation. Compared to other site-specific conjugation technologies such as Sortase A, SpyTag/SpyCatcher, or engineered cysteine/lysine coupling, the SNAP-tag system offers superior simplicity, stoichiometric control, and modularity under physiological conditions. Its covalent self-labeling mechanism ensures homogeneous conjugate populations and avoids complex purification steps or enzyme co-factors. Building on this platform, the present thesis will engineer SNAP-tagged antibody fusion proteins to target cervical cancer cells, enabling precise delivery of therapeutic agents to tumors.

### 1.2.6 Mesothelin as a Target in Cervical Cancer

Mesothelin (MSLN) is a cell-surface glycoprotein that has emerged as an attractive target in several cancers (*Lv et al., 2019*) [35]. MSLN is normally expressed at low levels on mesothelial cells lining the pleura, peritoneum, and pericardium, but is overexpressed in a variety of epithelial tumors—classically mesothelioma, pancreatic adenocarcinoma, and ovarian cancer (*Hassan et al., 2016*) [36]. The role of MSLN in tumor biology is not fully elucidated, but it may promote cell adhesion (via binding to CA125/MUC16) and facilitate tumor invasion and metastasis.

In cervical cancer, information on MSLN expression has until recently been limited. However, new evidence indicates that a substantial subset of cervical cancers express MSLN, albeit with differences between histological subtypes. A 2022 immunohistochemical study by Takamizawa et al. examined MSLN in 123 cervical cancer patients and found MSLN expression in 98% of tumors, with high expression (moderate to strong staining in  $\geq 30\%$  of cells) in 63% of cases (*Takamizawa et al., 2022*) [37]. Notably, MSLN overexpression was significantly more frequent in non-squamous tumors (approximately 80% of adenocarcinomas) compared to squamous cell carcinomas (about 49% of SCC cases showed high MSLN) [37].

This suggests that while many cervical cancers do express MSLN, squamous tumors, the majority of cases, often have lower MSLN levels relative to adenocarcinomas. Consistent with tissue data, cervical cancer cell lines show variable MSLN expression: for example, SiHa (HPV16 $^{+}$  squamous) and HeLa (HPV18 $^{+}$  adenocarcinoma) cells display

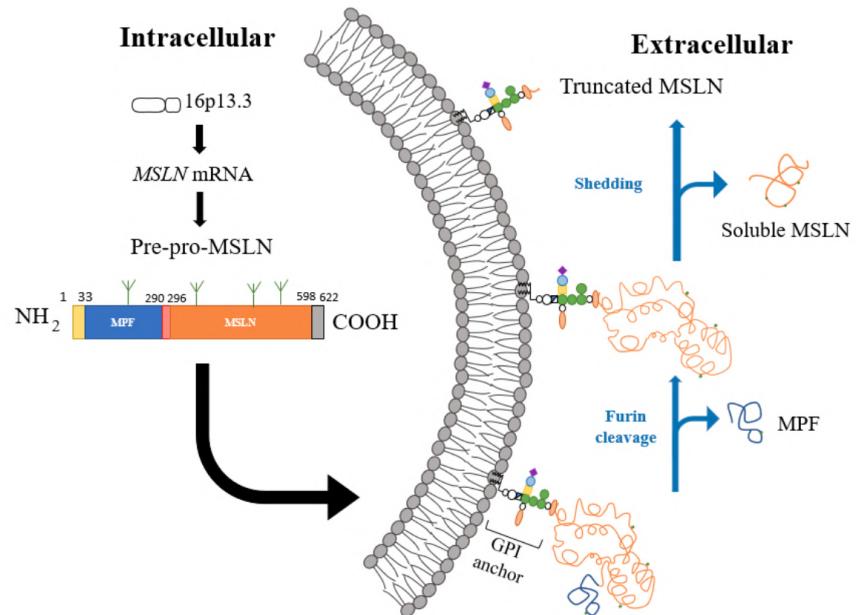
MSLN on their surface, while CaSki (HPV16<sup>+</sup> squamous) cells have minimal MSLN expression (*Kutle et al., 2024*) [38]. This variability implies that MSLN-targeted therapies could be very effective in MSLN-high cervical cancers, but may have limited impact in MSLN-low tumors.

**Therapeutic relevance:** Mesothelin's tumor-restricted expression and cell-surface localization have motivated multiple MSLN-targeted therapies (antibody-drug conjugates, immunotoxins, CAR-T/NK cells, and vaccines) in other cancers. In cervical cancer, early preclinical studies are promising. For instance, anti-MSLN CAR-NK cells showed potent killing of cervical cancer cell lines in vitro, confirming that MSLN can serve as a functional target on these cells [38].

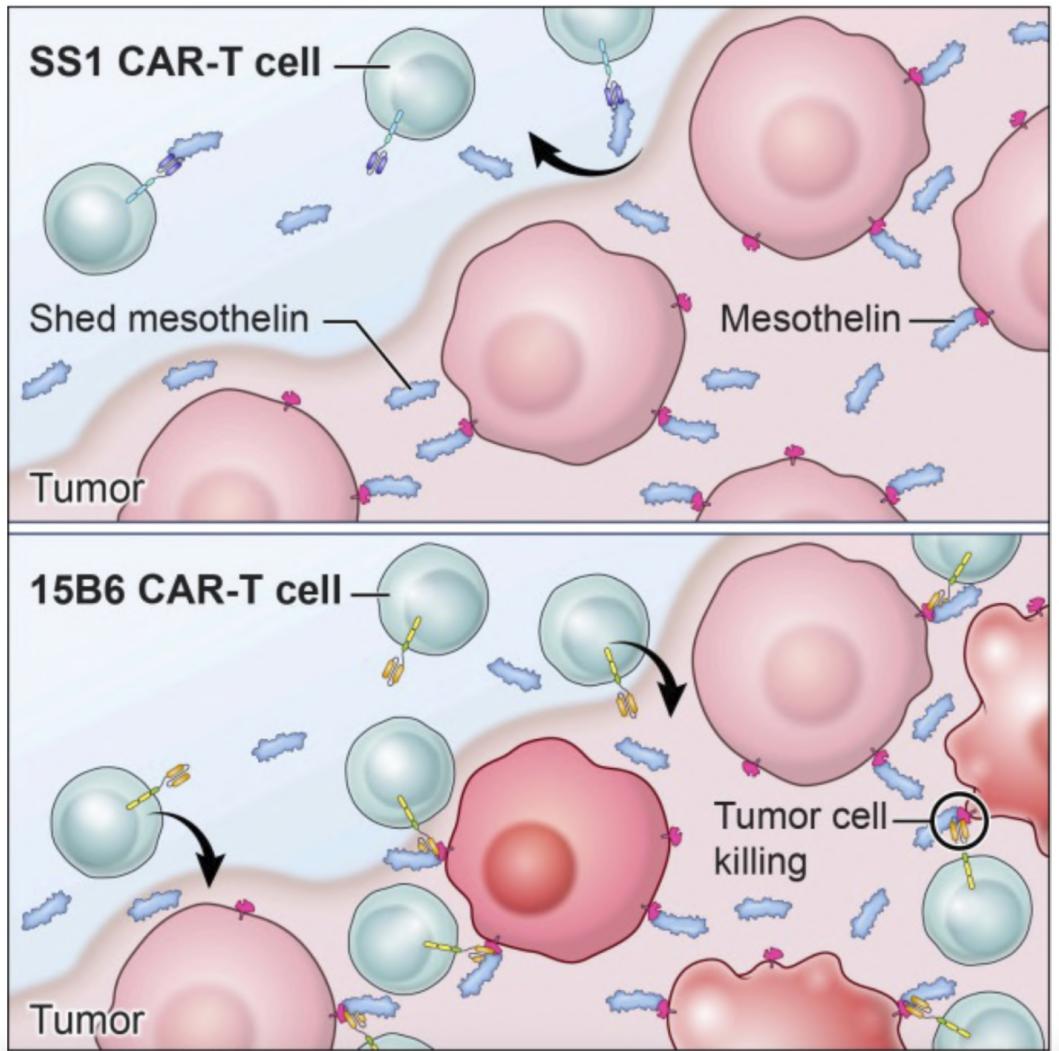
**Mesothelin shedding and therapeutic challenges:** One significant biological factor that can hinder mesothelin-targeted therapy is the shedding of MSLN from the tumor cell surface. MSLN is synthesized as a precursor protein and undergoes proteolytic cleavage into two fragments: a 40 kDa glycosylphosphatidylinositol (GPI)-anchored membrane-bound portion (the mature MSLN fragment) and a 30 kDa secreted fragment known as megakaryocyte potentiating factor (MPF). As illustrated in Figure 2, this process involves complex post-translational modification steps and contributes to the presence of soluble MSLN in extracellular fluids such as pleural effusions and serum (*Faust et al., 2022*) [39]. This soluble MSLN acts as a decoy, sequestering therapeutic antibodies and CAR T cells, thereby preventing them from efficiently binding to tumor cells expressing surface MSLN.

For example, shed MSLN has been shown to neutralize the MSLN-targeted immunotoxin SS1P and other antibody-based agents, reducing their therapeutic efficacy (*Liu et al., 2020*) [40]. Liu et al. demonstrated that inhibition of MSLN shedding through blockade of ADAM proteases, significantly enhanced the cytotoxic activity of an anti-MSLN immunotoxin, highlighting that shedding creates a “sink” effect that undermines antibody therapy.

More recently, as shown in Figure 3, Liu et al. (2024) illustrated how shed MSLN impairs CAR T cell recognition by saturating surface epitopes, and proposed that targeting a juxtamembrane epitope can overcome this barrier (*Liu et al., 2024*) [41]. The schematic shows that while traditional CARs bind to external epitopes blocked by soluble antigen, membrane-proximal targeting remains accessible and functionally active.



**Figure 2: Mesothelin maturation and shedding into membrane-bound and soluble fragments.** MSLN is translated as a precursor protein that undergoes cleavage and glycosylphosphatidylinositol (GPI) anchoring. Adapted from *Faust et al., 2022* [39].



**Figure 3: Shed mesothelin impairs CAR-T cell targeting and function.** Binding of SS1 CAR-T cells to shed mesothelin reduces tumor engagement, while targeting membrane-proximal epitopes with 15B6 CAR-T restores tumor cell killing. Adapted from *Liu et al., 2024* [41].

Most current MSLN-targeted therapies have been developed for tumors with uniformly high mesothelin expression, overlooking many cervical cancers—particularly squamous subtypes—that exhibit low or heterogeneous levels. To address this, the present work explores a bispecific antibody strategy combining MSLN with a secondary target, supported by bioinformatic analyses to identify co-expressed markers or actionable pathways. This approach aims to improve therapeutic coverage and mitigate challenges such as low target density or antigen shedding. For example, selecting membrane-proximal epitopes may help avoid cleavage-associated loss of surface MSLN (*Liu et al., 2024*) [41], (*Chakraborty et al., 2024*) [42]. While MSLN remains a promising and tumor-specific target, its limitations justify investigating additional antigens such as EGFR to broaden the efficacy of bispecific constructs in cervical cancer.

#### 1.2.7 Epidermal Growth Factor Receptor (EGFR) as a Co-Target for Bispecific Therapy

Cervical cancer is a molecularly heterogeneous disease, and the variability in mesothelin (MSLN) expression—particularly its low levels in many squamous cell carcinomas

(Takamizawa *et al.*, 2022) [37], limits the utility of MSLN as a standalone therapeutic target. This supports the rationale for bispecific antibody strategies that combine MSLN with an additional, broadly expressed antigen to enhance coverage while minimizing the need for multiple drug combinations.

Epidermal growth factor receptor (EGFR), a receptor tyrosine kinase implicated in epithelial tumorigenesis, emerges as a strong candidate for such co-targeting. EGFR is frequently overexpressed in cervical SCCs, with studies reporting protein-level prevalence from 50–60% [43] to as high as 90% [44]. Its elevated expression correlates with poor prognosis, resistance to chemoradiotherapy, and reduced survival [45, 46], reinforcing its relevance as a complementary target in bispecific therapeutic design.

Several EGFR-targeted therapies have been tested in cervical cancer, most notably monoclonal antibodies like cetuximab and small molecule inhibitors such as erlotinib and afatinib. These agents act by blocking EGFR signaling, leading to reduced proliferation and enhanced sensitivity to chemotherapeutic agents (*Lv et al.*, 2019; *Chen et al.*, 2020) [47, 48]. However, the clinical efficacy of these inhibitors in cervical cancer has been mixed, and to date, EGFR inhibition alone has not produced durable responses across patient populations.

In contrast, the approach explored in this project does not focus on blocking EGFR function, but rather leverages EGFR as a stable, overexpressed surface anchor to deliver cytotoxic payloads directly into tumor cells. By incorporating an EGFR-binding domain into a bispecific SNAP-tag antibody construct, EGFR becomes part of a targeting mechanism rather than a therapeutic target per se (*Huysamen et al.*, 2023) [34]. This strategy may offer several advantages: it allows selective delivery of toxin-conjugated agents even in the presence of MSLN heterogeneity or shedding, and may enhance tumor specificity, as dual expression of MSLN and EGFR is rare in normal tissues.

To support this dual-targeting strategy, a bioinformatics analysis of TCGA cervical cancer datasets was performed to assess the expression patterns of MSLN and EGFR across tumor samples. It also serves a complementary role by identifying other surface markers that could serve as future co-targets with improved selectivity and coverage in MSLN-low patient subgroups.

In summary, EGFR serves in this project not as a direct pathway to inhibit, but as a means to broaden tumor targeting and facilitate intracellular delivery of therapeutic agents within a bispecific SNAP-tag-based system. This design reflects a strategic pivot from classical EGFR blockade toward a modular targeting framework better suited for heterogeneous cervical tumors.

### 1.2.8 Therapeutic Payloads: Auristatin F

Auristatin F, a synthetic analog derived from dolastatin 10, is an antitumor agent known primarily for its incorporation into antibody-drug conjugates (ADCs). Its mechanism of action is predominantly through the induction of apoptosis via microtubule destabiliza-

tion, leading to cell cycle arrest.

*Mechanism of Action:* Auristatin F functions by inhibiting microtubule polymerization, which disrupts mitotic spindle formation, ultimately leading to cessation of cell division and induction of apoptosis (*Shor et al., 2016*) [49].

*Use in Antibody–Drug Conjugates (ADCs):* Auristatin F is often linked to monoclonal antibodies to create ADCs. For example, studies indicate that combining auristatin-based conjugates with inhibitors of the PI3K or mTOR pathways can yield synergistic effects (*Shor et al., 2016*) [49]. In cervical cancer, such strategies could improve targeting of tumors that express relevant surface antigens.

### 1.3 Additional study: Chondroitin sulfate proteoglycan 4 (CSPG4)

Chondroitin sulfate proteoglycan 4 (CSPG4), also recognized as NG2, is a glycoprotein significantly associated with various malignancies, particularly those of the central nervous system, melanoma, and breast cancer (*Ilieva et al., 2018*) [50]. Its expression is particularly notable in aggressive cancer phenotypes, where it is linked to key processes such as tumor angiogenesis, cellular proliferation, and metastasis (*Yang et al., 2022*) [51]. CSPG4 facilitates these malignant traits through the modulation of signaling pathways involved in cell survival and mobility, fostering an environment conducive to tumor progression [51]. Given its aberrant expression in tumors and its roles in enhancing invasiveness and resistance to conventional therapies, CSPG4 presents a compelling target for therapeutic strategies, including monoclonal antibodies and CAR-T cell therapies [50].

Previous studies have demonstrated the therapeutic potential of CSPG4-targeting strategies in cancer. Jordaan et al. (2017) [52] reported the successful delivery of cytolytic fusion proteins specifically to CSPG4-positive cancer cells, while Mungra et al. (2023) [53] developed a SNAP-tag-based antibody fusion protein targeting CSPG4, achieving specific killing of triple-negative breast cancer cells when conjugated to auristatin F. Building on this evidence, the current study aims to extend the application of this CSPG4-targeting strategy to cervical cancer. Although CSPG4 has not been extensively characterized in this cancer type, immunohistochemical data from the Human Protein Atlas indicate that CSPG4 is expressed in cervical cancer cell lines such as HeLa and CaSki, supporting its potential as a therapeutic target in this context.

### 1.4 Perspectives

Using click chemistry, a BG-modified auristatin F linker was synthesized (*Huysamen et al., 2023*) [34] for conjugation to the SNAP-tag on antibody fusions. Prior studies have shown the efficacy of scFv–SNAP conjugates carrying auristatin-based payloads in targeting EGFR<sup>+</sup> tumors both in vitro and in vivo [34, 54], supporting the therapeutic promise of this modular system.

In this project, a bispecific SNAP-tag-based antibody was designed to co-target mesothelin and EGFR, enabling selective delivery of cytotoxic agents even in heterogeneous or

MSLN-low cervical tumors. Auristatin F, a tubulin inhibitor, will serve as the primary payload. Together, these distinct mechanisms may provide insight into optimal ADC payload strategies for cervical cancer.

This approach addresses the central research question: *Can a bispecific antibody targeting EGFR and MSLN using a SNAP-tag fusion platform overcome the challenges of tumor heterogeneity and mesothelin shedding in cervical cancer?*

## 1.5 Aims and objectives

The overall goal of this project is to develop a bispecific antibody–drug conjugate (ADC) that targets cervical cancer through a SNAP-tag-based delivery platform. By combining two surface antigens—mesothelin (MSLN) and epidermal growth factor receptor (EGFR) into a single construct, the strategy aims to overcome limitations posed by antigen heterogeneity and mesothelin shedding, both of which can reduce the efficacy of conventional monotherapies.

This bispecific design could later be conjugated site-specifically to potent cytotoxic agents, including auristatin F and others, using benzylguanine (BG) linkers. The project also includes a parallel aim of designing a shedding-resistant anti-MSLN variant to further improve therapeutic stability.

### Hypotheses:

- A bispecific SNAP-tag antibody targeting MSLN and EGFR will enhance tumor cell coverage and selective toxicity in cervical cancer compared to monospecific constructs.
- EGFR can serve as a reliable co-target in MSLN-low tumors, expanding the therapeutic window.
- SNAP-tag-mediated conjugation of auristatin F will produce effective, well-defined ADCs with potent cytotoxicity in vitro.
- Engineering a shedding-resistant anti-MSLN antibody will increase the efficacy of mesothelin-directed ADCs by preserving surface antigen density.

This modular strategy, combining bispecific targeting, site-specific conjugation, and alternative payloads, has the potential to produce next-generation antibody therapeutics tailored to heterogeneous cervical cancers, particularly in high-burden, resource-limited settings.

## 2 Materials and Methods

### 2.1 Bioinformatic Analysis for Secondary Target Discovery

#### Software and Data

- **Python** (v3.9.18), Anaconda environment, Visual Studio Code (v1.99.3)
- **RStudio** (Version 2024.04.0+735)
- **Dataset:** TCGA cervical squamous cell carcinoma (PanCancer Atlas, cBioPortal, [https://www.cbiopal.org/study/summary?id=cesc\\_tcga\\_pan\\_can\\_atlas\\_2018](https://www.cbiopal.org/study/summary?id=cesc_tcga_pan_can_atlas_2018))

#### Libraries

- **Python:** pandas, numpy, matplotlib, seaborn, scipy.stats (pearsonr, spearmanr, ttest\_ind, mannwhitneyu), sklearn (PCA), umap-learn, CPTAC data
- **R:** DESeq2, limma, ggplot2, enrichR

#### Statistical Analyses

- Differential gene expression analysis performed with **DESeq2** (using raw counts)
- Differential protein expression analysis performed with **limma** (using log2-transformed and median centred raw RPPA counts)
- Chi-square tests utilized for categorical clinical variables
- Mann-Whitney U tests applied for continuous clinical variables
- Dimensionality reduction and visualization via PCA and UMAP (using z-score normalized data)
- Pearson and Spearman correlation analyses conducted for assessing mRNA-protein relationships
- Pathway enrichment analyses performed using **EnrichR** (KEGG 2021, WikiPathways 2024)

#### Selection Criteria

- **MSLN-high:** z-score  $> 1$ ; **MSLN-low:** z-score  $< -1$
- Clinical filtering restricted to cervical squamous cell carcinoma tumors to mitigate confounding variables

This structured and robust bioinformatics approach ensured the accurate identification of relevant biomarkers and biological pathways, contributing valuable insights into cervical cancer subtype differentiation and potential therapeutic targeting.

## 2.2 Cell culture

Human embryonic kidney cells (HEK293T) (ATCC: CRL-11268) were cultured in RPMI-1640 medium (containing 2 mM L-glutamine, 3.7 g/L NaHCO<sub>3</sub> and 15 mg/L phenol red), supplemented with 10% (v/v) heat-inactivated fetal bovine serum (FBS) and 1% (v/v) 100 U/ml penicillin-streptomycin.

The cells were maintained in a 5% CO<sub>2</sub> incubator with 95% humidity at 37 °C. Medium was changed every 3-4 days and cells were passaged when 90% confluent. All products were purchased from Gibco by Life Technologies® (RPMI-1640: Gibco #10566), supplied by Thermo Fisher Scientific, South Africa.

## 2.3 In silico design of mammalian expression vectors

The gene sequences of the variable heavy (VH) and light (VL) chains for the humanized  $\alpha$ MSLN (scFv) and  $\alpha$ EGFR (scFv) constructs were derived from previously characterized antibodies and sequence repositories. The  $\alpha$ EGFR (scFv) sequence was originally obtained from clone scFv-1711, while the  $\alpha$ MSLN (scFv) sequence corresponded to Human Mesothelin (residues 296–580). In parallel, sequences for the VHH-based constructs were acquired from VHH-G10 for  $\alpha$ EGFR(VHH) and VHH-SD2 for  $\alpha$ MSLN(VHH). These sequences had been previously verified by Sanger sequencing and were provided by collaborators, with validated clones stored in the laboratory plasmid collection. To confirm antibody gene identity and origin, all sequences were subjected to IgBLAST analysis.

**Table 1: Important components of the mammalian expression plasmid and their function**

Component	Function
Cytomegalovirus (CMV) promoter	Induces high-level expression of the fusion protein [55]
T7 promoter	Allows <i>in vivo</i> transcription in the sense orientation
<i>Ig-Kappa</i> leader	Crucial for the secretion of the fusion protein into the supernatant of HEK293T cells [56]
N-terminal polyhistidine tags (His × 6)	Useful for protein purification using immobilized metal affinity chromatography (IMAC) [57] and for detection of fusion proteins using an anti-His antibody
Enterokinase (EKS) cleavage site	Allows removal of the N-terminal His-tags and isolation of the fusion protein
SNAP-Tag	self-labeling protein derived from human O <sup>6</sup> -alkylguanine-DNA-alkyltransferase. SNAP-Tag reacts with covalently with O <sup>6</sup> -benzylguanine derivatives.
Chimeric intron	Enhances mRNA processing and increases the expression levels of the protein-coding genes [58]
bGH poly(A) signal	Allows efficient transcription termination and polyadenylation of mRNA
F1 origin	Allows rescue of single-stranded DNA

Continued on next page

Component	Function
SV40 Ori	Origin of replication
Internal ribosome entry site (IRES)	mRNA sequence which mediates expression of green fluorescent protein (GFP)
Green fluorescent protein (GFP)	Reporter protein showing successful transfection in mammalian cells, as well as <i>in cis</i> expression of the putative fusion protein
BleoR	Allows selection of transient transfectants in mammalian cells (using Zeocin)
Lac operator	Short region of DNA that interacts with the lac repressor to inhibit transcription of lac genes in the absence of lactose
Lac promoter	Modulates transcription of lac operon genes involved in the metabolism and absorption of lactose in bacteria
SV40 promoter	Allows high-level expression and replication in cell lines which express the large T antigen
Ampicillin	Resistance gene, allows selection in <i>E. coli</i>
M13 rev	Single-stranded oligonucleotide sequence (5'-CAG GAA ACA GCT ATG ACC-3') which can be used in polymerase chain reactions [59]
CAP binding site	Binding site for catabolite activator protein (CAP), which is involved in the transcription of several genes, including enzymes involved in sugar metabolism [60]
<i>SfiI/NotI/BlpI/EcoRI</i>	Restriction sites

## 2.4 Molecular cloning of SNAP-tag based fusion proteins

### 2.4.1 Transformation of *E. coli* with plasmid DNA

The exogenous DNA of interest (pCB- $\alpha$ MSLN(scFv)-SNAP- $\alpha$ MSLN(scFv), pCB- $\alpha$ EGFR (scFv)-SNAP- $\alpha$ MSLN(scFv), pUC57 -  $\alpha$ MSLN(m15B6) (scFv) ) was incorporated into chemically (calcium) competent DH5 $\alpha$  *E. coli* cells (K12 strain, genotype: fhuA2 (argF-lacZ)U169 phoA glnV44 80 (lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17, New England Biolabs, USA) using the heat shock method. The bacterial cells were firstly thawed on ice for 10 minutes. Thereafter, 0.5  $\mu$ l of plasmid DNA (100 ng/ $\mu$ l) (resuspended in molecular grade water) was added to 50  $\mu$ l of bacterial cells. The mixture was carefully mixed by flicking the tube, which was then placed on ice for 30 minutes and heat shocked at 42°C for 60 seconds in a water bath. The tube was then placed on ice for 5 minutes and 950  $\mu$ l of SOC (Super Optimal broth with Catabolite repression, containing 2% (w/v) vegetable peptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl<sub>2</sub>, 10 mM MgSO<sub>4</sub> and 20 mM glucose) outgrowth medium was added to the mixture to help in the recovery of the cells. The full 1 ml mixture was then inoculated in 50 ml LB (Luria broth, containing 1% (w/v) Casein Peptone, 0.5% (w/v) yeast extract and 1.0% (w/v) NaCl) liquid medium (Thermo Fisher Scientific, South Africa) supplemented with 200  $\mu$ g/ml ampicillin antibiotic and placed on a shaker at

37°C overnight.

#### 2.4.2 Isolation of plasmid DNA from *E. coli*

The plasmid DNA was isolated from *E. coli* and purified using the NucleoBond® plasmid purification kit (Product number: 740573, Macherey-Nagel, United Kingdom) according to the manufacturer's instructions. The principle of this assay relies on the alkaline lysis of bacterial cells for the extraction of plasmid DNA. The DNA is then bound to an anion-exchange resin in the NucleoBond® column, from where it is eluted off and precipitated using isopropanol. In the last step, 30  $\mu$ l of molecular grade water (pre-heated at 37°C for 10 minutes) was used to resuspend the DNA pellet instead of TE buffer. This step was done because the presence of chemicals in TE buffer (EDTA and Tris) can affect the quality of the DNA samples that are meant to be used in downstream applications. The amount of DNA recovered was then quantified using a NanoDrop One (Thermo Fisher Scientific, USA).

#### 2.4.3 Restriction enzyme digest

Restriction digestion was used for the generation of compatible ends suitable for ligation. Starter plasmid DNA (pCB-H22-SNAP-H22, pCB- $\alpha$ MSLN(scFv)-SNAP, pCB- $\alpha$ EGFR(scFv)-SNAP, pUC57 -  $\alpha$ MSLN(m15B6) (scFv) and the intermediate constructs) (in molecular grade water) was incubated successively with New England Biolabs (NEB, USA) restriction enzymes NotI (for 1:30 hours at 37°C), followed by SfiI (for 1:30 at 50°C). In case the section downstream the SNAP needed to be digested, the plasmid DNA was incubated simultaneously with New England Biolabs (NEB, USA) restriction enzymes EcoRI-HF and BlpI (for 1:30 hours at 37°C) as indicated in Table 2 below.

