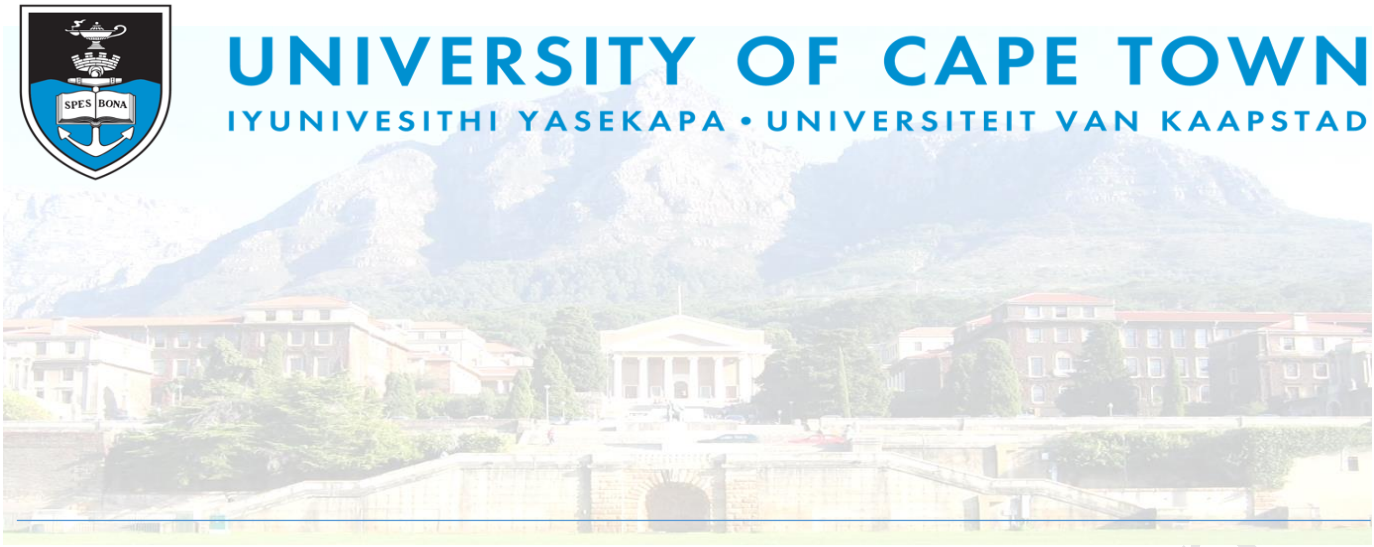


Evaluation of tumour-associated antigens to optically label cutaneous basal cell carcinoma for surgical excision

Suresh Madheswaran

Supervisor: Prof. Dr. Dr. Stefan Barth

SUBMITTED TO THE UNIVERSITY OF CAPE TOWN

In fulfilment of the requirements for the degree of

Doctor of Philosophy in Chemical Biology

12 Feburary 2022

DECLARATION

I declare this thesis is my unaided work. It has not been submitted for any degree or at any university other than for this Doctor of Philosophy at the University of Cape Town.

I have used numbered American Psychological Association (APA) convention for citation and referencing. All significant contributions to the thesis, and quoted content, if any, from the work or works of other authors and collaborators, have been duly attributed to, cited, and referenced.

Suresh Madheswaran

Signed by candidate

Signature

Date: 12 Feburary 2022

ACKNOWLEDGMENTS

First and foremost, I am truly appreciative of Prof. Dr. Dr. Stefan Barth, for having helped me engage with my cancer research journey and for having helped me bring out the best in me—and above all for his unwavering faith in me and for his utmost patience. You're a source of constant support and inspiration. I can't thank you enough.

I wish to thank the Cancer Association of South Africa (CANSA), the South African Research Chair in Cancer Biotechnology, and the University of Cape Town (UCT) whose generous funding made this study possible.

I thank Dr. Dharanidharan Ramamurthy for his guidance in the lab and for his timely thoughts and suggestions on the manuscript of this thesis. Your assiduousness and humility would never fail to astound me.

My thanks are due to the following members of MB&I group for their support and camaraderie: Mrs. Nyameka Mhlonyelwa-Mona, Dr. Eden Padayachee, Dr. Shivan Chetty, Dr. Krupa Naran, Ms. Neelakshi Mungra, Ms Trishana Nundalall, Dr. Alex Akinrinmade, Mrs. Adebukola Daramola, Mrs. Precious Hlongwane, Ms. Natasha Hardcastle, Dr. Sandra Jordaan, and Dr. Fleury Augustin Nsole Biteghe. I am so grateful to have all of you in my life—thank you!

I thank Prof. Roger Hunter for providing us with BG-modified substrates.

I thank Prof. Edward Sturrock, Sylva Schwager, Mrs. Susan Cooper, Mr. Tim Reid (University of Cape Town)—thank you for infrastructural and technical support, as well as for, in your own ways, encouraging me to embark on my research journey. I am truly blessed to have shared some of my life's adventures with each of you as a colleague as much as a friend.

Special thanks go to Dr. Kevin Dozobo, Dr. Arul Jothi, Mrs. Arul Jothi, Dr. Sudesh Sivarasu, and Dr. Rajesh Sharma for helping me keep my head above water on several occasions.

I am immensely grateful for the unconditional love and support of my family—my mother, Mrs. Meena, and father, Mr. Madheswaran, and my brother Mr. Arun Prasad.

I thank my best friend Mrs. Prema Lenin and lovable nephew A.N. Pavithran for adding flavor and color to my life. I thank my friends Dr. Manikandan Kasi and Mr. Karthik S for their timely help with preparing the manuscript at various levels.

Above all, I'd like to express my love and appreciation for my wife, Yuvarani. Her patience, advice, and love over the years have kept me going, particularly at trying times.

Last but not least, thank you, God, for all the blessings and for giving me the strength, patience, and intelligence to always believe in myself and overcome all the obstacles life has thrown at me.

ABSTRACT

Basal cell carcinoma (BCC) is the most common skin cancer worldwide, with South Africa having the highest incidence rate only after Australia. The most effective treatment modality for BCC is tumor excision via Mohs surgery (pioneered by Dr. Frederic Mohs of the University of Wisconsin in 1930), a microscopically controlled surgery that removes a tumor piecemeal in layers until each layer is free of any neoplastic tissue. The major drawback of Mohs excision is that the surgeon might miss any neoplastic tissue as the tumor margin is not always well defined, and the tumor often could extend beyond the superficial layers of skin. Moreover, it's a time-consuming, expensive procedure that takes generally 3-4 h, at times even more, if several rounds of excisions are warranted. In South Africa, at the time of writing, therapy using the surgery cost around R45,000. The status quo thus necessitates identifying BCC cells both in the superficial layers and beyond the layers of the skin in individual patients. Our aim was to identify BCC-specific cell surface proteins and design, engineer, and test a range of SNAP-tag–based antibody fusion proteins that would specifically bind to and detect such BCC cell surface receptors. The SNAP-tag antibody technology is based on the genetic fusion of a disease-specific ligand to a protein tag derived from O6-alkylguanine-DNA alkyltransferase, which would allow for covalent auto-labeling of the corresponding antibody based fusion proteins with benzylguanine-modified (BG) substrates (e.g., fluorophores) under physiological conditions with high efficiency at 1:1 stoichiometry. This would allow to develop a unique immunological screening modality which should allow to visually label BCC lesions for a more precise surgical excision. The best-performing SNAP-tag–based diagnostic antibodies resulting from these studies would be further evaluated in the future in suitable mouse models, thus aiming to reduce the time needed for surgical removal of BCC lesions and complete removal of the tumor from both superficial and deep layers of the skin by a single-excision procedure.

We used an integrated computational tool to re-analyze publicly available cDNA microarray data in combination with theoretical search to identify BCC-associated antigens. Accordingly, six different antigens were selected and single-chain variable fragments (scFv) targeting these antigens were cloned in fusion with SNAP-tag encoding gene into a custom expression vector for production in a secretory mammalian system. scFv-SNAP-tag protein was isolated from the cell culture supernatant by immobilized metal affinity chromatography and eluted protein samples were analyzed by gel electrophoresis and immunoblotting. The absolute amount of the full-length protein was quantified by densitometry. Purified scFv-SNAP-tag proteins were

validated for specific binding to corresponding antigen-positive cells by flow cytometry and confocal microscopy.

Of the six different scFv-SNAP-tag fusion proteins cloned, four were successfully expressed in HEK293T cells. The specific binding to EpCAM, EMA, CSPG4, and CD138 antigen-expressing cell lines was observed on incubation with scFvUBS54-SNAP-tag, scFvID405-SNAP-tag, 9.2.27scFv-SNAP-tag, and scFvh-STL002-SNAP-tag, respectively. In addition, we showed the selective cell killing effect of scFvUBS54-SNAP after conjugating it with the cytotoxic drug BG-modified auristatin-F (BG-AF).

In conclusion, we identified various cell surface antigens along with one possibly novel antigen for BCC detection and therapy. Further, we successfully designed and synthesized SNAP tag based antibody fusion proteins and showed their functional activity by selective binding to the corresponding antigens on the surface of tumor cells. Based on these findings, we presume that these antibodies can effectively bind to BCC and can confirm EpCAM as one of the target antigens, which has already been reported to be a standard immunophenotypic marker for differential BCC diagnosis.

TABLE OF CONTENTS

2.1.2. Microarray data preprocessing and analysis
2.1.3. Identification of DEGs and screening for cell surface markers
2.1.4. Validation of cell surface markers using alternative computational tools (SAM and TAC) and literature search

2.2. IN SILICO CLONING OF MAMMALIAN AND BACTERIAL EXPRESSION VECTOR SYSTEM
2.2.1. Identification of antibody sequence from the public repository
2.2.2. In silico design of mammalian and bacterial expression vector systems encoding scFv-SNAP-tag fusion protein
2.2.3. Removing or altering cryptic translation initiation site within the ORF of αEpCAM (scFvUBS54-SNAP-tag)

2.3. MOLECULAR CLONING OF MAMMALIAN AND BACTERIAL EXPRESSION VECTOR
2.3.1. Transformation of *E. coli* DH5α by heat shock
2.3.2. Isolation of plasmid DNA from *E. coli* DH5α
2.3.3. Restriction DNA
2.3.4. Ligation of DNA and transformation of recombinant clone in *E. coli*
2.3.5. Restriction mapping, DNA sequencing, and analysis

2.4. EXPRESSION OF FUSION PROTEINS (scFv-SNAP-tag)
2.4.1. Cell culture
2.4.2. scFvUBS54/ID405/h-STL002-SNAP-tag expression in mammalian system
2.4.3. Optimizing transfection efficiency
2.4.3.1. DNA Plasmid transfection using polymer transfection agent.
2.4.3.2. Analysis of transfection efficiency of DNA plasmid transfection based on manufacture instruction by flow cytometry
2.4.3.3. Effects on transfection efficiency of modified protocol
2.4.3.4. Analysis of transfection efficiency of optimized protocol
2.4.4. Zeocin™ selection
2.4.4.1. Standard Procedure
2.4.4.2. Optimized Procedure
2.4.5. Sandwich ELISA

2.5. PURIFICATION AND QUANTIFICATION OF MAMMALIAN EXPRESSED ANTIBODY FUSION PROTEINS
2.5.1. IMAC purification
2.5.2. Protein quantification
2.5.2.1. UV–visible spectrophotometer
2.5.2.2. BCA assay
2.5.2.3. Bradford assay
2.5.3. Estimating the integrity and purity of protein by SDS and WB
2.5.4. Densitometry analysis to quantify protein band of interest

2.6. VALIDATING THE FUNCTIONAL ACTIVITY OF FUSION PROTEIN
2.6.1. Conjugation of scFv-SNAP-tag with BG-modified Alexa 488
2.6.2. In vitro binding study of labeled scFvUBS54/ID405/h-STL002-SNAP-tag fusion protein
2.6.2.1. Confocal microscopy

CHAPTER 1
LITERATURE REVIEW

1.1. INTRODUCTION

1.1.1. Structure of the skin

The skin, otherwise known as the integument, is the largest organ of the human body covering a surface area of approximately 1.7 m^2; it weighs about 15% of the total body weight in an average adult.[1] It acts as an inert barrier against the external environment by shielding the body from toxic chemicals, ultraviolet radiation, microorganisms, and mechanical trauma.[2] It is an immunologically active sensory and excretory organ that assists in regulating body temperature.[3,4] It consists of three layers: epidermis, dermis, and subcutaneous fat/tissue.[5] The epidermis, the outermost layer of the skin, is composed mainly of keratinocytes (95%) and dendritic cells. Other cells in the epidermal layer are non-keratinocyte cells such as pigmented cells, immune cells, and also sensory cells such as melanocytes, Langerhans cells, and Merkel cells.[1] The dermis is a connective tissue layer that lies between the epidermis and subcutaneous tissue layer, composed mainly of collagen (type I and type III collagen) and elastic fibers (elastin and fibrillin microfibrils) produced by fibroblast cells, a primary cell type in the dermis.[6] The secondary cells in the dermis layer are histiocytes, adipocytes, mast cells, and Schwann cells. In addition, the dermis also houses hair follicles, nerves, lymphatic vessels, sweat glands, sebaceous glands, and a rector pili muscles.[7] The subcutaneous tissue lies deep under the dermis, formed mainly by adipocytes, which store fat and act as a repository for energy storage, provide buoyancy and cushion deep tissues from violent trauma and insulate them from cold, and act as an endocrine organ.[8]

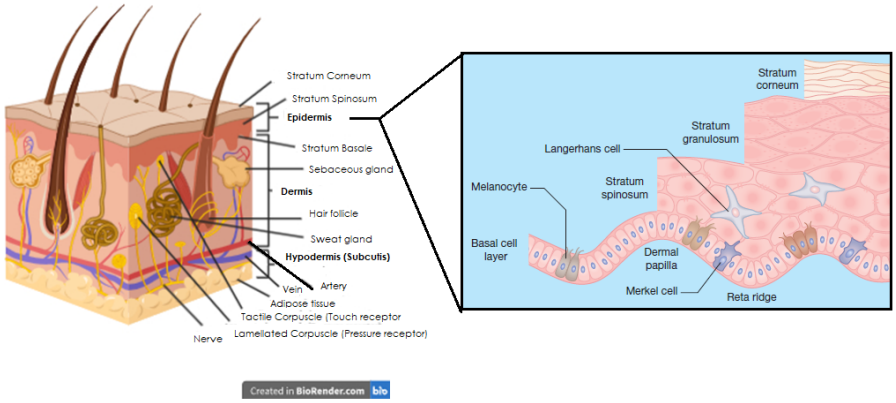


Figure 1: Anatomy of the human skin (left), created using BioRender & the figure epidermis (right) copied from Marks et al.,[9]

1.1.2. Cellular basis of skin cancer

Cutaneous tumors most commonly originate from the outermost layer of skin "epidermis" composed mainly of keratinocytes (basal keratinocytes, squamous keratinocytes) and melanocytes and minimally composed of Langerhans cells, Merkel cells, and epidermis stem cells. These cells perform various function such as keratin secretion, melanin secretion, immune surveillance, sensory function and differentiation to functional basal, spinous cells and non-proliferative stratum corneum.[1]

When these cells accumulate multiple mutations from a combination of extrinsic factors, e.g., environmental factors such as exposure to ultraviolet (UV) radiation and intrinsic factors such as genetic predisposition and phenotypic risk, they lead to disruption of cell homeostasis and eventually give rise to the most common type of skin cancer called basal cell carcinoma, squamous cell carcinoma, and melanoma, respectively. The uncommon skin cancer called Merkel cell carcinoma (MCC) arises from Merkel cell and Langerhans histiocytosis and or Langerhans cell sarcoma from Langerhans cells.[10,11,12]

Not only the cells in the epidermal layer get transformed into malignant, but cells in the dermis and subcutaneous layer such as dermal dendritic cells/dermal histiocytes and dermal fibroblasts may also give rise to dermatofibromas or cutaneous fibrous histiocytomas.[13] Whereas the adipocyte tumors are the neoplasm of the subcutaneous layer. This adipocyte tumor can be of benign (lipomas) type or malignant type (pleomorphic liposarcoma).

1.1.3. General classification and epidemiology of skin cancer

Skin cancer is broadly divided into keratinocytes tumor and melanocytic tumor, keratinocytes tumors commonly known as non-melanoma skin cancers (NMSCs). Melanocytic tumors are synonymous with the term "melanoma." This classification is based on the metastatic nature of

the tumors. Melanomas have a high risk of metastasis and nonmelanoma skin cancer; particularly squamous cell carcinomas (SCC) have a relatively low risk of metastasis. Basal cell carcinoma (BCC) has an extremely low risk of metastasis. Calling BCC and SCC as nonmelanoma skin cancer is a bit of a misnomer as other cutaneous malignancies are not melanomas, but because of their more aggressive clinical behavior, they tend not to be grouped under this moniker.[14]

Melanomas are further classified into four major subtypes based on clinical and histological features: superficial spreading melanoma, nodular melanoma, lentigo malignant, and acral lentiginous melanoma, all of which account for 5% of all skin cancers and responsible for the majority of skin cancer deaths.[15] Worldwide, melanomas account for more than 280,000 new cases and 60,000 deaths annually.

NMSCs account for the remaining 95% of skin cancer cases of which basal cell carcinoma (BCC) is the most common skin cancer, which represents between 70% and 80% of all keratinocyte carcinomas.[16] The remaining 20% to 30% are represented by cutaneous squamous cell carcinomas (cSCC).[17,18] NMSCs have a high (18–20 times greater) incidence rate (10,42,056 incident) compared to melanoma and account for more than 65,000 deaths annually.[19,20] BCC has a very low mortality rate of 0.12 per 100,000 despite being frequent and common skin cancer.[21] This reduced mortality rate is due to a reduced growth rate and metastasis rate of 0.00281–0.05% resulting from limited angiogenesis. Consequently, most cases are amenable to surgery.[21,22] The recurrence rate for incompletely and completely excised tumors is approximately 27% and 6% respectively.[23] However, despite the relatively low malignant/metastatic potential of BCC, these skin cancers are associated with the effects of remarkable deformity or morbidity and substantial cost of treatment.[24,25] The age-associated mortality rate is estimated to be 0.12 per 100,000 for BCC.[26]

Classifying BCC is challenging, as it is histopathologically diverse with more than 20 patterns reported in literature.[27] However, certain histological subtypes such as nodular, micronodular, superficial, infiltrating, and sclerosing/morphoeic tumors exhibit different clinical presentations and differ in their response to treatment.[28] The World Health Organization (WHO) classifies BCC into various subtypes based on the risk of recurrence. The low-risk subtypes include nodular, superficial pigmented, and fibroepithelial BCC as well as few other variants. More stringent histological criteria are followed for the diagnosis of high-risk subtypes such as superficially micronodular, infiltrating, and sclerosing/morphoeic BCC, as patient management is largely ranked by the risk of recurrence. Basosquamous carcinoma is a controversial entity as it shows features of both BCC and SCC. Subsequently, they present a higher risk of recurrence than other BCC variants. Despite the relatively low malignant potential of BCCs, these skin

cancers are associated with the effects of remarkable deformity or morbidity and substantial cost of treatment.[24,25]

In immunosuppressed populations, SCC is more common than BCC[29] and tends to grow and spread faster than BCC with metastatic potential ranging from 1.1% to 2.4%.[30] SCC is classified into various histological subtypes based on differentiation grade/aggressive nature, as some subtypes are harmless, and others are aggressive. This has significance in the management and prognosis of the disease. SCCs in situ (Bowen disease, actinic keratoses) remain unchanged for many years. Only 4% of SCC in situ evolve to become invasive.[31] Invasive SCCs tend to grow and spread much faster and are thus further classified as low-grade and high-grade invasive SCCs. Low-grade invasive SCCs include keratoacanthoma a well-differentiated variant and verrucous SCC. They have a lesser potential for metastasis but are locally destructive. More aggressive variants such as adenosquamous carcinoma and spindle cell SCC are poorly differentiated and have a higher tendency to spread to distant organs.[32] The disease-free survival rate is less than 42% for metastatic cSCC and 90% for localized cSCC.[33] The overall recurrence rate is 6.7% for cSCC.[34]

Categorizing the diagnosed lesions with accuracy into skin tumor subtypes is the first step to determine the choice and success of treatment. However, such classification is rarely an easy task as carcinomas of the skin may be morphologically and histologically mistaken for other skin diseases and more benign lesions. Tumor characterization is further complicated by difficulties in distinguishing the boundaries of skin tumors from surrounding healthy tissue, a factor that is one of the causes that contributes to incidences of recurrence and subsequent morbidity and mortality.

1.2. BIOLOGY OF BASAL CELL CARCINOMA

1.2.1. Histogenesis of BCC

Basal cell carcinoma (BCC) was first named *ulcus rodens* in 1827 by Arthur Jacob; in 1900 Krompecher coined the term *carcinoma epitheliale adenoides* and later the term *Basalzellenkrebs*, the latter referring to BCC's origin in stratum basale of the epidermis or stem cells of hair follicles. In contrast, in 1910, Mallory used the term "hair matrix tumor" to specify the follicular origin of BCC. However, the term that is currently in use, i.e., "basal cell carcinoma," has been prevalent only since 1974, after the WHO classification system of skin cancer was established. Interestingly, the origin of BCC has not been solved for more than a century. Multiple hypotheses have been postulated and proven for the origin of BCC by various

researchers. Wang et al. and Peterson et al. claim that BCCs arise from stem cells of the hair follicle.[35,36] Wang, Peterson, and Youssef et al. state that SmoM2 activation of hair follicle bulge stem cells did not elicit BCC, through clonal analysis they found that BCC arose from long-term resident progenitor cells located in the interfollicular epidermis or basal layer of the epidermis and upper infundibulum.[37,38] Danes et al. proved using a mouse model that the BCC originated from stem cells of the interfollicular epidermis (IFE) but not from progenitor cells by constitutive activation of oncogenic HH (hedgehog) signaling in mouse skin. Thus, the true cytologic origin of BCC remains in question and future research is warranted to address this histogenesis of BCC.[39]

1.2.2. Etiology and genetics of BCC

The exact cause of BCC is unknown, but there are various factors, such as environmental factors exposure to ultraviolet radiation (UVR), radiotherapy, viral infections human papilloma virus (HPV) and cytomegalovirus (CMV),[40,41] immunosuppression, and exposure to arsenic that are known risk factors.[42] One of the main carcinogenic factors are ultraviolet light (UV) and hereditary predisposition syndrome, which may explain why BCC is high in high UV index regions (Australia and New Zealand) and in surfaces of the body that have higher solar exposure (e.g., the face and neck). Genetic factors such as hereditary predisposition syndromes are also known to contribute to disease pathology.[43] People with familial disorder (germline mutation) such as basal cell nevus syndrome (BCNS) which comprises Gorlin syndrome, Gorlin–Goltz syndrome (GGS), and nevoid basal cell carcinoma are caused by mutations in PTCH1, PTCH2, and SUFU genes, that are expressed as proteins involved in the hedgehog signaling pathway which, in turn, controls growth and tissue development. Bazex–Dupré–Christol syndrome-(BDCS) caused due to germline defects in ARCT1 is again due to the gene's participation in hedgehog signaling pathway. Individuals who exhibit Oley syndrome and Rombo syndrome (caused by hitherto unknown genes), oculo-cutaneous albinism (TYR, OCA2, MATP/OCA4, TYRP1), xeroderma pigmentosum (XPA–XPG and XP variant) are highly prone to BCC.[44-45] People who are frequently diagnosed with BCCs may present the disease and its recurrence as an external clinical marker of inherited cancer risk.[46] Moreover, families with a history of skin cancer remained strongly associated with early-onset BCC.[47] There is also sporadic BCC due to somatic mutation in various genes involved in DNA repair (XPD and XRCC3 genes), cell growth, cell proliferation and differentiation (PTCH, SMO, SHH, Ras, MYCN, PPP6C, PTPN14, STK19, and LATS1), cell death (PTEN, p53) and chemical detoxication (cytochrome p450 (CYP) and glutathione S transferase (GST) genes such as GSTM1, GSSTT1, and

CYP2D6).[48,49,50,51] BCC is one of the most highly mutated human tumors (65 mutations/megabases)[52,53] and harbors a large percentage of UV-induced mutations (C:T or CC:TT transitions at dipyrimidine sites).[48,43,54,55]

Table 1: Genetics of BCC

Genetic skin disorder	Gene(s) involved	Function
Basal cell nevus syndrome (BCNS) • Gorlin syndrome • Gorlin–Goltz syndrome (GGS). • nevoid basal cell carcinoma	PTCH1, PTCH2, SUFU	Hedgehog signaling pathway for growth and tissue development
Bazex–Dupré–Christol syndrome-(BDCS)	ACTRT1	Hedgehog signaling pathway for growth and tissue development
Oculo-Cutaneous Albinism (OCA)	TYR, OCA2, MATP/OCA4, TYRP1	Melanin synthesis
Xeroderma pigmentosum	XPA, XPB, XPC, XPD, XPF, XPG, or POLH	DNA repair
Oley syndrome (possibly a subtype of Bazex–Dupré–Christol syndrome).	Unknown gene	NA
Rombo syndrome	Unknown gene	NA
Schöpf–Schulz–Passarge syndrome	WNT10A	Wnt signaling pathway: cell fate determination, cell migration, and organogenesis during embryonic development
Cartilage-hair hypoplasia	RMRP (noncoding RNA) forms mitochondrial RNA-processing endoribonuclease.	Replicate the DNA in mitochondria, assembling amino acids to a functional protein, control the cell cycle
Cowden syndrome	PTEN, SDHB, SDHD, and KLLN	Cell survival
Hermansky–Pudlak syndrome	HPS1 and HPS3	Protein trafficking
Muir–Torre syndrome	MLH1, MSH2, and MSH6	DNA mismatch repair DNA replication and repair
Brooke–Spiegler syndrome	CYLD	Nuclear factor-kappa-B for cell death

1.3. BCC STANDARD DIAGNOSIS AND TREATMENT

1.3.1. Diagnosis (distinction and characterization of BCC)

Skin tumors are identified, distinguished, and classified into various subtypes based on clinical or morphological features, histopathological features, and immunophenotyping for differential diagnosis.

1.3.1.1. Morphological classification

Based on its clinical presentation BCC can be divided into four major groups: nodular (with micronodular), superficial, infiltrative (morpheaform), and mixed subtype/other subtypes.[56] The classic clinical feature of BCC is pearly or flesh-colored papule with telangiectasia and may be eroded or ulcerated and occasionally accompanied by bleeding, scaling, or crusting.[57,58,59]

Nodular BCC: 80% of BCC cases are nodular subtype, simply characterized by the presence of large tumor nodule in dermis, or lesion appear in the face as pink or flesh-colored papule with a translucent quality, telangiectatic vessel, and central erosion. The term "rodent ulcer" refers to these ulcerated nodular BCCs.[60,57]

Superficial BCC: Approximately 15% of BCCs are superficial BCCs. Such lesions are more predominant in photo-protected areas of the body such as the trunk. They appear as multifocal erythematous with well-circumscribed macule. They may appear slightly scaly and erosion is less common in the center, whereas they may be rimmed in the periphery with fine translucent papules that may mimic eczema or psoriasis. Occasionally, their appearance may show spotty brown or black pigment, which could be mistaken as melanoma.[57, 60, 58]

Morpheaform/Scleorosing BCC constitutes of 5% to 10% of BCCs. The lesion manifests as a depressed scariform or fibrotic scar with ill-defined borders. In this subtype, ulceration, bleeding, and crusting are uncommon. It may be mistaken for scar tissue.[57-87]

Mixed subtype/other subtypes: Histologically such subtypes show both basal cell and squamous cell carcinoma differentiation. Some authors prefer the term basosquamous carcinoma for the mixed type,[61] a rare tumor that behaves aggressively (with higher risk of recurrence than that of any other variant). Clinically the lesion appears as a slow evolving papule or nodule, which may ulcerate.[59]

1.3.1.2. Histological classification

a) Histopathology of BCC

Histology is the microscopic study of animal and plant cells and tissues through sectioning staining and examining them under a microscope (electron or light microscope).[62] The gold standard of diagnosis for skin cancer has been skin biopsy with routine paraffin-embedded hematoxylin and eosin staining.[63]

Hematoxylin has a deep blue-purple color and stains nucleic acids by a complex, incompletely understood reaction. Eosin is pink and stains proteins nonspecifically and shows the cytoplasm and extracellular matrix in pink.[64] Cancer disrupts the normal control mechanisms of cell proliferation and differentiation. This affects the structure, appearance, and activity of the cell nuclei within the tissue, leading to cancer-types specific patterns such as condensation of heterochromatin,[64] anisokaryosis (variation in nuclear size), irregular chromatin distribution, pleomorphism (variable size and shape of cells and/or nuclei in cells of the same type), to name a few features.[65] BCC is morphologically variable, but they are histologically invariable showing islands or nests of peripherally palisaded basaloid cells with hyperchromatic nuclei and/or large elongated nuclei with scant, inconspicuous, pale, or lightly eosinophilic cytoplasm.[59] The cluster of basaloid cells is surrounded by a stromal reaction. Retraction artifact is an important microscopic feature of basal cell carcinoma, where the cells may appear to be separate from the stroma, unlike the basal layer of the skin which is firmly attached to its basement membrane.[66]

Table 2: Histological subtype of Basal cell carcinoma (BCC) ranked by risk factor

Lower risk	**Higher risk**
Nodular BCC	Basosquamous carcinoma
Superficial BCC	Sclerosing/Morphoeic BCC
Pigmented BCC	Infiltrative BCC
Infundibulocystic BCC (a variant of BCC with adnexal differentiation)	BCC with sarcomatoid differentiation
Fibroepithelial BCC	Micronodular BCC

Adopted from WHO classification of skin cancer

Nodular BCC: The nodules of tumor or islands of basaloid cells show a peripheral palisading extended into the dermis. Retraction space formed between tumor islands and stroma and amyloid deposition may be seen. Centrally, a haphazard nuclear arrangement with frequent apoptosis is seen.[59,58] Variants of nodular BCC include keratotic BCC; with central keratinization in tumor nest or within tumor nodule. Nodulocystic BCC or cystic BCC; with cystic degeneration (leaving nothing but space) within some of the tumor islands. Adenoid BCC: with cribriform nests or cribriform tumor islands.[59,58]

Superficial BCC is characterized by superficial lobules (nests) of basaloid cells that project from the epidermis or attach to the undersurface of the epidermis. Typically nests are surrounded by loose myxoid stroma, and lymphoid infiltrate may be present in opposition to the basaloid nests and under the surface of the epidermis.[58,67]

Micronodular basal cell carcinoma manifests tumor micronodules of size>0.15mm in diameter and is often shows asymmetric distribution throughout the dermis and subcutis. This might appear as separate satellites outlines by a thin rim of stroma, i.e., retraction spaces might be seen in some micronodules that are separated by a normal dermis. The peripheral edges of this tumor are irregular, tentacular, and infiltrative.[68]

Infiltrative BCC shows variable size and shape. They nest with a jagged contour that infiltrates the normal dermal collagen. The irregular nests are surrounded by abundant tumor stroma. The tumor nests are typical>5-8 cells thick with perineural invasion.[58,69]

Sclerosing/Morpheaform BCC: These tumors form small narrow cords (1-4 cells thick). Irregular (sharply angulated) tongues of neoplastic basaloid cells are embedded in the densely proplastic and heavily collagenized stroma or sclerotic stroma, which disrupt normal dermal architecture. Retraction artifact is uncommon[58]; however, single-cell pattern (2-cell clusters) is common a histological feature that is frequently seen in morpheaform BCC. Three types of the single-cell pattern were identified: intratumoral (single cells within the main tumor mass), peripheral, and distant tumor, i.e., cells seen at a significant distance from the main component are unique to morpheaform BCC.[70]

Basosquamous carcinoma (BsC) or metatypical BCC: BsC has mixed histopathological characteristics of both basal cell carcinoma and squamous cell carcinoma classical features include inflammatory islands of basaloid cells with nuclei that palisade at the periphery with atypical/malignant squamous cells with abundant cytoplasmic keratinization. A typical squamous cell can be focal or scattered throughout the lesions.[71]

Pigmented BCC: a BCC variant that shows melanin pigment deposition in a cystic basaloid nodule or tumor lobules and melanophages. Dendritic melanocytes are seen within tumor

islands.[72] Both nodular and superficial BCC may contain pigment and fall into the category of pigmented BCC.

Sarcomatoid basal cell carcinoma or BCC with sarcomatoid differentiation: BCC variant which has a nest of basaloid epithelial cell comingled with the atypical mesenchymal component, which may consist of pleiomorphic undifferentiated sarcoma, osteosarcoma, chondrosarcoma, leiomyosarcoma, rhabdomyosarcoma, and or sarcomatous stroma. Other names are carcinosarcoma, metaplastic carcinoma, sarcomatoid carcinoma, and malignant mixed tumors. All the above are terms interchangeably used to describe malignant tumors combining malignant epithelial and mesenchymal differentiation. The monoclonal theory has been postulated for the histogenesis of carcinosarcomas, which suggests that an origin could be from a single totipotent stem cell that differentiates into separate epithelial and mesenchymal directions (monoclonal theory).[73,74]

Infundibulocystic BCC: BCC variant that manifests with matrical cell differentiation of follicular, sebaceous, apocrine, or eccrine glands[75] within tumor lobules. Infundibulocystic BCC shows differentiation toward follicular infundibulum, where the nests of basaloid cells are punctured by small keratin-filled cyst.[76-77] BCC with sebaceous differentiation shows mature sebocytes that are positive for EMA.[78]

Fibroepithelial BCC (FEBCC): Strands of basaloid cells and or basaloid islands surrounded by abundant fibroblastic stroma, this tumor frequently connects to the overlying epidermis at multiple points, the other name for fibroepithelial BCC is fibroepithelioma of Pinkus or Pinkus tumor, named after Pinkus who first described them in 1953.[79-80]

Knowing these histological subtypes is important for risk assessment, example micronodular, infiltrating, sclerosing and basosquamous are associated with higher recurrence and invasion of the dermis.[81-82]

b) Histologic mimics of BCC

A wide spectrum of skin diseases including benign lesions such as trichoadenoma, dermatofibroma, trichoepithelioma, and cancerous lesions such as sebaceous carcinoma or Merkel cell carcinoma mimic basal cell carcinoma and vice versa based on morphologic and histological appearance.[83,121] Recently Singh et al., Abudu et al., and Hasbun et al. reported that pigmented basal cell carcinomas masquerade as melanomas (aggressive skin cancer)[84- 85]. The adenoid variant BCC resembles cutaneous ACC (adenoid cystic carcinoma)[86] and features of metatypical BCC overlaps with squamous cell carcinoma. Merkel cell carcinoma (MCC), a rare

aggressive primary cutaneous neuroendocrine carcinoma, is reminiscent of BCC.[87] For example, the histological features of benign tumor trichoepithelioma (TE) or trichoblastoma shows basaloid cells, peripheral palisading, mucin deposition in tumor nodules, apoptotic bodies, and connection to the epidermis. The general histological features of BCC are nests of peripherally palisaded basaloid cells, elongated nucleus, scant cytoplasm, and apoptotic bodies. This island of basaloid cells is surrounded by a stromal retraction, making both tumors' histological features overlap with each other. Interestingly, not only do histological features of benign condition overlap with BCC, but cancerous lesions of Merkel cell carcinoma, fast-spreading cancer, and squamous cell carcinoma with less aggressiveness also show potential overlapping features with basosquamous and other variants of BCC. The overlapping features of the tumor may have nested, nodular, or infiltrative growth patterns with stromal mucin retraction artifacts, tumor necrosis, and foci of squamous differentiation. The histological and morphological features mimicking BCC have been extensively covered by Stanoszek et al.[66] These overlapping features can lead to misdiagnosis. In addition, they might misguide the physician and consequently lead to overtreatment or undertreatment, as the clinician's choice of treatment is based on the type of skin cancer diagnosed.[66] Knowing these histological subtypes allow for the right choice of treatment, as more aggressive subtypes may not respond well to topical therapies compared to how superficial subtypes are treated.[63] Histologically confirmed, superficial BCC can be treated with the FDA-approved drug, fluorouracil (5-FU) and imiquimod creams, however, in practice, these agents can be used to treat other subtypes though not for treating aggressive cancer.[88] Moreover, the identification of certain histopathologic features is important for risk assessment. For example, micronodular, infiltrating, sclerosing and basosquamous are associated with higher recurrence and invasion of the dermis.[88] One of the main reasons for recurrence for certain types of BCC, such as morphoeic, infiltrative, or multifocal lesions is the incomplete excision of BCC. This is attributed to the difficulty in identifying peripheral margin as positive margin with hematoxylin and eosin (H&E) staining technique, leading to incomplete excision of BCC, which would increase the risk of recurrence.[89] Thus, relying only on H&E staining for the accurate diagnosis of BCC and its variants and locating the lateral margin in an aggressive type of BCC will lead to erroneous diagnosis and treatment. Therefore, the additional diagnostic procedures are more warranted to eliminate erroneous diagnosis. This could be made possible with antibody-based diagnostic methods, commonly called immunophenotyping.

1.3.1.3. Immunophenotyping

According to the NCI, immunophenotyping is defined as "A process that uses antibodies to identify cells based on the types of antigens or markers on the surface of the cells."[90] Immunophenotyping may also be used to separate cells into different groups based on the markers they have on the surface. In addition, intracellular markers in combination with cell surface markers can also be used to differentiate cells and to isolate cells from a mixture of cell populations. Immunophenotyping not only helps in accurate diagnosis of tumor subtypes but also plays a significant role in understanding cellular heterogeneity, prognosis and serves as a guide to determine the most appropriate treatment strategy. Immunohistochemistry (IHC), immunocytochemistry (ICC), and immunofluorescence (IF) are immunophenotyping techniques that utilize antibodies to provide visual details about protein abundance, distribution, and localization.

Certain antigens are highly specific to a particular type of tumor. In the context of distinguishing BCCs from SCCs, antigens EpCAM and EMA can be used, especially since the EpCAM is seen in BCCs, but not in cSCCs. Similarly, EMA is present in bSCCs (basaloid squamous cell carcinomas), but not in BCCs.[91,92] Thus, IHC of skin lesions with EpCAM- and EMA-specific antibodies will help distinguish BCCs from cSCCs, which does mimic the former. However, on rare occasions, basosquamous carcinomas or metatypical BCCs and basaloid squamous cell carcinomas (bSCCs) may stain for both BerEp4 and EMA.[93,92] Therefore, under certain conditions, multiple markers are needed to discriminate BCCs from other disease entities.

IHC involves labeling of antibodies or antigens with fluorophores, which has various drawbacks. These probes used for detection can cause background fluorescence from unreacted or non–specifically bound probes, thus interfering with tissue antigenicity. This may consequently give rise to false-positive or -negative results.[94,95] In addition, the excision of tissues for IHC-based diagnosis covers only a portion of lesions and not the entire lesions, which might result in misdiagnosis. For example, certain subtypes of BCCs (e.g., superficial multicentric BCCs) have skip lesions, i.e., new foci appear usually at about 0.5 mm intervals away from other foci, which might not be excised while sampling for IHC, leading to a misdiagnosis, which in turn affects the treatment outcome.[89,96] This challenge can be addressed using optical immunofluorescence imaging.

1.3.1.4. Biomarkers in BCC

Per the WHO, a biomarker is defined as "any substance, structure or process that can be measured in the body or its products and influence or predict the incidence of outcome or disease."[97,98]

Cancer biomarkers can be classified into the following categories based on their usage: predictive biomarkers, prognostic biomarkers, and diagnostic biomarkers.[99]

Predictive biomarkers aid in the selection and administration of contemporary treatments, tailored to a specific tumor profile or patient profile allowing for early assessment of potential treatment response and success rate.[100]

Prognostic biomarkers, on the other hand, do not predict the likelihood of treatment benefit but provides insights into the risk of overall disease outcome such as relapse or metastatic disease progression, drug resistance, and or resistance to anoikis.[101,102]

Diagnostic biomarkers on other hand are tumor-specific, used to establish the presence of a disease condition or to differentiate a specific disorder from another.[99]

Examples of predictive, prognostic, and diagnostic cell surface markers of BCC such as antigen/antigen type, the cellular location of antigens, and associated recognition antibodies are summarized in **Table 3**.

Table 3: Cell surface Biomarkers in BCC

Protein	Diagnostic	Prognostic	Treatment	Ref
EpCAM# (Cell surface)	Ber-EP4 stains are positive for BCC and SCC. **mAb Clone VU-1D9**	Downregulated in invasive BCC. **mAb Clone Ber-EP4, M0804**		92,103,104,93,105
EMA# (Cell surface)	Stains negative for BCC and Positive for SCC and Sebaceous carcinoma (SC). **mAb Clone E29**			92,106,107,108,93,109
CD44# (Cell surface)	CD44 stains positive in all bSCC and 50 % of BCC. **mAb Clone MRQ-13**			92,110

	CD44 is highly expressed in BCC compared with a normal cell. **mAb Clone MEM-263/MEM-85**			
CEA# (Cell surface)	BCC is negative for CEA and 35 % of SCC stains positive for CEA. **mAb Clone II-7, N1586** However, this is not statistically different.			106,93
CD10 # (Membrane and Cytoplasm)	NO cSCC stains positive for CD10, whereas 75% of BCCs stain positive for CD10, which is statistically significant when staining with clone. **mAb Clone M0727 DAKO**			106
Podoplanin # (Cell surface)	87.7% in SCC, and 64% of TE, and 32.1% of BCC stain positive for Podoplanin. There is no difference in expression of Podoplanin between BCC, AK, and normal skin. However, there is a significant difference between TE and SCC vs BCC. **mAb Clone D2-40**			107,111
PGRMC1# (Cell surface)	BCC and SCC did not stain positive for PGMRC1, whereas 80% of Sebaceous Carcinoma stains positive for PGMRC1. **mAb Clone NA**			109
MT1-MMP #	MT1-MMP is expressed in BCC, but not in the	MT1-MMP is expressed high in high-risk BCCs (mixed,		112

(Cell surface, cytoplasm, nucleus)	majority of epithelial cells.	infiltrative, morpheaform, micronodular, and basosquamous types), compared with the low-risk BCC (nodular type). **mAb Clone 113-5B7**		
CSPG4# (Cell surface)	Data were not available for BCC. However, its expression is seen in epidermal skin/IFE interfollicular epidermal stem cells and SCC and melanoma. **mAb Clone 18**			113,114,115
LGR5/GPR49# (Cell surface)	LGR5/GPR49 expressed high levels in BCC compared with normal skin. **Clone: NA**			110, 116
CD90# (Cell surface)	CD90 is moderately expressed in BCC compared with control (normal skin). **Clone: NA**			110
CD138# (Cell surface)	Basal cells and BCC has low expression of Sdc-1/CD138 (**mAb Clone MI15**), compared with differentiated keratinocytes **(anti-P117**)			117,118
EGFR* (Cell Surface)	High expression (**mRNA)** of EGFR in BCC compared to healthy tissues	Recurrent BCC shows much more high expression compared to nonrecurrent BCC		119
PTCH1*# (Cell surface cytosol/ nucleus)	85% BCC stains positive for PTCH1, compared with adjacent histologically normal epithelial cells.			120,121

	Ab53715 Polyclonal antibody			
nfP2X7 (Cell surface)			BIL010t: polyclonal Ab	[122]

* Gene expression; # protein expression.

Underline, not statistically significant; NA, not available.

1.3.2. Therapy

Surgical therapy is the mainstay treatment comprising conventional surgical excision, Mohs micrographic surgery (MMS), cryotherapy (CT), and electrodesiccation and curettage (ED&C). Nonsurgical treatment options for BCC include topical therapy (imiquimod, intralesional chemotherapy), the use of energy devices (photodynamic therapy, radiation therapy), systemic therapy encompassing targeted therapy (vismodegib, sonidegib). Immunotherapy with anti–programmed cell death 1 (PD-1) antibodies is a promising therapeutic option, currently being investigated in clinical trials.[123-124] This broad range of therapeutic modalities allows the practitioner to achieve outstanding results in eliminating BCC, preserve the function of normal parts, reduce the adverse effects and cost.[125] Here, we will cover the surgical treatment and systemic therapy in detail, two important treatment modalities relevant to our work.

1.3.2.1. Surgical treatment

a) Standard excision

Standard excision is the mainstay of treatment for primary BCCs with nonaggressive histology (nodular or superficial) as it is highly efficacious. It has been reported that the 5 year cure rates is > 95% and the recurrence rate is less than 5 % than those associated with nonsurgical treatment modalities.[123,126] During standard surgical excision of BCC, it is vital to mark the clinical margin of the tumor based on dermatoscopy, visual inspection, and/or reflectance confocal microscopy and then mark intended surgical margins approximately 4 mm margin around the clinically apparent border of the tumor, down to subcutaneous fat. This method is popularly known as conventional “bread loaf” histopathologic sectioning. Such marking is done before infiltration of anesthesia as this distorts the tissue.[127,128,125] The purpose of this standard surgical therapy is to eliminate both the clinically apparent tumor and its microscopic extension into normal-appearing skin.[127] However, standard surgical treatment has its own disadvantage if it is used to

treat an invasive type of BCC, as it is hard to point out the skin lesions outside the excision margins.[89]

b) Mohs micrographic surgery

Mohs micrographic surgery (MMS) is microscopic controlled tissue sparing surgical excision procedure that allows histological evaluation of the peripheral margin during the surgical procedure.[129] The tumor is surgically removed layer by layer and each layer is examined microscopically on-site, with the removal continuing until no abnormal cells remain.[130,131] The aggressive BCC types such as morpheaform, metatypical and infiltrating forms that show high recurrence rates are treated with MMS[132,133] because 5 and 10 year recurrence rates for primary BCCs treated by MMS is 2.5% and 4.4%, whereas recurrence rate is 4.1% and 12.2% by standard excision.[134-135] It has been demonstrated that MMS has a 99% cure rate for primary BCCs, compared with 91.3% for other methods.[136] MMS is an outpatient, time-consuming (generally taking 3–4 h or more if several rounds of excisions are needed) and expensive procedure, performed most commonly under local anesthesia.[137] The drawback of standard surgical and MMS excision is that the boundary of the tumor margin is not well defined and the tumor can be extended beyond the superficial layers of the skin.

1.3.2.2. Systemic targeted therapy

According to NCI (National Cancer Institute) definition, targeted cancer therapies are drugs or other substances that block the growth and spread of cancer by interfering with specific molecules ("molecular targets") that are involved in the growth, cell division, and spread of cancer. Targeted cancer therapies are synonymous with terms such as "molecularly targeted drugs," "molecularly targeted therapies," "precision medicines."[138]

There are two main types of targeted therapy. The first type is small molecule drugs or small molecule inhibitors (SMI), which are small enough to enter cells. The second type is monoclonal antibodies (mAb), which are too large to enter cells, so it targets molecules outside the cells or molecules on the surface of diseased cells.

a) Small-molecule inhibitor–based targeted therapy in BCC

In BCC, aberrant activation of the Hh signaling pathway is seen due to activating mutations in Hh pathway components such as SMO or PTCH1.[139,140] Normally the Hh pathway is activated during embryonic development for cell differentiation and proliferation. In adults this pathway

is silent under normal conditions, however, it is maintained in certain tissues and stem cell proliferation.[141,142] As Hh signaling participates in the pathogenesis of BCC and absents in adult cells, targeted inhibition of this pathway was hypothesized to offer a possibility of selectively killing BCC. Scientists around the globe have developed various small-molecule inhibitors such as vismodegib, sonidegib (LDE225), itraconazole, LY2940680, BMS-833923, and PF-04449913, which are in various stages of clinical development.[143] In 2012, vismodegib, a smoothened (SMO) antagonist that inhibits Hh signaling, was approved by FDA and became the first small molecule inhibitor used to treat advanced BCC. In 2015, the FDA approved sonidegib, a second Hh signaling inhibitor.[144] The detailed mechanism of action of this small molecule inhibitor is reported elsewhere.[145] There are pieces of evidence that some BCC patients may develop acquired drug resistance to Hh-pathway inhibitor (Vismodegib and sonidegib during the course of their therapy,[146] either by a gain of function of *Gli2*-transcription factors via amplification of *Gli2*, mutations in SMO or through the upregulation of PI3K signaling that reactivate Hh signaling.[147-148] Thus, alternative targeted treatment options should be employed to treat advanced BCC patients who are resistant to Hh-Inhibitor or further improvement in the existing Hh inhibitor drug is warranted[149] to treat SMI resistance in advanced stage BCC patients.

b) Antibody-based targeted therapy in BCC

The antigen nfP2X$_7$, with hidden epitope (E200) that distinguish them from the variant P2X$_7$ are reported to be expressed in BCC but not in normal (non-cancerous cell). The polyclonal antibody (BIL010t) shown to interact with antigen nfP2X$_7$ in BCC is used in Phase I clinical study as a novel antibody based ointment for treating BCC patients-superficially[122]. Immune checkpoint referring to a plethora of molecules (PD-1/CD279 & CTLA-4/CD152) that act as gatekeepers of immune responses and crucial for maintaining self-tolerance is often expressed in immune cells such as T cells binds to a cognate molecule (B7-H1/PD-L1& CD80/CD86) expressed in cancer cell and or antigen-presenting cells (APC). This interaction dampens the activation of T-cells by outcompeting the CD28 stimulatory coreceptor expressed in T-cell in recognizing CD80/B7-1 and CD86/B7-2, which is required for CD28 mediated T cell activation.[150] In other words this interaction between immune checkpoint and cognate protein/ligand trigger sends an "off" signal to the T cells, thereby not allowing T-cells to mount an attack on cancer. However, this immune checkpoint protein is blocked from binding to its partner protein using antibodies or peptides collectively known as immune checkpoint inhibitors, which prevent the "off" signal from being sent, allowing the T cells to kill cancer cells.[151] The antibody based immune checkpoint inhibitor

is the only intravenous-based immunotherapy that is currently available to treat advanced BCC, i.e., BCCs that are invasive and/or resistant to Hedgehog (Hh) inhibitor and radiotherapy.[152]

Monoclonal antibodies such as ipilimumab and nivolumab that block immune regulatory proteins such as the CTLA-4 receptor or programmed death-1 (PD-1) have demonstrated remarkable efficacy in controlling the growth of multiple tumor types. Interestingly, there are few studies reported the efficacy of ipilimumab and nivolumab in treating BCC.

Ipilimumab: Anti-CTLA4 blockade mAb, which binds to CTLA-4 and promotes T-cell recognition of tumors. Mohan et al. treated a 60-year-old melanoma patient who developed advanced BCC adjacent to recurrent nodular melanoma with ipilimumab. This patient subsequently showed regression of BCC. This observation suggests that ipilimumab might work for advanced BCC.[153]

Nivolumab is a fully human PD-1 inhibitor. When used off-label to treat advanced BCC in the year 2016, after exhaustion of available therapeutic options, it showed promising results.[154] Similarly, another anti-PD-1 drug pembrolizumab, a humanized PD-1 blocking antibody, has been tested in an advanced metastatic BCC patient who failed to respond to Hh signaling inhibitor and showed partial response for more than a year.[155] An independent phase-1 clinical study reported a similar outcome of partial response for more than 12 months when advanced BCC is treated with REGN2810, a fully human anti-PD-1 monoclonal antibody. From such clinical studies, it is clearer that ICIs such as anti-PD-1 and anti-CTLA4 either individually or in combination provide stable to partial response for metastatic BCC patients.

However not all BCC patients show a similar response to anti-PD-1 therapy. Some BCC patients develop resistance to anti-PD-1 and the reason for resistance is a “cold” tumor microenvironment, characterized by lack of human leukocyte antigen class I expression, low PD-L1 expression, a high number of immune regulatory cells,[155] and lower tumor mutational burden (45 mutations per megabase). Interestingly, Cohen et al. reported refractory metastatic BCC disease with high tumor mutational burden (103 mutations per megabase) showed favorable response to anti-PD-1.[156] This implies that antibody-based ICIs might work better for advanced disease with a high mutation rate compared to nonadvanced BCC (low mutational burden), which can show a better response to other targeted therapy strategies—the use of small molecule inhibitors. To date, no studies have reported the use of immunotoxins or antibody–drug conjugate in the treatment of advanced or nonadvanced BCC.

1.4. ANTIBODY STRUCTURE CHARACTERISTICS AND PRODUCTION

1.4.1. Antibody structure and characteristics

Antibodies are immunoglobulin (Ig) glycoproteins produced and secreted by specialized B lymphocytes known as plasma cells. Immunoglobulins consist of two heavy (H) and two light (L) chains in a light–heavy–heavy–light structure arrangement. The polypeptide chains are held together by disulfide and noncovalent bonds. L-chain can consist of either a κ or a λ chain, ever one of each. The light chain is made up of two immunoglobulin domains, one NH2-terminal "variable" (V) domain and the other one is the COOH-terminal "constant" (C) domain. The heavy chain of the IgG antibody contains four domains: one NH2-terminal "variable" (V) domain and three COOH-terminal "constant" (C) domains, named as the first (CH1), second (CH2), and third (CH3) constant domains that are numbered from the amino-terminal end to the carboxy terminus. The spacer hinge region (the stretch of polypeptide chain) located between the first (CH1) and second (CH2) domains helps in flexibility and acts as a spacer preventing the interference of the Fab arms with the Fc binding sites[157]. Each IgG will have two Fab (for **F**ragment **a**ntigen **b**inding) fragments, and two Fc fragments (for Fragment crystallizable or Fragment constant), each Fab will have one complete light chain paired with the V_H and C_H1 domains of the heavy chains that recognize the antigen. More specifically, variable loops (three each on the light (VL) and heavy (VH) chains) are responsible for binding to the antigen. These loops are referred to as the complementarity-determining regions (CDRs). Fc fragment corresponds to the paired C_H2 and C_H3 domains with no antigen-binding activity, but it interacts with effector molecules and cell receptors such as Fc receptors, to trigger complement-mediated cytolysis.[158,159]

1.4.2. Monoclonal antibody production

The monoclonal antibody can be produced by two different mechanisms: (i) Immunization-hybridoma technology and (ii) non-immunization-dependent recombinant antibody display technology (using for example bacteriophage).

1.4.2.1. Immunization-Hybridoma technology

Immunization of animals for polyclonal antibody production is mainstream antibody production for experiments involving the identification and quantification of specific proteins of interest within complex biological environments. However, for in vivo experiments or therapeutics monoclonal antibody (mAb) is preferred as it has reduced background reactivity or lack of non-

specific binding. In 1975, Georges Köhler and César Milstein found a technique for monoclonal antibody production known as hybridoma, for which they, along with Niels Jerne, received the Nobel Prize in Physiology or Medicine in 1984. Hybridomas are cells formed by fusing short-lived antibody-producing B cells from an immunized mouse or rabbit with immortal myeloma cells, which lack the HGPRT (hypoxanthine-guanine phosphoribosyltransferase) gene. The fusion is done using polyethylene glycol or inactivated Sendai virus, or electrofusion, and hybridomas are grown in HAT (hypoxanthine, aminopterin, and thymine) medium, which allows them to survive and grow; subsequently, the mAbs produced from hybridomas are purified by affinity chromatography.[160]

1.4.2.2. Non-immunization

a) Antibody phage display

The phage display library technology refers to the cloning and expression of antibodies variable region genes obtained from a B lymphocyte repertoire reactive to a specific antigen isolated from a certain host (either immunized or non-immunized) in a bacteriophage. Genotype and phenotype are linked: a bacteriophage displaying an antibody is carrying his specific genetic information. This is followed by the screening of the library on a specific antigen, thereby isolating corresponding specific antibodies. Successful generation of antibody fragments using phage display involves (i) construction of human immune, naïve, synthetic, and semisynthetic recombinant antibody gene libraries; (ii) amplification of gene library; and (iii) bio-panning, i.e., screening and selection. For more details on phage display, please refer to Ferro et al. and Almagro et al. [161,162].

b) Recombinant antibody technologies

Recombinant antibodies are monoclonal antibodies of full length or fragmented form, generated in vitro using synthetic genes that encode for the variable heavy and light chain which mediate antigen binding. The engineering of recombinant antibodies (rAb) begins with the determination of immunoglobulin (Ig) mRNA transcripts or cDNA sequence from a monoclonal antibody (mAb) producing hybridoma cell. Alternatively, one may retrieve antibody sequences from the public repository or published literature. To retrieve sequence from hybridoma clone, a PCR-based method combined with molecular cloning can be used to predict the variable region (VR) primary structure that corresponds to the parental mAb or one can use tandem mass spectrometry (MS) for confirmation of individual VR protein sequence from purified mAbs.[163,164] Subsequently, the identified antibody sequences are verified using IgBLAST to predict the

complementarity-determining region (CDR) of the variable region. On confirmation of CDR, in silico design of plasmid is performed using a plethora of vector mapping software (SNAPgene, Vector NTI, Serial cloner, etc.) to clone into a suitable expression vector for production and secretion of recombinant mAb. A variety of recombinant production systems are available, such as gram-negative and gram-positive bacteria, yeasts and filamentous fungi, insect cell lines, mammalian cells, and transgenic plants and animals.[165] Mammalian cell lines such as HEK293 (human embryonic kidney 293) and Chinese hamster ovary cells (CHO) are the preferred recombinant production systems that reduce the risk of antibody immunogenicity due to altered, nonhuman glycosylation patterns. However, smaller antibody fragments including bispecific antibodies without any glycosylation have been successfully produced in bacteria, which are a comparably cheap and fast expression host for large-scale antibody production. Schirrmann et al. have provided some examples of recombinant antibodies produced in different hosts.[166]

Recombinant DNA technology can also be employed in the conversion of mouse antibodies to chimeric and humanized antibody.[167] This is essential when a mAb is used for human application, especially since the human immune system tends to generate anti-mouse antibodies (HAMA) against mouse mAb. This leads to clearance of mouse mAbs from the bloodstream and possibly causes hypersensitivity reactions. They also prevent the mouse antibody from reaching its target and in some cases block its binding to antigen. Chimeric antibody constructs (i.e., humanized chimeric antibody) can be generated with recombinant DNA technology, where the variable region (V_H and V_L) of the mouse-derived antibody domain are spliced and ligated onto the human constant region of the human-derived antibody (C_H and C_L). Such chimeric antibodies show the same specificity to the target antigen and reduced immunogenicity in human host.[164,168] Another type of antibody construct known as a humanized antibody can be generated by replacing CDR_S (complementarity-determining region) of human variable region genes with specific CDR_S for mouse antibody is called CDR grafting or CDR transplantation.[169,170] Human antibodies are usually derived from human B cells; interestingly, human antibodies may also be derived from other animals such as a transgenic mouse that has human chromosomal fragments comprising the gene encoding the H chain and L chain of the human antibody or by other nonimmune resource methods such as phage display for/by naive library.[171]

1.5. RECOMBINANT PROTEIN TAGS FOR DIRECTED CONJUGATION

Antibody/ligand-based, molecular targeted imaging probes are being developed for optical imaging to allow visualization of disease-specific markers. One of the key steps in constructing antibody/ligand-based molecular-targeted imaging probes is conjugating fluorophores to the antibody via chemical conjugation or enzymatic process. Chemical conjugation can be carried out using natural amino acid residues such as cysteine (Cys), selenocysteine (Sec), and lysine (Lys), and unnatural amino acids (i.e., p-acetyl-phenylalanine or p-acetyl-Phe and p-azidomethyl -Phe for probe/drug conjugation.[172] The use of natural amino acids is popular because there is no need to reengineer the antibody for drug attachment. However, the drawback is that they result in heterogeneous product mixtures (0-8 drugs per antibody) that cannot be further purified.[172] With the advancement in protein engineering, enzymatic site-specific conjugation such as conjugation to interchain or engineer cysteine residues, glycans, unnatural or noncanonical amino acids, and modification of peptide tags yield more homogeneous products that have demonstrated benefits over their heterogeneous counterparts in vivo.[173] The enzyme tag-based labeling allows fluorophores to be attached at a specific location on the imaging probe, with minimal off-site labeling. Examples of enzyme tag-based labeling methods include SNAP-tag (O6-alkylguanine-DNA alkyltransferase),[174] CLIP-tag (a modified version of SNAP-tag, engineered to react with benzylcytosine rather than benzylguanine derivatives)[175]. ACP-tag, PCP-tags, Halo-tag, FLAG-tag, HA-tag, Myc-tag, Sfp phosphopantetheinyl transferase (CoA), Sortase A, and Avi biotin ligase,[176,177,178,179] are derived from non-human organisms and can also be used to selectively label corresponding fusion proteins with synthetic probes in both cell imaging and *in vitro* applications.

SNAP-tag is superior for being low immunogenicity and self -labelling, compared to other peptide tags especially to small peptide tags such as FLAG-tag, HA-Tag, ACP-tag, PCP-tags, or Myc-tag. The latter require multiple copies and the affinity of such tags is low and does not provide dense labeling to visualize a single molecule in high resolution.[180] Furthermore, peptide epitopes are not stably expressed in cells without fusion to a scaffold protein.[181] SNAP-tag in comparison to mTagBFP and other tags of 20-30 kDa size, shows the most active and least intrusive in terms of localization when expressed as endogenous protein in *E. coli*.[182] Moreover, SNAP-tag can be used to optically image tumors when conjugated to a fluorophore such as Alexa Fluor 647.[183] It can also be used in antibody–drug conjugates (ADCs), immunotoxins, and photodynamic therapy to induce apoptosis or necrosis in a tumor by conjugating with cytotoxic

drugs such as Auristatin F, and photosensitizers (e.g., IR700, chlorin e6) on exposure to a changeable wavelength of light.[184,183,185] The dual nature of SNAP-tag fusions as both a diagnostic and therapeutic tool with fluorophores and photosensitizers in a 1:1 stoichiometry reinforces its role in cancer treatment in the current era of precision medicine.[186]

SNAP-tag is 182 amino acid residue polypeptide, derived from human DNA repair enzyme (O(6)-alkylguanine-DNA alkyltransferase (AGT)), that is modified by introducing a total of 19 point mutations and deletion of 25 C-terminal residues in AGT encoding gene MGMT. This mutation will render the AGT to possess higher activity with BG (O(6)-benzylguanine (O6-BG) derivatives) substrate, which is a synthetic derivative of guanine and reduced binding affinity to DNA with reduced size (~20 kDa) compared with wild-type AGT. The SNAP-tag can react with benzylguanine (BG) and benzylchloropyrimidine (CP) derivatives by transferring the substituted benzyl group to its active site through a nucleophilic substitution reaction while releasing free guanine[187] (see **Fig. 2**). The addition of a 20th point mutation in amino acid position 30 gives rise to SNAPf, a fast-labeling variant that displays up to a tenfold increase in its reactivity toward BG substrates.[188]

Table 4: Summary of the 20-point mutation

1. 32: Increase the activity against BG derivatives
2. 33: Increase the activity against BG derivatives
3. 71: Increased half-life
4. 115: Suppress DNA binding and reactivity to nucleosides
5. 116: Suppress DNA binding and reactivity to nucleosides
6. 125: Suppress DNA binding and reactivity to nucleosides
7. 127: High stability (prevents proteasomal degradation)
8. 128: High stability (prevents proteasomal degradation)
9. 131: Stops triggering protein unfolding and degradation on alkyl transfer
10. 132: Stops triggering protein unfolding and degradation on alkyl transfer
11. 134: Avoid unfolding of an alkylated protein
12. 135: improved the overall protein stability
13. 150: Increase the activity against BG derivatives
14. 151: Increase the activity against BG derivatives
15. 152: Increase the activity against BG derivatives
16. 153: Increase the activity against BG derivatives

17. 154: Increase the activity against BG derivatives
18. 157: Significantly increased the reactivity toward BG and stability of protein
19. 159: Significantly increased the reactivity toward BG and stability of the protein.
20. 30: SNAPF (10-fold increase in reactivity compared to SNAP)

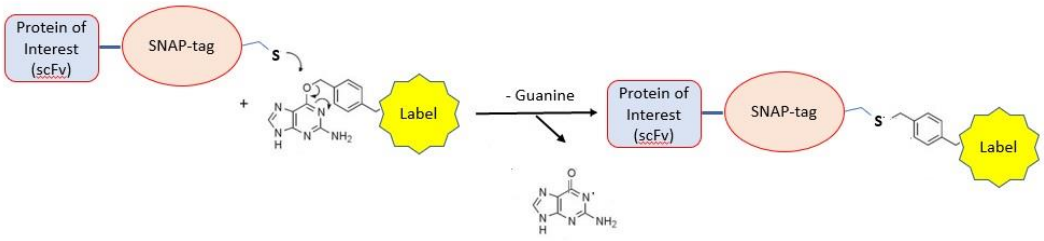


Figure 2: SNAP-tag reaction scheme with O6-benzyl guanine derivative

SNAP-tag fusion protein can be expressed along with full-length antibodies or with antibody fragments called single-chain variable fragments (scFvs), which are formed by the tandem arrangements of the heavy (VH) and light chain (VL) domains joined by a flexible serine/threonine linker a 15-amino acid residue linker of composition (Gly4Ser)3[189,190]. The scFv-SNAP-tag is more favorable for use as targeting elements in drug delivery system, because of their small size, scFvs are expected to display a higher cell penetration rate than their full-length counterparts. These scFv-SNAP-tag fusion proteins can be expressed in bacterial cell system[182] or mammalian cell system[183] to target various tumor-associated antigens (TAAs) specific for BCC and to distinguish BCC from normal tissue provide information for image-guided surgery and photodynamic therapy (PDT) or photoimmunotherapy (PIT)[191].

1.6. BIOINFORMATICS IN BIOMARKERS DISCOVERY

Bioinformatics and computational biology (BCB) combine elements of computer science, information technology, mathematics, statistics, and biotechnology, providing the methodology and in silico solutions to mine biological data and processes for knowledge discovery.[192] Identifying tumor targets is the main bottleneck in immunotherapy, that has been made easy with the availability of bioinformatics tools and data resources used in high throughput molecular

analyses range of data types, including those generated from microarray, whole-exome sequencing (WES), RNA-seq, DNA copy number, and DNA methylation assays [365].

