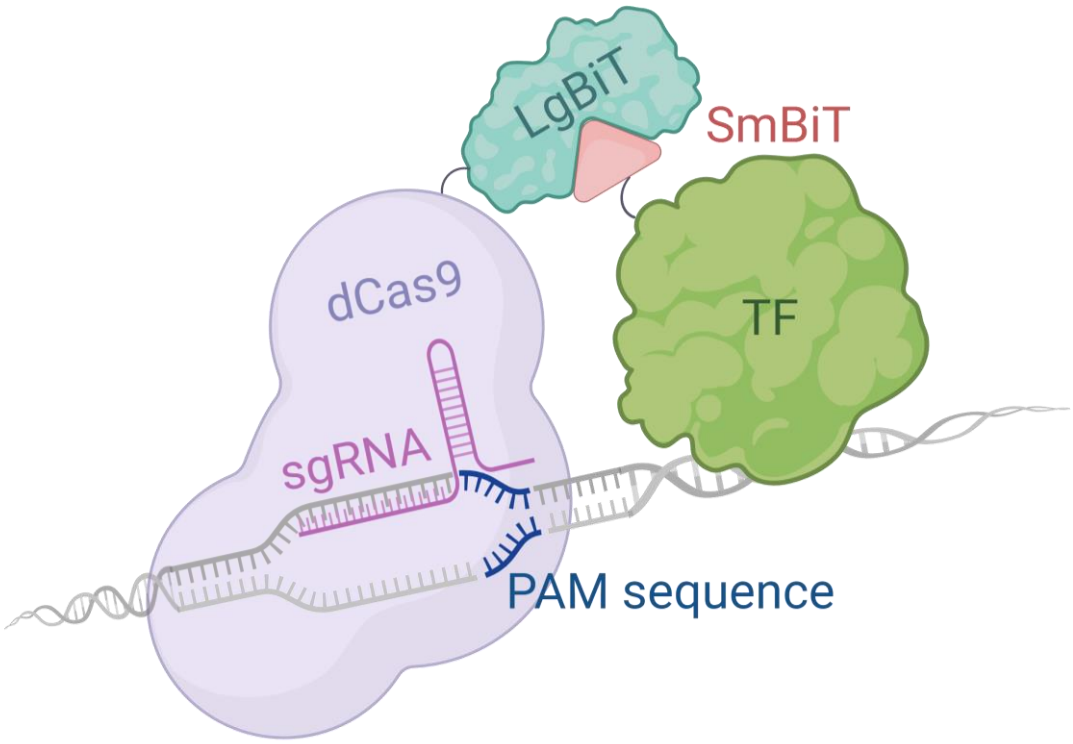
CHALMERS
UNIVERSITY OF TECHNOLOGY



Development of novel high-throughput assays to measure compound binding to transcription factors in cells

Master's thesis in Biotechnology

CLARA JOHNSSON

DEPARTMENT OF LIFE SCIENCES
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2024
www.chalmers.se

MASTER'S THESIS 2024

Development of novel high-throughput assays to measure compound binding to transcription factors in cells

CLARA JOHNSSON

CHALMERS
UNIVERSITY OF TECHNOLOGY

MSc Thesis 2024
Department of Life Sciences
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone +46 (0)31-772 1000

Sponsored by and performed in collaboration with the Cell Engineering Team
Discovery Biology Department
AstraZeneca AB R&D
SE-431 50 Mölndal
Sweden
Telephone +46 (0)31-776 1000

Development of novel high-throughput assays to measure compound binding to transcription factors in cells
CLARA JOHNSSON

Supervisors: Anna-Lina Cavallo, Associate Principal Scientist, AstraZeneca, Gothenburg & Sandra Stefanovic-Barrett, Associate Principal Scientist, AstraZeneca, Cambridge

Examiner: Yvonne Nygård, Associate Professor, Chalmers University of Technology, Gothenburg

Acknowledgements, dedications, and similar personal statements in this thesis, reflect the author's own views.