**Table 2: Setting up the restriction enzyme digest reaction mixture**

Reagents	Amount required
<i>SfiI</i> , <i>NotI-HF</i> , <i>BlpI</i> , or <i>EcoRI-HF</i> (2000 units/ml)	2 $\mu$ l
10× NEB CutSmart buffer	5 $\mu$ l
Plasmid DNA (added last)	2 $\mu$ g
Molecular grade water	Adjust to final volume of 50 $\mu$ l

#### 2.4.4 Agarose gel electrophoresis

Following restriction digest, the DNA fragments were separated using agarose gel electrophoresis. A 1.2% (w/v) agarose gel was made using a 1x TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA, pH 8.5). For visualization and tracking of the DNA bands, the agarose gel was mixed with SYBR® Safe DNA gel stain (Thermo Fisher Scientific, South Africa). Additionally, the DNA samples were mixed with 6x Purple DNA Loading Dye (New England Biolabs, USA) (supplied with the restriction enzymes) before loading into the wells. Electrophoresis was performed at 100 volts for 60 minutes using Wide Mini Sub-Cell systems (Bio-Rad, USA) and the DNA fragments were visualized

upon exposure to blue light excitation (using a Dark Reader Transilluminator from Clare Chemical Research, USA).

#### 2.4.5 Recovery of DNA fragments from agarose gels

The digested DNA fragments were excised from the agarose gels and purified using the QIAquick® Gel Extraction Kit (Product number: 28704, Qiagen, Netherlands) according to the manufacturer's instructions. This gel extraction method relies primarily on the binding of DNA fragments to a silica-membrane-based column. The DNA was eluted by adding 20  $\mu$ l of distilled water (pre-heated at 37°C) to the center of the membrane, before centrifuging the samples for 2 minutes at 14,000 rpm. Thereafter, the DNA recovered was quantified (as described above) and stored at 4°C until ready for use.

#### 2.4.6 Ligation reactions and transformation of recombinant plasmids in *E. coli*

Ligation of the DNA fragments to the pCB-SNAP vector backbone was carried out using T4 DNA ligase (400,000 units/ml) and 10x ligation buffer (both from New England Biolabs, USA) according to the manufacturer's instructions. Reactions were set up as indicated in Table 6 and incubated overnight at 16°C. The amount of insert required for a 1:1, 1:2 and 1:3 (vector: insert) ratio was calculated using the NEBioCalculator (<https://nebiocalculator.neb.com/#!/ligation>). This was done to maximize the success rate of the ligation experiments. A 'vector only' control double digested pCB-H22-SNAP-H22 / pCB- $\alpha$ MSLN(scFv) - SNAP -  $\alpha$ MSLN(scFv) was also set up to assess any possible vector re- ligation. Next day, the reaction mixture was heat-inactivated at 65°C for 10 minutes and 5  $\mu$ l of the recombinant DNA was then transformed into 50  $\mu$ l of chemically competent DH5 $\alpha$  E. coli cells (New England Biolabs, USA). The mixture was incubated on ice for 30 minutes, heat shocked for 60 seconds at 42°C in a water bath and cooled on ice for a further 5 minutes. Thereafter, 950  $\mu$ l of SOC solution was added to the mixture, which was then incubated at 37°C for 1 hour in a heating block. The cells were then mixed thoroughly, centrifuged (14,000 rpm for 2 minutes) and 900  $\mu$ l of supernatant was removed. The remaining pellet was resuspended in a final volume of 200  $\mu$ l with LB liquid medium and 100  $\mu$ l of cells were plated on LB agar plates (1% (w/v) Casein Peptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl and 1.2% (w/v) agar) supplemented with 200  $\mu$ g/ml ampicillin (including 'vector only' and bacterial controls). The plates were left overnight at 37°C. The next day, 3 single colonies were picked from each plate and grown overnight in 2 ml of LB medium containing ampicillin (200  $\mu$ g/ml). The transformation efficiency in colony forming units (cfu) per  $\mu$ g of DNA was calculated using the following formulae:

$$\text{Transformation efficiency} \left( \frac{\text{cfu}}{\mu\text{g}} \right) = \frac{\text{Number of transformed colonies}}{\text{Amount of DNA spread on the plate} (\mu\text{g})}$$

**Table 3: Setting up ligation reactions of digested insert with vector backbone**

Component	Amount required
Vector DNA	50 ng
Insert DNA	Varies depending on fragment size (use NEBioCalculator)
T4 DNA ligase buffer (10×)	2 $\mu$ l
T4 DNA ligase (400,000 units/ml)	1 $\mu$ l (added last)
Molecular grade water	Adjust to final volume of 20 $\mu$ l

#### 2.4.7 Small-scale DNA isolation, restriction mapping and DNA sequencing

The recombinant DNA plasmids were isolated and purified from *E. coli* using the ZippyTM Plasmid Miniprep Kit (Product number: D4036, Zymo Research, USA) according to the manufacturer's instructions, with minor alterations. In the first step, 1.2 ml of the bacterial culture was centrifuged (14,000 rpm for 2 minutes), 600  $\mu$ l of supernatant was removed and the pellet was resuspended in the remaining 600  $\mu$ l of LB medium. This was done to maximize the number of bacterial cells present and thus increase the resulting DNA yield. Single restriction enzyme digest simulations were carried out on SnapGene software (version 3.1.1, GSL Biotech, Chicago) to predict the cutting patterns of PvuII-HF for pCB- $\alpha$ MSLN(scFv)-SNAP- $\alpha$ MSLN(scFv) against pCB- $\alpha$ MSLN(scFv)-SNAP-H22, SacI for pCB- $\alpha$ EGFR(scFv)-SNAP- $\alpha$ MSLN(scFv) against pCB-H22-SNAP- $\alpha$ MSLN(scFv) and, NheI and BamHI for pCB- $\alpha$ MSLN(m15B6)(scFv)-SNAP against pCB-2F9(scFv)-SNAP. Digestion reactions were then set up with 2  $\mu$ g of purified DNA, 5  $\mu$ l of 10x NEB Cutsmart buffer, 2  $\mu$ l of restriction enzyme and topped up to 30  $\mu$ l with molecular grade water. Samples were incubated at 37°C for 1:30 hours and analyzed using agarose gel electrophoresis as described above. This allowed identification and confirmation of correct ligation of recombinant plasmids, by comparing with the SnapGene simulations. Correct recombinant clones were then grown in LB media and purified using the NucleoBond® plasmid purification kit as mentioned before (section 2.4.2). Construct integrity was validated by Sanger sequencing, performed by Inqaba Biotechnical Industries (Pretoria, South Africa). The following primers (5'-3') were used for sequencing:

- **CMV-for:** CGCAAATGGCGGTTAGCGTG
- **S-Tag:** GAACGCCAGCACATGGACA
- **SNAP-R:** GCTGGCGGCCCTGG
- **BGH-rev:** CTAGAAGGCACAGTCGAGGC

#### 2.4.8 PCR amplification of plasmid fragment

Site-directed mutagenesis was performed using insertion-PCR by designing primers incorporating the *BspI* and *EcoRI* restriction sites at the 5'- and 3'-ends of  $\alpha$ -MSLN(scFv) OFR, respectively. To amplify the  $\alpha$ -MSLN(scFv) antibody fragment for downstream cloning, PCR was performed using Phusion® High-Fidelity DNA Polymerase (NEB,

M0530L), following the manufacturer's standard protocol. Primers were designed to incorporate restriction sites compatible with the destination vector: a BlpI site (5'-**GCTNAGC**-3') in the forward primer and an EcoRI site (5'-**GAATTC**-3') in the reverse primer. Oligonucleotides were synthesized by Inqaba Biotechnical Industries (South Africa).

- Forward primer: 5'-TAAGCAG**GCTNAGC**GGCCAGGCCAT-3'
- Reverse primer: 5'-TAAGCAG**GAATT**CCTTTATTCCAACTTGTCCCAGCA-3'

PCR reactions (50 µL) contained 1× Phusion HF Buffer, 200 µM dNTPs, 0.5 µM each primer, 1 ng template plasmid, and 1 U polymerase. Thermal cycling: 98°C for 30 s; 30 cycles of 98°C for 10 s, 60°C for 20 s, 72°C for 30 s; and a final extension at 72°C for 5 min.

PCR products were run on a 1.2% agarose gel after digestion with BlpI and EcoRI prior to ligation into the expression vector (sections 2.4.4, 2.4.3) Following digestion with *BlpI* and *EcoRI*, the PCR product was gel-purified and ligated into the corresponding digested vector backbone to generate the downstream MSLN insert.

#### 2.4.9 Expression of the fusion proteins in HEK293T cells

Upon confirmation of the DNA sequences, the eukaryotic expression vector systems (1 µg/µl) were transiently transfected into HEK293T cells (at 70-80% confluence) using XtremeGENE™ transfection reagent (Roche, Switzerland) according to the manufacturer's instructions. This procedure utilizes lipids and polymers that are capable of complexing with DNA to form micelles, which in turn facilitate the uptake of DNA into mammalian cells. Different ratio of DNA to transfection reagent were used ( 1:3 - 3 µl of DNA and 9 µl of transfection reagent, 1:3 - 5 µl of DNA and 15 µl of transfection reagent, 1:2 - 5 µl of DNA and 10 µl of transfection reagent) and topped up to a 200 µl total volume with a serum-free and antibiotic-free RPMI-1640 (Gibco #10566, containing 2 mM L-glutamine, 3.7 g/L NaHCO<sub>3</sub> and 15 mg/L phenol red). Untransfected HEK293T cells were included as a negative control. The transfected cells were then grown in RPMI-1640 culture medium (Gibco #10566) supplemented with 10% (v/v) FBS and 1% (v/v) 100 U/ml penicillin-streptomycin. Plasmid uptake was then assessed after 3 days, through the microscopic visualization of enhanced green fluorescent protein (eGFP) expressed by the transfected cells, using a ZOE™ Fluorescent Cell Imager (Bio-Rad Laboratories, UK). To determine the transfection efficiency, 2 ml of the transfected cells and controls were subjected to flow cytometry using the BD™ LSR II flow cytometer (BD Biosciences, USA). Transfection efficiency was expressed as a percentage of the eGFP-positive cells present within the total population. Zeocin selection (100 mg/ml, ThermoFisher n°R25001) was then applied to enrich the eGFP-positive cells containing the recombinant plasmids. These cells were then grown at 90% confluence and the cell culture supernatant containing the secreted protein of interest, was harvested every 4 days. The optimal period of harvesting is 6± months or until sufficient protein ( ≥ 1 mg/ml) is obtained. Therefore this thesis did not go further after that step because of lack of time to collect sufficient proteins.

However, to gain practical experience with the downstream purification process, an additional experiment was carried out using a previously produced  $\alpha$ CSPG4(scFv)-SNAP construct, as described below.

#### 2.4.10 Protein purification using Immobilized Metal Affinity Chromatography (IMAC)

The cell culture supernatant (3 parts of cell culture supernatant was mixed with 1 part of 4 $\times$  incubation buffer: 200 mM NaH<sub>2</sub>PO<sub>4</sub>, 1.2 M NaCl, 40 mM imidazole, pH 8.0) of  $\alpha$ CSPG4(scFv)-SNAP was first filtered using the Nalgene<sup>TM</sup> vacuum filtration system (Sigma-Aldrich, South Africa) containing a 0.22  $\mu$ m Durapore<sup>®</sup> membrane filter (Millipore, USA) to exclude any microcellular debris, before purification by Immobilized Metal Affinity Chromatography (IMAC).

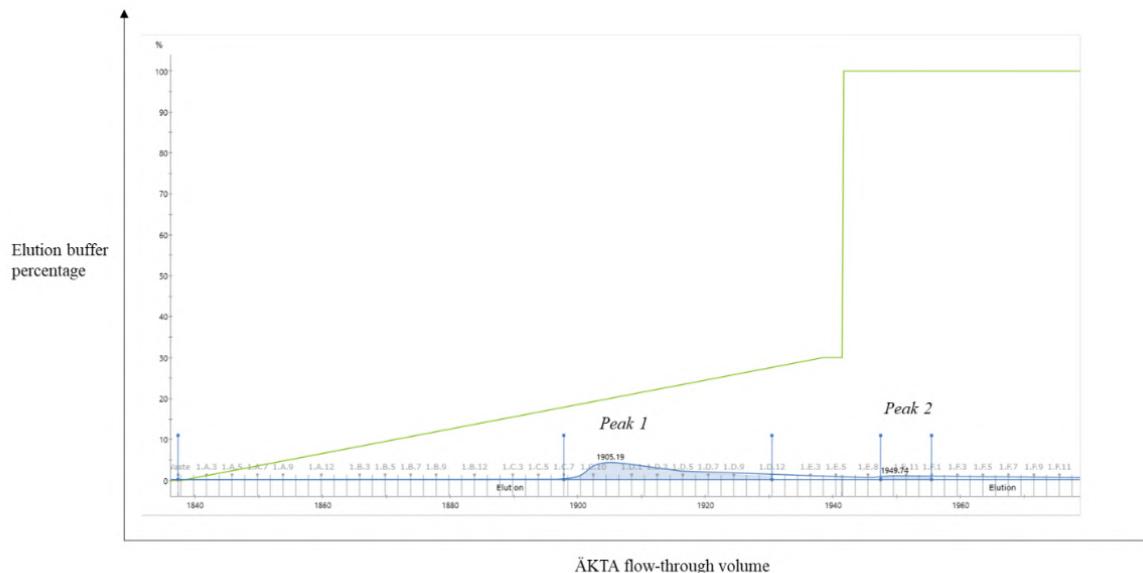
IMAC was carried out using a Ni<sup>2+</sup> sepharose affinity resin (packed in a HisTrap<sup>TM</sup> Excel column, GE Healthcare, USA) on an ÅKTA Avant protein purification system (GE Healthcare, USA). Initially, each clarified cell culture supernatant was applied to a pre-equilibrated HisTrap<sup>TM</sup> Excel column (equilibration buffer: 50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, pH 8.0) at a flow rate of 3 mL/min. The column was then washed with 20 column volumes of washing buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, pH 8.0), and the bound fusion proteins were eluted using elution buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 500 mM imidazole, pH 8.0). In principle, fractional elution of the His-tagged fusion proteins is made possible by increasing the concentration of imidazole, which competes with histidine for binding to the metal-charged resin. As illustrated in Figure 4, the combination of a gradient (0–30% imidazole) and step (100% imidazole) elution process (green graph) allows complete removal of the fusion proteins from the column, observed as two distinct peaks on the chromatogram (blue graph).

Thereafter, in order to concentrate the eluted fractions (as well as remove residual imidazole), 30K-sized Amicon filters (Sigma-Aldrich, South Africa) were used. The samples were centrifuged at 4500  $\times$  g for 20 minutes at 4 °C and washed three times in 1 $\times$  phosphate-buffered saline (PBS; pH 7.4) (containing 137 mM NaCl, 8.8 mM Na<sub>2</sub>HPO<sub>4</sub>, 2.7 mM KCl, and 1.75 mM KH<sub>2</sub>PO<sub>4</sub>) prior to downstream assays.

*Note:* 1 $\times$  PBS was also used as the protein storage buffer at 4 °C. Protein quantification was assessed using a NanoDrop One (Thermo Fisher Scientific, USA) prior to further characterization.

#### 2.4.11 Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and western blot analysis of recombinant protein fractions

A discontinuous 10% SDS-PAGE gel was used to resolve proteins electrophoretically on the basis of their molecular weights. For sample preparation, 15  $\mu$ l of the recombinant protein samples was mixed with 5  $\mu$ l of 4 $\times$  Laemmli protein sample buffer (Bio-Rad, USA) supplemented with 10% (v/v) 2-mercaptoethanol (Sigma-Aldrich, South Africa),



**Figure 4: Chromatogram exemplifying the applied imidazole concentrations and the expected elution peaks.** After application of the clarified cell culture supernatant onto the Ni<sup>2+</sup> sepharose column and the washing of low affinity proteins, the high affinity his-tagged fusion proteins were eluted using a high concentration of imidazole. This was done via combination of a washing gradient (0-30% imidazole) and step elution process (100%) (green graph), resulting in the formation of 2 distinct protein peaks on the chromatogram (blue graph).

and heated at 95 °C for 5 minutes. After loading the protein samples and the PageRuler™ prestained protein ladder (5 µl) (Thermo Fisher Scientific, South Africa), the SDS-PAGE gel was run at 100 V for 95 minutes on the Mini-PROTEAN™ Tetra Cell system (Bio-Rad, USA). The protein bands were visualized by staining the gel with Aqua Staining Solution (Vacute, South Africa). Densitometry measurements were carried out using *ImageJ* v1.52a software (<https://imagej.nih.gov/ij/download.html>), which compares the optical densities of the target band of interest against positive control bands (bovine serum albumin (BSA)) on the same gel. Two-fold serial dilutions of BSA (Thermo Fisher Scientific, South Africa) were used for the generation of a standard curve of optical colour intensity against the quantity of protein (in µg), allowing for the yield of the SNAP-tag fusion proteins to be estimated from the total protein concentration.

Subsequently, western blotting was used to confirm the functionality and integrity of the recombinant His-tagged fusion proteins. Protein bands were transferred from an unstained SDS-PAGE gel to a nitrocellulose membrane (PVDF transfer membrane) using the Trans-Blot Turbo™ Transfer System (Bio-Rad, USA) with the Standard SD protocol. The membrane was then blocked with non-fat milk for 1 hour at room temperature and incubated overnight with a 1:1000 dilution of an anti-His antibody directly conjugated to horseradish peroxidase (HRP) (Sigma-Aldrich, Clone A7058-1VL). . The PVDF membrane was activated by three cycles of 1-minute incubation in molecular-grade water followed by 1-minute incubation in methanol. Signal development was performed using Pierce™ 1-Step Ultra TMB-Blotting Solution (Thermo Fisher Scientific, South Africa), a chromogenic substrate that enables direct colorimetric detection.

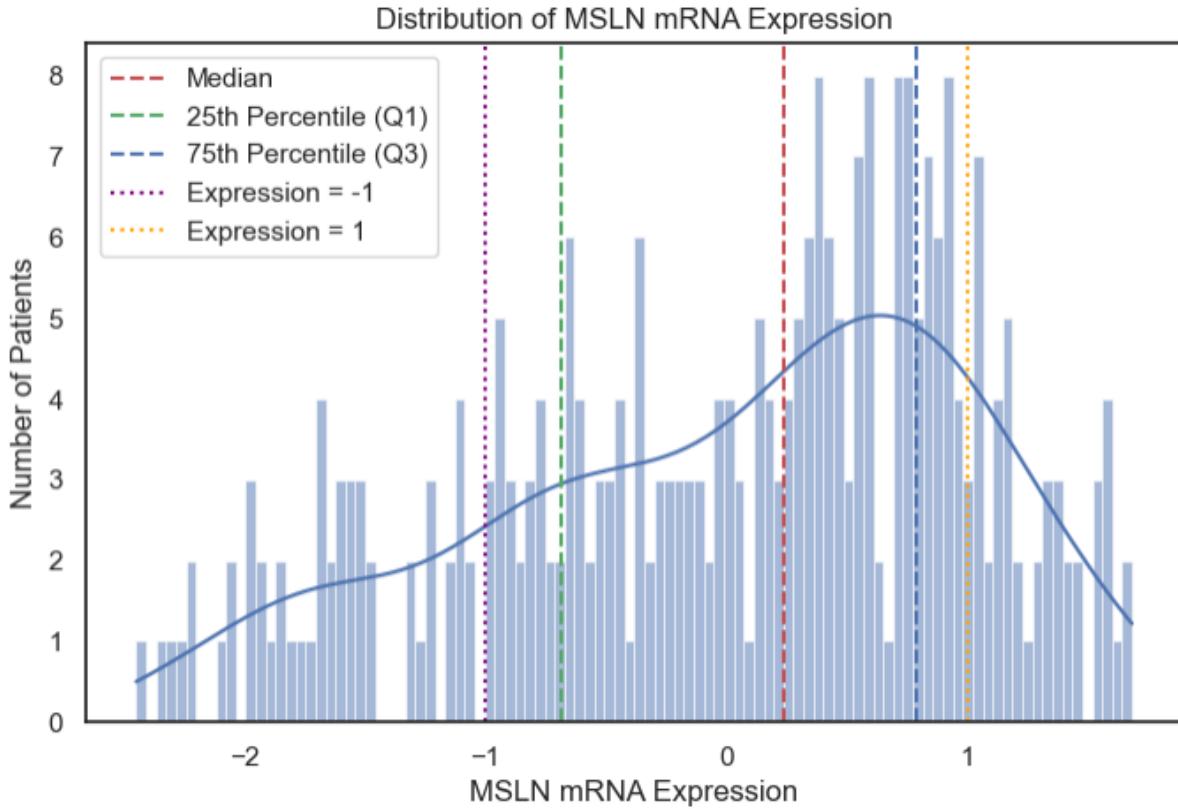
## 3 Results

### 3.1 Discovery and First Characterisation of a New Cervical Cancer Subtype: A Bioinformatic Analysis

The primary objective of this bioinformatic analysis was to enhance therapeutic coverage in cervical cancer by identifying complementary targets to mesothelin (MSLN). While MSLN represents a promising therapeutic target due to its frequent overexpression in cervical cancers, tumor antigen heterogeneity limits the effectiveness of MSLN-targeted approaches. A similar strategy has previously been explored in breast cancer, where co-targeting nucleolin (NCL) alongside MSLN significantly improved coverage in triple-negative breast cancer subtypes (*Thongchot et al., 2024* [6]). Surprisingly, when investigating the existence of a mesothelin-low or mesothelin-negative subtype within cervical cancer, a notable lack of dedicated research was identified. Only two recent publications noted the presence of cervical tumors with low or absent mesothelin expression, without characterizing these observations further (*Takamizawa et al., 2022*) [37], (*Kutle et al., 2024*) [38]. Given this gap, this bioinformatic analysis aims first to determine the validity and relevance of defining such a subtype within cervical cancer, and subsequently, if justified, to characterize its molecular signature by identifying relevant pathways and surface biomarkers for targeted therapeutic interventions.

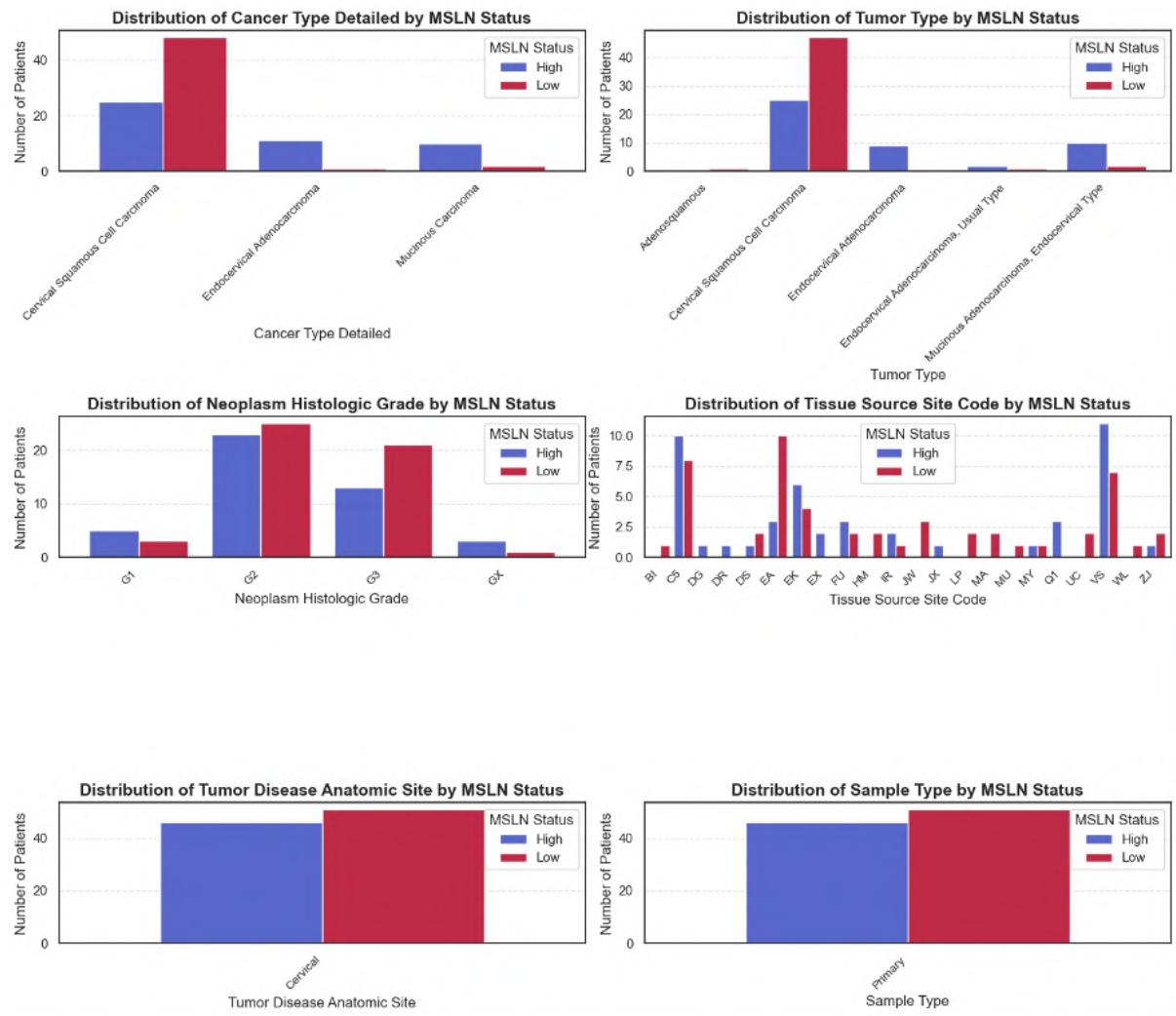
#### 3.1.1 Dataset cleaning and first observation

Data preprocessing involved removing low-quality gene data, resulting in a dataset containing 20,001 genes across 294 samples. MSLN mRNA expression levels z-score normalized dataset was used to facilitate comparisons by mitigating scale differences across samples. Samples with z-scores below -1 were classified as MSLN-low (n=51), while those above +1 were considered MSLN-high (n=46) (Figure 5), aligning with existing literature data [37] that suggested the same proportions of MSLN-low tumours.



**Figure 5: Distribution of MSLN mRNA expression in the dataset.** Histogram and kernel density plot of MSLN mRNA z-scores across all samples, annotated with percentile cutoffs and expression thresholds used for classification.

Clinical data analyses confirmed that MSLN expression significantly correlated with cancer and tumor types, notably adenocarcinoma, potentially confounding initial classifications (Chi-square test,  $p < 0.001$ ). To avoid this bias, only cervical squamous cell carcinoma (SCC) cases were selected, resulting in a refined cohort of 72 samples: 25 MSLN-high and 47 MSLN-low. This step ensured biological distinctions rather than clinical variability explaining the observed differences (Figure 6).



**Figure 6: Clinical characteristics of MSLN-high and MSLN-low tumor groups.** Bar plots comparing MSLN expression subgroups across clinical and histological parameters. Tumor/cancer type is correlated to MSLN expression.