1.6.1. cDNA-microarray

cDNA microarray is an innovative technology that facilitates the analysis of the expression of thousands of genes simultaneously. A microarray is typically a glass slide on which DNA molecules are fixed in an orderly manner at specific locations called spots (or features). A microarray may contain thousands of spots and each spot may contain a few million copies of identical DNA molecules that uniquely correspond to a gene. Affymetrix and Illumina are two microarray platform commonly employed in gene analysis, Affymetrix arrays use set of different 25-mer probes synthesised in situ to characterize gene expression.[193] Illumina arrays utilize multiple copies of single 50-mer probe attached to microbeads to quantify target level.[194]

The first step in the analysis of microarray data is to process the fluorescence image of the microarray, in which each spot that corresponds to a gene has an associated fluorescence value representing the relative expression level of mRNAs. Normalization describes the process of removing or minimizing nonbiological variation in measured signal intensity levels (mRNA expression values) so that real biological differences in mRNAs expression among samples can be appropriately detected. Some popular normalization methods are global normalization, LOWESS intensity-dependent nonlinear normalization, intensity-dependent linear normalization, and quantile normalization.[195-196] The data analysis process is typically involved filtering to identify those differentially expressed genes using a combination of criteria, including fold change, and statistical significance genes determined by t-statistics or statistical analysis software packages such as BRB-ArrayTools, Matlab, or R, where user can set pre-defined a *p*-value and FDR value. p is simply a measure of the probability that an observed difference in expression between diseased and normal could have occurred just by random chance. The lower the *p*-value, the greater the statistical significance of the observed difference. On the other hand, FDRs (false-discovery rates) represent the probability that any particular significant finding is a false positive.[197] Most authors refer to statistically significant as $P < 0.05$ (conventionally, less than 1 in 20 chances of being wrong), $p < 0.01$ (conventionally, less than 1 in 100 chance of being wrong), and statistically highly significant as $P < 0.001$ (less than one in a thousand chance of being wrong). A false discovery rate of <0.05 means that any declared significant finding has a 5% chance of being a false positive.[198,199]

1.7. AIMS OF THE STUDY

1.7.1. Aims

The overall aim of this study is to synthesize scFv-SNAP fusion proteins and label them with fluorophores to target tumor associated antigens highly expressed in BCC compared to normal skin tissue, further allowing to selectively excise entire BCC tissue by sparing the normal tissue in a single excision.

1.7.2. Objectives

In pursuit of the abovementioned overall aim, three phases of work were planned to be executed according to the following framework (**Fig. 3**).

Phase I: Identification/screening of differentially expressed BCC-associated cell surface antigen–encoding genes

a) Literature search approach is established using web-based medical journal search engines such as PubMed, Embase, and MEDLINE, with keywords skin cancer biomarker, non-melanoma skin cancer associated antigen, basal cell carcinoma cell surface antigen, BCC target, biomarkers for basal cell carcinoma, and so on.

b) Microarray data encompassing basal cell carcinoma and normal human epidermal keratinocytes are retrieved from the public repository "Gene Expression Omnibus." Affymetrix chip-based microarray data are reanalyzed using computational tools "BRB-ArrayTools" and "Transcriptome Analysis Console."

Phase II: Generation of scFv-SNAP fusion proteins

a) Identification of antibody sequence from the public repository "Google Patents" and DEPATISnet.

b) In silico cloning of expression constructs encoding scFv-SNAP fusion protein.

c) Molecular cloning of expression constructs and sequencing of open reading frames for mammalian expression of scFv-SNAP fusion protein.

d) Transient expression of scFv-SNAP in HEK293T mammalian expression system.

e) Purification of His-tagged protein from cell culture supernatants.

f) Electrophoretic and immunoblotting analysis of purified proteins.

Phase III: Characterization of scFv-SNAP proteins as a cell surface antigen–targeting agent for diagnosis and therapeutics

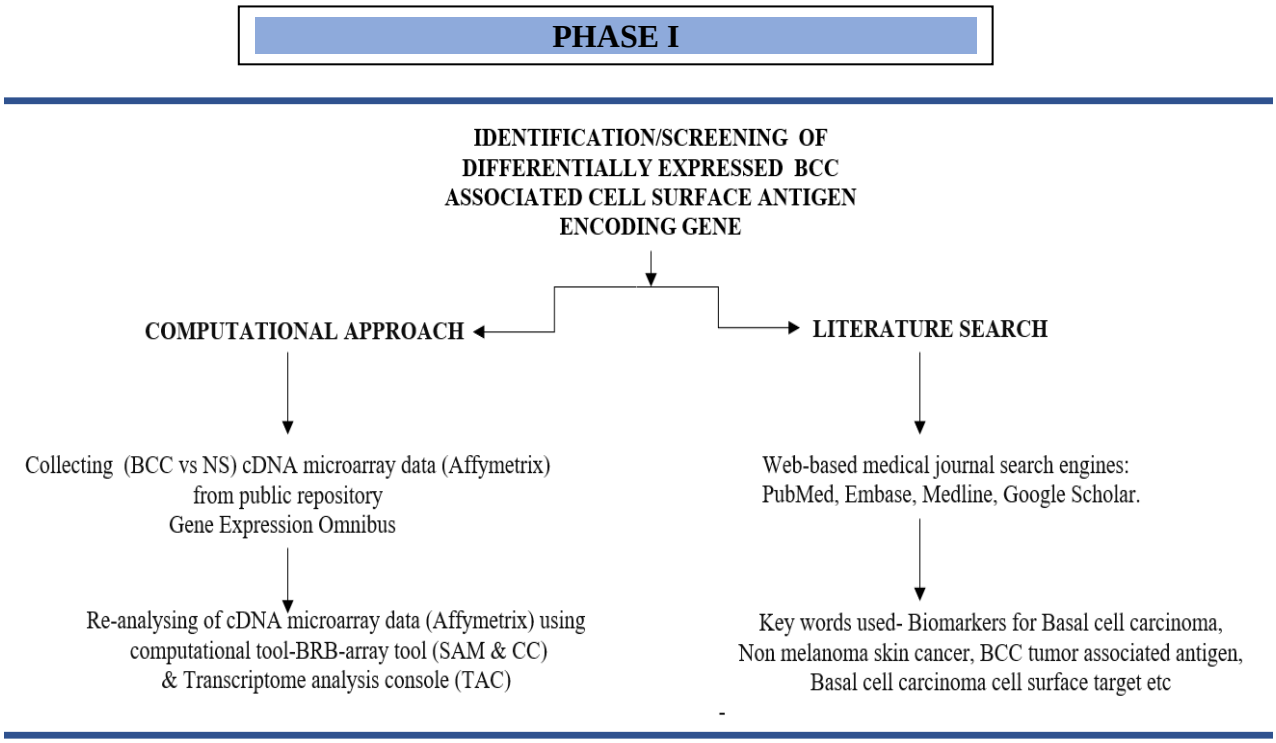
a) Validation of selective binding of fluorophore labelled scFv-SNAP proteins (UBS54, mAb 9.2.27, ID405, h-STL002) to EpCAM-, CSPG4-, EMA-, and CD138-positive tumor cells by confocal imaging, and optionally with flow cytometry.

b) Determination of scFv(UBS54)-SNAP-Auristatin F dose-dependent cytotoxic activity in EpCAM-positive tumor cells.

Future aims after proof of concept findings

a) To characterize the selective binding of labeled antibody fusion proteins in FFPE BCC tissue sections and subsequently generate a PDX mouse model and topically applied with fluorophore/chromophore labelled scFv-SNAP tag under fluorescent light or white light to selectively detect and excise the BCC tumor tissue by sparing the normal tissue.

b) To use antibody fusion proteins conjugated with a cytotoxic drug as a topical/systemic application as an alternative treatment option for invasive BCC.



PHASE II

GENERATION OF scFv-SNAP FUSION PROTEIN

↓

Identification of antibody sequence targeting the selected cell surface antigen from the public repository DEPATISnet, Google Patent search etc

↓

In silico design of scFv-SNAP fusion protein

↓

Requested synthesis of ScFv-SNAP fusion protein encoding open reading frame (ORFs)

↓

Molecular cloning of ORFs into custom vector for protein production in mammalian expression system –HEK293T

↓

Transfection and Expression of scFv-SNAP constructs into a mammalian expression system

↓

Enrichment of scFv-SNAP fusion protein from cell cultured supernatant by Immobilised metal affinity chromatography constructs into a mammalian expression system

↓

Analysis of scFv-SNAP fusion protein integrity, purity and quantity by SDS-PAGE electrophoresis and Immunoblotting

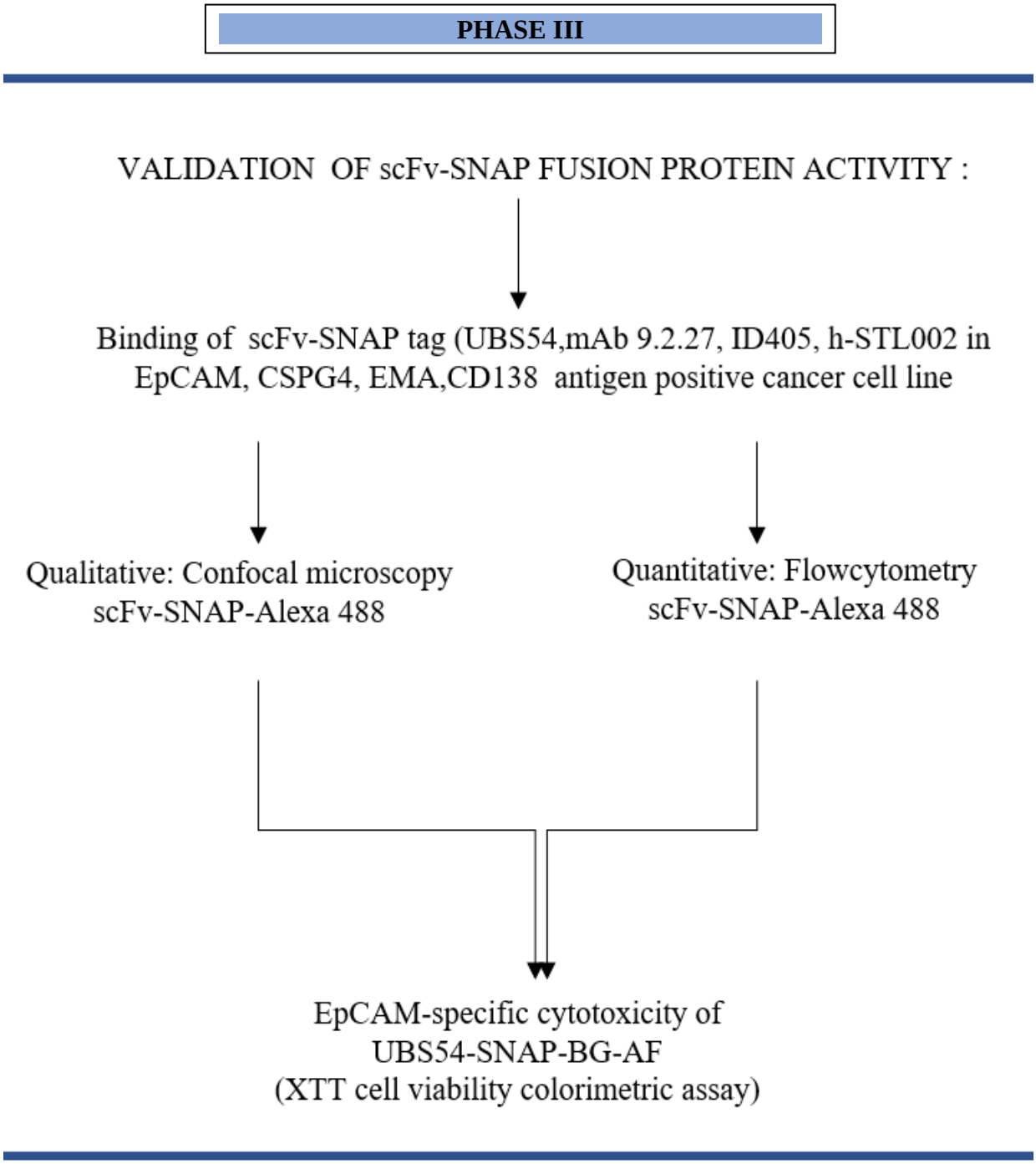


Figure 3: Schematic structure and workflow of the research work performed

CHAPTER 2: MATERIALS AND METHODS

2.1. COMPUTATIONAL BASED IDENTIFICATION OF BCC-ASSOCIATED CELL SURFACE ANTIGEN

2.1.1. Acquisition of microarray data

The GEO database (http://www.ncbi.nlm.nih.gov/geo) is a public functional genomics data repository that contains multifaceted data, including data derived from microarray and next-generation sequencing. GEO database was searched using the following keywords: “Basal cell carcinoma” (All Fields) and “Homo sapiens” (porgn) and for precise searching “gse” or “gds”(Filter) were used. The selection criteria for datasets were as follows: (i) the datasets were gene expression profiles of the Affymetrix platform, (ii) the data was obtained from BCC tumor tissues of BCC patients or normal skin tissues (human normal epidermal keratinocytes) from healthy individuals or diseased individuals (human healthy skin epidermal cells), and (iii) the patients did not receive medication or other treatments. According to the selection criteria, four datasets (GSE7553, GSE103439, GSE42109, and GSE39612) were chosen, as they are generated by the GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array platform, Human Genome U133 Plus 2.0 Array feature size of 11-micron and 11 probe pairs per probe set with over 54,000 probe sets representing approximately 38,500 genes on a single array. (Estimated by UniGene coverage). Various public data sources were used for the HG-U133 Plus 2.0 Array design. Sequence data for new content were obtained from dbEST (NCBI, Feb. 2003), GenBank® (NCBI, Feb. 2003, Release 134), and RefSeq (NCBI, March 2003). Additionally, a draft assembly of the human genome (NCBI, Nov. 2002, Build 31) was used to assess sequence orientation and quality. Therefore, Affymetrix is our choice of interest, rather than Illumina, to explore the gene* expression profiles in tumor tissues from BCC patients or normal skin (NS) tissue from healthy or diseased controls (**Table 5**). These datasets would be used for the following integrated analyses, which consisted of 21 BCC tissues and 80 normal tissue samples. Note: * Here we refer gene as the transcriptome or the set of messenger RNA (mRNA) transcripts expressed by a group of gene

Table 5: Microarray datasets for basal cell carcinoma and normal skin

GEO accession	BCC Case	Control	Year	Platform	Author
GSE7553	15	4	2008	GPL570[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	Sean Yoder
GSE39612	2	64	2012	GPL570[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	Paul Harms
GSE42109	0	10	2013	GPL570[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array GPL571 [HG-U133A_2] Affymetrix Human Genome U133A 2.0 Array	Mayte Suarez-Farinas
GSE103439	4	2	2017	GPL570[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	Yoshiaki Tabuchi

GEO: Gene Expression Omnibus; Norma skin represent human epidermal keratinocytes

2.1.2. Microarray data preprocessing and analysis

To decrease bias from chip types, we selected Affymetrix human genome U133 Plus 2.0 array as chip assay because several BCC and normal skin samples are performed by HG-U133_Plus_2 and have been stored in the GEO database.[200] The raw data file or the series matrix file(s) of GSE7553, GSE3961, GSE42109, and GSE103439 were downloaded from the GEO database. The software "Transcriptome Analysis Console" (TAC) version 4.0.2.15 is used to convert Affymetrix cell intensity (CEL) files to a probe summarization TXT file to import in BRB-ArrayTools.[201]

The BRB-ArrayTools v4.6.0 Beta 1(https://brb.nci.nih.gov/BRB-ArrayTools/download.html) is an integrated software for the visualization and statistical analysis of microarray gene*-expression data. The preprocessing step included selection and merging of expression array of basal cell carcinoma and normal skin from the text file, local background subtraction, averaging of intensities of duplicated probes, log2 transformation, and quantile normalization across multiple arrays. The merged dataset comprised 21 BCC samples and 80 normal skin samples (human epidermal keratinocytes). Chip data were filtered according to the following criteria: i) Genes were excluded if <20% of the expression data had 1.5-fold change (in either direction) compared to the gene's median value when comparing the two types of samples, and ii) missed or filtered out gene expression data should be ≤50%. iii) log intensity variation *p*-value >0.01. The filtered genes were tested with a random variance version of the t-test or F-test separately for each gene with the *p*-values below the threshold (*p.* value<0.001).

2.1.3. Identification of DEGs and screening for cell surface markers

To identify gene* expression between predefined classes "BCC vs Normal Skin," first we run the BRB array-tool for the filtered genes with a random variance version of the t-test or F-test separately for each gene with the *p*-values below the threshold (*p*-value <0.001). The random variance t-test is an improvement over the standard separate t-test as it permits sharing information among genes within-class variation without assuming that all genes have the same variance[202]. Class comparison is performed to identify differentially expressed genes between BCC and normal skin tissue ($p<0.001$). To further enhance the biological and statistical effect and to reduce the false-positive rate of test results, we determine the false discovery rate (FDR) for each gene by the method of Benjamini and Hochberg[203] by setting cutoff criteria by adjusting *p*.value, FDR< $1e^{-07}$ (0.0000001) & $\log_2$fold-change (FC) > 3 were used to identify gene which is differentially expressed among classes. A volcano plot, scatter plot, and heat map of the DEGs from each dataset was subsequently generated using BRB-Arraytools. To screen for cell surface markers**,** the DEGs with *p.* value <$1e^{-07}$ and fold change (FC) value >3 were selected and compared with the list of 2,886 surfaceome markers published by Fluck et al.,[204] using gene venny (https://bioinfogp.cnb.csic.es/tools/venny/), this would allow uncovering the possible targetable cell surface protein-encoding gene. The venn diagram illustrating the intersection of these DEGs and surfaceome markers was visualized using venny 2.1 software (https://bioinfogp.cnb.csic.es/tools/venny/).

2.1.4. Validation of cell surface markers using alternative computational tools (SAM and TAC) and literature search

To validate the identified cell surface protein-encoding genes as a biomarker for Basal cell carcinoma, two validation approaches were considered, (i) computational approach, using significance analysis of microarray (SAM) and transcriptome analysis console (TAC) software. (ii) literature search approach using web-based medical journal search engines such as PubMed, Embase, MEDLINE, with keywords skin cancer biomarker, non-melanoma skin cancer associated antigen, basal cell carcinoma cell surface antigen, BCC target, and so on.

SAM is a nonparametric, permutation-based method proposed especially for microarray data analysis[205]. The normalized and log2 transformed data of all the samples were imported into SAM software and analyzed using multiclass analysis with 1000 permutations was done. A delta value of 0.90 and a fold difference of 3 were used to identify the genes differentially expressed between BCC and normal skin. Transcriptome analysis console (TAC) software provided by Affymetrix that uses Fisher's exact test and FDR controlling using the Benjamini–Hochberg

procedure is used for the validation study. Further for this analysis, we used cutoff for FDR *p*-value <0.001 and the difference in fold change is 3. Finally, we performed a comprehensive literature search for cell surface protein–encoding genes using web-based search engines, i.e., Google, Google Scholar, and so on, and electronic source databases such as PubMed, MEDLINE, or EMBASE, Semantic Scholar using the keywords "Basal cell carcinoma marker" AND ("differentially expressed gene in non-melanoma skin cancer (NMSC)/BCC") AND ("cell surface antigen in basal cell carcinoma").

2.2. IN SILICO CLONING OF MAMMALIAN AND BACTERIAL EXPRESSION VECTOR SYSTEMS

2.2.1. Identification of antibody sequence from public repository

We used platform DEPATISnet " https://www.dpma.de" to perform online searches for the state of the art published in patent literature from all over the world. DEPATISnet, the German patent information system on the internet, provided by the German patent and trademark office (DPMA). We used the search word "anti-EpCAM/anti-EMA/anti-CD138" etc in the title search of beginner's search mode. We scrutinize randomly few patent documents and look for certain published criteria or properties of antibodies such as human, humanized antibody, high binding affinity, lack of cytotoxicity in a naked antibody format and transgene expression host to choose the right antibody from publicly available multiple antibody sequence targeting this aforementioned target antigen. In addition, we performed a comprehensive literature search in PubMed[206] using the keywords "immunoglobulin, monoclonal antibody, single-chain variable fragment for interested antigen. One can use ABCD database (for AntiBodies Chemically Defined) a repository of sequenced antibodies, to identify and retrieve antibody sequence targeting interested cell surface protein. [207]

2.2.2. In silico design of a mammalian and bacterial vector system encoding the scFv-SNAP-tag fusion protein

Here in silico refers to designing circular DNA molecules/plasmid by virtual cloning, as a first step after identifying and retrieving the variable regions of heavy and light chain gene sequences of antibodies targeting various cell surface antigens such as EpCAM, EMA, CD138, LGR5, CD90 [See table 10 for the list of monoclonal antibodies and its target antigen that are retrieved from respective parent file]. Next these heavy and light chain V gene sequences of αEpCAM

(UBS54), αEMA (ID405), αCD138(hSTL002), αLGR5(YW353), αCD90(h5-Thy1) were subjected to IgBLAST analysis to compare the extracted sequences against existing immunoglobulin germline variable region gene sequences.

In addition, this IgBLAST analysis further confirmed the existence of intact CDRs and FR regions of each variable heavy chain (VH) and variable light chain (VL). Once this is performed, each scFv variable heavy chain (VH) and variable light chain (VL), were aligned to their parental sequences to compare and confirm CDR sequence homology, using CLC genomic workbench 11 software. After that, carboxyl terminus of each variable heavy chain (VH) is linked with the corresponding amino terminus of variable light chain (VL), using a flexible peptide linker consisting primarily of stretches of glycine (G) and serine (S) residues $(GGGGS)_3$ peptide designed by Huston et al., [208] to avoid serious steric interference of variable domains of scFvs.

Then, the newly designed scFvUBS54, scFvID405, and scFvh-STL002 sequences were separately inserted into plasmid encoding for mammalian expression host "pCB-Annexin V-SNAP expression plasmid, by replacing annexin V with each scFv to produce pCB- scFvUBS54-SNAP, pCB-scFvID405-SNAP, and pCB-scFvh-STL002-SNAP plasmids using SnapGene® software. (Fig 4A).

Whereas for construct scFvOCAb9-1, scFvYW353, and scFvh5-Thy1 nucleotide sequences were codon-optimized to *E. coli* expression host and then separately inserted into pMT-J3-SNAP expression plasmid, by replacing J3 with each scFv to produce pMT-scFvOCAb9-1-SNAP, pMT-scFvYW353-SNAP, and pMT-scFvh5-Thy1-SNAP plasmids using SnapGene® (Fig 4B). pCB vector is derived from mammalian expression vector backbone pMS,[209] these vector is further modified from pSecTag2[210]. pMT vector is successfully used for periplasmic stress expression of recombinant αEpCAM(scFv)-ETA' in BL21 *E. coli*.[211] This vector was derived from the pET27b vector (Novagen).

The pCB vector (Fig. 4A) used for mammalian expression and pMT vector (Fig. 4B) for bacterial expression of scFv-SNAP fusion proteins, was customized using the Thermo-Fisher GeneArt® Elements™ vector customization system. The ORFs encoding the recombinant scFv-SNAPtag fusion proteins were designed using SnapGene® software. The (scFv) sequence was flanked by *Sfi*I and *Not*I restriction sites at the 5' and 3' ends, respectively. Finally, these scFv (UBS54, ID-405, h-STL002 and OCAb9-1, YW353, h5-Thy1) containing *SfiI* and *NotI* restriction enzymes sticky ends were sent to GenScript Biotech Corp (Nanjing, China) to be synthesized in commercial pUC57 plasmid. The synthesized commercial plasmid has an antibiotic (Ampicillin) resistant gene, and we have not done any restriction digestion and ligation in this commercial plasmid, therefore the plasmid where intact that enables selective growth of transformed *E.coli*

cells in Ampicillin containing media, thus eliminating the need of a single colony selection procedure.

The pCB vector (Fig. 4A), used for mammalian expression of (UBS54/ID405/h-STL002) scFv-SNAP fusion proteins, was customized using the Thermo-Fisher GeneArt® Elements™ portal to include the following sequences.

• *Ampicillin (Amp) resistance gene*: for selection of transformed 5-alpha *E. coli* cells.

• *Internal ribosome entry site (IRES)*: for translation initiation.

• *SV40 promoter*: drives gene expression in HEK293T cells, which stably express the SV40 large T-antigen.

• *CMV promoter*: drives expression of downstream genes in expression vector.

• *Mouse IgK VIII leader sequence*: signal for secretion of the recombinant protein.

• *Enhanced GFP*: for visual detection of plasmid expression.

• *Bleomycin resistance gene*: for selection for transfected cells in mammalian culture.

• *Poly(6x)-His tag*: for affinity chromatography and for detection by αHis immunoblotting.

• *Enterokinase cleavage site*: for removal of the N-terminal segment of the fusion protein.

The pMT vector (Fig. 4B), used for bacterial expression of (OCAb9-1/ YW353/h5-Thy1) scFv-SNAP fusion proteins, was customized using the Thermo-Fisher GeneArt® Elements™ portal to include the following sequences.

• *Kanamycin (Kan) resistance gene*: for selection of successfully transfected recombinant clone.

• *T7 promoter*: drives expression of pelB peptide signal/leader sequence, which directs the fusion protein to the bacterial periplasm, where the sequence is removed by signal peptidase

• *Poly(10x)-His tag*: for affinity chromatography and for detection by αHis immunoblotting.

• *Enterokinase cleavage site*: for removal of the N-terminal segment of the fusion protein.

• *lac repressor (lacI protein)*: derives repression of target gene transcription by binding to lac operator and inactivates lac promoter. In the presence of lactose or lactose analog IPTG, the lac repressor undergoes a conformational change that removes it from lacO sites within the promoter and ceases repression of the target gene.

• *Ribosomal Binding Site (RBS):* also known as Shine–Dalgarno (SD) sequence located upstream of start codon recruit the ribosome to the messenger RNA (mRNA) and promotes efficient and accurate translation of mRNA.

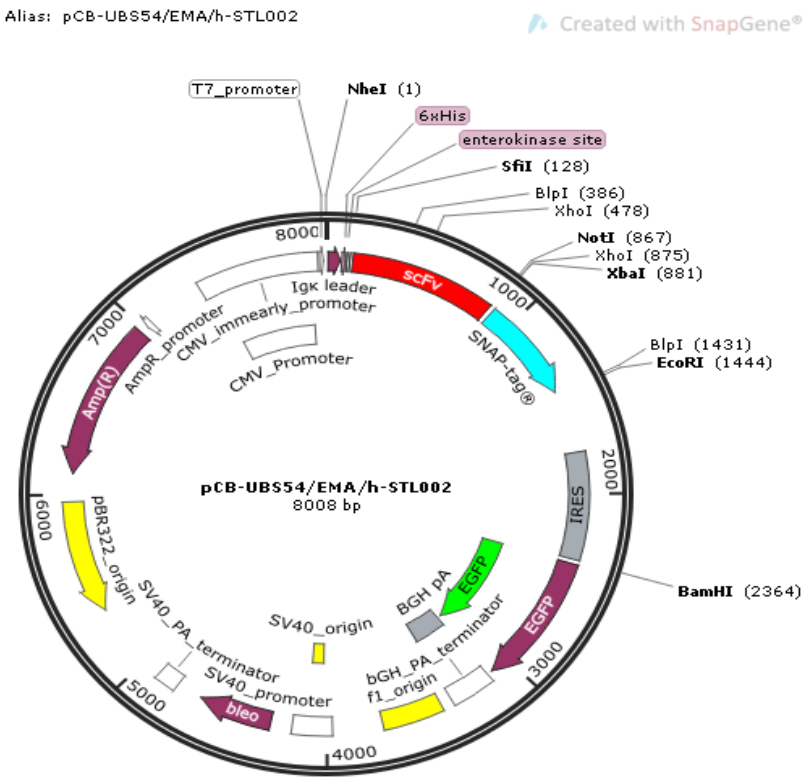


Figure 4A. Mammalian expression vector designed for targeted scFv-SNAP production. The pCB expression vector customized for the expression of recombinant fusion proteins in a HEK293T mammalian expression system, using ampicillin and bleomycin antibiotic selection and for monitoring of expression via GFP imaging. The image in this figure was generated using SnapGene® software.

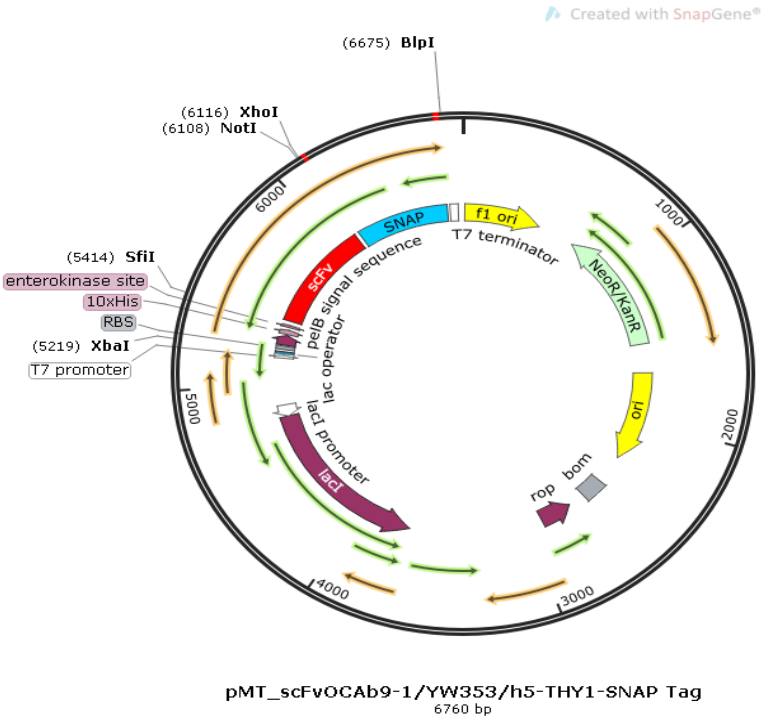


Figure 4B. Bacterial expression vector designed for targeted scFv-SNAP production. The pMT expression vector customized for the expression of recombinant fusion proteins in a *E.coli* (BL21-DE3) expression system, using kanamycin antibiotic selection the detection of successfully transformed recombinant clone is selected. The image in this figure was generated using SnapGene® software.

2.2.3. Removing or altering cryptic translation initiation site within the ORF of αEpCAM (scFvUBS54)-SNAP

Protein-coding sequences have traditionally been defined as uninterrupted ORFs that begin with the universal AUG start codon and end with one of three stop codons (UAA, UGA, and UAG). On construction of recombinant plasmid, it is indispensable to cross-check the ORF sequence within the plasmid and screen for factors that could interrupt the ORF in translation such as (i) internal stop codon (UAA, UGA, and UAG), (ii) non-AUG start codons, e.g., CUG, UUG, GUG, ACG, AUA, and AUU,[212] iii) Kozak motif "A/GXXaugG".[213,214] As these factors individually or in combination may contribute to the generation of truncated protein by ribosome stalling and leaking. With the use of SnapGene (https://www.snapgene.com/) a molecular biology software for in silico gene analysis, molecular cloning, and PCR developed by GSL Biotech, USA. we examined the ORF for internal stop codon and screened for non-AUG start codon and Kozak

consensus sequence "A/GXXaugG" within the ORF. If ORF appear with an internal STOP codon, it can be eliminated by changing the nucleotide codon without affecting the amino acid sequence. However, the ORF might have a Kozak sequence with varying scores [weak Kozak (score<80) and strong Kozak (score>80)], a stronger Kozak sequence should be replaced as a weaker Kozak sequence or remove Kozak sequence without affecting amino acid sequence. The list of strong and weak Kozak sequences is kindly provided by Aurelio A. Teleman[215].The consensus Kozak sequence within the ORF of αEpCAM-(scFvUBS54)-SNAP is identified using "SnapGene® software and compared it with the list of Kozak sequences with varying strength scores.

2.3. MOLECULAR CLONING OF MAMMALIAN AND BACTERIAL EXPRESSION PLASMID

2.3.1. Transformation of *E. coli* DH5α by heat shock

Chemically (calcium) competent *E. coli* DH5α cells (50 µL aliquots) are thawed on ice and mixed gently with 0.5 µL of 100 ng/µL plasmid DNA [(pCB-AnnexinV-SNAP or pUC57-scFvUBS54 or pUC57-scFvID405 or pUC57-scFvh-STL002) and (pMT-J3(scFv)-SNAP or pUC57-scFvOCAb9-1 or pUC57-scFvYW353 or scFvh-5Thy1)]. The cell–DNA mixture is incubated for 30 min on ice and then heat-shocked at 42°C for 30 s; afterward, the bacteria are transferred back on ice for 5 min. Lastly, 950 µL of SOC (Super Optimal Broth with catabolite repression) (NEB #B9020S) is added to the cells and the entire 1 mL is inoculated into 50 mL Luria Broth (LB) or Luria–Bertani (LB) broth flask containing (200 ng/µL) ampicillin antibiotics for plasmids pCB-AnnexinV-SNAP, pUC57-scFvUBS54, pUC57-scFvID405, scFvh-STL002, pUC57-scFvOCAb9-1, pUC57-scFvYW353, scFvh-5Thy1 and kanamycin for plasmid pMT-J3(scFv)-SNAP and incubated on a shaker at 37°C overnight.

2.3.2. Isolation of plasmid DNA from *E. coli* DH5α

Plasmid DNA (pCB-AnnexinV-SNAP or pUC57-scFvUBS54 or pUC57-scFvID405 or scFvh-STL002) and (pMT-J3(scFv)-SNAP or pUC57-scFvOCAb9-1 or pUC57-scFvYW353 or scFvh5-Thy1) is purified from *E.coli DH5-α*, using the NucleoBond plasmid kit/NucleoBond PC 100 Midiprep kit (740573, Macherey-Nagel) "Midi protocol" according to the manufacturer's instructions and stored at 4°C.

2.3.3. Restriction DNA

During this protocol, plasmids DNA (pCB-AnnexinV-SNAP or pUC57-scFvUBS54 or pUC57-scFvID405 or scFvh-STL002) and (pMT-J3(scFv)-SNAP or pUC57-scFvOCAb9-1 or pUC57-scFvYW353 or scFvh5-Thy1) are successively digested with restriction enzymes SfiI (for 3 h at 50^{o}C) and NotI (overnight at 37^{o}C) in New England Biolabs master mix buffer reaction, as illustrated in **Table 6**.

On completion of the restriction enzyme digest, the digested DNA is analyzed and separated by agarose gel electrophoresis using 1.2% (W/V) agarose gel (100 ml) supplemented with 8ul SYBR safe nucleic acid staining (ThermoFisher Scientific, catalog number: S33102) in 1xTAE buffer using wide mini sub-cell systems (Bio-Rad, USA) set at 100 V for 60 min. Visualization of the DNA fragments is performed on exposure to blue light excitation. Finally, DNA is extracted from the agarose gel and purified using Nucleospin Extract II kit (Macherey-Nagel, Duren, Germany).

Table 6: Restriction enzymes digest reaction mixture

Reaction mixture	**pUC57-scFv/ pCB-AnnexinV-SNAP/ pMT-J3(vhh)-SNAP.**
*Sfi*I (3 h at 50^{o}C)	2000 U/mL
*Not*I (37^{o}C overnight)	2000 U/mL
(10×) NEB cut smart buffer	1× NEB cut smart buffer
DNA/Plasmid	2 µg
ddH_2O	adjust to the final volume
Final volume	50 µL

2.3.4. Ligation of DNA and transformation of recombinant clones in *E. coli*

Digested DNA fragments for mammalian expression (scFvUBS54 or scFvID405 or scFvh-STL002) are ligated to pCB-Annexin V-SNAP vector backbone or pCB-SNAP vector backbone, and fragments for bacterial expression (scFvOCAb9-1 or scFvYW353 or scFvh5-Thy1) are ligated to pMT-J3(vhh)-SNAP vector backbone or pMT-SNAP using T4 DNA Ligase (400,000 U/ml; New England BioLabs) and 10x ligase buffer as recommended in the manufacturer's

protocol in a final volume of 20 µl. Ligation reactions are achieved using three insert–plasmid vector molar ratios (1:1, 3:1, and 5:1) at 16°C overnight (**Table 7**). The next day, the transformation of *E. coli* DH5α with recombinant DNA is performed. During this experiment, 5µl of recombinant DNA is added to 50 µL of *E. coli* DH5α calcium competent cells, which are then incubated on ice and mixed thoroughly. Then, the cells are subjected to heat shock at 42^{o}C for 60 seconds and cooled on ice for 5 min respectively. Thereafter, 950µl SOC is added to the cells to produce a 1 mL mixture which is incubated at 37^{o}C for 60 min. Following that, the cells are mixed thoroughly, spun down and 800 µl of supernatant is removed, while the remaining pellet is resuspended into a final volume of 200 µl with LB. Finally, the 100µl of cells are plated onto LB agar plates supplemented with ampicillin (200 ng/µl) for mammalian expression construct and kanamycin (200 ng/µl) (including vector-only and bacterial controls) for bacterial expression construct and incubated at 37^{o}C overnight. The next day, several single colonies are picked and placed in 2 ml LB containing ampicillin for pCB construct and kanamycin for pMT construct and grown overnight at 37^{o}C on the shaker.

Table 7: DNA ligation reaction master mix using different insert-to-vector DNA molar ratios.

Component	**Final reaction Volume (20µl)**		
T4 DNA ligase buffer (10x)	1x T4 DNA ligase buffer		
Vector DNA (1:1) or (3:1) or (5:1)	50 ng	50 ng	50 ng
Insert DNA (1:1) or (3:1) or (5:1)	5 ng	15 ng	25 ng
T4 DNA ligase	4000,000 units/ml		
Nuclease-free water	Adjust to final volume 20ul		

Calculation of (scFvOCAb9-1: 687bp) insert and plasmid (pMT-SNAP: 6073bp) vector DNA mass was performed using the online New England Biolabs calculator (NEBio calculator) before ligation reaction.

Note: The calculation is almost similar for other inserts "scFvUBS54/scFvID405/scFvh-STL002, and scFvYW353/scFvh5-Thy1 and vector plasmids (pCB-SNAP: 7276bp) and (pMT-SNAP: 6073bp).

2.3.5. Restriction mapping, DNA sequencing, and analysis

To confirm successful isolation of recombinant bacteria clones, recombinant plasmids (pCB-scFvUBS54 or pCB-scFvID405 or pCB-scFvh-STL002-SNAP) & (pMT- scFvOCAb9-1 or pMT-scFvYW353 or pMT-scFvh5-Thy1-SNAP) were extracted from bacterial cells using Zippy plasmid mini kits (D4036 Zymo Research, USA) according to manufacturer's protocols. Thereafter, using SnapGene software, agarose gel simulation for recombinant and parental plasmid maps were generated and several restriction enzymes (BlpI, BamHI, PvuII) cutting within the insert fragment and the vector backbone was selected to generate different fragment sizes. This method enables digestion of the parental plasmid (pCB-AnnexinV-SNAP) and corresponding recombinant plasmids harboring the insert (scFv) and vector backbone (pCB-scFvUBS54/ID405/h-STL002-SNAP) for mammalian expression. Similarly, parental plasmid (pMT-J3-SNAP) of bacterial expression and its corresponding recombinant plasmids harboring insert and vector backbone (pMT-scFvOCAb9-1/YW353/h5-Thy1-SNAP) are digested with multiple restriction enzymes (PvuII, PstI), thus generating different patterns of digested fragments between the two plasmids. Finally, recombinant plasmid DNA sequences were determined by the Inqaba Biotec sequencing facility, and the recombinant DNA sequences were aligned to the in silico designed sequences in SnapGene to confirm successful insertion of the scFv DNA fragments in the pCB-SNAP vector and the pMT-SNAP vector.

2.4. EXPRESSION OF FUSION PROTEINS (scFv-SNAP-tag)

2.4.1. Cell culture

For all experiments, human embryonic kidney (HEK-293T) cell line (ATCC: CRL-3216^{TM}), were cultured in RPMI-1640 medium (Gibco #61870), supplemented with 2 mM L-glutamine (GlutaMAX^{TM}), 10% (v/v) fetal bovine serum (FBS) (Thermo-Fisher, USA) and 100 IU/mL penicillin (P3032 Sigma, USA) and 100 µg/mL streptomycin (S-91370 Sigma, USA) and incubated in 5% CO_2 with 95% humidity at 37^{o}C. The human cell line;MDA-MB-468 (ATCC:HTB-132), MCF-7(ATCC:HTB-22), HEK293T(ATCC:CRL-3216), PANC-1(ATCC: CRL-1469), Hs578T (ATCC:HTB-126) were cultured in Dulbecco's Modified Eagle's Medium (DMEM; Gibco #10566 Highveld Biological (Pty) Ltd., RSA), supplemented with 10% (v/v) heat-inactivated fetal bovine serum (Highveld Biological (Pty) Ltd., RSA), 100 IU/mL penicillin (P3032 Sigma, USA) and 100 µg/mL streptomycin (S-91370 Sigma, USA), and incubated in 5% CO_2 with 95% humidity at 37^{o}C.

2.4.2. scFvUBS54/ID405/h-STL002-SNAP-tag expression in mammalian (HEK293T) system

2.4.3. Optimizing transfection efficiency

2.4.3.1. DNA plasmid transfection using polymer transfection agent

HEK293T cells are transfected with constructed pCB-scFv-SNAP vectors using X-tremeGENE HP DNA transfection reagent (Roche, Germany). Before transfection, plasmids are purified using plasmid mini-kit (Roche, Germany) and added 0.3×10^6 cells/35 mm dish and allowed to reach 90% confluency (1.0×10^6 cell) by growing them at 37°C in a CO_2 incubator. Plasmid DNA is mixed with DNA transfection reagent at a ratio of 1:3 according to the manufacturer's instructions. The mixture of 3 µg plasmid DNA and 9 µl transfection reagents in the final volume of 200µl serum-free media (SFM) is prepared. After 20 min of incubation at room temperature, the complex solution is added dropwise to each dish and gently swirl the dish clockwise and counterclockwise. Since green fluorescent protein (GFP) is coexpressed on the same expression vectors, transfection efficiency in the transfected cell population is analzed by the flow cytometry method. All transfections are carried out in triplicate.

2.4.3.2. Analysis of transfection efficiency of DNA plasmid by flow cytometry

After 72–96 h of posttransfection, transfection efficiency and transient eGFP expression levels were analyzed by assessing the fluorescence intensity of transfected cells using flow cytometry. Positive cells expressing the target protein can be easily distinguished by the presence of intracellular green fluorescent protein, and transfection efficiency can be evaluated by flow cytometry. Briefly, after 72–96 h posttransfection the attached cells are harvested from a 35 mm dish by gentle aspiration with 2ml media. Aspirated cells are collected in FACS tube and centrifuged at 1000 rpm for 5 min and the supernatant is discarded, to the pellet 500ul of FACS buffer is added before analysis with a FACSCalibur flow cytometer (Becton Dickinson). The data are analyzed using the CellQuest software package provided with the instrument.

2.4.3.3. Effects on transfection efficiency of a modified protocol

Twenty-four hours before transfection, HEK293 cells (1×10^6 cells) are plated on a T25 flask containing RPMI with 10% FBS and 5% antibiotics and allowed to reach 90% confluency (2.6×10^6 cells) by growing them at 37 °C in a CO_2 incubator. On the day of transfection, 4.0 µg plasmid DNA is diluted with 560 µL of various solvents such as nuclease-free water, 1× PBS,

Opti-MEM® I (1×) reduced serum medium, and plain RPMI medium in a 1.5 ml sterile plastic tube. Controlled cells are processed with the same protocol, but in a plasmid DNA–free manner. Transfection lipoplex complex formation is prepared by adding 36 µL of X-treme gene to the 1.5 mL sterile plastic tube containing the mixture of 4 µg plasmid DNA to bring to the final volume of 600 µL to give a reaction mixture ratio of 1:9 (plasmid–transfection agent) and incubated for 20 min at RT. After 20 min 600 µL of the reaction mix is administered to the T25 flask cells followed by 6 h incubation at 37°C in 5% CO_2; growing medium was not removed from the T25 flask before transfection. However, after 6 h incubation growing medium in T25 flask is removed and replaced with fresh growth medium and left for transient expression of green fluorescent protein (GFP) is coexpressed on the same expression vectors with target protein through integration with vector or driving by internal ribosome entry sequence (IRES), transfection efficiency in transfected cell population was analyzed by flow cytometry method as described in Section **2.4.3.4**. All transfections are carried out in triplicate.

In parallel, we followed another method where the plasmid DNA concentration is increased 10-fold, i.e., 40 µg of plasmid DNA is added to 120 µL of X-treme gene to give plasmid: transfection agent ratio 1:3 in 1.5 mL sterile Eppendorf tube and brought the final volume of the reaction mixture to 1 mL and incubated for 20 min at RT. After 20 min 1 mL of the reaction mix was administrated to the T25 flask cells and incubated at 37°C and 5% CO_2, growing medium was not removed from the T25 flask before transfection and after transfection. After 72–96 h transfection, transfection efficiency in transfected cell population was analyzed by flow cytometry method as described in section (**2.4.3.4**).

2.4.3.4. Analysis of transfection efficiency of optimized protocol

Briefly, attached cells were harvested from T25 by gentle aspiration after 72hr-96hr post transfection. Collected cells are centrifuged in 15ml falcon tube and the supernatant is discarded, to the pellet 1ml of media is added and aspirated to dissolve the pellet, from dissolved pellet 100µL of cell suspension is taken in BD FACS tube with and without 7-AAD and cells were fixed in 2% paraformaldehyde before analysis with a FACS LSRII. The remaining 900µL suspension is put back to the T25 flask for continuous growth and maintenance and antibiotics selection of clones with transient expression. The numbers of eGFP-positive and eGFP-negative cells were calculated according to the flow cytometry results, and the transfection efficiency was represented by the ratio of the number of eGFP-positive cells to the total cell number. The percentage of live cells and dead cells are calculated based on uptake of the 7-AAD dye, as transfection agent has a considerable impact on cell survival. This method allowed us to easily

distinguish positive cells (labeled with green, fluorescent dye) that are alive and dead. We therefore used this method to evaluate transfection efficiency under various conditions in subsequent experiments. The GFP fluorescence profiles of each clone were examined using flow cytometry and an FACS LSRII. The 488 nm laser was used to detect GFP fluorescence (505 LP, 530/30). Forward scattering had a voltage setting of 82, side scattering of 181, GFP fluorescence of 381. Forward and side scattering data were linear, and fluorescence was collected on a logarithmic scale. The analysis of flow-cytometry data was performed using FlowJo version 7.6.

2.4.4. Zeocin Selection

Zeocin™ is a formulation of phleomycin D1, a basic, water-soluble, copper-chelated glycopeptide isolated from *Streptomyces verticillus* and shows strong toxicity against bacteria, fungi (including yeast), plants, and mammalian cell lines. On purchasing the Zeocin™ is inactive, due to presence of copper in the solution hence the solution is blue color. Once the antibiotics enters the cell the copper cation is reduced from Cu^{2+} to Cu^{1+} and removed by sulfhydryl compounds in the cell. On copper removal, Zeocin™ gets activated and binds and cleaves DNA, causing cell death.

A Zeocin™ resistance protein of 13,665 Da, has been isolated and characterized. The protein is the product of the Sh ble gene (*Streptoalloteichus hindustanus* bleomycin gene), binds stoichiometrically to Zeocin™ and inhibits its DNA strand cleavage activity. Expression of this protein in eukaryotic and prokaryotic hosts confers resistance to Zeocin™.

2.4.4.1. Standard procedure

Zeocin™ selection antibiotic is purchased at a 100 mg/mL stock concentration and used in cell culture to a 100 μg/ml final concentration (Thermo-Fisher, USA) by directly adding to cultured cells maintained in complete media[216].

2.4.4.2. Optimizing Zeocin selection

Before transfection, we determined the optimal antibiotic selection with the optimal concentration required to kill nontransfected HEK-293T. Interestingly, there are reports that HEK293 is susceptible to certain antibiotics but not its derivative HEK293T, so optimal concentration of antibiotic must be determined for a given cell to generate a kill curve. In addition, it is essential to estimate the minimal concentration of Zeocin to kill non transfected cells, as high zeocin concentration might create DNA strand breaks in the recombinant cells,

despite the stable expression of the Sh ble gene in the recombinant clones. This could lead to silencing of transgene. Moreover, chronic exposure of transfected cells, harboring Zeocin resistant gene to Zeocin might amenable to DNA damage (Mutagenesis), as Zeocin is not fully detoxified[217]. Therefore, it is uttermost important to determine the optimal concentration required to kill nontransfected HEK-293T and spare the transfected HEK-293T. After plating HEK293T in 96-well plates and allowing cells to attach overnight, Zeocin is added at a concentration ranging from 5 mg to 0.15 mg in 2-fold serial dilution and incubated for 72hr. Cell viability is assessed based on XTT assay. Repeating this selection process will result in cell clones with a high number of copies of the marker gene, co-amplification of the gene of interest, and high levels of protein expression.

2.4.5. Sandwich ELISA

The enzyme-linked immunosorbent assay (ELISA) is an immunological technique that is used to determine the presence or quantity of an antigen within a sample. Here, the antigen is SNAPtag fusion protein to be detected and quantified in HEK293T cultured cell-free supernatant. Briefly, maleic anhydride activated microplates are treated with a capture probe (BG-PEG12-NH_2 (1002 $g.mol^{-1}$) 20 ug/ml in immobilization buffer to each well that is specific for a protein of interest (SNAP-tag). All incubation steps were performed in the wells of a 96-well plate at a 200 µL scale at room temperature for 1 h or overnight at 4°C. The 96-well plates are washed with PBST three times after every incubation step. Before the addition of the capture probe, plates were washed 6 times with wash buffer (Product No. 28320-ThermoFisher) or 1× Elisa wash buffer (Product No.421601-BioLegend/Biocom). After the addition of the capture probe (first step) and completion of the incubation step, the plate was blocked with Protein Blocking Buffer (100 µL) per well (ThermoScientific Superblock Blocking Buffer, Product No. 37515 or 37535) and incubate the plate for 1 h at room temperature. This second step quenches remaining reactive maleic anhydride groups and blocks remaining open sites on the plate surface. As a third incubation step, the purified SNAPtag protein is used as standard (35µg/well as starting concentration, 2-fold serial diluted) and cell-free supernatant (100ul) as test samples are added to each well and incubation at room temperature for 3 h. In the fourth step, the HRP-conjugated primary antibody (Anti-6x His tag antibody (HRP) was added at a dilution of 1:5000 in PBS or in blocking buffer. Finally, 100 µl/well TMB substrate was added and the absorption at 450 nm or 510 nm was measured. The TMB reaction was stopped with 100 µL/well in 1 M HCl and the absorbance was measured at 450 nm.

2.5. PURIFICATION AND QUANTIFICATION OF MAMMALIAN EXPRESSED ANTIBODY BASED FUSION PROTEINS

2.5.1. IMAC purification

On successful transfection in mammalian cells (HEK293T), the recombinant fusion proteins (scFvUBS54-SNAP or scFvID405-SNAP or scFvh-STL002-SNAP) secreted by transfected HEK293T cell were collected as a cell culture supernatant (CCSN) and purified using IMAC as described below. Similarly, clear bacterial lysate for other recombinant proteins (scFvOCAb9-1 or scFvYW353 or scFvh5-Thy1) expressed by BL21-DE3 were collected and subjected to immediate purification via the C-terminal His-tag by fractional elution on an immobilized metal

affinity chromatography (IMAC), using HisTrap™ excel column containing 5ml Ni(2+) Sepharose® affinity resin (GE Healthcare, USA) equilibrated by using equilibration buffer (pH 8.00) on an ÄKTA protein purification system (GE Healthcare Europe GmbH, Freiburg, Germany).

After washing the column with 20 CV of equilibration buffer, His-tagged recombinant proteins bound to the stationary phase (the resin) were eluted by a stepwise increase in the concentration of high-imidazole elution buffer (pH 8.00). which competitively displaced the histidine of the recombinant protein on the column through interaction with nickel on the column. The titration of elution buffer was optimized for each fusion protein to obtain a narrow peak (maximum yield) and the column was cleared with a chaser of 100 % elution buffer, followed by washing and re-equilibration. The eluted fractions were subsequently concentrated and buffer exchange using a 10 kDa cutoff amicon filter (Ref: UFC901008, Merck, South Africa).

2.5.2. Protein quantification

2.5.2.1. UV–visible spectrophotometer

Direct protein quantification is performed using UV–visible Spectrophotometer /Fluorometer (DeNovix. DS-11) at 280 nm, with an estimated percent extinction coefficient (ε1%) of 6.67. The analysis was performed in triplicate on undiluted protein fraction for various construct and PBS were used as blank. The purity of the protein was determined by measuring the OD at 260/280 nm ratio of 0.6–0.7. The Beer–Lambert law was used for the calculation of the final protein concentration of each solution using the percent extinction coefficient (ε1%) of each protein for a wavelength of 280 nm (**Equation 1**).

$$C = \frac{A280}{\epsilon 1\%/} \qquad \text{Equation 1}$$

2.5.2.2. BCA assay

Purified protein concentration is determined using ready-to-use kits: The Pierce™ BCA protein assay kit (Thermo Fisher Scientific, Waltham, MA, USA; 23225) according to manufactures instructions. The stock concentration of BSA 2000 µg/mL is serially diluted in the concentration range from 1 to 2000 µg/mL are used to generate the linear regression standard curve by calibrating nine-point in triplicate. scFv-SNAP purified total protein of 20 µl is used as a test sample in triplicate and PBS as a diluent. The analyses are performed in triplicate. After a 30 min incubation of the samples with working reagent (WR) at 37°C, the absorbance of the samples

(both standard and test sample and appropriate blank at 562 nm.) is measured on a spectrophotometer (Bio-Rad, iMark^{TM}) at 562 nm.

2.5.2.3. Bradford assay

Total cellular proteins after IMAC purification were concentrated using Amicon® Ultra-15 Centrifugal filter, following centrifugation at 12,000 × *g* at 4°C for 45 min, the total protein concentrations of the supernatants were determined using Bradford Protein Assay (Bio-Rad). A set of BSA (ThermoFisher) standards at 10.0, 5.0, 2.5, 1.25, and 0.6 µg/well in PBS was used in 96-well plate and another 10 µl of the PBS solution (BSA solution concentration of 0 mg/ml) was used as a control, 10 µl of a test sample in triplicate is added to the remaining well. 200 µl of diluted dye reagent (Coomassie Brilliant Blue G-250 dye) is added to each well and mixed with a multichannel pipette. Incubate the plate at room temperature for at least 10 min and not more than 1 h. After incubating the plate for 10 min at 37°C, the optical density values of the BSA standards were measured at 595 nm using a microplate reader. A standard curve was generated, and the protein concentrations were calculated based on the obtained equation.

2.5.3. Estimating the integrity and purity of protein by SDS and WB

Discontinuous 10% SDS-PAGE was used for the analysis of both mammalian and bacterial expressed purified recombinant protein samples obtained after IMAC purification and amicon concentration. Before loading onto the gel, recombinant protein samples (20µg) were diluted in PBS (pH 7.4) with (5µl) 5x protein loading dye and heated at 95°C for 5 min. Thereafter, the recombinant proteins were sequentially separated by electrophoresis at 120V for 1hr or until the dye front reaches the bottom of the gel. Protein bands were visualized by staining gel with Acqua staining solution (Vacutec, South Africa) for 20 min or silver staining solution (Thermo Fisher Scientific, catalogue number: 24612), according to manufacturer protocols. Additionally, transfer of recombinant protein from unstained gel to nitrocellulose membrane (Roche, Switzerland) was performed at 100V for 1 h 30 min using the electro-tank blotting, followed by blocking with milk for 1 h at room temperature. After blocking, the membrane was washed three times with 1x PBS/Tween. The recombinant proteins were detected using rabbit monoclonal anti-his-tag (1:1000, Qiagen, Hilden, Germany) and Goat anti-rabbit horseradish peroxidase-conjugate antibody (1: 5000) (Cat: 170-6515, Bio-Rad Laboratories, USA). Finally, SuperSignal West Femto chemiluminescent substrate was used for detection. Protein was filter-sterilized and quantified using a UV–visible spectrophotometer (Denovix, Alliance Global).

2.5.4. Densitometric analysis to quantify protein band of interest

Bovine serum albumin (BSA) is a cheap, well-known protein commonly used in protein quantification (Bradford, 1976) as a standard. BSA stock 2 mg/mL is serially diluted in PBS in the following concentration 7.9, 2.6, 0.8, and 0.2 μg and loaded in the first 4 lanes of SDS-PAGE from the order of left to right. Various scFv-SNAP purified protein is used as a test sample in remaining lanes with total protein concentration estimated by UV–visible spectrophotometer (DeNovix. DS-11) at a wavelength of 280 nm. lane 5; h-STL002-SNAP (3.2 μg), lane 6; UBS54-SNAP (3.0 μg), lane 7; ID405-SNAP (4.3 μg), lane 8; rGFP, lane 9; UBS54-SNAP (11.8 μg) to the total volume of 10 μL. In addition, lane 10 is loaded with 5 μL of the ready-to-use precision plus protein unstained protein standards (Bio-Rad) was used as molecular weight markers. Before loading the protein sample into the well all protein samples were diluted at a ratio of 4:1 with Laemmli 4× buffer solution (Bio-Rad) with 5% 2-mercaptoethanol (Sigma-Aldrich) as a denaturing agent and heated in a water bath at 90°C for 10 min. For the SDS-PAGE, the ready-to-use precision plus protein unstained protein standards (Bio-Rad) were used as molecular weight markers. The marker contains ten recombinant protein bands of 250, 150, 100, 75, 50, 37, 25, 20, 15, and 10 kDa. According to the manufacturer, the protein bands of 75, 50, and 25 kDa are reference markers within the molecular weight marker, as they have three times the intensity of the other bands.

Electrophoresis was run at 200 V for 60 min in a Mini-Protean Tetra cell (Bio-Rad) using 1x Tris-Buffered Saline (1× TBS), TGX running buffer (Bio-Rad). Gels were washed with distilled water and stained with AcquaStain a single-step Coomassie blue protein gel dye for 1 h. Gels were imaged in a molecular imager ChemiDoc XRS System (170-8070, Bio-Rad) under white light epi-illumination. Images were saved as a TIFF file with a size of 8-bit and a resolution of 1,392 × 1,040 pixels. The bands obtained were analysed by ImageJ a public domain program from the national institutes of health that allows image processing. Subsequently, a correlation analysis was carried out for the peak that is corresponding not to the total protein, but for a band of interest in standard and test to define the area of the curve to quantify the protein bands of interest. The average relative lane density of the total protein (Standard BSA) loaded in a linear dynamic range that extends from high to low signal intensities give a trend line with **R2 values >0.95** (c). To solve for the quantity of target proteins, i.e., the concentration of test for bands of interest by using the trendline formula that is generated using standard curve is displayed in result section (section 3.6.6 figure 62).

2.6. VALIDATING THE FUNCTIONAL ACTIVITY OF FUSION PROTEIN

2.6.1. Conjugation of scFv-SNAP with BG-modified Alexa 488

Initially purified scFv-SNAP fusion proteins of 5 µM were conjugated with 10 µM of BG-modified Alexa 488, in the presence of 1 mM DTT and 1xPBS by incubation in the dark for 1 h at room temperature or as described by manufacturer recommendation. To further optimise the concentration of labelling reagent, we used various concentration of BG-alexa 488 (1, 2.5, 5 µM) to conjugate with purified scFv-SNAP fusion proteins of 5 µM in the presence of 1 mM DTT by incubation in the dark for 1 h at room temperature. Labelled proteins were visualized after separation by SDS-PAGE, and visualization of the fluorescence signal was performed using blue light from dark reader transilluminator DR-89X system (Inqaba Biotec, SA). The coupling efficiency was photometrically determined, using the theoretical extinction coefficients of the proteins and the extinction coefficients of the fluorescence dyes.

2.6.2. In vitro binding study of labeled scFvUBS54/ID405/h-STL002-SNAP fusion protein

2.6.2.1. Confocal Microscopy

The in vitro binding activities of the labeled scFv-SNAPtag fusion protein are determined using confocal microscopy live imaging. The $EpCAM^+$ and EMA^+ cell lines MCF-7 are used to confirm the functional activity of the EpCAM-targeting scFv-UBS54-SNAP-tag fusion protein (figure 71) and EMA-specific scFv-ID405-SNAP-tag fusion protein. HEK293T negative for EpCAM and EMA is used as a control, Finally $CD138^+$ cell line HEK293T and HS578T is used to confirm binding of the scFvh-STL002-SNAP-tag fusion protein. All adherent cells are detached from the culture flask by adding 2 ml of Accutase solution or 1x TE buffer and collect the cells into 8 mL of growth medium. Determine the number of the cells using an automated cell counter and dispense 4×10^5 cells into four chamber 35 mm glass-bottom dish and allow them to attach and grow for 24–48 h. After 24 h wash the cells twice with 300 µL of plain RPMI and add 1 µg (5 µM) of the respective fluorescence-labeled fusion protein and incubate the cells for 30 to 40 min on ice for a binding study and then wash two times with plain RPMI and add 5 µg/ml of Hoechst 33342 fluorescent nuclear counterstain and incubate for 10 min on ice. Finally,

wash twice with 300 µL RPMI or phenol red–free incomplete media. Immediately visualize the live cells and monitor the cell membrane staining with the Zeiss LSM880 Airyscan with Fast Airyscan module confocal microscope equipped with the following lasers: Argon laser with 458 nm; 488 nm and 514 nm lines; solid-state laser 561 nm and Helium-neon laser for far-red imaging, 633 nm. Also, a Mai Tai infrared pulsed laser using the blue channel for Hoechst 33342 dye imaging and the green channel for Alexa Fluor 488 dye imaging.

2.6.2.2. Flow cytometry

The binding efficiency of anti-scFvUBS54-SNAP, scFvID405-SNAP fusion proteins to the respective antigen (EpCAM, EMA) positive cell line (PANC-1, MCF-7), and negative cell line (HS578T, HEK293T) was determined by flow cytometry, FACS LSRII (Becton Dickinson, Heidelberg, Germany) and data were analysed using FlowJo software (CyFlo Ltd, Turku, Finland). A total of 3×10^5 cells were washed with PBS and incubated with 3µg of purified protein labeled with Alexa 488 in 100µl of plain RPMI for 40 min at 4 °C. After 40 min of incubation, cells are incubated for 7 mins with and without 7-AAD (2µl). Finally, cells are washed twice in 1× PBS, resuspended the pellet in 100 µl PBS, with an equal volume of 1% PFA incubated in dark for 5 min at 4°C and then washed gently by adding 1ml PBS or FACS buffer containing 1× PBS and 2% FBS by centrifuging at 2000 RPM for 5 min. Discard the supernatant and resuspend pellet in 100–300 µL PBS or staining buffer. (Final volume depends on your cell concentration.) The resuspended pellet are stored in ice under minimal light/dark condition and then acquired and analysed using flow cytometer, within 1hr from the preparation. The saturation analysis was performed as described previously.[218] Briefly, three different concentrations (3 µg, 12 µg, 18 µg) of ID405-SNAP-Alexa488' were incubated for 1 h at 4 °C with MDA-MB-468 cells followed by 10 min incubation at 4°C with and without 7-AAD. Finally, cells are washed twice in 1× PBS, resuspended the pellet in 100 µl PBS, with an equal volume of 1% PFA incubated in dark for 5 min at 4°C and then washed again gently by adding 1 mL PBS or FACS buffer (PBS + 2% FBS) and by centrifuging at 2000 RPM for 5 min. Discard the supernatant and resuspend pellet in 100–300 µL PBS or staining buffer. (The final volume depends on your cell concentration.) The resuspended pellet are stored in ice under minimal light/dark condition and then acquired and analysed using flow cytometer, within 1hr from the preparation.

2.7. IN VITRO CYTOTOXIC STUDY OF SCFV-SNAP-AURIF FUSION PROTEIN

2.7.1. Generation of monomethyl auristatin-F-SNAP fusion proteins

Briefly, the purified UBS54scFv-SNAP proteins were incubated with a twofold molar excess of BG-modified AURIF in 1× PBS (BG-modified by Prof. Roger Hunter and AURIF sourced by Brightgene Bio-Medical Technol, China) for 2 h at room temperature. The remaining 1449 unconjugated BG-AURIF were removed using, Zeba™ spin Desalting column 7 kDa MWCO (89893, Thermo Scientific, USA) according to the manufacturer's protocol.

2.7.2. Cytotoxic analysis of UBS54scFv-SNAP-AURIF fusion protein on EpCAM positive and negative cell lines

Briefly, 5×10^3 (MCF-7, MDA-MB468, MDA-MB-231) cells are seeded 48 h prior in 96-well plate and allow to adhere and grow at 37ºC and 5% CO_2. The following day, these cells are incubated with decreasing concentrations of BG-AURIF (2000 nM) alone and UBS54scFv-SNAP-AURIF (4000, 2000, 1000, 500 nM) for 72 h in triplicate. At 68 h posttreatment, electron coupling reagent "phenanzine methosulfate" and XTT labeling/solubilizing reagent (containing yellow tetrazolium salt) are combined according to the manufacturer's instructions and immediately 50 µL of the mixture is applied to each experimental well. After application, plates are incubated at 37 °C for a further 4-6 h, following which colorimetric absorption of the metabolic product (orange formazin crystals) is measured at 450 nm, with reference wavelength of 650 nm, on an iMark spectrophotometer using MPM 6 software (Bio-Rad, USA). IC_{50} values are calculated relative to PBS/RPMI-treated cells, control cells, and Zeocin-treated cells; statistical analysis of nonlinear regression is done using GraphPad Prism v5 software. All experiments are carried out in triplicate.

CHAPTER 3: RESULTS

The broad context of this work pertains to screening for differentially overexpressed BCC-specific cell surface antigens by literature search, and optionally by computational approach, and subsequently design and synthesize a range of SNAP-tag–based antibody fusion proteins to target these selective gene-encoding products in an antigen-positive cancer cell line, thereby fulfilling the search of a state-of-the-art technology that could be later applied in targeting PDX-BCC tumor models to reduce the time needed for surgical excision of BCCs by single excision. Down the line this state-of-the-art technology can be further developed into a systemic or topical application modality by switching the coloring/fluorescence substance with a cytotoxic drug to treat invasive cases of BCC where surgical intervention is not possible.

Chapter 3 is divided into two parts.

I. The first part (**Section 3.1**) describes the Identification of BCC-specific cell surface antigens.
II. The second part (**Section 3.2**) demonstrates the design, generation, and characterization of all the fusion proteins used in this study, and finally the biological activity of the purified targeted SNAP-tag–based antibody fusion proteins in the corresponding antigen-positive cell lines.