Cover: Image illustrating the split-reporter system developed in this project. Created with BioRender.com.
Typeset in Word

Printed by Chalmers Reproservice
Gothenburg, Sweden 2024

Development of novel high-throughput assays to measure compound binding to transcription factors in cells

CLARA JOHNSSON

Department of Life Sciences
Chalmers University of Technology

Abstract

Transcription factors (TFs) are proteins possessing domains that recognizes and bind to specific DNA sequences. These proteins also possess domains that recruit various collaborator proteins and the transcriptional regulatory machinery, thus regulating the rate of transcription in nearby genes. Dysregulations of gene expression mediated by TFs is associated with numerous diseases, making them attractive targets for research. However, the understanding of the intricate mechanisms governing TF-DNA binding remains limited. This lack of knowledge complicates the study of TF-DNA binding, making it both labour-intensive and costly. To overcome these challenges, there is a need for high-throughput screening (HTS) methods that can efficiently identify TF-DNA interactions, characterize functional binding site motifs, and evaluate the impact of compounds on the TF-DNA binding. This thesis project thereby aimed to develop a novel cell-based assay to measure TF-DNA interactions in high throughput by leveraging the split-reporter system NanoBiT®. This system comprises two subunits: the Large BiT (LgBiT) and Small BiT (SmBiT), that upon proximity forms an active enzyme that produces luminescence. The concept involved tagging dead Cas9 (dCas9) with a LgBiT, tagging various TFs with SmBiTs, and providing single guide RNA (sgRNA), directing dCas9 to target DNA sites. The study focused on two disease related TFs: signal transducers and activators of transcription 3 (STAT3) and androgen receptor (AR). Two distinct cell models were used: HEK-293 for STAT3 and LNCaP for AR. Plasmids expressing dCas9 linked to N- or C-terminus of the LgBiT were designed and synthesized. Following transfections with these constructs, the work of generating stable pools using geneticin antibiotic selection was initiated. Subsequently, TF-SmBiT plasmids were designed and synthesized, along with a range of sgRNAs. These guides were designed to target promoter sequences of downstream genes associated with the TF. The established cell lines expressing dCas9-LgBiT were then transiently transfected with the TF-SmBiT plasmid and various sgRNAs, enabling subsequent measurement of the TF-DNA interactions. In this project, stable pools of HEK-293 cells expressing dCas9-LgBiT-N/C-terminus were established and validated through western blots. Despite several attempts, generation of LNCaP cell lines stably expressing these constructs was never achieved. After generating the assay reagents, the NanoBiT® assay was developed and tested. Specific interactions were not evident in the initial testing phase, when testing the first set of sgRNAs. By designing and testing more sgRNAs, the developed system holds potential for evaluating TF-DNA binding in a high-throughput manner, identifying transcription factor binding sites (TFBSs), and assessing the impact of compounds on TF-DNA interactions. Nonetheless, further validation of the assay is imperative to ensure its reliability and applicability in future studies.

Keywords: Transcription factor, High-throughput screening, Cell-based assay, Split-reporter system, dCas9, sgRNA, STAT3, AR, HEK-293, LNCaP

Acknowledgements

First and foremost, I would like to thank my two supervisors, Anna-Lina Cavallo and Sandra Stefanovic-Barrett. I would like to express my sincere gratitude for their guidance and support throughout this thesis project. They have made this experience incredibly enriching.

I extend my thanks to the members of the Cell Engineering Team at AstraZeneca: Susanna Engberg, Neil Hattersley, Ella Quist, Markus Nordberg, Hyunsoo Park, Patricia Mendoza-Garcia, and Anja Will. Their expertise in cell engineering and assistance in troubleshooting have been instrumental whenever parts of the project have encountered challenges.

Special appreciation goes to Evangelia Anagnostou for acquainting me with laboratory procedures and analyses relevant to this thesis project. Additionally, I am grateful to the Protein Expression and Molecular Biology Team and the Cell Banking Team for providing essential reagents throughout the project. I also want to acknowledge Mei Ding for her support in identifying TF downstream target genes and tool compounds. Moreover, I am thankful for Riccardo Panero from the Target Identification and Validation Informatics Team for his support in bioinformatic analysis in the design of the sgRNAs.

I would also like to thank my examiner, Yvonne Nygård, for her valuable feedback and assistance with the administrative and academic prospects of my thesis work.

Finally, I would like to give a special thanks to my family and my friends for their encouragement and endless support throughout my years of studies and during my thesis work.