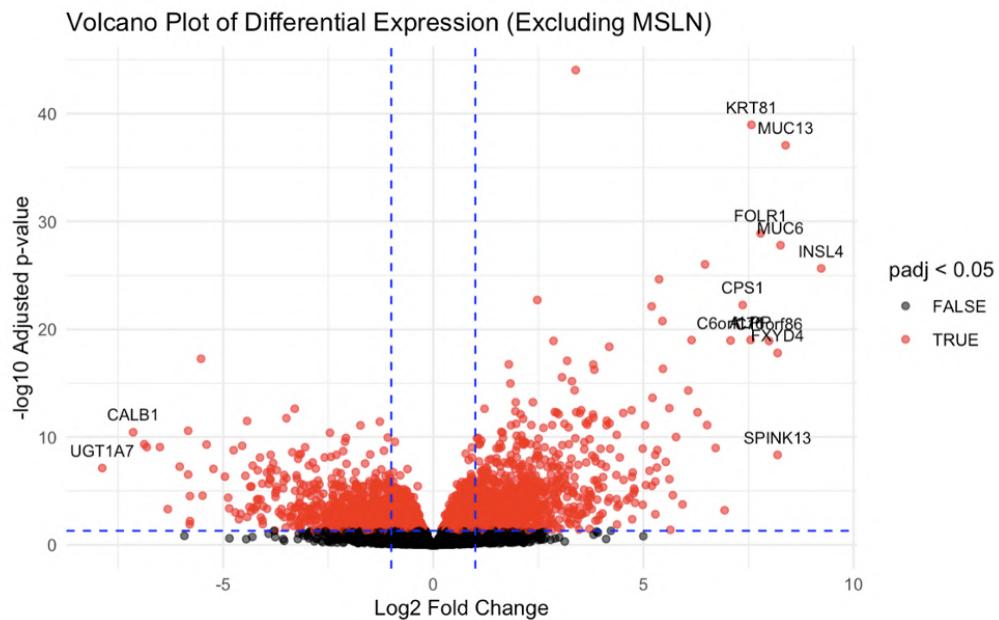
Quantitative clinical variables revealed significant differences in Aneuploidy and MSIensor scores (Mann-Whitney U test,  $p = 0.0025$  and  $p = 0.0071$ , respectively). These differences reflect genomic instability and mutation-driven tumor biology in the groups, warranting further investigation. However, no additional data were excluded, as these findings were not considered confounders but rather potential insights into cervical cancer biology that may be valuable for future studies.

### 3.1.2 Subtypes identification

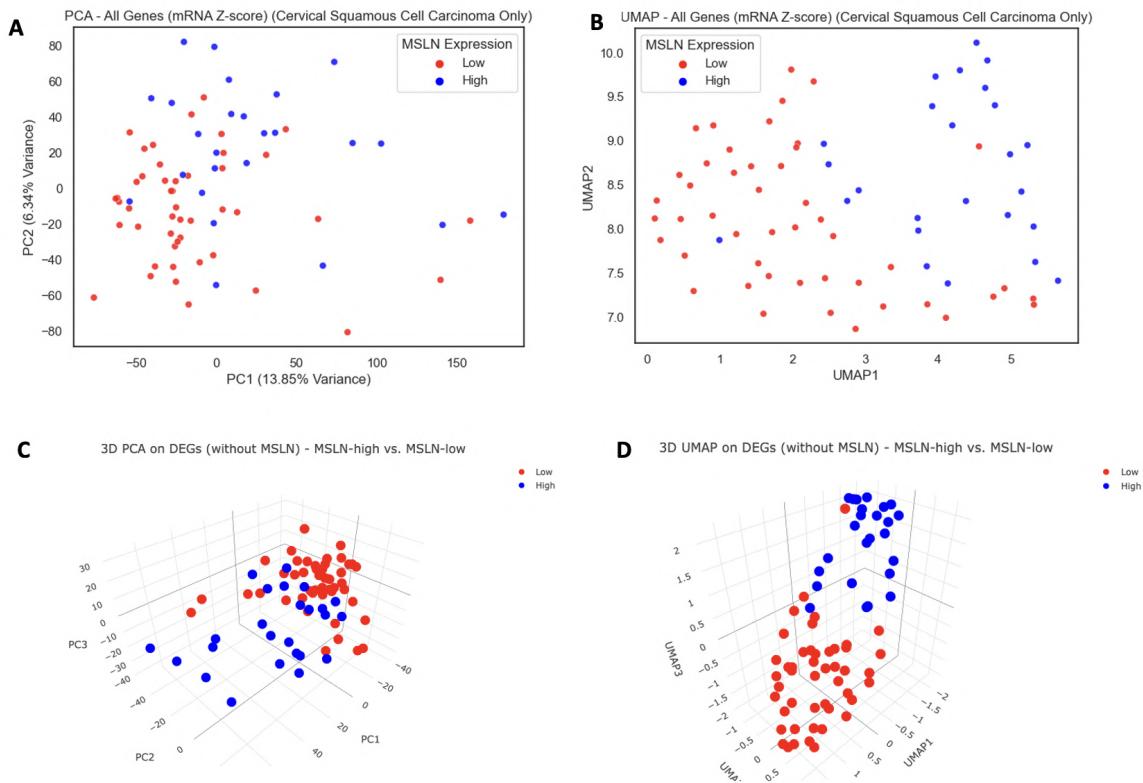
Differential gene expression analysis via DESeq2 clearly distinguished the two groups. The top differentially expressed genes (DEGs) logically included MSLN itself, alongside genes such as *ALDH3B1*, *KRT81*, *MUC13*, and *FOLR1* (Table 4) (Figure 7). PCA and UMAP visualizations robustly confirmed clear separation between MSLN-high and MSLN-low groups, especially evident when considering only DEGs (Figure 8).

**Table 4: Top 10 Differentially Expressed Genes (DEGs) between MSLN-High and MSLN-Low Groups**

Gene	Base Mean	$\log_2$ Fold Change	IfcSE	Wald statistic	p-value	Adjusted p-value (padj)
MSLN	7954.31	10.63	0.301	35.29	$8.34 \times 10^{-273}$	$1.54 \times 10^{-268}$
ALDH3B1	673.87	3.39	0.231	14.67	$1.02 \times 10^{-48}$	$9.42 \times 10^{-45}$
KRT81	6060.83	7.57	0.548	13.83	$1.78 \times 10^{-43}$	$1.10 \times 10^{-39}$
MUC13	1079.02	8.38	0.622	13.49	$1.90 \times 10^{-41}$	$8.77 \times 10^{-38}$
FOLR1	495.22	7.79	0.649	12.00	$3.48 \times 10^{-33}$	$1.28 \times 10^{-29}$
MUC6	377.35	8.26	0.702	11.77	$5.33 \times 10^{-32}$	$1.63 \times 10^{-28}$
MSLNL	13.29	6.47	0.567	11.41	$3.71 \times 10^{-30}$	$9.76 \times 10^{-27}$
INSL4	42.92	9.23	0.815	11.32	$1.01 \times 10^{-29}$	$2.33 \times 10^{-26}$
TNNC1	83.76	5.37	0.484	11.11	$1.15 \times 10^{-28}$	$2.36 \times 10^{-25}$
ACSF2	1043.78	2.47	0.231	10.70	$1.07 \times 10^{-26}$	$1.96 \times 10^{-23}$



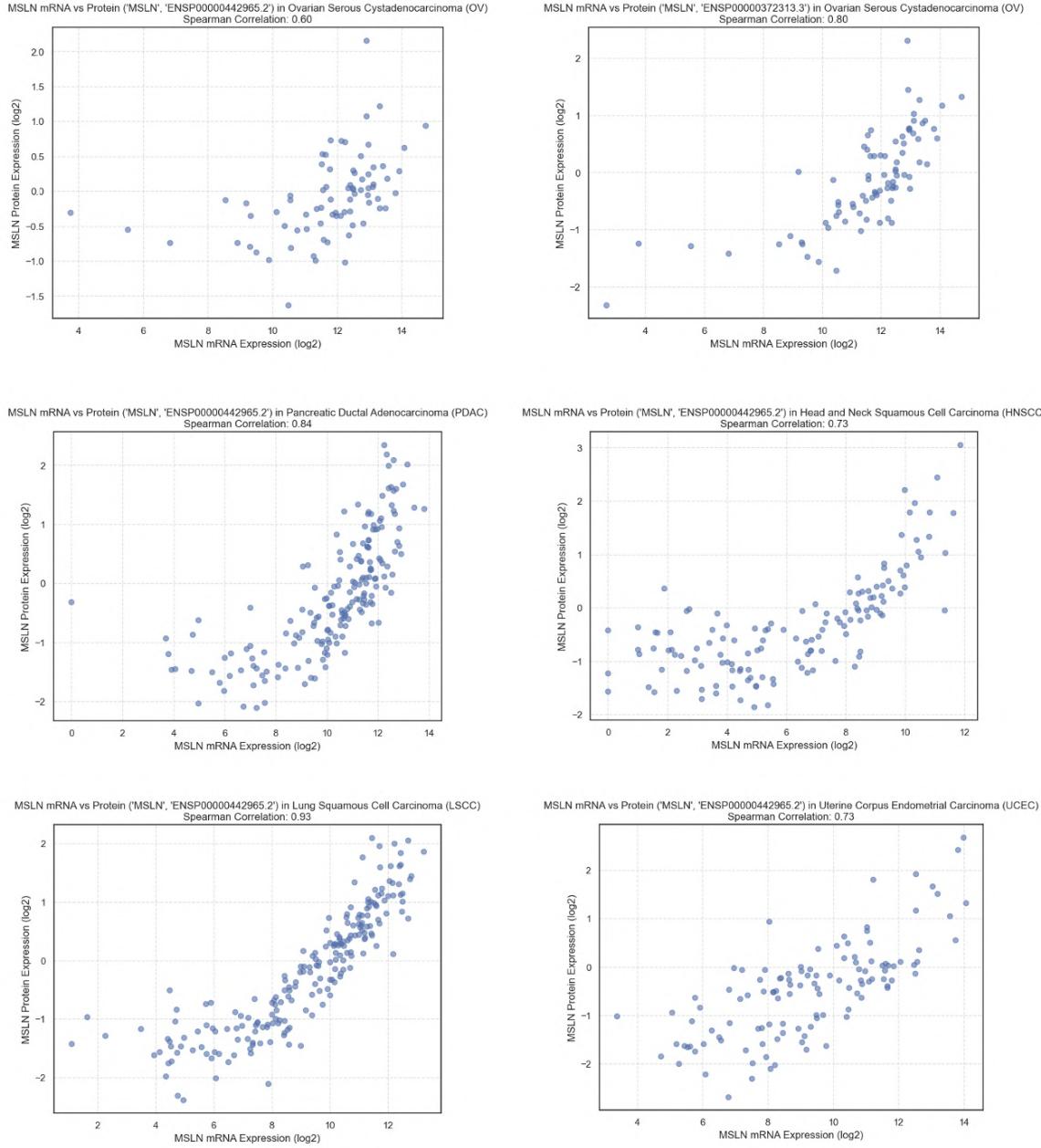
**Figure 7: Volcano plot of differentially expressed genes (DEGs).** Genes differentially expressed between MSLN-high and MSLN-low tumors, excluding MSLN itself. Significance threshold set at adjusted  $p$ -value  $< 0.05$  and absolute  $\log_2$  fold change  $> 1$ . Red dots are genes that are statistically significant.



**Figure 8: Dimensionality reduction of tumor transcriptomic profiles.** PCA and UMAP analyses show separation of MSLN-high and MSLN-low tumors using all genes (A–B) or only DEGs (C–D), highlighting transcriptomic differences between groups.

### 3.1.3 Proteomics analysis and identification of potential targets

Given the absence of direct MSLN data in the Reverse Phase Protein Array (RPPA), we employed a comparative analysis using CPTAC datasets from related cancer types to validate the assumption of consistent mRNA-protein expression relationships for MSLN. This approach supports the rationale that the tumor classification established using mRNA expression data can be reliably translated to RPPA data by matching tumor samples based on their IDs. (Figure 9).



**Figure 9: Correlation between MSLN mRNA and protein expression in multiple tumor types.** Scatter plots showing Spearman correlations across six cancer types from the CPTAC dataset, confirming linear association between transcriptomic and proteomic MSLN levels.

Differential protein expression analysis using the *limma* library revealed that *NOTCH1*, *FN1* (Fibronectin), and *TFRC* (Transferrin Receptor) were upregulated in MSLN-low tumors (Table 5). Among these, *FN1* showed modest expression in MSLN-low tumors (mean = 0.18) versus strong downregulation in MSLN-high (mean = -0.47), yielding a moderate difference of -0.65, though with limited statistical significance (adj. p = 0.18). *NOTCH1* displayed a larger shift (0.71 vs -0.45; difference = -1.16), but with high variability (adj. p = 0.37), and poses safety concerns due to its known expression in normal tissues. While both proteins are cell surface-associated and remain attractive, their clinical translatability differs. In contrast, the highest-ranking proteins, such as *MYH11*, *PRKAA1\_AMPK\_pT172*, and *RICTOR*, are intracellular and thus not exploitable for ADC strategies. However, their consistent overexpression highlights mechanistic features

of MSLN-low tumors suggesting an aggressive or adaptive phenotype that could be relevant for future pathway-targeted strategies.

**Table 5: Top RPPA Proteins Upregulated in MSLN-low Tumors (selected targets highlighted)**

Protein	Mean MSLN-Low	Mean MSLN-High	Difference	logFC	P-Value	Adj. P-Value
MYH11	0.7222	-0.2900	-1.0122	-1.0122	$7.79 \times 10^{-4}$	0.0361
PRKAA1_AMPK_pT172	2.2023	1.3149	-0.8873	-0.8873	$1.23 \times 10^{-3}$	0.0011
RICTOR	0.8895	0.0493	-0.8402	-0.8402	$7.37 \times 10^{-3}$	0.0731
STAT5-alpha	1.6130	0.9552	-0.6578	-0.6578	$7.87 \times 10^{-2}$	0.1012
FN1 (Fibronectin)	0.1822	-0.4679	-0.6507	-0.6507	0.0419	0.1827
...						
CTNNB1 ( $\beta$ -catenin)	2.1502	1.8218	-0.3284	-0.3284	0.1358	0.3447
...						
NOTCH1	0.7176	-0.4451	-1.1627	-0.2960	0.0260	0.3759
...						
EGFR	0.6389	0.4754	-0.1634	-0.1634	0.4285	0.6591

### 3.1.4 EGFR as a target in MSLN-low tumors

At the mRNA level, *EGFR* expression was negatively correlated with MSLN expression ( $r = -0.123$ ,  $p = 0.0343$ ), and its transcript levels were significantly higher in MSLN-low tumors (mean = 0.059) compared to MSLN-high (mean = -0.421,  $p = 0.0144$ ), reinforcing its relevance as a complementary target. At the protein level, *EGFR* was modestly elevated in MSLN-low tumors (0.64 vs 0.48), but without statistical significance (adj.  $p = 0.66$ ). Despite this, *EGFR* remains a highly actionable candidate due to its established role in oncology, validated antibody-drug conjugate platforms, and compatibility with SNAP-tag bispecific designs—offering a pragmatic balance between biological rationale and translational feasibility (*Huysamen et al., 2023*) [34].

### 3.1.5 MSLN-low subtype characterisation

Pathway enrichment analyses revealed significant activation of Wnt/ $\beta$ -catenin signalling in MSLN-low tumours (Table 6), indicating a potential vulnerability to next-generation Wnt antagonists, including the condensate modulator DPTX3186, which sequesters  $\beta$ -catenin into inactive nuclear condensates, and upstream porcupine inhibitors such as LGK-974 or ETC-159 (*Li et al., 2021*) [62]; (*Zhang et al., 2020*) [63]. Several enriched terms were associated with canonical Wnt signaling, including upstream ligands (e.g., WNT2B, WNT5A), receptors (FZD1, FZD7), and core regulators (GSK3B, LEF1, CTNNB1). This implies a possible increase in transcriptional activity of  $\beta$ -catenin-driven gene expression programs involved in proliferation and epithelial-mesenchymal transition (EMT). RPPA data further supported this trend:  $\beta$ -catenin (CTNNB1) protein levels were higher in MSLN-Low tumors (mean = 2.15) compared to MSLN-High tumors (mean = 1.82), with a difference of -0.328. Although not statistically significant ( $p = 0.1358$ , adj. $p = 0.3447$ ) (Table 5), this observation aligns with the enrichment results and warrants future functional validation to assess pathway activity and drug responsiveness.

**Table 6: Selected significantly enriched pathways in MSLN-Low tumors.** Identified via pathway enrichment analysis using *EnrichR* [64]. The input was the list of 2154 upregulated genes in MSLN-Low tumors.

Database	Pathway	Rank	Adj. P-value	Combined Score	Genes
KEGG 2021	Wnt signaling pathway	2	0.00158	27.04	a
WikiPathways	Regulation of Wnt/ $\beta$ -catenin Signaling	1	0.00113	156.23	b
WikiPathways	Wnt Signaling Pathway	4	0.00337	29.99	c
WikiPathways	lncRNA in Canonical Wnt / CRC	7	0.00365	29.78	d

<sup>a</sup> GSK3B, CAMK2D, WNT2B, CHD8, CAMK2A, LEF1, PRICKLE1, ZNRF3, WIF1, DVL3, RSPO1, GPC4, PRKACA, MAP3K7, TLE4, FZD1, APC2, SMAD3, CSNK2A1, FBXW11, CSNK1A1, FZD7, CTNNBIP1, WNT5A, FZD6, SIAH1, NFATC3, SENP2, WNT16, DKK4, SFRP4, SFRP2, PLCB4, APC, ROR1, LGR5.

<sup>b</sup> FZD1, GSK3B, SFRP4, LRP1, APC, CSNK1A1, FZD7, LEF1, TCF4, DKK3.

<sup>c</sup> GSK3B, CAMK2D, WNT2B, CHD8, CAMK2A, LEF1, PRICKLE1, WIF1, DVL3, GPC4, MAP3K7, FZD1, CSNK2A1, CSNK1A1, FZD7, CTNNBIP1, WNT5A, KREMEN1, SENP2, WNT16, DKK4, SFRP4, SFRP2, PLCB4, APC, ROR1.

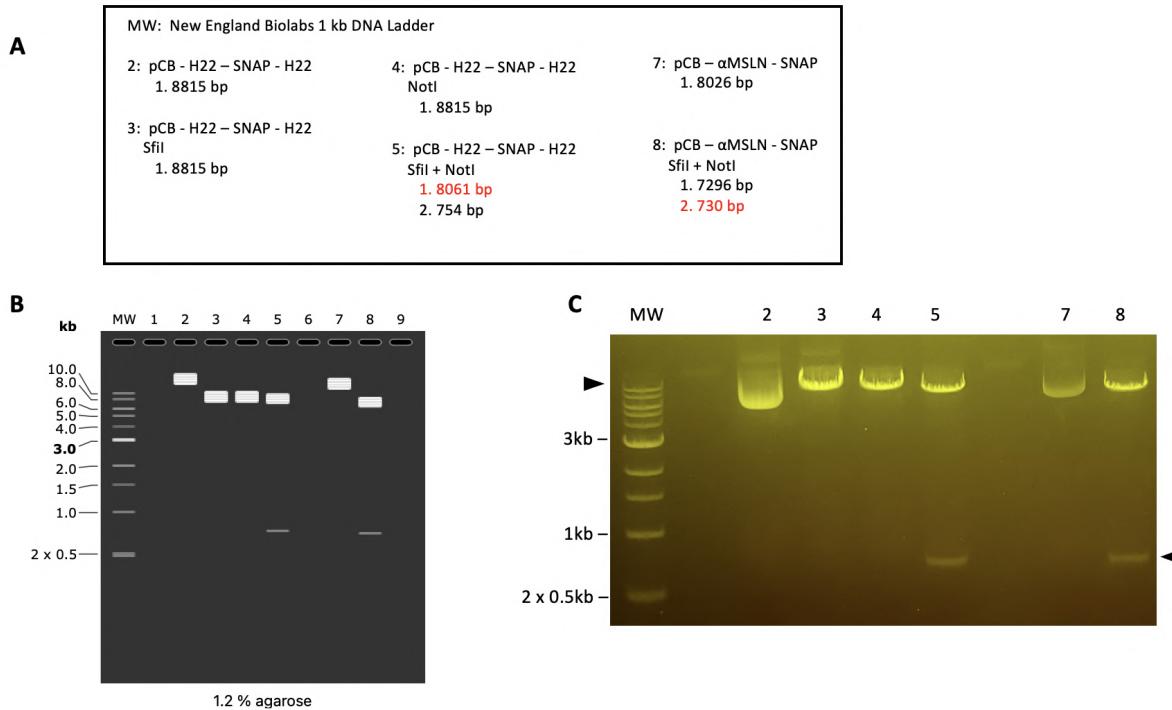
<sup>d</sup> APC2, FZD1, GSK3B, WNT2B, CSNK2A1, CSNK1A1, CHD8, FZD7, CTNNBIP1, WNT5A, LEF1, FZD6, KREMEN1, SENP2, WNT16, DKK4, SFRP4, SFRP2, APC, WIF1, DVL3, ROR1, MAP3K7, EZH2.

## 3.2 Bispecific SNAP-tag Antibody Cloning and Production

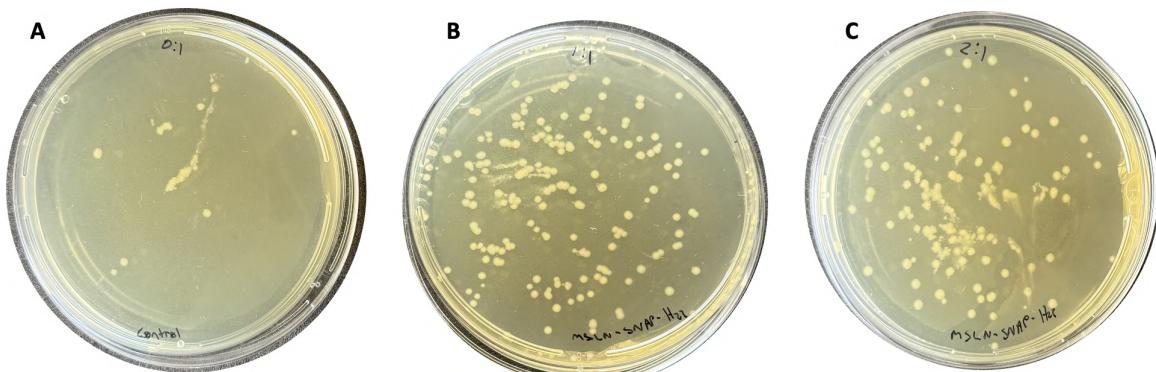
As an initial training step, I shadowed Thabo Matshoba, a PhD student in the lab, who was constructing a bivalent anti-mesothelin antibody ( $\alpha$ MSLN (scFv)–SNAP– $\alpha$ MSLN (scFv)). This construct was generated by first replacing the anti-CD64 (H22) scFv in an existing H22–SNAP–H22 plasmid with an  $\alpha$ -MSLN(scFv), using *SfiI* and *NotI* restriction sites flanking the upstream insertion site. This cloning strategy served both as a hands-on introduction to the lab’s workflow and as a stepping stone toward my own bispecific construct:  $\alpha$ EGFR (scFv) –SNAP– $\alpha$ MSLN (scFv). In this design, the  $\alpha$ -EGFR (scFv) is inserted upstream of SNAP, while the  $\alpha$ -MSLN (scFv) is inserted downstream. The upstream insertion was facilitated by the presence of conserved *SfiI*–*NotI* sites in lab-standard plasmids, while the downstream insertion required PCR amplification of the  $\alpha$ MSLN (scFv) fragment with flanking *BspI* and *EcoRI* restriction sites. These primers had already been designed by Thabo, making the EGFR–SNAP–MSLN configuration technically more accessible than a reversed MSLN–SNAP–EGFR setup.

### 3.2.1 Cloning of pCB- $\alpha$ MSLN-SNAP-H22 (training phase with Thabo Matshoba)

Following double digestion of pCB-H22–SNAP–H22 and pCB- $\alpha$ MSLN(scFv)–SNAP (both plasmid in use and priorly validated in the lab) with *SfiI* and *NotI*, the insert and vector fragments were gel-purified and ligated (Figure 10, Section 2.4.6). The ligation products were transformed into chemically competent DH5 $\alpha$  *E. coli* and plated on LB agar containing 200  $\mu$ g/ml ampicillin. As expected, colonies grew on all insert:vector plates, while the ‘vector only’ control showed low-level growth, indicating background re-ligation (Figure 11A). A total of 11, 197, and 98 colonies were obtained on the 0:1 (vector-only), 1:1, and 2:1 insert:vector plates, respectively. Given the amount of DNA used, this corresponds to transformation efficiencies of  $3.67 \times 10^2$ ,  $6.57 \times 10^3$ , and  $3.27 \times 10^3$  cfu/ $\mu$ g DNA.



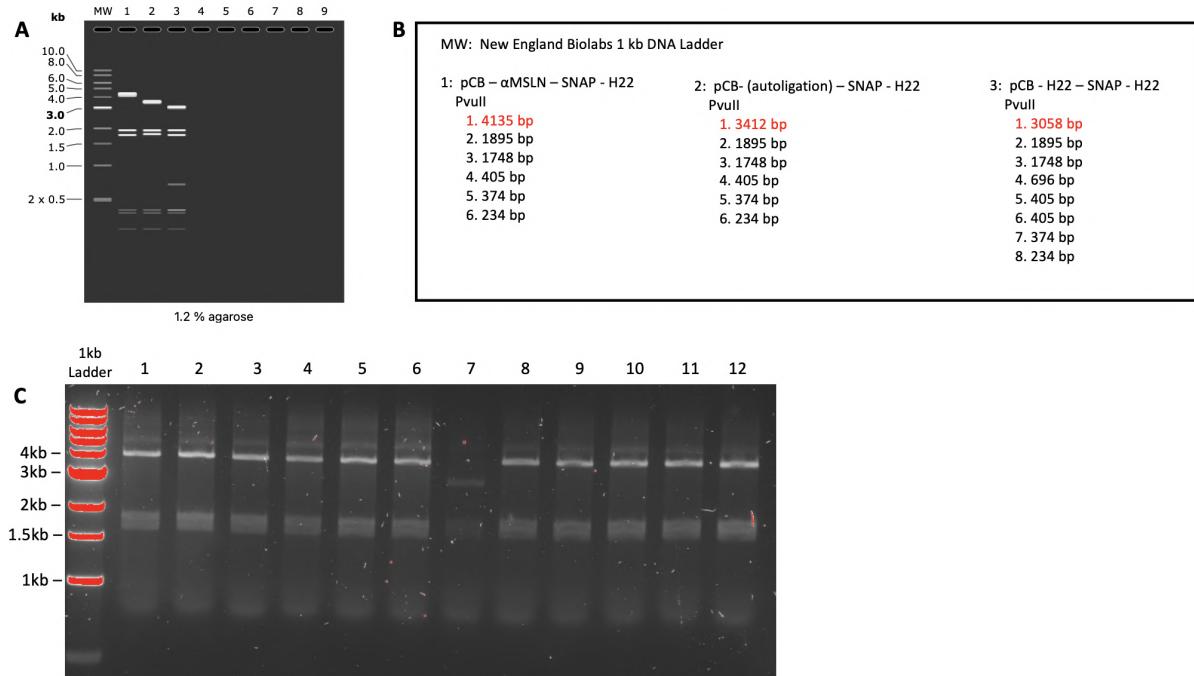
**Figure 10: Restriction digestion of pCB-H22-SNAP-H22 and pCB- $\alpha$ MSLN (scFv) -SNAP using agarose gel electrophoresis.** (A) Legend of the samples loaded in each well and expected digestion patterns, including single and double digests with *SfiI* and *NotI*. The extracted fragments used for ligation are highlighted in red. (B) Simulated digestion results using SnapGene showing expected banding patterns on a 1.2% agarose gel. (C) Agarose gel electrophoresis of digested constructs, photographed under blue-light transillumination with an amber (orange) filter. Bands corresponding to expected digestion fragments are indicated by black arrowheads in lanes 5 (top) and 8 (bottom).