3.1. IDENTIFICATION OF BCC-SPECIFIC TUMOR-ASSOCIATED ANTIGENS

Two different approaches were followed to identify BCC-specific tumor associated antigens: (1) computational approach and (2) literature search, with the primary aim to find cell surface antigens that had not been reported yet. A recent study by Gaiser et al. reported loss of EpCAM expression at the invasive front of BCC lesions,[105] which might impede the complete excision of the lesions when EpCAM alone is considered as a target in guided excision of the lesions; therefore, there is a pressing need to find alternative tumor-associated antigens in both invasive and non-invasive BCC lesions. The literature search and alternative computational tool analysis

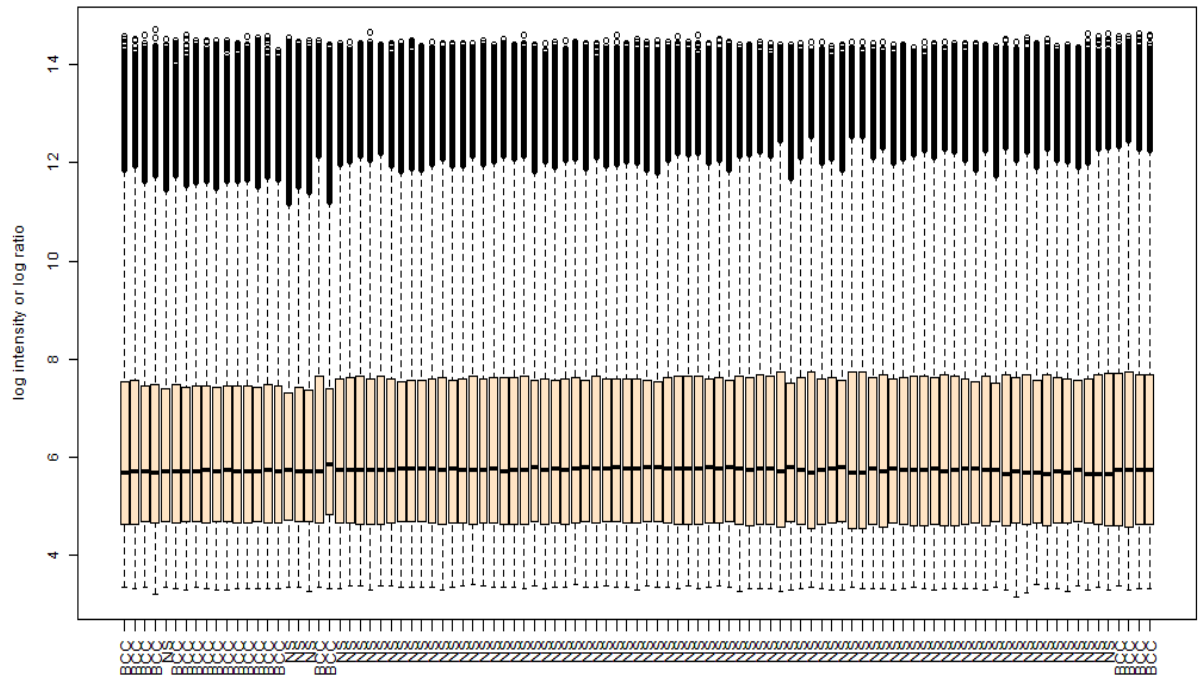
("Transcriptome Analysis Console") helped us to compare the BRB-Class comparison outcome list with published information to further validate the robustness of the computational analysis.

3.1.1. Identification of differentially expressed genes between BCC and NS by Integrated reanalysis of published microarray datasets.

In this computational study, a gene set that correlated with basal cell carcinoma samples was first identified from the gene expression profiling of four different datasets (GSE7553, GSE103439, GSE42109, GSE39612), as described in the methods section (see **Section 2.1.1**) To decrease bias from chip types, we selected the Affymetrix human genome U133 Plus 2.0 array as chip assay because several BCC and normal skin samples are assayed with HG-U133_Plus_2 and stored in the GEO database. In addition Affymetrix arrays use set of different 25-mer probes synthesized in situ to characterize gene expression of thousands of genes [193]. Microarray data encompass information on more than ≈22,000 genes, whereas conventional quantitative PCR would provide information about only a handful of genes, which might be in single digit or double digit.[219] Analyzing thousands of gene by microarray and its relation by clustering analysis, pathway analysis will further pave way to find hub genes that could have profound effect on cellular phenotype and function when suppressed or induced by siRNA. CRISPR-CAS, CRISPR activator (CRISPRa), small activating RNAs (saRNAs).[220,221]

To increase the utility of microarray analysis, the background adjustment is made for these datasets by merging the gene expression array of basal cell carcinoma and normal skin, and then the data are log-2 transformed with Quantile normalization, as described in **Section 2.1.2** and shown in **Fig. 5**. In **Fig. 5** the black horizontal lines in the colored box indicate the median expression value, one could observe that the boxes are more aligned at their medians on the bottom plot, compared with top plot, which indicates Quantile normalization imposes the same empirical distribution of intensities on each array after good batch effect removal in 21 BCC and 80 normal skin samples. Whereas in the top plot before normalization that the boxes are not aligned at their medians indicates there is no uniform distribution of intensities on each array as batch effect is not normalized, which might increase false-positive/false-negative results in the analysis. The merged dataset comprised 21 BCC samples and 80 normal skins. For further computational data preprocessing analysis, certain criteria are followed such as genes were excluded if <20% of the expression data had 1.5-fold change (in either direction) compared to the gene's median value, the missing or filtered out data was >50% and log intensity variation

p-value >0.01. After date preprocessing and normalization, we found 7553 gene among the 54,613 probesets passes the gene filters and subsetting criteria, To identify gene that are differentially expressed between predefined classes "BCC vs Normal Skin" first we run the BRB-ArrayTools with a random variance version of the t-test or F-test separately for each gene and obtained 3745 genes with the *p*-values below the threshold (*p*-value<0.001) (Complete table could not be shown due to its large size, however this data are available on the MB&I server and can be provided upon request) are shown in **Fig. 6** the dotted line represents the *p*-value of 0.001, genes with high significance-dots that are blue, above, and farther away from the *p*-value line are highly significant when compared with the blue dots, above, and closer to the *p*-value line. Of these 3745 genes, 213 genes are threefold upregulated and 188 genes are threefold downregulated with significant *p*-value and FDR <$1e^{-07}$, as shown in **Fig. 7;** the 213-green circles above the top black line represent the genes that are ≥3 fold up in BCC and 188 green circles below the bottom black line indicate genes that are ≥3.0-fold down in BCC compared with the control (normal skin). Therefore these 401 DEG genes are statistically significant differential gene between diseased (BCC) and normal which are considered for downstream analysis of choosing cell surface genes.



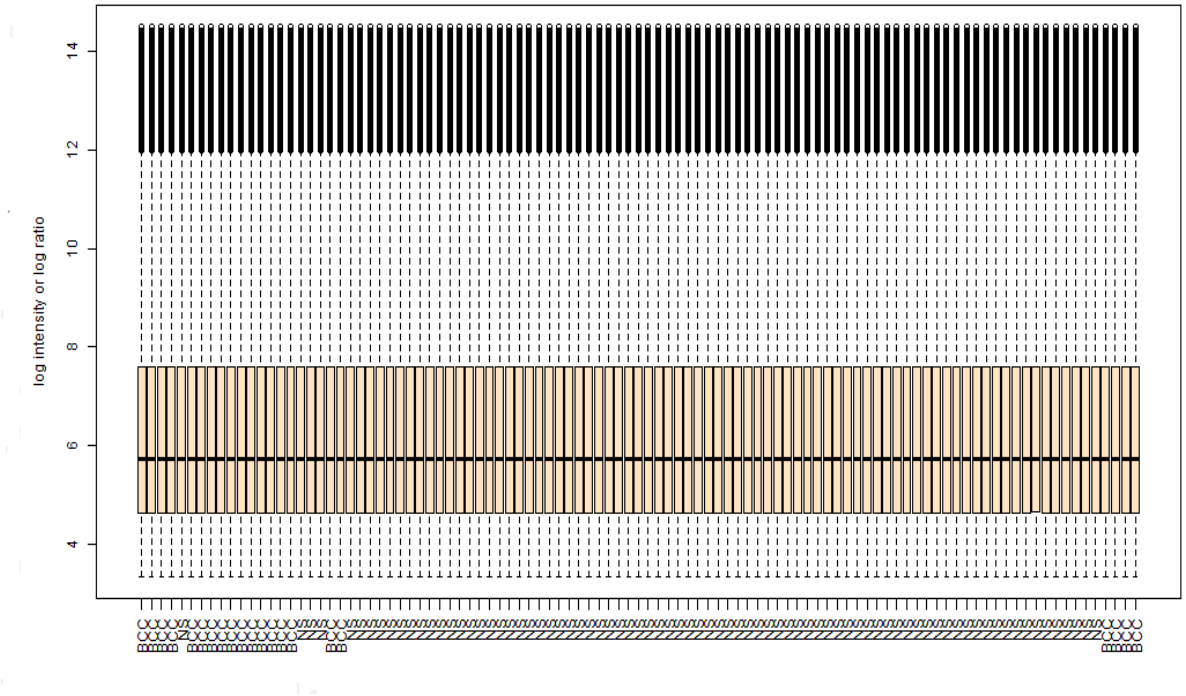


Figure 5. Box plots of gene expression data before normalization (top) and after normalization (bottom) of 21 BCC and 80 Normal skin arrays of the combined dataset (GSE7553, GSE103439, GSE42109, GSE3961). The horizontal axis, sample; vertical axis, expression value. The box stretches from the lower hinge (defined as the 25th percentile) to the upper hinge (the 75th percentile) The black horizontal lines in the colored box indicate the median expression value. Observe that the boxes are more aligned at their medians on the bottom plot, *which* indicates Quantile normalization imposes the same empirical distribution of intensities on each array after good batch effect removal in 21 BCC and 80 normal skin samples.

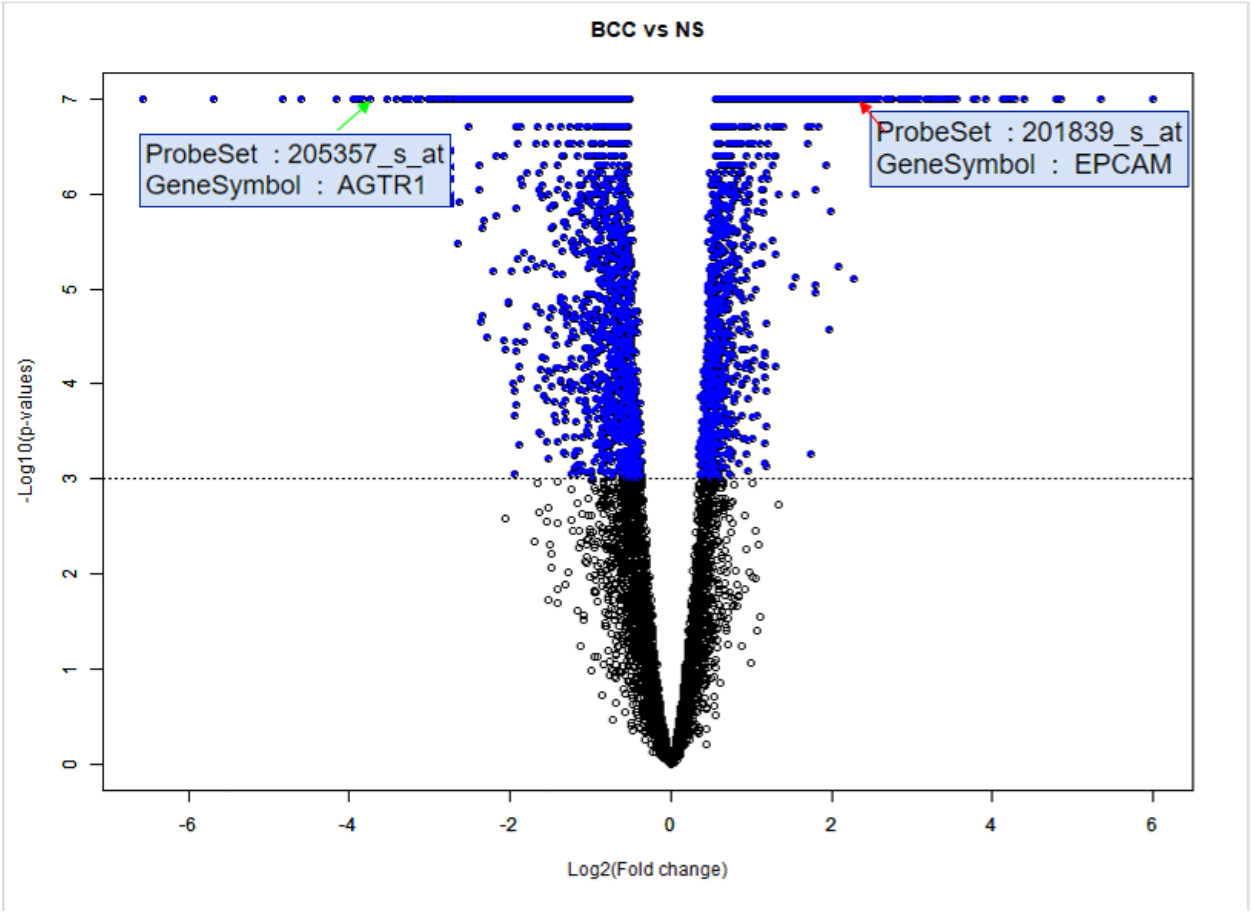


Figure 6: Volcano plots are a graphical representation of the results of t-tests. The x-axis is the log of the fold change "biological effect" and the y-axis is the log of the p-value "statistical effect." The dotted line represents the *p*-value of 0.001. These plots allow for the visualization of genes with high significance-dots that are blue, above, and farther away from the *p*-value line are highly significant when compared with the blue dots, above, and closer to the *p*-value line. Each dot corresponds to a probe set which corresponds to a specific gene. The blue dot to the left represents the downregulated gene and the right represents upregulated gene in the disease (BCC). The green arrow points to one of the 3-fold downregulated genes and the red arrow point to one of the threefold upregulated genes in a disease condition.

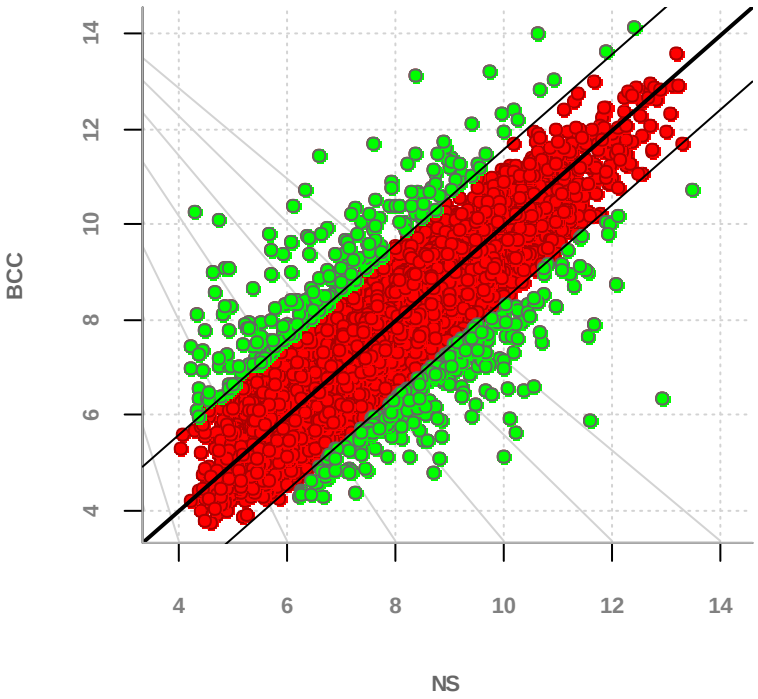


Figure 7: Scatterplot of mRNA signal value variation between the control-normal skin sample (x-axis), and basal cell carcinoma (y-axis). The values of the x-axis and y-axis in the scatterplot are the averaged normalized signal values of BCC and the control (log2 scaled). The black lines are folded change lines. The 213-green circle above the top black line and 188-green circle below the bottom black line indicate gene that are ≥3.0-fold change difference between BCC and the control (normal skin).

3.1.2. Screening for BCC associated cell surface markers

To rule out the gene that encodes for cell surface protein, these 401 genes, i.e., genes with 3-fold change differentially expressed and having *p*-value and FDR value $<1e^{-07}$ are compared with 2,799 surfaceome markers retrieved from Fluck et al.,[204] to uncover the possible targetable cell surface protein, as described in **Section 2.1.3**. Subsequently, we found 63 differentially expressed gene that encode for cell surface protein within these 401 DEG (**Fig. 8**). Interestingly of these 63 differentially expressed cell surface protein-encoding genes, 34 genes where 3-fold upregulated in BCC and 29 genes are 3-fold downregulated in BCC (**Table 8**). To express these results intuitively, the 63 DEGs were visualized with Heatmap (**Fig. 9**) and the gene of interest for further downstream analysis was pointed in red arrow.

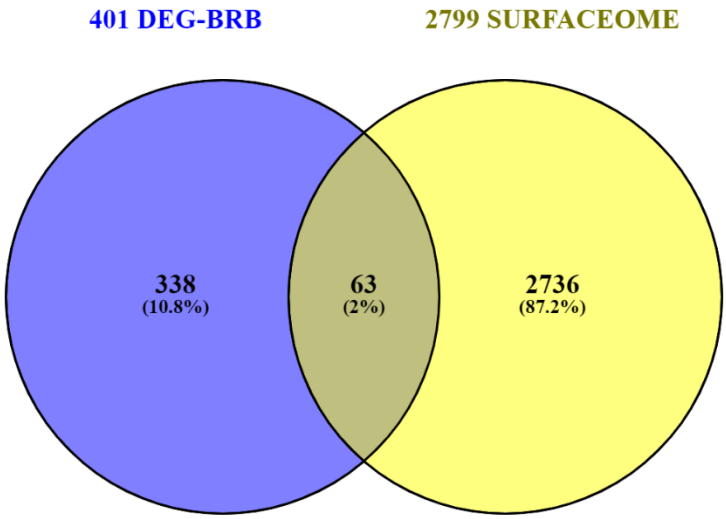


Figure 8: Venn diagram of intersections of 401 differentially expressed genes (3-FC, *p*-value and FDR **< 1e^{-07}**) with 2799 cell surface protein–encoding genes obtained from surfaceome. We found 63 genes are common between these two categories.

Table 8: 63 cell surface protein–encoding gene that are differentially expressed between BCC and NS analyzed using BRB-ArrayTools

p-value	FDR	FC	ProbeSet	Symbol	Name
< 1e-07	< 1e-07	17.86	236029_at	**FAT3**	FAT atypical cadherin 3
< 1e-07	< 1e-07	10.9	214297_at	**CSPG4**	chondroitin sulfate proteoglycan 4
< 1e-07	< 1e-07	10.78	208212_s_at	**ALK**	anaplastic lymphoma receptor tyrosine kinase
< 1e-07	< 1e-07	9.01	226145_s_at	**FRAS1**	Fraser extracellular matrix complex subunit 1
< 1e-07	< 1e-07	8.36	209815_at	**PTCH1**	patched 1
< 1e-07	< 1e-07	7.96	208433_s_at	**LRP8**	LDL receptor related protein 8
< 1e-07	< 1e-07	7.61	213880_at	**LGR5**	leucine-rich repeat containing G protein-coupled receptor 5
< 1e-07	< 1e-07	6.47	212110_at	**SLC39A14**	solute carrier family 39 (zinc transporter), member 14
< 1e-07	< 1e-07	6.03	203903_s_at	**HEPH**	hephaestin
< 1e-07	< 1e-07	5.79	214104_at	**GPR161**	G protein-coupled receptor 161
< 1e-07	< 1e-07	5.71	201195_s_at	**SLC7A5**	solute carrier family 7 (amino acid transporter light chain, L system), member 5
< 1e-07	< 1e-07	5.7	239572_at	**GJA3**	gap junction protein alpha 3
< 1e-07	< 1e-07	5.22	203256_at	**CDH3**	cadherin 3, type 1, P-cadherin (placental)
< 1e-07	< 1e-07	5.21	231867_at	**TENM2**	teneurin transmembrane protein 2
< 1e-07	< 1e-07	5.17	223278_at	**GJB2**	gap junction protein beta 2

< 1e-07	< 1e-07	5.06	203509_at	**SORL1**	sortilin-related receptor, L(DLR class) A repeats containing
< 1e-07	< 1e-07	4.99	204984_at	**GPC4**	glypican 4
< 1e-07	< 1e-07	4.77	207602_at	**TMPRSS11D**	transmembrane protease, serine 11D
< 1e-07	< 1e-07	4.72	201839_s_at	**EPCAM**	epithelial cell adhesion molecule
< 1e-07	< 1e-07	4.51	231771_at	**GJB6**	gap junction protein beta 6
< 1e-07	< 1e-07	4.23	227307_at	**TSPAN18**	tetraspanin-18
< 1e-07	< 1e-07	3.85	217028_at	**CXCR4**	chemokine (C-X-C motif) receptor 4
< 1e-07	< 1e-07	3.81	239229_at	**PHEX**	phosphate regulating endopeptidase homolog, X-linked
< 1e-07	< 1e-07	3.78	225516_at	**SLC7A2**	solute carrier family 7 (cationic amino acid transporter, y+ system), member 2
< 1e-07	< 1e-07	3.78	229649_at	**NRXN3**	neurexin 3
< 1e-07	< 1e-07	3.64	218856_at	**TNFRSF21**	tumor necrosis factor receptor superfamily, member 21
< 1e-07	< 1e-07	3.64	203706_s_at	**FZD7**	frizzled class receptor 7
< 1e-07	< 1e-07	3.59	226777_at	**ADAM12**	ADAM metallopeptidase domain 12
< 1e-07	< 1e-07	3.56	210220_at	**FZD2**	frizzled class receptor 2
< 1e-07	< 1e-07	3.54	203029_s_at	**PTPRN2**	protein tyrosine phosphatase, receptor type, N polypeptide 2
< 1e-07	< 1e-07	3.51	217033_x_at	**NTRK3**	neurotrophic tyrosine kinase, receptor, type 3
< 1e-07	< 1e-07	3.49	232195_at	**GPR158**	G protein-coupled receptor 158
< 1e-07	< 1e-07	3.32	234973_at	**SLC38A5**	solute carrier family 38, member 5
< 1e-07	< 1e-07	3.15	204368_at	**SLCO2A1**	solute carrier organic anion transporter family, member 2A1
< 1e-07	< 1e-07	0.33	219615_s_at	KCNK5	potassium channel, two pore domain subfamily K, member 5
< 1e-07	< 1e-07	0.31	230425_at	EPHB1	EPH receptor B1
< 1e-07	< 1e-07	0.3	205440_s_at	NPY1R	neuropeptide Y receptor Y1
< 1e-07	< 1e-07	0.3	235911_at	MFI2	antigen p97 (melanoma associated) identified by monoclonal antibodies 133.2 and 96.5
< 1e-07	< 1e-07	0.29	227526_at	CDON	cell adhesion associated, oncogene regulated
< 1e-07	< 1e-07	0.28	225524_at	ANTXR2	anthrax toxin receptor 2
< 1e-07	< 1e-07	0.28	235746_s_at	PLA2R1	phospholipase A2 receptor 1
< 1e-07	< 1e-07	0.27	223620_at	GPR34	G protein-coupled receptor 34
< 1e-07	< 1e-07	0.27	239185_at	ABCA9	ATP binding cassette subfamily A member 9
< 1e-07	< 1e-07	0.27	205498_at	GHR	growth hormone receptor
< 1e-07	< 1e-07	0.26	1556427_s_at	LRRN4CL	LRRN4 C-terminal like
< 1e-07	< 1e-07	0.25	214265_at	ITGA8	Integrin alpha 8
< 1e-07	< 1e-07	0.25	229254_at	MFSD4	major facilitator superfamily domain containing 4
< 1e-07	< 1e-07	0.24	205328_at	CLDN10	claudin 10
< 1e-07	< 1e-07	0.24	241412_at	BTC	betacellulin
< 1e-07	< 1e-07	0.24	228575_at	IL20RB	interleukin 20 receptor beta
< 1e-07	< 1e-07	0.22	206637_at	P2RY14	purinergic receptor P2Y, G-protein coupled, 14

< 1e-07	< 1e-07	0.22	220351_at	ACKR4	atypical chemokine receptor 4
< 1e-07	< 1e-07	0.22	219263_at	RNF128	ring finger protein 128, E3 ubiquitin protein ligase
< 1e-07	< 1e-07	0.2	211734_s_at	FCER1A	Fc fragment of IgE, high affinity I, receptor for; alpha polypeptide
< 1e-07	< 1e-07	0.19	213369_at	CDHR1	cadherin-related family member 1
< 1e-07	< 1e-07	0.17	219059_s_at	LYVE1	lymphatic vessel endothelial hyaluronan receptor 1
< 1e-07	< 1e-07	0.16	226625_at	TGFBR3	transforming growth factor beta receptor III
< 1e-07	< 1e-07	0.15	228640_at	PCDH7	protocadherin 7
< 1e-07	< 1e-07	0.13	228335_at	CLDN11	claudin 11
< 1e-07	< 1e-07	0.13	209894_at	LEPR	leptin receptor
< 1e-07	< 1e-07	0.13	220356_at	CORIN	corin, serine peptidase
< 1e-07	< 1e-07	0.12	214598_at	CLDN8	claudin 8
< 1e-07	< 1e-07	0.075	205357_s_at	AGTR1	angiotensin II receptor, type 1

Note: False Discovery rate, FDR. Fold Change, FC. Bold and underlined genes represent cell surface genes that are 3-fold up in BCC compared with NS.

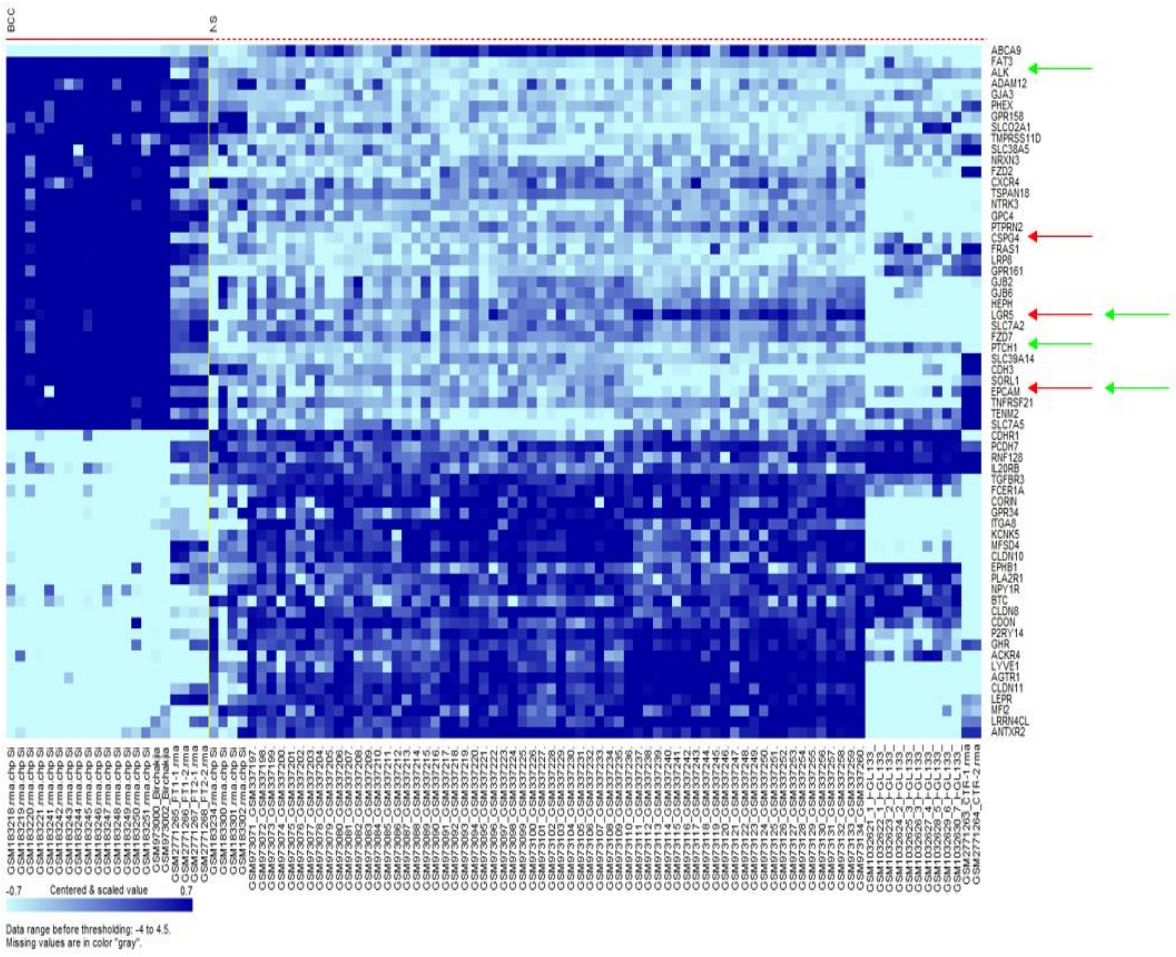


Figure 9: A heat map showing the expression of cell surface protein-encoding genes that are differentially expressed between normal human epidermal keratinocytes (normal) and cutaneous basal cell carcinoma cells (cancer). The straight red line represents the basal cell carcinoma samples and the dotted red line for normal skin samples. Dark blue represents gene upregulated; light blue represents gene downregulated in respective conditions (disease or normal). The red arrow indicates the gene of interest and its encoding protein as a target antigen for this study, while the green arrow represents the genes that are already reported to be a potential diagnostic marker for BCC.

3.1.3. Validation of 63 cell surface markers obtained from BRB-ArrayTools using alternative additional computational tools (SAM and TAC) and literature search

In the second analysis, we verified whether these 63 cell surface markers, that have a stronger statistical basis obtained from BRB-ArrayTools analysis are compared with alternative popular computational tools Significance Analysis of Microarrays (SAM) and Transcriptome Analysis Console (TAC), see **Section 2.1.4**. The Significance Analysis of Microarrays (SAM) Tool is based on a popular algorithm developed by Tusher et al.[205] for identifying significant genes in microarray data (www-stat.stanford.edu/~tibs/SAM). The SAM algorithm is based on the method of controlling the False Discovery Rate (FDR), which is defined in SAM as the median number of false-positive genes divided by the number of significant genes between the diseased (BCC) and normal skin. The false discovery rate cutoff value is one of the criteria that we chose for this analysis is **FDR <0.01**. There are 4284 differentially expressed genes estimated to have a false discovery rate of 0.0098 and are significant by SAM shown in the output file (Supplementary file2). Of this 4284 there are 232 genes 3-fold up in BCC and 263 genes were 3-fold significantly suppressed in BCC, owing to total of 495 DEG as shown in **Fig. 10**. In SAM plot positive significant genes are represented by red colour (red dots, n = 232) whereas negative significant genes are represented by green color (green dots, n = 263). Genes outside the dotted lines in red are significantly overexpressed in BCC ("positive significant") compared with normal skin and the one in green are significantly overexpressed in normal skin compared with BCC.

Subsequently, we looked for 63 cell surface protein encoding genes in this 495 differentially expressed gene reported by SAM, interestingly all the 63 cell surface protein encoding genes are matching with the 3-fold DEG obtained from SAM (**Fig. 11**). Adding to this, 34 BCC upregulated cell surface protein encoding genes are 100% matching with 232 BCC upregulated

genes reported in the SAM output file (figure not shown, the reader is asked to refer to the supplementary file 2). In parallel to this SAM analysis, we used TAC (Transcriptome Analysis Console 4.0.2) software[222,223] to identify genes that are differentially expressed in 21 BCC vs 80 NS, this TAC software also uses the FDR stat method, analysis of variance (ANOVA), and ebayes. Which uses, the information from all the probesets to yield an improved estimate for the variance, as sample being analyzed is small (21 BCC vs 80 NS). In addition, RMA (Robust Multi-array Average (RMA) method is employed for background correction and normalization and the outcome of TAC reports is 781 gene from 53,832 probe set are differentially expressed in BCC and normal Skin, which includes 266 gene of 3-fold up in BCC and 515 gene of 3-fold down in BCC with FDR *p*-value <0.001, as shown in **Fig. 12** every point in the scatter plot shows the expression of gene in two different condition diseased (BCC) and normal (skin). Red dot represent the gene that are 3-fold upregulated in BCC compared with normal skin, few interesting cell surface protein encoding genes such as EpCAM, LGR5, PTCH1, ALK, FAS, and GPR34 that are reported by others such as Linskey et al., Jia et al., Wan et al., Ning et al., Berman et al.,[116,224–227]are seen in the TAC scatter plot output, with CSPG4 as a possible new positive target for BCC and GPR34 as a negative target for BCC. Later we compared 63 cell surface protein encoding genes (BRB-ArrayTools) with 781 DEGs obtained from the TAC tool, we found 57 cell surface protein genes are overlapping, and 6 genes, namely, CXCR4, PHEX, SLC7A2, FZD7, PTPRN2, and IL20RB, are not matching with TAC based DEG. The possible loss of expression of these 6 genes in the TAC output file could be due to the selection of highly stringent criteria of FDR-corrected *p*-value <0.001 (**Fig. 13**). In parallel, we looked for BCC-associated cell surface markers that are reported in various published articles (see **Table 3**, biomarkers in BCC in **Section 1.3.1.4**) and we found EpCAM, LGR5, PTCH1, ALK obtained from the final list of computational analysis matching with the list of cell surface markers reported in literature, as highlighted by green arrows in **Fig. 9** heat map. In addition, CD90, EGFR, CD44, EMA, CD138, and CEA markers are not in the final list; however, EGFR and CD44 are seen in the list of differentially expressed genes between BCC and NS with *p*-value $>1e^{-07}$ and fold change less than 3. CD90, EMA, and CD138 are not seen in the output of differentially expressed genes in BCC vs Normal skin as this (raw data) probe set 208851_s_at, for CD90;213693_s_at, for MUC1; 201287_s_at, for CD138, did not fulfill the filtering and subsetting criteria. CEA (which might include any of this gene CEACAM5, CEACAM7, CEACAM1, CEACAM6) also failed to fulfill the chosen statistical criteria.

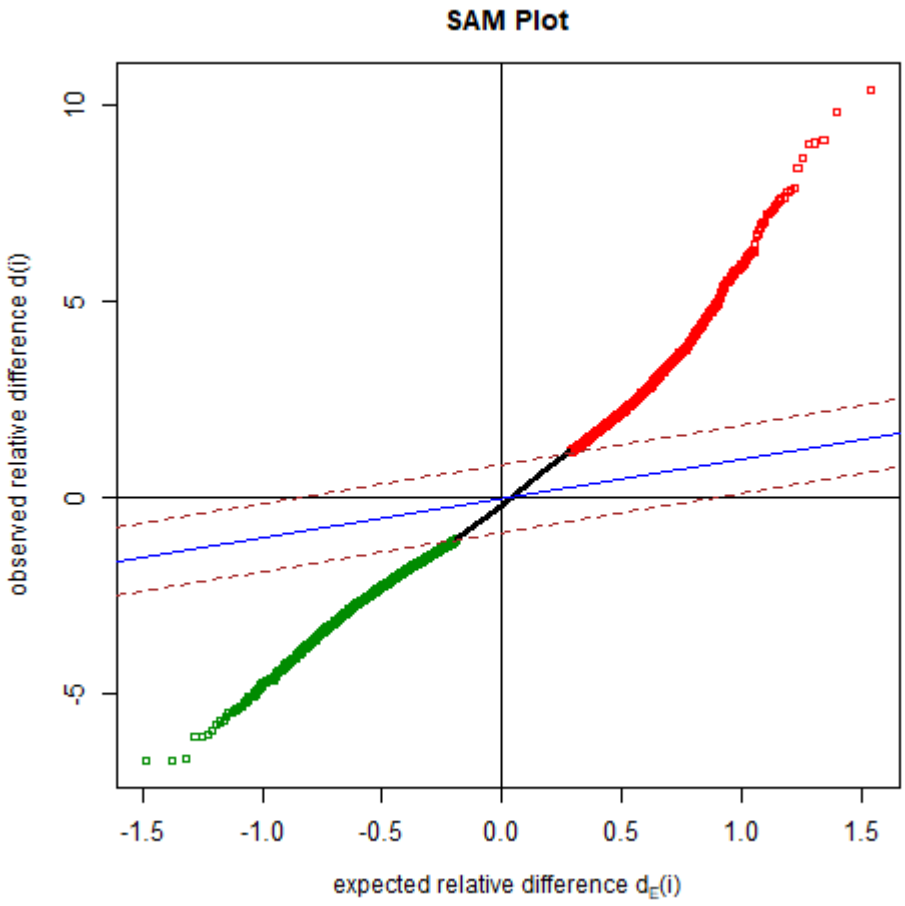


Figure 10: **SAM scatter plot:** SAM scatter plot of the observed d(i) versus the expected $d_E(i)$ relative difference between diseased (n = 21) and normal controls (n = 80) for the 4284 data using the median estimated FDR was <1%. In this SAM plot, the significant genes are represented by square symbols. Positive significant genes are represented by red color, whereas negative significant genes are represented by green color. The solid blue line represents no difference between d(i) and dE(i), and the broken red lines show the delta limit of 1.0 from the solid blue line. A total of 232 probe sets outside the broken lines were 3-fold significantly induced in diseased (red dots, n = 232) and 263 probe sets were 3-fold significantly suppressed (green dots, n = 263) in diseased.

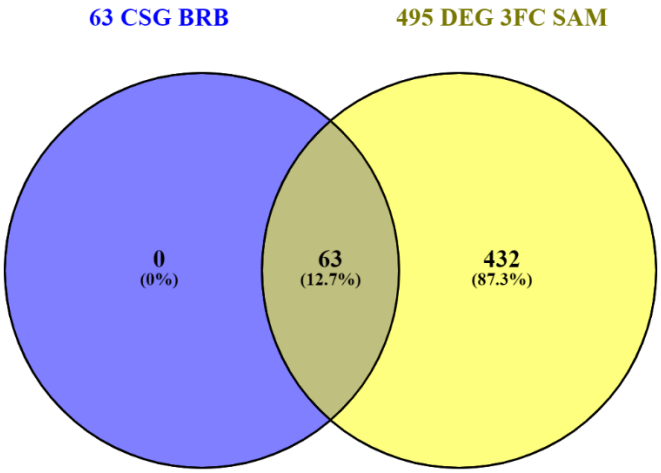


Figure 11: **Venn diagram of intersections of cell surface protein–encoding genes between data obtained from BRB-ArrayTools and SAM tool**. All the 63-cell surface protein–encoding genes are present within the 3-fold upregulated and downregulated genes of basal cell carcinoma obtained from the SAM tool, further strengthening the premise that these differentially expressed mRNAs are truly significant in diseased vs. normal.

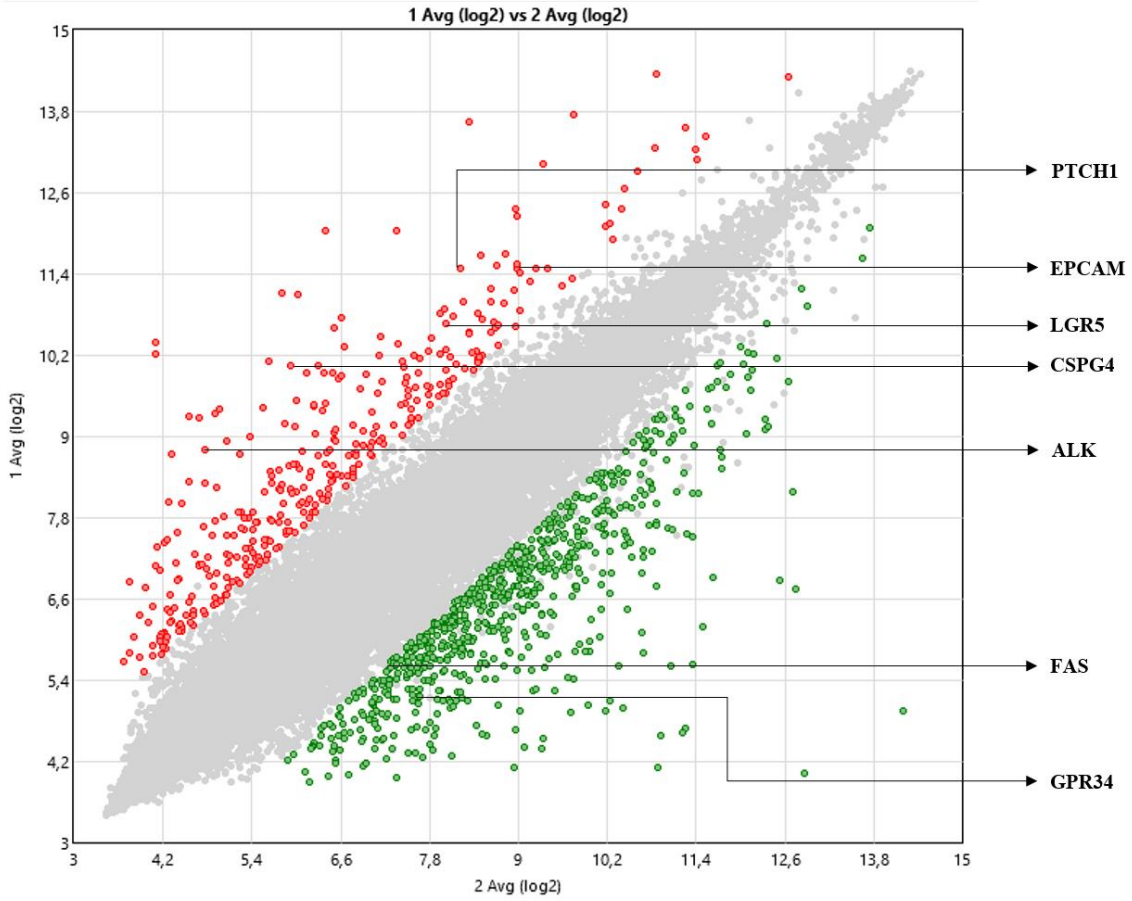


Figure 12: Transcriptomic data between basal cell carcinoma and normal skin

In the scatter plot, the values plotted on the X and Y axis are the averaged normalized signal values in each group (log2scaled). 1 Avg (log 2) in Y-axis represents 21 basal cell carcinoma samples and 2 Avg(log2) in X-axis represent 80 normal skin samples. The red points in the plots indicate >3.0-fold upregulation of expression in BCC, green points indicate >3.0-fold downregulation of corresponding mRNAs in BCC, and gray points indicate < 3.0-fold change in expression. FDR-corrected *p*-value < 0.001. Arrows highlight some key mRNAs discussed in the text and or reported by others [116,224–227] and **Table 3.**

Table 9: Selective cell surface protein-encoding genes that are differentially expressed between BCC and NS-analyzed using TAC software

Gene Symbol	ProbeSet ID	FDR *p*-value	*p*-value	Fold Change in BCC
CSPG4	214297_at	2.84E-27	2.81E-30	17.65
ALK	208212_s_at	8.42E-28	6.94E-31	16.47
PTCH1	209815_at	4.06E-29	2.40E-32	9.59
LGR5	210393_at	1.66E-08	1.06E-09	6.26
EPCAM	201839_s_at	2.66E-18	1.73E-20	5.66
FAS	215719_x_at	7.66E-11	1.61E-09	-3.25
GPR34	223620_at	5.93E-08	4.44E-09	-5.52
MUC1	**213693_s_at**	**0.0438**	**0.0221**	**-2.23**

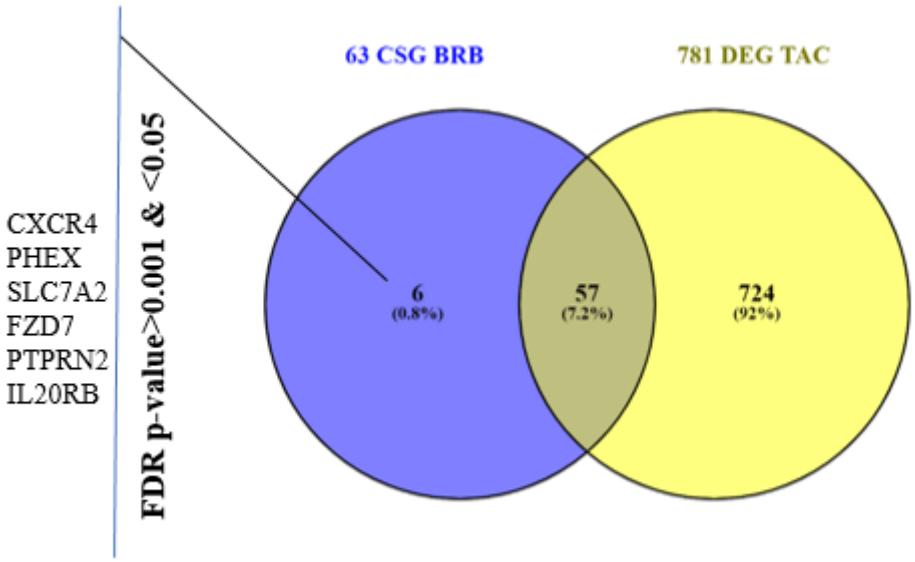


Figure 13: **Venn diagram of intersections of cell surface protein–encoding genes between data obtained from BRB-ArrayTools and TAC tool**. A total of 63 cell surface protein–encoding genes, 57 genes are common in both the tools, the remaining 6 genes that are not seen in data obtained from TAC were identified with FDR corrected p >0.001 and p <0.05.

Taken together, the results described indicate that the selective BCC upregulated mRNAs (such as EpCAM, CSPG4, PTCH1, LGR5, ALK) and BCC downregulated mRNAs (such as GPR34, FAS) have been found to be overlap with both literature[116,224–227] and a couple of computational tools (e.g., SAM, TAC), further confirming that these statistically significant upregulated genes when expressed in BCC could be a promising candidate for targeted BCC diagnosis and therapy. In addition, we speculate FAS and GPR34 could be applied as a negative markers for BCC diagnosis. Subsequently literature reported BCC downregulated mRNAs such as CD90, EMA, and CD138 are not seen in the output of differentially expressed mRNAs in BCC vs Normal skin as this probe set 208851_s_at, for CD90;213693_s_at, for MUC1; 201287_s_at, for CD138, did not fulfill the filtering and subsetting criteria. We assumed further relaxing the filtering criteria from FDR *p*-value <0.001 to <0.05 and fold change 3 to 1.5 might allow these mRNAs to appear in class comparison output and to our surprise we found the MUC1 (EMA) BCC downregulated mRNAs appeared in the TAC output, after relaxing the criteria, data shown in **Table 9**.

3.2. IN SILICO DESIGN OF MAMMALIAN AND BACTERIAL EXPRESSION VECTOR SYSTEMS

3.2.1. Identification of antibody sequences from the public repository

To develop a monoclonal antibody against the target antigen by immunization and to characterize the produced antibody will reflect years of effort. With readily available information such as antibody sequences, species of origin such as human or mouse, epitope recognition within an antigen, and with the application of recombinant DNA (rDNA) technology, mAb against target antigen can be made in a short time. The foremost prerequisite for rDNA-based scFv production is knowing the antibody sequences that recognizes the target antigen. This information can be readily obtained from patented documents and published literature will reflect years of efforts in developing monoclonal antibodies for diagnosis and therapeutics. Therefore, we looked for readily available antibody sequence for the target antigen from major jurisdictions (e.g., USPTO, WIPO, **DEPATISnet**) from a published scientific article (PUBMED), from antibody database such as ABCD (for AntiBodies Chemically Defined) using keywords such as mAb, scFv, and Anti-TAAs (αEpCAM, αMUC1, αLGR5, αCD90, αCD138, αCSPG4), see materials and methods section 2.2.1. Concomitantly, we found antibody sequence (VH-VL) targeting EpCAM, LGR5, and CD90 from a patent file, whereas sequence for anti-CD138 and anti-CSPG4 is

obtained from a published article Chen et al. and Morgan et al.[228,229] (**Table 10**). Fortunately, we got anti-MUC1 sequences provided by Gerard Bos, UMC Maastricht.

Table 10. Target antigen and corresponding antibody from respective parent file

Cell surface antigen	scFv	Reference	Theoretical pI/Mw of full-length protein scFv-SNAP
EpCAM	UBS54 and OCAb9-1*	US7429486B2 and US20150017230A1*	**UBS54-SNAP** (5.89/50502.03) **OCAb9-1-SNAP** (6.23/50974.62)
EMA/MUC1	ID405	Gerard Bos, UMC Maastricht	**ID405-SNAP** (6.28/51733.32)
LGR5/GPR49	YW353	US 9175089 B2	**YW353-SNAP** (6.30/50142.65)
CD90/Thy1	h5-Thy1	WO 2017214050 A1	**h5-Thy1-SNAP** (6.64/52306.14)
CD138/SDC	h-STL002	Chen et al., 2016	**h-STL002-SNAP** (6.22/50897.48)
CSPG4**	mAb9.2.27	Morgan et al., 1981	**mAb9.2.27 -SNAP** (5.97/51364.87)

We have chosen the above cell surface markers as target antigens to distinguish and characterize BCC from other skin lesions and normal skin tissues. These markers were selected based on literature reports and computational analysis using the BRB-ArrayTools, SAM, and TAC (see **Section 3.1.2**).

*Antibody fragment described by Liao et al.; recognizes EpCAM expressed in cancer cells but not in cultured normal cells.

**Antibody for CSPG4 was provided by Siyabulela Magugu, a former MSc student of MB&I

3.2.2. Construction of mammalian and bacterial expression vector

After accessing antibody sequence (VH & VL) from the public repository, we cross-verified or identified complementarity-determining regions (CDRs) and framework region (FRs) for each VH and VL chain using IgBlast tool (**Fig. 14**, see **Section 2.2.1 and 2.2.2**). Concomitantly, the mammalian expression plasmid encoding for scFv-SNAP is constructed by genetical modification of the VH-VL coding sequence of the αEpCAM/EMA/CD138 fragment [scFv-UBS54/ID405/h-STL002] with three iterations of a $(G_4S)_3$ linker and theoretically cloned (see **Section 2.2.2**) to the mammalian expression vector "pMS-SNAP" bicistronic vector using SnapGene® software (GSL Biotech, available from snapgene.com) to generate the complete scFv-SNAP cassettes, as shown in **Fig. 15**. Similarly, the bacterial expression plasmid is constructed by genetically modifying the VH-VL fragments of αEpCAM*/LGR5/CD90 [scFv-OCAb9-1*/ YW353/ h5-Thy1] with three iterations of a $(G_4S)_3$ linker and theoretically cloned to the bacterial expression vector "pMT-SNAP" to generate the complete scFv-SNAP cassettes, as shown in **Fig. 16**. pMT vector is successfully used for periplasmic stress expression of recombinant αEpCAM(scFv)-ETA' in BL21 *E. coli*[211]. This vector was derived from pET27b vector (Novagen).

In mammalian expression vector, an N-terminal mouse IgK V-III secretory leader sequence was included to the upstream of the scFv, which promotes the secretion of the expressed protein by the host cell[230], followed by (in order, N → C) the two recommended alanine (AA) residues, a poly(6x)-histidine (HHHHHH) tag, or His-tag, and an enterokinase cleavage site (DDDDK), which allows digestive cleavage of the fusion protein for optional removal of the N-terminal elements (**Fig. 15**).[231] In the bacterial expression vector, upstream of the scFv, an N-terminal pelB peptide signal/leader sequence was included, which directs the fusion protein to the bacterial periplasm, where the sequence is removed by signal peptidase,[232] followed by (in order, N → C) the two recommended alanine (AA) residues, a poly(10x)-histidine (HHHHHHHHHH) tag, or His-tag[233] and an enterokinase cleavage site (DDDDK), which allows digestive cleavage of the fusion protein for optional removal of the N-terminal elements (**Fig. 16).**[231]

A **H UBS54**

```
                <-------------------------------FR1-IMGT-------------------------------><-------CDR1-IM
                 Q  V  Q  L  V  Q  S  G  A  E  V  K  K  P  G  S  S  V  R  V  S  C  K  A  S  G  G  T  F  S
Query_1       1  CAGGTGCAGCTGGTGCAGTCTGGGGCTGAGGTGAAGAAGCCTGGGTCCTCGGTGAGGGTCTCCTGCAAGGCTTCTGGAGGCACCTTCAGC  90
IGHV1-69*01   1  ..................................................A.................................  90

                GT------><--------------------FR2-IMGT--------------------><-------CDR2-IMGT------><-----
                 S  Y  A  I  S  W  V  R  Q  A  P  G  Q  G  L  E  W  M  G  G     I  P  I  F  G  T  A  N  Y
Query_1       91  AGCTATGCTATCAGCTGGGTGCGACAGGCCCCTGGACAAGGGCTTGAGTGGATGGGAGGG---ATCCCTATCTTTGGTACAGCAAACTAC  177
IGHV1-69*01   91  ..........................................................ATC.........................  180

                ------------------------------------------------FR3-IMGT-----------------------------------
                 A  Q  K  F  Q  G  R  V  T  I  T  A  D  E  S  T  S  T  A  Y  M  E  L  S  S  L  R  S  E  D
Query_1      178  GCACAGAAGTTCCAGGGCAGAGTCACGATTACCGCGGACGAATCCACGAGCACAGCCTACATGGAGCTGAGCAGCCTGAGATCTGAGGAC  267
IGHV1-69*01  181  ..................................................................................................  270
                 -----------------><-------CDR3-IMGT------><----------FR4-IMGT---------->
                  T  A  V  Y  Y  C  A  R  D  P  F  L  H  Y  W  G  Q  G  T  L  V  T  V  S
Query_1      268  ACGGCTGTGTATTACTGTGCAAGAGACCCGTTTCTTCACTATTGGGGCCAAGGTACCCTGGTCACCGTCTC  338
IGHV1-69*01  271  .....C..............G.....----------------------------------------------  296
```

VL UBS54

```
                <---------------------------------FR1-IMGT---------------------------------><-----------
                 E  I  E  L  T  Q  S  P  L  S  L  P  V  T  P  G  E  P  A  S  I  S  C  R  S  S  Q  S  L  L
Query_1       1  GAAATTGAGCTCACTCAGTCTCCACTCTCCCTGCCCGTCACCCCTGGAGAGCCGGCCTCCATCTCCTGCAGGTCTAGTCAGAGCCTCCTG  90
IGKV2-28*01   1  ..T....T.A.G.....................................................................  90

                --CDR1-IMGT------------><--------------------FR2-IMGT---------------------><CDR2-IM><-----
                 H  S  N     G  Y  N  Y  L  D  W  Y  L  Q  K  P  G  Q  S  P  Q  L  L  I  Y  L  G  S  N  R
Query_1       91  CATAGTAATG---GATACAACTATTTGGATTGGTACCTGCAGAAGCCAGGGCAGTCTCCACAGCTCCTGATCTATTTGGGTTCTAATCGG  177
IGKV2-28*01   91  ..........---.....................................................................  177

                ------------------------------------------FR3-IMGT-------------------------------------
                 A  S  G  V  P  D  R  F  S  G  S  G  S  G  T  D  F  T  L  K  I  S  R  V  E  A  E  D  V  G
Query_1      178  GCCTCCGGGGTCCCTGACAGGTTCAGTGGCAGTGGATCAGGCACAGATTTTACACTGAAAATCAGCAGAGTGGAGGCTGAGGATGTTGGG  267
IGKV2-28*01  178  ..................................................................................  267

                -----------><--------CDR3-IMGT------><----------FR4-IMGT---------->
                 V  Y  Y  C  M  Q  A  L  Q  T  F  T  F  G  P  G  T  K  V  E  I  K
Query_1      268  GTTTATTACTGCATGCAAGCTCTACAAACTTTCACTTTCGGCCCTGGGACCAAGGTGGAGATCAAA  333
IGKV2-28*01  268  ..........................----------------------------------  297
```

B

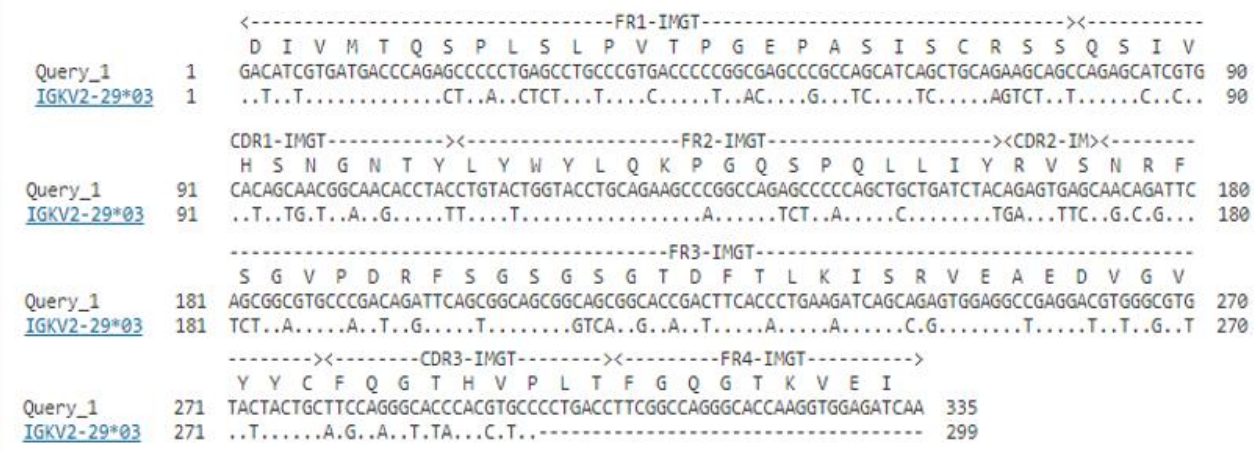
VH- ID405

VL-ID405

C

(VH) h-STL002

(VL) h-STL002



D

(VH) OCAb9-1

```
                <--------------------------------FR1-IMGT--------------------------------><-------CDR1-IMG
                   V  Q  L  V  E  S  G  P  E  L  K  K  P  G  E  T  V  K  I  S  C  K  A  S  G  Y  T  F  T
Query_1      2     AGGTGCAGCTGGTGGAGAGCGGCCCCGAGCTGAAGAAGCCCGGCGAGACCGTGAAGATCAGCTGCAAGGCCAGCGGCTACACCTTCACCG  91
IGHV7-81*01  2     ..............C..TCT....AT...G.....C....T..G.CCT.A......G..TC.........TTCT..T....GT......A  91

             T------><--------------------FR2-IMGT--------------------><-------CDR2-IMGT------><------
             D  Y  S  M  H  W  V  K  Q  A  P  G  K  G  L  K  W  M  G  W  I  N  T  E  T  G  E  P  T  F
Query_1      92    ACTACAGCATGCACTGGGTGAAGCAGGCCCCCGGCAAGGGCCTGAAGTGGATGGGCTGGATCAACACCGAGACCGGCGAGCCCACCTTCG  181
IGHV7-81*01  92    C...TG.T...A.T......CCA........T..AC.A..G..TG.........A...T........T.C..T..GA.C..A..A.AT.  181

             ------------------------------------------------FR3-IMGT-------------------------------------
             A  D  D  F  K  G  R  F  A  F  S  L  E  T  S  A  R  T  T  Y  L  Q  I  N  N  L  K  N  E  D
Query_1      182   CCGACGACTTCAAGGGCAGATTCGCCTTCAGCCTGGAGACCAGCGCCAGAACCACCTACCTGCAGATCAACAACCTGAAGAACGAGGACA  271
IGHV7-81*01  182   ..C.G.G.....CA..AC.G..T.T....TC.A....C...TCT.....C..AG.A.............G..G...A...GCT.......  271

             ---------------><----CDR3-IMGT--->
             T  A  T  Y  F  C  A  R  T  A  V  Y  W  G  Q  G  T
Query_1      272   CCGCCACCTACTTCTGCGCCAGAACCGCCGTGTACTGGGGCCAGGGCACC  321
IGHV7-81*01  272   TG....TG..T.A...T..G..----------------------------  294
```

(VL) OCAb9-1

```
             <---------------------------------FR1-IMGT---------------------------------><----CDR1-IM
              D  I  Q  L  T  Q  S  P  S  S  L  S  A  S  L  G  E  R  V  S  L  T  C  R  A  S  Q  E  I  S
Query_1      1     GACATCCAGCTGACCCAGAGCCCCAGCAGCCTGAGCGCCAGCCTGGGCGAGAGAGTGAGCCTGACCTGCAGAGCCAGCCAGGAGATCAGC  90
IGKV1-5*01   1     .........A........TCT..TTC..C....TCT..ATCTG.A..A..C.....C.C.A.C..T...C.G.....T...AGT..T..T  90

             GT--->&lt;--------------------FR2-IMGT--------------------><CDR2-IM><----------------------
              V  S  L  S  W  L  Q  Q  K  P  D  G  T  I  K  R  L  I  Y  A  T  S  T  L  D  S  G  V  P  K
Query_1      91    GTGAGCCTGAGCTGGCTGCAGCAGAAGCCCGACGGCACCATCAAGAGACTGATCTACGCCACCAGCACCCTGGACAGCGGCGTGCCCAAG  180
IGKV1-5*01   91    AGCT.GT..GC....TAT........A..A.GGAAAG..CCT...CTC........T.ATG..TC..GTT....A..T..G..C..ATCA  180

             -------------------------FR3-IMGT-----------------------------------------------------><-----
              R  F  S  G  S  R  S  G  S  D  Y  S  L  T  I  S  S  L  E  S  E  D  F  A  D  Y  Y  C  L  Q
Query_1      181   AGATTCAGCGGCAGCAGAAGCGGCAGCGACTACAGCCTGACCATCAGCAGCCTGGAGAGCGAGGACTTCGCCGACTACTACTGCCTGCAG  270
IGKV1-5*01   181   ..G...........TG..TCT..G.CA..A.T..CT..C..............C..CCT..T..T..T..AACT..T.......AA...  270

             ---CDR3-IMGT-------->
              Y  A  S  Y  P  W  T  F  G  G  G  T  K  L  E  I
Query_1      271   TACGCCAGCTACCCCTGGACCTTCGGCGGCGGCACCAAGCTGGAGATCAA  320
IGKV1-5*01   271   ..-------------------------------------------------  272
```

E

(VH) YW353

```
                  <-------------------------------FR1-IMGT-------------------------------><-------CDR1-IM
                   E  V  Q  L  V  E  S  G  G  G  L  V  Q  P  G  G  S  L  R  L  S  C  A  A  S  G  F  T  F  T
Query_1       1   GAGGTGCAGCTGGTGGAGAGCGGCGGCGGCCTGGTGCAGCCCGGCGGCAGCCTGAGACTGAGCTGCGCCGCCAGCGGCTTCACCTTCACC  90
IGHV3-53*02   1   ...................CT..A..A...T..A.C.....T..G..GTC.........CTC...T..A...TCT..G......G...GT  90

                  GT------><--------------------FR2-IMGT--------------------><-------CDR2-IMGT------><-----
                    S  Y  S  I  S  W  V  R  Q  A  P  G  K  G  L  E  W  V  A  E  I  Y  P  P  G  G  Y  T  D  Y
Query_1       91  AGCTACAGCATCAGCTGGGTGAGACAGGCCCCCGGCAAGGGCCTGGAGTGGGTGGCCGAGATCTACCCCCCCGGCGGCTACACCGACTAC  180
IGHV3-53*02   91  ...A..TA...G........CC.C.....T..A..G.....G...........CT.A.TT..T..---TAG...T..TAG...AT.....  177
                  ------------------------------------------------FR3-IMGT---------------------------------
                   A  D  S  V  K  G  R  F  T  I  S  A  D  T  S  K  N  T  A  Y  L  Q  M  N  S  L  R  A  E  D
Query_1       181 GCCGACAGCGTGAAGGGCAGATTCACCATCAGCGCCGACACCAGCAAGAACACCGCCTACCTGCAGATGAACAGCCTGAGAGCCGAGGAC  270
IGHV3-53*02   178 ..A...TC.........C...........TC.AGA....ATTC.........GCTG..T..T..A.......................  267

                  -----------------><--------CDR3-IMGT--------><-------FR4-IMGT-------->
                   T  A  V  Y  Y  C  A  K  A  R  L  F  F  D  Y  W  G  Q  G  T  L  V  T
Query_1       271 ACCGCCGTGTACTACTGCGCCAAGGCCAGACTGTTCTTCGACTACTGGGGCCAGGGCACCCTGGTGACCG  340
IGHV3-53*02   268 ..G........T.....T..-------------------------------------------------  287
```

(VL) YW353

```
                  <---------------------------------FR1-IMGT---------------------------------><----CDR1-IM
                   D  I  Q  M  T  Q  S  P  S  S  L  S  A  S  V  G  D  R  V  T  I  T  C  R  A  S  Q  D  V  S
Query_1       45  GACATCCAGATGACCCAGAGCCCCAGCAGCCTGAGCGCCAGCGTGGGCGACAGAGTGACCATCACCTGCAGAGCCAGCCAGGACGTGAGC  134
IGKV1-5*01    1   ...................TCT..TTC..C....TCT..ATCT..A..A........C........T...C.G.....T...AGTA.T..T  90

                  GT--->&lt;--------------------FR2-IMGT--------------------><CDR2-IM><----------------------
                   T  A  V  A  W  Y  Q  Q  K  P  G  K  A  P  K  L  L  I  Y  S  A  S  F  L  Y  S  G  V  P  S
Query_1       135 ACCGCCGTGGCCTGGTACCAGCAGAAGCCCGGCAAGGCCCCCAAGCTGCTGATCTACAGCGCCAGCTTCCTGTACAGCGGCGTGCCCAGC  224
IGKV1-5*01    91  .G.TGGT..........T........A..A..G..A.....T.....C........TGAT...TC.AGTT..G.A..T..G..C..ATCA  180

                  -------------------------FR3-IMGT---------------------------------------------------><-----
                   R  F  S  G  S  G  S  G  T  D  F  T  L  T  I  S  S  L  Q  P  E  D  F  A  T  Y  Y  C  Q  Q
Query_1       225 AGATTCAGCGGCAGCGGCAGCGGCACCGACTTCACCCTGACCATCAGCAGCCTGCAGCCCGAGGACTTCGCCACCTACTACTGCCAGCAG  314
IGKV1-5*01    181 ..G..........T..ATCT..G..A..A.....T..C.....................T..T..T..T..A..T..T........A...  270


                    ---CDR3-IMGT--------><F>
                     S  Y  T  T  P  P  T  F
Query_1       315   AGCTACACCACCCCCCCCACCTTC  338
IGKJ5*01      4     ------------------.......  10
```

F

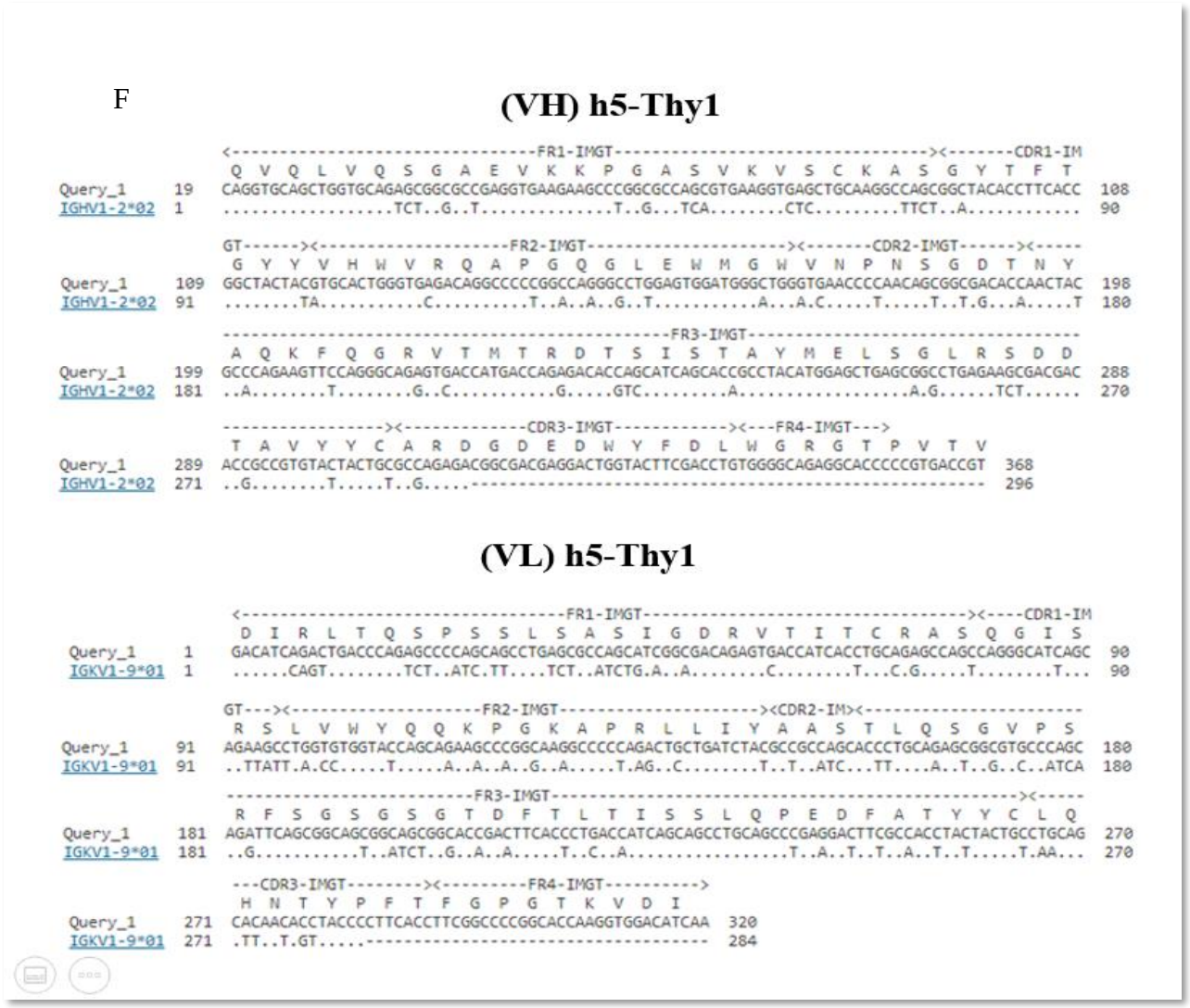


Figure 14: The delineation of IGV domain framework regions and complementarity-determining regions of query antibody sequence. Sequences were analyzed individually using the IgBLAST tool and compared against the best matching germline V gene. Dots represent matching nucleotides and mismatched nucleotides are represented with an acronym of nucleobases. A) UBS54 scFv, B) ID405 scFv, C) h-STL002 scFv, D) OCAb9-1 scFv, E) YW353 scFv, F) h5-Thy1scFv.