Clara Johnsson, Gothenburg, May 2024

List of Acronyms

The following is a list of acronyms used throughout this thesis in alphabetic order.

AR	Androgen receptor
ChIP	Chromatin immunoprecipitation
CMV	Cytomegalovirus
CRISPR	Clustered regularly interspaced short palindromic repeats
CRPC	Castration-resistant prostate cancer
crRNA	CRISPR RNA
DBD	DNA-binding domain
DBS	DNA-binding site
dCas9	Dead Cas9
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
EF-1α	Elongation factor 1-alpha
EMSA	Electrophoretic mobility shift assay
ER	Oestrogen receptor
FBS	Fetal bovine serum
G418	Geneticin
GFP	Green fluorescent protein
GPCR	G protein-coupled receptor
HEK-293	Human embryonic kidney 293
HTS	High-throughput screening
IDR	Intrinsically disordered region
IL-6	Interleukin 6
LAF	Laminar air flow
LB	Lysogeny broth
LBD	Ligand binding domain
LgBiT	Large Binary Technology, Large BiT
LLPS	Liquid-liquid phase separation
LNCaP	Lymph node carcinoma of the prostate
MoA	Mechanism of action
NanoBiT®	NanoLuc® Binary Technology
NLS	Nuclear localization signal
OD	Optical density
ORF	Open reading frame
PAM	Protospacer adjacent motif
PBM	Protein binding microarrays
PBS	Phosphate buffered saline

PIC	Pre-initiation complex
PoC	Proof of concept
POI	Protein of interest
PPE	Personal protective equipment
PPI	Protein-protein interaction
PROTAC	Proteolysis targeting chimera
PTM	Post-translational modifications
RH	Relative humidity
RLU	Relative light unit
RPMI 1640	Rosewell Park Memorial Institute 1640
SELEX	Systematic evolution of ligands by exponential enrichment
sgRNA	Single guide RNA
SHE	Safety, health, and environment
SmBiT	Small Binary Technology, Small BiT
STAT	Signal transducers and activators of transcription
STR	Short tandem repeat
TF	Transcription factor
TFBS	Transcription factor binding site
tracrRNA	Trans-activating CRISPR RNA
TSS	Transcription starting site

Table of Contents

1

Introduction

In this chapter, the thesis begins with an introductory overview. Initially, a project background is presented. Following this, are chapters outlining the aim, research questions, project scope, disposition of the thesis, and author contribution.

1.1 Project Background

Transcription factors (TFs) are DNA-binding proteins that can modulate gene expression by regulating the transcription rate of genes. Dysregulation of gene expression mediated by TFs is implicated in various diseases, with approximately 19% of the 1,600 TFs identified in the human genome associated with disease phenotypes [1, 2]. Despite substantial efforts, the targeting of these proteins has remained limited [1, 3]. Conventional approaches for modulating TF activity have primarily focused on compounds mimicking endogenous ligands, often lacking the desired specificity. Nevertheless, direct targeting of TFs could offer unique potential for highly specific modulation, surpassing the selectivity of commonly targeted upstream signalling proteins like G-protein-coupled receptors (GPCRs) or kinases. Furthermore, direct targeting of TFs holds promise for reducing toxicity and enhancing drug efficacy by selectively regulating transcription driven by specific TFs. This approach of regulation could potentially avoid interference with unrelated signalling pathways of the disease [1, 2]. Therefore, TFs constitutes a significant class of potential therapeutic targets. [1, 2, 4]

Traditionally, TFs have been deemed as “undruggable” by small molecule ligands due to their involvement in protein-protein interactions (PPIs) and lack of structural properties, including well-defined ligand binding pockets [1]. However, recent decades have witnessed progress in targeting TFs, revitalizing confidence in pursuing further research of this class of proteins. Various approaches to modify TFs activity have been demonstrated, both preclinically and, in some cases, clinically. This includes compounds inhibiting TF-cofactor PPIs and compounds inhibiting TF-DNA binding [5]. Furthermore, degrader molecules targeting TFs, such as monomeric degraders, molecular glues and PROTACs (proteolysis targeting chimera) are also being developed. [1]