**Figure 11: Growth of E. coli cells transformed with potential recombinant pCB- $\alpha$ MSLN(scFv)-SNAP-H22 plasmid DNA.** Following transformation, 100  $\mu$ l of bacterial mixture was plated onto LB agar plates, which were then incubated overnight at 37°C (section 2.4.6). The concentration of ampicillin used was 200  $\mu$ /ml. (A) 0:1 vector only plate, (B) 1:1 ratio of insert to vector, (C) 2:1 ratio of insert to vector.

To identify colonies containing correctly assembled constructs, restriction screening was performed using *PvuII*. As shown in Figure 12, all the colonies, except colony 7, showed the expected banding pattern corresponding to the successful insertion of MSLN and SNAP. Colony 7 displayed a digestion profile consistent with vector autoligation, as confirmed by SnapGene simulation. Based on these results, colonies 1 and 2 were selected

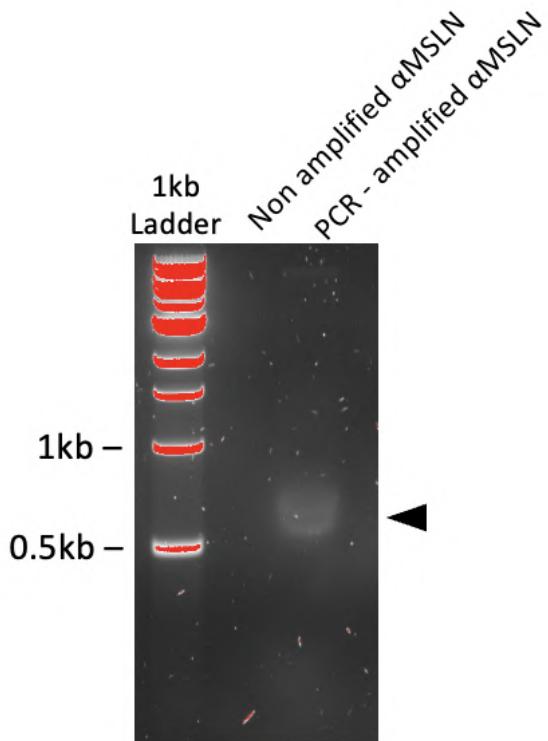
for further cloning steps.



**Figure 12: Colony screening by *PvuII* digestion to confirm insertion of MSLN and SNAP in pCB- $\alpha$ MSLN (scFv) -SNAP-H22.** (A) Simulated digestion patterns generated using SnapGene for pCB- $\alpha$ MSLN (scFv) -SNAP-H22, autoligated vector, and pCB-H22-SNAP-H22. Diagnostic fragments used to differentiate constructs are highlighted in red. (B) Summary table of expected fragment sizes following *PvuII* digestion. Diagnostic fragments are highlighted in red. (C) Agarose gel electrophoresis of miniprep DNA digested with *PvuII*. All colonies except colony 7 exhibit the expected pattern for correct insertion of both MSLN and SNAP.

### 3.2.2 PCR Amplification and Restriction Digest of $\alpha$ MSLN (scFv) Insert

To enable directional cloning of the  $\alpha$ MSLN(scFv) downstream of the SNAP-tag, the insert was PCR-amplified using primers encoding *EcoRI* and *BspI* restriction sites (Section 2.4.8). As a control, a reaction containing the same PCR mix but without primers was run in parallel. Since the initial template DNA was present at low concentration, it was not expected to be visible on the gel unless amplified. The clear presence of a band in the sample with primers, and its absence in the control, confirms successful amplification of the  $\sim$ 736 bp product.

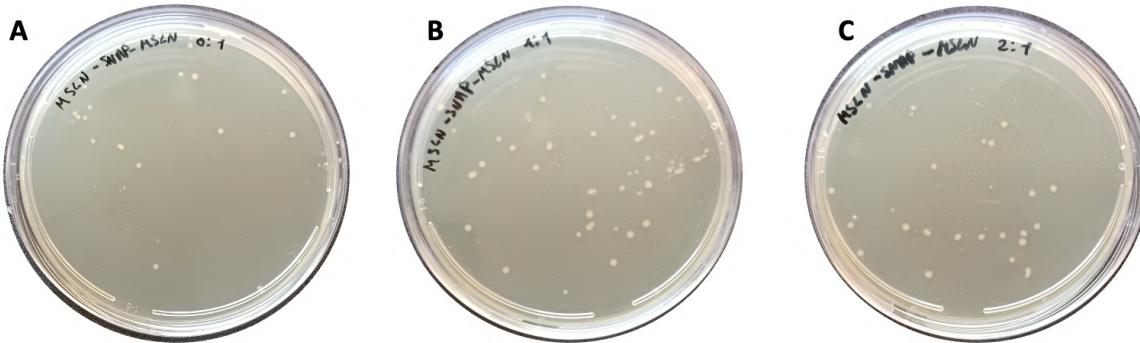


**Figure 13: PCR amplification of  $\alpha$ MSLN (scFv) insert.** 1.2% agarose gel image. Lane 1 corresponds to unamplified  $\alpha$ MSLN (scFv), loaded as a negative control and not visible as expected. Lane 2 shows the PCR-amplified, with a visible band at 736 bp (arrowhead), confirming successful amplification.

### 3.2.3 Cloning and screening of pCB- $\alpha$ MSLN (scFv) -SNAP- $\alpha$ MSLN (scFv) (training phase with Thabo Matshoba)

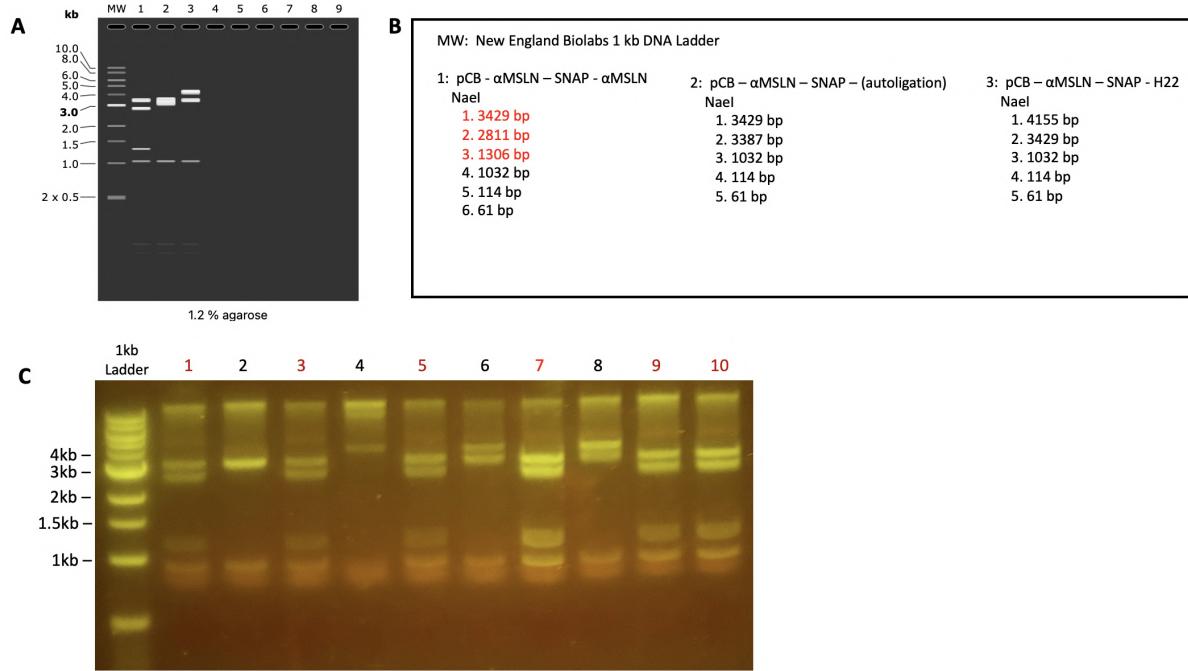
The digested PCR product encoding the  $\alpha$ -MSLN (scFv) was ligated into the pCB- $\alpha$ MSLN(scFv)-SNAP-H22 vector backbone downstream of the SNAP-tag, replacing the original H22 sequence. The ligation mixtures were transformed into chemically competent DH5 $\alpha$  *E. coli* and plated on LB agar containing 200  $\mu$ g/ml ampicillin. Colony growth was observed on all insert:vector plates, as well as the vector-only control, indicating background re-ligation (Figure 14).

A total of 18, 46, and 33 colonies were observed for the 0:1 (vector-only), 1:1, and 2:1 insert:vector conditions, respectively. Given the amount of DNA used, this corresponds to transformation efficiencies of  $6.0 \times 10^2$ ,  $1.53 \times 10^3$ , and  $1.1 \times 10^3$  cfu/ $\mu$ g DNA, respectively.



**Figure 14: Growth of *E. coli* cells transformed with potential recombinant pCB- $\alpha$ MSLN(scFv)-SNAP- $\alpha$ MSLN(scFv) plasmid DNA.** Following transformation, 100  $\mu$ l of bacterial mixture was plated onto LB agar plates, which were then incubated overnight at 37°C (section 2.4.6). The concentration of ampicillin used was 200  $\mu$ /ml. (A) 0:1 vector only plate, (B) 1:1 ratio of insert to vector, (C) 2:1 ratio of insert to vector.

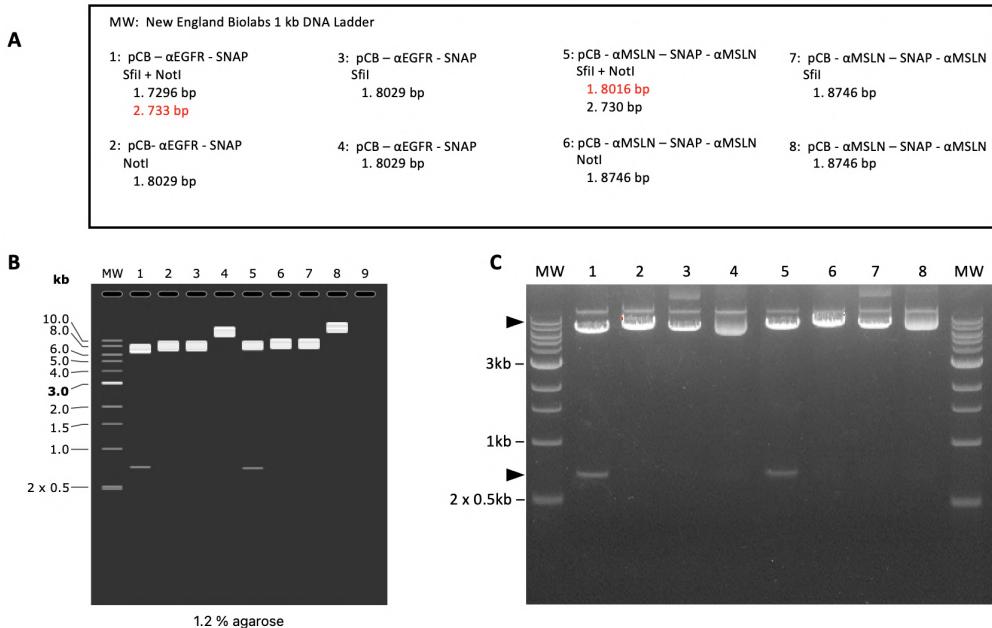
To confirm successful cloning, restriction digest screening with *NaeI* was performed on selected colonies. As shown in Figure 15, colonies 1, 3, 5, 7, 9, and 10 displayed the expected digestion pattern corresponding to insertion of both MSLN fragments. Colony 4 showed a digestion profile consistent with autoligation. Colonies 6 and 8 corresponded to the original pCB- $\alpha$ MSLN(scFv)-SNAP-H22 construct and were retained for use in parallel experiments. Colony 7 was selected for downstream cloning due to clear diagnostic band separation.



**Figure 15: Screening of pCB- $\alpha$ MSLN(scFv)-SNAP- $\alpha$ MSLN(scFv) by *NaeI* digestion.** (A) SnapGene simulation of expected digestion patterns for the bispecific construct, autoligated vector, and control vector lacking the second MSLN insertion. (B) Expected fragment sizes after *NaeI* digestion. Diagnostic fragments used to differentiate constructs are highlighted in red. (C) Agarose gel electrophoresis of digested miniprep DNA from colonies 1–10. Photographed under blue-light transillumination with an amber (orange) filter. Lanes 1, 3, 5, 7, 9, and 10 show the correct banding pattern corresponding to successful insertion of both MSLN arms. Colony 4 shows a banding profile consistent with autoligation. Lanes 6 and 8 correspond to pCB- $\alpha$ MSLN(scFv)-SNAP-H22, and were kept in culture for use in parallel experiments. Colony 7 was selected for downstream cloning due to optimal fragment separation.

### 3.2.4 Cloning of pCB- $\alpha$ EGFR(scFv)-SNAP- $\alpha$ MSLN(scFv)

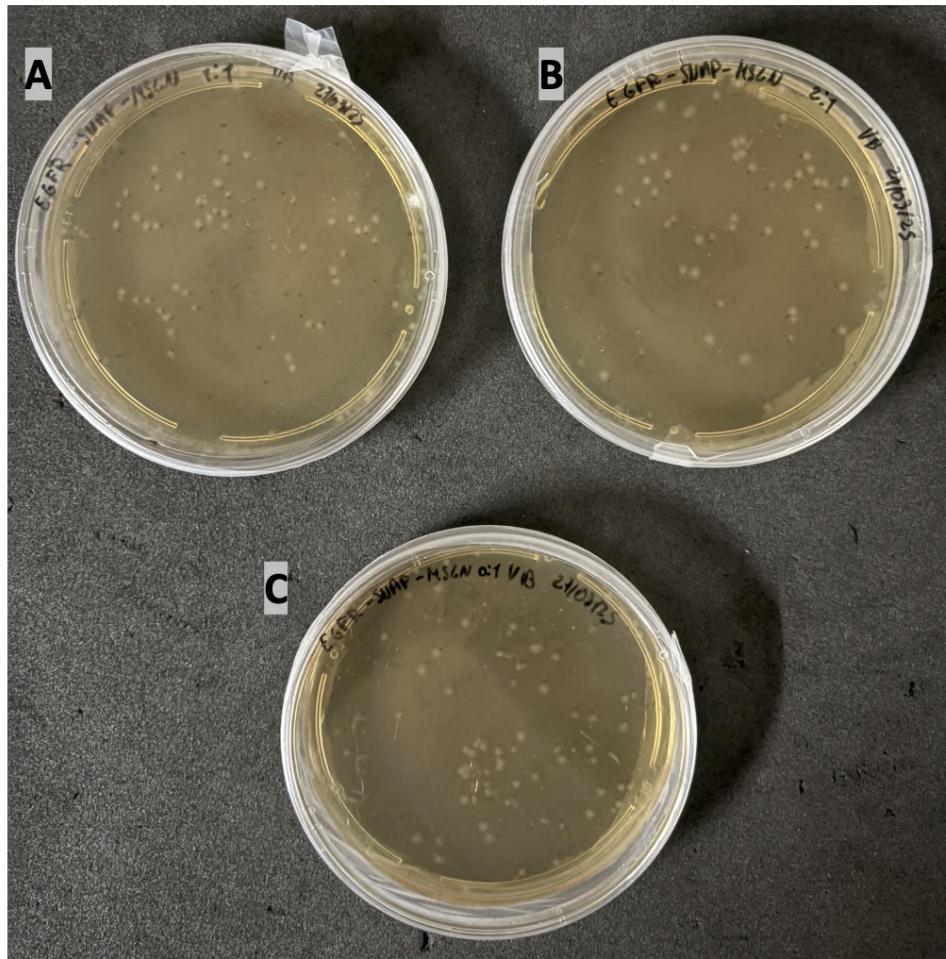
To construct the bispecific EGFR-SNAP-MSLN fusion protein, the upstream  $\alpha$ -MSLN(scFv) insert was excised from the validated pCB- $\alpha$ MSLN(scFv)-SNAP- $\alpha$ MSLN(scFv) vector using *SfiI* and *NotI*. The pCB- $\alpha$ EGFR(scFv)-SNAP plasmid (in use and priorly validated in the lab) was digested with the same enzymes, and both fragments were purified by agarose gel extraction. SnapGene simulation was used to confirm the expected fragment sizes and validate the restriction strategy (Figure 16).



**Figure 16: Restriction digestion of pCB- $\alpha$ EGFR(scFv)-SNAP and pCB- $\alpha$ MSLN(scFv)-SNAP- $\alpha$ MSLN(scFv) using agarose gel electrophoresis.** (A) Legend of the samples loaded in each well. The extracted fragments used for ligation are highlighted in red. (B) Simulated digestion using SnapGene. (C) Agarose gel electrophoresis results. DNA fragments were separated on a 1.2% agarose gel for 1 hour at 100 volts. The extracted fragments used for ligation are indicated by black arrows in lanes 1 (bottom band) and 5 (top band).

The ligation reaction was performed using T4 DNA ligase, and recombinant DNA mixtures were transformed into chemically competent DH5 $\alpha$  *E. coli*. Transformed cells were plated on LB agar supplemented with 200  $\mu$ g/ml ampicillin. Colony growth was observed on all conditions, including the vector-only control plate, indicating some background re-ligation (Figure 17A).

A total of 80, 90, and 57 colonies were obtained on the 0:1 (vector-only), 1:1, and 2:1 insert:vector plates, respectively. These values correspond to transformation efficiencies of  $2.67 \times 10^3$ ,  $3.0 \times 10^3$ , and  $1.9 \times 10^3$  cfu/ $\mu$ g DNA.

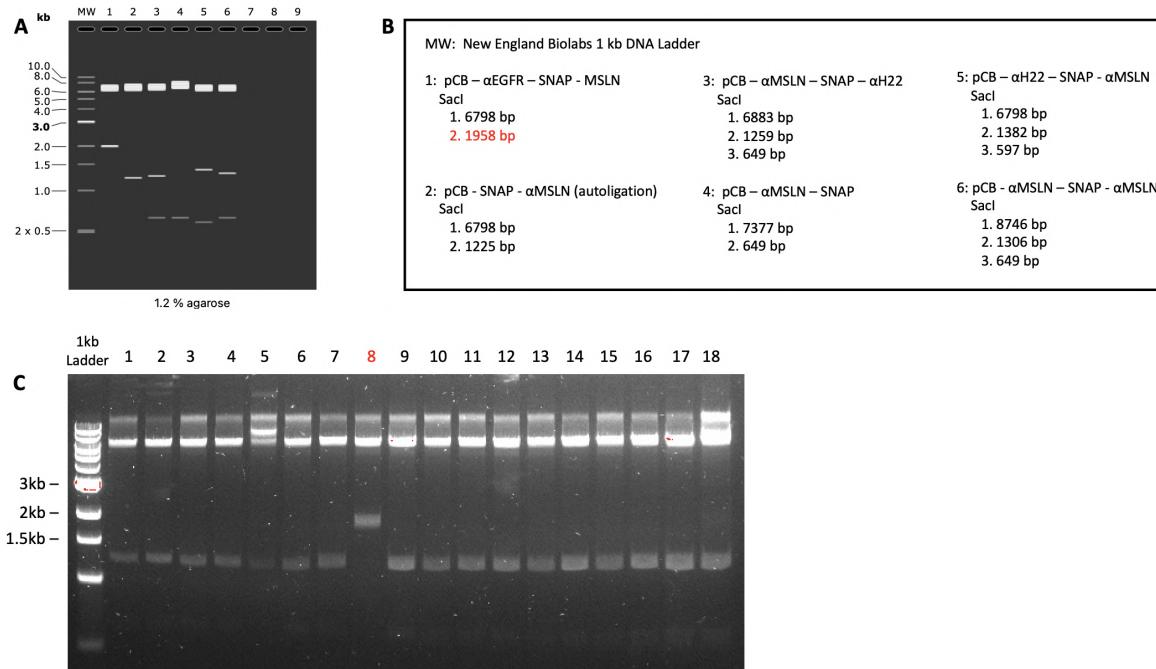


**Figure 17: Growth of *E. coli* cells transformed with potential recombinant pCB- $\alpha$ EGFR(scFv)-SNAP- $\alpha$ MSLN(scFv) plasmid DNA.** (A) 1:1 insert:vector ratio; (B) 2:1 insert:vector ratio; (C) vector-only control. Colonies were grown overnight at 37°C on LB agar containing 200  $\mu$ g/ml ampicillin.

### 3.2.5 Colony screening and validation of pCB- $\alpha$ EGFR(scFv)-SNAP- $\alpha$ MSLN(scFv)

To identify colonies harboring correctly assembled constructs, restriction screening was performed on miniprep plasmid DNA using *SacI*. SnapGene was used to simulate the expected digestion patterns for the correctly ligated bispecific vector, autoligated backbone, and other potential misligation products.

As shown in Figure 18, colony 8 displayed a digestion profile matching the expected 6798 bp and 1958 bp fragments, confirming the presence of the full-length pCB- $\alpha$ EGFR(scFv)-SNAP- $\alpha$ MSLN(scFv) construct. All other colonies showed alternative banding profiles, consistent with incomplete or incorrect ligation events. Colony 8 was therefore selected and cultured for plasmid DNA extraction and expansion for downstream applications such as mammalian cell transfection.



**Figure 18: Colony screening by *SacI* digestion to identify correct ligation of pCB- $\alpha$ EGFR(scFv)-SNAP- $\alpha$ MSLN(scFv).** (A) SnapGene simulation of expected banding patterns for the correct construct (lane 1) and alternative outcomes (lanes 2–6). (B) Summary of expected fragment sizes. Diagnostic fragment used to differentiate constructs is highlighted in red. (C) Agarose gel of *SacI*-digested plasmid DNA from colonies 1–18. Only colony 8 shows the expected 6798 bp and 1958 bp fragments, confirming successful ligation.

### 3.2.6 Sequencing confirmation of pCB- $\alpha$ EGFR(scFv) -SNAP- $\alpha$ MSLN (scFv) integrity

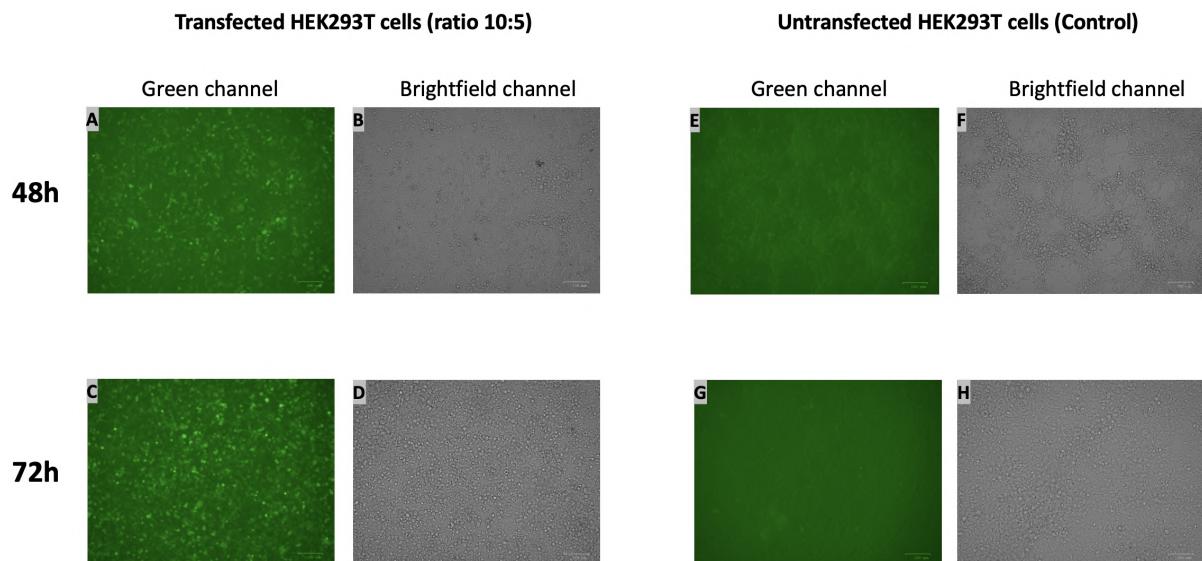
To validate the integrity of the recombinant construct, Sanger sequencing was performed using four primers targeting distinct regions of the expression cassette (see section 2.4.7). The resulting sequences were aligned to the original *in silico* pCB- $\alpha$ EGFR (scFv) -SNAP- $\alpha$ MSLN (scFc) design obtained from GenScript.

Due to the limitations of Sanger sequencing on large plasmids, several ambiguous base calls (e.g., ‘K’, ‘M’, ‘W’) were observed in the raw alignments, corresponding to mixed peaks. These IUPAC codes were manually inspected and replaced with the most likely nucleotide when supported by overlapping regions sequenced by other primers. For instance, if a ‘K’ (G/T) ambiguity occurred and the corresponding overlapping region showed a G, the manual correction was accepted. When ambiguities could not be resolved or the substitution conflicted with the reference, the mismatch was left unaltered.

A summary of alignment quality before and after manual review is presented in the appendix, Table 1A. This review helped clarify base identities and confirmed that the final construct contained no frameshifts or deleterious mutations in the coding region. Final SnapGene alignments are shown in Figure 1A.

### 3.2.7 Transfection into HEK293T cells for protein expression

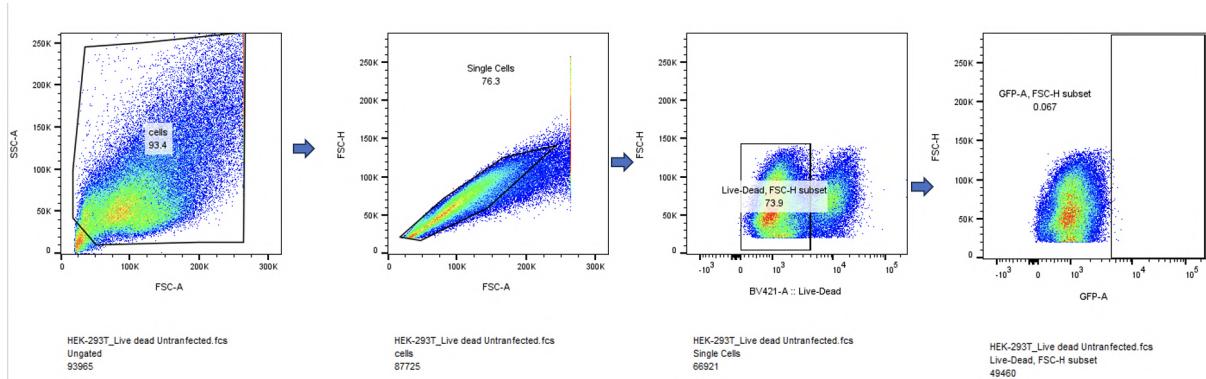
To determine the efficiency of transient transfection in HEK293T cells with the SNAP-tag bispecific antibody construct (pCB- $\alpha$ EGFR(scFv)-SNAP -  $\alpha$ MSLN(scFv)). The enhanced green fluorescent protein (eGFP) of positively transfected cells was monitored every 24 hours for 3 days using a ZOE™ Fluorescence Cell Imager (BioRad, USA) as seen in (Figure 19). Cells were transfected using different DNA:XtremeGENE reagent ratios: 3  $\mu$ g:9  $\mu$ l (ESM93), 5  $\mu$ g:10  $\mu$ l (ESM105), and 5  $\mu$ g:15  $\mu$ l (ESM155). Figure 19 illustrates results from HEK-293T cells transfected with 5  $\mu$ g of plasmid DNA using 10  $\mu$ l of XtremeGENE transfection reagent (DNA: reagent ratio 1:2), as per protocol. Untransfected cells served as a negative control.