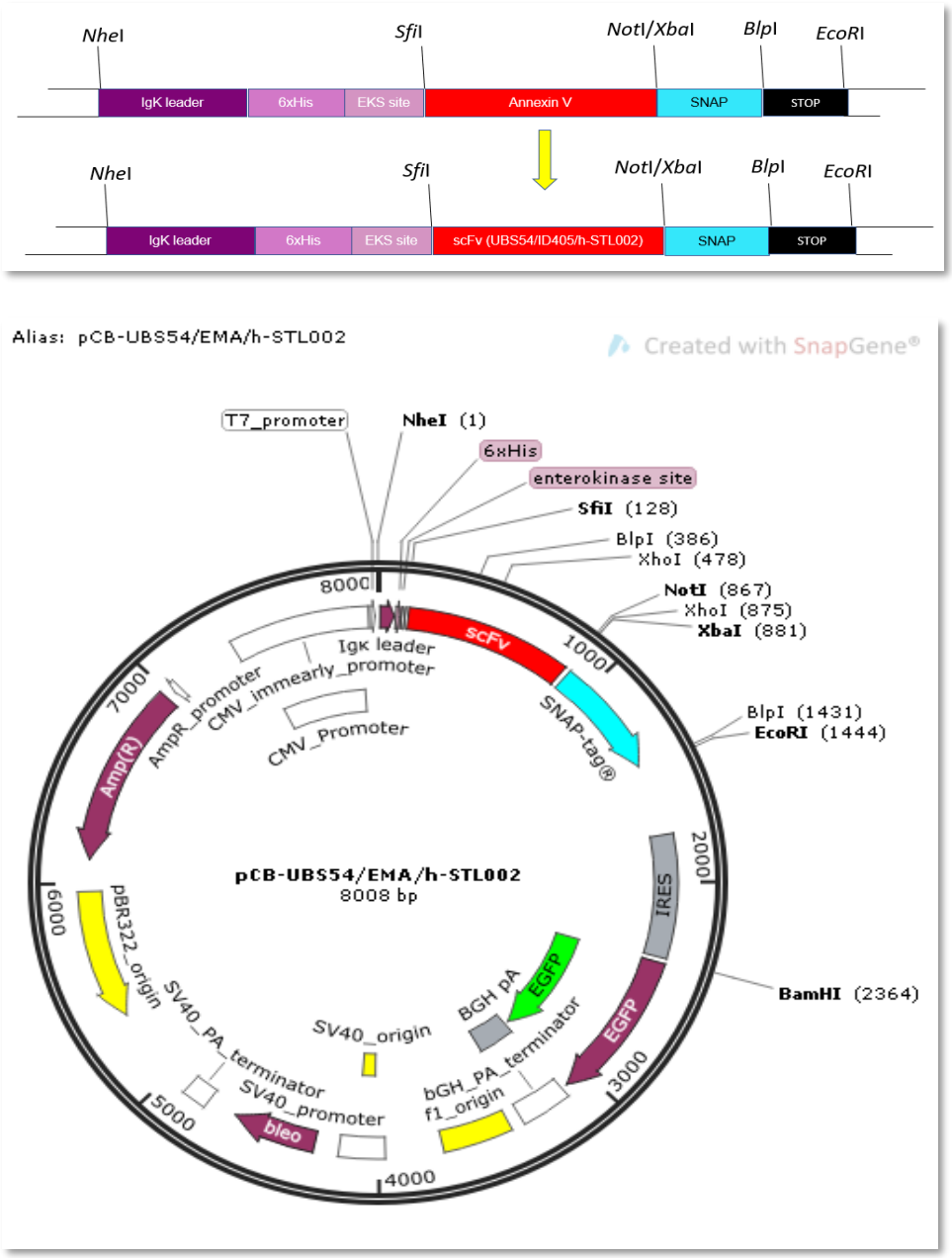
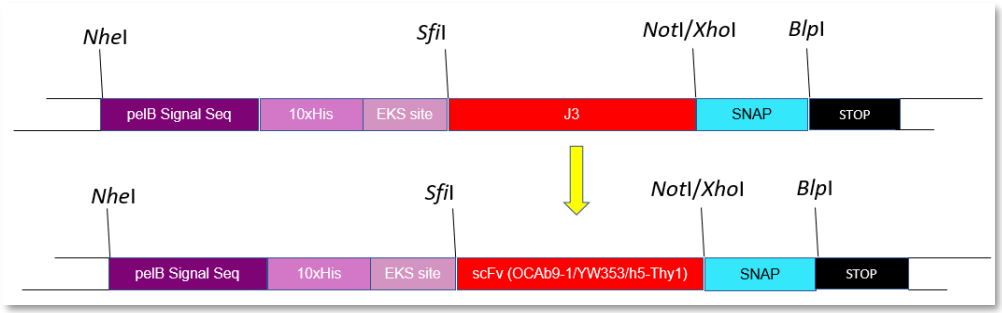


Figure 15: In silico design of mammalian expression plasmids; vector map of pCB-scFvUBS54-SNAP, pCB-scFvID405-SNAP and pCB-scFvh-STL002-SNAP. The screenshot obtained from the molecular biology software, SNAPGene, used for performing - *in silico* designing of expression plasmids illustrates the vector design which contains the **scFvUBS54** or **scFvID405** or **scFvh-STL002** at the N-terminus and SNAP at the C-terminus. F1 origin of replication for bleomycin (zeocin) resistance gene; pBR322 origin for *E. coli*, Amp (R) ampicillin-resistant gene; T7 promoter controlling IgK leader peptide signal for extracellular secretion of the fusion protein; polyhistidine (6xHis) sequence used for purification, enterokinase site to remove putative phosphorylation sites, (IRES) which mediates the cotranslational

cytosolic expression of enhanced green fluorescent protein (EGFP) which is used for the detection of successfully transfected recombinant protein.



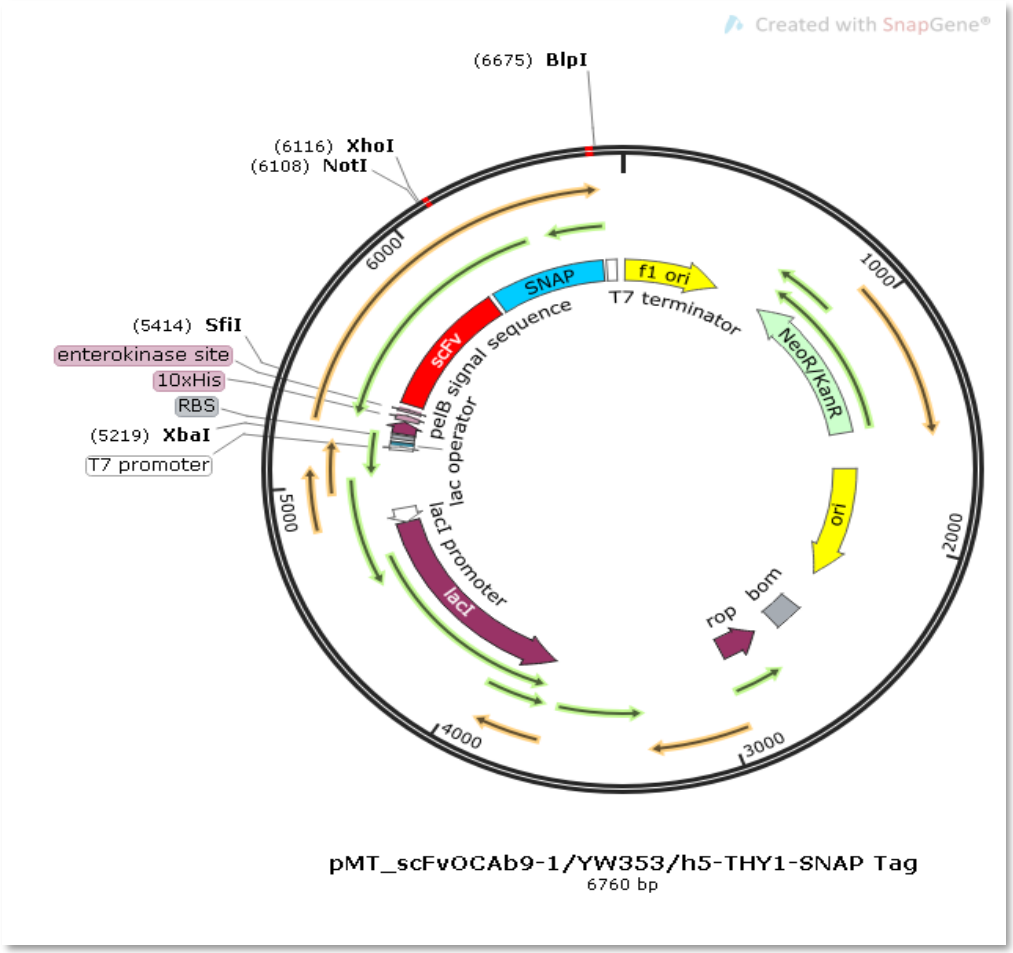


Figure 16: In silico design of bacterial expression plasmids; Vector map of pMT-scFvUOCAb9-1 SNAP, pMT-scFvYW353-SNAP, and pMT-scFvh5-Thy1 SNAP plasmids for bacterial expression. The vector contains the **scFvOCAb9-1** or **scFvYW353** or **scFvh5-Thy1** at the N-terminus and SNAP at the C-terminus. F1 origin of replication for kanamycin resistance gene; pBR322 origin for *E. coli*, kanamycin-resistant (Kan R) gene used for the

detection of successfully transfected recombinant clone; T7 promoter controlling pelB peptide signal/leader sequence, directs the fusion protein to the bacterial periplasm, where the sequence is removed by signal peptidase; polyhistidine sequence (10xHis), used for purification, enterokinase site to remove putative phosphorylation sites. lac repressor (*lac*I protein) binds to lac operator and inactivates Lac promoter and represses target gene transcription. In the presence of lactose or lactose analog IPTG, the lac repressor undergoes a conformational change that removes it from lacO sites within the promoter and ceases repression of the target gene. Rop gene, which maintains plasmid at high copy number.

3.2.3. Removing cryptic translation initiation site within the ORF of αEpCAM-UBS54-SNAP

Protein truncation could be due to the presence of translation initiation site downstream of annotated start codon in human mRNAs, which coevolves with the Kozak sequence. The translation initiation could be due to ribosome shunting, discontinuous scanning, and ribosome leaking, giving rise to two protein products translated from a single transcript.[234,235,236] Therefore, we examined for cryptic translation initiation sites caused by the Kozak motif "A/GXXaugG" within the ORF of pCB-scFvUBS54-SNAP, see **Section 2.2.3**. We found the sequence CCACCATGG, a near-perfect match to the consensus eukaryote Kozak sequence "A/GXXaugG", positioned 6 bp upstream of the first ATG; this Kozak sequence is essential for enhancing the translation from the correct initiation codon. In addition to this Kozak sequence, we found Kozak sequence downstream of first "ATG" positioned at 42 amino acids named as Kozak 2, at 124 as Kozak 3, at 267 as Kozak 4, at 293 as Kozak 5, at 299 as Kozak 6, and at 352 as Kozak 7. Kozak 2, 5, and 6 are strong Kozak sequences (Kozak score >80) and the Kozak 3, 4, and 7 are weak Kozak sequences (Kozak score<80) (**Table 11**).[215,237] We observed that the stronger the Kozak sequence, the stronger the level of ribosome initiation or stalling. From this, we hypothesized that there is the possibility that the 30–35 kDa truncated product can be generated after translating the transcript encoded for scFv-UBS54-SNAP. To test this hypothesis, we engineered translationally silent mutations in the Kozak 5 and Kozak 6 region to remove the cryptic Kozak sequence and to reduce the strength of Kozak sequence by manipulating the nucleotide sequence in positioned Kozak 5; ultimately, we derived four constructs from parental pCB-scFv-UBS54-SNAP, with the intention to generate a single full-length protein translated from a single transcript (see **Fig. 17),** which include 1) Kozak 5 and Kozak 6 deletion, 2) Kozak 5 del, 3) Kozak 6 del, 4) Kozak 5 low score.

Table 11: Transcripts UBS54-SNAP with weak Kozak sequences and strong Kozak sequences

KOZAK SEQ 5′-A/GXXAUGG-3′	**ORF "pCB-scFvUBS54-SNAP"**	**MET position in Kozak consensus within ORF**	**KOZAK QUALITY SCORE**
KZ1	CCACCATGG	1	85.75
KZ2	CGGCCATGG	42	92.25
KZ3	CCTACATGG	124	73.50
KZ4	ACTGCATGC	267	33.00
KZ5	CTAGAATGG	293	98.80
KZ6	GCGAAATGA	299	91.00
KZ7	CACTGATGC	352	69.06

KZ represents stronger Kozak, KZ represents weaker Kozak.

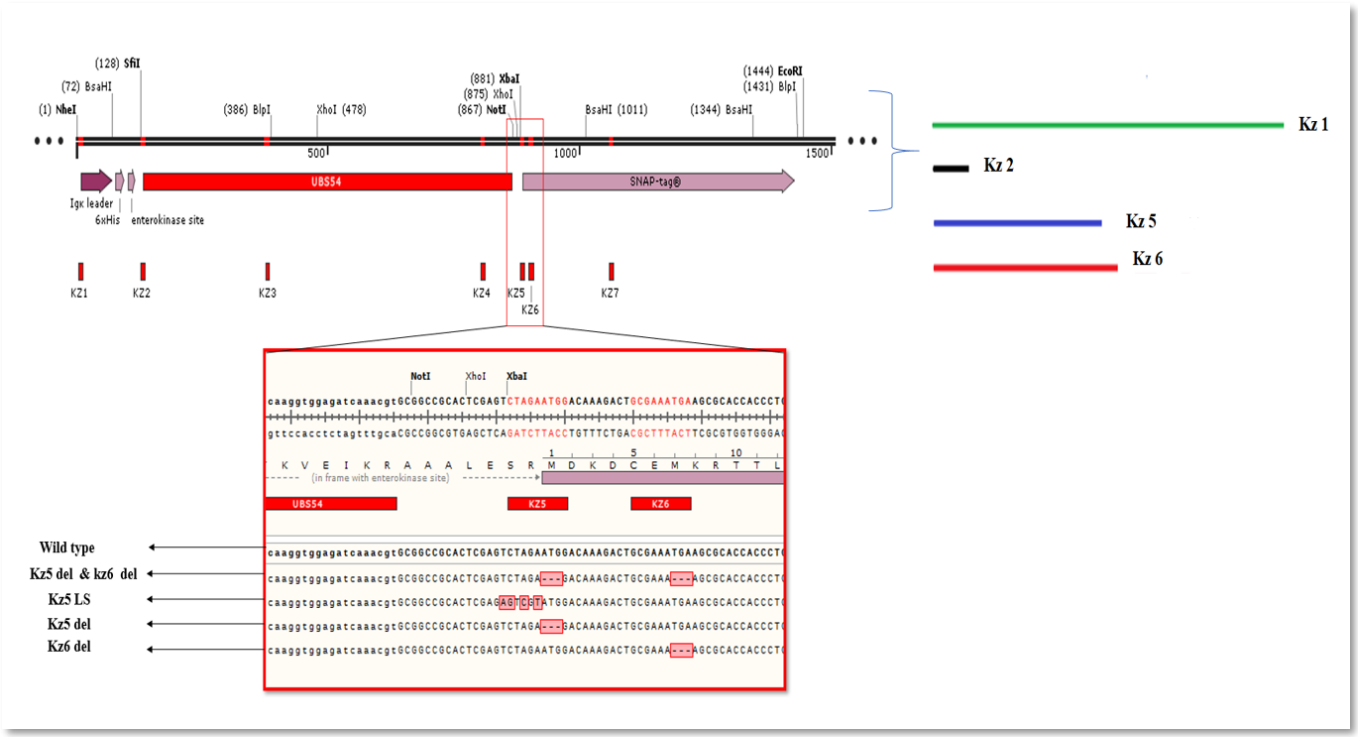


Figure 17: Schematic representation of plasmids encoding full length (green line) and C-terminally truncated proteins (black, blue, and red lines). We found three Kozak sequences with high score preceding the first Kozak sequence (TISs), one at the N-terminal end of scFv (UBS54) might give 4 kDa truncated protein (black line), and another two at the N-terminal region of SNAP might give 30 kDa truncated protein (blue and red line). We were interested in avoiding 30 kDa truncated protein. Hence, we designed four different constructs (black arrows): (1) Plasmids pkz5del and kz6 del, (2) pkz5LS, (3) pkz5del, and (4) pkz6del were obtained from the pCB-UBS54-SNAP original sequence by deletions including the in-frame ATG at 293 and 299 and replacing ATG in 293 to give a low score as described in the materials and methods section (**see Section 2.2.3**). The Kozak sequences are numbered from the N-terminal end or 5′ positions including 1st AUG or TISs.

3.3 PRODUCTION OF RECOMBINANT PLASMIDS THROUGH MOLECULAR CLONING

3.3.1 Isolation and restriction digestion of plasmid pUC57, pCB, and pMT

After in silico design**,** these predesigned scFv sequences were aligned to their parental patent sequences using CLC genomic workbench 11 software to confirm the integrity of CDRs and FRs regions. Then, the newly designed scFvUBS54, OCAb9-1*, ID405/YW353/h5-Thy1/h-STL002 were sent to a company (GenScript) for synthesis into a commercial pUC57 vector. Once synthesized plasmid "scFvUBS54/ID405/hSTL002 and OCAb9-1/YW353/h5-Thy1" is obtained from GeneScript, the chemically competent *E. coli* DH5α cells are transformed with commercial plasmid vectors (pUC57-scFvUBS54/ID405/hSTL002/OCAb9-1/YW353/h5-Thy1) through heat shock, as discussed in **Section 2.3.1**. In parallel, the backbone plasmid (vector) for mammalian expression (pCB-AnnexinV-SNAP) and *E. coli* expression (pMT-J3-SNAP) is isolated from glycerol stock of *E. coli* DH5-α donated by Krupa and Sanele (The University of Cape Town, Faculty of Health Sciences, Medical Biotechnology and Immunotherapy Unit). The pCB-Annexin V-SNAP, the backbone of mammalian expression vector, is a modified version of parental plasmid pSNAP$_f$ Vector (NEB: N9183) and pSNAP-tag® (T7)-2-Vector (NEB: N9181) or pMS[209] vector, a derivative of the pSecTag2 vector. pMT vector, the backbone of bacterial expression vector, was derived from pET27b vector (Novagen). Thereafter, inserts (scFvUBS54/ID405/h-STL002 and OCAb9-1/YW353/h5-Thy1) to be used for molecular cloning were isolated from glycerol stock of *E. coli* DH5-α, and vector plasmids were extracted

as described in **Section 2.3.1**. The extracted insert and vectors are enzymatically digested with *Sf*i (20000U/ml) and *Not*I (20000 U/ml) restriction. Once the enzymatic digestion reactions were completed, digested DNA fragments were loaded and visualized onto 1.2% agarose gel. Identification of digested scFv fragments was performed through direct visualization of 732 bp for scFvUBS54 (**Fig. 18**), 747 bp for scFvID405 (**Fig. 20**), 729 bp for scFvh-STL002 (**Fig. 21**). In addition, corresponding backbone pCB-annexin V-SNAP digested fragments giving rise to 956 bp (**Fig. 19**) for Annexin V and 7276 bp backbone (**Fig. 19**) molecular weights respectively are visualized. Identification of digested scFvOCAb9-1, scFvYW353, and scFvh5-Thy1 fragments was performed through direct visualization of 687 bp (**Fig. 22**), 678 bp, and 732 bp (**Fig. 23**) molecular weights, respectively. In addition, corresponding backbone pMT-J3-SNAP was digested to give rise to 957 bp fragments of J3 and 7279 bp (**Fig. 22**) of backbone pMT Vector.

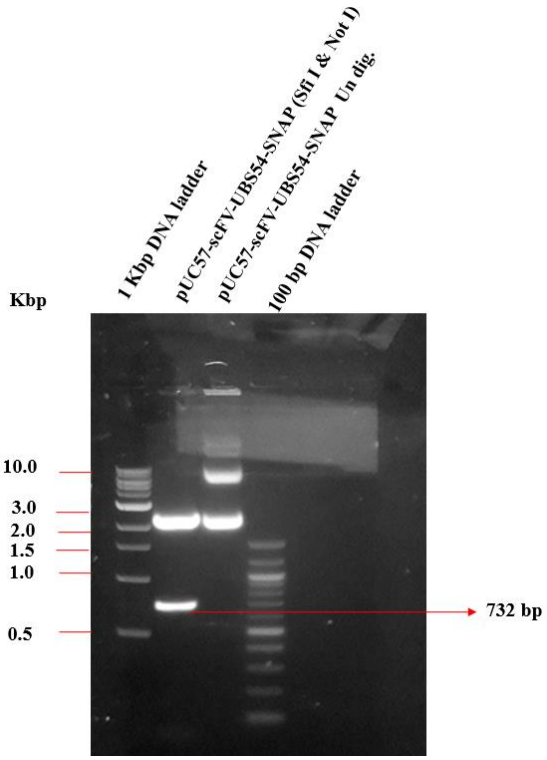


Figure 18: Agarose gel electrophoresis illustrating the restriction digestion pattern of pUC57-scFvUBS54

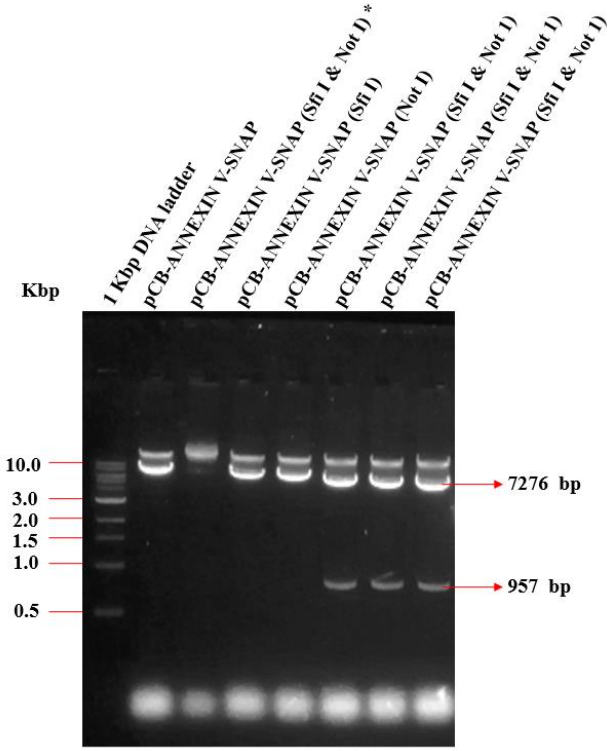


Figure 19: Agarose gel electrophoresis illustrating the restriction digestion pattern of pCB-ANNEXIN V-SNAP

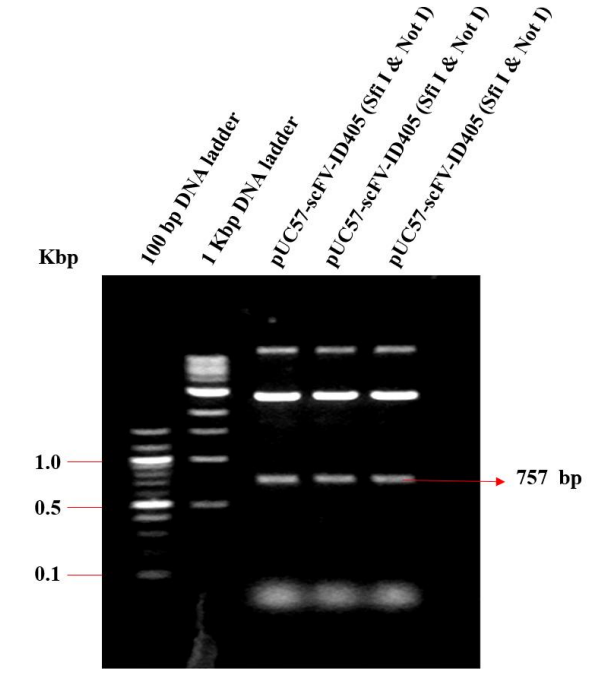


Figure 20: Agarose gel electrophoresis illustrating the restriction digestion pattern of pUC57-scFvID405

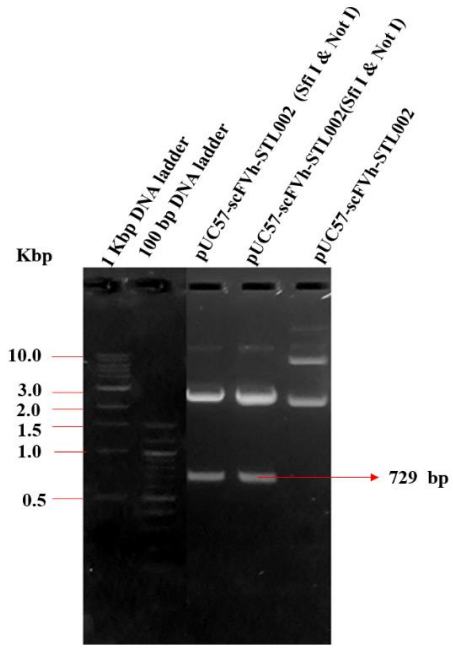


Figure 21: Agarose gel electrophoresis illustrating the restriction digestion pattern of pUC57-scFvh-STL002

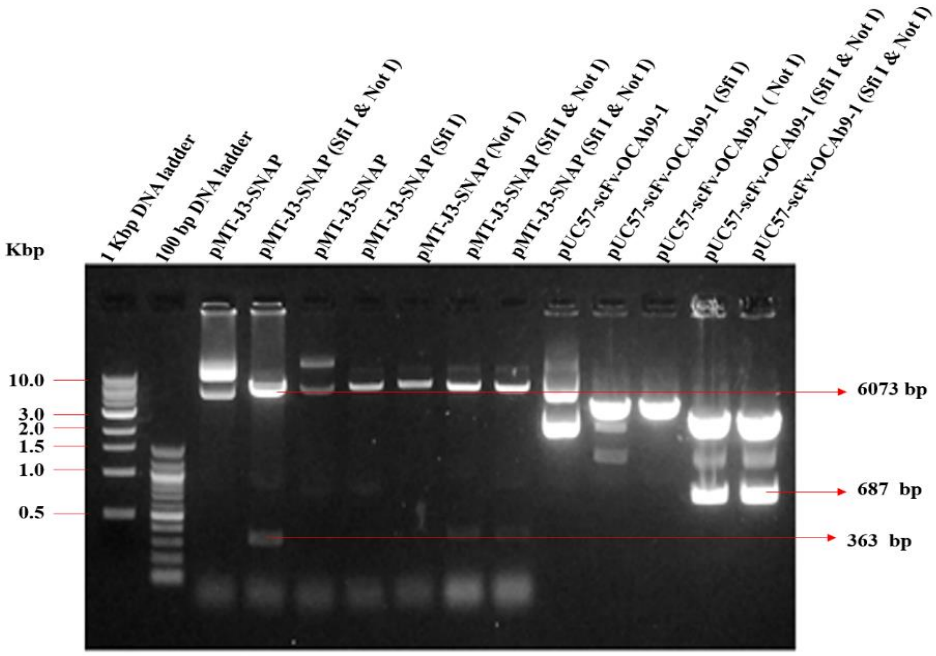


Figure 22: Agarose gel electrophoresis illustrating the restriction digestion pattern of pMT-J3-SNAP and pUC57-scFv-OCAb9-1

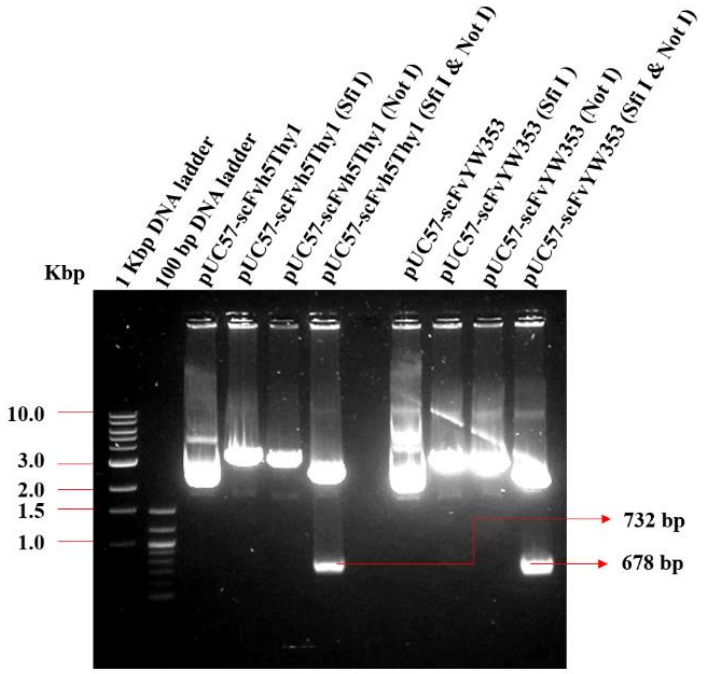


Figure 23: Agarose gel electrophoresis illustrating the restriction digestion pattern of pUC57-scFvh5-Thy1 and pUC57-scFvYW353 plasmids

Figure 18–23: Agarose gel images electrophoresis results of digested pUC57-scFvUBS54 (18). pCB-ANNEXIN V-SNAP (19) and pUC57-scFvID405 (20). pUC57-scFvh-STL002 (21). pMT-J3-SNAP and pUC57-scFvOCAb9-1 (22). pUC57-scFvh5-Thy1/scFvYW353 (23).

Two micrograms of pCB-ANNEXIN V-SNAP, pMT-J3-SNAP, and pUC57-scFvUBS54, pUC57-scFvID405, pUC57-scFvh-STL002, pUC57-scFvOCAb9-1, pUC57-scFvh5-Thy1, & pUC57-scFvYW353 plasmid DNA were enzymatically digested with *Sfi*I and *Not*I restriction enzymes as described in **Section 2.15.3**. Fifty microliters (with 10 µl loading 6x loading dye) corresponding to 2 µg of digested products plasmid DNA is loaded and visualized onto 1.2% agarose, as described in **Section 2.3.3**.

3.3.2. Ligation of DNA and transformation of recombinant clone in *E. coli*

Identified fragments are purified from gel and ligation of scFv fragments scFvUBS54/ID405/h-STL002 to the corresponding backbone pCB-SNAP vector (mammalian expression vector) was carried out using T4 DNA ligase. Similarly, scFvOCAb9-1/scFvYW353/scFvh5-Thy1 fragments purified from gel and ligated to the backbone of pMT-SNAP vector (Bacterial expression vector) using T4 DNA ligase as described in **Section 2.3.4**. This reaction was

successfully performed by using different ratios (1:1, 3:1, and 5:1) of inserts to plasmid vector DNA calculated using the NEB calculator (**Table 12**).

Table 12: The ratio of insert (scFv) to plasmid vector DNA (pCB-SNAP, pMT-SNAP) was calculated using the NEB calculator

INSERT–VECTOR RATIO		
1:1	**3:1:**	**5:1**
scFvUBS54 (5.0 ng): pCB-SNAP (50ng)	scFvUBS54(15.0ng): pCB-SNAP (50ng)	scFvUBS54(30.0ng): pCB-SNAP (50ng)
scFvID405(5.0ng): pCB-SNAP (50ng)	scFvID405 (15.0 ng): pCB-SNAP (50ng)	scFvID405 (26.0 ng): pCB-SNAP (50ng)
scFvh-STL002(5.0 ng): pCB-SNAP (50ng)	scFvh-STL002(15.0ng): pCB-SNAP (50ng)	scFvh-STL002(25.0ng): pCB-SNAP (50ng)
scFvOCAb9-1(6.0ng): pMT-SNAP (50ng)	scFvOCAb9-1(17.0ng): pMT-SNAP (50ng)	scFvOCAb9-1(28.0ng): pMT-SNAP (50ng)
scFvYW353(6.0ng): pMT-SNAP (50ng)	scFvYW353(17.0ng): pMT-SNAP (50ng)	scFvYW353(28.0ng): pMT-SNAP (50ng)
scFvh5-Thy1(6.0ng): pMT-SNAP (50ng)	scFvh5-Thy1(17.0ng): pMT-SNAP (50ng)	scFvh5-Thy1(28.0ng): pMT-SNAP (50ng)

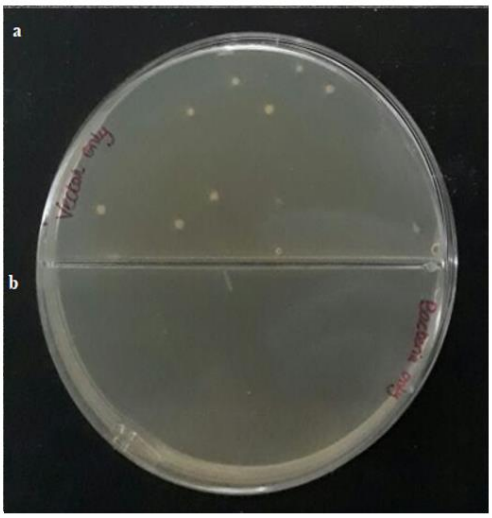
During this reaction, T4 DNA ligase covalently joined each digested insert fragment (scFv UBS54, scFv ID405, scFvh-STL002) to pCB-SNAP vector DNA and other insert fragments (scFvOCAb9-1, scFvYW353, scFvh5-Thy1) to pMT-SNAP vector DNA through their 5´-phosphorylated and 3´-hydroxylated compatible cohesive ends produced by restriction enzyme digestion. Once ligation was completed, successful transformation of ligated plasmid into *E. coli* was initially confirmed by bacterial growth on antibiotic selection medium as described in section 2.3.4.

These antibiotics provided a stable and controlled environment for the growth of bacteria possessing ampicillin-resistant genes and kanamycin-resistant genes in recombinant plasmids. Hence, bacterial cells deprived of recombinant plasmids were unable to grow due to the absence of the ampicillin resistance (AmpR) gene and kanamycin resistance gene (KanR). However, the religated plasmid vector (pCB-SNAP) can grow in ampicillin and (pMT-SNAP) vector in kanamycin-containing media. Some explanations involve the presence of 5′-phosphate groups from vector molecules or incomplete restriction digest of vector to be used for ligation (Fast-

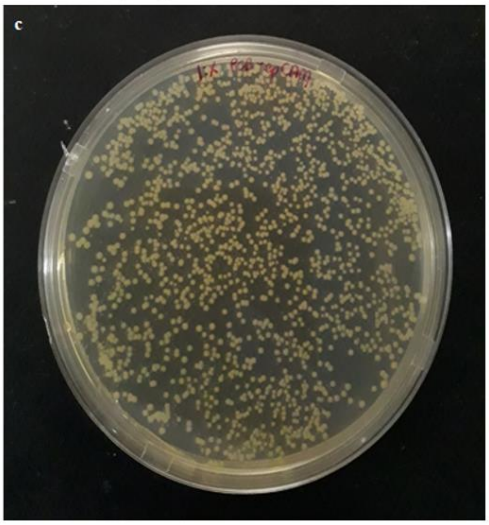
Link™ DNA Ligation Kit protocol, 2012). Therefore, an improvement would be to treat the digested vector backbone with a phosphatase before the ligation reaction (Addgene, 2016).

However, only bacterial clones possessing the pCB-scFvUBS54-SNAP, pCB-scFvID405-SNAP, and pCB-scFvh-STL002-SNAP plasmids were able to grow due to their capacity to express the ampicillin-resistant gene (AmpR) breaking down ampicillin as show in **Fig 24.** Likewise, bacterial clones possessing pMT-scFvOCAb9-1-SNAP, pMT-scFvYW353-SNAP, and pMT-scFvh5-Thy1-SNAP plasmids were able to grow due to their capacity to express the kanamycin-resistant gene (KanR) breaking down Kanamycin as shown in representative **Fig 24**. On completion of the cloning, additional screening of the recombinant plasmids (pCB-scFv UBS54-SNAP, pCB-scFvID405-SNAP, pCB-scFvh-STL002-SNAP and pMT-scFvOCAb9-1-SNAP, pMT-scFvYW353-SNAP, pMT-scFvh5-Thy1-SNAP) were performed using restriction mapping to discriminate them from parental backbone plasmid pCB-AnnexinV-SNAP and pMT-J3-SNAP.

a) Vector alone (pCB-SNAP)
b) Bacteria alone (E. coli *DH5α*)



pCB-scFvUBS54-SNAP
(1:1) Insert: Vector



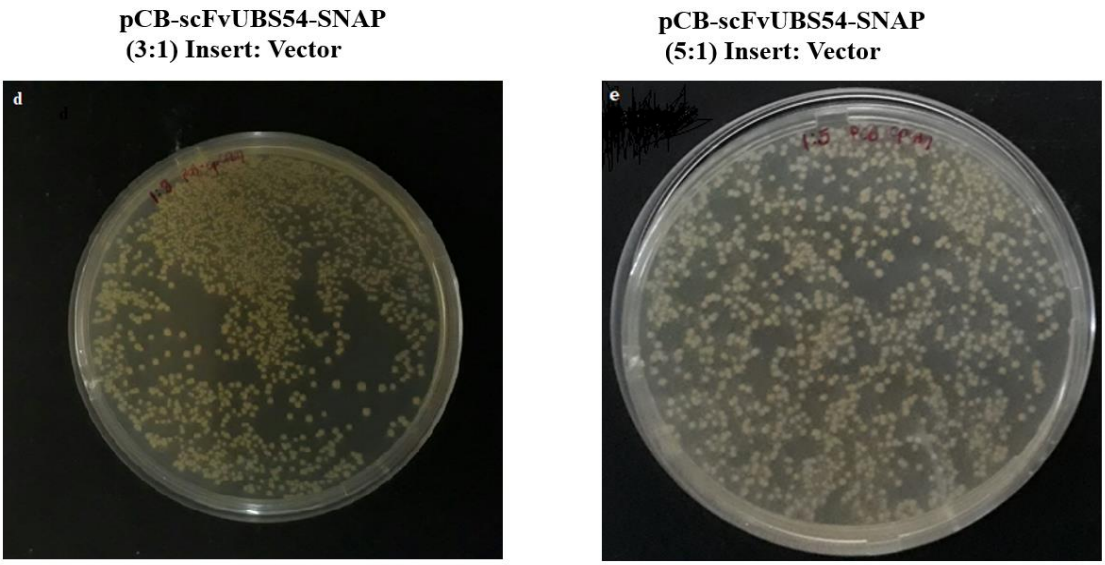


Figure 24: Transformed pCB-scFvUBS54-SNAP bacterial colonies on agar plates containing ampicillin

Figure 24: Transformed pCB-scFvUBS54-SNAP bacterial colonies on agar plates containing ampicillin (200 ng/µL). Recombinant plasmids DNA were obtained through ligation of scFvUBS54/ID405/ h-STL002 insert fragments to pCB-SNAP vector DNA at different molar ratios (1:1, 1:3, and 1:5) and **scFvOCAb9-1/YW353/h5-Thy1-SNAP fragments** were ligated to pMT-SNAP vector DNA at different molar ratios (1:1, 1:3, and 1:5) using T4 DNA ligase as previously described (**Section 2.3.4**). Then, chemically competent *E. coli* DH5α cells were transformed with the ligated product as described in **Section 2.3.4**. Then, transformed bacterial cells harboring the plasmid pCB-scFv-SNAP were plated on agar plates containing ampicillin, and bacterial cells harboring pMT-scFv-SNAP plasmid were plated on agar plates containing kanamycin. Ampicillin and kanamycin selectively promoted the growth of bacterial cells transformed with desired recombinant plasmids at the expense of untransformed bacterial cells or those transformed without the insert or religated vector plasmid vectors only.

Note: The figure of t**ransformed pCB-scFvUBS54-SNAP bacterial colonies** shown as a representation example will hold true for other constructs such as **pCB-scFv/ID405/h-STL002-SNAP** & **pMT-scFvOCAb9-1/YW353/h5-Thy1-SNAP**.

3.4. VALIDATION OF SYNTHESIZED RECOMBINANT PLASMID

3.4.1. Restriction mapping

Restriction mapping is a molecular biology technique, based on the measurement of DNA fragment lengths after restriction endonuclease digestion.[238] Hence, this technique was used to confirm successful cloning by comparing recombinant digested DNA fragments to their parental digested counterpart. To achieve it, both recombinant and parental plasmids DNA are enzymatically digested at unique short, specific sequences called restriction enzyme consensus sites. The best restriction enzyme consensus sites chosen occurred inside the insert fragment and in the vector (making it possible to determine if the fragment had been incorporated and determine the orientation of the fragment if the same sticky ends were used). Hence, during this experiment both parental and recombinant plasmids were successively digested with multiple restriction enzymes, resulting in different digested fragment patterns in both digested plasmids. To successfully performed restriction mapping, an agarose simulation of the endonucleases reactions were performed on SnapGene software.

From this, *Blp*I (6963 and 1045 bp digested fragments) was identified as the restriction enzyme of choice for pCB-scFvUBS54-SNAP, whereas for pCB-scFv-ID405-SNAP, *BamH*I (6156 and 1867 bp digested fragments) and *Pvu*II (3412, 1895, 1748, 723, and 245 bp digested fragments) are the best choices. For the plasmid pCB-scFvh-STL002-SNAP, *BamH*I is a good choice (6162 and 1846 bp fragments) based on SnapGene simulation, as seen in **Figs. 25, 27, and 29**. Subsequently for the bacterial expression plasmids pMT-OCAb9-1-SNAP, pMT-scFvh5-THY1-SNAP, and pMT-scFv-YW353-SNAP, *Pst*I (4348, 1562, 460, 231, and 159 bp digested fragments), *Pvu*II (4102, 1611, 999, 93 bp fragments and 4066, 1593, 999, and 93 bp digested fragments) are found to be best suited based on SnapGene simulation, as seen in **Figs. 31 and 33**. After that, corresponding recombinant plasmids were subjected to particular restriction enzyme digestion and run on a 1.2% agarose gel (see **Section 2.3.5**). Then, digested DNA fragment sizes were confirmed for each enzyme digest and compared with the SnapGene simulation (**Figs. 25-34**). There is a difference in fragment size between digested parental plasmid pCB-Annexin V-SNAP, pMT-J3-SNAP, and plasmid of scFv-SNAP, thus confirming successful cloning.

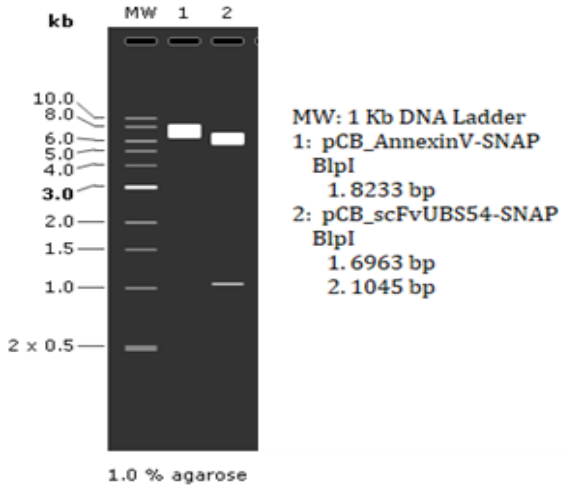


Figure 25: Restriction digestion simulation of the predicted pCB-Annexin V-SNAP and pCB-scFv-UBS54-SNAP enzymatic digestion pattern

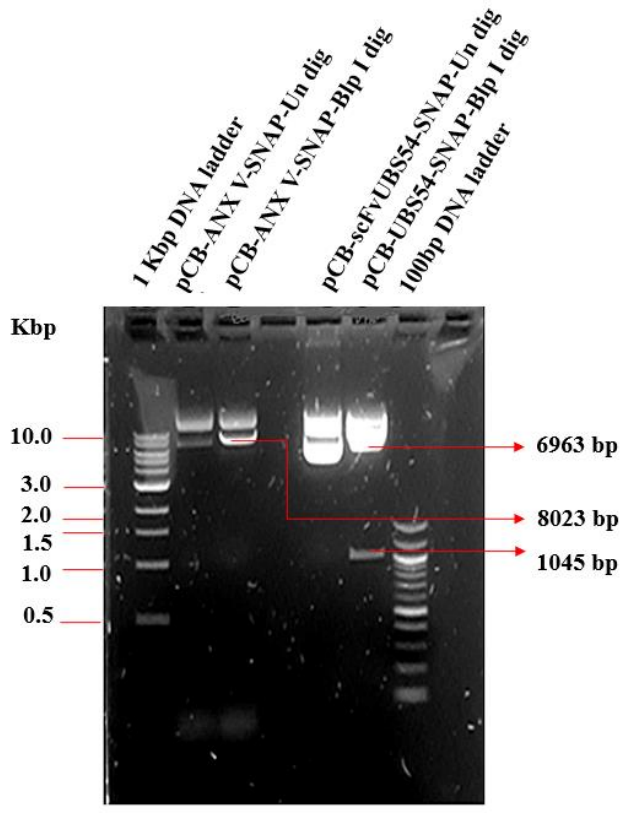


Figure 26: Restriction digestion of the predicted pCB-Annexin V-SNAP and pCB-scFv-UBS54-SNAP with *Blp*I restriction enzyme

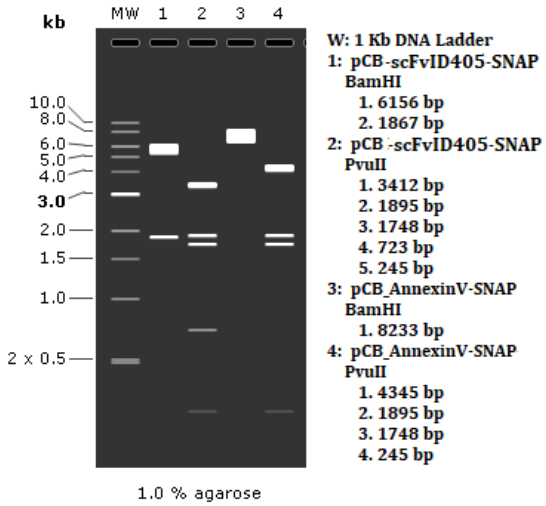


Figure 27: Restriction digestion simulation of the predicted pCB-scFvID405-SNAP and pCB-Annexin V-SNAP enzymatic digestion pattern

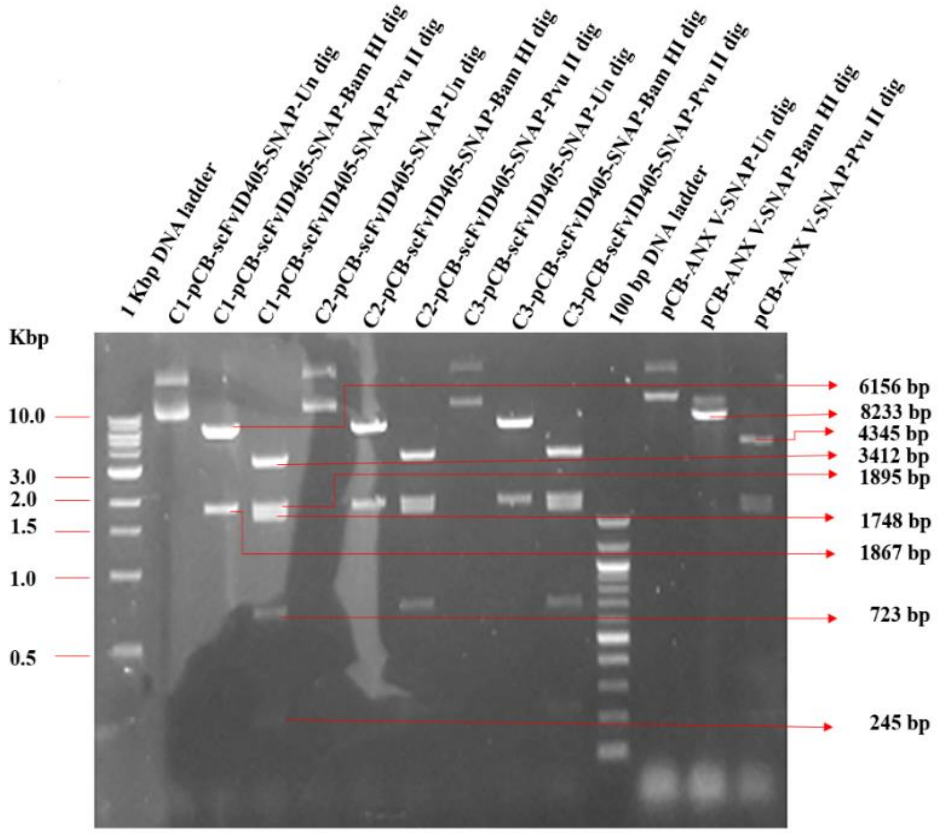


Figure 28: Restriction digestion of the predicted pCB-scFvID405-SNAP and pCB-Annexin V-SNAP with *BamH*I and *Pvu*II restriction enzymes

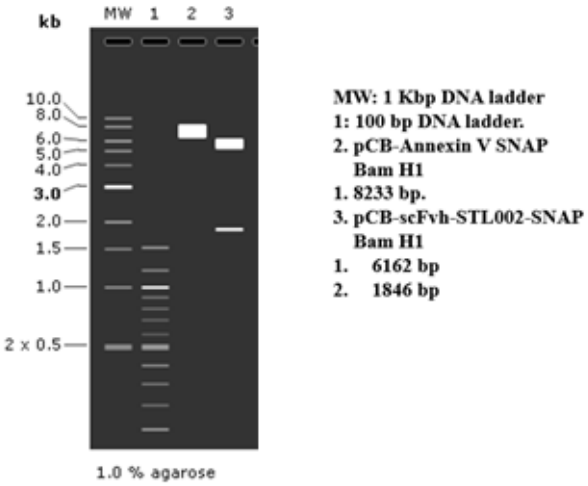


Figure 29: Restriction digestion simulation of the predicted pCB-scFvh-STL002-SNAP and pCB-Annexin V-SNAP enzymatic digestion pattern

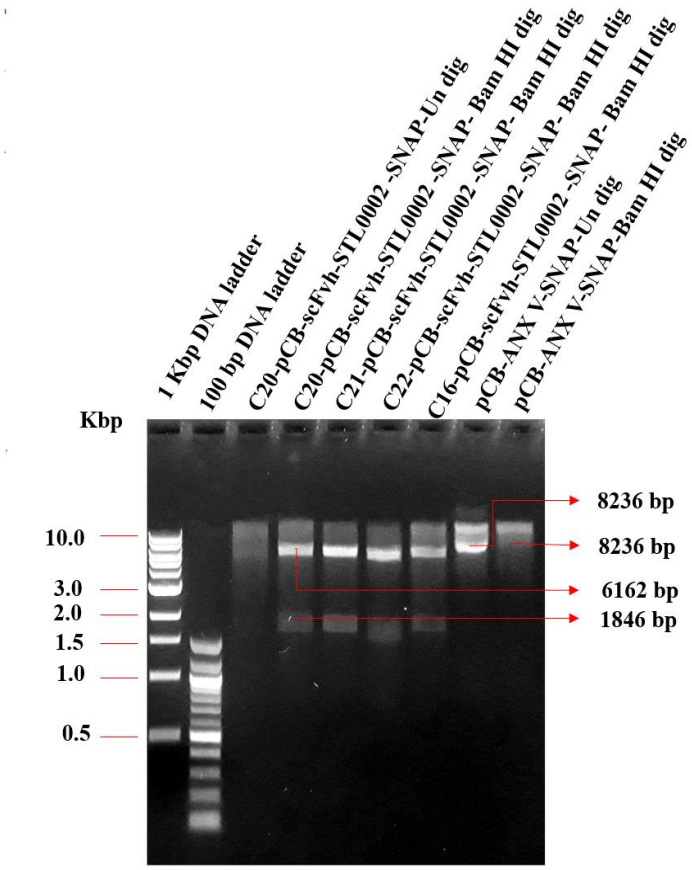


Figure 30: Restriction digestion of the predicted pCB-scFvh-STL002-SNAP and pCB-Annexin V-SNAP with *BamH*I

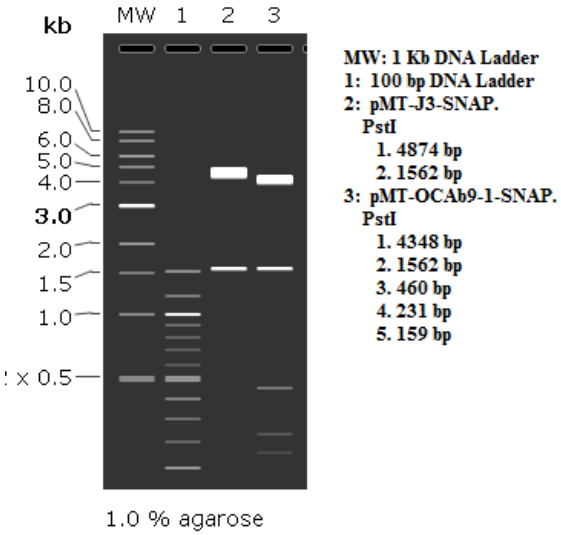


Figure 31: Restriction digestion simulation of the prediction of pMT-J3-SNAP and pMT-OCAb9-1-SNAP enzymatic digestion pattern

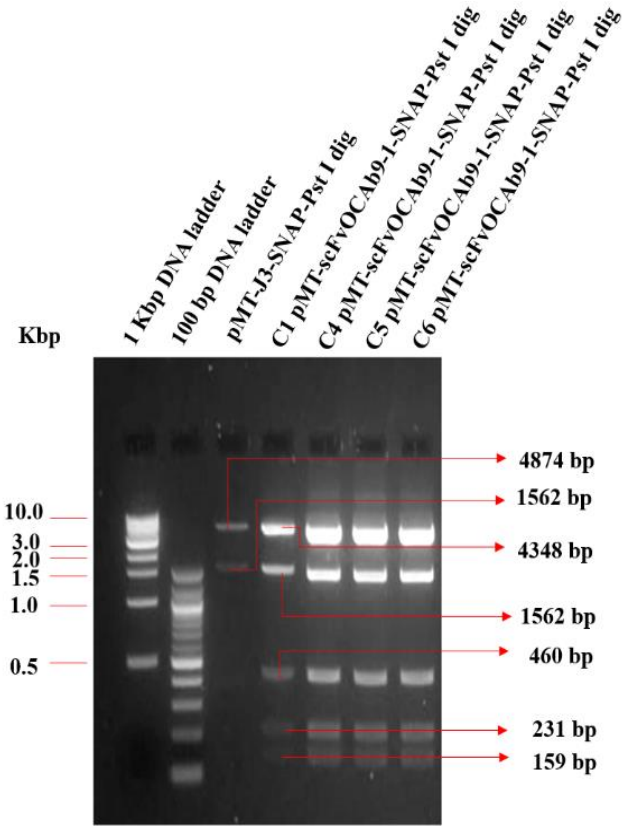


Figure 32: Restriction digestion of the predicted of pMT-J3-SNAP and pMT-OCAb9-1-SNAP with restriction enzyme *Pst*I

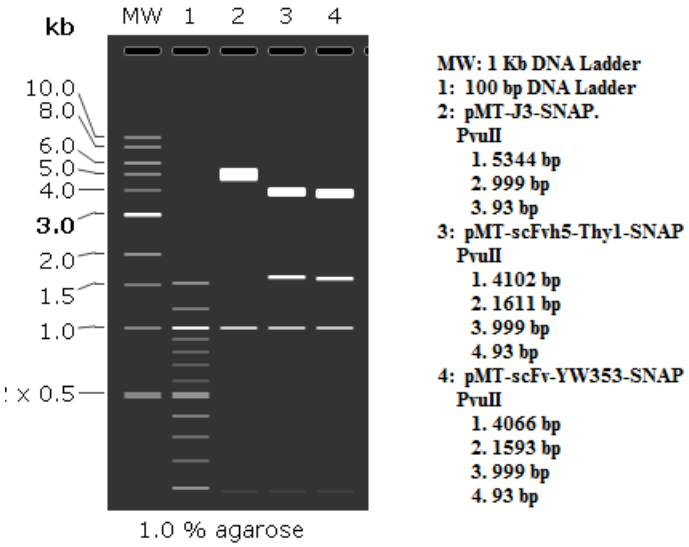


Figure 33: Restriction digestion simulation of the predicted of pMT-J3-SNAP, pMT-scFvh5-THY1-SNAP, pMT-scFv-YW353-SNAP enzymatic digestion pattern

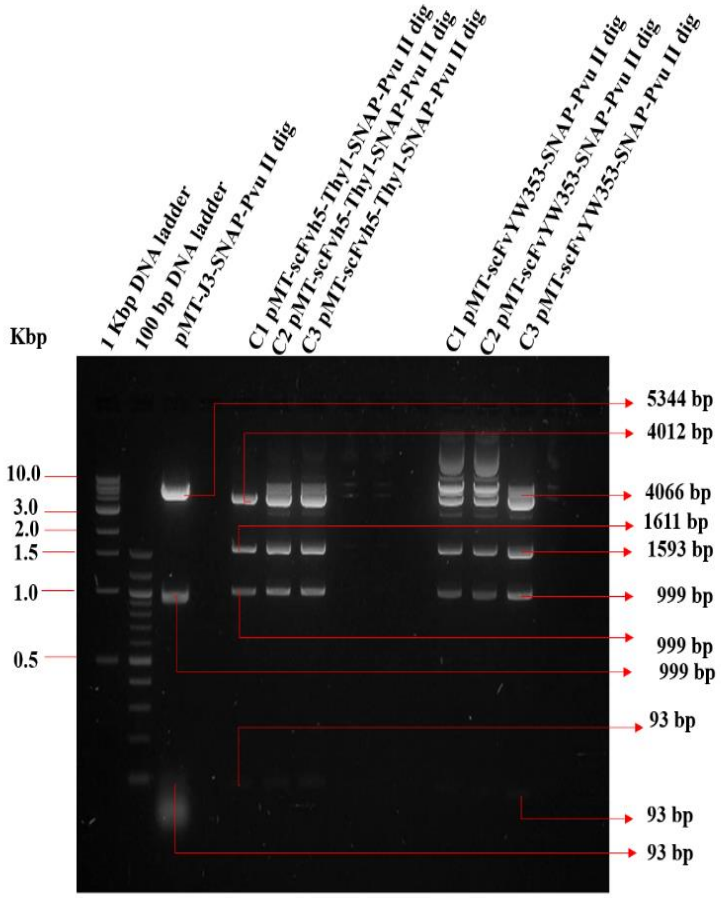


Figure 34: Restriction digestion of the predicted pMT-J3-SNAP, pMT-scFvh5-THY1-SNAP, pMT-scFv-YW353-SNAP with *Pvu*II.

Figure 25-34: Images of agarose gel electrophoresis of digested pCB-scFv-SNAP plasmids and pMT-scFv-SNAP plasmids. In silico digestion of 25) pCB-Annexin V-SNAP and pCB-scFvUBS54-SNAP with *Blp*I restriction enzymes. 27) pCB-Annexin V-SNAP and pCB-scFv-ID405-SNAP with *BamH*I and *Pvu*II restriction enzymes. 29) pCB-Annexin V-SNAP and pCB-scFvhSTL002-SNAP with *BamH*I restriction enzymes. 31) pMT-J3-SNAP and pMT-scFvOCAb9-1-SNAP with *Pst*I restriction enzymes. 33) pMT-J3-SNAP and pMT-scFvh5-THY1, pMT-YW353-SNAP with *Pvu*II restriction enzymes was simulated on SnapGene software to predict the sizes of different digested DNA fragments as shown in (25, 27, 29 and 31, 33) parental plasmids pCB-Annexin V-SNAP and pMT-J3-SNAP, recombinant plasmid pCB-scFvUBS54/ID405/h-STL002-SNAP and pMT-scFvOCAb9-1/h5THY1/YW353-SNAP were extracted from bacterial cells and digested with *Blp*1, *BamH*I and *Pvu*II, *Pst*I and *Pvu*II restriction enzymes at 37^{o}C overnight. Thereafter, reaction products were loaded onto 1% agarose gel, for visualization of pCB-ANNEXINV-SNAP and pCB-scFvUBS54-SNAP (26), pCB-ANNEXIN V-SNAP and pCB-scFvID405-SNAP (28), pCB-ANNEXIN V-SNAP and pCB-scFvhSTL002-SNAP (30), pMT-J3-SNAP and pMT-scFvOCAB9-1-SNAP (32) pMT-J3-SNAP and pMT-scFv-h5THY1/YW353-SNAP (34) digested fragments on blue light excitation (see **Section 2.3.5**).

3.4.2. Sequencing ORF region of plasmid

3.4.2.1. Sequencing selected region (ORF) of selected clones harboring pCB-scFv-SNAP

The selected recombinant plasmids were sent to Inqaba Biotechnical Industries for DNA sequencing using the Sanger sequencing method. For this purpose, a CMV-F primer, as well as specific internal primers (scFvUB54/scFvID405/scFvh-STL002 and SNAP), were used to help in the sequencing of the cloned product. SnapGene® (version 3.1.1, GSL Biotech, Chicago) was used to analyze the resulting sequences. The clones of all the three recombinant plasmid sequences (scFvUBS54, scFvID405, and scFvh-STL002) showed 100% sequence homology to their corresponding in silico ORF sequences (**Figs. 35–37**).

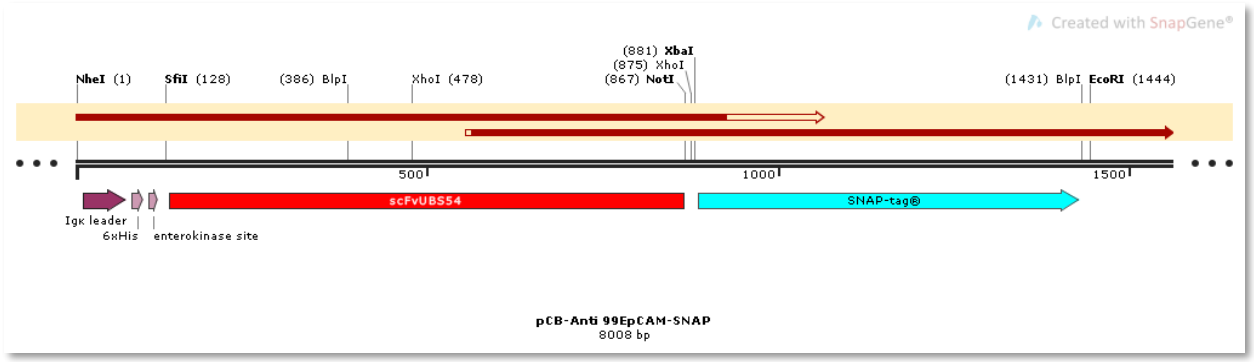


Figure 35: Sequencing map covering ORF of **cloned pCB-scFv-UBS54-SNAP**

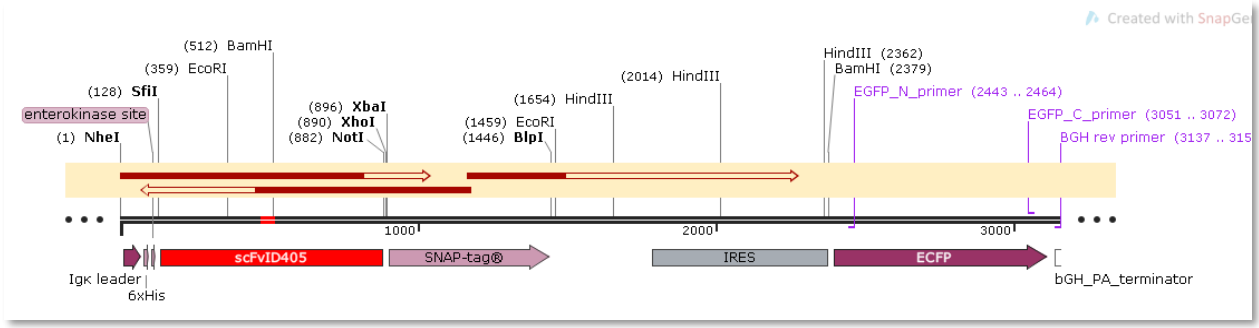


Figure 36: Sequencing map covering ORF of **cloned pCB-scFvID405-SNAP**

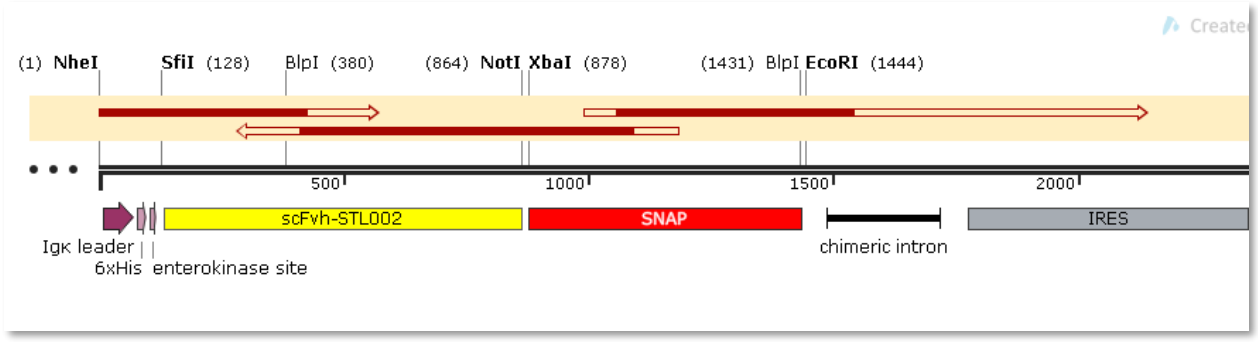


Figure 37: Sequencing map covering ORF of **cloned pCB-scFvh-STL002-SNAP**

Figure. 35–37: Sequencing map covering ORF of pCB-scFv-UBS54-SNAP (35), pCB-scFvID405-SNAP (36), and scFvh-STL002-SNAP (37) red arrow describing 100% homology of the sequencing results to the corresponding ORF of the consensus sequence.

3.4.2.2. Sequencing selected region (ORF) of selected clones harboring pMT-scFv-SNAP

The selected bacterial expression recombinant plasmids were sent to Inqaba Biotechnical Industries for DNA sequencing using the Sanger sequencing method. For this purpose, a T7primer, as well as specific internal primers (scFvOCAb9-1/scFvh5-Thy1/scFvYW353 and SNAP), were used to help in the sequencing of the cloned product. SnapGene® (version 3.1.1,

GSL Biotech, Chicago) was used to analyze the resulting sequences. The clones of all the three recombinant plasmid sequences (scFvOCAb9-1, scFvh5-Thy1, and scFvYW353) showed 100% sequence homology to their corresponding in silico ORF sequence (**Figs. 38–40**).

In spite of successful cloning of bacterial expression vector, we are unable to express the fusion protein in *E. coli* bacterial expression host. Thus, further discussion will be amenable to mammalian expression vector.

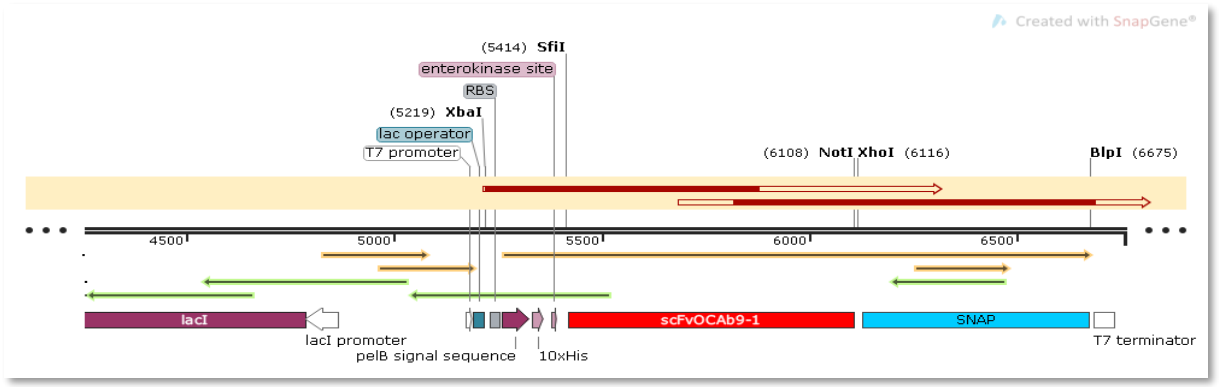


Figure 38: Sequencing map covering ORF of **cloned pMT-scFvOCAb9-1-SNAP (6760 bp)**

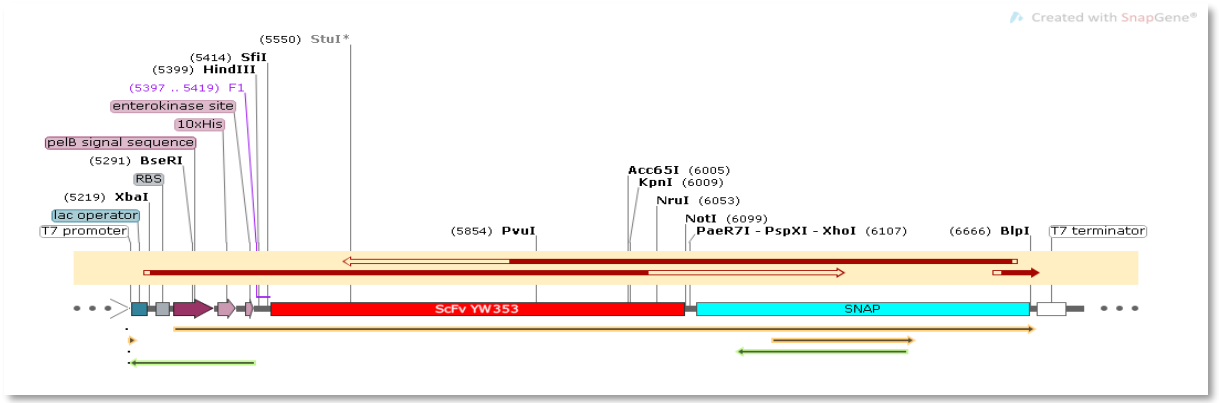


Figure 39: Sequencing map covering ORF of **cloned pMT-scFvYW353-SNAP (6751 bp)**

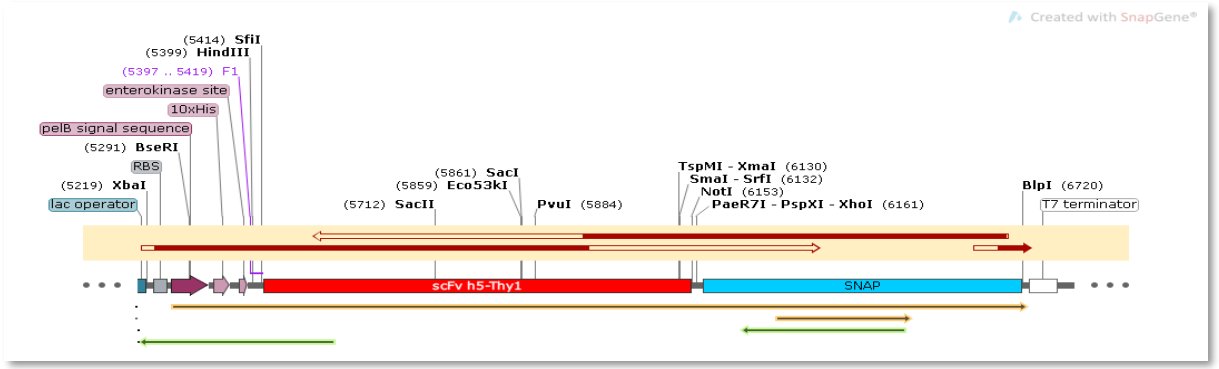


Figure 40: Sequencing map covering ORF of **cloned pMT-scFvh5-Thy1-SNAP (6805 bp)**

Figure 38–40: Sequencing map covering ORF of pMT-scFvOCAb9-1-SNAP (38), pMT-scFvYW353-SNAP (39), and scFvh5-Thy1-SNAP (40), dark red arrow describing 100% homology of the sequencing results to the corresponding ORF of the consensus sequence.

3.5. MAMMALIAN EXPRESSION OF ANTIBODY-BASED FUSION PROTEIN

3.5.1. Selecting the best transfection agent

The verified recombinant scFvUBS54-SNAP transgene of interest is chemically transfected to mammalian cells (HEK293T) for ectopic expression. There are multiple transfection reagents commercially available, we chose X-treme Gene HP and TurboFect both works by two different mechanisms, the former works on liposome-based method and the latter one works on non-liposome (Cationic polymer)–based method. However, most of the transfection reagents have cytotoxic effects on the cells, especially when transfection reagent/plasmid amounts are not optimized for a given cell line.[239] We initially estimated transfection efficiency using a flow cytometer (FACSCalibur) for these two different transfection agents in the HEK293T cell line by following prescribed protocol and we found TurboFect cationic polymer-based transfection agent shows low transfection rate compared to Xtremegene HP (see **Table 13**). Therefore, we chose Xtreme gene HP for further analysis.

Table 13: Transfection efficiency of two different transfection agents

	X-treme gene HP	**Turbofect**
Construct	scFvUBS54-SNAP	scFvUBS54-SNAP
DNA purity(absorbance)	260/280 (1.91) 260/230 (2.13)	260/280 (1.91) 260/230 (2.18)
DNA Concentration	350 ng/µL	953 ng/µL
No of cells seeded	2.3×10^5 cells/dish	3×10^5 cells/dish
% of HEK293T cells (Visual estimation)	70 to 90%	70 to 90%
Ratio of plasmid: Transfection agent	1:3 (3 µg:9 µLL)	1:3 (3 µg:9 µL)
Incubation time	15 min	15 min
Transfection efficiency	53%	49%
Instrument	FACSCalibur	FACSCalibur

3.5.2 Transfection of HEK293T mammalian cells with recombinant plasmids and production of recombinant scFv-UBS54-SNAP, scFvID405-SNAP, and scFvh-STL002-SNAP fusion proteins

Recombinant plasmids (pCB-scFvUBS54-SNAP, pCB-scFvID405-SNAP, and pCB-scFvh-STL002-SNAP) were introduced in HEK293T cells by transfection. Positively transfected cells were identified through microscopic visualization of cytosolic enhanced green fluorescent protein (eGFP) expression, using ZOE fluorescent cell imager 1450031 microscopes (Biorad, USA) (**Section 2.4.2** and **Figs. 41, 43, and 45**). However, the transfection efficiency of recombinant pCB-scFvUBS54-SNAP, pCB-scFvID405-SNAP, and pCB-scFvh-STL002-SNAP plasmids into HEK293T cells (revealing 4.63%, 4.99%, and 9.72% transfection efficiency) was assessed using flow cytometry, based on the cells' viability and ability to express eGFP or not (**Section 2.4.3.2** and **Figs. 41, 43, and 45**). One of the possible reasons for poor transfection is the amount of transfection agent and the plasmid is used at a ratio of 1:3 suggested in the data sheet. Due to the very low transfection efficiency, enrichment of eGFP-positive HEK293T cells was performed by zeocin selection for several weeks. This antibiotic (Zeocin) promoted the growth of eGFP-positive HEK293T cells (possessing zeocin resistant gene) at the expanse of eGFP-negative cells deprived of the zeocin resistant gene. These positively transfected HEK293T cells were optimally maintained in culture over two months under zeocin selection (see **Section 2.4.4.1 and 2.4.4.2**) and a positive population of eGFP-expressing cells has enriched from 4.63 % to 89.76%, from 4.9% to 72.94, and from 9.72% to 77.42% for scFvUBS54-SNAP, scFvID405-SNAP, and scFv-h-STL002-SNAP prior supernatant harvest (**Figs. 42, 44, and 46** and **Table 14**). Then, scFvUBS54-SNAP, scFvID405-SNAP, and scFvh-STL002-SNAP fusion proteins were extracellularly expressed and secreted in culture media through the IgK-leader sequence. Once in culture media, these recombinant fusion proteins were harvested every 4 days (500 mL) for 8 weeks and its presence was detected using Sandwich ELISA (see Section 2.4.5 and **Fig. 50**). Finally, isolation of scFv-SNAP fusion proteins from cell culture supernatant was performed through affinity chromatography using 6xHis, polyhistidine (see **Section 2.5.1**. and **Fig. 51**).

Based on GFP protein expression, transfection efficiency was quantified in HEK293T cells transfected after 72–96 h of chemical transfection (Xtreme Gene Hp) with DNA- plasmid containing eGFP reporter. The % of live cells is determined based on the exclusion of 7-AAD

and % of GFP expressing cells calculated before and after Zeocin selection for 8–12 weeks (300 µg/mL for first 2 weeks and 100 ug/ml for remaining selection time) or till at the time of harvesting cell-free supernatant: Untransfected HEK293T cells (negative control) maintained in subculture for several weeks without Zeocin selection. HEK293T cells were transfected with a plasmid expressing the scFvUBS54/ID405/h-STL002-SNAP fusion protein and the transfection agent Xtreme Gene HP at a ratio of 1:3.

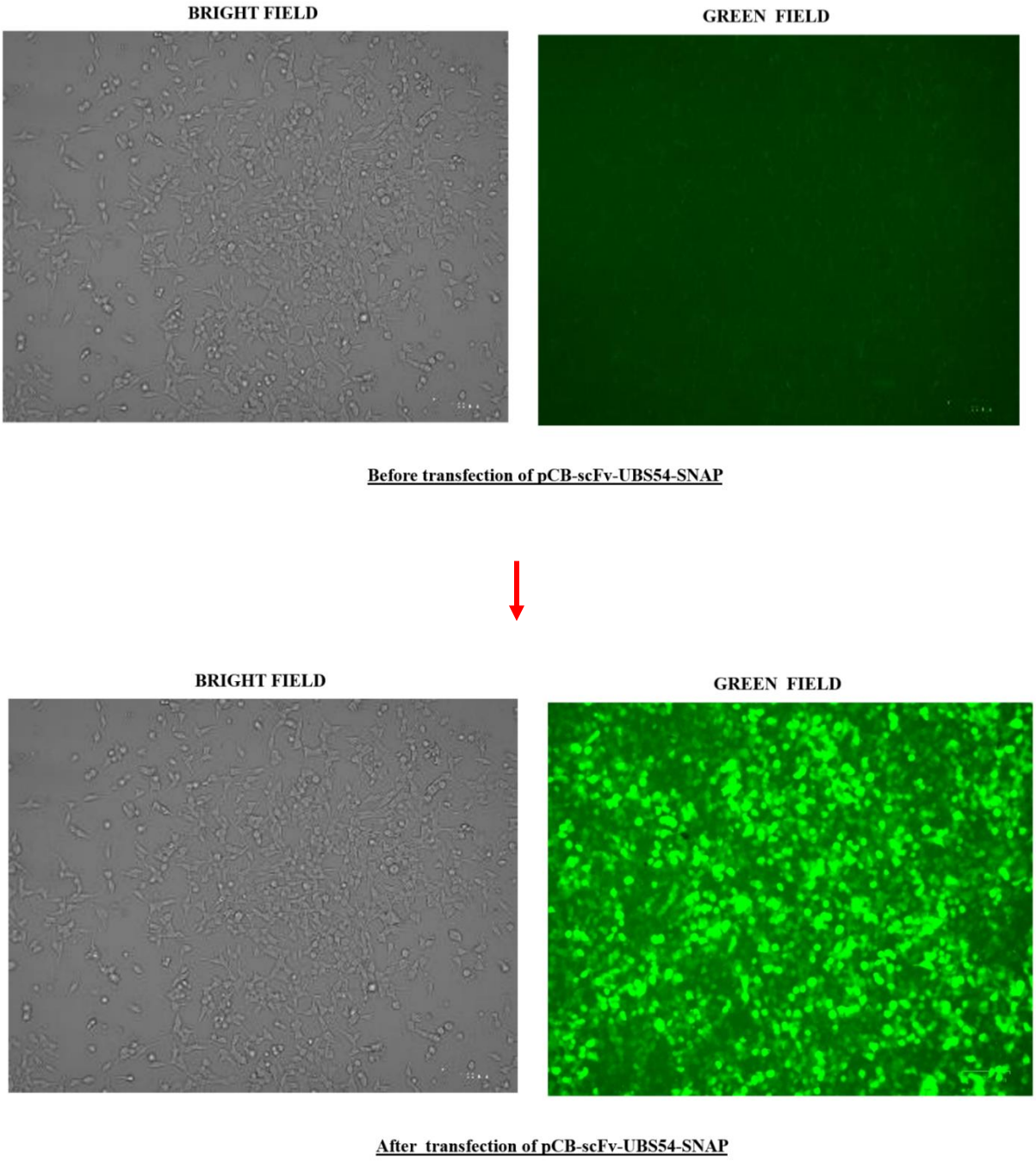


Figure 41: Transfection of pCB-scFv-UBS54-SNAP

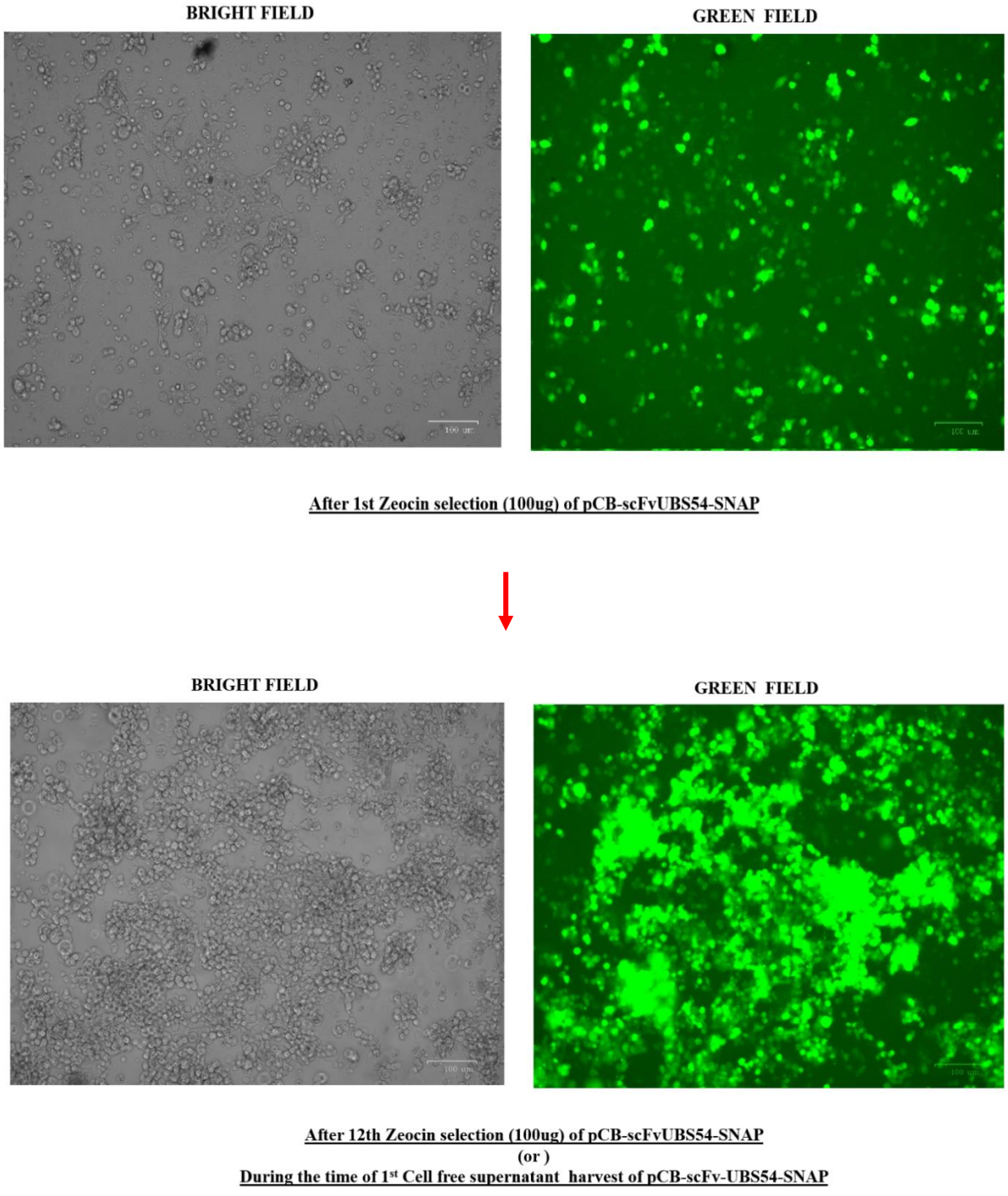


Figure 42: Enriching the population of pCB-scFvUBS54-SNAP transfected cell under zeocin selection pressure.

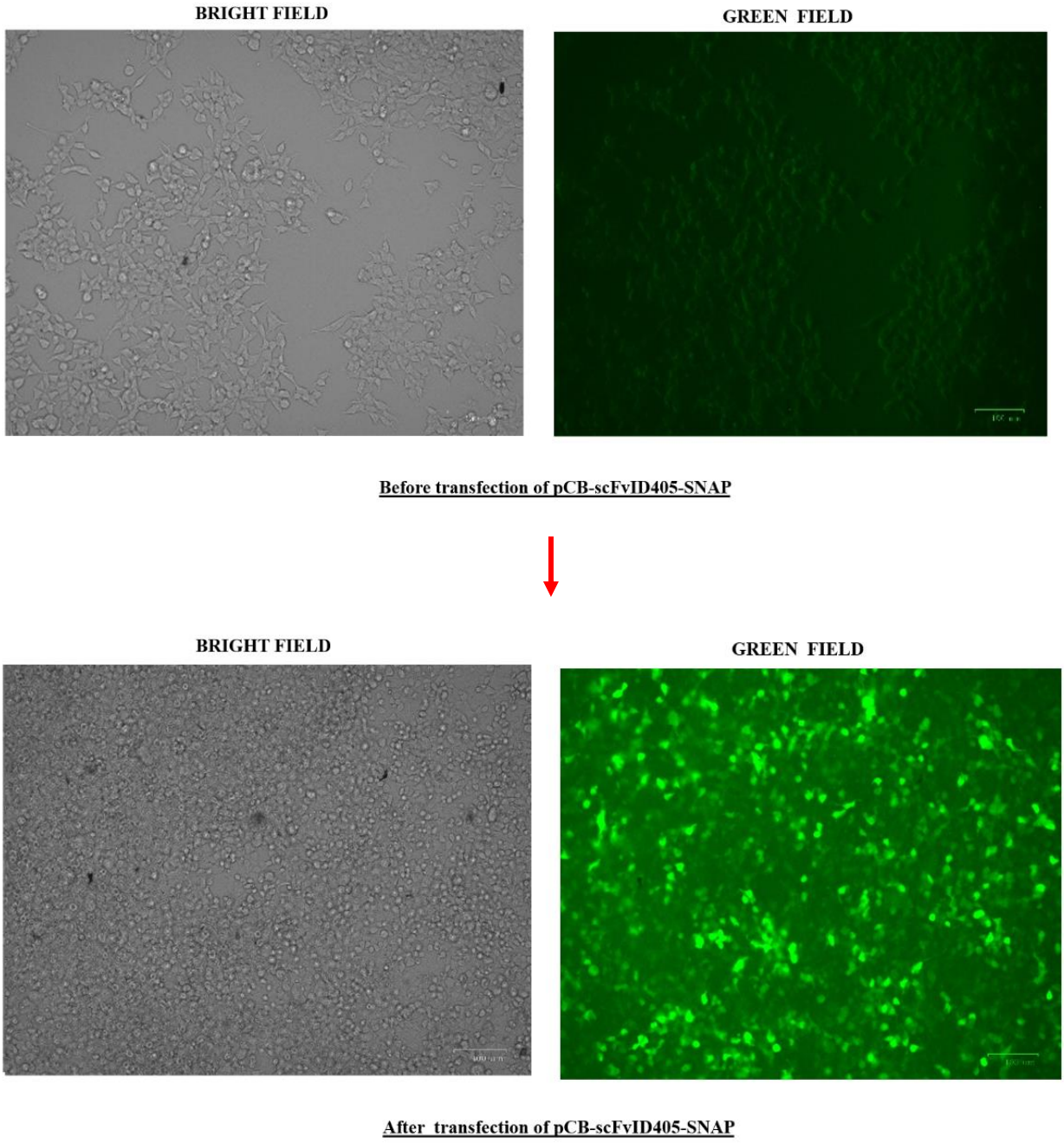


Figure 43: Transfection of pCB-scFv-ID405-SNAP in HEK293T cell line

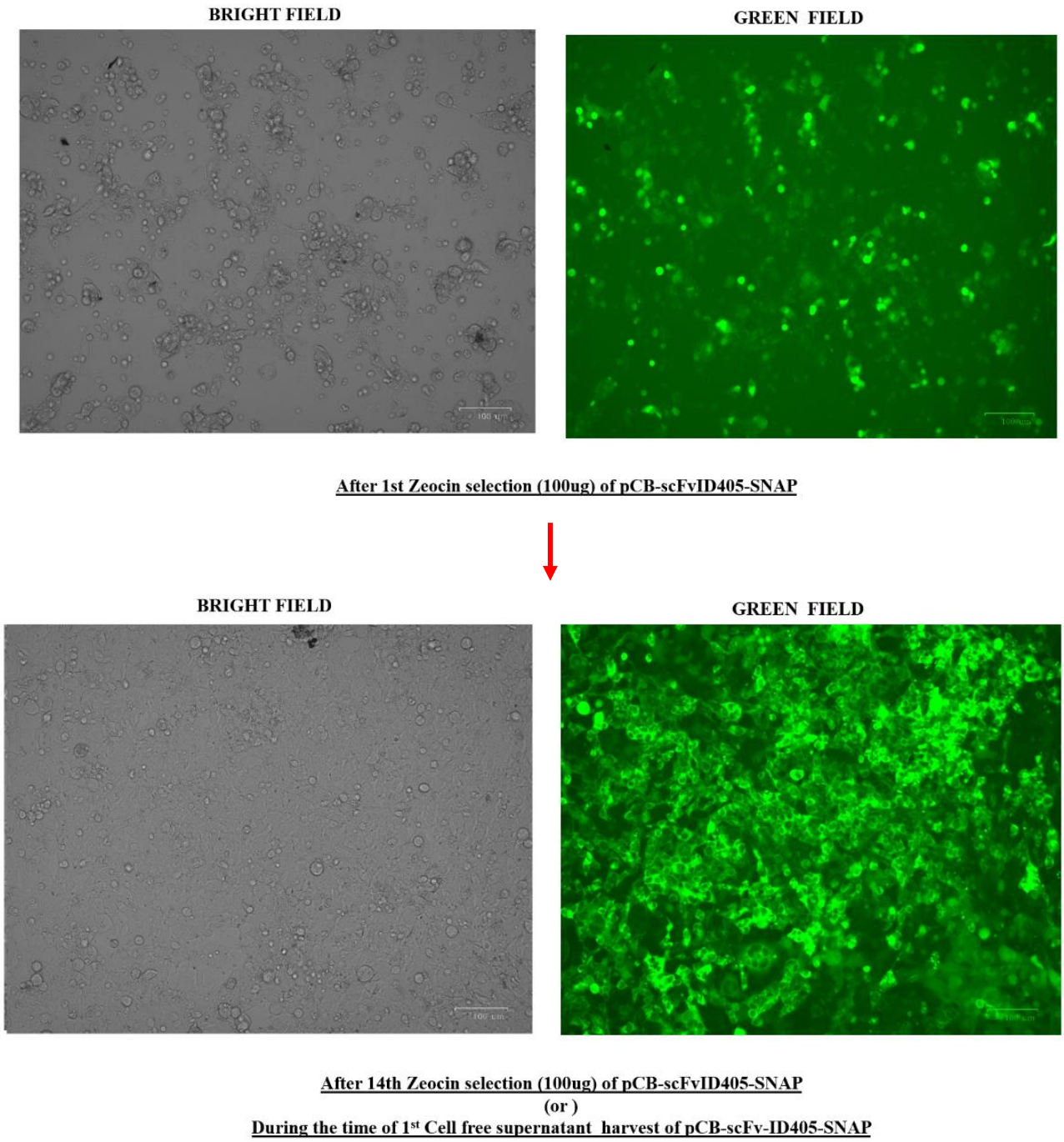


Figure 44: Enriching the population of pCB-scFv-ID405-SNAP cell by Zeocin selection

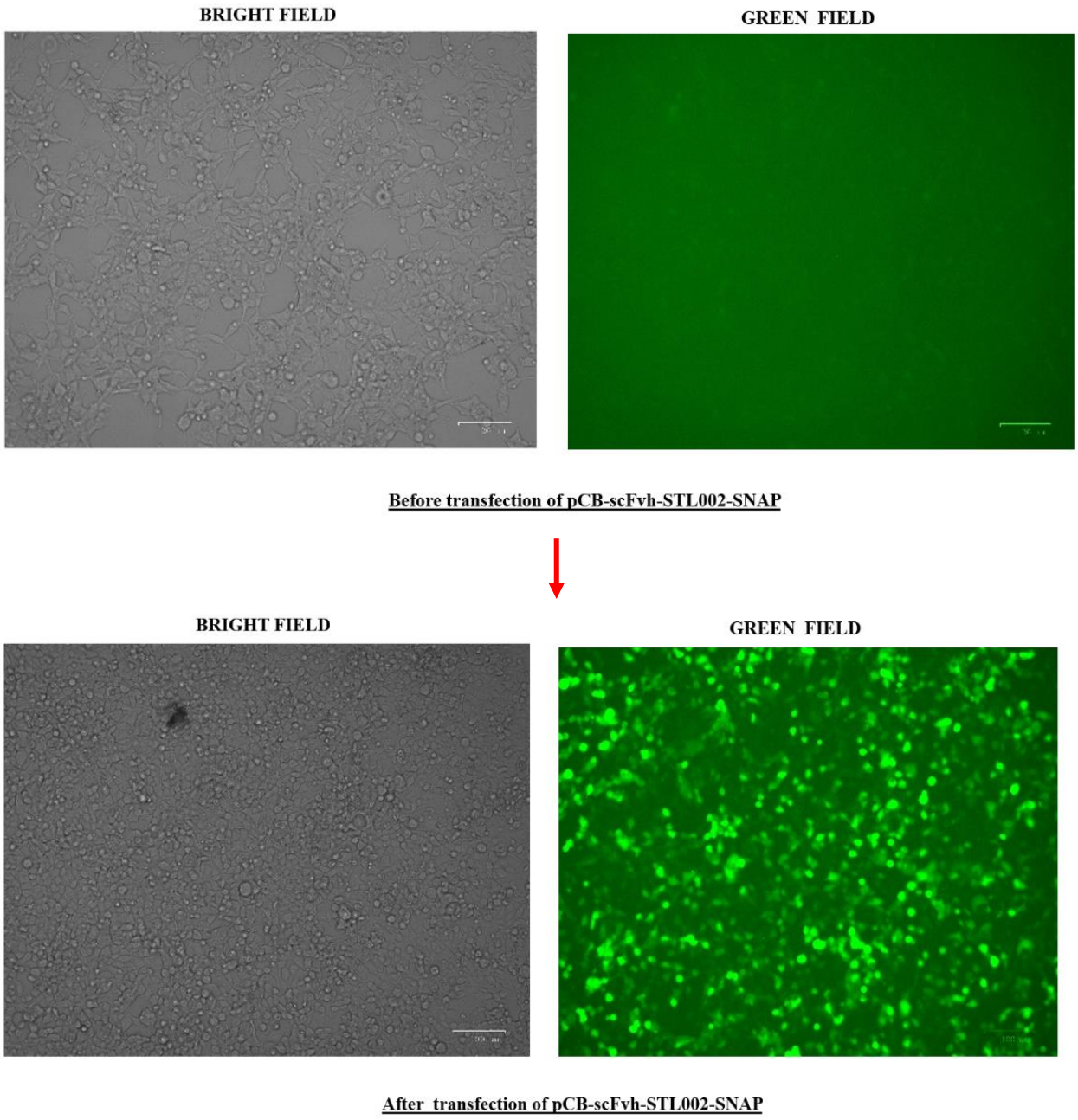


Figure 45: Transfection of pCB-scFvh-STL002-SNAP

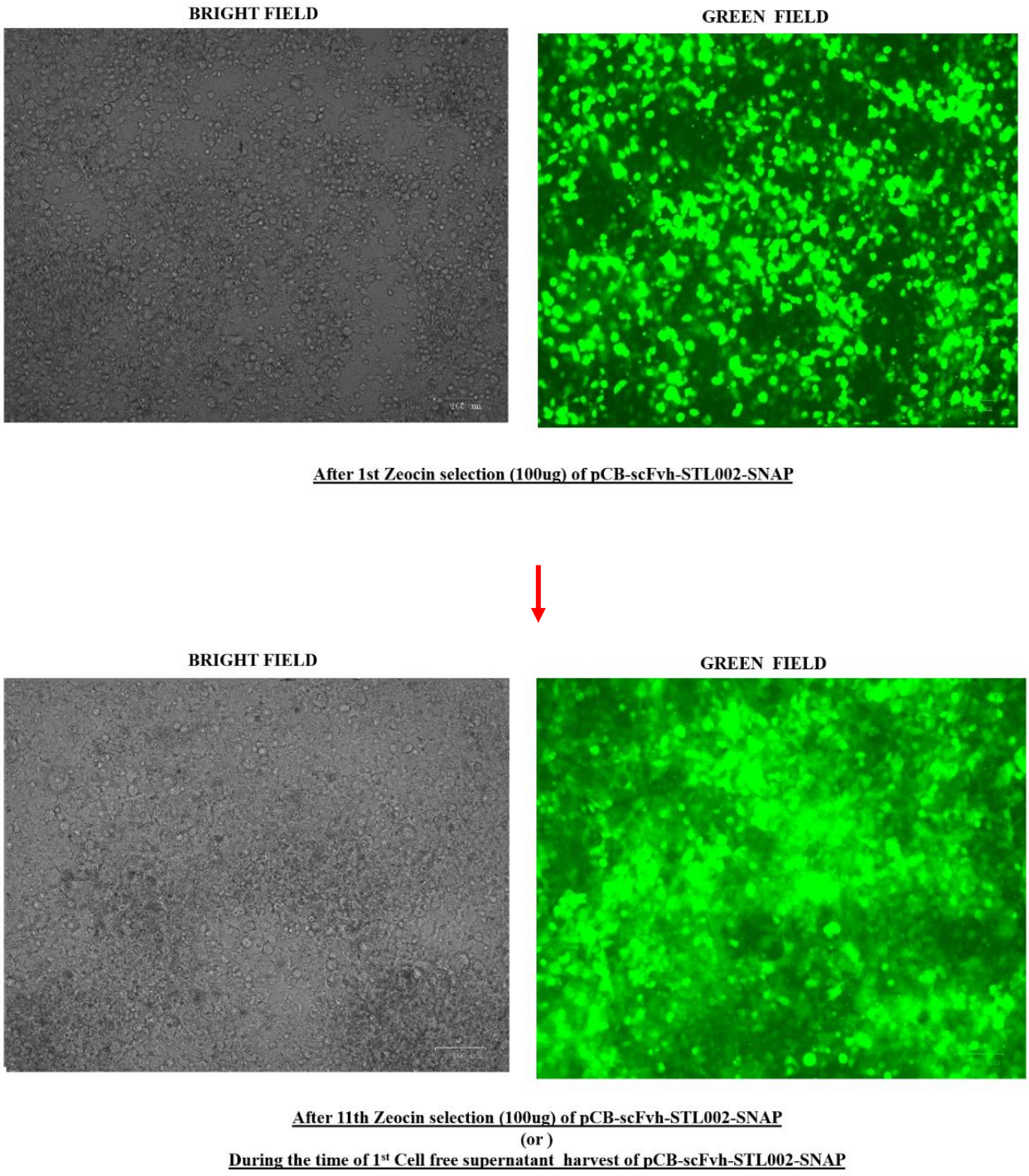


Figure 46: Enriching the population of the pCB-scFvh-STL002-SNAP transfected cell by Zeocin selection

Figure: 41,43,45 top represent the images of un-transfected HEK293T cells, tope left represent the phase contrast and top right represent fluorescent images of un-transfected cells that are visualized and captured under ZOE™ Fluorescent Cell Imager at 100 µm magnification. The absence of green signal in the fluorescent images (top right of 41,43,45) indicates that there is no autofluorescence in un-transfected HEK293T

Figure: 41,43,45 bottom represent the images of plasmid pCB-scFvuBS54-SNAP (41), pCB-scFvID405-SNAP (43), and pCB-scFvh-STL002-SNAP (45) transfected into HEK293T cells, tope left represent the phase contrast and top right represent fluorescent images of transfected cells and are visualized and captured under ZOE™ Fluorescent Cell Imager at 100 µm magnification. The presence of green signal in the fluorescent images (bottom right for 41,43,45) indicates that there is GFP expression in transfected HEK293T.

Figure: 42,44,46 top represents images of transfected HEK293T cells kept under first time Zeocin selection to enrich the population of GFP expressing HEK293T cells.

Figure: 42 ,44, 46 bottom represent the enriched population of GFP expressing HEK293T, that are treated multiple times with Zecoin, this enrichment take place by allowing the GFP expressing cells to grow and proliferate and does not allow the un-transfected cells that do not express Zeocin resistant gene. All images were visualized and captured under ZOE™ Fluorescent Cell Imager at 100 µm magnification.

Table 14: Percentage of live cells and GFP expressing cell before and after Zeocin selection

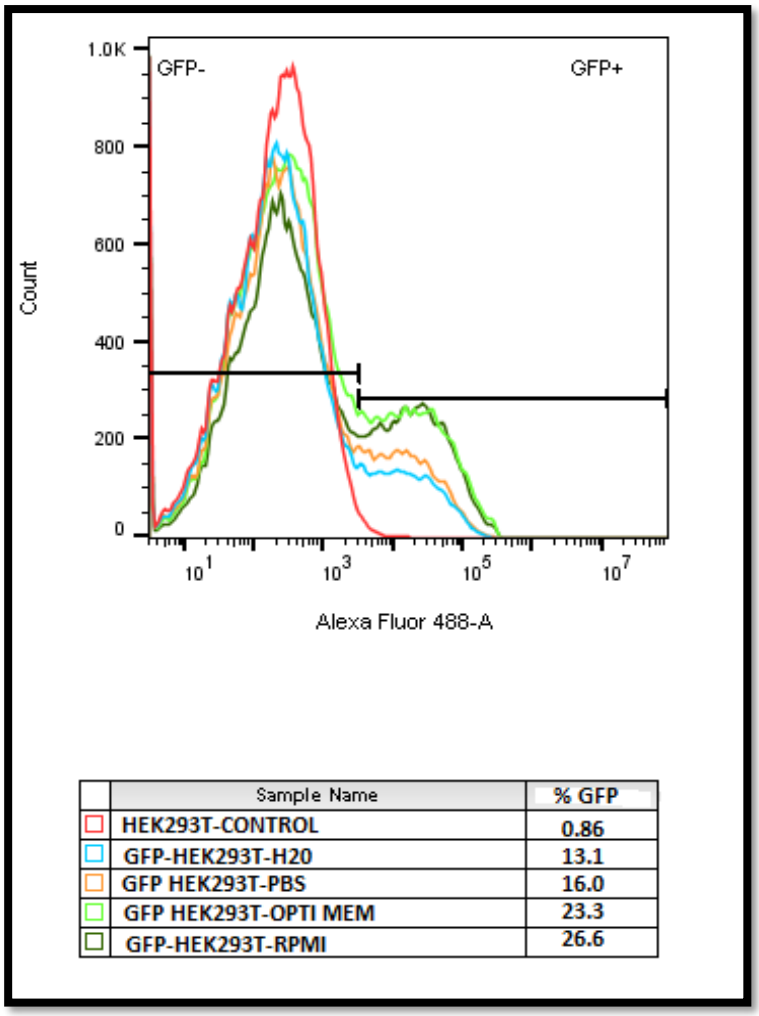
Plasmid	**% of live cells before zeocin selection**	**% of live GFP cells before Zeocin selection**	**% of live cells after multiple Zeocin selections**	**% of live GFP cells after multiple Zeocin selections.**
Un-Transfected HEK293T*	38.6%	0.12%	74.31%	0.0%
pCB-scFvUBS54-SNAP Transfected HEK293T	35.6%	4.63%	54.00%	89.0%
pCB-scFvID405-SNAP Transfected HEK293T	13.2%	4.99%	34.00%	72.0%
pCB-scFvh-STL002-SNAP Transfected HEK293T	23.4%	9.72%	40.00%	77.0%

* No Zeocin selection

Protocol - Manufacturer’s instructions, transfection mixture DNA: Transfection agent in 1:3 ratio

3.5.3 Optimizing Transfection efficiency.

Various factors might impede transfection rates, such as plasmid quantity and quality, cell confluency, antimicrobial agent, media component, serum availability in the culture dish, and the quantity and quality of transfection reagent in transfection mixture. Based on the transfection rate for the construct scFvUBS54-SNAP we chose the Xtreme gene transfection agent and tried to improve transfection efficiency by adjusting key factors, mainly medium formulations and concentration used for transfection DNA complex mixture formation. We found increasing plasmid–transfection agent ratio from 1:3 to 2:3 using RPMI media formulation without serum resulted in higher transfection efficiency, followed by low serum-containing medium Opti-MEM. One possible reason is serum is known to interfere with DNA complex formation, and the presence of para-aminobenzoic acid (PABA) in RPMI might enhance the delivery of nucleic acid to cells as indicated/shown by Reddy et al.[240] In addition, increasing twice the total concentration of nucleic acid per dish enhances efficiency by twofold (see **Section 2.4.3.3** and **Fig. 47**). Further increasing the plasmid and transfection reagent concentration improved the efficiency tenfold; however, the transfected cells died (data not shown here), possibly due to a tenfold increased concentration of transfection agent acting as a cytotoxic agent. We assume removing the transfection reagent after 4–6 h of transfection by replacing them with fresh media could be a possible solution to avert this toxic effect of the transfection reagent.



	Sample Name	% GFP
☐	HEK293T-CONTROL	0.86
☐	GFP-HEK293T-H20	13.1
☐	GFP HEK293T-PBS	16.0
☐	GFP HEK293T-OPTI MEM	23.3
☐	GFP-HEK293T-RPMI	26.6

Figure 47: Transfection efficiency of UBS54-SNAP in various solvent mixtures using XtremeGENE-HP reagent. The highest transfection efficiency of 26% was achieved when the reaction mixture was prepared with RPMI media.

3.5.4. Optimizing Zeocin concentration for selection of GFP clones

The transgenic construct containing both the transgene of interest and a selection marker "antibiotic-resistant gene [Zeocin™-resistance gene Sh ble (Zeo^R)] with a green fluorescent protein (GFP) as a reporter gene is introduced into HEK293T. This hybrid resistance gene combines the functional properties of the Zeocin-resistance marker and GFP reporter enabling to identify populations with higher fluorescence signals that are resistant to Zeocin, which allow for the isolation of better clonal populations and fewer false positives. Sh ble is the gene encoding antibiotic resistance to Zeocin™, This Sh ble gene is driven by the SV40 promoter in tandem with the bacterial EM7 promoter to allow Zeocin selection in mammalian cells. whereas enhanced GFP reporter gene is expressed under the control of T7 promoter as shown in backbone plasmid pCB-Annexin V-SNAP (**Fig. 48**).

To generate a stable line expressing a transgene it is important to eliminate non-transfected or wild-type cells. Not all wild type cell lines are susceptible to antibiotics, some cell lines such as HEK-293T are resistant to certain drugs, so knowing the minimal amount of drug required to kill nontransfected/transduced cells that do not possess a drug resistance marker, will reduce the time and cost to harvest protein and increases the yield.

We treated wild-type HEK293T with various concentrations of Zeocin™ for 72 h (see section 2.4.4.2) and we found 50% inhibition of cell viability is seen at 264 µg concentration (See **Fig 49**). Later based on Zeocin™ viability assay, we started treating scFv-ID405-SNAP transfected cells grown and maintain in T25 flask with 300 µg for 4 weeks and another group of scFv-ID405-SNAP transfected cells in T25 flask are treated with 100 µg of Zeocin for 4 weeks and we compared the percentage of viable cells and GFP cells after 4 weeks and we found 100 µg treated batch showed higher viability and GFP-positive population compared to 300 µg treated batch (see **Table 15**). A possible explanation for low GFP cell densities and low cell viability in high concentration (300 µg) of Zeocin™ could be due to the formation of DNA strand breaks by a high concentration of Zeocin™ despite the expression of the Sh ble gene in the recombinant clone.[217]

Therefore we decided to treat transfected cells (scFvUBS54/ID405/h-STL002-SNAP) initially 2–3 weeks with 300 µg and then switching from 300 µg to 100 µg for long term maintenance, because Hunter et al.[241] reported that untransfected cells (HEK293T) could die when these cells are exposed to 100 µg Zeocin™ for continuously 6 days, this would allow us to reduce the Zeocin usage nearly less than half the concentration for long term maintenance and give similar % of

GFP cells and a little higher percentage of live cells by minimizing the formation of DNA strand breaks in the recombinant clone.

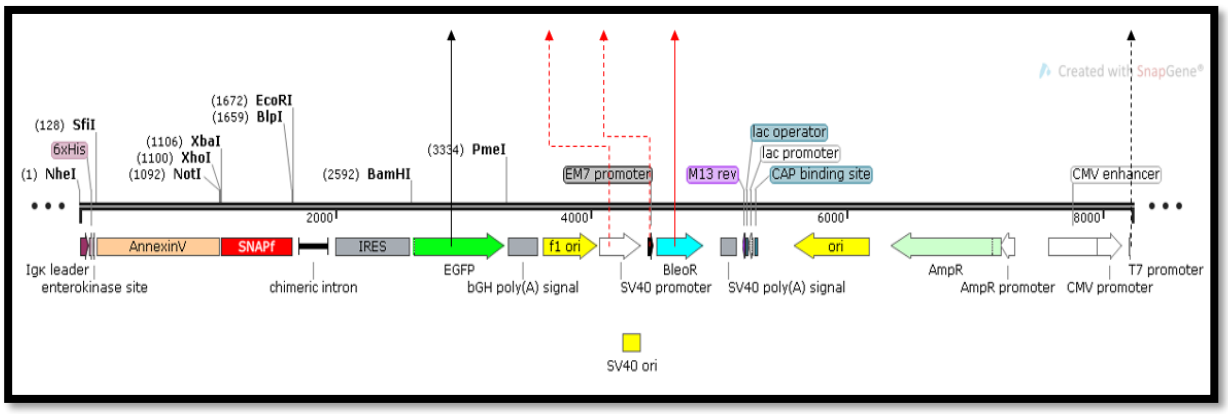
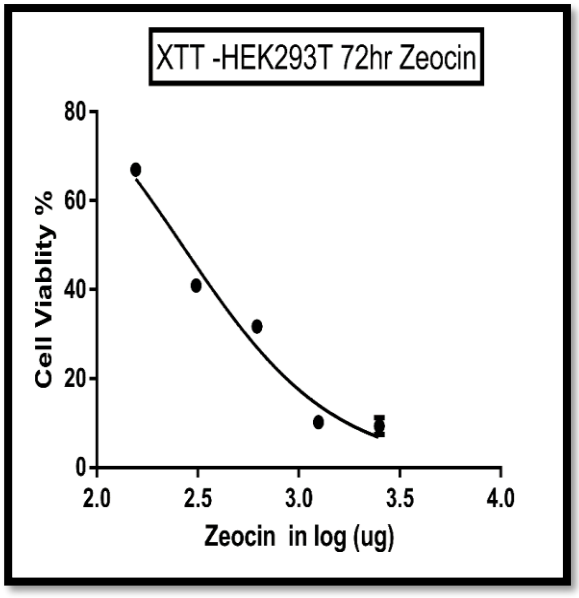


Figure 48: The expression vector gene cassette: CMV promoter -T7 Promoter- gene of interest-chimeric intron-IRES-GFP-signal-SV40 and EM7 promoter -Bleomycin resistance gene. The continuous black arrow represents the enhanced GFP gene, dashed upward facing black arrow represents the promotor for GFP. The upward facing red arrow represents the bleomycin resistance gene, and the dashed upward red arrow represents the corresponding promotor for bleomycin.



IC:50 = 264 ug
R square 0.9694

Figure 49: HEK293T cell culture with 70%-90% confluency is treated with Zeocin at various concentrations. Cell viability was measured 3 days posttreatment. The percentage

of viability is shown for various concentrations along the y-axis. Untreated HEK293T cells are used as positive controls. IC_{50} value achieved at 264 µg/mL.

Table 15: Percentage of live cell and GFP cells of pCB-scFv-ID405-SNAP Transfected cells subject to prolonged Zeocin treatment maintained under two distinct concentrations (100 µg and 300 µg).

	% of live cells	% of live GFP cells
Un-transfected HEK293T*	74.31%	0.00%
pCB-scFvID405-SNAP transfected HEK293T maintained in 100 µg/ml of Zeocin	54.24%	89.76%
pCB-scFvID405-SNAP transfected HEK293T maintained in 300 µg/ml of Zeocin	40.32%	77.42%

* Cells not treated with Zeocin were used as control for flow cytometry.

scFvID405-SNAP transfected cells were treated with two different concentrations of Zeocin for 8 weeks. The percentage of viability estimated using 7-AAD staining and estimation of the number of GFP-positive cells based on reporter gene expression via flow cytometry. Cells treated with 100 µg/ml Zeocin showed a higher population percentage of viable cells and GFP-positive cells compared to the cell culture treated with 300 µg/ml Zeocin.

3.5.5 Determining the harvesting time of CFS via ELISA

After the establishment of stably transfected HEK293T cell lines, it is essential to know whether the transfected HEK293T cell lines are transiently expressing and secreting functional protein of interest, this can be verified by ELISA. It is essential to do ELISA before IMAC (immobilized metal ion affinity chromatography) purification of mass culture supernatant harvested at different time points. because the transient expression might be reduced or lost after prolong culturing, which might impede good yield, so knowing the right time point when to start harvest and stop harvest is key for good yield. Therefore, it is appropriate to determine the availability and quantity, and functionality of recombinant protein at a different batch of harvest with ELISA.

Here the ELISA is sandwich ELISA, where the bottom of the ELISA plate is activated by Pierce™ Maleic anhydride, that can react with primary amines ($-NH_2$) of capture probe (BG-PEG12-NH2) keeping free BG for covalent coupling with SNAP protein (antigen) that are

detected using primary anti-His tag antibody, that is later recognized by secondary HRP conjugated anti-mouse antibody as described in Section 2.4.5.

The 100 µL of cell-free supernatant of the scFvID405-SNAP transfected HEK293T is harvested at two different time points. (a) Immediately after first week of transfection and (b) after second week of transfection added to the respective wells, and purified UBS54-SNAP total protein is used as standard (35 µg/well as starting concentration, 2-fold serial diluted) to create a standard curve (**Fig. 50 and Table 16**).

Table 16: Absorbance value of standard and test determined at 450 nm using HRP anti-His tag-antibody

Standard and Test Sample (µg)	OD VALUE	ELISA (µg)/100 µL
scFvUBS54-SNAP total protein (8.7)	2.405	8.060161
scFvUBS54-SNAP total protein (4.3)	1.742	5.628026
scFvUBS54-SNAP total protein (2.1)	0.845	2.337491
scFvUBS54-SNAP total protein (1.0)	0.381	0.635363
scFvUBS54-SNAP total protein (0.5)	0.191	-0.06163
scFvID405-SNAP cell free supernatant (first week)	0.191667	-0.05918
scFvID405-SNAP cell free supernatant (second week)	0.29333	0.313756
scFvUBS54-SNAP total protein (35)	2.21333	7.357043
scFv-ID405-SNAP total protein (70)	2.356667	7.882858

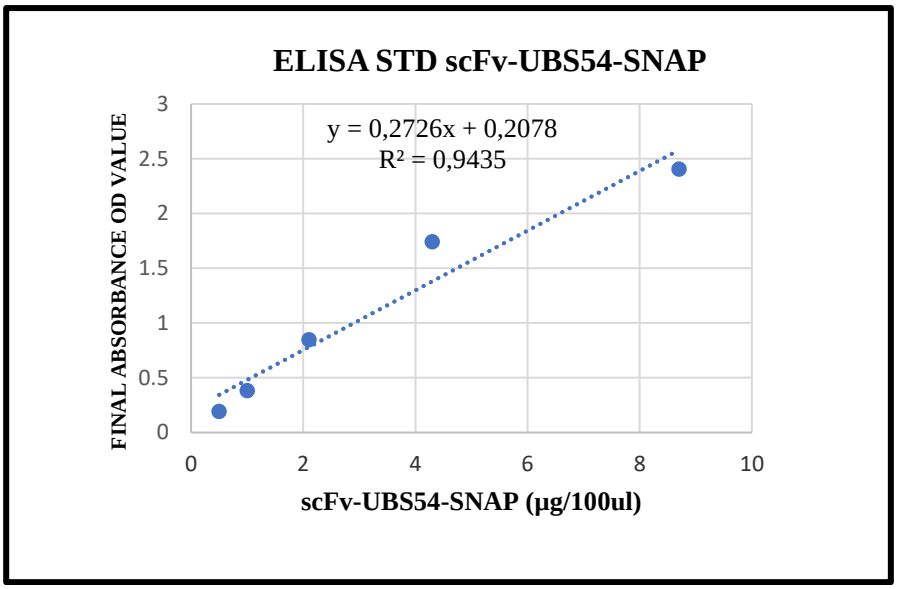


Figure 50: Detection of the scFv-SNAP-tag fusion protein by conventional sandwich ELISA. The linear segment of the curve was used to generate a regression straight line. This standard curve can then be extrapolated, using the equation of the straight line above to find the concentration of the unknown samples.

Standard curves for scFvUBS54-SNAP obtained after an enzymatic reaction was visualized through the addition of TMB substrate catalyzed by HRP produces a blue-colored product that changes yellow after the addition of an acidic stop solution. The Intensity of yellow coloration read by absorbance at 450 nm is quantitatively proportional to the amount of 2-fold serially diluted standard sample "scFv-UBS54-SNAP" captured in the corresponding well. The concentration of UBS54-SNAP is proportional to the concentration of colored by products of the HRP-catalyzed reaction. In this sandwich ELISA, SNAP-tag fusion protein was successfully detected at concentrations as low as 1 µg/100 µl standard. 100 µl of test samples (cell-free supernatant) harvested from two different cell cultures (HEK293T cultures expressing UBS54-SNAP and ID405-SNAP) were assayed by ELISA. The concentration of scFvID405-SNAP recombinant protein is detected in the corresponding cultured cell-free supernatant, collected at two different time points (early and late) after transfection and Zeocin selection. Interestingly the SNAP fusion protein in the supernatant of ID405-SNAP transfected culture is not seen immediately after first week post transfection, but seen at later time points. This indicating the production and secretion of SNAP fusion protein at detectable levels (0.3 µg/100 µL) would

require 2 weeks post transfection. In addition, we found that the plate got saturated at 8 µg of full-length SNAP fusion protein.

3.6. PURIFICATION AND QUANTIFICATION OF ANTIBODY BASED FUSION PROTEIN

3.6.1. Purification of scFv-SNAP recombinant fusion proteins by IMAC

Fusion proteins, scFvUBS54-SNAP (5 mg/mL in 500 mL), scFvID405-SNAP (10 mg/mL in 500 mL) and scFvh-STL002-SNAP (3 mg/mL in 350 mL) were enriched from cell culture supernatant by affinity chromatography (using the N-terminal His_6 tag), using immobilized Metal Ion (Nickel ion: Ni^{2+}) Affinity (IMAC). During IMAC, recombinant proteins were specifically bound to the column (Ni^{2+}) due to their N-terminal His_6 tag affinity for Ni^{2+}. Thus, elution of these recombinant fusion proteins (scFvUBS54-SNAP, scFvID405-SNAP, and scFvh-STL002-SNAP) was performed through competitive binding between His_6 tag and imidazole for Ni^{2+} on the column. The collection of the eluted fractions was monitored at wavelength 280 nm (see **Fig. 51**, **Section 2.5.1**).

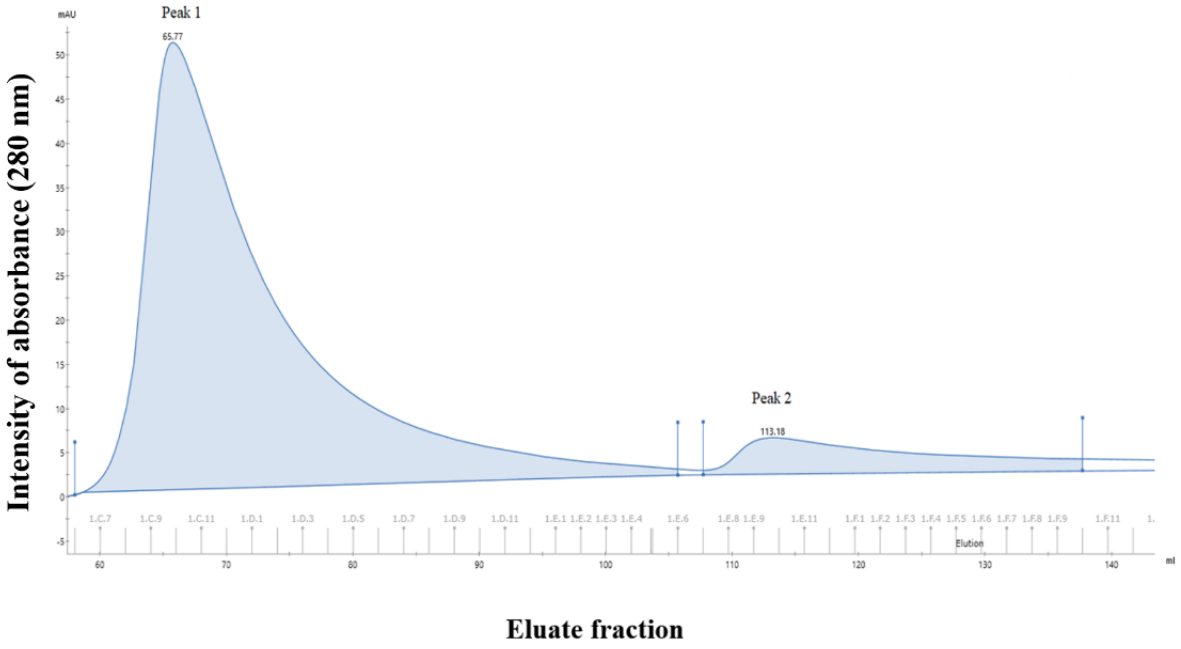


Figure 51: Chromatogram illustration of the absorbance profile of scFvUBS54-SNAP, at 280 nm, during the elution step of the IMAC purification performed. The elution profile is the same for the other two constructs scFvID405-SNAP, scFvh-STL002-SNAP (Image not shown). Cell culture supernatant was purified from HEK293T cells transiently expressing scFvUBS54-SNAP/scFvID405-SNAP/ scFvh-STL002-SNAP. Both peaks contained purified protein fractions as illustrated in the SDS-PAGE and immunoblot results provided in the subsequent sections. Peak 1 was eluted in 0-75 mM and peak 2 in 250 mM imidazole.

3.6.2. Quantification of scFv-SNAP fusion proteins by UV–visible spectrophotometer

Quantitate the concentration of total protein by Nanodrop. The total protein from eluted fractions corresponds to each peak is estimated by Nanodrop before Amicon concentration and after amicon concentration (see **Table 17**). The absolute amount of total protein concentrations (scFvUBS54-SNAP, scFvID405-SNAP, and scFvh-STL002-SNAP) were 5.8, 4.1, and 3.2 mg in about 600 µL determined based on UV–visible spectrophotometer.

Table 17: Quantitation of the total protein concentration by UV spectrophotometry

scFv-SNAP	**BEFORE AMICON CONCENTRATION**	**AFTER AMICON CONCENTRATION**		
	Conc. (mg/mL)	Conc. (mg/mL)	A280	A260/280
scFv-UBS54-SNAP-p1	0.1	5.8	3.8776	0.59
scFv-UBS54-SNAP-p2	0.03	2.1	1.4111	0.63
scFv-ID405-SNAP-p1	0.18	4.1	2.7971	0.58
scFv-ID405-SNAP-p2	0.06	2.0	1.3658	0.58
scFv-h-STL002-SNAP-p1	0.13	3.2	2.18	0.61
scFv-h-STL002-SNAP-p2	0.04	3.0	2.03	0.61

3.6.3. Quantification of of scFv-SNAP fusion proteins by BCA assay.

Total cellular proteins after IMAC purification were concentrated using Amicon® Ultra-15* Centrifugal filter, following centrifugation at 12,000 × g at 4°C for 45 min, the total protein concentrations of the supernatants were determined using Pierce™ BCA Protein Assay Kit (ThermoFisher) standards at 40.0, 20.0, 10.0, 5.0, 2.50, 1.25, 0.62, 0.3, 0.1 µg/well in PBS was

used in 96-well plate (see methods 2.5.2.2) and another 20 µl of the PBS solution (BSA solution concentration of 0 mg/ml) was used as a control, 20 µl of the test sample in triplicate is added to the remaining well. 200 µl of working reagent is added to each well and mixed with a multichannel pipette. Incubate the plate at room temperature for at least 30 min and not more than 1 h. After incubating the plate for 30 min at 37°C, the optical density values of the BSA standards were measured at 562 nm using a microplate reader. A standard curve was generated, and the protein concentrations were calculated based on the obtained equation.

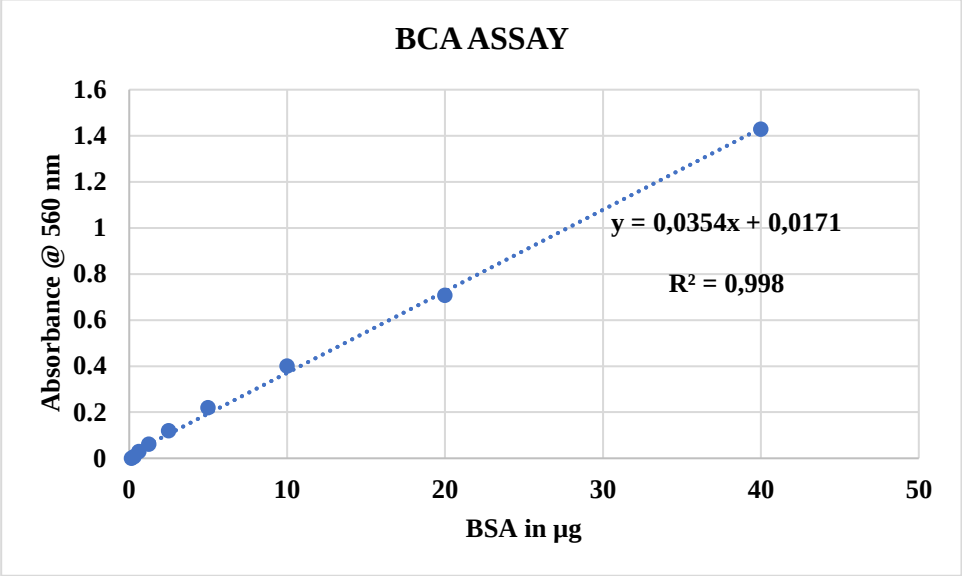


Figure 52: A sample micro-BCA assay standard curve. BSA standard was added in duplicate from, 0.1 to 40 µg, the assay responds linearly to the concentration of standard. The data is fit with the equation y = 0,0354x + 0,0171 and has an R^2 value of 0.998. The data table used to generate the figure and depiction of a typical micro-BCA assay is shown in **Table 18**. Test samples were used in their diluted and undiluted forms. The lowest detection limit was 0.3 µg of total protein/well.

Table 18: Absorbance values for various concentrations of standard and test samples at 562 nm

STANDARD AND CONC scFv-SNAP	OD 1	OD 2	AVERAGE-OD	AVERAGE-BLANK	y = 0.0354x + 0.0171
STD 1 BSA (40.00 µg)	1.439	1.564	1.5015	1.4295	39.89
STD 2 BSA (20.00 µg)	0.771	0.789	0.78	0.708	19.51
STD 3 BSA (10.00 µg)	0.477	0.469	0.473	0.401	10.9
STD 4 BSA (05.00 µg)	0.294	0.289	0.2915	0.2195	5.71
STD 5 BSA (02.50 µg)	0.192	0.193	0.1925	0.1205	2.92
STD 6 BSA (01.25 µg)	0.133	0.135	0.134	0.062	1.26
STD 7 BSA (00.62 µg)	0.104	0.1	0.102	0.03	0.36
STD 8 BSA (00.31 µg)	0.08	0.08	0.08	0.008	-0.2
STD 9 BSA (00.15 µg)	0.073	0.071	0.072	0	-
PBS- BLANK	0.072	0.072	0.072	0	-
scFv-UBS54-SNAP (190 µg) 20µl	1.574	1.544	1.559	1.487	41.52
scFv-ID405-SNAP (140 µg) 20µl	1.157	1.183	1.17	1.098	30.53
scFv-h-STL002-SNAP CD138 (52 µg) 20µl	0.69	0.683	0.6865	0.6145	17.59
scFv-UBS54-SNAP Diluted (23.7 µg)	0.54	0.558	0.549	0.477	12.96
scFv-ID405-SNAP Diluted (17.5 µg)	0.432	0.431	0.4315	0.3595	9.65
scFv-h-STL002-SNAP CD138 (6.5 µg)	0.233	0.232	0.2325	0.1605	4.03

3.6.4. Quantification of scFv-SNAP fusion proteins by Bradford assay

Total cellular proteins after IMAC purification are concentrated using Amicon® Ultra-15* Centrifugal filter, following centrifugation at 12,000 × g at 4°C for 45 min, the total protein concentrations of the supernatants are determined using Bradford Protein Assay (Bio-Rad) A set of BSA (ThermoFisher) standards at 10.0, 5.0, 2.5, 1.25, and 0.6 µg/well in PBS was used in 96-well plate and another 10 ul of the PBS solution (BSA solution concentration of 0 mg/ml) was used as a control, 10 µl of the test sample in triplicate is added to the remaining well (**Section**

2.5.2.3). 200 µl of diluted dye reagent is added to each well and mixed with a multichannel pipette. Incubate the plate at room temperature for at least 10 min and not more than 1 h. After incubating the plate for 10 min at 37°C, the optical density values of the BSA standards were measured at 595 nm using a microplate reader. A standard curve was generated, and the protein concentrations were calculated based on the obtained equation.