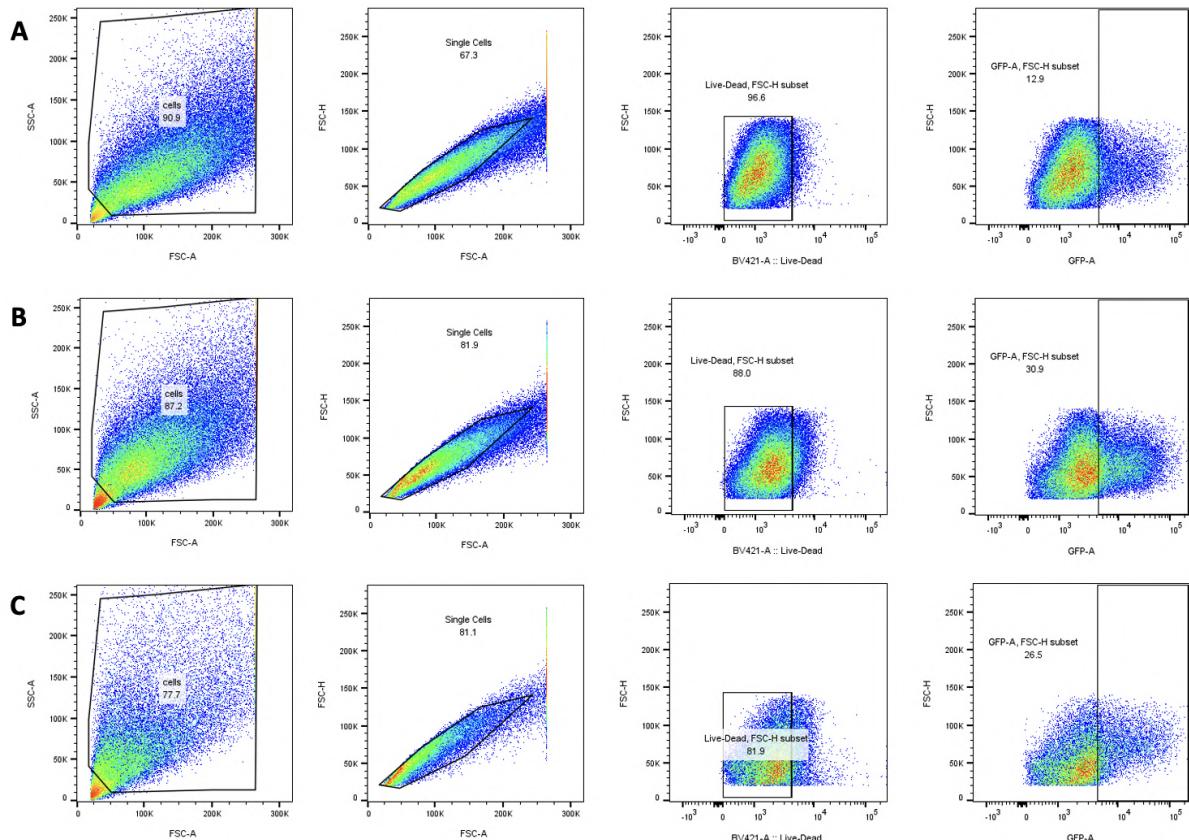
**Figure 19: Visualization of transient GFP expression at 48 and 72 hours post-transfection.** HEK293T cells were transfected with the  $\alpha$ -EGFR(scFv)-SNAP- $\alpha$ -MSLN(scFv) construct at a DNA:XtremeGENE ratio of 5  $\mu$ g:10  $\mu$ l and imaged at 48 h and 72 h post-transfection using the ZOE™ fluorescent cell imager. (A–D) Transfected cells: (A) green channel at 48 h, (B) brightfield at 48 h, (C) green channel at 72 h, (D) brightfield at 72 h. (E–H) Untransfected control cells: (E) green channel at 48 h, (F) brightfield at 48 h, (G) green channel at 72 h, (H) brightfield at 72 h. Fluorescence from GFP is detectable at both timepoints only in transfected samples. Scale bar = 100  $\mu$ m.

GFP expression was analyzed by flow cytometry (section 2.4.9) three days post-transfection. A standardized gating strategy was applied using FlowJo™ software (v10.6.1) consisting of: (1) FSC-A vs SSC-A to gate cells, (2) FSC-H vs FSC-A to exclude doublets, (3) BV421-A viability dye to gate live cells, and (4) GFP-A intensity to quantify eGFP-positive populations (Figure 20).

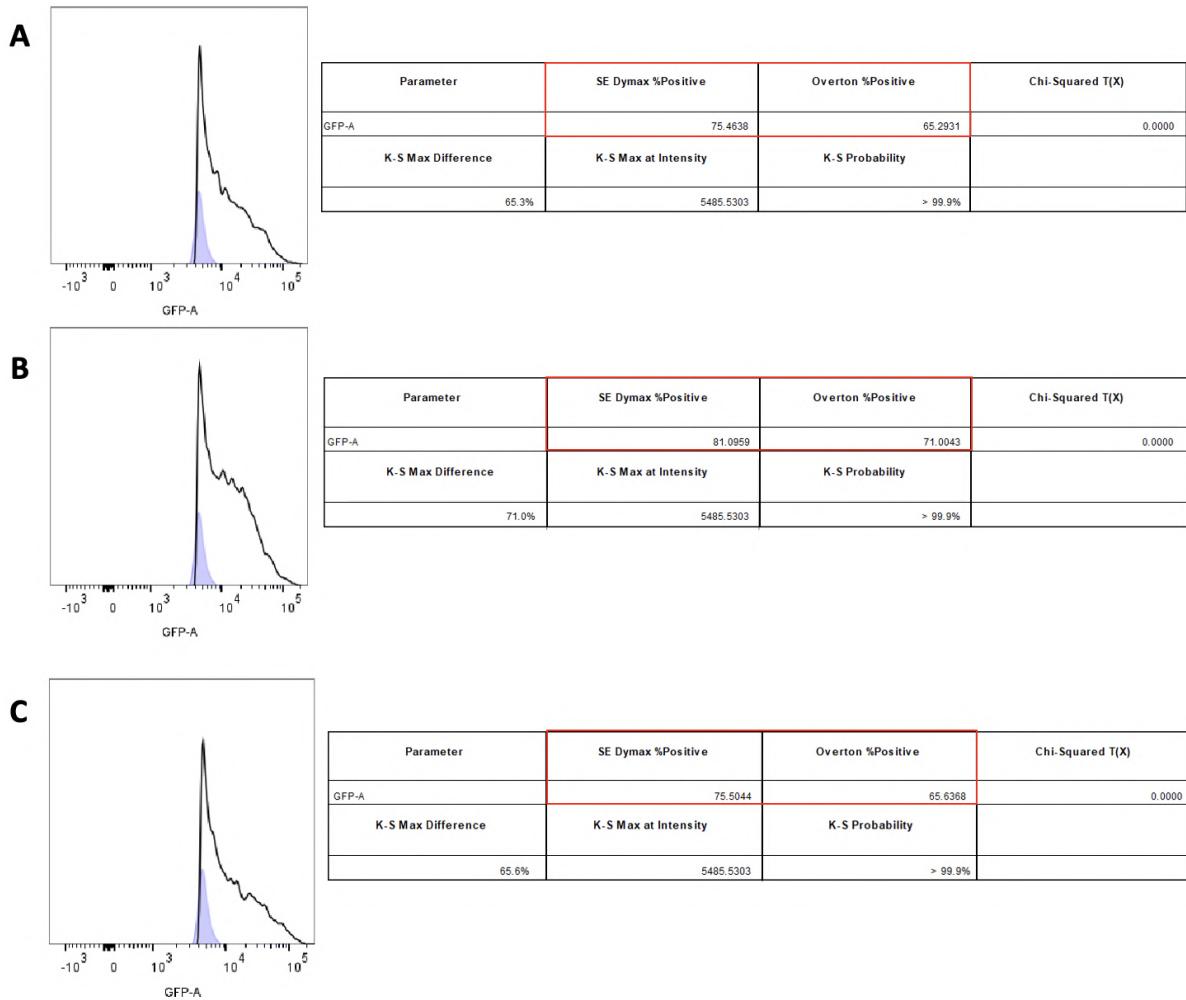
Representative dot plots for each transfected sample are shown in Figure 21, and histogram overlays of GFP intensity comparing transfected samples to the untransfected control are provided in Figure 22. Quantitative analysis is summarized in Table 7.



**Figure 20: Gating strategy for quantifying transfection efficiency.** (A) FSC-A vs SSC-A to select cells. (B) FSC-H vs FSC-A to exclude doublets. (C) Live/Dead staining using BV421-A. (D) GFP-A expression from live single cells. Data shown from untransfected HEK293T cells.



**Figure 21: GFP-based quantification of transfection efficiency.** Representative flow cytometry plots of live single-cell populations gated for GFP-A signal for constructs: (A) ESM93 (12.9%), (B) ESM105 (30.9%), and (C) ESM155 (26.5%).

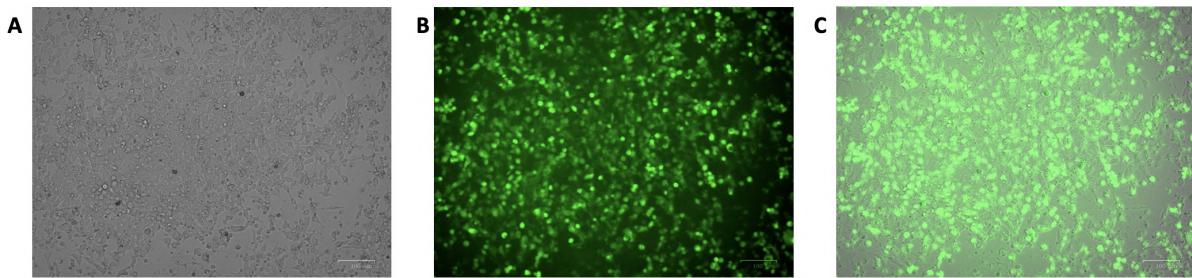


**Figure 22: Histogram analysis of GFP expression.** Overlay of GFP-A intensity between untransfected cells and transfected samples: (A) ESM93, (B) ESM105, and (C) ESM155. Histogram-based Overton subtraction estimates transfection efficiencies of ~65%, ~71%, and ~66%, respectively.

**Table 7: Transfection efficiencies of HEK293T cells with bispecific SNAP-tag antibody constructs, measured by GFP expression**

Sample	DNA:XtremeGENE ( $\mu\text{g}:\mu\text{l}$ )	GFP+ (%)	Histogram Shift (%)
ESM93	3:9	12.9%	~65%
ESM105	5:10	30.9%	~71%
ESM155	5:15	26.5%	~66%

Transfection efficiencies are summarized in Table 7. Based on the Table 7 values, Zeocin selection (100 mg/ml) was applied to enrich the eGFP-positive cell population carrying the bleomycin resistance gene. These cells were then maintained in culture (ideally at 90% confluence) until only transfected cells remained ((Figure 23) and cell culture supernatant (containing the fusion protein of interest) was ready for collection.



**Figure 23: Stable expression of GFP following one month of zeocin selection in HEK293T cells.** (A) Brightfield channel image shows general cell morphology. (B) Fluorescence image captured using the green channel (excitation: 480/17 nm, emission: 517/23 nm) reveals GFP expression in a majority of the cell population. (C) Merged image confirms colocalization of fluorescence with viable adherent cells. Cells were transfected with the  $\alpha$ -EGFR(scFv)-SNAP- $\alpha$ -MSLN(scFv) construct and maintained under zeocin selection for 4 weeks using the ZOE™ fluorescent cell imager (Bio-Rad). Selection pressure was gradually increased using Zeocin™ (Thermo Fisher Scientific; 100 mg/mL stock) added to 5 mL culture medium in a T25 flask, beginning with 5  $\mu$ L (100  $\mu$ g/mL), then 10  $\mu$ L (200  $\mu$ g/mL), and finally 15  $\mu$ L (300  $\mu$ g/mL) added weekly. This approach successfully enriched the transfected population over time. Scale bar = 100  $\mu$ m.

### 3.3 Design of Shedding-resistant Anti-mesothelin Antibody

Mesothelin (MSLN) is a glycosylphosphatidylinositol (GPI)-linked cell-surface protein overexpressed in various cancers, including mesothelioma, pancreatic, ovarian, and cervical cancers, making it a promising target for antibody-based therapies (*Chakraborty et al., 2024*) [42]. However, MSLN undergoes significant shedding from cancer cells due to proteolytic cleavage in its membrane-proximal C-terminal region. This shedding phenomenon can reduce the efficacy of antibodies targeting mesothelin, as antibodies binding to shed mesothelin are neutralized before reaching tumor cells (*Liu et al., 2024*) [41].

To address this issue, Liu et al. (2024) [41] developed the 15B6 monoclonal antibody (mAb) targeting a unique juxtamembrane epitope near the protease-sensitive region of mesothelin. This antibody, whose amino acid sequence is provided (Figure 24), demonstrated resistance to neutralization by shed mesothelin and significantly improved antitumor efficacy in preclinical models. Subsequently, Chakraborty et al. (2024) [42] leveraged this antibody to design a bispecific antibody (BsAb) format, showing enhanced T-cell-mediated cytotoxicity against MSLN-expressing tumors without interference from shed mesothelin. They also provided the amino acid sequences of their different constructs (Figure 25)



**Figure 24: Variable region sequences of m15B6.** Amino acid alignment of the m15B6 light and heavy chain variable regions (VL and VH), showing complementarity-determining regions (CDRs) used for antibody design. Adapted from *Liu et al., 2024* [41].

Name	Sequence
BsAb 1	<b>SP</b> - m15B6 VL - linker - m15B6 VH - linker - CD3 VH - linker - CD3 VL - His tag  <b>MGWGSCILFLVATATGVHQS</b> QAVTQE <b>SALSTP</b> SG <b>PGETV</b> T <b>LTCRSTS</b> G <b>ATV</b> T <b>GNGPNWV</b> Q <b>EKPDFHLFTGL</b> Q <b>AGTNNR</b> A <b>PGVP</b> F <b>ARSFSGL</b> I <b>DKAALT</b> Q <b>GAQTE</b> D <b>EVY</b> <b>FCA</b> LWFSS <b>HV</b> F <b>GGGT</b> K <b>LTV</b> <b>GGGG</b> S <b>GGG</b> <b>GSGGGGSE</b> V <b>QLQNSQGPVLVKPGVKSIC</b> K <b>ASG</b> <b>YPTG</b> <b>YMMH</b> W <b>VRQSLV</b> R <b>LEW</b> <b>GIRN</b> P <b>TYG</b> <b>PSV</b> <b>YHN</b> <b>HKD</b> <b>ASL</b> <b>T</b> <b>DVSS</b> <b>TA</b> <b>YML</b> <b>HSLS</b> <b>T</b> <b>EDS</b> <b>A</b> <b>YV</b> <b>CARE</b> <b>LGYY</b> <b>WQGT</b> <b>ITLV</b> <b>SGGG</b> <b>SEVQ</b> <b>LVE</b> <b>GGGLVOPG</b> <b>PLRS</b> <b>L</b> <b>CSA</b> <b>G</b> <b>ASF</b> <b>G</b> <b>TF</b> <b>D</b> <b>DY</b> <b>T</b> <b>M</b> <b>HWR</b> <b>QAP</b> <b>GK</b> <b>LEW</b> <b>WGS</b> <b>WS</b> <b>NC</b> <b>S</b> <b>I</b> <b>G</b> <b>AD</b> <b>S</b> <b>V</b> <b>KGR</b> <b>F</b> <b>T</b> <b>SD</b> <b>N</b> <b>AK</b> <b>SS</b> <b>Y</b> <b>Q</b> <b>M</b> <b>NS</b> <b>L</b> <b>R</b> <b>A</b> <b>E</b> <b>D</b> <b>T</b> <b>AL</b> <b>Y</b> <b>C</b> <b>AD</b> <b>NS</b> <b>G</b> <b>Y</b> <b>H</b> <b>Y</b> <b>Y</b> <b>G</b> <b>D</b> <b>W</b> <b>W</b> <b>Q</b> <b>G</b> <b>GT</b> <b>TV</b> <b>TA</b> <b>S</b> <b>GGG</b> <b>S</b> <b>G</b> <b>G</b> <b>S</b> <b>G</b> <b>G</b> <b>S</b> <b>EIVMTQSPATL</b> <b>V</b> <b>S</b> <b>P</b> <b>G</b> <b>E</b> <b>R</b> <b>A</b> <b>T</b> <b>L</b> <b>S</b> <b>C</b> <b>R</b> <b>A</b> <b>S</b> <b>Q</b> <b>V</b> <b>S</b> <b>S</b> <b>N</b> <b>L</b> <b>W</b> <b>Y</b> <b>Q</b> <b>K</b> <b>P</b> <b>Q</b> <b>G</b> <b>A</b> <b>R</b> <b>P</b> <b>L</b> <b>T</b> <b>G</b> <b>A</b> <b>R</b> <b>T</b> <b>G</b> <b>I</b> <b>P</b> <b>R</b> <b>S</b> <b>G</b> <b>S</b> <b>G</b> <b>S</b> <b>T</b> <b>E</b> <b>F</b> <b>L</b> <b>T</b> <b>I</b> <b>S</b> <b>L</b> <b>Q</b> <b>S</b> <b>E</b> <b>D</b> <b>I</b> <b>A</b> <b>V</b> <b>Y</b> <b>C</b> <b>Q</b> <b>H</b> <b>Y</b> <b>I</b> <b>N</b> <b>W</b> <b>P</b> <b>L</b> <b>F</b> <b>G</b> <b>G</b> <b>G</b> <b>K</b> <b>V</b> <b>E</b> <b>I</b> <b>H</b> <b>HHHHH</b>

**Figure 25: Full amino acid sequence of the bispecific m15B6-based antibody.** Sequence of BsAb1 incorporating m15B6 VL and VH domains, CD3-binding regions, linkers, and purification tags. Sequence shows 100% identity with the construct described by Liu et al., 2024 [41]. Adapted from Chakraborty et al., 2024 [42].

For our master's thesis, we adopted the murine variant of 15B6 (m15B6) for initial validation due to its robust characterization and preclinical efficacy [42][41]. We confirmed the identity of the m15B6 antibody sequences from both sources using NCBI BLAST alignment (Figure 26), ensuring perfect correspondence between the sequences from Liu et al. (2024, Supplementary Fig. 1) and Chakraborty et al. (2024, Supplementary Table S1).

Score	Expect	Method	Identities	Positives	Gaps
459 bits(1182)	3e-172	Compositional matrix adjust.	223/223(100%)	223/223(100%)	0/223(0%)
Query 1	QAVVTQESALTTSPGETVLTCRSSTGAVTTGNYPNWQVEKPDLHFTGLIAGTNRNPGV				60
Sbjct 1	QAVVTQESALTTSPGETVLTCRSSTGAVTTGNYPNWQVEKPDLHFTGLIAGTNRNPGV				60
Query 61	PARFSGSLIGDKAALTTITGAQTEDEAIYFCALWFSSHVFGGGTQLTVLEVLQ0QSGPVL				120
Sbjct 61	PARFSGSLIGDKAALTTITGAQTEDEAIYFCALWFSSHVFGGGTQLTVLEVLQ0QSGPVL				120
Query 121	VKPGASVKISCKASGYSTGTGYYMHWRQSLVKRLEWIGRINPYTGVPYSYKHNFKDKAASLT				180
Sbjct 121	VKPGASVKISCKASGYSTGTGYYMHWRQSLVKRLEWIGRINPYTGVPYSYKHNFKDKAASLT				180
Query 181	VDKSSSTAYMELHSLTSEDSAVYYCARELGGYGWGOGTTLVSS	223			
Sbjct 181	VDKSSSTAYMELHSLTSEDSAVYYCARELGGYGWGOGTTLVSS	223			
Description		Scientific Name	Max Score	Total Score	E value
▼		▼	▼	▼	▼
unnamed protein product			459	459	100%
			3e-172	100.00%	223
			Query_7614659		

**Figure 26: BLAST alignment of m15B6 antibody sequences.** NCBI BLAST analysis confirms 100% identity between the variable region sequences of m15B6 antibodies from *Liu et al.* (2024) and *Chakraborty et al.* (2024).

The amino acid sequence for the m15B6 variable light (VL) and heavy (VH) chains is as follows:

**m15B6 VL (1–109):**

```
QAVVTQESALTTSPGETVTLTCRSSTGAVTTGNYPNWVQEKPDLFTGLIAGTNNR  
APGVPARFSGSLIGDKAALTITGAQTEDEAIYFCALWFSSHWFGGTKLTVL
```

**m15B6 VH (1–114):**

```
EVQLQQSGPVLVKPGASVKISCKASGYSFTGYYMHWVRQSLVKRLEWIGRINPYTG  
VPSYKHNFKDKASLTVDKSSSTAYMELHSLTSEDSAVYYCARELGGYWQGTTLVSS
```

These sequences were merged into a single-chain format using a flexible linker:

```
GGGGSGGGGGGGGGGS
```

Final configuration: VL--linker--VH

The construct was then optimized at the nucleotide level for expression in human embryonic kidney (HEK293T) cells via human codon optimization. This step ensures efficient translation and proper protein folding within human cell lines used for protein production.

For verification purposes, the designed DNA sequences of the light and heavy chains were verified using the IgBlast online tool [65]. As Figure 27 and Figure 28 shows, IgBlast correctly identified our DNA sequence as the light and heavy chains of an antibody. The CDR regions match in both number and location. Although Figure 24 from *Liu et al., 2024 supp1* displays a larger CDR region window, it always includes all the amino acids detected by IgBlast as part of the CDRs. The translated sequence also corresponds. However, the last nucleotides of both the light and heavy chains do not appear to be recognized by IgBlast as essential and are therefore not included in the translation, resulting in missing the last 2 to 5 amino acids. This discrepancy, located outside the CDRs, is attributed to IgBlast's specific operational characteristics rather than any critical issue with the designed sequences.

V-(D)-J rearrangement summary for query sequence (multiple equivalent top matches, if present, are separated by a comma):

Top V gene	match	Top J gene	match	Chain type	stop codon	V-J frame	Productive	Strand	V frame shift
IGLV1*01		IGLJ1*01		VL	No	In-frame	Yes	+	No

## Alignments

**Figure 27: IgBlast analysis of the m15B6 light chain (VL).** Germline V(D)J rearrangement and CDR region annotation of the m15B6 VL sequence, confirming productive, in-frame recombination.

V-(D)-J rearrangement summary for query sequence (multiple equivalent top matches, if present, are separated by a comma):

Top V gene match	Top D gene match	Top J gene match	Chain type	stop codon	V-J frame	Productive	Strand	V frame shift
IGHV1-18-11*01	IGHD6-1-1*01	IGHJ2*01	VH	No	In-frame	Yes	+	No

## Alignments

**Figure 28: IgBlast analysis of the m15B6 heavy chain (VH). Germline V(D)J rearrangement and CDR region annotation of the m15B6 VH sequence, confirming productive, in-frame recombination.**

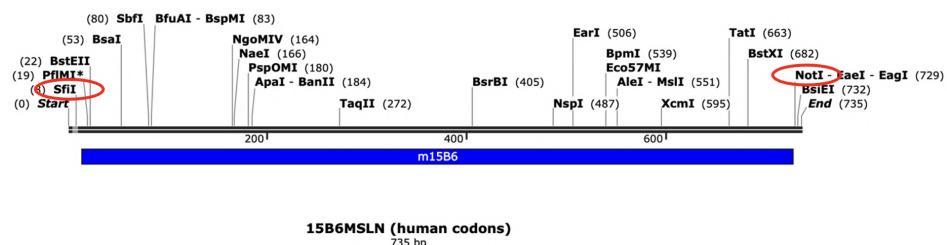
To facilitate subsequent cloning into the SNAP-tag plasmid available in our laboratory, we incorporated specific restriction enzyme sites (*SfiI* and *NotI*) flanking the antibody sequence. These restriction sites were selected due to their absence within the antibody sequence, ensuring accurate insertion without unintended digestion of internal sites.

The final optimized DNA construct is:

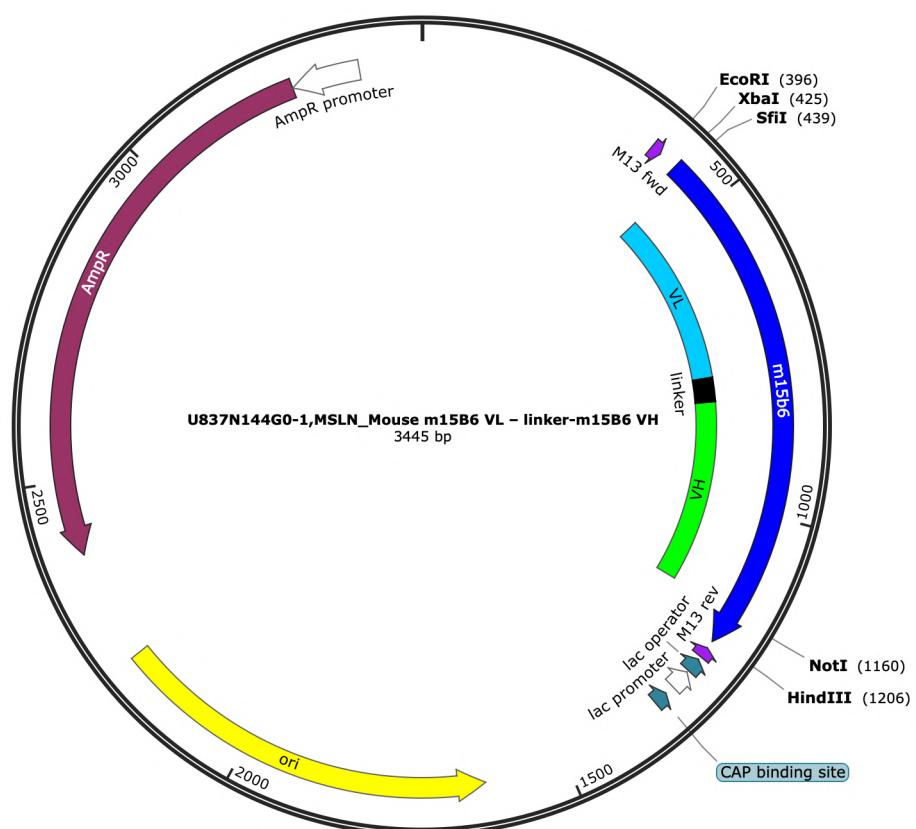
S<sub>fi</sub>I -- m15B6 VL -- linker -- m15B6 VH -- N<sub>ot</sub>I

(Full DNA sequence provided in the appendix.)