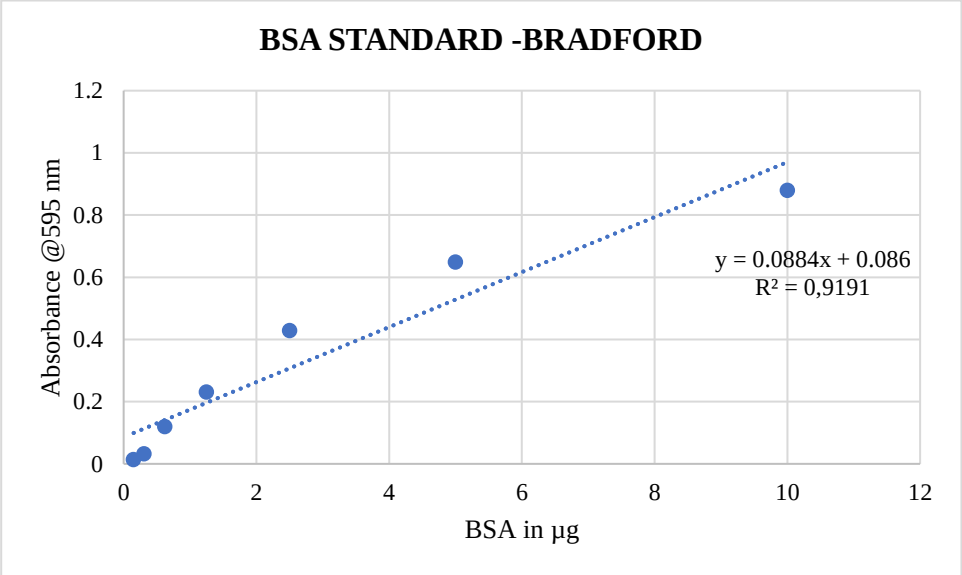


Figure 53: A sample micro-Bradford assay standard curve. BSA standards were added in duplicate from 0.1 to 10 µg. The assays responded linearly to the concentration of standard. The data is fit with the equation y = 0.0884x + 0.086 and had an R^2 value of 0.9191. The data table used to generate the figure and depiction of a typical micro-Bradford assay is shown in **Table 19**. Test samples were used in both diluted and undiluted forms. The lowest detection limit is 0.6 µg or total protein per well.

Table 19: Absorbance values for various concentrations of standard and test samples at 595 nm

STANDARD AND CONC scFv-SNAP	OD 1	OD 2	AVERAGE	AVERAGE-BLANK	y=0.0884x + 0.086
STD 3 BSA (10.00 µg)	1.143	1.15	1.1465	0.88	8.98
STD 4 BSA (05.00 µg)	0.917	0.915	0.916	0.6495	6.36
STD 5 BSA (02.50 µg)	0.695	0.695	0.695	0.4285	3.86
STD 6 BSA (01.25 µg)	0.499	0.496	0.4975	0.231	1.64
STD 7 BSA (00.62 µg)	0.388	0.385	0.3865	0.12	0.38
STD 8 BSA (00.31 µg)	0.295	0.303	0.299	0.0325	-
STD 9 BSA (00.15 µg)	0.28	0.281	0.2805	0.014	-
PBS BLANK	0.268	0.265	0.2665	0	
EPCAM (190 µg) 20 µl	1.567	1.405	1.486	1.2195	12.8
EMA (140 µg) 20 µl	1.327	1.315	1.321	1.0545	10.9
CD138 (52 µg) 20 µl	1.351	1.321	1.336	1.0695	11.1
EPCAM diluted (23.7 µg)	1.035	1.086	1.0605	0.794	8
EMA Diluted (17.5 µg)	1.155	1.137	1.146	0.8795	9
CD138 (6.5 µg)	0.603	0.668	0.6355	0.369	3.2

3.6.5. Estimating the integrity or purity of scFv-SNAP fusion protein by SDS and WB.

The fractions which corresponded to each peak on the chromatogram (**Fig. 51**) were concentrated and desalted to remove any non-specific protein contaminants by Amicon® Ultra -15* centrifugal filter subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) to confirm the presence of the recombinant fusion proteins in eluted fractions based on molecular size (see methods 2.5.3). During this SDS-PAGE experiment, 20µl of each fusion protein collected fraction was loaded onto a 7.5% acrylamide gel with 5µl of 4x loading dye. This experiment enabled the electrophoretic separation of the scFvUBS54-SNAP, scFvID405-SNAP, and scFvh-STL002-SNAP fusion proteins based on their respective sizes 50.50 kDa, 51.73kDa and 50.90kDa. Thereafter, visualization of the fusion proteins on the gel was performed post acquaStain dye incubation which stained the fusion proteins (**Fig. 54**)

To further confirm the successful purification of the fusion proteins, a western blot was performed by transferring the fusion proteins from another SDS-PAGE gel to a nitrocellulose membrane. Once on the nitrocellulose membrane, the fusion proteins were sequentially

incubated with primary rabbit monoclonal anti-his-tag to detect the C-terminal His_6 tag of the fusion proteins and secondary antibody anti-rabbit horseradish peroxidase-conjugate which reacted with the chemiluminescent substrate to allow visualization of the fusion proteins as seen in **Fig 55**. These fusion proteins were directly used for binding studies. **Note:** The protein yield can be increased 10 fold by switching to suspension culture and by using Gal6-pGEN2 vector [242] backbone**.**

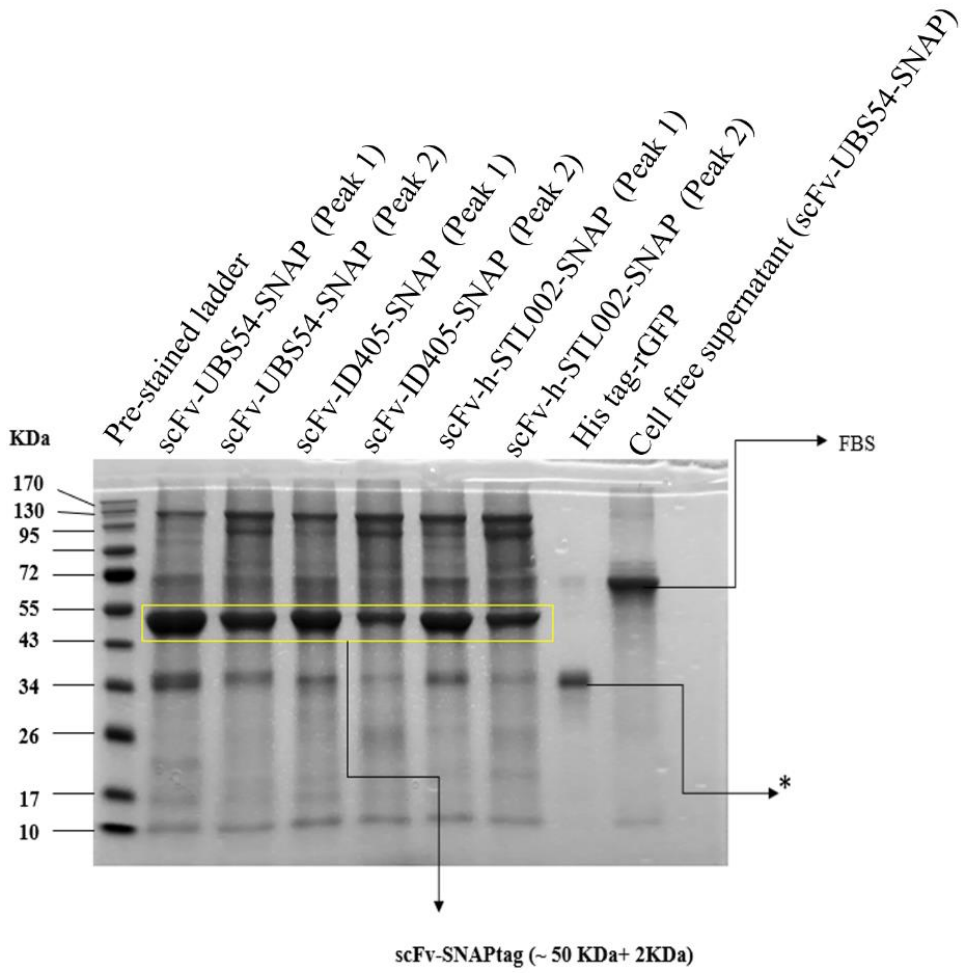


Figure 54: scFv-SNAP fusion protein Identity, purity, and integrity: SDS-PAGE of scFvUBS54-SNAP, scFvID405-SNAP, and scFvh-STL002-SNAP enriched fusion protein fractions.

Note: * The recombinant GFP consisting of 253 amino acids and has a calculated molecular mass of 28.7 kDa. It migrates as an approximately 34 kDa band in SDS-PAGE under reducing conditions as predicted.

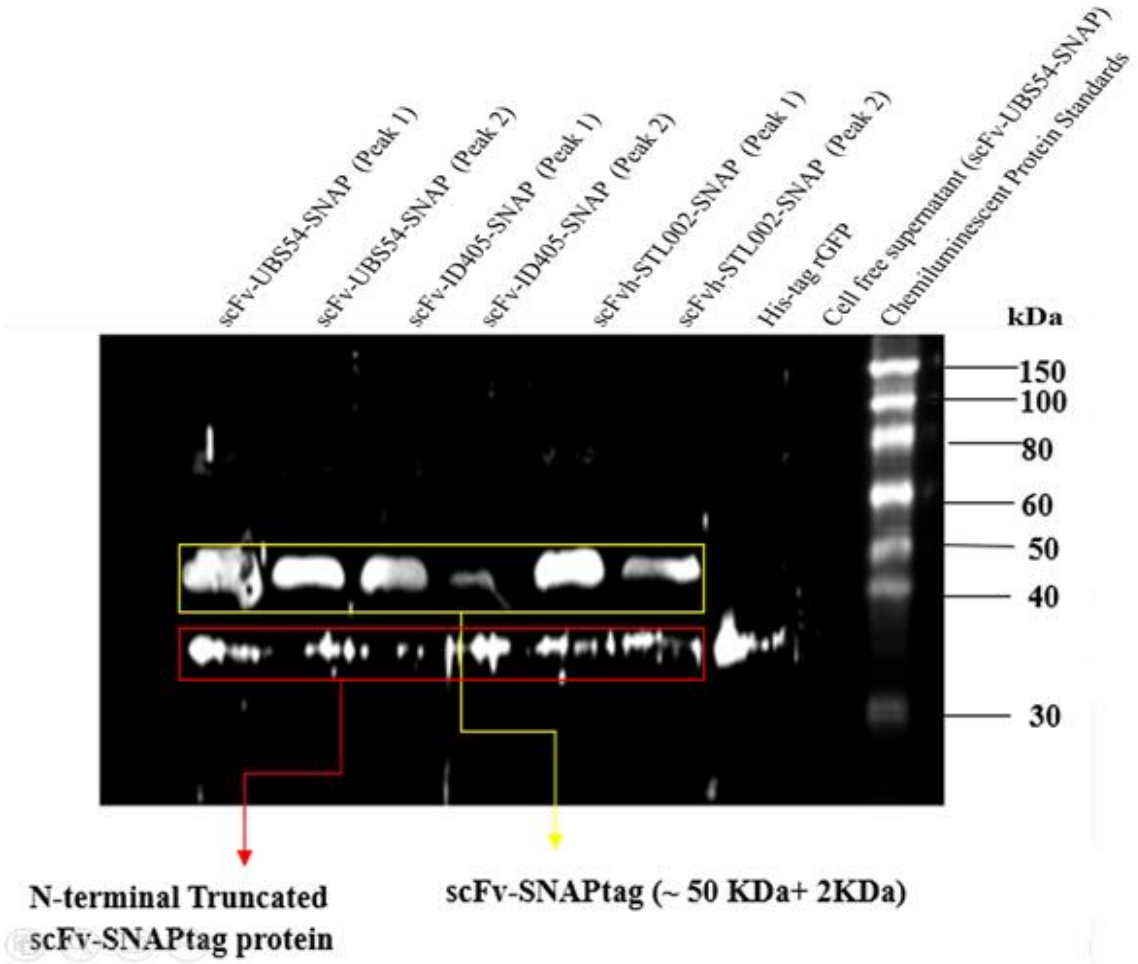
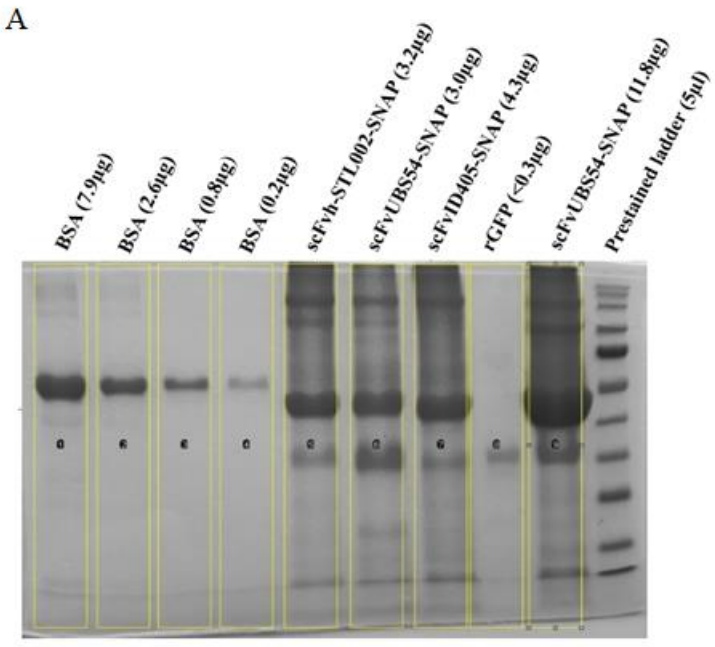


Figure 55: scFv-SNAP fusion protein identity, purity, and integrity.

Western blot images of scFv-UBS54-SNAP, scFvID405-SNAP, and scFvh-STL002-SNAP fusion protein. 5 µl of protein ladder (M) were loaded beside 20 µg of eluted protein fractions representing scFvUBS54-SNAP, scFvID405-SNAP, and scFvh-STL002-SNAP fusion proteins. The protein fraction was run on SDS-PAGE gel and transferred to a nitrocellulose membrane and subsequently probed with rabbit monoclonal anti-his-tag and anti-rabbit horseradish peroxidase-conjugate antibodies as the primary and secondary antibodies used for the method. SuperSignal West Femto Chemiluminescent Substrate was used for detection at multiple time points (0, 5, 10, 15 min) and the figure corresponds to 10 min exposure of the membrane when imaging Note: P1 & P2 represent a pooled fraction of 1st peak and 2nd peak in shown in IMAC chromatogram. The truncated protein at 35 kDa size was detected which could be due to the Kozak sequence or by degradation of the fusion protein due to reason yet unknown. This could be eliminated by implementing additional purification steps.

3.6.6. Densitometric analyses to quantify full-length protein by SDS-PAGE

The protein purity is generally checked using SDS-PAGE, where densitometry could be used to quantify the protein bands of interest. The quantification of an unknown protein could be made using the correlations established using the protein standards ultra-pure B.S.A, or the molecular weight markers: 100, 50, and 20 kDa. The choice depends on the type of molecule to be quantified and size, in this case, the single-chain variable fragment of 3 different constructs with a molecular weight of approx. 50 kDa is analysed using an imaging system or separate software (Image J), see methods 2.5.4. Software algorithms can compare the measured density to a background area typically a protein standard B.S.A (66 kDa) loaded in a linear dynamic range that extends from high to low signal intensities. B.S.A supports accurate quantitation across an equally broad range of protein concentrations (Fig 56) typically adjacent to the target band, to produce a relative measurement.



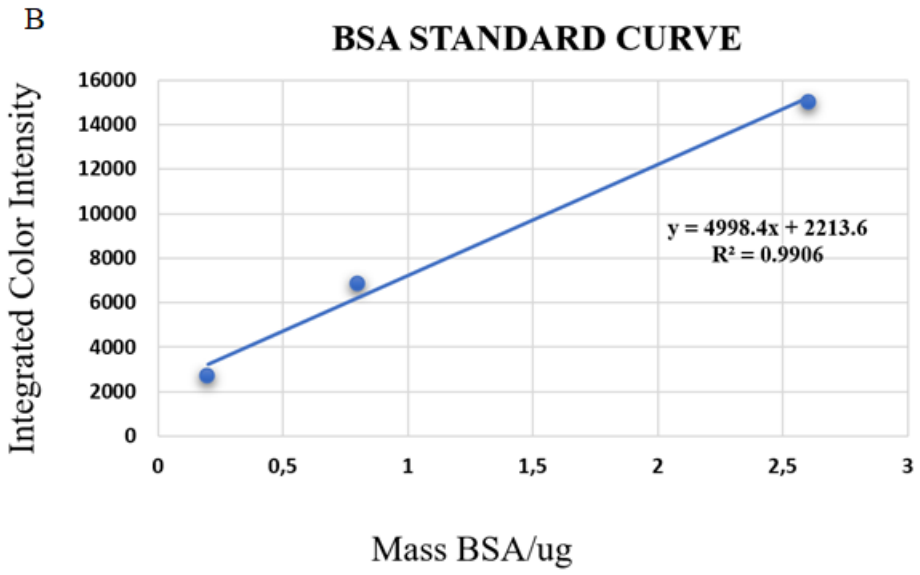


Figure 56: Quantitation of protein band of interest using the linear dynamic range of BSA standard.

The SDS-PAGE image recorded was transformed to 8-bit images. Lanes from left to right represent the highest to the lowest concentration of BSA, serially diluted from the stock solution and loaded in the first 4 lanes in the following order of quantity: 7.9, 2.6, 0.8, and 0.2 µg. Test samples were loaded in next 3 lane with total protein concentration estimated by nanodrop (lane 5; h-STL002-SNAP (3.2 µg), lane 6; UBS54-SNAP (3.0 µg), lane 7; ID405-SNAP (4.3 µg), lane 8; rGFP, lane 9; UBS54-SNAP (11.8 µg). The molecular weight marker was run in the last lane of the gel (from right) (A). The bands obtained were analysed by ImageJ and a correlation analysis was carried out for the peak that corresponded to the band of interest in the serially diluted standards and test samples to define the area of the curve. This aided in the quantification of the protein bands of interest (B). Average relative lane density of the total proteins (Standard BSA) loaded in a linear dynamic range that extended from high to low signal intensities and provided a trend line with **R2 values >0.95**. This was used to obtain the concentration of the bands of interest in the test samples using the displayed equation.

Table 20: Determining the concentration of full-length protein from total protein

lane	Standard and Test Sample	Image. J	Total Protein (µg/well) (Nanodrop)	Peak area (Image J)	Y= peak area (Full length protein)
1	BSA	Uncalibrated	7.9		
2	BSA	Calibrated	2.6	15040.368	2.566174776
3	BSA	Calibrated	0.8	6888.083	0.935195863
4	BSA	Calibrated	0.2	2706.426	0.098596751
5	h-STL002-SNAP		3.2	12382.054	2.034341789
6	UBS54-SNAP		3.0	9542.347	1.46621859
7	ID405-SNAP		4.3	10704.125	1.698648568
8	rGFP		NA	2212.134	-0.000293294
9	UBS54-SNAP		11.8	11818.388	1.921572503
10	Pre-stained ladder	-	-	-	-

Table 20: The quantities of target proteins loaded in lanes 5–7 and 9 were determined with reasonable accuracy (values obtained in the last column). If the highest or lowest quantity did not fit in the liner part of the standard curve generated, i.e., the quantity of protein loaded on lane-8 was less than the lowest dilution of standard, the obtained values were not considered as the quantities could not be reliably determined. The BSA standard with the highest concentration that was loaded on lane-1 did not fit in the trendline of the standard curve and was thus not considered for quantitation.

3.7 CONJUGATION OF ANTIBODY FUSION PROTEIN WITH BG-MODIFIED FLUOROPHORE

3.7.1 Conjugation of scFv-SNAP with BG-Alexa 488

Initial investigation of the self-labeling activity of 5 µM of scFv-SNAP such as scFv-UBS54-SNAP, scFvID405-SNAP, and scFvh-STL002-SNAP with 10 µM of BG-Alexa 488 results in conjugation of labelled fluorescent proteins (scFvUBS54/scFvID405/scFvh-STL002-SNAP-Alexa 488) that were observed to migrate at a size corresponding to their protein molecular weight as shown in Fig. 57, left panel. In addition, we noticed that there are unconjugated

labelling reagent (BG-Alexa 488) present at the bottom of the gel, with very high intensity of green fluorescence as shown in Fig 57, right panel. This indicates that the concentration of labeling reagent should be reduced or optimized further according to the need. Therefore, we used various concentrations of unconjugated BG-Alexa 488 (1, 2.5, 5 µM) and found 2.5 µM should be more than enough to conjugate 5 µM of total protein scFv-SNAP (Fig. 58). The coupling efficiency was determined photometrically, based on the theoretical extinction coefficients of the recombinant fusion proteins and the extinction coefficients of the fluorescent dyes. This revealed a labeling efficiency of 80%-90% after 1 h of incubation at room temperature.

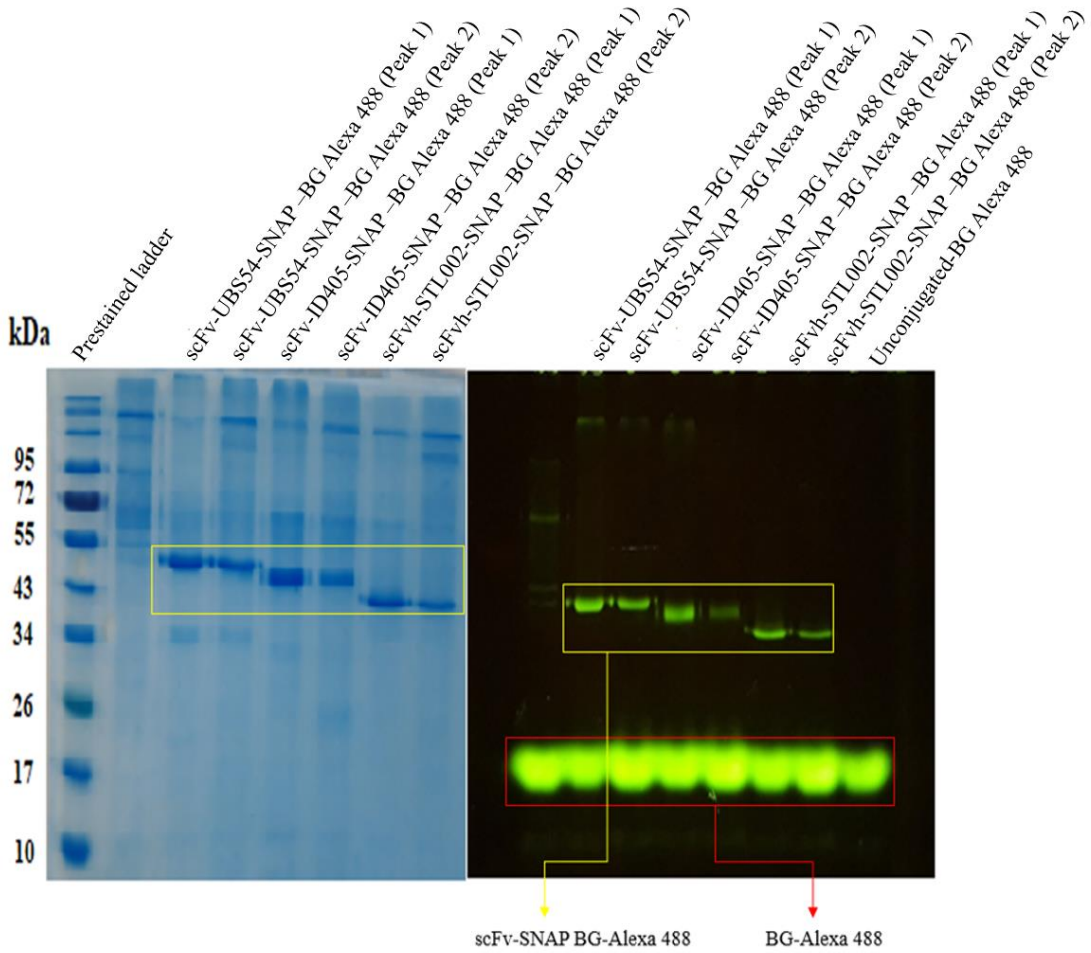


Figure 57: Fusion proteins scFv-UBS54-SNAP, scFvID405-SNAP, and scFvh-STL002-SNAP were labeled with BG-Alexa 488. SDS-PAGE and fluorescence visualization of BG-

Alexa 488 labeled scFvUBS54/scFvID405/scFvh-STL002-SNAP-tag. 5 µM of the fusion protein was incubated with 10 µM of BG-Alexa 488 Green and 50 mM DTT for 30 min at 37^0 C. The fusion protein was loaded without the labeling reagent BG-Alexa 488 was loaded as a negative control. (a) The visualization was performed after staining of the SDS gel and (b) exposure to blue light excitation between 450 and 495 nm. The unconjugated labeling reagent is seen at the bottom of the gel and the labeled fluorescent proteins (scFvUBS54/scFvID405/scFvh-STL002-SNAP-Alexa 488) were observed to migrate at a size corresponding to their protein molecular weight.

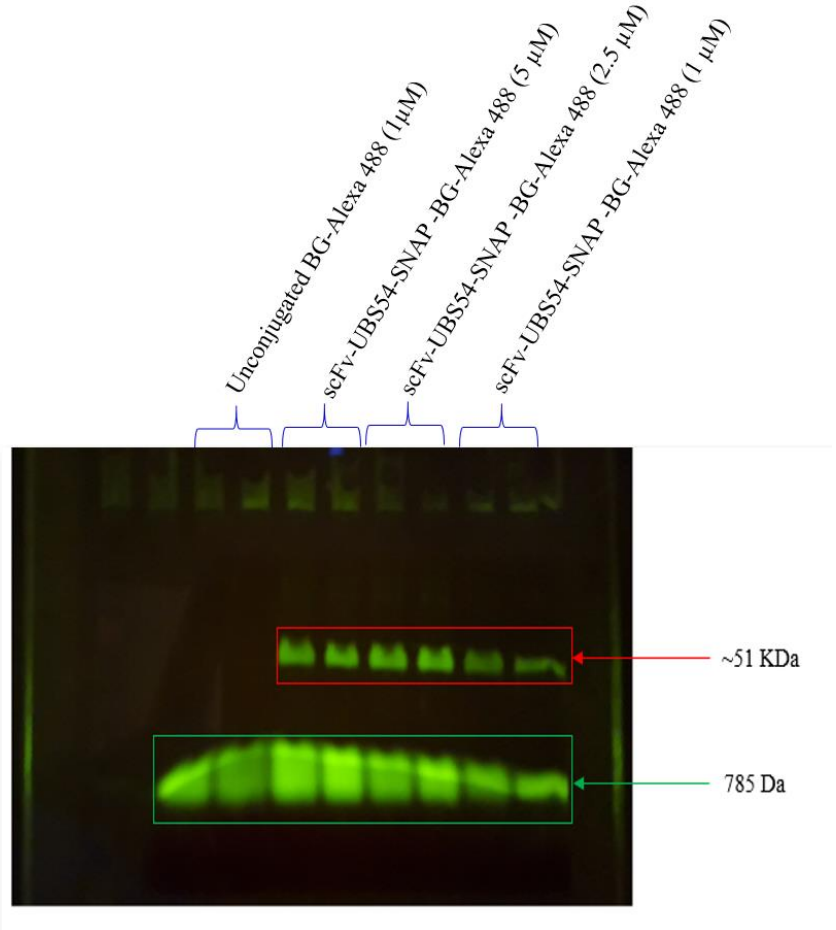


Figure 58: Optimizing the concentration of BG-Alexa 488 to be used for conjugation with scFv-SNAP. The high intensity of unconjugated BG-Alexa 488 shows that the recommended 10 µM of BG-Alexa 488 is too excessive to label 5 µM of total protein (scFv-SNAPtag). We hence tried to estimate the minimum concentration of BG-Alexa 488 required for successful labeling scFv-SNAPtag protein. We found 2.5 µM of BG-Alexa 488 is enough to label 5 µM of

scFv-SNAP. Interestingly we tried to increase the concentration of total protein from 5 µM of scFv-SNAP to 15 µM and keeping the concentration of BG-Alexa 488 as 2.5 µM. This still showed efficient labeling through conjugation (data not shown). The bands highlighted in the red box and indicated the BG-Alexa conjugated scFv-SNAP. Bands highlighted with the green box represent unconjugated BG-Alexa 488.

3.8. BINDING ACTIVITY OF FLUOROPHORE-CONJUGATED ANTIBODY BASED FUSION PROTEIN TO TARGET ANTIGEN OF INTEREST

3.8.1. Confocal Imaging

After successful conjugation of the fusion proteins (scFvUBS54-SNAP, ID405-SNAP, and scFvh-STL002-SNAP) to Alexa 488, binding to their cognate antigens EpCAM, EMA, and CD138 expressing cells were assessed using confocal microscopy. The binding efficiency of the fusion proteins (scFvUBS54-SNAP, scFvID405-SNAP, scFvh-STL002-SNAP) labeled with BG- modified Alexa 488 was determined by a Zeiss LSM880 Airyscan with Fast Airyscan module confocal microscope. Cells expressing EpCAM, EMA, and CD138 were prepared by culturing them at 37°C until it reaches 70-80% confluency in DMEM complete media. Binding efficiency was determined by incubating cells with 5µM-10µM labeled fusion proteins for 30-40 min on ice (see **Section 2.6.2.1**). Confocal microscopy results confirmed the homogeneous and strong membrane staining on $EpCAM^+$ cells MDA-MB-468 using scFvUBS54-SNAP but not in $EpCAM^-$ cells HS578T (**Figs. 59 and 60**). Similarly, EMA^+ cells MCF-7 cell surface is labeled with scFvID405-SNAP, but not in EMA^- cells HEK293T (**Figs. 61 and 62**). Finally, we tested the binding activity of the scFvh-STL002-SNAP protein in HEK293T, The scFvh-STL002-SNAP stains positive for CD138 in HEK293T (**Fig. 63**). In addition, we showed the binding of (αCSPG4) mAb9.2.27 (scFv)-SNAPtag labeled with Alexa 488 (green) in CSPG4-positive cell line HS578T (**Fig. 64**). Unfortunately, due to unforeseen circumstances and other logistic issue, we were unable to procure a BCC cell line ("TE 354.T ") from ATCC.

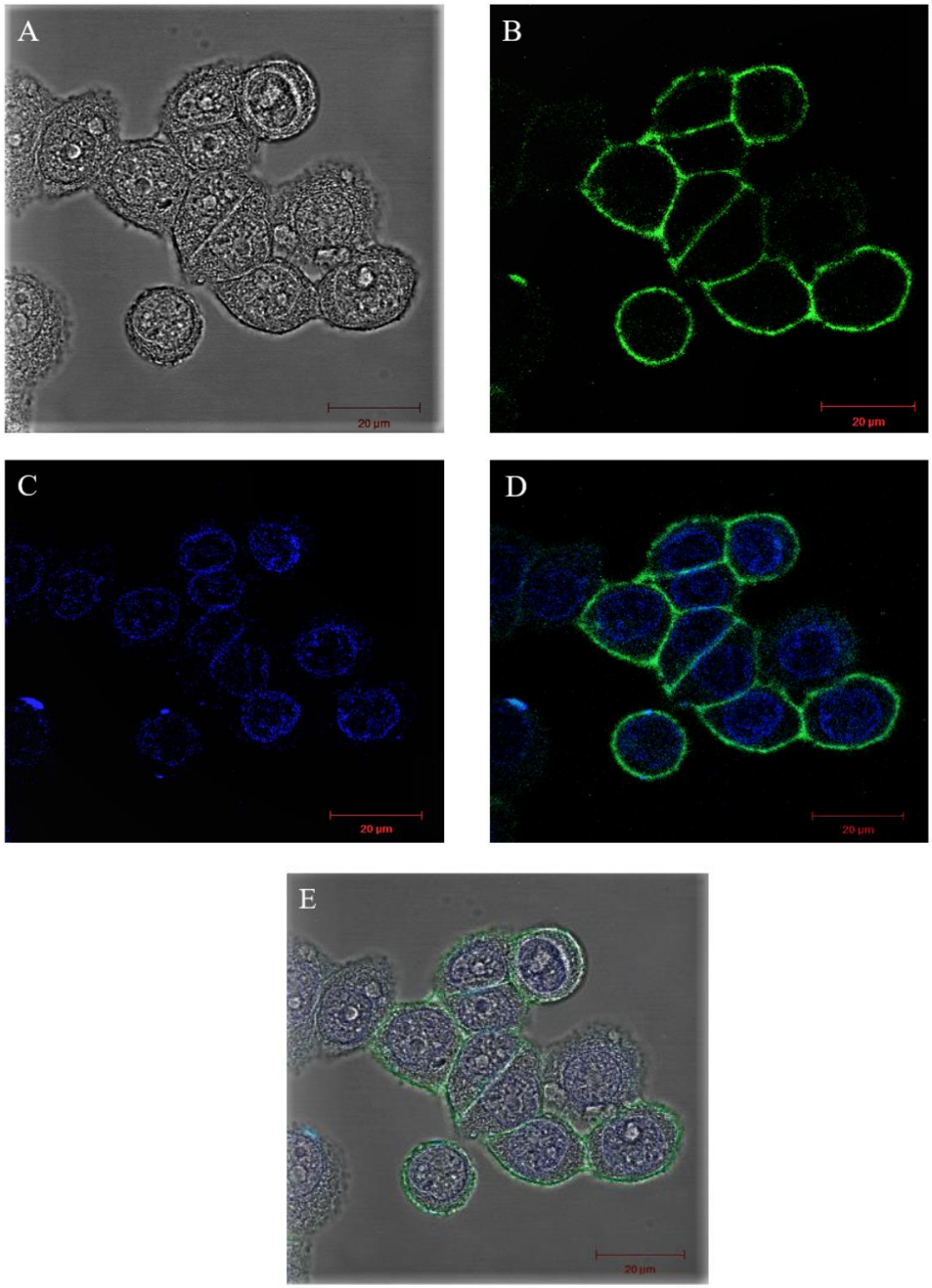


Figure 59: Staining of EpCAM-positive MDA-MB468 cells with UBS54(scFv)-SNAPtag labeled with Alexa 488 (green) for 1 h at 4°C. Nuclear staining was achieved with Hoechst (blue), which is **excited** by ultraviolet light and emits blue light at 460 to 490 nm. The top left image (A) represents a bright field, top right (B) is cell surface staining (greenfield) with Alexa 488, excited by laser light at 499 nm. Middle left (C) is nucleus staining, middle right (D) is a superimposition of cell surface staining and nucleus staining images, and bottom image (E) is a superimposition of bright field, greenfield, and nucleus staining images.

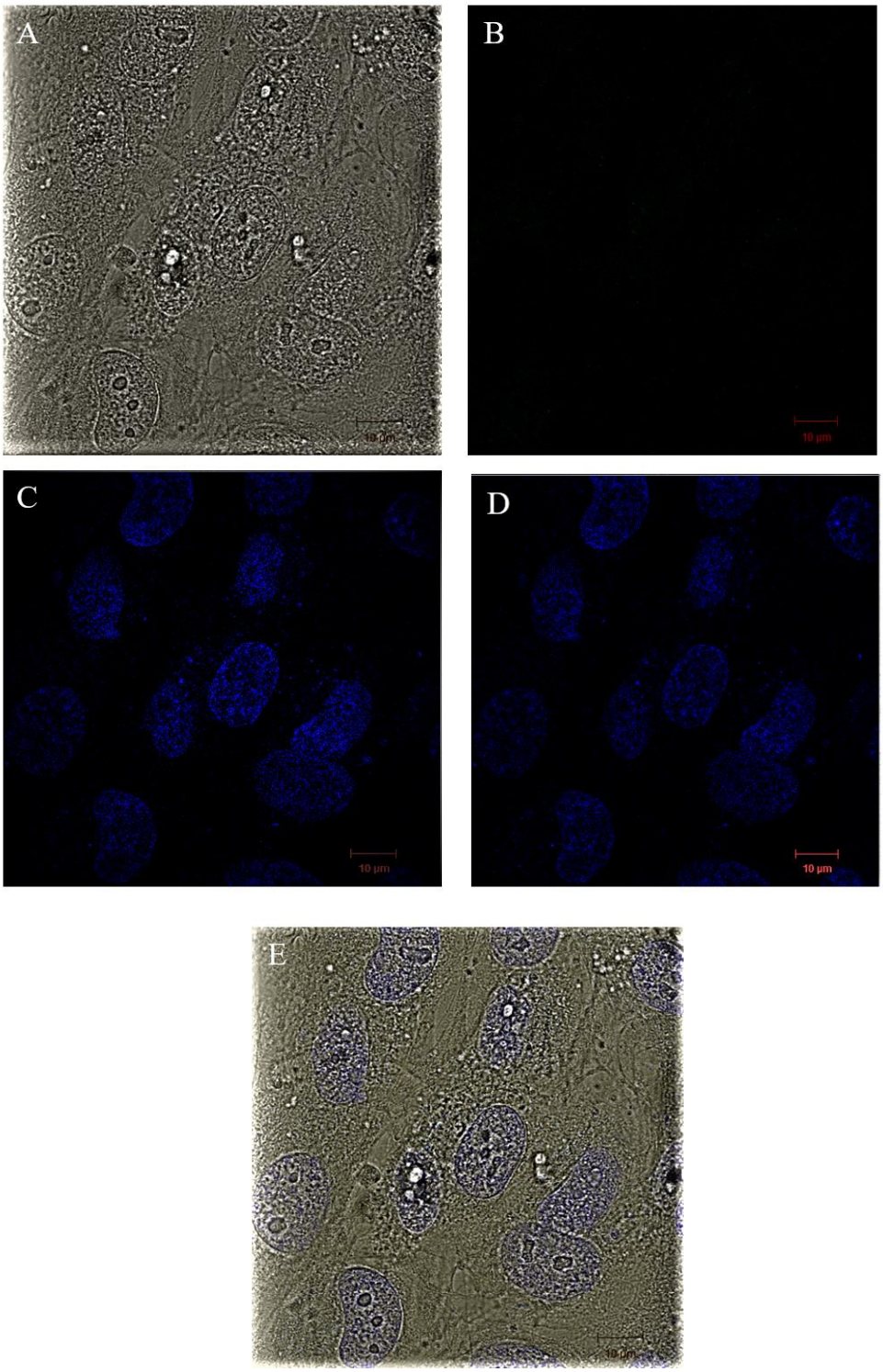


Figure 60: Staining of EpCAM-negative Hs578T cells with UBS54(scFv)-SNAPtag' labeled with Alexa 488 (green) for 1 h at 4°C. The nucleus was stained with Hoechst (blue) **excited** by ultraviolet light and emits blue light at 460 to 490 nm. Top left image (A) represents a bright field, top right (B) is cell surface staining (greenfield) excited by laser light at 499 nm, middle left (C) nucleus staining, middle right (D) is a superimposition of cell surface staining and

nucleus staining images, and finally, bottom image (E) is a superimposition of bright field, greenfield, and nucleus staining images.

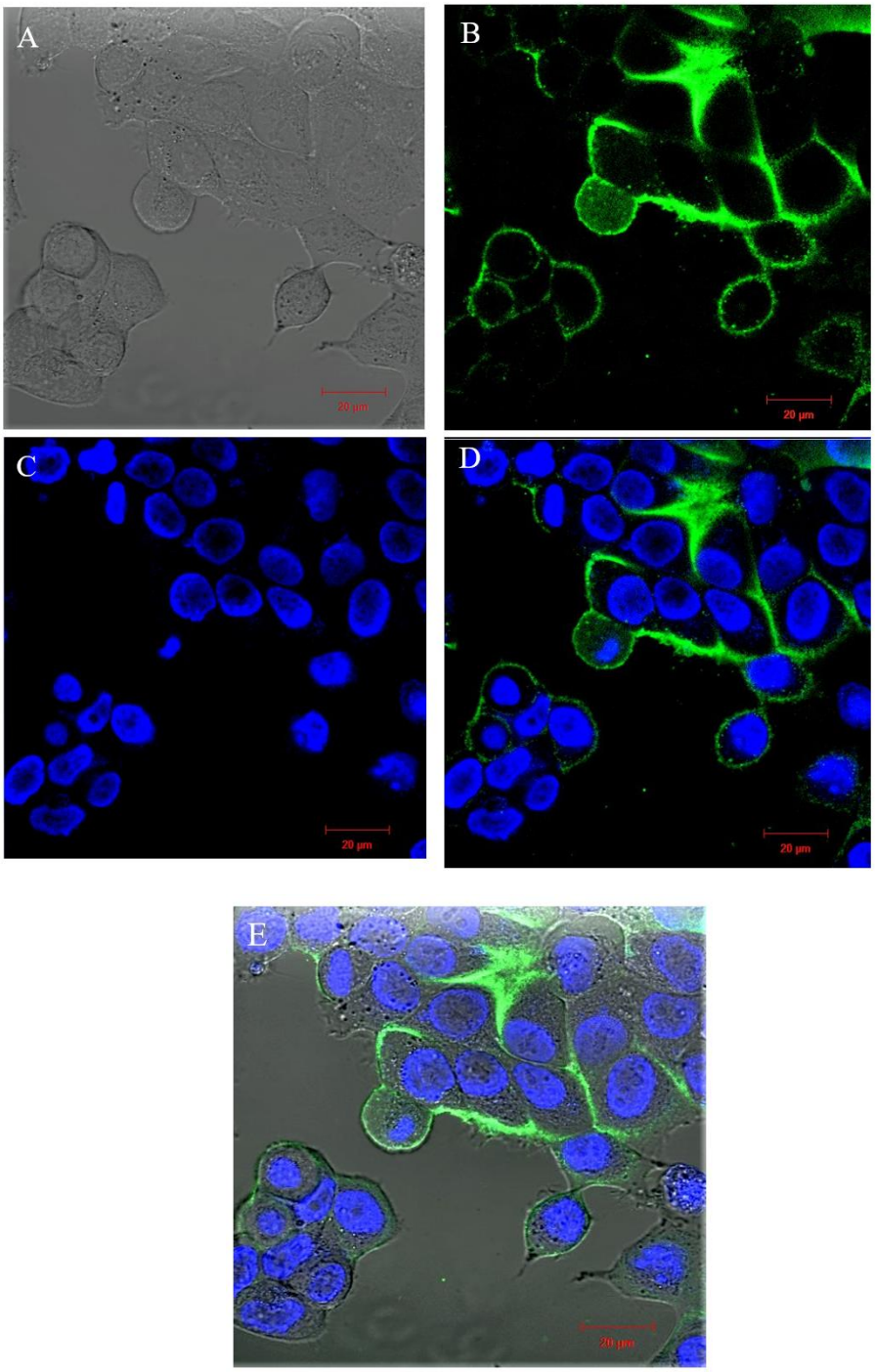


Figure 61: Staining of EMA-positive MCF-7 cells with ID405(scFv)-SNAPtag labeled with Alexa 488 (green) for 1 h at 4°C. Nuclear staining was achieved with Hoechst (blue), which is **excited** by ultraviolet light and emits blue light at 460 to 490 nm. Top left image (A) represents a bright field, top right (B) is cell surface staining (greenfield) excited by laser light at 499 nm,

middle left (C) nucleus staining, middle right (D) is a superimposition of cell surface staining and nucleus staining images, and finally, bottom image (E) is a superimposition of bright field, greenfield, and nucleus staining images.

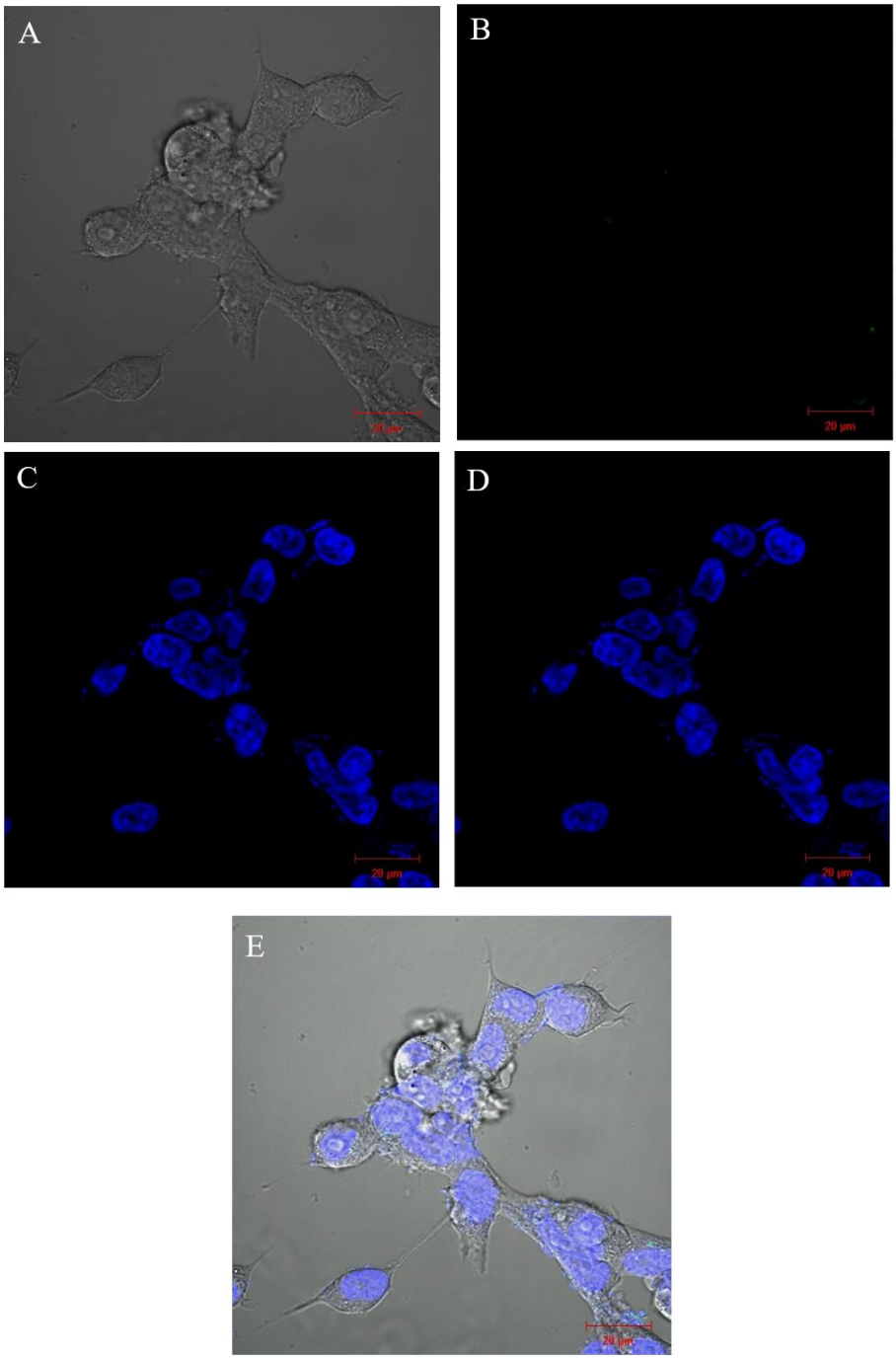


Figure 62: Staining of EMA-negative HEK293T cells with ID405(scFv)-SNAPtag labeled with Alexa 488 (green) for 1 h at 4°C Nuclear staining was achieved with Hoechst (blue) **excited** by ultraviolet light and emits blue light at 460 to 490 nm. Top left image (A) represents a bright

field, top right (B) is cell surface staining (greenfield) excited by laser light at 499 nm, middle left (C) nucleus staining, middle right (D) is a superimposition of cell surface staining and nucleus staining images, and finally, bottom image (E) is a superimposition of bright field, greenfield, and nucleus staining images.

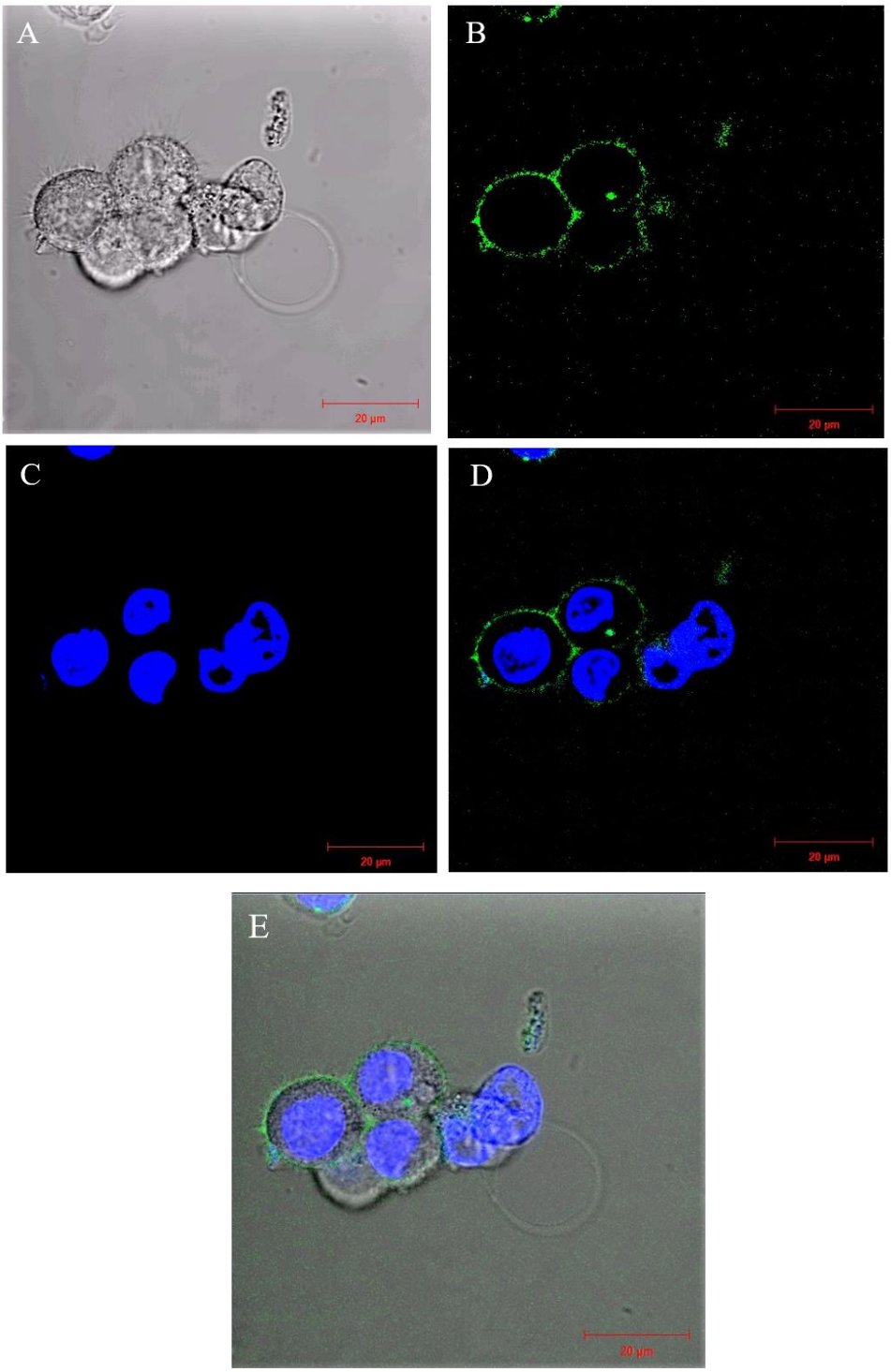


Figure 63: Staining of CD138-positive HEK293T cells with h-STL002 (scFv)-SNAPtag labeled with Alexa 488 (green) for 1 h at 4°C. Nuclear staining was achieved with Hoechst (blue)

excited by ultraviolet light and emits blue light at 460 to 490 nm. Top left image (A) represents a bright field, top right (B) is cell surface staining (greenfield) excited by laser light at 499 nm, middle left (C) nucleus staining, middle right (D) is a superimposition of cell surface staining and nucleus staining images, and finally, bottom image (E) is a superimposition of bright field, greenfield, and nucleus staining images

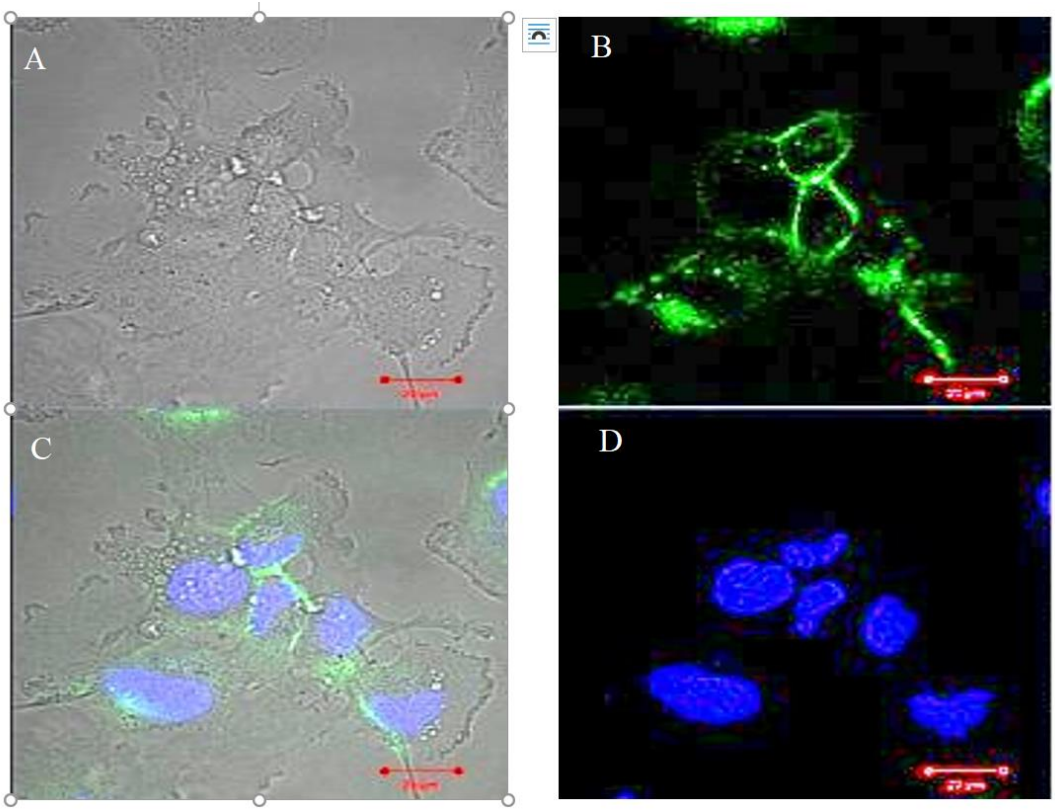


Figure 64: Staining of CSPG4-positive HS578T cells with mAb9.2.27 (scFv)-SNAPtag labeled with Alexa 488 (green) for 15 min at 4°C. Nuclear staining was achieved with Hoechst (blue), which is **excited** by ultraviolet light and emits blue light at 460 to 490 nm. The top left image (A) represents a bright field, top right (B) is cell surface staining (greenfield) excited by laser light at 499 nm, bottom left (C) is a superimposition of cell surface staining and nucleus staining images, bottom right (D) is nucleus staining.

3.8.2. Flow cytometry

Once successful binding of the BG-Alexa 488 labeled fusion proteins (scFvUBS54-SNAP, ID405-SNAP, and scFvh-STL002-SNAP) to their cognate antigens EpCAM, EMA, and CD138-expressing cells was shown by confocal imaging, we were interested to know the percentage of antigen-positive cells, and therefore tried to quantitate the percentage of cells that are positive for the antigen in a selected cell line that is already reported to be positive and negative for the antigen using flow cytometry (see Section 2.6.2.2.). As shown in Fig 65, the binding of the scFvUBS54-SNAP labeled protein was depicted on EpCAM positive cell lines (PANC-1), but not on control cells (HS578T). Interestingly, scFv UBS54-SNAP labeled protein could show 55% viable EpCAM-positive cell population in PANC-1 cell line, and viable cells are detected based on the exclusion of (7-AAD) staining (data not shown). Surprisingly, Pozza et al. have shown 41% of PANC-1 are positive for EpCAM which correlates with our study; when tested with **s**cFvUBS54-SNAP it was shown 55% of PANC-1 express EpCAM.[243]. Next, we investigated the binding activity of scFv9.2.27-SNAP in CSPG4-positive cell line HS578T and -negative cell line MDA-MB-468.

As shown in Fig 67, the binding of the scFv9.2.27-SNAP labeled protein was depicted on CSPG4-positive cell lines (HS578T), but not on control cells (MDA-MB-468), which corroborates the findings of a study by Amoury et al., that CSPG4 is moderately expressed in HS578T and absent in MDA-MB468.[244]

Finally, the binding activity of the scFvID405-SNAP labeled protein was depicted on EMA positive cell lines (MCF-7), but not on control cells (HEK293T) Fig 66. Interestingly, scFvID405-SNAP labeled protein could detect 29% viable EMA positive cell population in MCF-7 cell line, viable cells are detected based on the exclusion of (7-AAD) staining (data not shown). Moreover, our results of low expression of EMA in MCF-7 correlates with the study reported by Solatycka et al. and Kim et al.[245,246] We doubt whether the increase in the concentration of antibodies would further increase the proportion of the antigen-positive population. So, we added various concentrations of scFvID405-SNAP antibody total protein 3µg, 12µg, and 18µg. As shown in Fig 68, The histogram plot from left to right depicting 0% positive population for unstained and 9%, 30% 30% positive population when stained with different concentrations of scFvID405-SNAP (3 µg, 12 µg, and 18 µg). Mouffouk et al.[247] reported anti-MUC1/EMA with high affinity conferred increased fluorescence to target cells MCF-7 and MDA-MB468. In our study increasing the concentration from 3 µg to 12 µg

increases the positive population went from 10% to 30%. However, a further increase in concentration to 18 µg did no effect on the percentage of the positive population, showing that MUC1 receptor gets saturated at 12 µg.

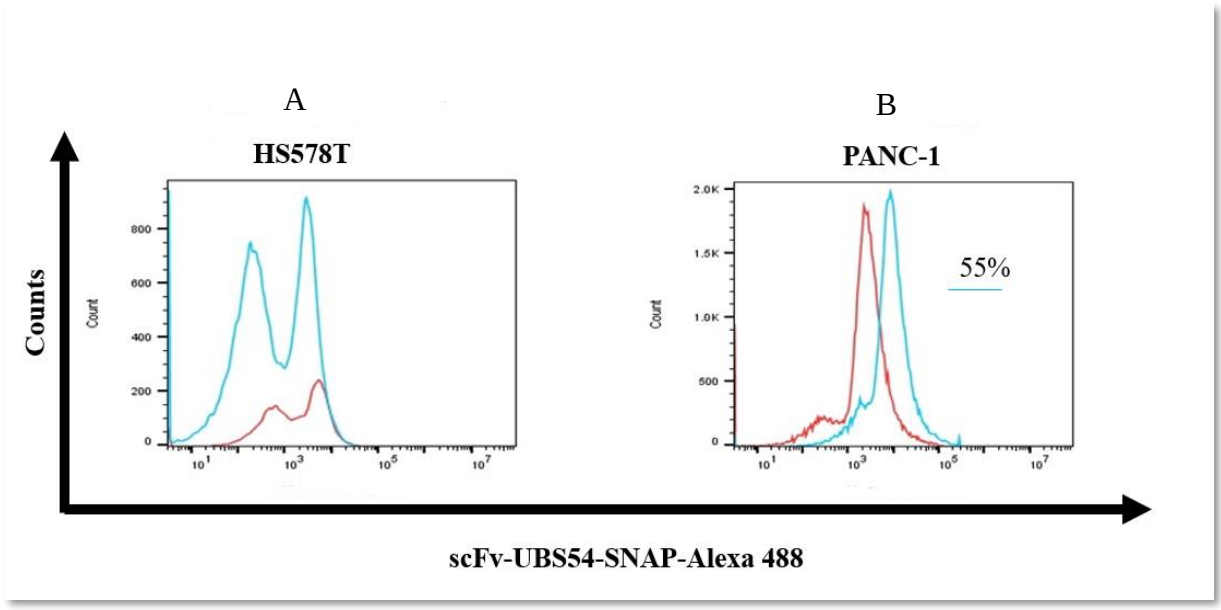


Figure 65: Representative flow cytometry curves of HS578T and PANC1.
PANC-1 and HS578T stained with the indicated antibodies scFvUBS54-SNAP (blue histogram) or unstained (red histogram). A) Double peak observed in the histogram for stained and unstained in HS578T could be due to cell clumps. The size of the double peak histogram (red) for unstained HS578T is small due to the low number of cells acquired during flow. B) The percentage of positive cells is reported on the flow cytometry curves for antigen-positive cells lines PANC1.

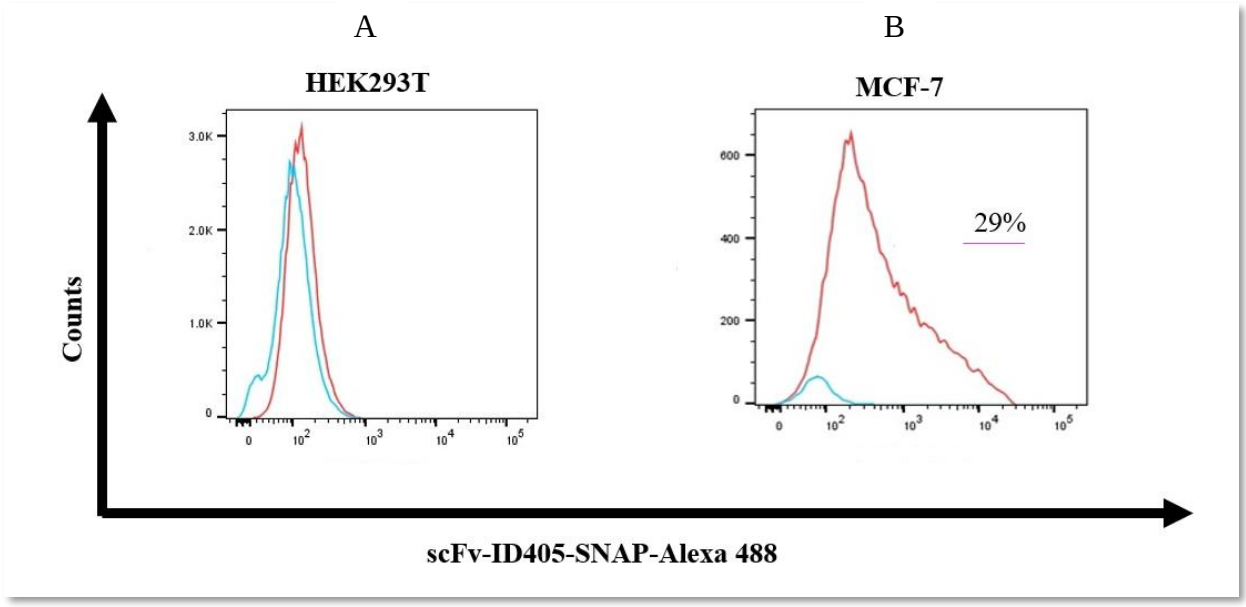


Figure 66: Representative flow cytometry curves of HEK293T and MCF-7.
HEK293T and MCF-7 is stained with the indicated antibodies scFvID405-SNAP (red histogram) or unstained (blue histogram). A) The size of the histogram (blue) for unstained in MCF-7 is small due to a low number of cells acquired in flow. B) The percentage of positive cells is reported on the flow cytometry curves for antigen-positive cells line MCF-7.

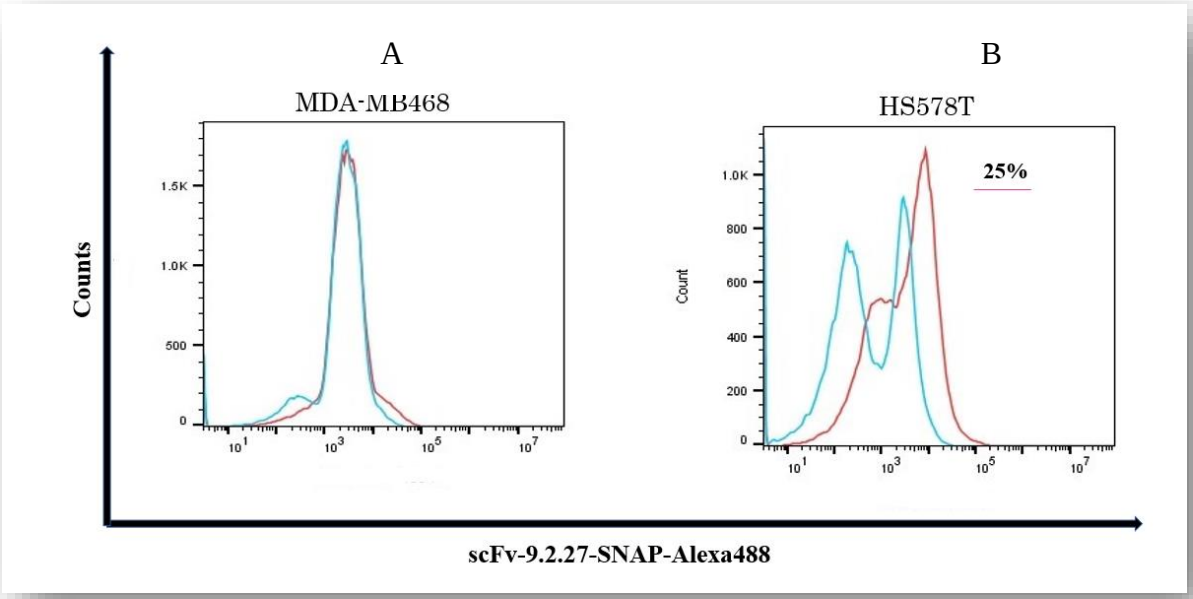


Figure 67: Representative flow cytometry curves of MDA-MB-468 and HS578T.