This finalized design (Figure 29) was handed over to GenScript in the Netherlands for synthesis and subsequent shipment to our laboratory in South Africa. The construct was provided in a pUC57 vector with ampicillin resistance, and it was shipped freeze-dried following mutagenesis at the XbaI/HindIII cloning sites. (Figure 30)

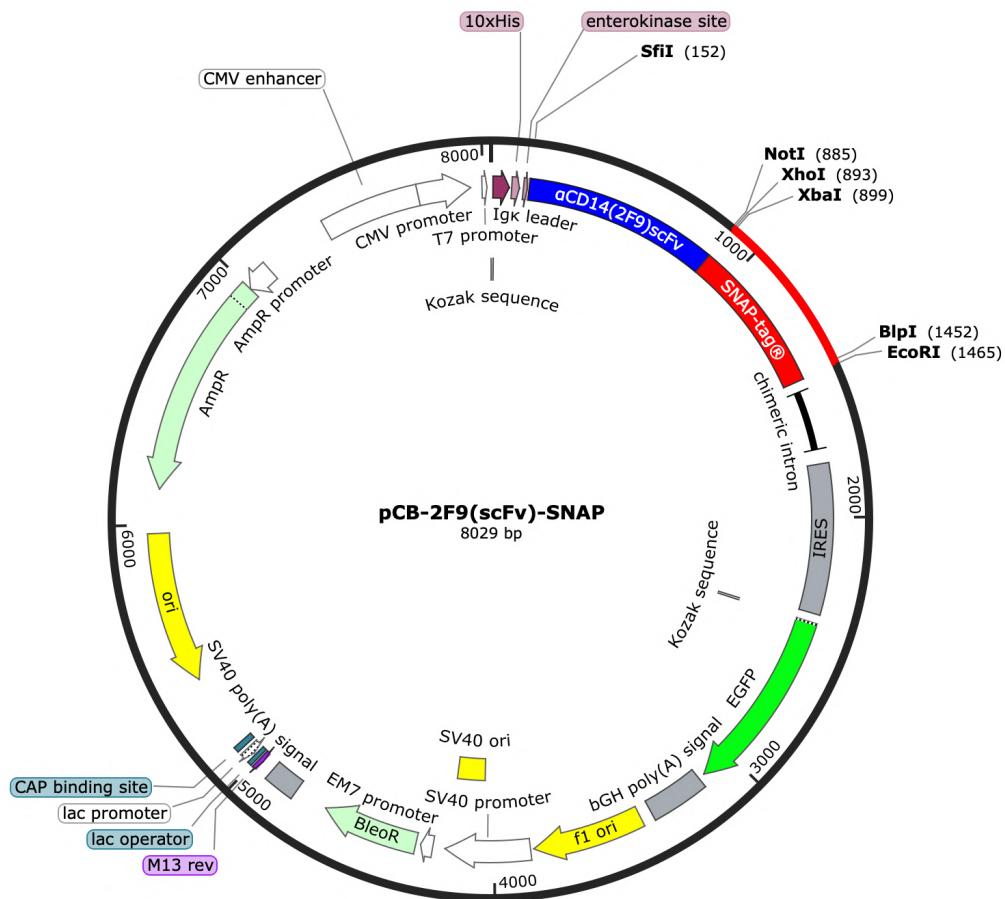


**Figure 29: SnapGene map of the m15B6 antibody cassette.** Schematic showing the insertion of the m15B6 VL–linker–VH antibody sequence between the *SfiI* and *NotI* restriction sites highlighted.

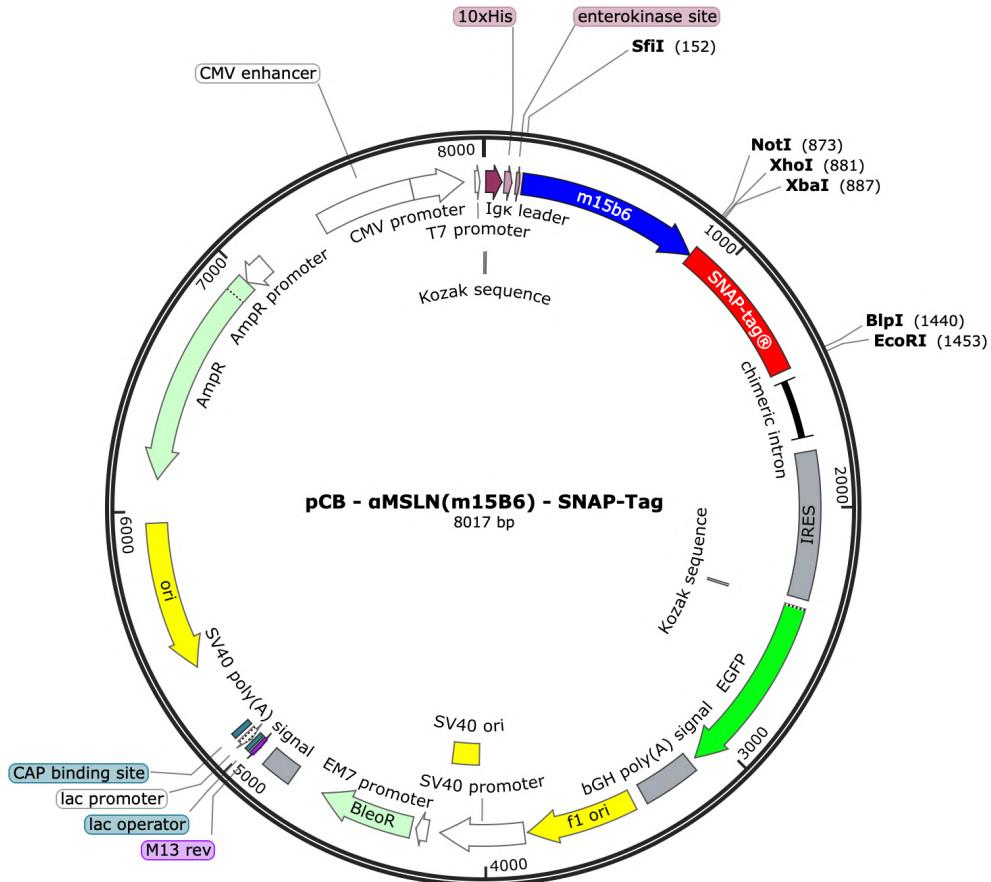


**Figure 30: pUC57 -  $\alpha$ MSLN (m15B6) (scFv).** SnapGene schematic of the m15B6 plasmid construct cloned by GeneScript in a pUC57 vector

The downstream cloning strategy for the SNAP-tag fusion involves using an existing  $\alpha$ CD14(2F9)(scFv)–SNAP-tag construct (Figure 31). This plasmid will be digested using *SfiI* and *NotI* restriction enzymes to enable insertion of the m15B6 antibody sequence. SnapGene simulations confirmed the feasibility of this strategy, illustrating the final designed plasmid (Figure 32) suitable for subsequent transfection and expression experiments. The main characteristics of the final plasmid include the m15B6 antibody sequence fused in-frame to the SNAP-Tag sequence, facilitated by an IgK leader peptide and an enterokinase cleavage site. Additional features include a 10xHis-tag for protein purification, EGFP for monitoring transfection efficiency, and a Zeocin resistance gene (BleoR) for stable cell selection. Further details of this cloning strategy are described comprehensively in the Materials and Methods section.



**Figure 31: pCB -  $\alpha$ CD14(2F9)(scFv) - SNAP-Tag construct.** Image obtained from *Snapgene*.

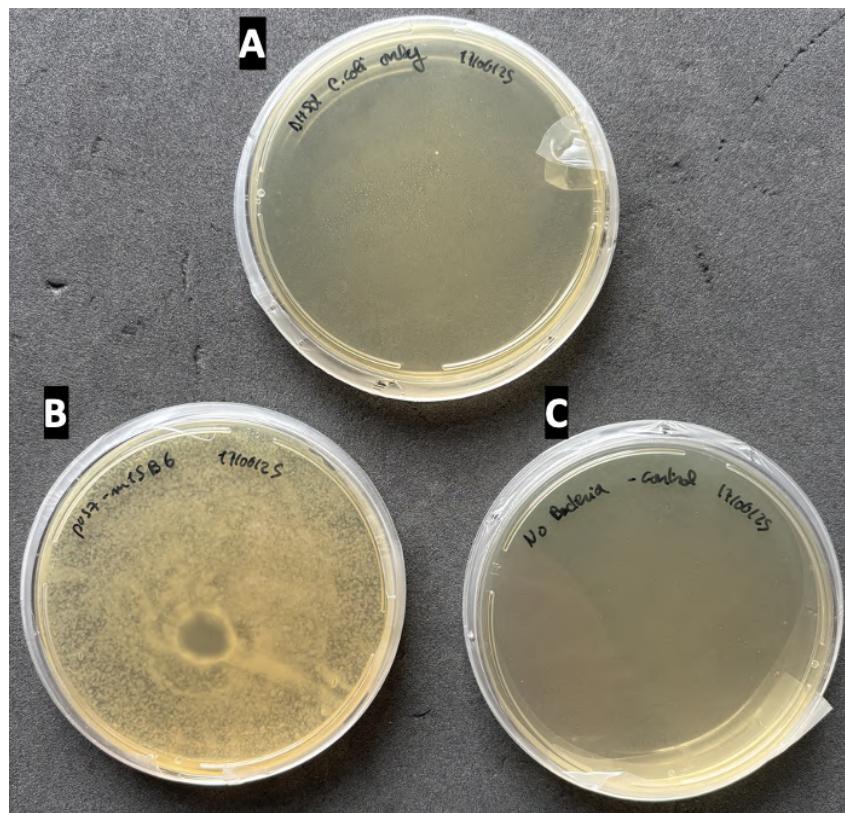


**Figure 32: pCB -  $\alpha$ MSLN(m15B6)(scFv) - SNAP-Tag construct.** Image obtained from *Snapgene*.

### 3.3.1 Cloning of pCB - $\alpha$ MSLN(m15B6)(scFv) - SNAP

The following cloning experiments were conducted in collaboration with two students I supervised: Samantha Levetan (University of Cape Town), as part of her Honours thesis, and Emma Jakubicka (Massachusetts Institute of Technology), as part of her summer research project. Under my guidance, they completed the initial stages of cloning, as detailed below. Both students will carry forward the downstream steps as part of their respective academic work.

To begin the cloning process, the lyophilized pUC57-  $\alpha$ MSLN(m15B6)(scFv) plasmid (Figure 30) received from GenScript was resuspended and directly transformed into *E. coli* DH5 $\alpha$  competent cells, following GenScript and NEB transformation protocols. As shown in Figure 33, no colonies were observed in the negative controls: (A) transformation without plasmid DNA and (C) LB-agar plate with no bacteria added—indicating no background growth or contamination. In contrast, Panel B confirms successful transformation of pUC57- $\alpha$ MSLN(m15B6)(scFv), as evidenced by the high density of colonies. A single colony was selected and expanded for plasmid extraction and downstream digestion.

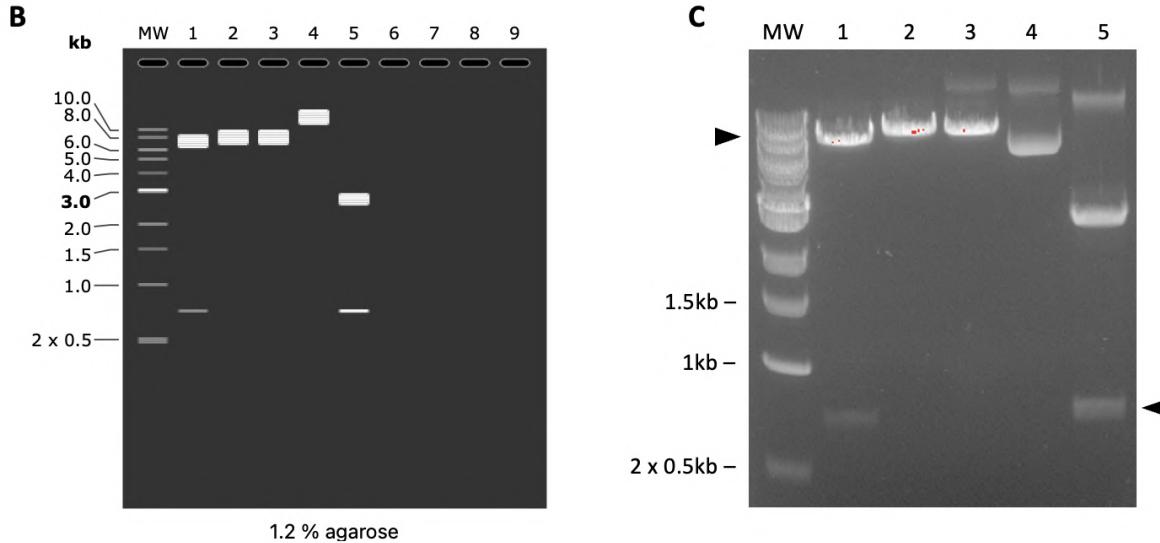


**Figure 33: Transformation controls for pUC57- $\alpha$ MSLN(m15B6)(scFv).** (A) *E. coli* DH5 $\alpha$  only (no plasmid); no colony growth observed. (B) pUC57-m15B6 transformation; high colony count confirms successful uptake. (C) No bacteria control; no colony growth, confirming sterility of LB-agar plates.

pCB- $\alpha$ CD14(2F9)(scFv)-SNAP and pUC57- $\alpha$ MSLN(m15B6)(scFv) were digested with *Sfi*I and *Not*I restriction enzymes. The expected insert and vector fragments were successfully generated and confirmed by agarose gel electrophoresis (Figure 34). Specifically, the top band in lane 1 corresponds to the SNAP vector backbone ( 7.3 kb), and the bottom band in lane 5 to the  $\alpha$ MSLN(m15B6)(scFv) insert ( 721 bp). These fragments were excised and purified for downstream ligation.

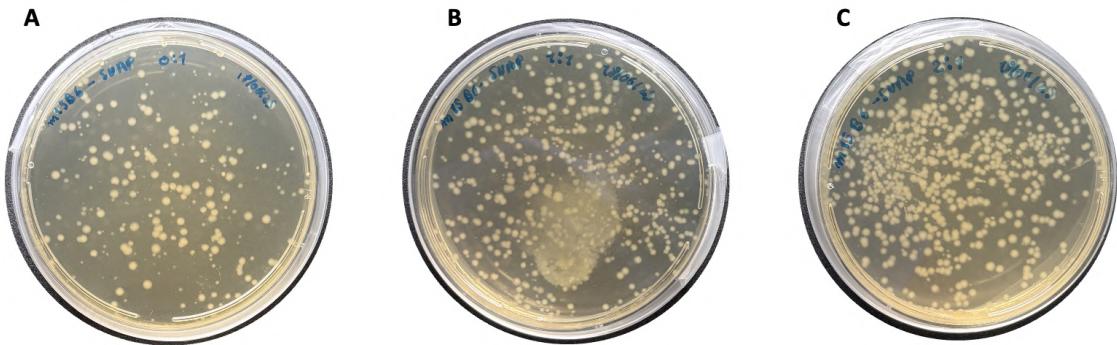
**A**

MW: New England Biolabs 1 kb DNA Ladder		
1: pCB - $\alpha$ CD14(2F9) - SNAP Sfil + NotI <b>1.7296 bp</b> 2.733 bp	3: pCB - $\alpha$ CD14(2F9) - SNAP NotI 1.8029 bp	5: pCB - $\alpha$ m15B6 - SNAP Sfil + NotI 1.2724 bp <b>2.721 bp</b>
2: pCB - $\alpha$ CD14(2F9) - SNAP Sfil 1.8029 bp	4: pCB - $\alpha$ CD14(2F9) - SNAP (SUPERCOILED) 1.8029 bp	



**Figure 34: Restriction digestion of pCB- $\alpha$ CD14(2F9)(scFv)-SNAP and pUC57- $\alpha$ MSLN(m15B6)(scFv) using agarose gel electrophoresis.** (A) Legend of the samples loaded in each well. The extracted fragments used for ligation are highlighted in red. (B) Simulated digestion using SnapGene. (C) Agarose gel electrophoresis results. DNA fragments were separated on a 1.2% agarose gel for 1 hour at 100 volts. The extracted fragments used for ligation are indicated by arrowheads in lanes 1 (top band) and 5 (bottom band).

Following plasmid extraction and restriction digestion (Figure 34), ligation reactions were set up with varying insert-to-vector molar ratios (0:1, 1:1, and 2:1) (section 2.4.6). The transformed ligation products were plated and yielded the following colony counts: 0:1 (no insert) – 163 colonies, 1:1 – 360 colonies, and 2:1 – 307 colonies (Figure 35). The presence of colonies in the 0:1 condition may reflect re-ligation of vector backbone, while the elevated colony counts in the 1:1 and 2:1 conditions suggest successful ligation and transformation events. The calculated transformation efficiencies were:  $2.72 \times 10^3$ ,  $6.00 \times 10^3$ , and  $5.12 \times 10^3$  cfu/ $\mu$ g DNA.

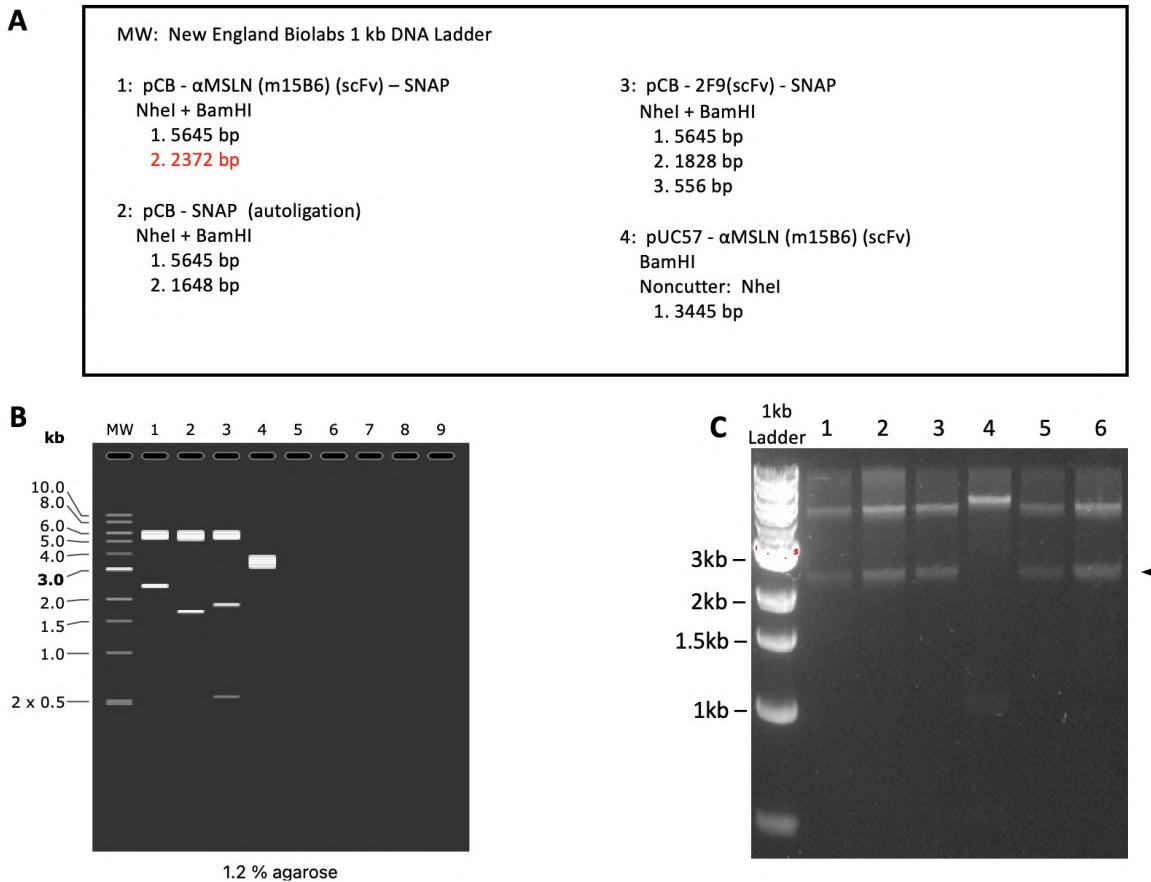


**Figure 35: Growth of *E. coli* cells transformed with potential recombinant pCB- $\alpha$ MSLN(m15B6)(scFv)-SNAP plasmid DNA.** (A) vector-only control; moderate colony growth suggests re-ligation. (B) 1:1 insert:vector ratio; highest number of colonies observed. (C) 2:1 insert:vector ratio; similar colony density to 1:1. Colonies were grown overnight at 37°C on LB agar containing 200  $\mu$ g/ml ampicillin.

### 3.3.2 Colony screening and validation of pCB- $\alpha$ MSLN(m15B6)(scFv)-SNAP

To identify colonies harboring correctly assembled constructs, restriction screening was performed on miniprep plasmid DNA using *NheI* and *BamHI*. SnapGene was used to simulate the expected digestion patterns for the correctly ligated bispecific vector, auto-ligated backbone, and other potential misligation products.

As shown in Figure 36, only colony 4 displayed an unexpected digestion profile, all the other colonies displayed the desired digestion profile, matching the expected 2372 bp fragment, confirming the presence of the full-length pCB- $\alpha$ MSLN(m15B6)(scFv)-SNAP construct. Colony 2 was therefore selected and cultured for plasmid DNA extraction and expansion for downstream applications such as mammalian cell transfection.

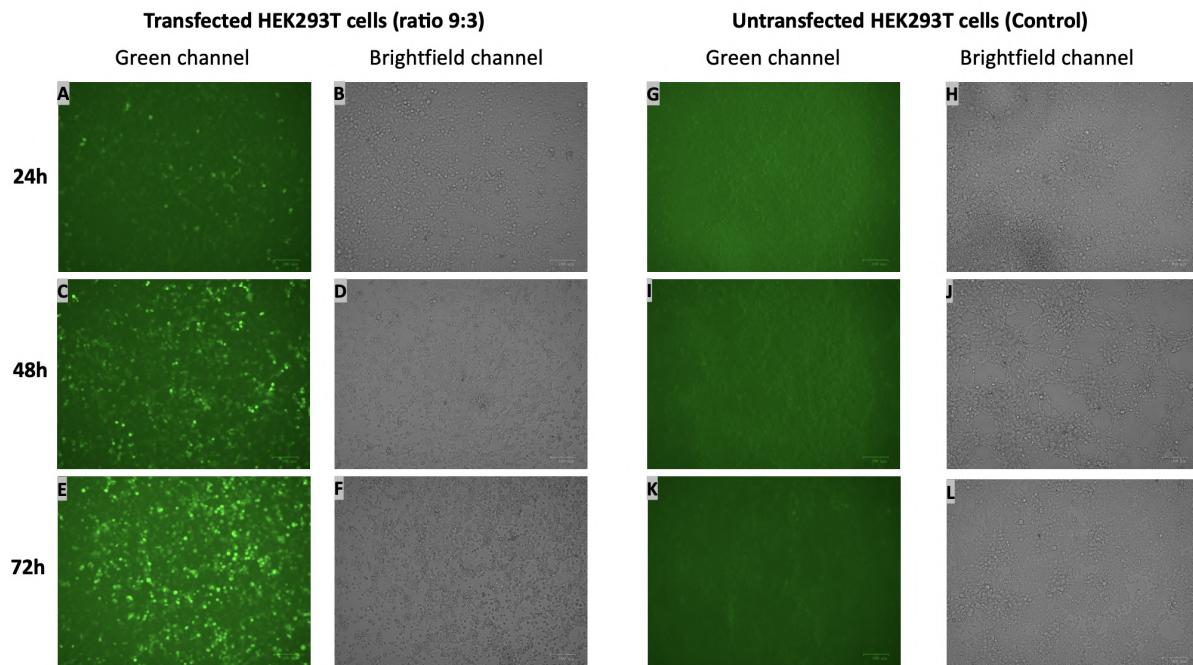


**Figure 36: Colony screening by *NheI* and *BamHI* digestion to identify correct ligation of pCB- $\alpha$ MSLN(m15B6)(scFv)-SNAP.** (A) Summary of expected fragment sizes. Diagnostic fragment used to differentiate constructs is highlighted in red. (B) SnapGene simulation of expected banding patterns for the correct construct (lane 1) and alternative outcomes (lanes 2–4). (C) Agarose gel of *NheI* and *BamHI*-digested plasmid DNA from colonies 1–6. All colonies apart colony 4, show the expected 2372 bp fragment, confirming successful ligation.

### 3.3.3 Transfection of pCB- $\alpha$ MSLN(m15B6)(scFv)-SNAP into HEK293T cells

To assess the expression of the newly cloned construct pCB- $\alpha$ MSLN(m15B6)(scFv)-SNAP, transient transfection was performed in HEK293T cells using a DNA:XtremeGENE transfection ratio of 3:1 (3  $\mu$ g DNA: 9  $\mu$ l reagent). Expression of the GFP reporter was monitored at 24, 48, and 72 hours post-transfection using the ZOE<sup>TM</sup> Fluorescence Cell Imager (Bio-Rad, USA). Representative images from both the green fluorescence and brightfield channels are shown in Figure 37. Untransfected HEK293T cells served as a negative control.

A clear increase in GFP fluorescence intensity was observed over time in the transfected condition, indicating successful expression of the SNAP-tag fusion protein. No fluorescence was detected in the untransfected control cells. These microscopy results confirm effective plasmid delivery and initial expression in HEK293T cells.

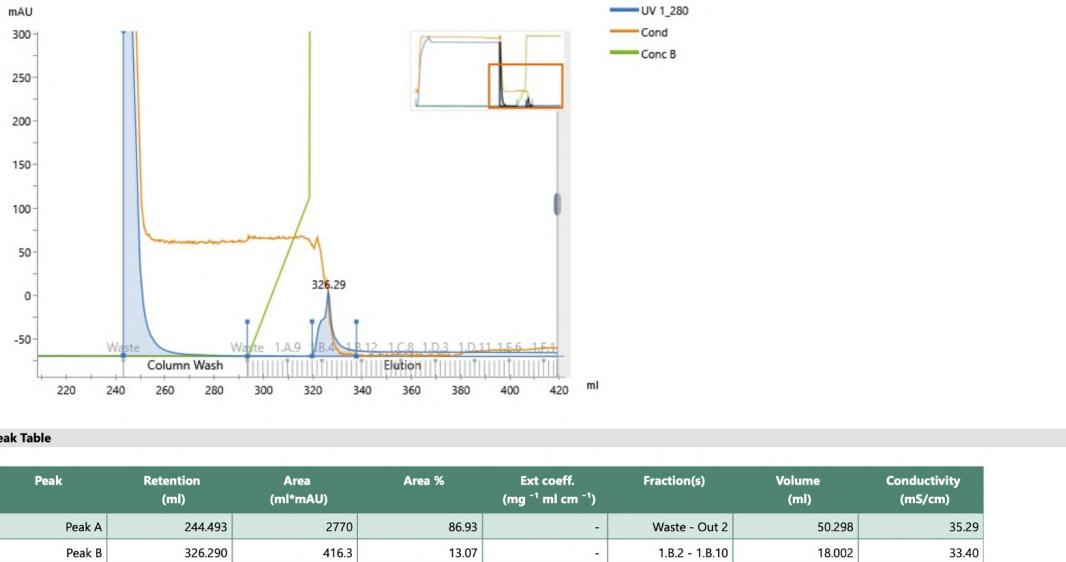


**Figure 37: Visualization of transient GFP expression in HEK293T cells transfected with pCB- $\alpha$ MSLN(m15B6)(scFv)-SNAP.** HEK293T cells were transfected with the pCB- $\alpha$ MSLN(m15B6)(scFv)-SNAP construct using a DNA:XtremeGENE ratio of 3:1. Images were acquired at 24 h, 48 h, and 72 h post-transfection using the ZOE<sup>TM</sup> fluorescent cell imager. (A–F) Transfected cells: (A, C, E) green fluorescence channel at 24, 48, and 72 h, respectively; (B, D, F) corresponding brightfield images. (G–L) Untransfected control cells: (G, I, K) green fluorescence channel at 24, 48, and 72 h, respectively; (H, J, L) corresponding brightfield images. Increased GFP signal over time was only observed in transfected cells, confirming successful expression. Scale bar = 100  $\mu$ m.