MDA-MB-468 & HS578T are stained with the indicated antibodies scFv9.2.27-SNAP (red histogram) or unstained (blue histogram). A) The histogram (red) for stained in MDA-MB-468 overlap with histogram (blue) of unstained B) The percentage of positive cells is reported on the flow cytometry curves for antigen-positive cells line HS578T, where the stained histogram (red) shows 1 log shift with unstained histogram (blue) in X-axis.

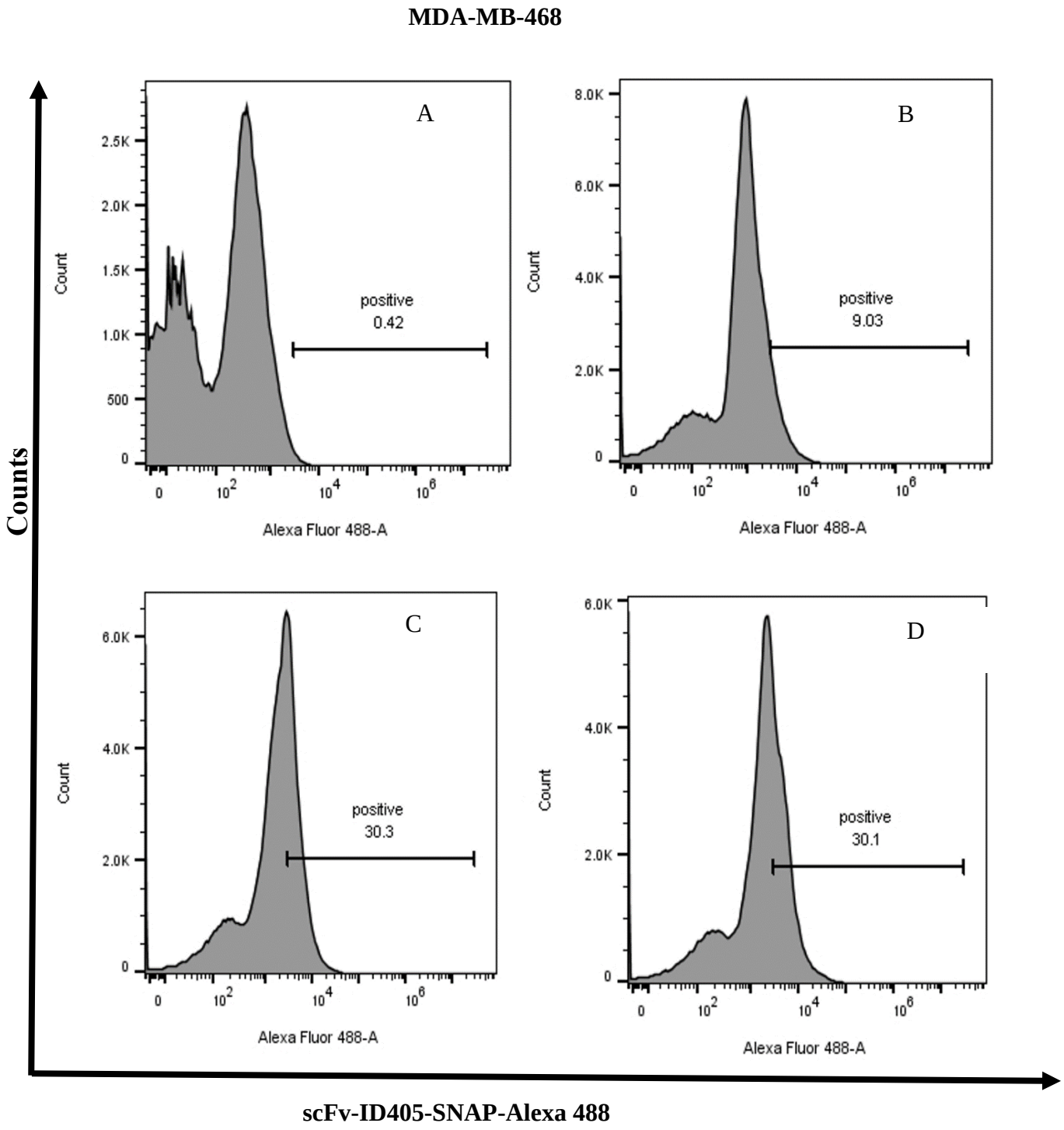


Figure 68: Representative flow cytometry curves of MDA-MB-468 stained with different concentrations of scFvID405-SNAP. A) Histogram plot from top left depicting 0% positive cell population for unstained MDA-MB-468, followed by the top right (B), bottom left (C), and bottom right (D). Cells were stained with different concentrations of scFvID405-SNAP (3µg, 12µg, and 18µg) to get 9%, 30% 30% positive population in MDA-MB468 (B, C, D). A) Double peak histogram in unstained MDA-MB-468 could be due to cell clumps.

3.9. Cytotoxic effect of αEpCAM(scFv)-SNAP-MMAF in EpCAM-positive and -negative cell lines

Cell lines are an indispensable tool to study the in vitro toxicity of drug of interest. In this in vitro cytotoxic study we chose EpCAM-positive cell line MCF-7 and MDA-MB-468 and the EpCAM-negative cell line MDA-MB231 to evaluate cytotoxic activity of an ADC (αEpCAM(scFv)-SNAP-MMAF). Both the unconjugated (drug alone) and conjugated form (targeting moiety and drug) are investigated to show the proof of concept that the fusion protein UBS54scFv-SNAP-AURIF can induce toxic effect in EpCAM enriched cell line, by which we are opening new therapeutic window for BCC patient who cannot undergo surgical excision of primary BCC tumor and who developed vismodegib-small molecular inhibitors resistance BCC.

The cytotoxic effects of UBS54scFv-SNAP-AURIF fusion proteins and unconjugated BG-monomethyl auristatin F (MMAF) were assessed using an XTT-based colorimetric cell proliferation assay. Briefly, MCF-7, MDA-MB-468, and MDA-MB-231 were incubated at 37°C for 24 h and then treated with series of concentration of unconjugated BG-MMAF, and UBS54scFv-SNAP conjugated BG-MMAF. After 72 h incubation with the unconjugated MMAF, the viability of MCF-7, MDA MB-468, and MDA-MB-231 cells was reduced in a concentration dependent manner (see **Section 2.7.2** and **Fig 69**). Adversely, unconjugated BG-MMAF (BG-AF) was toxic towards all cell lines irrespective of differential EpCAM expression with IC_{50} values of 494 nM (MCF-7), 152 nM (MDA-MB468), 297 nM (MDA-MB-231) as shown in **Fig 69** and **Table 21**.

Next, we investigated whether the concentration of full-length protein of UBS54-scFv-SNAP-BG-MMAF has any dose dependent cytotoxic effect on EpCAM positive MCF-7 and MDA-MB-468 and EpCAM negative MDA-MB-231 cell line. As speculated UBS54-SNAP-BG-MMAF was specifically killing MCF-7 and MDA-MB-468 in a dose-dependent manner with IC_{50} values of 1668 nM (MCF-7) and 3350 nM (MDA-MB-468), whereas the MDA-MB-231, an EpCAM-negative cell line (**Fig 70**), was not affected within the stipulated concentration, thus ensuring no off-target toxicity was detected in this cell line.

In addition to EpCAM targeting ADC, we were interested to test the cytotoxic effect of anti-EMA-SNAP (ID405-scFv-SNAP-BG-MMAF) both as a total protein and full length protein in EMA positive and negative cell line MCF-7 and MDA-MB-231 (**Fig73)**. As anticipated total protein concentration is 5 times more than full length protein with IC:50 value of 5110 nM and 1184 nM in MCF-7. This evidence strongly supports the interference of N-terminal truncated

protein in estimating the absolute toxic effect by occupying the EMA receptors. Next we envisaged the toxic effect of total protein and full length protein in in EMA negative cell line, without any surprise both total protein and full length protein did not induce any toxic effect in EMA negative cell line MDA-MB-231, which further emphasize that the ADC ID405scFv-SNAP-AF is selective to antigen expressing cells (**Fig 73**).

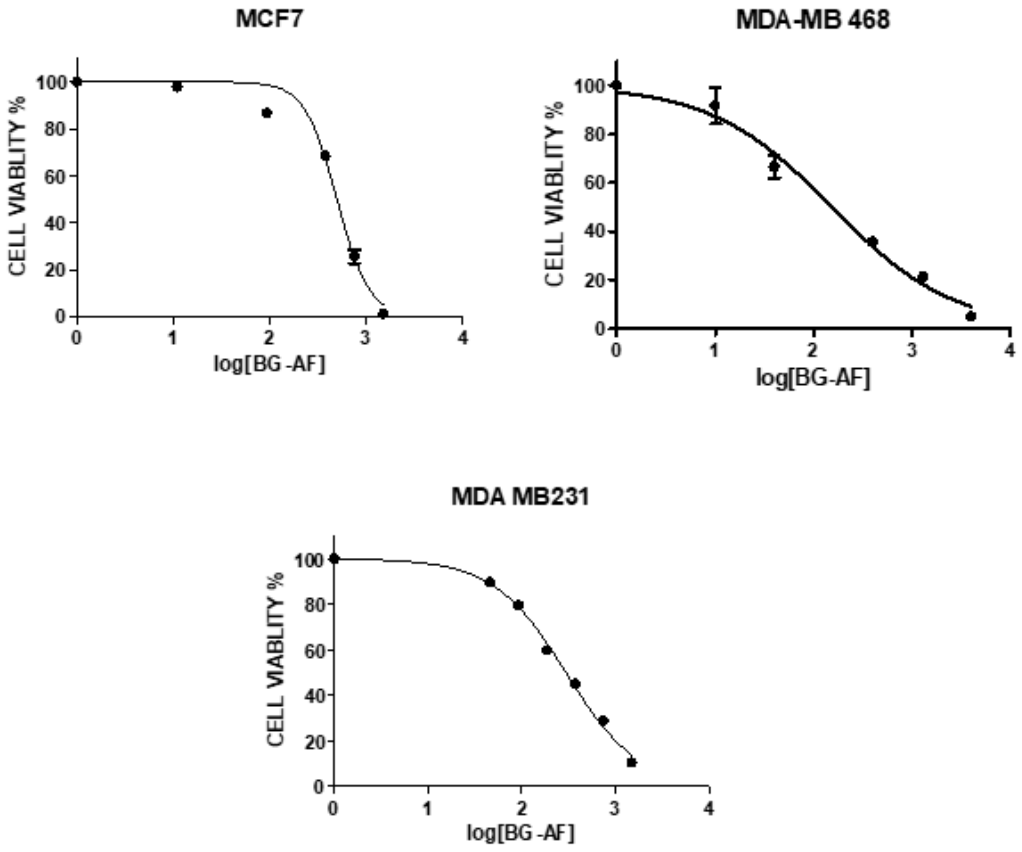


Figure 69: Cell viability assay 72 h of MMAF treatment on EpCAM$^{+/-}$cell lines

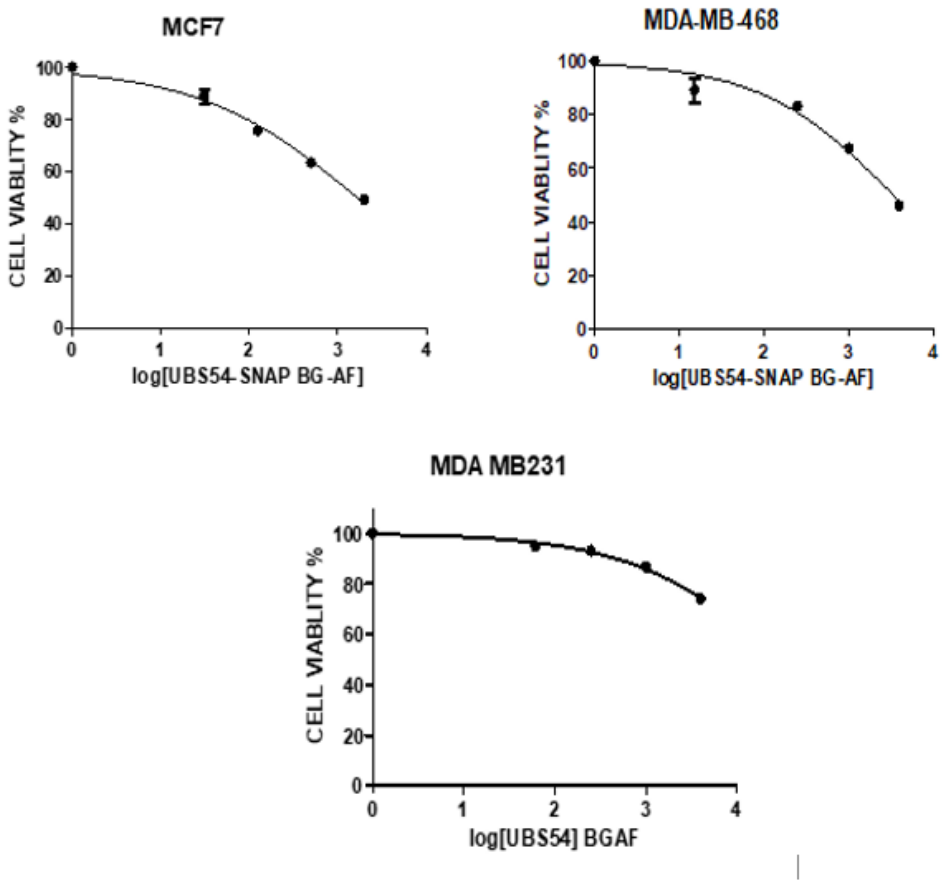


Figure 70: Cell viability assay 72 h of UBS54-SNAP-MMAF treatment on EpCAM$^{+/-}$cell lines

TREATMENT	**MCF-7**	**MDA-MB468**	**MDA-MB-231**
EpCAM EXPRESSION LEVEL*	HIGH	MODERATE	LOW/NO EXPRESSION
BG-AF/BG-MMAF (BG-AURISTATIN F) **(IC_{50}, nmol/L)**	493.7	152.4	296.7
UBS54scFv-SNAP-MMAF **(IC_{50}, nmol/L)**	1668	3350	ND

Table 21: IC_{50} values for BG-MMAF and UBS54scFv-SNAP-MMAF on EpCAM$^{+/-}$ cell lines: the IC_{50} values are derived from XTT based cell viability assays and indicate the UBS54scFv-SNAP-BG-MMAF concentration required to achieve a 50% reduction in cell viability in EpCAM^{+} relative to the vehicle control cells. (ND: Not Defined). *As per Amoury et al. 2016 report[244]

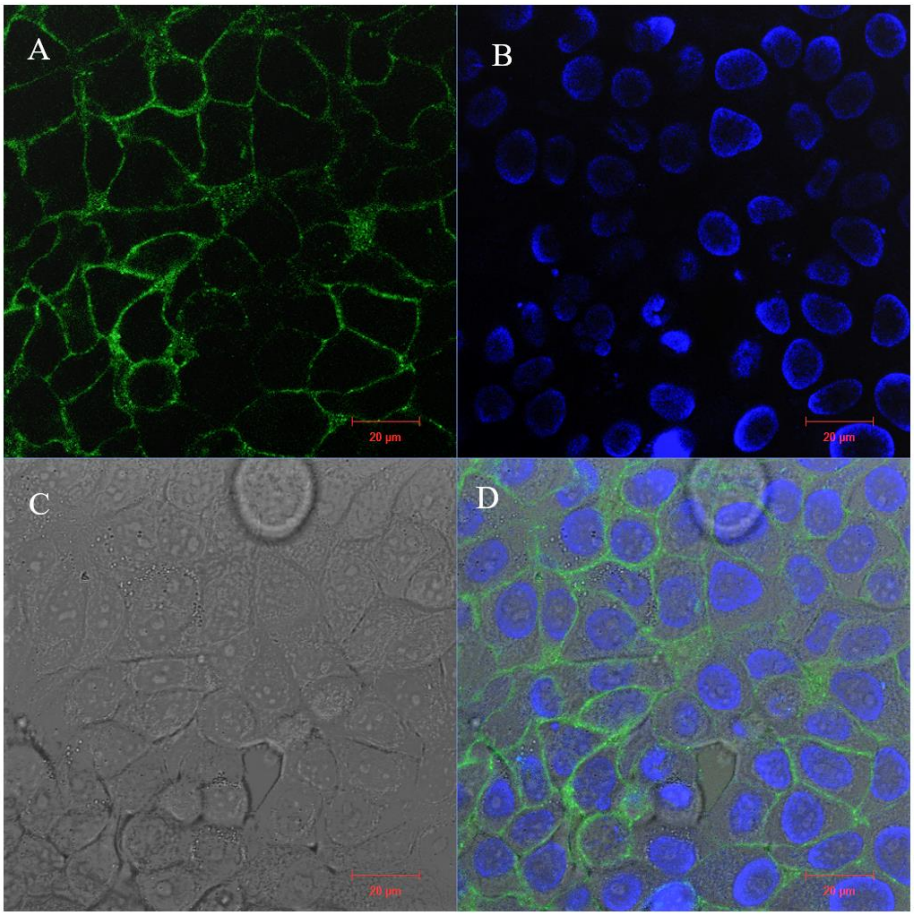


Figure 71: Staining of EpCAM-positive MCF-7 cells with UBS54(scFv)-SNAPtag labeled with Alexa 488 (green) for 1 h at 4°C. EpCAM positive MCF-7 cells were stained with UBS54(scFv)-SNAPtag labeled with Alexa 488 (green) for 1 h at 4°C. Nuclear staining was achieved with Hoechst (blue) excited by ultraviolet light and emits blue light at 460 to 490 nm. The top left image (A) represents cell surface staining (greenfield) excited by laser light at 499 nm, top right (B) is nucleus staining, bottom left (C) represents a bright field, bottom right (D) is a superimposition of bright field, greenfield, and nucleus staining images.

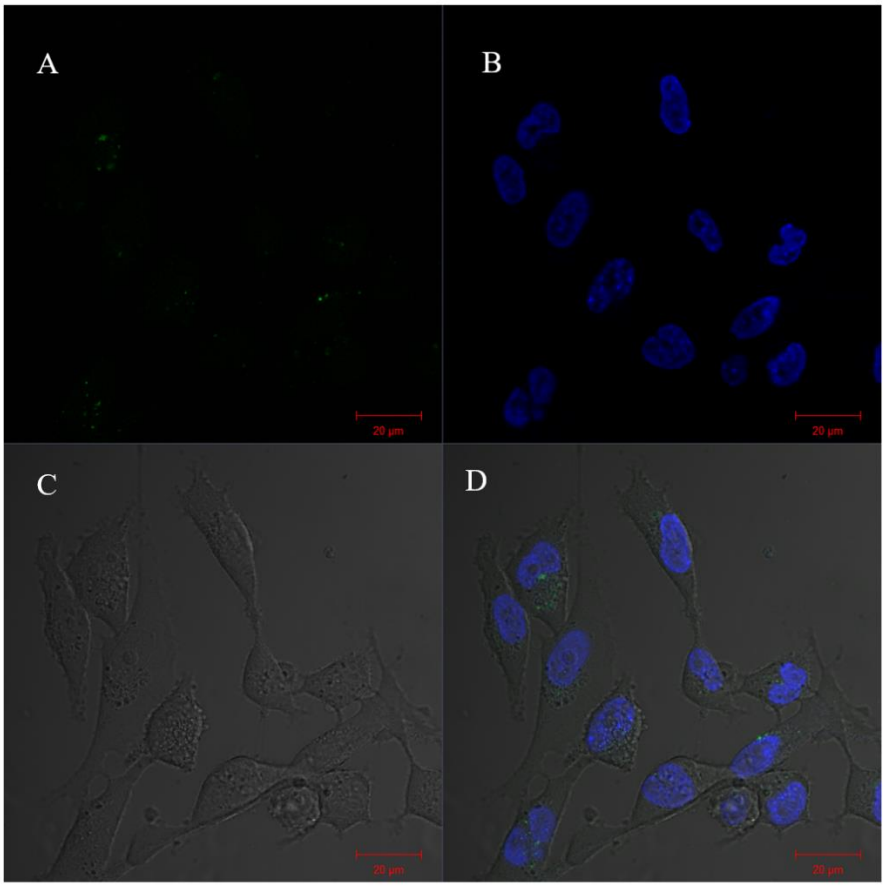


Figure 72: Staining of EpCAM-negative MDA-MB-231 cells with UBS54(scFv)-SNAPtag' labeled with Alexa 488 (green), for 1 h at 4°C. EpCAM-negative MDA-MB-231 cells were stained with UBS54(scFv)-SNAPtag' labeled with Alexa 488 (green), for 1 h at 4°C and the nucleus was stained with Hoechst (blue) **excited** by ultraviolet light and emits blue light at 460–490 nm. Top left image (A) represents cell surface staining (greenfield) excited by laser light at 499 nm, top right (B) is nucleus staining, bottom left (C) represents a bright field, bottom right (D) is a superimposition of bright field, greenfield, and nucleus staining images.

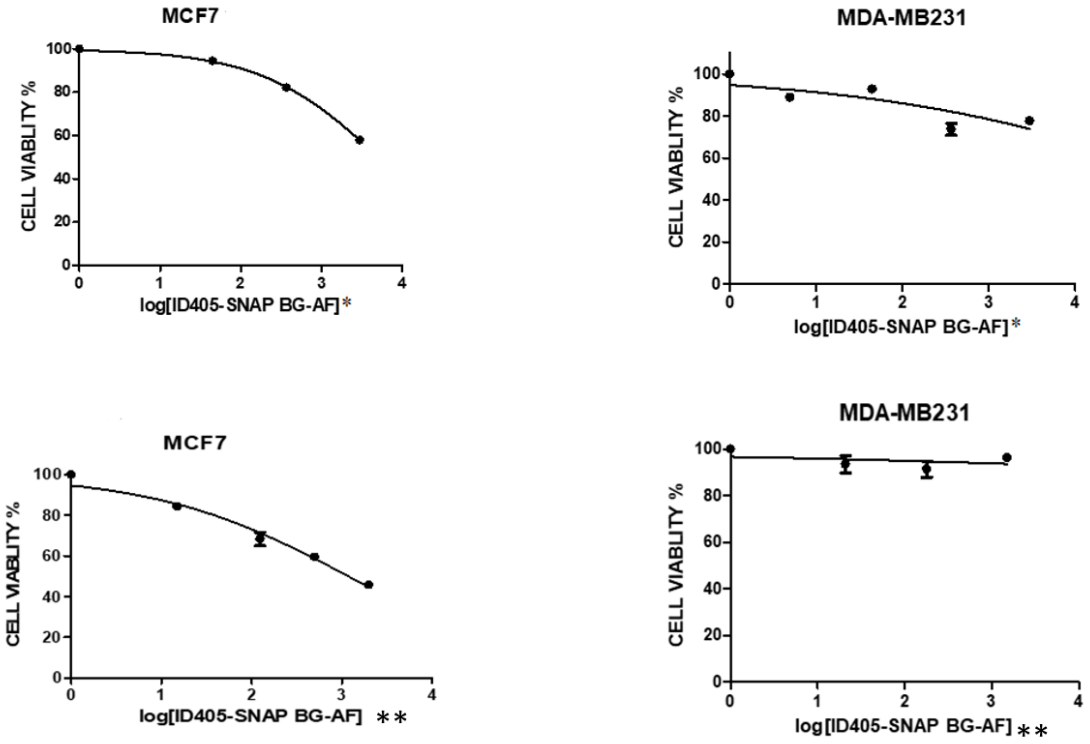


Figure 73: Cell viability assay 72 h of ID405-SNAP-AF treatment on EMA$^{+/-}$cell lines

Treatment	**MCF-7**	**MDA-MB-231**
EMA EXPRESSION LEVEL	HIGH	No expression
ID405scFv-SNAP-MMAF* **(IC_{50}, nmol/L)**	5110	ND
ID405scFv-SNAP-MMAF **(IC_{50}, nmol/L)**	1184	ND

Table 22: IC_{50} values for ID405scFv-SNAP-MMAF* & ID405scFv-SNAP-MMAF on EMA$^{+/-}$ cell lines: the IC_{50} values are derived from XTT based cell viability assays and indicate the ID405scFv-SNAP-BG-MMAF concentration required to achieve a 50% reduction in cell viability in EMA^{+} relative to the vehicle control cells (MDA-MB-231). (ND: Not Defined).

*Total Protein concentration estimated by UV–visible spectrophotometer

**Full-length protein concentration estimated by densitometry

Chapter 4: Discussion

This chapter is divided into two parts. The first part (Section 4.1) describes the identification of BCC-specific cell surface antigens, while the second part (Section 4.2) demonstrates the design, generation, and characterization of all the fusion proteins used in this study. Finally, the biological activity of the purified targeted SNAP-tag–based antibody fusion proteins has been described.

4.1. QUEST FOR NOVEL BCC BIOMARKERS

BCC is the least invasive of all cancers, with a low death rate of 0.12 per 100,000.[16,26,21,22] However, it has a very high incidence and relapse rate and thus remains a substantial burden on the South African health care system, accounting for a total annual cost of ZAR 78.4 to 113.5 million ($13.3–19.3 million) and affecting people from low socioeconomic backgrounds.[16,23,248] Therefore, there is a pressing need to appropriately reduce this burden on health care resources. Biomedical scientists and physicians need to adopt cost-effective paradigms for managing BCC. BCCs with an aggressive histology (micronodular, morpheaform, those with squamous differentiation, infiltrative) are widely managed by Mohs micrographic surgery (MMS), which costs around R45,000 per treatment. MMS involves repeated excision of a thin layer of tissue under local anesthesia. Each layer is examined under the microscope by a pathologist at an on-site laboratory. The excision is repeated until the layers are free from tumor cells.[130] Though MMS has been reported to show a high cure rate and spares maximal amount of normal tissue, the technique does have its drawbacks. For instance, this procedure is not as effective when patients present skip lesions or multiple micro lesions that account for a high relapse rate[249,89]. Knowing the morphology of entire tumor lesions both superficially and beneath the superficial layers irrespective of various subtypes of BCC would eliminate the need for multiple rounds of excision and reduce the relapse rate and overall cost of the surgery. With the advent of tumor-specific markers or tumor-associated markers and the increased availability of antibodies specific to these markers, one can label the tumor specifically with a dye that can be excited under UV, visible, infrared, and far-infrared light, allowing for accurate detection and consequently improved treatment of the tumor, with a single round of excision. Parallelly, the payload of such antibodies (that are labeling dyes) can be replaced with a photosensitizer (or) cytotoxic drug to treat surgically inaccessible or complex tumors. Until now, the two prominent cell surface markers that are commonly used in BCC differential diagnosis are EpCAM (TROP1/CD326) and EMA (MUC-1). EpCAM is a positive marker expressed in BCCs, irrespective of its subtype,

and EMA serves as a negative marker for BCCs.[250] By far, research has shown that any BCC is EpCAM positive and EMA negative. We thus chose EpCAM as our primary marker for our investigation. However, few studies have reported loss of EpCAM expression in BCCs. For example, Ansai et al. and Rajabi et al.[251,252] reported that a very limited number of BCC cases are BerEP4-negative. BerEP4 is a monoclonal antibody that binds to EpCAM.[253] A very recent study by Gaiser et al.[105] showed that nearly 12% of nodular BCC and 60% of infiltrative BCC showed loss of EpCAM expression at the invasive tumor front. EMA, the negative marker, is reported to be focally expressed in 8–10% of BCC.[254,255,256,253] Bartoš et al. claim that this limited number of controversial cases (loss of EpCAM expression and expression of EMA in a small population of BCCs) might represent a technical error, say, in immunohistochemical processing, rather than true negativity[253]. However, we speculate that this loss of EpCAM in BCCs could be truly negative based on the study conducted by Gaiser et al., where the investigators showed nearly 60% of infiltrative BCCs were negative for EpCAM at the invasive front when stained with BerEP4. A technical inconsistency may not entirely explain a variation as high as a 60% proportion of false results.[105] In addition, some epithelial cancers such as breast cancer that are positive for EpCAM lose their expression when they transition from epithelial to mesenchymal state, and have a high potential to metastasize.[257] Therefore, to improve diagnosis accuracy and outcome of BCC patients, there is an urgent demand for new diagnostic, prognostic, and therapeutic targets that can be used individually or in combination with existing primary markers such as EpCAM and EMA to stain the entire BCC tissue in 100% of the patient population with high specificity and selectivity, which would allow for complete removal of BCC tissues, no matter the subtype, through a single excision procedure.

4.2 IDENTIFICATION OF BCC-ASSOCIATED BIOMARKERS

We performed an integrated cDNA-microarray analysis of BCC using four different microarray datasets (**Table 5**). In total, 3,745 genes were identified to be differentially expressed between BCC and nontumor control tissues (1764 upregulated and 1981 downregulated genes) with the *p*-values below our defined threshold (*p*-value <0.001). Out of these 3,745 differentially expressed genes (DEGs), there are 401 DEG (213 genes are 3FC [fold change] upregulated and 188 genes are 3FC [fold change] downregulated with significant *p*-values and FDR < 1e-^{07} compared to normal skin [human epidermal keratinocytes]). One should be highly attentive to setting right fold change and statistical cutoffs, as it might significantly alter the outcome of microarray interpretation. In general, fold change >2 and *p*-values of <0.02 are recommended

for most studies.[258] However, we chose a fold change >3 in consideration of Zhao et al.'s findings that TROP2 gene is 2.7-fold upregulated in triple-negative breast cancer (TNBC) compared with other BC (breast cancer) subtypes and adjacent normal breast tissue. The TROP2 gene, which encodes a cell surface protein, has been targeted with an ADC, Sacituzumab govitecan-hziy, in TNBC which was approved by FDA for the treatment of TNBC[259]. Thus, we envisaged that fold change >3 would be the optimal threshold for this BCC study, Moreover, these differentially expressed genes satisfy both statistical significance (such as *p*-value and FDR) and biological relevance (such as fold change) simultaneously give high assurance that the differences found are large enough to be biologically meaningful.

Consequently, the aforementioned 401 DEG genes are considered to be highly statistically significant and biologically relevant differentially expressed genes between diseased (basal cell carcinoma) and normal as they satisfy the chosen criteria (fold change >3 or <3, *p*-values and FDR <1e-^{07}**)**, and thus considered for downstream analysis.

Screening for cell surface antigens/receptors for the 401 DEGs manually is difficult and time-consuming, let alone do the same for all the 3745 DEGs. To circumvent this laborious process, we compared the 401 DEGs with 2,886 proteins that are exposed at the cell surface based on SURFY, an *in silico* prediction surfaceome, which corresponds to 14.3% of the entire human proteome. SURFY, a public resource developed by Damaris Bausch-Fluck et al., works by machine-learning approach utilizing "transmembrane" domain-specific features to predict possible cell surface proteins from the human proteome of 20,193 proteins. Close to three thousand (2,886) proteins were found to be exposed at the cell surface, which amounts to 14.3% of the entire human proteome.[204] Subsequently, this comparison resulted in our prediction of 63 genes out of the 401 DEGs coded for cell surface markers. Of these 63 genes, 34 genes were at least 3-fold upregulated in BCCs, and 29 genes are 3-fold downregulated in BCCs. Also, we carried out a comprehensive literature search to identify any antibodies used to target these 34 genes encoding proteins in other cancers, and to our surprise we found that various groups have used antibodies to target ALK, LGR5, CSPG4, PTCH1, LRP8, SLC7A5/LAT-1, CDH3, SORL1, CXCR4, SLC7A2, TNFRSF21, FZD7, TSPAN8, and EpCAM, paving the way to use them to discriminate BCCs from nontumor control tissues. These antibodies can also be used for ADC, PDT, and CAR T-cell–based therapies to selectively kill BCC tumor cells when MMS (Mohs micrographic surgery) or wide-excision surgery is not possible for selected BCC cases. Additionally, all these upregulated genes and their translated proteins have been used to target

and treat various human cancers, such as neuroblastoma; colorectal cancer; breast cancer; melanoma; non–small cell lung cancer (NSCLC); ovarian, prostate, pancreatic cancers; and renal cell carcinoma. Generating antibodies for such antigens can not only be useful for treating BCC but can also be used to treat other malignancies that overexpress such cell surface proteins. Interestingly, various authors have reported the increased expression of ALK, LGR5, PTCH1, and EpCAM in BCC, further validating our computational findings and suggesting that the described genes and their peptides can be an effective target for BCC diagnosis and therapy. These upregulated genes are mostly involved in cell growth, proliferation, differentiation, migration, cell survival, and anticancer treatment resistance. Therefore, these genes and the encoded proteins are not only candidates for differential diagnosis and targeted therapy but can also be considered as prognostic markers to predict the clinical outcomes in BCC patients. For example, we speculate that the loss of EpCAM expression in BCC, as reported by Gaiser et al., could be associated with epithelial–mesenchymal transition (EMT), which has already been reported to cause drug resistance, local invasion, and metastasis.[105,260,261]

Vismodegib is the hedgehog (Hh) signaling inhibitor drug used to treat BCCs. Most BCCs appear with aberrant activation of hedgehog signaling due to mutation in PTCH1 and SMO.[262,263] In certain cases, BCC tumors tend to relapse after withdrawal of vismodegib treatment. It was found that a small population of $LGR5^{+}$ BCC cells with active Wnt signaling evade vismodegib-induced differentiation and mediate tumor relapse after treatment discontinuation. [264] One possible reason behind this observation could be Wnt signaling pathway's ability to cross-talk with Hh signaling pathways in the pathology of BCC and provide resistance to vismodegib.[265]

In our research, we noted that LGR5 is 7-fold upregulated and the CTNNB1 gene that encodes for B-catenin involved in Wnt signaling are 2-fold upregulated in BCC compared with normal skin (human epidermal keratinocytes); these findings indirectly provide additional evidence that the outcome of our computational gene expression analysis is robust and our data was not an observation by chance. Further, we emphasize targeting LGR5 with ADCs in BCC, in combination with vismodegib or without the drug, could decimate any relapsed BCC tumor. It is worth mentioning that nearly 10% to 67% of BCC cases appear after complete or incomplete excision.[266,23] Thus, we need additional cell surface markers to reduce the recurrence rate by improving the outcomes of surgical excision, to detect BCC cells at the invasive front that appear with no EpCAM expression, and to treat vismodegib-resistant BCCs. We chose to explore the expression of LGR5, CSPG4 along with EpCAM in BCC. Although LGR5 has previously been reported to be overexpressed in localized and metastatic BCCs compared to normal skin tissue, its expression is poor along the margins compared to the central region of BCC tumors.[110] This

holds true for EpCAM expression as well.[105] For reasons unknown we were not able to generate anti-LGR5 in mammalian host expression systems. However, we could successfully develop anti-EpCAM and anti-CSPG4 from mammalian host systems. Interestingly, till date, CSPG4 was not reported by others as an overexpressed marker for BCC, which led us to explore whether this cell surface antigen can be a potential marker for BCC diagnosis, prognosis, or treatment. If it is proven to be a useful differential marker for BCC, understanding the expression levels of CSPG4 in the tumor periphery and core region of BCCs. In our integrated microarray reanalysis study, we found that CSPG4 was upregulated 10-fold in BCC tumors compared with normal skin, which corroborates the findings of a previously published computational study done by Litvinov et al.; however, they have not mentioned CSPG4 in the publication.[267] The authors of the study were more interested in new targetable signal transduction pathways rather than targetable cell surface receptors, and thus did not focus on reporting the expression of EpCAM, CSPG4, or LGR5 as a target for BCC treatment. CSPG4 overexpression is reported in a wide range of human cancers such as melanoma, breast cancer, soft-tissue sarcoma, pancreatic cancer, and glioblastoma. The protein plays a central role in the oncogenic pathways required for malignant progression, metastasis, and survival of cancer stem-like cells. This feature and its poor expression in healthy and normal adult tissues strongly suggests CSPG4 is an ideal oncoantigen for anti-tumor immunotherapy.[268] Another study by Wan et al. showed that the expression of PTCH1 and GPR161 was upregulated in BCC compared with adjacent normal tissues by both RNA-seq and real-time PCR, which is in agreement with our findings. PTCH1 and GPR161 were upregulated 3-fold in BCCs.[225]

Downregulated cell surface protein–encoding genes can also be used as a marker for differential diagnosis and prognosis. EMA, the negative marker for BCC, can assist in distinguishing the BCC lesions from SCC lesions when coupled with EpCAM.[91,92] CD138, also known as SDC-1, was reported to have low expression in BCCs compared to normal keratinocytes.[117,118] A combination of these two markers can thus be used to distinguish BCC lesions from other lesions. We could not find evidence of differential expression of EMA or CD138 from our computational analysis of the differentially expressed genes when comparing BCC with normal skin. However, we observed that the probeset 213693_s_at for MUC1 and 201287_s_at for CD138 that represented raw signals did not fulfill the filtering and subsetting criteria that we chose. Additionally, we searched for CD90 in our class comparison output file, as this marker had previously been reported to be moderately expressed in BCC tissue compared with adjacent normal skin by Milosevic et al.[110] We could not find the CD90 gene in output list, but the probeset 208851_s_at that corresponds to CD90 was observed in the raw data, indicating once again that

the criteria that we chose to get differentially expressed genes were tightly controlled with a high statistical threshold (nominal significance level of each univariate test: 0.001), which might be the reason for CD90's absence in the output; this further emphasizes the importance of threshold setting—more stringent threshold setting might lead to loss of few truly differential genes, and also avoid false-positive genes. Lastly, we examined the expression level comparison for EGFR, CD44, and CEA markers, which were reported to be expressed in BCC. All the above markers were seen in the list of differentially expressed genes between BCC and NS with *p*-value <1e-^{07} and fold change >3.

Consecutively, our class comparison results matched with the extant data for most of the cell surface markers (EpCAM, LGR5/GPR49, ALK, PTCH1, GPR161, CD44) that were reported to be upregulated in BCC by others. In our study, we chose 3-fold change and *p*-value<1e-^{07}. Our results thus reinforce our inference that the computational results we have obtained are robust; most notably, our computational analysis using BRB-ArrayTools, and alternative computational tools such as SAM and TAC, and supplementary data published by Litvinov et al. favor our hypothesis that CSPG4 could be among the novel markers employed for the accurate distinction of BCC.[267]

4.3. DESIGN, EXPRESSION, AND PURIFICATION OF RECOMBINANT SCFV-SNAP-TAG

After designing and synthesizing the plasmid by molecular cloning, the corresponding plasmids encoding for the recombinant antibody fusion proteins (as scFv-SNAP- fusions) against various target antigens of interest, namely, EpCAM, EMA, CD90, LGR5, CD138, and CSPG4 are transfected into mammalian cells for the production and secretion of antibody-SNAP fusion proteins. We were successful in obtaining the full-length functional proteins of UBS54(scFv)-SNAP, ID405(scFv)-SNAP, h-STL002(scFv)-SNAP and mAb9.2.27(scFv)-SNAP. However, for unknown reason we failed in generating YW353(scFv)-SNAP and h5-Thy1 (scFv)-SNAP that targets LGR5 and CD90.

The recombinant plasmid UBS54(scFv)-SNAP that encodes for anti-EpCAM fusion protein are reported by others to recognize EpCAM expressed both in diseased (BCC) and normal cell, we have shown its binding activity in various breast cancer that are reported to be EpCAM positive but not in EpCAM negative cell line. In addition, we designed an expression plasmid for another antibody fusion protein named OCAb9-1(scFv)-SNAP that is specific to EpCAM expressed in the diseased cells, but not in normal tissues, as reported by Liao et al. The authors of the study

do not provide a reason for the preferential recognition of EpCAM in tumor cells by OCAb9-1.[261]

We examined the antibody recognition sites for UBS54 and OCAb9-1 to understand the difference between the two antibodies. OCAb9-1 recognizes the epitope EGF-like domain II of EpCAM, whereas UBS54 binds to the first EGF-like repeat of EpCAM. We do not know whether this recognition site could have impacted preferential binding to tumor cells. However, we can speculate from the study reported by Garrett et al. that mAb806 recognizes EGFR expressed in tumor, but not in normal cells, due to a cryptic epitope in EGFR.[262] We hypothesize that similar to EGFR, EpCAM expressed in cancers might have some hidden mutation in the EpCAM exon-encoding gene that might change the conformation in the epitope recognition site for OCAb9-1.

The recombinant plasmid coding for OCAb9-1(scFv)-SNAP was transiently transfected into HEK293T cell culture expression system and scFv-SNAP–expressing cells were enriched by antibiotic selection. We preferred mammalian cells (HEK293T) as the expression host, as it has in-built mechanisms required for proper posttranslational modification of recombinant proteins, enabling them to be functional.[263] The transgene is transfected into HEK293T cells using a chemical agent (X-tremeGENE HP), comprising nonliposomal cationic lipids. We estimated the transfection efficiency of the plasmids encoding UBS54-SNAP, ID405-SNAP, h-STL002-SNAP by quantitating the percentage of eGFP-expressing cells, using flow cytometry in the presence and absence of live–dead cell discrimination dye 7-AAD. Initially, we obtained 50% ±5% transfection efficiency in the absence of viable–dead cell discrimination dye. However, in the presence of viable–dead cell discrimination dye, the percentage of live eGFP cells is in the range of 5%–10% ±2. This agrees with the report of Meisel et al. and Sork et al., who showed a transfection efficiency of 10% and below in HEK293T, when using X-tremeGene HP.[264,265] Later we tried to enhance the transfection efficiency by modifying the DNA–transfection reagent ratio from 1:3 to 2:3 and we achieved a maximum transfection efficiency of 26%. Any further increase in the ratio of transfection agent and nucleic acid did not improve the transfection rate. Transfected cells were found to die within 1 or 2 weeks. There could be two possible explanations for such cell death. First, the increased concentration of transfection agent might be toxic to cells.[266] Second, based on the study by Mori et al., the high concentration of plasmid DNA increases the high copy number in transfected cells. Consequently, the protein expression levels of the transgene are increased. The sustained overexpression of eGFP in the transfected cell will deplete the cellular resources available for native protein production, which would cause ER stress–induced cell death.[267]

In future, we would like to do two things to enhance the transfection rate and viability: first, we would like to switch from the adherent HEK293T cell line to a suspension HEK293 cell line, as suspension HEK293 cells with early passage number show greater transfection efficiency (70-96%), as reported by Molinas et al.[268] Second, we would like to opt for electroporation-mediated transfection rather than chemical-mediated transfection. Electroporation-mediated transfection in HEK293 has been reported to provide more than 50% transfection efficiency, as described by Sherba et al.[269]

After transfection of all constructs with transfection efficiencies ranging from 5% to 26%, we were able to enrich the population of UBS54-SNAP, ID405-SNAP, and h-STL002-SNAP-transfected (recombinant gene) cells to more than 70% of GFP-expressing cells when maintained under antibiotic selection for 8–12 weeks. Unfortunately, the transfected cells that expressed OCAb9-1-SNAP, YW353-SNAP, h5-Thy1-SNAP unable to survive after 3–4 weeks of maintenance. The population of eGFP-positive cells depleted after few weeks of culturing. Although we don't know the exact cause of this transfected cell death, we can speculate from various literature reports that protein aggregation, misfolding, and mistargeting of proteins from the Golgi apparatus to the vacuole/lysosome, where they are proteolytically degraded, could be the possible cause for the deleterious effect of the transgene in HEK cells. Even single amino-acid substitutions can increase the aggregation propensity of immunoglobulin proteins by affecting folding stability.[270,271] Therefore, the possible solution could be switching the expression host from mammalian to other eukaryotic systems such as the yeast *Pichia pastoris* or prokaryotic systems such as *Escherichia coli* and *Brevibacillus choshinensis*.[165,272]

At the stage of CCSN harvesting, cells transfected with UBS54-SNAP, ID405-SNAP, and h-STL002-SNAP had more than 80% of high-intensity eGFP fluorescence, appeared to mostly retain their natural spindle-shaped morphology, and even attained the high multilayered density characteristic of parental HEK293T cultures. For the other constructs (OCAb9-1-SNAP, YW353-SNAP, and h5-Thy1-SNAP), we could not find any eGFP-expressing cells after a few weeks of maintenance in Zeocin.™ This could be due to various factors such as protein aggregation, degradation, and toxicity due to misfolded protein expression.

Harvested CFS was stored in 4°C until the total collected volume ranged from 250 mL to 500 mL, to isolate scFv-SNAP, which is extracellularly secreted due to the presence of Ig Kappa leader sequence. In addition, all the scFv-SNAP used in this study contained a synthetic poly(6x)-

His tag sequence, which provides enriched negative charge to the N-terminal end of the recombinant protein, thereby allowing it to bind with high affinity to a positively charged metal ion (e.g., nickel)–containing column. Ni^{2+} Sepharose columns are commonly employed for immobilized metal-affinity chromatography. While most endogenous proteins will either bind with low affinity or not at all in near-neutral buffer conditions at low imidazole salt concentrations (25 mM), the His-tag-containing protein bound to the column is eluted by raising imidazole concentration to at least 200 mM, in low pH (e.g., 0.1 M glycine hydrochloride, pH 2.5) buffer.

In our study, we found bovine serum albumin (BSA) at 66 kDa in our eluted fraction along with the His-tag recombinant protein (~50 kDa). Literature claims that BSA also has multiple histidines and can bind to IMAC supports in the absence or low yield of His-tagged proteins in the sample. We obtained a very low yield (i.e., the absolute amount of total protein was ~5 mg for 500 mL of CFS) measured by photometric method (UV-Visible spectrophotometer) and by biochemical assay (BCA and Bradford assay), this total protein quantification method further allowed to check (or) to detect the interference of buffer (salt, detergents, surfactants) and other organic compounds in quantitating the total protein concentration. We found there is no interference of any of this reagents/salts with total protein quantification that are used during protein purification and concentration procedure, this further emphasize that the measured concentration is absolute amount of total protein i.e ~5 mg for 500 mL of CFS. whereas Jäger et al. reported the yield of 20 different scFv-Fc antibodies in the range of 400 mg/L–600 mg/L, which is 10-fold higher than what we have obtained; the low yield in our case is most likely the result of low cell density, viability, and reduced biomass production in HEK293T adherent cells.[273] To further quantitate the expressed scFv-SNAP, and truncated protein contaminants, we performed densitometry for the IMAC purified proteins and determined the absolute amount of protein of interest. On an average, we obtained yields of ~3 mg/L for all the scFv-SNAP constructs. In the future, we would like to opt for suspension HEK293 culture or CHO cell line, or PER.C6 (Crucell, Leiden, The Netherlands) cell line in bioreactors to enhance the protein yield to a range of 5 g/L–12 g/L. Parallelly, to further reduce BSA elution, we would like to implement additional steps to narrow down the purification by incorporating ion-exchange and/or size exclusion chromatography to isolate suitably pure fractions of scFv-SNAP or culture-adapt HEK293 cells in serum-free media. Furthermore, we speculate that N-terminal truncated protein (scFv), which elutes along with the full-length protein (scFv-SNAP) and BSA, could further affect the purity of full-length protein scFv-SNAP, thereby interfering in downstream analyses.

In addition, a protein band of ~35 kDa was detected by anti-His-tag immunoblotting. The band was detected in all the SNAP-fusion constructs that were expressed successfully in HEK293T. We were not able to remove this 35 kDa nonspecific protein, using 50 kD molecular weight cutoff (MWCO) centrifugal filters, which theoretically should exclude proteins below 50 kDa in desalted samples. We hypothesized that this N-terminal truncation could be caused due to ribosome stalling in the middle of ORF and not by protein degradation. To study this hypothesis, as a proof of concept we chose UBS54-SNAP and we designed four different constructs (Kz 1 del, Kz 2 del, Kz1&2 del, and Kz low score) from parental pCB-scFv-UBS54-SNAP to avoid ribosome stalling or collision in Kozak consensus region within the ORF. We anticipated that this would generate a single full-length protein translated from a single transcript. We have completed in silico and molecular cloning for these Kozak constructs; however, at the time of reporting this study, due to low transfection efficiency, we were unable to harvest CFS of this Kozak-modified UBS54-SNAP construct.

Finally, we conclude that densitometry-based quantification of a protein band of interest from purified and concentrated protein is far better than other biochemical assays like Bradford and BCA assays for estimating total protein, and ELISA for estimating full-length proteins from CFS. Low scFv-SNAP concentrations in combination with high contaminant (BSA and truncated protein) concentrations renders protein quantification by biochemical analytical assays such as those mentioned above unreliable.

4.4. FUNCTIONAL ACTIVITY OF FUSION PROTEINS IN ANTIGEN-POSITIVE CANCERS

On estimating the absolute amount of full-length protein by densitometry, we sought to validate the functional activity of SNAP-tag (UBS54-SNAP, ID405-SNAP, h-STL002-SNAP) by conjugating with BG modified fluorophore, or BG-containing substrates that were commercially available (https://international.neb.com/products/cellular-analysis/snap-tag-substrates). Under physiological conditions in a one-step reaction, the self-labeling efficiency of the scFv-SNAP-tag fusion protein with BG-modified fluorophore (BG-Alexa 488) was analyzed photometrically. We obtained more than ~90% labeling efficiency after 2 h labeling reaction at room temperature. In addition, we did not find any difference in labeling efficiency when BG-modified fluorophore is incubated with scFv-SNAP both at 37°C and at room temperature (26-30°C). Moreover, we found 4:1 (scFv-SNAP:BG-488) ratio is enough to obtain 90% labeling efficiency, instead of 1:1 or 1:2 ratio described by the manufacturer. One possible explanation is the concentration of full-

length protein from the total protein might be too low, deduced based on densitometry could be one of the reasons for using 4:1 ratio of scFv:BG-Alexa488 in our experiment. In addition, the conjugation efficiency might depend on incubation time, the longer the incubation the more saturated the reaction. Interestingly, we found protein bands above and below the required molecular size, which indicates that there may be some aggregated full-length protein above 100 kDa and unconjugated Alexa fluorophores. Using WB, we were unable to confirm the presence of the aggregated protein of size more than 100 kDa, as high molecular weight proteins are often resistant to their transfer from SDS gel to PVDF membrane.[274] However, the labeling with BG-modified fluorophore lends clarity to the nature of the high molecular weight polypeptide, that is, aggregated SNAP-fusion protein. This further supports our claim that a second protein purification step such as size-exclusion chromatography is essential to abrogate this contaminant protein, both truncated and aggregated forms that might affect the effector function of scFv.

After successful conjugation of the fusion proteins (scFvUBS54-SNAP, ID405-SNAP, and scFvh-STL002-SNAP) to Alexa 488 fluorophore, binding to cells expressing their cognate antigens (EpCAM, EMA, and CD138) was assessed using confocal microscopy.

We found strong cell surface membrane labeling of EpCAM in MDA-M-468, but not in HS578T. Similarly, uniform staining of EpCAM was seen in the majority of MCF-7 cells, but not in MDA-MB-231 (**Fig. 71 and 72**). The positive and negative staining of EpCAM in the aforementioned cell lines is in agreement with the findings reported by Amoury et al.[275] Similarly, scFvID405-SNAP bound to EMA-positive cell line MCF-7 and failed to bind to EMA-negative HEK293T cell line. When we compared the distribution of EpCAM and EMA in MCF-7, EpCAM showed uniform binding covering the entire membrane of the MCF-7 cell, whereas the EMA stain was intense in certain regions of the cell membrane and very weak in other regions, which is contradictory to the findings of Kufe et al., Hilkens et al., and Cascio et al., who stated that EMA, also known as MUC1, expression was observed in apical membrane surface, whereas in cancer it showed a diffuse pattern with loss of apical polarity.[276,277,278] Interestingly, Truant et al. reported that HT-29 5M21 cells, a highly mucin-secreting clone, and a derivative of colon cancer cell line HT-29, showed apical expression of MUC1 in standard culture condition.[269] In contrast, HT-29 parental cells showed MUC1 distribution throughout the surface of the cell in standard culture conditions. However, when HT-29 5M21 was cultured in the presence of collagen, the apical expression of MUC1 was lost and mucins were redistributed over the entire cell surface. This clearly indicates that there is a difference in localization of MUC1 expression between noninvasive tumors and invasive tumors. A further thorough investigation is needed to precisely

state the location of MUC1 in MCF-7 under the standard conditions as our results are contradictory to the findings of Kufe et al. and Hopkinson et al.[276,280] We could not ascertain whether the antibody recognition site might cause this difference. Pemberton et al. showed that MUC1 can exist in two discrete motifs: one at the junction of the transmembrane and cytoplasmic domains that are necessary for surface expression and the other in the extracellular domain (outside of the tandem repeat region) involved in apical localization.[281] Our antibody scFvID405 might recognize the extracellular domain, rather than the motif between transmembrane and cytoplasmic domains, which could cause this difference in the distribution of MUC1.

After studying the functional activity of scFv-SNAP, which targets the well-known primary marker EpCAM and EMA, we investigated the functional activity of scFvh-STL002-SNAP in CD138 a positive cell line. CD138 was reported to be expressed in low amounts or absent in BCC.[117] We have shown the functional activity of scFvh-STL002-SNAP in CD138-positive HEK293T, which matched the findings of Orecchia et al..[282] Unfortunately, we were unable to obtain a CD138-negative cell line by the time of the concluding our study. We have thus not shown a lack of binding of scFvh-STL002-SNAP in a CD138-negative cell line. Since we were interested in novel biomarkers for BCC, we hypothesized and chose to target CSPG4; the antibody fusion protein mAb9.2.27 scFv-SNAP was kindly provided by Siyabulela Magugu, a former MSc student of MB&I. We found that activity of this antibody fusion protein in HS578T cell line matched to that of Amoury et al.[244] Based on our confocal microscopy studies, there is clear evidence that the antibody fusion protein generated to target EpCAM, EMA, CD18, and CSPG4 were functional, as they recognized the target antigen expressed on the surface of cells of various cell lines.

The intensity of surface immunostaining depends on the quality, quantity, and specificity of the antibodies, target antigen expression level, internalization rate of the receptor, fluorophore brightness, and photobleaching effect. In our study, we used Alexa 488 that emits a green signal that is less bright when compared to the stronger red signal emitted by Alexa 647, for the reason that the former does not require expensive or sophisticated instrumentation settings and the green channel is ubiquitous across instruments. Alexa 488 shows more persistent and brighter fluorescence than FITC.

Once the functional activity of scFv-SNAP was determined, the next step was to estimate the total population of cells that were positive for the antigen, via flow cytometry. Quantification of the total population of tumor cells that are positive for antigen expression would allow for patient-specific tailored therapy. Generally, tumors with low expression of target antigens show

poor response to antibody–drug conjugate compared to tumors with moderate to strong expression of antigen.[283] Ideally, to determine the applicability of such a method, one would need to screen the expression of EpCAM, EMA, CD138, and CSPG4 in a BCC-positive cell line, However, to establish the technique, the antigen expression level of EpCAM and EMA is determined in other solid tumors (in which it was originally reported to express the respective TAA). We were able to show that 55% of live PANC-1 were positive for EpCAM, matching the findings of Pozza et al., who had shown that 41% of PANC-1 are positive for EpCAM.[243] Similarly, Amoury et al. reported that CSPG4 is moderately expressed in HS578T and absent in MDA-MB468, which corelates with our result.[244] However, the percentage of CSPG4-positive population in HS578T is low and this could be attributed to difference in passage number, purity of fusion protein, and concentration of fusion protein.

To find out whether the concentration of fusion protein would have any profound effect on determining the percentage of positive population, we used the EMA-targeting fusion protein "scFvID405-SNAP."

scFvID405-SNAP-labeling could detect 29% viable EMA-positive cell population in MCF-7 cell line, but not in control cells (HEK293T).The low expression of EMA in MCF-7 correlates with the findings of Solatycka et al. and Kim et al.[245,246] A further increase in the concentration of antibody (ID405-SNAP) to estimate the positive population in MCF-7 did not show any changes. This indicates that the concentration that we employed is enough to saturate the target receptors on the cell surface. Pan et al. have reported that MCF-7 and SKBR3 (breast cancer cell line) showed nearly 90% EMA expression when targeted with 16A mAb that recognized the glycosylated tandem repeat region of MUC1/EMA.[284] This indicates that the epitope recognized by the antibody can influence the expression signature, particularly when one targets a glycosylated antigen. It would be interesting to know whether the 16A mAb that recognizes glycosylated region in EMA exhibits a similar negative expression signature in BCC, when stained with mAb E29.[285] The latter is commonly used in IHC procedure to detect EMA in tissue sections and the antibody ID405-SNAP, which we used in our study, is known to recognize EMA, but we dont know the specific epitope region that the antibody binds to.

Once we prove the binding functionality of scFv-SNAP, we intend to see the cytotoxic effect of scFv-SNAP drug conjugates in a EpCAM-positive cell line.

4.5. CYTOTOXICITY OF TARGETED scFv DRUG CONJUGATES IN EpCAM-POSITIVE CANCER CELLS

Investigation of scFv-SNAP-AF cytotoxicity was used as a proof of concept, providing an alternative treatment option for patients who cannot undergo surgical excision of primary BCC tumor and who have developed resistance to vesmodigib small-molecular inhibitors. The BG-modified AF is conjugated with scFv-SNAP-tag to derive small format drug conjugates "scFv-SNAP-BG-AF" which have better tumor-penetrating properties, micro-distribution as well as more rapid pharmacokinetics (PK), systemic clearance in in vivo model compared with free AF.[184]

Here, we showed the in vitro cytotoxic effects of UBS54scFv-SNAP fusion proteins and unconjugated BG-AURIF (BG-AF) using an XTT based colorimetric cell proliferation assay. Interestingly, MDA-MB-468 cells (Fig. 37, right) are twofold more susceptible to BG-AF compared with MCF-7 cells; this could be due, in part, to increased expression and activity of drug efflux protein MDR1 in MCF-7 compared to MDA-MB-468. MDR1 is a drug transporter known to efflux MMAE out of cells, leading to drug resistance against monomethyl auristatin E (MMAE).[286] Monomethyl auristatin F (MMAF) is a derivative of monomethyl auristatin E (MMAE); they both comprise five amino acids, namely, norephedrine, dolaproine, dolaisoleuine, valine, and monomethyl valine. In MMAF, the C-terminal norephedrine is replaced by phenylalanine, contributing to its membrane impermeability.[287] Hence, we reasoned that the resistance to MMAF developed by MCF-7 could be due to increased MDR1 expression compared to MDA-MB-48. However, further study is warranted to prove our notion that MDR1 can efflux MMAF as equally as MMAE despite differences in chemical composition and cell permeability.

Quantitative determination of cell viability indicated that relatively more drug of UBS54scFv-SNAP-AF is required to kill MDA-MB-468 cells compared to MCF-7 (**Fig. 75**, **Table 21**); this could be due to low expression of EpCAM in MDA-MB-468 compared to MCF-7, which has been reported by others.[244,288] Surprisingly, MCF-7 that are reported to express increased MDR1 are more susceptible to ADC (UBS54scFv-SNAP-BG-MMAF) compared to MDA-MB-468. We do not know the exact reason for these outcomes; however, we speculate from study reported by Loganzo et al. and Kovtun et al. that ADCs containing cleavable mcValCitPABC-linked auristatins and anti-EpCAM–DM1 conjugate with PEG4Mal linker have the ability to bypass multidrug resistance compared to non–cleavable-linked ADCs such as TM-ADC and anti-EpCAM–SMCC–DM1.[289,290] In our study the UBS54scFv-SNAP-BG-MMAF has a protease cleavable site (EKS); thus, we assume it might be involved in mitigating MDR1-mediated resistance in MCF-7.

The generated MMAF based ADC showed selective cytotoxic activity in EpCAM expressing cancer cells, these might allow application in BCC patients, who are not eligible for surgical excision such as invasive/metastatic BCC patient, and patient whose tumor is located in surgically critical area such as eyelid, lips etc as well as BCC patients resistant to Vismodgib

The proof-of-concept data provided in this thesis shows that the antibody SNAP-fusion protein designed and synthesized by us is functional as it recognizes the target antigen and specifically induces cell death whenever the scFv-SNAP fusion protein is conjugated with the cytotoxic drug. Despite protein fragmentation, the efficient cell binding via the scFv domain is not compromised; however, the cellular toxicity is compromised (data not shown) and therefore a further purification step is essential to remove the truncated protein. Thus, we describe for the first time the successful generation of scFv-UBS54-SNAP that has potential applications in molecular diagnosis and targeted therapy of BCC.

CHAPTER 5: CONCLUSIONS AND FUTURE DIRECTIONS

We present here a reanalysis of a comprehensive collection of microarray datasets and literature search for discovery of BCC cell surface antigens. Parallelly, the synthesis of functional targeting moiety for the selected cell surface antigens, namely, EpCAM, EMA, CSPG4, further strengthens the use of biomarkers for targeted diagnosis, targeted surgical excision (therapy) for noninvasive BCC and systemic targeted therapy (ADC) for invasive BCC. Some of our findings such as CSPG4 expression in BCC and possible truncation of scFv-SNAP that might be attributed to the likelihood of a strong Kozak sequence are unreported. Interestingly, despite minimal protein fragmentation, the efficient cell binding via the scFv domain is not compromised. However, the construct with a modified/deleted Kozak sequence will need further structural and functional validation to prove the integrity of the protein. Despite such limitations, the study can serve as a valuable resource and a proof-of-concept study for future research to accurately identify BCC lesions and achieve complete excision of the lesions in a single procedure. The use of the ScFv-SNAP antibody-based diagnostics may enable tailored therapeutic decisions for patients with advanced BCC or inoperable BCC (for instance, locally advanced or metastatic BCC, resistant to several Sonic Hedgehog inhibitors) who might exhibit differential overexpression of the EpCAM/CSPG4 antigen on the tumor cell surface. This thesis has two major limitations: first, the expression of cell surface antigens EpCAM, CSPG, and EMA has not been delineated in BCC tissue samples (both invasive and noninvasive). Nevertheless, EpCAM expression in BCC is well documented by various authors and is used as a standard antibody-based diagnosis both in research and clinical studies. Second, we ought to improve the purity of scFv-SNAP protein by removing the truncated scFv-SNAP to obtain the best possible dose–response curve when targeted with an ADC.

In conclusion, we have performed a comprehensive bioinformatics analysis of DEGs obtained from 21 BCC and 80 normal skin tissues. Six genes have been identified to play a critical role in BCC pathogenesis. Targeting these protein-encoding genes by employing scFv-fusion proteins has further laid the foundation for exploring effective molecular targets for the diagnosis and treatment of BCC. The efficient transient expression of the scFv-fusion proteins in the HEK293T cell line currently serves as a valuable niche to produce biotherapeutics. However, not all fusion proteins were produced successfully in the mammalian expression system used in this study. Of the six fusion proteins, four were expressed in sufficient quantities as functional proteins.

The production of recombinant proteins in HEK293T cells is influenced by complex molecular events occurring at the level of transcription, posttranscriptional processing, translation, posttranslational processing, and secretion. Each of these steps represents a potential challenge to achieve the desired level of protein expression. The secondary structure of the plasmid DNA, for example, can limit the level of transcription initiation, while misfolding within the ER can shut down the production of the protein. Alternative hosts such as yeasts and prokaryotic organisms can be used for the expression of other scFv-fusion proteins in near future. The quantity and purity of the fusion proteins can be further enhanced with the use of the latest suspension culture and serum-free methods. To conclude, the genes identified from our BCC gene expression analysis could be tested in routine biopsy samples obtained from patients in the clinic. A larger validation cohort is likely needed to further characterize such expression and confirm the utility of these biomarkers. Also, our results have further laid the foundation for exploring effective molecular targets for the diagnosis and treatment of BCC. However, future preclinical experiments are required to confirm these findings in one or more animal models and patient samples. The SNAP-tag technology that we have established in this study has the potential to be applied for BCC diagnosis, image-guided surgery, and development of therapeutics (possibly if imaging payloads are replaced with cytotoxic payloads, e.g., chemotherapeutic drugs, photosensitizers, radionuclides, or immunotoxins).