Sequencing results and flow cytometry analysis to confirm construct integrity and transfection efficiency are currently pending and will be completed by the students I supervised as part of the ongoing project continuation.

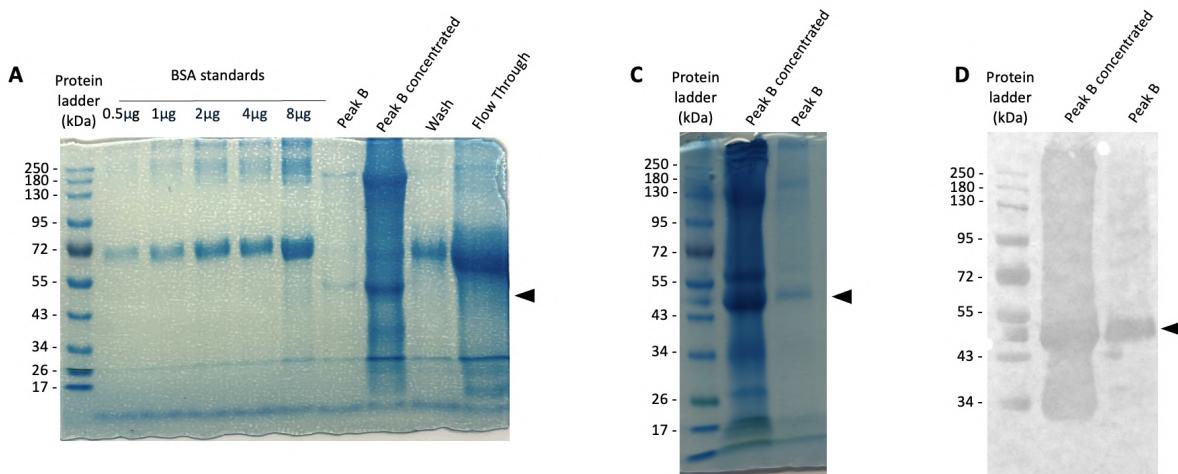
### 3.4 Additional experiment: Purification and Characterization of $\alpha$ CSPG4(scFv)-SNAP Fusion Protein

Due to the time required for the production of the bispecific construct (pCB- $\alpha$ EGFR(scFv)-SNAP- $\alpha$ MSLN(scFv)) which is constrained by the slow expression kinetics in incubator-based culture systems lacking bioreactor support, an additional experiment was undertaken to gain hands-on experience with protein purification. To this end, a previously produced recombinant  $\alpha$ CSPG4(scFv)-SNAP fusion protein, generated by former lab member Dr. Neelakshi Mungra, was used to evaluate the purification workflow on the AKTA Avant system. The protein was purified via immobilized metal affinity chromatography (IMAC), as described in the methods section.



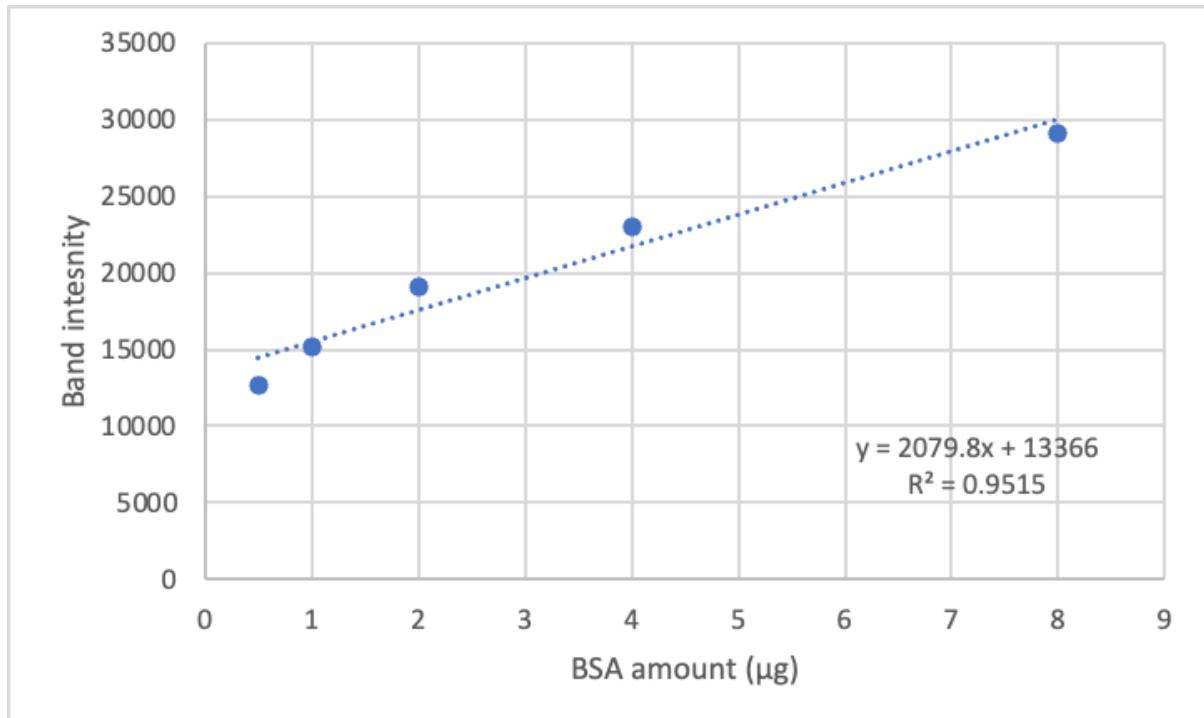
**Figure 38: Chromatogram of  $\alpha$ CSPG4(scFv)-SNAP protein purification.** UV absorbance at 280 nm (blue), conductivity (orange), and imidazole gradient (green) are shown. Two elution peaks were observed, with Peak A corresponding to the column wash and Peak B corresponding to the target protein as indicated by its retention time and conductivity profile. Fraction collection range is highlighted.

As shown in Figure 38, Peak B (326.29 mL) was identified as the elution peak corresponding to the His-tagged  $\alpha$ CSPG4(scFv)-SNAP protein. According to the peak table, it represented approximately 13% of the total absorbance area and was eluted in fractions 1.B.2 to 1.B.10. These fractions were collected and retained as they correspond to the most enriched portion of the target protein. To remove residual imidazole and concentrate the protein, the pooled fractions were further processed using 30 kDa Amicon<sup>TM</sup> centrifugal filters, as detailed in section 2.4.10. For confirmation of purification, protein was analysed using SDS-PAGE and western blot, using the IMAC elution profile to identify fractions for evaluation (Figure 38). The mammalian expression plasmid contained a 10x histidine-encoding tag, which allowed for the purification and visualisation of the expressed protein.



**Figure 39: Coomassie staining and western blot analysis of purified  $\alpha$ CSPG4(scFv)-SNAP fractions.** (A) SDS-PAGE gel showing BSA standards (0.5–8  $\mu$ g) and  $\alpha$ CSPG4(scFv)-SNAP fractions. A clear band at 51 kDa is visible in both Peak B and Peak B concentrated lanes, and absent in the wash and flow-through lanes, confirming specific retention and elution of the His-tagged protein (arrowhead). (B) SDS-PAGE comparison of concentrated and non-concentrated Peak B fractions. (C) Western blot of the same samples using  $\alpha$ -His-HRP antibody, validating the identity of the fusion protein at 51 kDa.

SDS-PAGE analysis confirmed the presence of a distinct protein band at approximately 51 kDa, matching the expected size of the recombinant  $\alpha$ CSPG4(scFv)-SNAP construct (Figure 39A,B). Notably, this band was observed only in the Peak B and Peak B concentrated fractions and was absent from the wash and flow-through lanes, supporting successful column retention and specific elution. A diffuse vertical smear was visible in the concentrated lane, likely due to protein overload or minor buffer contamination; however, this did not affect band integrity. Western blot analysis using an  $\alpha$ -His-HRP antibody further confirmed the identity of the His-tagged protein in both samples (Figure 39C). Protein yield was estimated by densitometric comparison to the BSA standard curve at approximately 6.99  $\mu$ g (see also Figure 40).



**Figure 40: Generation of a BSA standard curve for the determination of  $\alpha$ CSPG4(scFv)-SNAP protein concentration.** Following measurement of optical colour intensity for the BSA standard bands (using ImageJ software), a graph of optical colour intensity (in units) was plotted against the corresponding amount of protein (in  $\mu\text{g}$ ). Using the resulting graphical equation, the amount of scFv-SNAP protein in the total protein sample, was estimated, and allowed determination of the protein yield and percentage purity

## 4 Discussion

### 4.1 Unmet clinical need and the promise of targeted biologics

Cervical cancer remains the *fourth* most common malignancy in women worldwide, with striking disparities between high-income and low-/middle-income regions, including South Africa (*WHO, 2022*) [1]; (*IARC, 2020*) [2]. Persistent infection with oncogenic human papillomavirus (HPV) drives disease; however, clearance versus persistence is dictated by a complex interplay between local immunity and the vaginal microbiome (*Ntuli et al., 2022*) [14]. For metastatic or recurrent disease, platinum-based chemoradiotherapy yields a median overall survival below 17 months and substantial morbidity (*Cohen et al., 2019*) [3]. Immune-checkpoint inhibitors produce durable responses in a small subset of patients but are limited by low tumour mutational burden and immunosuppressive stroma (*Colombo et al., 2024*) [66]. Together, these factors underscore an urgent need for precision therapeutics with higher efficacy and lower systemic toxicity. ADCs, already practice-changing in breast, bladder and lung cancers, represent a logical next leap for cervical cancer (*Drago et al., 2021*) [67].

## 4.2 Mesothelin and the rationale for bispecific strategies

In cervical cancer, information on MSLN expression has until recently been limited. However, new evidence indicates that a substantial subset of cervical cancers express MSLN, albeit with differences between histological subtypes. A 2022 immunohistochemical study by Takamizawa et al. examined MSLN in 123 cervical cancer patients and found MSLN expression in 98% of tumors, with high expression (moderate to strong staining in  $\geq 30\%$  of cells) in 63% of cases (*Takamizawa et al., 2022*) [37]. Notably, high MSLN expression was significantly more frequent in adenocarcinomas ( $\sim 80\%$ ) compared to SCC ( $\sim 49\%$ ) [37]. This highlights the risk of target escape in monovalent ADC formats and justifies the design of dual-targeting constructs.

In the present study, an EGFR co-targeting strategy was adopted to address this heterogeneity, supported by bioinformatic identification of a distinct **MSLN-low/EGFR-high** tumour subtype. EGFR has previously been implicated in driving cell proliferation, resistance to radiotherapy, and immune evasion in multiple epithelial cancers including cervical SCC (*Yarden et al., 2021*) [68]. Co-targeting MSLN and EGFR thus expands tumour coverage and may reduce the risk of therapeutic failure. Such bispecific ADC approaches have already shown preclinical promise in other heterogeneous tumours like triple-negative breast cancer (TNBC), where dual targeting of MSLN and nucleolin (NCL) has demonstrated strong efficacy (*Thongchot et al., 2024*) [61].

## 4.3 Production of bispecific SNAP-tag fusion proteins

Despite achieving construct integrity through directional cloning with *SfiI* and *NotI*, protein production remained a key bottleneck in the timeline of this project. Vector self-ligation was encountered during cloning, as reported in other SNAP-tag workflows, and required enzymatic dephosphorylation and optimization of insert-to-vector ratios. Moving to seamless assembly techniques such as Gibson or Golden Gate cloning would streamline construct generation in future iterations (*Garcia-Ruiz et al., 2025*) [69].

Once constructs were sequence-verified, transient transfection in HEK293T cells yielded detectable protein levels and confirmed GFP co-expression, indicating reasonable transfection efficiency. However, the main constraint was not efficiency per se, but the time required for protein accumulation in static flask-based incubator systems. These conditions limit cell density and nutrient exchange, making it necessary to culture large volumes, often several hundred millilitres to a litre, to reach sufficient quantities for downstream applications such as His-tag purification via IMAC.

While protein expression was successful, this time-intensive format delayed conjugation and functional assays. Transitioning to suspension-adapted systems such as Expi293F or CHO-GS, which allow for higher cell densities and improved scalability, could shorten production timelines significantly (*Sargon et al., 2024*) [70]. Moreover, the use of stirred-tank bioreactors would enable finer control over pH, oxygenation, and feeding schedules, improving reproducibility and potentially boosting protein concentration—thereby reducing the volume required for purification (*Marienberg et al., 2021*) [71].

Altogether, while the expression system used here proved functional and sufficient for early-stage validation, future iterations of this work would greatly benefit from adopting high-density or bioreactor-based production platforms to accelerate throughput and enable more rapid iteration on bispecific antibody formats.

#### 4.4 SNAP-tag scaffolds for next-generation ADCs and precision delivery

While the current study did not include conjugation or cytotoxicity testing, the choice of the SNAP-tag platform offers several theoretical advantages over conventional ADC formats that warrant further development. Classical ADCs, including those used in cervical cancer such as tisotumab vedotin, suffer from heterogeneity due to stochastic lysine or cysteine conjugation, resulting in variable drug-to-antibody ratios (DAR), unpredictable pharmacokinetics, and increased off-target toxicity (*Okojie et al., 2023*) [26]. In contrast, the SNAP-tag system enables covalent 1:1 conjugation via a genetically encoded O<sup>6</sup>-alkylguanine-DNA alkyltransferase domain that selectively binds benzylguanine (BG)-linked payloads under physiological conditions (*Keppler et al., 2003*) [31]. This feature produces highly uniform ADC populations with improved stability, pharmacodynamics, and manufacturing reproducibility (*Woitok et al., 2017*) [54].

The pCB- $\alpha$ EGFR(scFv)-SNAP- $\alpha$ MSLN(scFv) construct designed in this thesis leverages this modularity to accommodate future site-specific conjugation to diverse effector classes. In addition to classic cytotoxins such as auristatin F, BG-linked photosensitizers or radioisotopes could also be attached in a plug-and-play manner, enabling theranostic applications or combinatorial killing mechanisms (*Hussain et al., 2019*) [33]. Notably, pathway enrichment analyses from this project revealed upregulation of Wnt/ $\beta$ -catenin signaling in MSLN-low cervical tumors, suggesting that condensate-modulating inhibitors of this pathway may offer a particularly effective payload option (*Li et al., 2021*) [62]; (*Zhang et al., 2020*) [63]. The bispecific SNAP-tag format could thus serve as a delivery vehicle not only for broadly cytotoxic agents but also for precision payloads targeting molecular vulnerabilities unique to each tumour subtype. This strategy may be especially valuable in cervical cancer, where tumor heterogeneity and inconsistent biomarker expression often undermine monovalent ADC efficacy (*Qiu et al., 2025*) [27]; (*Takamizawa et al., 2022*) [37].

Taken together, the integration of biomarker-driven bioinformatic analysis with modular construct design exemplifies a rational pipeline for developing next-generation biologics (*Frullanti et al., 2024*) [72]. Rather than relying on a one-size-fits-all cytotoxic approach, this platform enables tailoring both the targeting domains and payload class to the tumor's molecular landscape (*Okojie et al., 2023*) [26]; (*Qiu et al., 2025*) [27]. Such adaptability aligns with the emerging paradigm of precision oncology, in which therapies are designed around individual tumor features and immune contexture (*Kurzrock et al., 2022*) [73]; (*Kurzrock and Subbiah, 2018*) [74]. By enabling highly specific, stoichiometric, and interchangeable conjugation, SNAP-tag bispecifics may become key tools in

future immunotherapy regimens, either as standalone agents or in synergy with checkpoint inhibitors, radiosensitizers, or engineered T-cell platforms.

Given these considerations, one would expect the SNAP-tag system to outperform traditional conjugation chemistries in generating homogeneous, functionally active antibody-drug conjugates with reduced immunogenicity and improved target engagement. While experimental validation is still required, the rational design of the constructs in this study lays a strong foundation for subsequent in vitro and in vivo testing. Moreover, the platform is readily extensible to multiplexed payload formats or photochemical strategies, which may help overcome current therapeutic ceilings in advanced cervical cancer (*Drago et al., 2021*) [67]; (*Muppa et al., 2025*) [30].

## 4.5 Proteomic landscape of MSLN-low tumors supports pathway-guided targeting

Beyond surface marker prioritization, analysis of RPPA data revealed significant upregulation of proteins such as STAT5A, PRKAA1 (AMPK $\alpha$ 1), and RICTOR in MSLN-low cervical tumors, despite their intracellular localization limiting direct ADC application. These proteins are key regulators of oncogenic and metabolic pathways. STAT5, for instance, is a pivotal transcription factor activated by IL-2 in cervical cancer, mediating a metabolic shift toward aerobic glycolysis (*Rocha-Zavaleta et al., 2024*) [75]; (*Halim et al., 2020*) [76]. Similarly, PRKAA1 (AMPK $\alpha$ 1) plays a dual role in immunosuppression and metabolic regulation within the tumor microenvironment (*Sierra et al., 2019*) [77], while RICTOR, a component of the mTORC2 complex, has been implicated in survival signaling and cancer progression (*Vicentini et al., 2018*) [78].

These enrichments suggest that while not suitable as ADC targets, these pathways may define a distinct metabolic and immunomodulatory phenotype in MSLN-low tumors. This opens avenues for combination therapies e.g., pairing bispecific ADCs with STAT5 or AMPK inhibitors, or using these proteins as mechanistic readouts to monitor response to targeted payloads. This interpretation is further supported by our identification of FN1 (Fibronectin) as one of the few upregulated and accessible surface proteins in MSLN-low tumors. FN1 promotes tumorigenesis via FAK activation and may itself contribute to invasiveness and resistance (*Huang et al., 2019*) [79]. Its overexpression (0.1822 vs. -0.4679) and moderate significance ( $p = 0.0419$ ) support its inclusion in future targeting panels.

Although EGFR showed only a small protein-level difference (0.6389 vs. 0.4754;  $p = 0.4285$ ), its mRNA significant upregulation and negative correlation with MSLN expression justify its continued prioritization as a co-target. This is bolstered by prior success in EGFR (scFv)-SNAP conjugates (*Huysamen et al., 2023*) [34] and broad antibody availability, making EGFR a pragmatically valuable target despite limited proteomic discrimination. Nonetheless, these conclusions must be interpreted cautiously. Our analysis was based on a single dataset (TCGA PanCancer Atlas), and the RPPA platform lacks direct MSLN quantification, requiring indirect assumptions based on mRNA stratification.

Moreover, RPPA protein levels may be influenced by antibody specificity, batch effects, or detection sensitivity. The proteomic trend in EGFR, although not statistically significant, remains directionally consistent with mRNA results, suggesting that broader multi-cohort analyses or orthogonal proteomic validations may reinforce these findings and improve confidence in EGFR as a co-target. Importantly, this caveat extends to other findings as well, including the pathway enrichment analyses that implicated Wnt/β-catenin signalling and metabolic rewiring in MSLN-low tumors. While such pathways offer compelling mechanistic hypotheses and potential payload directions, their inference relies on the same dataset and cross-platform assumptions. As such, further experimental validation is necessary to determine the true functional relevance and therapeutic exploitability of these molecular features.

## 5 Future Work and Perspectives

This project reflects a broader paradigm shift in cancer immunotherapy: from single-target, one-size-fits-all approaches toward modular, subtype-adapted biologics. By integrating bioinformatics with rational construct design, a tailored dual-targeting system was built to address two major barriers in cervical cancer: MSLN heterogeneity and shedding.

The next phase of this project will focus on completing the collection and purification of the bispecific SNAP-tag constructs, particularly the αEGFR (scFv) -SNAP-αMSLN (scFv) antibody, as well as the anti-shedding αMSLN(m15B6) (scFv) -SNAP construct. Once purified, these fusion proteins will undergo functional evaluation and be benchmarked against existing monoclonal antibodies available in the laboratory, including the standard αMSLN (scFv), to assess potential advantages in specificity, efficacy, or resistance to antigen shedding.

Although time constraints prevented completion of these downstream assays during the course of this thesis, the experimental pipeline has been successfully established. Importantly, two students—**Samantha Levetan** (University of Cape Town) and **Emma Jakubicka** (Massachusetts Institute of Technology)—have been trained in the workflow and are now equipped to continue this research as part of their respective Honours and summer research projects.

In parallel, the bioinformatic analyses performed here support the existence of an MSLN-low molecular subtype in cervical cancer and identify EGFR, Notch1, and FN1 as potential co-targets. While encouraging, these findings remain exploratory and require further validation through orthogonal datasets and in vitro functional experiments.

Collectively, the cloning, expression, and in silico exploration presented in this work establish a foundation for future therapeutic development in MSLN-low tumors. It is hoped that these efforts will contribute to expanding the reach of targeted immunotherapy in cervical cancer, especially in contexts where antigen heterogeneity or shedding limit current treatment options.

## 6 Conclusion

This thesis presents the successful design and initial validation of two modular SNAP-tag-based antibody constructs: a bispecific  $\alpha$ EGFR (scFv) - SNAP -  $\alpha$ MSLN (scFv) format to broaden therapeutic reach in MSLN-low cervical cancers, and an anti-shedding  $\alpha$ MSLN (m15B6) (scFv) - SNAP construct targeting membrane-proximal epitopes. The constructs were cloned, expressed, and are ready for purification and downstream cytotoxicity testing.

In parallel, bioinformatic analysis confirmed the presence of an MSLN-low tumor subtype and identified EGFR as a promising co-target. These findings support the central hypothesis that bispecific targeting can enhance selectivity and overcome antigen-related resistance mechanisms. Beyond target selection, the integrative analysis revealed additional differentially expressed proteins and enriched signaling pathways, such as Wnt/ $\beta$ -catenin, that may underlie tumor aggressiveness or therapeutic response. Together, these insights underscore the broader value of bioinformatic stratification for rational therapeutic design.

The continuation of this work—including functional validation of both constructs ,with auristatin F firstly, is now in the hands of two students I trained, Samantha Levetan and Emma Jakubicka, who will carry out protein purification and in vitro assays. Together, these efforts lay a foundation for the development of more adaptable, affordable ADCs for heterogeneous cervical cancers.

# 7 Appendix

## 7.1 Full DNA Sequence of m15B6 Construct

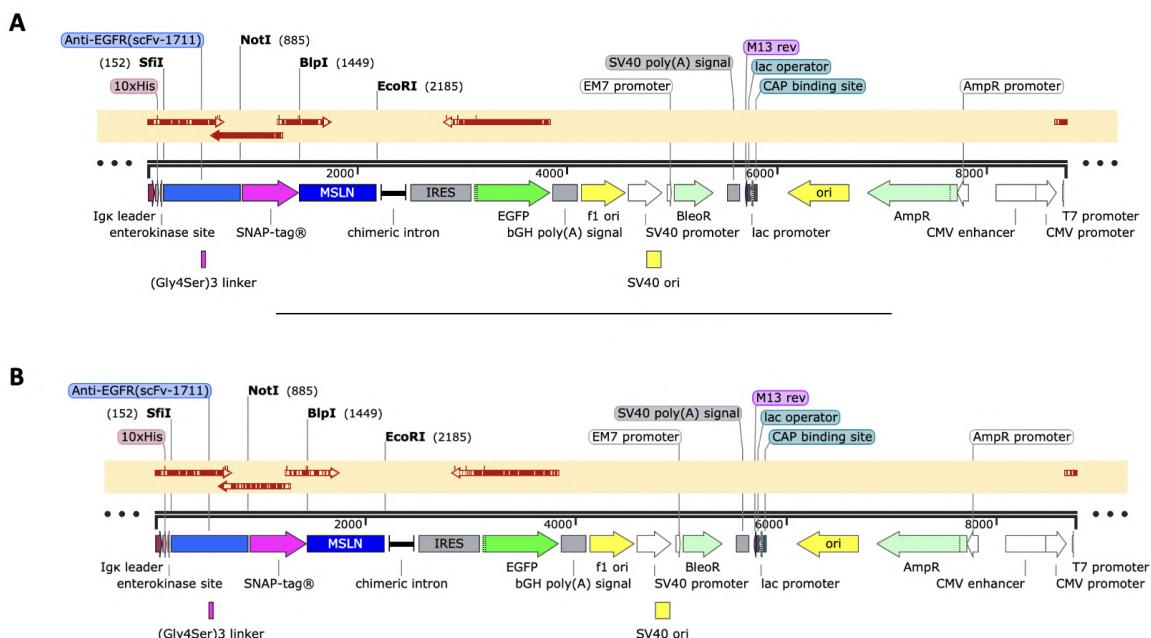
The sequence below represents the full VL-linker-VH construct of m15B6, flanked by SfiI and NotI restriction sites. Color coding: **SfiI**, **VL**, **black** (linker), **VH**,**NotI**

```

GGCCCAGCCGGCC CAGGCCGTGGT GACCCAGGAGAGCGCCCTGACCACCAGCCCCGGCAGACCGTGACC
CTGACCTGCAGGAGCAGCACCGGCCGTGACCACCGCAACTACCCCAACTGGGTGCAGGAGAACGCCG
ACCACCTGTTCACCGGCCGTATGCCGGCACCAACAAACAGGGCCCCGGCGTGCCTGCCAGGTTCAGCGG
CAGCCTGATCGGCACAAGGCCCTGACCATCACCGCGCCAGACCGAGGACGAGGCCATCTACTTC
TGCGCCCTGTGGTTAGCAGCCACTGGGTGTTGGCGCCAGCAAGCTGACCGTGCTGGCGGCG
GCAGCGGGCGGGCAGCG CCGGGCGGCAGCG AGGTGCAGCTGCAGCAGAGCGGCCGTGCTGGT
GAAGCCCGCGCAGCGTGAAGATCAGCTGCAAGGCCAGCGGCTACAGCTTACCGGCTACTACATGCAC
TGGGTGAGGCAGAGCCTGGTGAAGAGGCTGGAGTGGATGGCAGGATCAACCCCTACACCGCGTGCCTA
GCTACAAGCACAACCAAGGACAAGGCCAGCCTGACCGTGGAAGAGCAGCAGCACCGCCTACATGGA
GCTGCACAGCCTGACCGAGCGAGGACAGCGCCGTGTACTACTGCGCAGGGAGCTGGCGGCTACTGGG
CAGGGCACCAACCTGACCGTGAGCAGCAGCGCCGC

```

## 7.2 Sequencing



**Figure 1A: Sanger sequencing alignment of pCB- $\alpha$ EGFR (scFv) -SNAP- $\alpha$ MSLN (scFv) to the reference construct. (A) Alignment after manual review and replacement of ambiguous IUPAC codes. (B) Original automated alignment prior to correction. All four sequencing primers were used to ensure coverage across the construct.**

**Table 1A: Summary of alignment results per sequencing primer, before and after manual correction.**

Primer (5'-3')	Aligned Range	Length (bp)	Mismatches Before	Mismatches After	Gaps Before	Gaps After
CMV-for	14–802	808	16	13	6	6
SNAP-R	8–516	517	32	6	7	7
S-Tag	19–685	685	28	3	0	0
BGH-rev	18–1011	1021	29	4	6	6

# References

- [1] World Health Organization. Cervical cancer fact sheet, 2022. Accessed: 25-03-2025, Available at: <https://www.who.int/news-room/fact-sheets/detail/cervical-cancer>.
- [2] International Agency for Research on Cancer. South africa cervical cancer statistics - globocan 2020, 2020. Accessed: 25-03-2025, Available at: <https://gco.iarc.who.int/media/globocan/factsheets/populations/710-south-africa-fact-sheet.pdf>.
- [3] Peter A. Cohen, Anuja Jhingran, Ana Oaknin, and Lynette Denny. Cervical cancer. *The Lancet*, 393(10167):169–182, 2019. [https://doi.org/10.1016/S0140-6736\(18\)32470-X](https://doi.org/10.1016/S0140-6736(18)32470-X).
- [4] Lama Alzamil, Konstantina Nikolakopoulou, and Margherita Y. Turco. Organoid systems to study the human female reproductive tract and pregnancy. *Cell Death & Differentiation*, 28:35–51, 2021. <https://doi.org/10.1038/s41418-020-0565-5>.
- [5] Mengyun Wang, Kejia Huang, Man Chi Sin Wong, Li Zhang, Yang Yang, Rongshou Zheng, Freddie Bray, Wanqing Chen, and Weiqi Chen. Global cervical cancer incidence by histological subtype and implications for screening methods. *Journal of Epidemiology and Global Health*, 14:94–101, 2024. <https://doi.org/10.1007/s44197-023-00172-7>.
- [6] A. Muntyanu, V. Nechaev, E. Pastukhova, J. Logan, E. Rahme, E. Netchiporouk, and I. Litvinov. Risk factors and communities disproportionately affected by cervical cancer in the russian federation: a national population-based study. *The Lancet Regional Health - Europe*, 20:100454, 2022. <https://doi.org/10.1016/j.lanepe.2022.100454>.
- [7] M. Kang, S. Ha, H. Cho, D. Chung, N. Kim, J. An, and J. Jeong. Comparison of papanicolaou smear and human papillomavirus (hpv) test as cervical screening tools: can we rely on hpv test alone as a screening method? an 11-year retrospective experience at a single institution. *Journal of Pathology and Translational Medicine*, 54(1):112–118, 2020. <https://doi.org/10.4132/jptm.2019.11.29>.
- [8] Kimlin T. Ashing-Giwa, Jiyeon W. Lim, Jiali Tang, and Sheetal Surya. Surviving cervical cancer: Does health-related quality of life influence survival? *Gynecologic Oncology*, 118(1):35–42, 2010. <https://doi.org/10.1016/j.ygyno.2010.02.027>.
- [9] Lydie Lorin, Ousmane Boussari, Henri Marret, Jean Leveque, Catherine Bergeron, Sylvie Casteleyn, and Xavier Fritel. About invasive cervical cancer: A french population based study between 1998 and 2010. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 191:1–6, 2015. <https://doi.org/10.1016/j.ejogrb.2015.04.007>.
- [10] Q. Xiong, T. Zhang, and S. Su. Efficacy of the inducing chemotherapy combined with synchronized chemotherapy and radiotherapy on the local advanced cervical cancer. *International Journal of Radiation Research*, 19(4):1001–1007, 2021. <https://doi.org/10.52547/ijrr.19.4.29>.
- [11] M. Kusakabe, A. Taguchi, K. Sone, M. Mori, and Y. Osuga. Carcinogenesis and management of human papillomavirus-associated cervical cancer. *International Journal of Clinical Oncology*, 28(8):965–974, 2023. <https://doi.org/10.1007/s10147-023-02337-7>.