CHAPTER 6: REFERENCES

1. Lanigan S., Zaidi Z. Dermatology in Clinical Practice. *Dermatology Clin Pract*. Published online 2010:1-15. doi:10.1007/978-1-84882-862-9
2. Amsden BG, Goosen MFA. Transdermal delivery of peptide and protein drugs: An overview. *AIChE J*. 1995;41(8):1972-1997. doi:10.1002/aic.690410814
3. Harding CR. The stratum corneum: Structure and function in health and disease. *Dermatol Ther*. 2004;17(1):6-15. doi:10.1111/j.1396-0296.2004.04s1001.x
4. Archer CB. Functions of the skin. In: Burns DA, Breathnach SM, Cox NH, Griffiths CEM. Eds: Rook's Textbook of Dermatology, 8th edition, Blackwells (Oxford) 2010; pp 4.1-4.11. In: ; 2010.
5. Butnaru C-A, Kanitakis J. Structure of normal human skin. *Eur J Dermatol*. 2002;12(6):II-IV.
6. Lopez-Ojeda W, Oakley AM. *Anatomy, Skin (Integument)*.; 2018. http://www.ncbi.nlm.nih.gov/pubmed/28723009
7. Lattimer C, Faiz O. The skin and subcutaneous tissues. *Key Top Gen Surg*. Published online 2002:269-272. doi:10.1201/b17362-4
8. James G. Marks J and JJM. *Lookingbill and Marks' Principles of Dermatology,(Sixth Edition),*.; 2019. doi:10.1016/c2015-0-00881-4
9. Marks JG, Miller JJ. 2 - Structure and Function of the Skin. In: Marks JG, Miller JJBT-L and MP of D (Fifth E, eds. W.B. Saunders; 2013:2-10. doi:https://doi.org/10.1016/B978-1-4557-2875-6.00002-9
10. Barile P, Leroy C, Bolle S, et al. Merkel cell carcinoma. *J Fr Ophtalmol*. 2004;27(4):432-436. doi:10.1016/s0181-5512(04)96159-1
11. Allen CE, Merad M, McClain KL. Langerhans-Cell Histiocytosis. Longo DL, ed. *N Engl J Med*. 2018;379(9):856-868. doi:10.1056/NEJMra1607548
12. Yi W, Chen WY, Yang TX, Lan JP, Liang WN. Langerhans cell sarcoma arising from antecedent langerhans cell histiocytosis: A case report. *Medicine (Baltimore)*. 2019;98(10):e14531. doi:10.1097/MD.0000000000014531
13. Kuroda K, Tajima S. Proliferation of HSP47-positive skin fibroblasts in dermatofibroma. *J Cutan Pathol*. 2008;35(1):21-26. doi:10.1111/j.1600-0560.2007.00768.x
14. Leiter U, Eigentler T, Garbe C. Epidemiology of skin cancer. In: *Advances in Experimental Medicine and Biology*. Vol 810. Springer New York; 2014:120-140. doi:10.1016/b978-1-4377-1788-4.00005-8
15. Linos E, Swetter SM, Cockburn MG, Colditz A, Clarke CA. Increasing Burden of Melanoma. *J Invest Dermatol*. 2010;129(7):1666-1674. doi:10.1038/jid.2008.423.Increasing
16. Perera E, Gnaneswaran N, Perera M, Sinclair R. Validating the use of Medicare Australia billing data to examine trends in skin cancer. *F1000Research*. 2015;4(0):1-9. doi:10.12688/f1000research.7161.1
17. 2014 Rastrelli. Melanoma Risk Factor. In: *Definitions*. Vol 1012. Qeios; 2020:1005-1011. doi:10.32388/7XJ0GW
18. Cai W. Anatomical and molecular imaging of skin cancer. *Clin Cosmet Investig Dermatol*. Published online October 2008:1. doi:10.2147/CCID.S4249
19. Eide MJ, Krajenta R, Johnson D, et al. Identification of Patients With Nonmelanoma Skin Cancer Using Health Maintenance Organization Claims Data. *Am J Epidemiol*. 2010;171(1):123-128. doi:10.1093/aje/kwp352
20. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394-424. doi:10.3322/caac.21492
21. Lara F, Garbers LEF de M, Santamaría JR. Recurrence rate of basal cell carcinoma with positive histopathological margins and related risk factors. *An Bras Dermatol*. 2017;92(1):58-62. doi:10.1590/abd1806-4841.20174867
22. Matos I, Machado M, Semedo C, Santos J, Sousa S. A rare case of metastatic basal cell carcinoma. *Int J Oral Maxillofac Surg*. 2019;48:219. doi:10.1016/j.ijom.2019.03.676
23. Codazzi D, Van Der Velden J, Carminati M, et al. Positive compared with negative margins in a single-centre retrospective study on 3957 consecutive excisions of basal cell carcinomas. Associated risk factors and preferred surgical management. *J Plast Surg Hand Surg*. 2014;48(1):38-43. doi:10.3109/2000656X.2013.800526
24. Verkouteren JAC, Ramdas KHR, Wakkee M, Nijsten T. Epidemiology of basal cell carcinoma: scholarly review. *Br J Dermatol*. 2017;177(2):359-372. doi:10.1111/bjd.15321
25. Apalla Z, Lallas A, Sotiriou E, Lazaridou E, Ioannides D. Epidemiological trends in skin cancer. *Dermatol Pract Concept*. 2017;7(2):1-6. doi:10.5826/dpc.0702a01
26. McDaniel B, Badri T, Steele RB. *Basal Cell Carcinoma*.; 2021.

http://www.ncbi.nlm.nih.gov/pubmed/29494046
27. Saldanha G, Fletcher A, Slater DN. Basal cell carcinoma: A dermatopathological and molecular biological update. *Br J Dermatol*. 2003;148(2):195-202. doi:10.1046/j.1365-2133.2003.05151.x
28. Stockfleth E, Sterry W. New Treatment Modalities for Basal Cell Carcinoma. In: *Recent Results in Cancer Research. Fortschritte Der Krebsforschung. Progres Dans Les Recherches Sur Le Cancer*. Vol 160. ; 2002:259-268. doi:10.1007/978-3-642-59410-6_31
29. Ponticelli C, Cucchiari D, Bencini P. Skin cancer in kidney transplant recipients. *J Nephrol*. 2014;27(4):385-394. doi:10.1007/s40620-014-0098-4
30. Venables ZC, Autier P, Nijsten T, et al. Nationwide Incidence of Metastatic Cutaneous Squamous Cell Carcinoma in England. *JAMA Dermatology*. 2019;155(3):298-306. doi:10.1001/jamadermatol.2018.4219
31. Cox NH, Eedy DJ, Morton CA. Guidelines for management of Bowen's disease: 2006 Update. *Br J Dermatol*. 2007;156(1):11-21. doi:10.1111/j.1365-2133.2006.07610.x
32. Elder DE, Massi D, Scolyer RA, Willemze R. WHO Classification of skin tumors, 4th edition. In: Lyon:IARC; 2018.
33. Que SKT, Zwald FO, Schmults CD. Cutaneous squamous cell carcinoma: Incidence, risk factors, diagnosis, and staging. *J Am Acad Dermatol*. 2018;78(2):237-247. doi:10.1016/j.jaad.2017.08.059
34. Khan K, Mykula R, Kerstein R, et al. A 5-year follow-up study of 633 cutaneous SCC excisions: Rates of local recurrence and lymph node metastasis. *J Plast Reconstr Aesthetic Surg*. 2018;71(8):1153-1158. doi:10.1016/j.bjps.2018.03.019
35. Wang GY, Wang J, Mancianti ML, Epstein EH. Basal cell carcinomas arise from hair follicle stem cells in Ptch1+/- mice. *Cancer Cell*. 2011;19(1):114-124. doi:10.1016/j.ccr.2010.11.007
36. Peterson SC, Eberl M, Vagnozzi AN, et al. Basal cell carcinoma preferentially arises from stem cells within hair follicle and mechanosensory niches. *Cell Stem Cell*. 2015;16(4):400-412. doi:10.1016/j.stem.2015.02.006
37. Youssef KK, Van Keymeulen A, Lapouge G, et al. Identification of the cell lineage at the origin of basal cell carcinoma. *Nat Cell Biol*. 2010;12(3):299-305. doi:10.1038/ncb2031
38. Tan ST, Ghaznawie M, Heenan PJ, Dosan R. Basal cell carcinoma arises from interfollicular layer of epidermis. *J Oncol*. 2018;2018. doi:10.1155/2018/3098940
39. Sánchez-Danés A, Hannezo E, Larsimont JC, et al. Defining the clonal dynamics leading to mouse skin tumour initiation. *Nature*. 2016;536(7616):298-303. doi:10.1038/nature19069
40. Zafiropoulos A, Tsentelierou E, Billiri K, Spandidos DA. Human herpes viruses in non-melanoma skin cancers. *Cancer Lett*. 2003;198(1):77-81. doi:10.1016/S0304-3835(03)00269-6
41. Zaravinos A, Kanellou P, Spandidos DA. Viral DNA detection and RAS mutations in actinic keratosis and nonmelanoma skin cancers. *Br J Dermatol*. 2010;162(2):325-331. doi:10.1111/j.1365-2133.2009.09480.x
42. Didona D, Paolino G, Bottoni U, Cantisani C. Non melanoma skin cancer pathogenesis overview. *Biomedicines*. 2018;6(1):1-15. doi:10.3390/biomedicines6010006
43. Nikolaou 2012. Hereditary Nonmelanoma Skin Cancer. *Bone*. 2011;72(2):132-135. doi:10.1016/j.sder.2012.08.005.Hereditary
44. Schierbeck J, Vestergaard T, Bygum A. Skin cancer associated genodermatoses: A literature review. *Acta Derm Venereol*. 2019;99(4):360-369. doi:10.2340/00015555-3123
45. Furudate S, Fujimura T, Tojo GI, Haga T, Aiba S. Basal cell carcinoma arising from xeroderma pigmentosum: A case report and an immunohistochemical study. *Case Rep Dermatol*. 2013;5(1):64-68. doi:10.1159/000350182
46. Cho HG, Kuo KY, Li S, et al. Frequent basal cell cancer development is a clinical marker for inherited cancer susceptibility. *JCI Insight*. 2018;3(15). doi:10.1172/jci.insight.122744
47. Berlin NL, Cartmel B, Leffell DJ, Bale AE, Mayne ST, Ferrucci LM. Family history of skin cancer is associated with early-onset basal cell carcinoma independent of MC1R genotype. *Cancer Epidemiol*. 2015;39(6):1078-1083. doi:10.1016/j.canep.2015.09.005
48. Pellegrini C, Maturo MG, Di Nardo L, Ciciarelli V, Gutiérrez García-Rodrigo C, Fargnoli MC. Understanding the molecular genetics of basal cell carcinoma. *Int J Mol Sci*. 2017;18(11). doi:10.3390/ijms18112485
49. Madan V, Hoban P, Strange RC, Fryer AA, Lear JT. Genetics and risk factors for basal cell carcinoma. *Br J Dermatol*. 2006;154 Suppl:5-7. doi:10.1111/j.1365-2133.2006.07229.x
50. Couvé-Privat S, Le Bret M, Traiffort E, et al. Functional analysis of novel sonic hedgehog gene mutations identified in basal cell carcinomas from xeroderma pigmentosum patients. *Cancer Res*. 2004;64(10):3559-3565. doi:10.1158/0008-5472.CAN-03-4040
51. De Zwaan SE, Haass NK. Genetics of basal cell carcinoma. *Australas J Dermatol*. 2010;51(2):81-92. doi:10.1111/j.1440-0960.2009.00579.x
52. Atwood SX, Sarin KY, Whitson RJ, et al. Smoothened Variants Explain the Majority of Drug Resistance in Basal Cell Carcinoma. *Cancer Cell*. 2015;27(3):342-353. doi:10.1016/j.ccell.2015.02.002

53. Sharpe HJ, Pau G, Dijkgraaf GJ, et al. Genomic Analysis of Smoothened Inhibitor Resistance in Basal Cell Carcinoma. *Cancer Cell*. 2015;27(3):327-341. doi:10.1016/j.ccell.2015.02.001
54. Kobayashi H, Choyke PL. Near-Infrared Photoimmunotherapy of Cancer. *Acc Chem Res*. 2019;52(8):2332-2339. doi:10.1021/acs.accounts.9b00273
55. Maturo MG, Rachakonda S, Heidenreich B, et al. Coding and noncoding somatic mutations in basal cell carcinoma. *bioRxiv*. Published online 2019. doi:10.1101/807313
56. Rippey JJ. Why classify basal cell carcinomas? *Histopathology*. 1998;32(5):393-398. doi:10.1046/j.1365-2559.1998.00431.x
57. Mackiewicz-Wysocka M, Bowszyc-Dmochowska M, Strzelecka-Weklar D, Dańczak-Pazdrowska A, Adamski Z. Basal cell carcinoma - Diagnosis. *Wspolczesna Onkol*. 2013;17(4):337-342. doi:10.5114/wo.2013.35684
58. Crowson AN. Basal cell carcinoma: Biology, morphology and clinical implications. *Mod Pathol*. 2006;19(SUPPL. 2). doi:10.1038/modpathol.3800512
59. 2018 Elder Ebook WHO SKIN CANCER.pdf.
60. Scrivener Y, Grosshans E, Cribier B. Variations of basal cell carcinomas according to gender, age, location and histopathological subtype. *Br J Dermatol*. 2002;147(1):41-47. doi:10.1046/j.1365-2133.2002.04804.x
61. De Faria JL. Basal cell carcinoma of the skin with areas of squamous cell carcinoma: A basosquamous cell carcinoma? *J Clin Pathol*. 1985;38(11):1273-1277. doi:10.1136/jcp.38.11.1273
62. Alturkistani HA, Tashkandi FM, Mohammedsaleh ZM. Histological Stains: A Literature Review and Case Study. *Glob J Health Sci*. 2015;8(3):72-79. doi:10.5539/gjhs.v8n3p72
63. Rossi AM, Sierra H, Rajadhyaksha M, Nehal K. Novel approaches to imaging basal cell carcinoma. *Futur Oncol*. 2015;11(22):3039-3046. doi:10.2217/fon.15.231
64. Fischer AH, Jacobson KA, Rose J, Zeller R. Hematoxylin and eosin staining of tissueand cell sections. *Cold Spring Harb Protoc*. 2008;3(5):4986-4988. doi:10.1101/pdb.prot4986
65. Zimmermann A. Tumors and Tumor-Like Lesions of the Hepatobiliary Tract. *Tumors and Tumor-Like Lesions of the Hepatobiliary Tract*. Published online 2017. doi:10.1007/978-3-319-26956-6
66. Stanoszek LM, Wang GY, Harms PW. Histologic mimics of basal cell carcinoma. *Arch Pathol Lab Med*. 2017;141(11):1490-1502. doi:10.5858/arpa.2017-0222-RA
67. Alter M, Hillen U, Leiter U, Sachse M, Gutzmer R. Current diagnosis and treatment of basal cell carcinoma. *JDDG J der Dtsch Dermatologischen Gesellschaft*. 2015;13(9):863-875. doi:10.1111/ddg.12798
68. McKenzie CA, Chen AC, Choy B, Fernández-Peñas P, Damian DL, Scolyer RA. Classification of high risk basal cell carcinoma subtypes: experience of the ONTRAC study with proposed definitions and guidelines for pathological reporting. *Pathology*. 2016;48(5):395-397. doi:10.1016/j.pathol.2016.05.005
69. Leibovitch I, Huilgol SC, Selva D, Richards S, Paver R. Basal cell carcinoma treated with Mohs surgery in Australia I. Experience over 10 years. *J Am Acad Dermatol*. 2005;53(3):445-451. doi:10.1016/j.jaad.2005.04.083
70. East E, Fullen DR, Arps D, et al. Morpheaform Basal Cell Carcinomas with Areas of Predominantly Single-Cell Pattern of Infiltration: Diagnostic Utility of p63 and Cytokeratin. *Am J Dermatopathol*. 2016;38(10):744-750. doi:10.1097/DAD.0000000000000541
71. Anand RL, Collins D, Chapman A. Basosquamous carcinoma: appearance and reality. *Oxford Med Case Reports*. 2017;2017(1):omw095. doi:10.1093/omcr/omw095
72. Brankov N, Prodanovic EM, Hurley MY. Pigmented basal cell carcinoma: increased melanin or increased melanocytes? *J Cutan Pathol*. 2016;43(12):1139-1142. doi:10.1111/cup.12819
73. Mc Menamin ME, Goh SGN, Poblet E, Gostelow BE, Robson A, Calonje E. Sarcomatoid Basal Cell Carcinoma—Predilection for Osteosarcomatous Differentiation: A Series of 11 Cases. *Am J Surg Pathol*. 2006;30(10).
74. Tse JY, Pawlak AC, Boussahmain C, et al. Basal cell carcinoma with osteosarcomatous component. *Am J Dermatopathol*. 2013;35(2):261-265. doi:10.1097/DAD.0b013e31826b352b
75. Chu SWW, Biswas A. Basal cell carcinomas showing histological features generally associated with cutaneous adnexal neoplasms. *J Cutan Pathol*. 2015;42(12):1049-1062. doi:10.1111/cup.12577
76. Kato N, Ueno H. Infundibulocystic basal cell carcinoma. *Am J Dermatopathol*. 1993;15(3):265-267. doi:10.1097/00000372-199306000-00013
77. Onishi M, Takahashi K, Maeda F, Akasaka T. A Case of Basal Cell Carcinoma with Outer Hair Follicle Sheath Differentiation. *Case Rep Dermatol*. 2015;7(3):352-357. doi:10.1159/000442704
78. Misago N, Suse T, Uemura T, Narisawa Y. Basal cell carcinoma with sebaceous differentiation. *Am J Dermatopathol*. 2004;26(4):298-303. doi:10.1097/00000372-200408000-00006
79. Pinkus H. Premalignant fibroepithelial tumors of skin. *A M A Arch Dermatology Syphilol*. 1953;67(6):598-615. doi:10.1001/archderm.1953.01540060060009

80. Haddock ES, Cohen PR. Fibroepithelioma of Pinkus Revisited. *Dermatol Ther (Heidelb)*. 2016;6(3):347-362. doi:10.1007/s13555-016-0123-8
81. Dixon AY, Lee SH, McGregor DH. Histologic features predictive of basal cell carcinoma recurrence: results of a multivariate analysis. *J Cutan Pathol*. 1993;20(2):137-142. doi:10.1111/j.1600-0560.1993.tb00230.x
82. Haws AL, Rojano R, Tahan SR, Phung TL. Accuracy of biopsy sampling for subtyping basal cell carcinoma. *J Am Acad Dermatol*. 2012;66(1):106-111. doi:10.1016/j.jaad.2011.02.042
83. Pampena 2019. Broadening the List of Basal Cell Carcinoma Mimickers: Dermoscopic Features of Trichoadenoma. *Dermatol Pract Concept*. 2019;9(2):160-161. doi:10.5826/dpc.0902a17
84. Singh S, Zafar A, Khan S, Naseem I. Towards therapeutic advances in melanoma management: An overview. *Life Sci*. 2017;174:50-58. doi:10.1016/j.lfs.2017.02.011
85. Hasbún Acuña P, Cullen Aravena R, Maturana Donaire C, Ares Mora R, Porras Kusmanic N. Pigmented basal cell carcinoma mimicking a superficial spreading melanoma. *Medwave*. 2016;16(11):e6805. doi:10.5867/medwave.2016.11.6805
86. Jetley S, Jairajpuri ZS, Rana S, Talikoti MA. Adenoid basal cell carcinoma and its mimics. *Indian J Dermatol*. 2013;58(3):244. doi:10.4103/0019-5154.110874
87. Ball NJ, Tanhuanco-Kho G. Merkel cell carcinoma frequently shows histologic features of basal cell carcinoma: A study of 30 cases. *J Cutan Pathol*. 2007;34(8):612-619. doi:10.1111/j.1600-0560.2006.00674.x
88. Wetzel M, Strickley J, Haeberle MT, Brown TS. Depth of invasion of aggressive and nonaggressive basal cell carcinoma. *J Clin Aesthet Dermatol*. 2019;12(3):12-14.
89. Patel SS, Cliff SH, Ward Booth P. Incomplete removal of basal cell carcinoma: What is the value of further surgery? *Oral Maxillofac Surg*. 2013;17(2):115-118. doi:10.1007/s10006-012-0348-3
90. Definition of immunophenotyping - NCI Dictionary of Cancer Terms - National Cancer Institute. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/immunophenotyping
91. Yu L, Galan A, McNiff JM. Caveats in BerEP4 staining to differentiate basal and squamous cell carcinoma. *J Cutan Pathol*. 2009;36(10):1074-1176. doi:10.1111/j.1600-0560.2008.01223.x
92. Villada G, Kryvenko ON, Campuzano-Zuluaga G, Kovacs C, Chapman J, Gomez-Fernandez C. A Limited Immunohistochemical Panel to Distinguish Basal Cell Carcinoma of Cutaneous Origin from Basaloid Squamous Cell Carcinoma of the Head and Neck. *Appl Immunohistochem Mol Morphol*. 2016;26(2):126-131. doi:10.1097/PAI.0000000000000394
93. Beer TW, Shepherd P, Theaker JM. Ber EP4 and epithelial membrane antigen aid distinction of basal cell, squamous cell and basosquamous carcinomas of the skin. *Histopathology*. 2000;37(3):218-223. doi:10.1046/j.1365-2559.2000.00999.x
94. Yaziji H, Barry T. Diagnostic immunohistochemistry: What can go wrong? *Adv Anat Pathol*. 2006;13(5):238-246. doi:10.1097/01.pap.0000213041.39070.2f
95. Kim JH, Song DH, Youn SJ, et al. Crystal structures of mono- and bi-specific diabodies and reduction of their structural flexibility by introduction of disulfide bridges at the Fv interface. *Sci Rep*. 2016;6(June):1-12. doi:10.1038/srep34515
96. Jadotte YT, Sarkissian NA, Kadire H, Lambert WC. CASE REPORT Superficial Spreading Basal Cell Carcinoma of the Face: A Surgical Challenge. *Eplasty*. 2010;10(June 2010):e46.
97. Lassere MN. The biomarker-surrogacy evaluation schema: A review of the biomarker-surrogate literature and a proposal for a criterion-based, quantitative, multidimensional hierarchical levels of evidence schema for evaluating the status of biomarkers as surrogate endp. *Stat Methods Med Res*. 2008;17(3):303-340. doi:10.1177/0962280207082719
98. Organization WH, Safety IP on C. Biomarkers in risk assessment : validity and validation. Published online 2001. https://apps.who.int/iris/handle/10665/42363
99. Goossens N, Nakagawa S, Sun X, Hoshida Y. Cancer biomarker discovery and validation. *Transl Cancer Res*. 2015;4(3):256-269. doi:10.3978/j.issn.2218-676X.2015.06.04
100. Hyams DM, Cook RW, Buzaid AC. Identification of risk in cutaneous melanoma patients: Prognostic and predictive markers. *J Surg Oncol*. 2019;119(2):175-186. doi:10.1002/jso.25319
101. Zhao Q, Piyush T, Chen C, et al. MUC1 extracellular domain confers resistance of epithelial cancer cells to anoikis. *Cell Death Dis*. 2014;5(10):1-10. doi:10.1038/cddis.2014.421
102. Tarhini A, Kudchadkar RR. Predictive and on-treatment monitoring biomarkers in advanced melanoma: Moving toward personalized medicine. *Cancer Treat Rev*. 2018;71(March):8-18. doi:10.1016/j.ctrv.2018.09.005
103. Webb D V., Mentrikoski MJ, Verduin L, Brill LB, Wick MR. Basal cell carcinoma vs basaloid squamous cell carcinoma of the skin: An immunohistochemical reappraisal. *Ann Diagn Pathol*. 2015;19(2):70-75. doi:10.1016/j.anndiagpath.2015.01.004
104. Linskey KR, Gimbel DC, Zukerberg LR, Duncan LM, Sadow PM, Nazarian RM. BerEp4, cytokeratin 14,

and cytokeratin 17 immunohistochemical staining aid in differentiation of basaloid squamous cell carcinoma from basal cell carcinoma with squamous metaplasia. *Arch Pathol Lab Med*. 2013;137(11):1591-1598. doi:10.5858/arpa.2012-0424-OA
105. Gaiser MR, Hirsch D, Gaiser T. Loss of epithelial cell adhesion molecule (EpCAM) in infiltrative basal cell carcinoma. *Int J Clin Exp Pathol*. 2018;11(1):406-412.
106. Ramezani M, Mohamadzaheri E, Khazaei S, et al. Comparison of EMA, CEA, CD10 and Bcl-2 biomarkers by immunohistochemistry in squamous cell carcinoma and basal cell carcinoma of the skin. *Asian Pacific J Cancer Prev*. 2016;17(3):1379-1383. doi:10.7314/APJCP.2016.17.3.1379
107. Tebcherani AJ, De Andrade HF, Sotto MN. Diagnostic utility of immunohistochemistry in distinguishing trichoepithelioma and basal cell carcinoma: Evaluation using tissue microarray samples. *Mod Pathol*. 2012;25(10):1345-1353. doi:10.1038/modpathol.2012.96
108. 2004 miller. Immunohistochemistry. 2004;(December).
109. Plaza JA, Mackinnon A, Carrillo L, Prieto VG, Sangueza M, Suster S. Role of Immunohistochemistry in the Diagnosis of Sebaceous Carcinoma: A Clinicopathologic and Immunohistochemical Study. *Am J Dermatopathol*. 2015;37(11):809-821. doi:10.1097/DAD.0000000000000255
110. Milosevic M, Lazarevic M, Toljic B, et al. Characterization of stem-like cancer cells in basal cell carcinoma and its surgical margins. *Exp Dermatol*. 2018;27(10):1160-1165. doi:10.1111/exd.13755
111. Wojciechowska-Zdrojowy M, Szepietowski JC, Matusiak A, Dzigiel P, Pua B. Expression of podoplanin in non-melanoma skin cancers and actinic keratosis. *Anticancer Res*. 2016;36(4):1591-1597.
112. Oh ST, Kim HS, Yoo NJ, Lee WS, Cho BK, Reichrath J. Increased immunoreactivity of membrane type-1 matrix metalloproteinase (MT1-MMP) and β-catenin in high-risk basal cell carcinoma. *Br J Dermatol*. 2011;165(6):1197-1204. doi:10.1111/j.1365-2133.2011.10506.x
113. Kadoya K, Fukushi JI, Matsumoto Y, Yamaguchi Y, Stallcup WB. NG2 proteoglycan expression in mouse skin: Altered postnatal skin development in the NG2 null mouse. *J Histochem Cytochem*. 2008;56(3):295-303. doi:10.1369/jhc.7A7349.2007
114. Jensen KB, Jones J, Watt FM. A stem cell gene expression profile of human squamous cell carcinomas. *Cancer Lett*. 2008;272(1):23-31. doi:10.1016/j.canlet.2008.06.014
115. de Bruyn M, Rybczynska AA, Wei Y, et al. Melanoma-associated Chondroitin Sulfate Proteoglycan (MCSP)-targeted delivery of soluble TRAIL potently inhibits melanoma outgrowth in vitro and in vivo. *Mol Cancer*. 2010;9:1-14. doi:10.1186/1476-4598-9-301
116. Jia J, Shi Y, Yan B, et al. LGR5 expression is controled by IKKα in basal cell carcinoma through activating STAT3 signaling pathway. *Oncotarget*. 2016;7(19):27280-27294. doi:10.18632/oncotarget.8465
117. Barto V, Kullov M. Epithelial and Stromal Expression of Syndecan-1 (CD138) in Cutaneous Basal Cell Carcinoma. *J Res Med Dent Sci*. 2017;5(2):21. doi:10.5455/jrmds.2017524
118. Inki P, Larjava H, Haapasalmi K, Miettinen HM, Grenman R, Jalkanen M. Expression of syndecan-1 is induced by differentiation and suppressed by malignant transformation of human keratinocytes. *Eur J Cell Biol*. 1994;63(1):43-51.
119. Biray Avci C, Kaya I, Ozturk A, et al. The role of EGFR overexpression on the recurrence of basal cell carcinomas with positive surgical margins. *Gene*. 2019;687:35-38. doi:10.1016/j.gene.2018.11.024
120. Kim HS, Kim YS, Lee C, Shin MS, Kim JW, Jang BG. Expression profile of sonic hedgehog signaling-related molecules in basal cell carcinoma. *PLoS One*. 2019;14(11):1-12. doi:10.1371/journal.pone.0225511
121. Almomani R, Khanfar M, Bodoor K, et al. Evaluation of patched-1 protein expression level in low risk and high risk basal cell carcinoma subtypes. *Asian Pacific J Cancer Prev*. 2019;20(9):2851-2857. doi:10.31557/APJCP.2019.20.9.2851
122. Gilbert SM, Gidley Baird A, Glazer S, et al. A phase I clinical trial demonstrates that nfP2X7-targeted antibodies provide a novel, safe and tolerable topical therapy for basal cell carcinoma. *Br J Dermatol*. 2017;177(1):117-124. doi:10.1111/bjd.15364
123. Tanese K. Diagnosis and management of basal cell carcinoma. *Curr Treat Options Oncol*. 2019;20(2). doi:10.1007/s11864-019-0610-0
124. Fischer S, Ali OH, Jochum W, Kluckert T, Flatz L, Siano M. Anti-PD-1 Therapy Leads to Near-Complete Remission in a Patient with Metastatic Basal Cell Carcinoma. *Oncol Res Treat*. 2018;41(6):391-394. doi:10.1159/000487084
125. Bichakjian C, Armstrong A, Baum C, et al. Guidelines of care for the management of basal cell carcinoma. *J Am Acad Dermatol*. 2018;78(3):540-559. doi:10.1016/j.jaad.2017.10.006
126. Aasi SZ. Treatment and prognosis of basal cell carcinoma at low risk of recurrence. *UpToDate*. Published online 2020:1-37.
127. Peris K, Fargnoli MC, Garbe C, et al. Diagnosis and treatment of basal cell carcinoma: European consensus–based interdisciplinary guidelines. *Eur J Cancer*. 2019;118:10-34.

doi:10.1016/j.ejca.2019.06.003
128. Lewin J, Carucci J. Advances in the management of basal cell carcinoma. *F1000Prime Rep*. 2015;7(May). doi:10.12703/P7-53
129. Muller FM, Dawe RS, Moseley H, Fleming CJ. Randomized comparison of mohs micrographic surgery and surgical excision for small nodular basal cell carcinoma: Tissue-sparing outcome. *Dermatologic Surg*. 2009;35(9):1349-1354. doi:10.1111/j.1524-4725.2009.01240.x
130. Finley EM. The principles of Mohs micrographic surgery for cutaneous neoplasia. *Ochsner J*. 2003;5(2):22-33.
131. Mansouri B, Bicknell LM, Hill D, Walker GD, Fiala K, Housewright C. Mohs Micrographic Surgery for the Management of Cutaneous Malignancies. *Facial Plast Surg Clin North Am*. 2017;25(3):291-301. doi:10.1016/j.fsc.2017.03.002
132. Snow SN, Madjar DD. Mohs surgery in the management of cutaneous malignancies. *Clin Dermatol*. 2001;19(3):339-347. doi:10.1016/S0738-081X(01)00169-9
133. Thissen MRTM, Neumann MHA, Schouten LJ. A systematic review of treatment modalities for primary basal cell carcinomas. *Arch Dermatol*. 1999;135(10):1177-1183. doi:10.1001/archderm.135.10.1177
134. Van Loo E, Mosterd K, Krekels GAM, et al. Surgical excision versus Mohs' micrographic surgery for basal cell carcinoma of the face: A randomised clinical trial with 10 year follow-up. *Eur J Cancer*. 2014;50(17):3011-3020. doi:10.1016/j.ejca.2014.08.018
135. Smeets NWJ, Krekels GAM, Ostertag JU, et al. Surgical excision vs Mohs' micrographic surgery for basal-cell carcinoma of the face: Randomised controlled trial. *Lancet*. 2004;364(9447):1766-1772. doi:10.1016/S0140-6736(04)17399-6
136. Rowe DE, Carroll RJ, Day CL. Long-term recurrence rates in previously untreated (primary) basal cell carcinoma: implications for patient follow-up. *J Dermatol Surg Oncol*. 1989;15(3):315-328. doi:10.1111/j.1524-4725.1989.tb03166.x
137. Sebaratnam DF, Choy B, Lee M, Paver R, Fernández Peñas P. Direct cost-analysis of mohs micrographic surgery and traditional excision for basal cell carcinoma at initial margin clearance. *Dermatologic Surg*. 2016;42(5):633-638. doi:10.1097/DSS.0000000000000756
138. Institute TCTFS-NC. Targeted Cancer Therapies Fact Sheet - National Cancer Institute. https://www.cancer.gov/about-cancer/treatment/types/targeted-therapies/targeted-therapies-fact-sheet
139. Li C, Chi S, Xie J. Hedgehog signaling in skin cancers. *Cell Signal*. 2011;23(8):1235-1243. doi:10.1016/j.cellsig.2011.03.002
140. Caro I, Low JA. The role of the hedgehog signaling pathway in the development of basal cell carcinoma and opportunities for treatment. *Clin Cancer Res*. 2010;16(13):3335-3339. doi:10.1158/1078-0432.CCR-09-2570
141. McMahon AP, Ingham PW, Tabin CJ. 1 Developmental roles and clinical significance of Hedgehog signaling. *Curr Top Dev Biol*. 2003;53:1-114. doi:10.1016/s0070-2153(03)53002-2
142. Teglund S, Toftgård R. Hedgehog beyond medulloblastoma and basal cell carcinoma. *Biochim Biophys Acta - Rev Cancer*. 2010;1805(2):181-208. doi:10.1016/j.bbcan.2010.01.003
143. Migden MR, Chang ALS, Dirix L, Stratigos AJ, Lear JT. Emerging trends in the treatment of advanced basal cell carcinoma. *Cancer Treat Rev*. 2018;64:1-10. doi:10.1016/j.ctrv.2017.12.009
144. Jain S, Song R, Xie J. Sonidegib: Mechanism of action, pharmacology, and clinical utility for advanced basal cell carcinomas. *Onco Targets Ther*. 2017;10:1645-1653. doi:10.2147/OTT.S130910
145. Wahid M, Jawed A, Mandal RK, et al. Vismodegib, itraconazole and sonidegib as hedgehog pathway inhibitors and their relative competencies in the treatment of basal cell carcinomas. *Crit Rev Oncol Hematol*. 2016;98:235-241. doi:10.1016/j.critrevonc.2015.11.006
146. Amin SH, Motamedi KK, Ochsner MC, Song TE, Hybarger CP. Mechanisms and efficacy of vismodegib in the treatment of basal cell carcinoma. *Discov Med*. 2013;16(89):229-232.
147. Sharpe HJ, Pau G, Yauch RL, et al. Genomic Analysis of Smoothened Inhibitor Resistance in Basal Cell Carcinoma Article Genomic Analysis of Smoothened Inhibitor Resistance in Basal Cell Carcinoma. *Cancer Cell*. 2015;27(3):327-341. doi:10.1016/j.ccell.2015.02.001.Genomic
148. Chen L, Aria AB, Silapunt S, Lee HH, Migden MR. Treatment of advanced basal cell carcinoma with sonidegib: Perspective from the 30-month update of the BOLT trial. *Futur Oncol*. 2018;14(6):515-525. doi:10.2217/fon-2017-0457
149. Peer E, Tesanovic S, Aberger F. Next-generation hedgehog/GLI pathway inhibitors for cancer therapy. *Cancers (Basel)*. 2019;11(4):3-8. doi:10.3390/cancers11040538
150. Sansom DM. CD28, CTLA-4 and their ligands: who does what and to whom? *Immunology*. 2000;101(2):169-177. doi:10.1046/j.1365-2567.2000.00121.x
151. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer*. 2012;12(4):252-264. doi:10.1038/nrc3239
152. Falchook GS, Leidner R, Stankevich E, et al. Responses of metastatic basal cell and cutaneous squamous

cell carcinomas to anti-PD1 monoclonal antibody REGN2810. *J Immunother Cancer*. 2016;4(1):1-5. doi:10.1186/s40425-016-0176-3
153. Mohan S V., Kuo KY, Chang ALS. Incidental regression of an advanced basal cell carcinoma after ipilimumab exposure for metastatic melanoma. *JAAD Case Reports*. 2016;2(1):13-15. doi:10.1016/j.jdcr.2015.11.007
154. Ikeda S, Goodman AM, Cohen PR, et al. Metastatic basal cell carcinoma with amplification of PD-L1: Exceptional response to anti-PD1 therapy. *npj Genomic Med*. 2016;1(September):1-5. doi:10.1038/npjgenmed.2016.37
155. Sabbatino F, Marra A, Liguori L, et al. Resistance to anti-PD-1-based immunotherapy in basal cell carcinoma: a case report and review of the literature. *J Immunother cancer*. 2018;6(1):126. doi:10.1186/s40425-018-0439-2
156. Cohen PR, Kato S, Goodman AM, Ikeda S, Kurzrock R. Appearance of new cutaneous superficial basal cell carcinomas during successful nivolumab treatment of refractory metastatic disease: Implications for immunotherapy in early versus late disease. *Int J Mol Sci*. 2017;18(8):1-7. doi:10.3390/ijms18081663
157. Dall'Acqua WF, Cook KE, Damschroder MM, Woods RM, Wu H. Modulation of the Effector Functions of a Human IgG1 through Engineering of Its Hinge Region. *J Immunol*. 2006;177(2):1129-1138. doi:10.4049/jimmunol.177.2.1129
158. Janeway CAJ, Travers P, Walport M. The generation of diversity in immunoglobulins Virtually. *Immunobiol Immune Syst Heal Dis 5th Ed*. Published online 2001. https://www.ncbi.nlm.nih.gov/books/NBK27140/
159. Schroeder HW, Cavacini L. Structure and function of immunoglobulins. *J Allergy Clin Immunol*. 2010;125(2 SUPPL. 2):237-249. doi:10.1016/j.jaci.2009.09.046
160. Zaroff S, Tan G. Reports Expert Opinion monoclonal antibody generation for in vivo applications. *Biotechniques*. 2019;67(3):90-92.
161. Dal Ferro M, Rizzo S, Rizzo E, et al. Phage Display Technology for Human Monoclonal Antibodies. In: Vol 1904. ; 2019:319-338. doi:10.1007/978-1-4939-8958-4_15
162. Almagro JC, Pedraza-Escalona M, Arrieta HI, Pérez-Tapia SM. Phage Display Libraries for Antibody Therapeutic Discovery and Development. *Antibodies*. 2019;8(3):44. doi:10.3390/antib8030044
163. Kuniyoshi Y, Maehara K, Iwasaki T, et al. Identification of immunoglobulin gene sequences from a small read number of mRNA-seq using hybridomas. *PLoS One*. 2016;11(10):1-10. doi:10.1371/journal.pone.0165473
164. Babrak L, McGarvey JA, Stanker LH, Hnasko R. Identification and verification of hybridoma-derived monoclonal antibody variable region sequences using recombinant DNA technology and mass spectrometry. *Mol Immunol*. 2017;90(September):287-294. doi:10.1016/j.molimm.2017.08.014
165. Frenzel A, Hust M, Schirrmann T. Expression of recombinant antibodies. *Front Immunol*. 2013;4(JUL):1-20. doi:10.3389/fimmu.2013.00217
166. Schirrmann T, Al-Halabi L, Dübel S, Hust M. Production systems for recombinant antibodies. *Front Biosci*. 2008;13(6):4576-4594. doi:10.2741/3024
167. Peter J. Delves, Seamus J. Martin, Dennis R. Burton IMR. Immunological methods and application. In: *Roitt's Essential Immunology,*. 12th Editi. Wiley-Blackwell; 2011:141-187.
168. Morrison SL, Johnson MJ, Herzenberg LA, Oi VT. Chimeric human antibody molecules: Mouse antigen-binding domains with human constant region domains. *Proc Natl Acad Sci U S A*. 1984;81(21 I):6851-6855. doi:10.1073/pnas.81.21.6851
169. Jones PT, Dear PH, Foote J, Neuberger MS, Winter G. Replacing the complementarity-determining regions in a human antibody with those from a mouse. *Nature*. 1986;321(6069):522-525. doi:10.1038/321522a0
170. Chothia C, Lesk AM. Canonical structures for the hypervariable regions of immunoglobulins. *J Mol Biol*. 1987;196(4):901-917. doi:10.1016/0022-2836(87)90412-8
171. Tomizuka K, Ishida I, Lonberg N, HALK E. Transgenic transchromosomal rodents for making human antibodies. Published online 2003. https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2002043478
172. Nanna A, Roush W, Rader C. Chemical Assembly of Antibody-Drug Conjugates. In: ; 2017:1-28. doi:10.1007/978-3-319-46877-8_1
173. Agarwal P, Bertozzi CR. Site-specific antibody-drug conjugates: The nexus of bioorthogonal chemistry, protein engineering, and drug development. *Bioconjug Chem*. 2015;26(2):176-192. doi:10.1021/bc5004982
174. Cole NB. Site-specific protein labeling with SNAP-tags. *Curr Protoc Protein Sci*. 2013;2013:30.1.1-30.1.16. doi:10.1002/0471140864.ps3001s73
175. England CG, Luo H, Cai W. HaloTag Technology: A Versatile Platform for Biomedical Applications. *Bioconjug Chem*. 2015;26(6):975-986. doi:10.1021/acs.bioconjchem.5b00191

176. Zhou Z, Cironi P, Lin AJ, et al. Genetically encoded short peptide tags for orthogonal labelling by Sfp and AcpS phosphopantetheinyl transferases. *ACS Chem Biol*. 2007;2(5):337-346. doi:10.1021/cb700054k
177. Alam MK, El-Sayed A, Barreto K, Bernhard W, Fonge H, Geyer CR. Site-Specific Fluorescent Labeling of Antibodies and Diabodies Using SpyTag/SpyCatcher System for In Vivo Optical Imaging. *Mol Imaging Biol*. 2019;21(1):54-66. doi:10.1007/s11307-018-1222-y
178. Gautier A, Hinner MJ. *Site-Specific Protein Labeling: Methods and Protocols*. Vol 1266. (Gautier A, Hinner MJ, eds.). Springer New York; 2015. doi:10.1007/978-1-4939-2272-7
179. Lotze J, Reinhardt U, Seitz O, Beck-Sickinger AG. Peptide-tags for site-specific protein labelling: In vitro and in vivo. *Mol Biosyst*. 2016;12(6):1731-1745. doi:10.1039/c6mb00023a
180. Virant D, Traenkle B, Maier J, et al. A peptide tag-specific nanobody enables high-quality labeling for dSTORM imaging. *Nat Commun*. 2018;9(1):1-14. doi:10.1038/s41467-018-03191-2
181. Viswanathan S, Williams ME, Bloss EB, et al. High-performance probes for light and electron microscopy. *Nat Methods*. 2015;12(6):568-576. doi:10.1038/nmeth.3365
182. Landgraf D, Okumus B, Chien P, Baker TA, Paulsson J. Segregation of molecules at cell division reveals native protein localization. *Nat Methods*. 2012;9(5):480-482. doi:10.1038/nmeth.1955
183. Hussain AF, Heppenstall PA, Kampmeier F, Meinhold-Heerlein I, Barth S. One-step site-specific antibody fragment auto-conjugation using SNAP-tag technology. *Nat Protoc*. 2019;14(11):3101-3125. doi:10.1038/s41596-019-0214-y
184. Woitok M, Klose D, Di Fiore S, et al. Comparison of a mouse and a novel human scFv-SNAP-auristatin F drug conjugate with potent activity against EGFR-overexpressing human solid tumor cells. *Onco Targets Ther*. 2017;10:3313-3327. doi:10.2147/OTT.S140492
185. Hussain AF, Kampmeier F, Von Felbert V, Merk HF, Tur MK, Barth S. SNAP-tag technology mediates site specific conjugation of antibody fragments with a photosensitizer and improves target specific phototoxicity in tumor cells. *Bioconjug Chem*. 2011;22(12):2487-2495. doi:10.1021/bc200304k
186. Padayachee ER, Biteghe FAN, Malindi Z, Bauerschlag D, Barth S. Human Antibody Fusion Proteins/Antibody Drug Conjugates in Breast and Ovarian Cancer. *Transfus Med Hemotherapy*. 2017;44(5):303-310. doi:10.1159/000479979
187. Mollwitz B. Structural Studies and Protein Engineering of Human O 6 -Alkylguanine-DNA Alkyltransferase PAR. 2012;5343.
188. Sun X, Zhang A, Baker B, et al. Development of SNAP-tag fluorogenic probes for wash-free fluorescence imaging. *ChemBioChem*. 2011;12(14):2217-2226. doi:10.1002/cbic.201100173
189. Voölkel T, Korn T, Bach M, Muöller R, Kontermann RE. Optimized linker sequences for the expression of monomeric and dimeric bispecific single-chain diabodies. *Protein Eng*. 2001;14(10):815-823. doi:10.1093/protein/14.10.815
190. Kampmeier F, Ribbert M, Nachreiner T, et al. Site-specific, covalent labeling of recombinant antibody fragments via fusion to an engineered version of 6-O-alkylguanine DNA alkyltransferase. *Bioconjug Chem*. 2009;20(5):1010-1015. doi:10.1021/bc9000257
191. Padayachee ER, Adeola HA, Van Wyk JC, et al. Applications of SNAP-tag technology in skin cancer therapy. *Heal Sci Reports*. 2019;2(2):e103. doi:10.1002/hsr2.103
192. Ranganathan S, Gribskov M, Nakai K, Schönbach C. *Encyclopedia of Bioinformatics and Computational Biology*. Vol 1-3.; 2018. doi:10.1016/c2016-1-00174-8
193. Lockhart DJ, Dong H, Byrne MC, et al. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol*. 1996;14(13):1675-1680. doi:10.1038/nbt1296-1675
194. Gunderson KL, Kruglyak S, Graige MS, et al. Decoding randomly ordered DNA arrays. *Genome Res*. 2004;14(5):870-877. doi:10.1101/gr.2255804
195. Park T, Yi SG, Kang SH, Lee SY, Lee YS, Simon R. Evaluation of normalization methods for microarray data. *BMC Bioinformatics*. 2003;4:1-13. doi:10.1186/1471-2105-4-33
196. Hicks SC, Irizarry RA. quantro: A data-driven approach to guide the choice of an appropriate normalization method. *Genome Biol*. 2015;16(1):1-8. doi:10.1186/s13059-015-0679-0
197. Zhang Y, Szustakowski J, Schinke M. Bioinformatics analysis of microarray data. *Methods Mol Biol*. 2009;573:259-284. doi:10.1007/978-1-60761-247-6_15
198. Fay DS. A biologist's guide to statistical thinking and analysis. *WormBook*. Published online July 9, 2013:1-54. doi:10.1895/wormbook.1.159.1
199. Colquhoun D. An investigation of the false discovery rate and the misinterpretation of p-values. *R Soc Open Sci*. 2014;1(3). doi:10.1098/rsos.140216
200. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res*. 2013;41(D1):D991-D995. doi:10.1093/nar/gks1193
201. Simon R, Lam A, Li MC, Ngan M, Menenzes S, Zhao Y. Analysis of gene expression data using BRB-array tools. *Cancer Inform*. 2007;3(February 2007):11-17. doi:10.1177/117693510700300022
202. Wright GW, Simon RM. A random variance model for detection of differential gene expression in small

microarray experiments. *Bioinformatics*. 2003;19(18):2448-2455. doi:10.1093/bioinformatics/btg345
203. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B*. 1995;57(1):289-300. doi:10.1111/j.2517-6161.1995.tb02031.x
204. Bausch-Fluck D, Goldmann U, Müller S, et al. The in silico human surfaceome. *Proc Natl Acad Sci U S A*. 2018;115(46):E10988-E10997. doi:10.1073/pnas.1808790115
205. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A*. 2001;98(9):5116-5121. doi:10.1073/pnas.091062498
206. McEntyre 2001. Bridging the information gap. doi:10.1086/208703
207. Lima WC, Gasteiger E, Marcatili P, Duek P, Bairoch A, Cosson P. The ABCD database: A repository for chemically defined antibodies. *Nucleic Acids Res*. 2020;48(D1):D261-D264. doi:10.1093/nar/gkz714
208. Huston JS, Levinson D, Mudgett-Hunter M, et al. Protein engineering of antibody binding sites: Recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in Escherichia coli. *Proc Natl Acad Sci U S A*. 1988;85(16):5879-5883. doi:10.1073/pnas.85.16.5879
209. Stöcker M, Tur MK, Sasse S, Krüßmann A, Barth S, Engert A. Secretion of functional anti-CD30-angiogenin immunotoxins into the supernatant of transfected 293T-cells. *Protein Expr Purif*. 2003;28(2):211-219. doi:10.1016/S1046-5928(02)00709-X
210. Cao Y, Mei L, Wu D, Zhao C. [The construction of pSecTag2/B vector modified by osteoprotegerin]. *Shanghai Kou Qiang Yi Xue*. 2005;14(3):273-276.
211. Barth S, Huhn M, Wels W, Diehl V, Engert A. Construction and in vitro evaluation of RFT5(scFv)-ETA', a new recombinant single-chain immunotoxin with specific cytotoxicity toward CD25+ Hodgkin-derived cell lines. *Int J Mol Med*. 1998;1(1):249-256. doi:10.3892/ijmm.1.1.249
212. Kearse MG, Wilusz JE. Non-AUG translation: A new start for protein synthesis in eukaryotes. *Genes Dev*. 2017;31(17):1717-1731. doi:10.1101/gad.305250.117
213. Tholen M, Hillebrand LE, Tholen S, Sedelmeier O, Arnold SJ, Reinheckel T. Out-of-frame start codons prevent translation of truncated nucleo-cytosolic cathepsin L in vivo. *Nat Commun*. 2014;5:1-12. doi:10.1038/ncomms5931
214. Noderer WL, Flockhart RJ, Bhaduri A, et al. Noderer WL,,, and Wang CL (2014) Quantitative analysis of mammalian translation. Mol Sys Bio. *Mol Syst Biol*. 2014;10(8):1-14.
215. Acevedo JM, Hoermann B, Schlimbach T, Teleman AA. Changes in global translation elongation or initiation rates shape the proteome via the Kozak sequence. *Sci Rep*. 2018;8(1):1-12. doi:10.1038/s41598-018-22330-9
216. Hristodorov D, Amoury M, Mladenov R, et al. EpCAM-Selective Elimination of Carcinoma Cells by a Novel MAP-Based Cytolytic Fusion Protein. *Mol Cancer Ther*. 2014;13(9):2194-2202. doi:10.1158/1535-7163.MCT-13-0781
217. Trastoy MO, Defais M, Larminat F. Resistance to the antibiotic Zeocin by stable expression of the Sh ble gene does not fully suppress Zeocin-induced DNA cleavage in human cells. *Mutagenesis*. 2005;20(2):111-114. doi:10.1093/mutage/gei016
218. Krauss J, Arndt MAE, Martin ACR, Liu H, Rybak SM. Specificity grafting of human antibody frameworks selected from a phage display library: generation of a highly stable humanized anti-CD22 single-chain Fv fragment. *Protein Eng*. 2003;16(10):753-759. doi:10.1093/protein/gzg096
219. Cousins RJ, Blanchard RK, Popp MP, et al. A global view of the selectivity of zinc deprivation and excess on genes expressed in human THP-1 mononuclear cells. *Proc Natl Acad Sci U S A*. 2003;100(12):6952-6957. doi:10.1073/pnas.0732111100
220. Armando Casas-Mollano J, Zinselmeier MH, Erickson SE, Smanski MJ. CRISPR-Cas Activators for Engineering Gene Expression in Higher Eukaryotes. *Cris J*. 2020;3(5):350-364. doi:10.1089/crispr.2020.0064
221. Jin B, Wang W, Du G, et al. Identifying hub genes and dysregulated pathways in hepatocellular carcinoma. *Eur Rev Med Pharmacol Sci*. 2015;19(4):592-601.
222. de Araújo LJT, Lerario AM, de Castro M, et al. Transcriptome Analysis Showed a Differential Signature between Invasive and Non-invasive Corticotrophinomas. *Front Endocrinol (Lausanne)*. 2017;8(MAR):1-12. doi:10.3389/fendo.2017.00055
223. Di Mauro S, Scamporrino A, Fruciano M, et al. Circulating Coding and Long Non-Coding RNAs as Potential Biomarkers of Idiopathic Pulmonary Fibrosis. *Int J Mol Sci*. 2020;21(22):1-12. doi:10.3390/ijms21228812
224. Linskey KR, Gimbel DC, Zukerberg LR, Duncan LM, Sadow PM, Nazarian RM. BerEp4, cytokeratin 14, and cytokeratin 17 immunohistochemical staining aid in differentiation of basaloid squamous cell carcinoma from basal cell carcinoma with squamous metaplasia. *Arch Pathol Lab Med*. 2013;137(11):1591-1598. doi:10.5858/arpa.2012-0424-OA
225. Wan J, Dai H, Zhang X, et al. Distinct transcriptomic landscapes of cutaneous basal cell carcinomas and squamous cell carcinomas. *Genes Dis*. 2021;8(2):181-192. doi:10.1016/j.gendis.2019.10.004

226. Ning H, Mitsui H, Wang CQF, et al. Identification of anaplastic lymphoma kinase as a potential therapeutic target in Basal Cell Carcinoma. *Oncotarget*. 2013;4(12):2237-2248. doi:10.18632/oncotarget.1357
227. Sullivan TP, Dearaujo T, Vincek V, Berman B. Evaluation of Superficial Basal Cell Carcinomas after Treatment with Imiquimod 5% Cream or Vehicle for Apoptosis and Lymphocyte Phenotyping. *Dermatologic Surg*. 2003;29(12):1181-1186. doi:10.1111/j.1524-4725.2003.29399.x
228. Chen D, Zou J, Zong Y, Meng H, An G, Yang L. Anti-human CD138 monoclonal antibodies and their bispecific formats: generation and characterization. *Immunopharmacol Immunotoxicol*. 2016;38(3):175-183. doi:10.3109/08923973.2016.1153110
229. MORGAN AC, GALLOWAY DR, REISFELD RA. Production and Characterization of Monoclonal Antibody to a Melanoma Specific Glycoprotein. *Hybridoma*. 1981;1(1):27-36. doi:10.1089/hyb.1.1981.1.27
230. Güler-Gane G, Kidd S, Sridharan S, Vaughan TJ, Wilkinson TCI, Tigue NJ. Overcoming the refractory expression of secreted recombinant proteins in mammalian cells through modification of the signal peptide and adjacent amino acids. *PLoS One*. 2016;11(5):1-15. doi:10.1371/journal.pone.0155340
231. Arnau J, Lauritzen C, Petersen GE, Pedersen J. Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein Expr Purif*. 2006;48(1):1-13. doi:10.1016/j.pep.2005.12.002
232. Wang P, Shim E, Cravatt B, et al. Escherichia coli Signal Peptide Peptidase A Is a Serine-Lysine Protease with a Lysine Recruited to the Nonconserved Amino-Terminal Domain in the S49 Protease Family. *Biochemistry*. 2008;47(24):6361-6369. doi:10.1021/bi800657p
233. Mohanty AK, Wiener MC. Membrane protein expression and production: Effects of polyhistidine tag length and position. *Protein Expr Purif*. 2004;33(2):311-325. doi:10.1016/j.pep.2003.10.010
234. Schneider R. New Ways of Initiating Translation in Eukaryotes? *Mol Cell Biol*. 2001;21(23):8238-8246. doi:10.1128/mcb.21.23.8238-8246.2001
235. Kochetov A V. Alternative translation start sites and hidden coding potential of eukaryotic mRNAs. *BioEssays*. 2008;30(7):683-691. doi:10.1002/bies.20771
236. Xu H, Wang P, Fu Y, et al. Length of the ORF, position of the first AUG and the Kozak motif are important factors in potential dual-coding transcripts. *Cell Res*. 2010;20(4):445-457. doi:10.1038/cr.2010.25
237. Hernández G, Osnaya VG, Pérez-Martínez X. Conservation and Variability of the AUG Initiation Codon Context in Eukaryotes. *Trends Biochem Sci*. 2019;44(12):1009-1021. doi:10.1016/j.tibs.2019.07.001
238. Roberts RJ, Danna K. How restriction enzymes became the workhorses of molecular biology. *Proc Natl Acad Sci U S A*. 2005;102(17):5905-5908.
239. Stepanenko AA, Heng HH. Transient and stable vector transfection: Pitfalls, off-target effects, artifacts. *Mutat Res - Rev Mutat Res*. 2017;773:91-103. doi:10.1016/j.mrrev.2017.05.002
240. Reddy TL, Krishnarao PS, Rao GK, et al. Para amino benzoic acid-derived self-assembled biocompatible nanoparticles for efficient delivery of siRNA. *Int J Nanomedicine*. 2015;10:6411-6424. doi:10.2147/IJN.S86238
241. Hunter M, Yuan P, Vavilala D, Fox M. Optimization of Protein Expression in Mammalian Cells. *Curr Protoc Protein Sci*. 2019;95(1):1-28. doi:10.1002/cpps.77
242. Subedi GP, Johnson RW, Moniz HA, Moremen KW, Barb A. High yield expression of recombinant human proteins with the transient transfection of HEK293 cells in suspension. *J Vis Exp*. 2015;2015(106):1-10. doi:10.3791/53568
243. Pozza ED, Dando I, Biondani G, et al. Pancreatic ductal adenocarcinoma cell lines display a plastic ability to bi-directionally convert into cancer stem cells. *Int J Oncol*. 2015;46(3):1099-1108. doi:10.3892/ijo.2014.2796
244. Amoury M, Bauerschlag D, Zeppernick F, et al. Photoimmunotheranostic agents for triple-negative breast cancer diagnosis and therapy that can be activated on demand. *Oncotarget*. 2016;7(34):54925-54936. doi:10.18632/oncotarget.10705
245. Solatycka A, Owczarek T, Piller F, et al. MUC1 in human and murine mammary carcinoma cells decreases the expression of core 2 β1,6-N-acetylglucosaminyltransferase and β-galactoside α2,3-sialyltransferase. *Glycobiology*. 2012;22(8):1042-1054. doi:10.1093/glycob/cws075
246. Kim MJ, Choi JR, Tae N, et al. Novel antibodies targeting MUC1-C showed anti-metastasis and growth-inhibitory effects on human breast cancer cells. *Int J Mol Sci*. 2020;21(9):1-18. doi:10.3390/ijms21093258
247. Mouffouk F, Simão T, Dornelles DF, et al. Self-assembled polymeric nanoparticles as new, smart contrast agents for cancer early detection using magnetic resonance imaging. *Int J Nanomedicine*. 2014;10:63-76. doi:10.2147/IJN.S71190
248. Gordon LG, Elliott TM, Wright CY, Deghaye N, Visser W. Modelling the healthcare costs of skin cancer in South Africa. *BMC Health Serv Res*. 2016;16(1):1-9. doi:10.1186/s12913-016-1364-z

249. Tolkachjov SN, Brodland DG, Coldiron BM, et al. Understanding Mohs Micrographic Surgery: A Review and Practical Guide for the Nondermatologist. *Mayo Clin Proc*. 2017;92(8):1261-1271. doi:10.1016/j.mayocp.2017.04.009
250. Sunjaya AP, Sunjaya AF, Tan ST. The Use of BEREP4 Immunohistochemistry Staining for Detection of Basal Cell Carcinoma. *J Skin Cancer*. 2017;2017. doi:10.1155/2017/2692604
251. Ansai S, Takayama R, Kimura T, Kawana S. Ber-EP4 is a useful marker for follicular germinative cell differentiation of cutaneous epithelial neoplasms. *J Dermatol*. 2012;39(8):688-692. doi:10.1111/j.1346-8138.2011.01494.x
252. Evaluation of Diagnostic Values of EMA and Ber-Ep4 in Distinction between Basal Cell Carcinoma and Squamous Cell Carcinoma of the Skin. *Rajabi*. 2007;2(January):119-123. doi:10.4314/ajcem.v12i3.
253. Bartoš V. Immunohistochemical expression of BerEP4, CK7 and EMA in basal cell carcinoma of the skin. *Bangabandhu Sheikh Mujib Med Univ J*. 2019;12(4):163-166. doi:10.3329/bsmmuj.v12i4.43721
254. Dasgeb B, Mohammadi TM, Mehregan DR. Use of Ber-EP4 and Epithelial Specific Antigen to Differentiate Clinical Simulators of Basal Cell Carcinoma. *Biomark Cancer*. 2013;5:BIC.S11856. doi:10.4137/BIC.S11856
255. Sinard JH. Immunohistochemical distinction of ocular sebaceous carcinoma from basal cell and squamous cell carcinoma. *Arch Ophthalmol*. 1999;117(6):776-783. doi:10.1001/archopht.117.6.776
256. Fan YS, Carr RA, Sanders DSA, Smith AP, Lazar AJF, Calonje E. Characteristic Ber-EP4 and EMA expression in sebaceoma is immunohistochemically distinct from basal cell carcinoma. *Histopathology*. 2007;51(1):80-86. doi:10.1111/j.1365-2559.2007.02722.x
257. Kyung-A Hyun, Goo KB, Han H, et al. Epithelial-to-mesenchymal transition leads to loss of EpCAM and different physical properties in circulating tumor cells from metastatic breast cancer. *Oncotarget*. 2016;7(17):24677-24687. doi:10.18632/oncotarget.8250
258. Dalman MR, Deeter A, Nimishakavi G, Duan ZH. Fold change and p-value cutoffs significantly alter microarray interpretations. *BMC Bioinformatics*. 2012;13 Suppl 2(Suppl 2):11-14. doi:10.1186/1471-2105-13-S2-S11
259. Zhao CY, Hwang SJE, Anforth R, et al. Incidence of Basal Cell Carcinoma and Squamous Cell Carcinoma in Patients on Antiprogrammed Cell Death-1 Therapy for Metastatic Melanoma. *J Immunother*. 2018;41(7):343-349. doi:10.1097/CJI.0000000000000237
260. Gracia-Cazaña T, Mascaraque M, Lucena SR, et al. Biomarkers of basal cell carcinoma resistance to methyl-aminolevulinate photodynamic therapy. *PLoS One*. 2019;14(4):1-14. doi:10.1371/journal.pone.0215537
261. Hill C, Wang Y. The importance of epithelial-mesenchymal transition and autophagy in cancer drug resistance. *Cancer Drug Resist*. 2020;3(1):38-47. doi:10.20517/cdr.2019.75
262. Anshika Bakshi1,2, Sandeep C. Chaudhary1, Mehtab Rana1, Craig A. Elmets1 and MA. Basal cell carcinoma pathogenesis and therapy involving hedgehog signaling and beyond. 2017;344(6188):1173-1178. doi:10.1126/science.1249098.Sleep
263. Otsuka A, Levesque MP, Dummer R, Kabashima K. Hedgehog signaling in basal cell carcinoma. *J Dermatol Sci*. 2015;78(2):95-100. doi:10.1016/j.jdermsci.2015.02.007
264. Zhong Z, Virshup DM. Wnt signaling and drug resistance in cancer. *Mol Pharmacol*. 2020;97(2):72-89. doi:10.1124/MOL.119.117978
265. Noubissi FK, Yedjou CG, Spiegelman VS, Tchounwou PB. Cross-talk between Wnt and Hh signaling pathways in the pathology of basal cell carcinoma. *Int J Environ Res Public Health*. 2018;15(7). doi:10.3390/ijerph15071442
266. Lara F, Eduardo L, Melo F De. Recurrence rate of basal cell carcinoma with positive histopathological margins and related risk factors. Published online 2017:58-62.
267. Litvinov I V., Xie P, Gunn S, Sasseville D, Lefrançois P. The transcriptional landscape analysis of basal cell carcinomas reveals novel signalling pathways and actionable targets. *Life Sci Alliance*. 2021;4(7):1-12. doi:10.26508/LSA.202000651
268. Rolih V, Barutello G, Iussich S, et al. CSPG4: A prototype oncoantigen for translational immunotherapy studies. *J Transl Med*. 2017;15(1):1-14. doi:10.1186/s12967-017-1250-4
269. Truant S, Bruyneel E, Gouyer V, et al. Requirement of both mucins and proteoglycans in cell-cell dissociation and invasiveness of colon carcinoma HT-29 cells. *Int J Cancer*. 2003;104(6):683-694. doi:10.1002/ijc.11011

CHAPTER 7: APPENDIX

7.1. LIST OF ABBREVIATIONS

Ab: antibody
ADC: antibody–drug conjugate
ADCC: antibody-dependent cellular cytotoxicity
Ag: antigen
AGT: O(6)-alkylguanine-DNA alkyltransferase
APC: antigen-presenting cell
ATCC: American Type Culture Collection
AV: annexin V
BC: Breast cancer
BCA: bicinchoninic acid
BCC: basal cell carcinoma
BG: benzylguanine
BG-AF: BG-modified auristatin-F
bp: base pair
BRB: Biometric Research Branch
BSA: bovine serum albumin
BSC biosafety cabinet
bSCCs: basaloid squamous cell carcinomas
CANSA: Cancer Association of South Africa
CAR T-cell: chimeric antigen receptor T cell
CCSN: cell culture supernatant
CDC: complement-dependent cytotoxicity
cDNA: complementary DNA
CDR: complementarity-determining region
CFS: cell-free supernatant
CH: constant domain, heavy chain
CHO: Chinese hamster ovary cells
CL: constant domain, light chain
CMV: Cytomegalovirus
cSCC: cutaneous squamous cell carcinomas (cSCC)
CSPG4: chondroitin sulfate proteoglycan
CTLA4: cytotoxic T lymphocyte associated protein 4
CV: column volume
DEGs: differential expressed genes
del: deletion
DMEM: Dulbecco's Modified Eagle's Medium
DMSO: dimethyl sulfoxide
DNA: deoxyribonucleic acid
E. coli: Escherichia coli
EDTA: ethylenediaminetetraacetic acid
eGFP: enhanced green fluorescent protein
EKS: enterokinase cleavage site
ELISA: enzyme-linked immunosorbent assay
EMA: epithelial membrane antigen

EMT: epithelial–mesenchymal transition
EpCAM: epithelial cell adhesion molecule
ER: endoplasmic reticulum
et al.: *et alia*
ETA: Pseudomonas aeruginosa exotoxin A
EtOH: ethyl alcohol/Ethanol
Fab /F(ab)2: antigen binding fragment
Fc: crystallizable fragment
FC: fold change
FDA: Food and Drug Administration
FDRs: false-discovery rates
FITC: fluorescein isothiocyanate
G_4S: glycine (x4)–serine linker
GEO: gene Expression Omnibus
GFP: green fluorescent protein
GrB: granzyme B
HAMA: human anti-mouse antibody
HEK: human embryonic kidney
HGPRT: hypoxanthine–guanine phosphoribosyltransferase
HH: hedgehog
HPV: Human papilloma virus
IC_{50}: half maximal inhibitory concentration
ICC: immunocytochemistry (ICC)
IF: immunofluorescence (IF)
IFE: interfollicular epidermis
IFN-γ: interferon-γ
IgG: immunoglobulin G
IHC: immunohistochemistry
IMAC immobilized metal affinity chromatography
IPTG: isopropyl-beta-D-thiogalactopyranoside
IRES internal ribosome entry site
Kan: kanamycin
Kb: kilobase
kDa kilodalton
KZ: Kozak
LB: Luria–Bertani broth
mAb: monoclonal antibody
MAL: methyl-aminolevulinate
MDR: multidrug resistance
mL: milliliter
MMAE: monomethyl auristatin E
MMAF: monomethyl auristatin F
MMS: Mohs micrographic surgery
MT: mutant
MUC-1: mucin 1
MW: molecular weight
MWCO: molecular weight cutoff
ND: not defined
Ng: nanogram
NIR: near-infrared

Nm: nanometer
NMR: nuclear magnetic resonance
NMSCs: Non-melanoma skin cancers
NS: normal skin
Nt: nucleotides
OD: optical density
ORF: open reading frame
ori origin of replication
PBS: phosphate-buffered saline
PCR: polymerase chain reaction
PD-1: programmed death receptor 1
PD-L1: programmed death ligand 1
PDT: photodynamic therapy
PDX: patient-derived xenograft
pH: negative log of the concentration of hydrogen ions in solution
PI: propidium iodide
PIT: photoimmunotherapy
PK: pharmacokinetics
PVDF polyvinylidene difluoride
RNA: ribonucleic acid
rpm: revolutions per minute
RPMI-1640 Roswell Park Memorial Institute medium 1640
SAM: significance analysis of microarrays
SCC: squamous cell carcinomas
scFv: single-chain variable fragment
SDS PAGE: sodium dodecyl–sulfate polyacrylamide gel electrophoresis
SEC: size-exclusion chromatography
SMI: small-molecule inhibitors
TAA: tumor-associated antigen
TAC: Transcriptome Analysis Console
TAE: Tris–acetate–EDTA buffer
Taq: *Thermus aquaticus*
TBS: Tris-buffered saline
TE: trichoepithelioma
TEMED: tetramethylethylenediamine
TGX: Tris–glycine extended
TISs: translation initiation sites
UV: ultraviolet
VH: variable domain, heavy chain
VL: variable domain, light chain
WB: Western blot
WHO: World Health Organization
WT: Wild type
XTT: sodium 3'-[1-(phenylaminocarbonyl)-3,4-tetrazolium]-bis(4-methoxy-6-nitro)benzene-sulfonic acid hydrate
× *g*: relative centrifugal force
°C: degree Celsius
µg: microgram
µl: microliter

7.2 LIST OF TABLES

7.3 LIST OF FIGURES