- [12] T. Asmiati. Human papilloma virus type 16 & 18 gene identification in plasma of cervical squamous cell carcinoma and adenocarcinoma patients at haji adam malik general hospital medan. *International Journal of Chemtech Research*, 11(11):379–385, 2018. <https://doi.org/10.20902/ijctr.2018.111143>.
- [13] L. Su, H. Peng, and X. He. Detection of hpv infection and the current status of vaccine research: a review. *Clinical and Experimental Obstetrics & Gynecology*, 50(3), 2023. <https://doi.org/10.31083/j.ceog5003048>.
- [14] L. Ntuli, A. Mtshali, G. Mzobe, L. Liebenberg, and S. Ngcapu. Role of immunity and vaginal microbiome in clearance and persistence of human papillomavirus infection. *Frontiers in Cellular and Infection Microbiology*, 12, 2022. <https://doi.org/10.3389/fcimb.2022.927131>.
- [15] E. Broni, C. Ashley, M. Velazquez, P. Sakyi, S. Kwofie, and W. Miller. Structure-based discovery of potential hpv e6 and ebnal1 inhibitors: Implications for cervical cancer treatment. *Computation*, 12(6):112, 2024. <https://doi.org/10.3390/computation12060112>.
- [16] J. Zhang. Progress in the research on the relationship between infection with human papillomavirus and tumorigenesis. *Infection International*, 5(4):99–104, 2016. <https://doi.org/10.1515/ii-2017-0138>.
- [17] Q. Fan, T. Huang, X. Sun, Y. Wang, J. Wang, Y. Liu, and Y. Wang. Hpv-16/18 e6-induced apobec3b expression associates with proliferation of cervical cancer cells and hypomethylation of cyclin d1. *Molecular Carcinogenesis*, 60(5):313–330, 2021. <https://doi.org/10.1002/mc.23292>.
- [18] M. Kaczmarek, M. Baj-Krzyworzeka, L. Bogucki, and M. Dutsch-Wicherek. Hpv-related cervical cancer and extracellular vesicles. *Diagnostics*, 12(11):2584, 2022. <https://doi.org/10.3390/diagnostics12112584>.
- [19] M. Silva, B. Albuquerque, T. Allyrio, V. Almeida, R. Cobucci, F. Bezerra, and T. Fernandes. The role of hpv-induced epigenetic changes in cervical carcinogenesis (review). *Biomedical Reports*, 15(1), 2021. <https://doi.org/10.3892/br.2021.1436>.
- [20] Ezichi F. Ofoezie et al. Role of infectious agents in cancer pathogenesis and therapy. *The Microbe*, 6:100284, March 2025. <https://doi.org/10.1016/j.microb.2025.100284>.
- [21] R. L. Coleman, D. Lorusso, S. Gennatas, D. Aoki, A. Oaknin, E. Guardino, and N. et al. Colombo. Efficacy and safety of tisotumab vedotin in previously treated recurrent or metastatic cervical cancer (innovativ 204/gog-3023/engot-cx6): a single-arm, multicentre, phase 2 study. *The Lancet Oncology*, 22(5):609–619, 2021. [https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045\(21\)00056-5/fulltext](https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045(21)00056-5/fulltext).
- [22] D. Cibula, R. Pötter, F. Planchamp, E. Åvall Lundqvist, D. Fischerová, C. Meder, and M. Raspollini. The esgo/estro/esp guidelines for the management of patients with cervical cancer. *International Journal of Gynecological Cancer*, 28(4):641–655, 2018. <https://doi.org/10.1097/igc.0000000000001216>.
- [23] S. Lu, S. Wang, R. Kim, J. Zhang, X. Wang, W. Wang, and J. Lang. Clinical outcomes of conventional hdr intracavitary brachytherapy combined with complementary applicator-guided imrt boost in bulky cervical tumours. *BJOG: An International Journal of Obstetrics & Gynaecology*, 130(2):231–237, 2022. <https://doi.org/10.1111/1471-0528.17340>.
- [24] A. Beddoe. Elimination of cervical cancer: challenges for developing countries. *Ecancermedicalscience*, 13, 2019. <https://doi.org/10.3332/ecancer.2019.975>.
- [25] Y. Irwanto, A. Wulandari, I. Permatasari, A. Ganisia, L. Zuhriyah, I. Indrawan, and S. Sutrisno. Survival and clinicopathological features of cervical carcinoma patients receiving radiation therapy. *Medical Laboratory Technology Journal*, 8(1):84–93, 2022. <https://doi.org/10.31964/mltj.v0i0.452>.

- [26] J. Okojie, S. McCollum, and J. Barrott. The future of antibody drug conjugation by comparing various methods of site-specific conjugation. *Discovery Medicine*, 35(179):921, 2023. <https://pubmed.ncbi.nlm.nih.gov/38058056/>
- [27] Y. Qiu, Y. Shi, Z. Chao, X. Zhu, Y. Chen, and L. Lu. Recent advances of antibody–drug conjugates in treating breast cancer with different her2 status. *Therapeutic Advances in Medical Oncology*, 17, 2025. <https://doi.org/10.1177/17588359241311379>.
- [28] J. Katrini, L. Boldrini, C. Santoro, C. Valenza, D. Trapani, and G. Curigliano. Biomarkers for antibody–drug conjugates in solid tumors. *Molecular Cancer Therapeutics*, 23(4):436–446, 2024. <https://doi.org/10.1158/1535-7163.MCT-23-0482>
- [29] J. Hurwitz, L. Haggstrom, and E. Lim. Antibody–drug conjugates: Ushering in a new era of cancer therapy. *Pharmaceutics*, 15(8):2017, 2023. <https://doi.org/10.3390/pharmaceutics15082017>
- [30] L. Muppa, S. Ravindranathan, S. Sivaram, A. Jinnah, and A. Shibu. Comprehensive overview on antibody drug conjugates – a promising approach in cancer therapy. *Journal of Drug Delivery and Therapeutics*, 15(3):125–132, 2025. <https://doi.org/10.22270/jddt.v15i3.7011>
- [31] Antje Keppler, Stefan Gendreizig, Tim Gronemeyer, Heiko Pick, Horst Vogel, and Kai Johnsson. A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nature Biotechnology*, 21(1):86–89, January 2003. <https://doi.org/10.1038/nbt765>
- [32] Arnaud Gautier, Alexandre Juillerat, Christian Heinis, Ivan R. Corrêa, Markus Kindermann, François Beaufils, and Kai Johnsson. An engineered protein tag for multiprotein labeling in living cells. *Chemistry & Biology*, 15(2):128–136, February 2008. <https://doi.org/10.1016/j.chembiol.2008.01.007>
- [33] A. F. Hussain, M. A. Leitenstorfer, A. Schwarz, J. T. Scheicher, I. Neundorf, V. Albrecht, and S. Barth. One-step site-specific conjugation of cytotoxic drugs to therapeutic antibodies for cancer therapy: application of snap-tag technology. *Frontiers in Pharmacology*, 10:1244, 2019. <https://doi.org/10.3389/fphar.2019.01244>
- [34] C. Huysamen, J. T. Scheicher, A. F. Hussain, I. Neundorf, and S. Barth. Development of a novel snap-tag based antibody-drug conjugate for egfr-positive cancers. *Biomolecules*, 13(4):644, 2023. <https://doi.org/10.3390/biom13040644>
- [35] Jiang Lv and Peng Li. Mesothelin as a biomarker for targeted therapy. *Biomarker Research*, 7, 2019. <https://doi.org/10.1186/s40364-019-0169-8>
- [36] Raffit Hassan, Alvaro Thomas, Christine Alewine, Dao Le, Robert J. Kreitman, and Ira Pastan. Mesothelin immunotherapy for cancer: Ready for prime time? *Journal of Clinical Oncology*, 34(34):4171–4179, 2016. <https://doi.org/10.1200/JCO.2016.68.3672>
- [37] S. Takamizawa, S. Yazaki, Y. Kojima, and et al. High mesothelin expression is correlated with non-squamous cell histology and poor survival in cervical cancer: a retrospective study. *BMC Cancer*, 22:1215, 2022. <https://doi.org/10.1186/s12885-022-10277-0>
- [38] Iva Kutle, Ricarda Polten, Jens L. Stalp, Jonas Hachenberg, Fiona Todzey, Ralf Hass, Katja Zimmermann, Julius von der Ohe, Christian von Kaisenberg, Lea Neubert, Jens-Christian Kamp, Dirk Schaudien, Anna-Katharina Seyda, Peter Hillemanns, Ralf Klapdor, Matthias A. Morgan, and Axel Schambach. Anti-mesothelin car-nk cells as a novel targeted therapy against cervical cancer. *Frontiers in Immunology*, 15:1485461, 2024. <https://doi.org/10.3389/fimmu.2024.1485461>
- [39] Joshua R. Faust, Darcy Hamill, Edward Anders Kolb, Anilkumar Gopalakrishnapillai, and Sonali P. Barwe. Mesothelin: An immunotherapeutic target beyond solid tumors. *Cancers*, 14(6):1550, 2022. <https://doi.org/10.3390/cancers14061550>

- [40] Xiufen Liu, Aileen Chan, Chien-Hsing Tai, and et al. Multiple proteases are involved in mesothelin shedding by cancer cells. *Communications Biology*, 3:728, 2020. <https://doi.org/10.1038/s42003-020-01464-5>.
- [41] Xiufen Liu, Masanori Onda, Johannes Schloemer, and et al. Tumor resistance to anti-mesothelin car-t cells caused by binding to shed mesothelin is overcome by targeting a juxtamembrane epitope. *Proceedings of the National Academy of Sciences (PNAS)*, 121(4):e2317283121, 2024. <https://doi.org/10.1073/pnas.2317283121>.
- [42] Anirban Chakraborty, Masanori Onda, Tara O'Shea, Junxia Wei, Xiufen Liu, Tapan K. Bera, and Ira Pastan. A bispecific antibody that targets the membrane-proximal region of mesothelin and retains high anticancer activity in the presence of shed mesothelin. *Molecular Cancer Therapeutics*, 23(7):1021–1030, 2024. <https://doi.org/10.1158/1535-7163.MCT-23-0233>.
- [43] A. M. Kersemaekers, G. J. Fleuren, G. G. Kenter, A. Gorter, J. B. Trimbos, and M. J. van de Vijver. Oncogene alterations in carcinomas of the uterine cervix: overexpression of the epidermal growth factor receptor is associated with poor prognosis. *Clinical Cancer Research*, 5(3):577–586, 1999. <https://pubmed.ncbi.nlm.nih.gov/10100709/>.
- [44] Ferdinando Cerciello, Paolo Chieffi, Vittorio R. Iaffaioli, Giovanni De Rosa, Luigi Montella, Licia Del Vecchio, and Carlo Vecchione. Is egfr a moving target during radiotherapy of carcinoma of the uterine cervix? *Gynecologic Oncology*, 106(2):394–399, 2007. <https://doi.org/10.1016/j.ygyno.2007.04.019>.
- [45] H. Mohammed, H. Atwa, A. A. Hegazy, A. Elwan, N. Nawar, M. A. A. Salem, and I. Khalil. Prognostic significance and therapeutic implications of pten, egfr and mapk/erk in cervical squamous cell carcinoma. *American Journal of Research in Medical Sciences*, page 1, 2018. [https://www.researchgate.net/publication/322216899\\_Prognostic\\_significance\\_and\\_therapeutic\\_implications\\_of\\_PTEN\\_EGFR\\_and\\_MAPK/ERK\\_in\\_cervical\\_squamous\\_cell\\_carcinoma](https://www.researchgate.net/publication/322216899_Prognostic_significance_and_therapeutic_implications_of_PTEN_EGFR_and_MAPK/ERK_in_cervical_squamous_cell_carcinoma).
- [46] X. Gao, J. Kang, X. Li, C. Chen, and D. Luo. Deletion of the tyrosine phosphatase shp2 in cervical cancer cells promotes reprogramming of glutamine metabolism. *The FASEB Journal*, 37(4), 2023. <https://doi.org/10.1096/fj.202202078rr>.
- [47] Y. Lv, W. Cang, Q. Li, X. Liao, M. Zhan, H. Deng, S. Li, W. Jin, Z. Pang, X. Qiu, K. Zhao, G. Chen, L. Qiu, and L. Huang. Erlotinib overcomes paclitaxel-resistant cancer stem cells by blocking the egfr-creb/gr-il-6 axis in muc1-positive cervical cancer. *Oncogenesis*, 8(12), 2019. <https://doi.org/10.1038/s41389-019-0179-2>.
- [48] Q. Chen, Y. Huang, L. Shao, H. Han-Zhang, F. Yang, Y. Wang, J. Liu, and J. Gan. An egfr-amplified cervical squamous cell carcinoma patient with pulmonary metastasis benefits from afatinib: a case report. *Oncotargets and Therapy*, 13:1845–1849, 2020. <https://doi.org/10.2147/ott.s236382>.
- [49] B. Shor, J. Kahler, M. Dougher, J. Xu, M. Mack, E. Rosfjord, and P. Sapra. Enhanced antitumor activity of an anti-5t4 antibody-drug conjugate in combination with pi3k/mtor inhibitors or taxanes. *Clinical Cancer Research*, 22(2):383–394, 2016. <https://doi.org/10.1158/1078-0432.CCR-15-1166>.
- [50] Kristina M. Ilieva, Judit Fazekas-Singer, Dariya Achkova, Toma Dodev, Salvatore Mele, Silvia Crescioli, James F. Spicer, Sophia N. Karagiannis, and Panagiotis Karagiannis. Chondroitin sulfate proteoglycan 4 and its potential as an antibody immunotherapy target across different tumor types. *Frontiers in Immunology*, 8:1911, 2018. <https://doi.org/10.3389/fimmu.2017.01911>.
- [51] Jianbo Yang, Shuyuan He, Xiaoliang Xiong, Ziyang Zhou, Yuchun Li, Bin Tan, Zhiqiang Xu, and Xiaolin Chen. Chondroitin sulfate proteoglycan 4, a targetable oncoantigen that promotes ovarian cancer growth, invasion, cisplatin resistance and spheroid formation. *Translational Oncology*, 16:101318, 2022. <https://doi.org/10.1016/j.tranon.2021.101318>.

- [52] S. Jordaan, S. Chetty, N. Mungra, I. Koopmans, P. van Bommel, W. Helfrich, and S. Barth. Cspg4: A target for selective delivery of human cytolytic fusion proteins and trail. *Biomedicines*, 5(3):37, 2017. <https://doi.org/10.3390/biomedicines5030037>.
- [53] N. Mungra, F.A. Nsole Biteghe, S. Jordaan, A. Huysamen, K. Naran, S. Chetty, E. Padayachee, S. Cooper, D. Lang, R. Bunjun, D. Govender, W. Richter, R. Hunter, and S. Barth. Cspg4 as a target for specific killing of triple-negative breast cancer cells by a snap tag-based antibody fusion protein conjugated to auristatin f. *International Journal of Cancer*, 2023. <https://doi.org/10.1007/s00432-023-05031-3>.
- [54] M. Woitok, D. Klose, S. Di Fiore, W. Richter, R. Mladenov, S. Krämer, D. Schumacher, U. Rothbauer, K. Schneider, S. Becker, and R. Kontermann. An egfr-specific antibody–drug conjugate based on snap-tag technology eradicates egfr+ solid tumors in mice. *Oncotarget*, 8(26):42571–42581, 2017. <https://doi.org/10.1016/j.canlet.2016.08.003>.
- [55] R. L. Smith, D. L. Traul, J. Schaack, G. H. Clayton, and K. J. Staley. Characterization of promoter function and cell-type-specific expression from viral vectors in the nervous system. *Journal of Virology*, 74:11254–11261, 2000. <https://doi.org/10.1128/JVI.74.23.11254-11261.2000>.
- [56] W. Arafat, J. Gómez-Navarro, J. Xiang, G. P. Siegal, R. D. Alvarez, and D. T. Curiel. Antineoplastic effect of anti-erbB-2 intrabody is not correlated with scfv affinity for its target. *Cancer Gene Therapy*, 7:1250–1256, 2000. <https://doi.org/10.1038/sj.cgt.7700228>.
- [57] A. K. Mohanty and M. C. Wiener. Membrane protein expression and production: Effects of polyhistidine tag length and position. *Protein Expression and Purification*, 33:311–325, 2004. <https://doi.org/10.1016/j.pep.2003.10.010>.
- [58] S. Hermening, S. Kügler, M. Bähr, and S. Isenmann. Increased protein expression from adenoviral shuttle plasmids and vectors by insertion of a small chimeric intron sequence. *Journal of Virological Methods*, 122:73–77, 2004. <https://doi.org/10.1016/j.jviromet.2004.08.005>.
- [59] J. C. Schmit, J. Cogniaux, P. Hermans, C. Van Vaeck, S. Sprecher, B. Van Remoortel, M. Witvrouw, J. Balzarini, J. Desmyter, E. De Clercq, and A. M. Vandamme. Multiple drug resistance to nucleoside analogues and nonnucleoside reverse transcriptase inhibitors in an efficiently replicating human immunodeficiency virus type 1 patient strain. *Journal of Infectious Diseases*, 174:962–968, 1996. <https://doi.org/10.1093/infdis/174.5.962>.
- [60] S. Busby and R. H. Ebright. Transcription activation by catabolite activator protein (cap). *Journal of Molecular Biology*, 293:199–213, 1999. <https://doi.org/10.1006/jmbi.1999.3161>.
- [61] Suyanee Thongchot et al. Mesothelin- and nucleolin-specific t cells from combined short peptides effectively kill triple-negative breast cancer cells. *BMC Medicine*, 22(1):400, 2024. <https://doi.org/10.1186/s12916-024-03625-3>.
- [62] Zilu Li, Zhen Wang, and Haitao Ji. Direct targeting of -catenin in the wnt signaling pathway: Current progress and perspectives. *Medicinal Research Reviews*, 41:2109 – 2129, 2021. <https://doi.org/10.1002/med.21787>.
- [63] Ya Zhang and Xin Wang. Targeting the wnt/-catenin signaling pathway in cancer. *Journal of Hematology Oncology*, 13, 2020. <https://doi.org/10.1186/s13045-020-00990-3>.
- [64] Maxim V Kuleshov, Matthew R Jones, Andrew D Rouillard, Nicolas F Fernandez, Qiaonan Duan, Zichen Wang, Simon Koplev, Samantha L Jenkins, Kathleen M Jagodnik, Alexander Lachmann, Michael G McDermott, Caroline D Monteiro, Gregory W Gundersen, and Avi Ma'ayan. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research*, 44(W1):W90–W97, 2016. <https://maayanlab.cloud/Enrichr/enrich>.
- [65] NCBI IgBlast, National Center for Biotechnology Information, Bethesda (MD). Accessed: 06-05-2025, Available at: <https://www.ncbi.nlm.nih.gov/igblast/>

- [66] Isabella Colombo and Berenice Martinez-Cannon. The evolving role of immune checkpoint inhibitors in cervical and endometrial cancer. *Cancer Drug Resistance*, 7, 2024. <https://doi.org/10.20517/cdr.2023.120>
- [67] Joshua Z. Drago, Sarah Modi, David B. Page, Sarat Chandarlapaty, José Baselga, Charles M. Rudin, Benjamin L. O’Neil, Kevin Kalinsky, and Sharu Modi. Unlocking the potential of antibody-drug conjugates for cancer therapy. *Nature Reviews Clinical Oncology*, 18(6):327–344, 2021. <https://doi.org/10.1038/s41571-021-00470-8>
- [68] Yosef Yarden, Ilaria Marrocco, and Maria L. Uribe. Egfr in cancer: Signaling mechanisms, drugs, and acquired resistance. *Cancers*, 13(11), 2021. <https://doi.org/10.3390/cancers13112748>
- [69] Eva García-Ruiz and Jesús Laborda-Mansilla. Advancements in golden gate cloning: A comprehensive review. *Methods in Molecular Biology*, 2850:481–500, 2025. [https://doi.org/10.1007/978-1-0716-4220-7\\_27](https://doi.org/10.1007/978-1-0716-4220-7_27)
- [70] Alyssa Sargon, Peter Harms, Adrian Nava, Joni Tsukuda, Gavin C. Barnard, Kavya Ganapathy, S. Misaghi, Cynthia Lam, and Amy Shen. Speed-mode cell line development (cld): Reducing chinese hamster ovary (cho) cld timelines via earlier suspension adaptation and maximizing time spent in the exponential growth phase. *Biotechnology Progress*, 40, 2024. <https://doi.org/10.1002/btpr.3479>
- [71] Hannah Marienberg, N. von den Eichen, V. Sidarava, D. Weuster-Botz, and L. Bromig. Automated multi-scale cascade of parallel stirred-tank bioreactors for fast protein expression studies. *Journal of biotechnology*, 2021. <https://doi.org/10.1016/j.jbiotec.2021.03.021>
- [72] Elisa Frullanti, Antonio Giordano, Giulia Brunelli, Diletta Rosati, Andrea Morrione, Maria Palmieri, and Francesco Iannelli. Differential gene expression analysis pipelines and bioinformatic tools for the identification of specific biomarkers: A review. *Computational and Structural Biotechnology Journal*, 23:1154–1168, 2024. <https://doi.org/10.1016/j.csbj.2024.02.018>
- [73] R. Kurzrock, P. Jost, A. Wahida, L. Buschhorn, S. Fröhling, A. Schneeweiss, and P. Lichter. The coming decade in precision oncology: six riddles. *Nature Reviews Cancer*, 23:43–54, 2022. <https://doi.org/10.1038/s41568-022-00529-3>
- [74] R. Kurzrock and V. Subbiah. Challenging standard-of-care paradigms in the precision oncology era. *Trends in Cancer*, 4(2):101–109, 2018. <https://doi.org/10.1016/j.trecan.2017.12.004>
- [75] Leticia Rocha-Zavaleta, Arturo Valle-Mendiola, I. Soto-Cruz, Vilma Maldonado-Lagunas, Diego Morelos-Laguna, Benny Weiss-Steider, and Adriana Gutiérrez-Hoya. Stat5 is necessary for the metabolic switch induced by il-2 in cervical cancer cell line siha. *International Journal of Molecular Sciences*, 25(13):6835, 2024. <https://doi.org/10.3390/ijms25136835>
- [76] Clarissa Esmeralda Halim, C. Yap, Shuo Deng, and M. S. Ong. Involvement of stat5 in oncogenesis. *Biomedicines*, 8(9):316, 2020. <https://doi.org/10.3390/biomedicines8090316>
- [77] Rosa A. Sierra, D. Gilvary, J. Trillo-Tinoco, P. Rodriguez, B. Ruffell, Yu Cao, E. Flores, Álvaro de Mingo-Pulido, C. Anadon, Sheng Wei, T. Costich, J. Conejo-Garcia, and Eslam Mohamed. Ampk alpha-1 intrinsically regulates the function and differentiation of tumor myeloid-derived suppressor cells. *Cancer Research*, 2019. <https://doi.org/10.1158/0008-5472.CAN-19-0880>
- [78] C. Vicentini, A. Scarpa, A. Mafficini, G. Tortora, M. Milella, E. Bria, M. Simbolo, S. Pilotto, A. Gkountakos, and V. Corbo. Unmasking the impact of rictor in cancer: novel insights of mtorc2 complex. *Carcinogenesis*, 39:971–980, 2018. <https://doi.org/10.1093/carcin/bgj086>
- [79] Yan Huang, Yuzhen Zhou, and C. Shu. Fibronectin promotes cervical cancer tumorigenesis through activating fak signaling pathway. *Journal of Cellular Biochemistry*, 120:10988–10997, 2019. <https://doi.org/10.1002/jcb.28282>