To understand the intricate mechanisms of TFs and develop novel drugs targeting these proteins, it is important to study TFs and determine the direct binding of TFs to DNA. Several promising methods for studying protein-DNA interactions exist. A widely utilized *in vivo* technique is the antibody-based assay known as chromatin immunoprecipitation (ChIP). ChIP has proven to be a powerful analytical tool for investigating endogenous protein-DNA

interactions and identifying TFs DNA binding sites (DBSs) [6]. Moreover, techniques such as Protein Binding Microarrays (PBMs), Electrophoretic Mobility Shift Assay (EMSA) and Systematic Evolution of Ligands by Exponential Enrichment (SELEX) have all been proven useful in studying TF-DNA interactions *in vitro* [7, 8]. Despite the emergence of many promising methods to study TF-DNA interactions, several of these techniques have inherent limitations, and prior knowledge of expected DBSs and structural properties of TFs are often preferable. Furthermore, many techniques are not compatible for measuring TF-DNA interactions and assessing compounds effects in high throughput.

Successful targeting has been achieved for only a small fraction of the approximately 300 TFs associated with disease phenotypes. This challenge arises from various obstacles associated with targeting TFs, including the difficulty in directly monitoring TFs activity in cells and understanding their structural properties, such as the lack of well-defined ligand binding pockets and intrinsically disordered regions (IDRs) [1, 3]. Another bottleneck in targeting these proteins is predicting functional TF DBSs throughout the genome, as the genome-wide abundance of non-specific DBSs and the three-dimensional structure of chromatin can complicate this process [1]. Furthermore, the promoter regions of TFs are not always well-known, especially for poorly characterized TFs. In such cases, identifying and characterizing DBSs becomes crucial in comprehending the regulatory role of TFs [9]. Understanding the molecular modalities of highly dynamic interactions between TFs, genomic DNA, and their protein partners remains limited [9]. Consequently, establishing high throughput screening (HTS) systems for TFs, understanding their mechanism of action (MoA) and identifying novel drugs for this protein class, have proven challenging [1]. Understanding how TFs specifically bind to genomic DNA and how this specificity relates to the regulation of transcription are paramount in efforts to target TFs.

Due to their inherent properties, small molecules have been a dominant modality in drug discovery. With a low molecular weight and ability to penetrate cell membranes to reach targets intracellularly, small molecules constitute promising therapeutics. However, developing small molecule drugs is a complex and time-consuming process. HTS methods for drug candidates plays a vital role in expediting this process [10]. Presently, there are limited options for directly measuring TF-DNA interaction and compound binding to TFs, in high throughput and in complex *in vitro* and cellular environments. These systems are necessary to represent the disease in a relevant setting. In addition, indirect functional cellular responses can be influenced by interactions with co-regulators and other components in associated pathways, rather than directly impacting the intended target. The low success rates of these indirect cellular responses highlight the urgent need for disease-relevant screening assays. [11]

1.2 Aim

Currently, there exist a gap in available assays to assess the direct binding of TFs to DNA, particularly in high throughput in cells, which the pharmaceutical industry heavily relies on. The master's thesis project thereby aimed to develop a novel cell-based assay for measuring the binding of TFs to DNA. Furthermore, the study aimed to explore how the assay could be optimized for a 384-well plate configuration. Another key objective, if the time allowed, was to understand signal changes upon different compound treatments. The hypothesis was that using a split-reporter system, could facilitate high-throughput and high-sensitivity readouts of TF-DNA binding. Another potential outcome was that the search for TF binding regions, could be expedited.

1.3 Research Questions

The overarching goal of this project was to develop an assay for studying TF-DNA interactions in cells. In pursuit of the aims of this study, four research questions were outlined:

1. Can stable cell lines expressing dead Cas9 (dCas9) tagged with Large BiT (dCas9-LgBiT) be established?
2. Can a cell-based NanoBiT® assay accurately measure TF-DNA interactions with high specificity?
3. After establishing proof of concept (PoC), can the assay be optimized for a 384-well plate format?
4. Will the assay allow sensitive detection of small molecule modulation on TF-DNA interactions?

1.4 Project Scope

This project was defined by the Cell Engineering- and Cellular Assay Development Teams at AstraZeneca. The project was scoped to address the existing gap in available DNA-binding assays for HTS of TFs. To achieve this, the split-reporter system NanoBiT® was employed to develop a novel cell-based assay. Various cell models were transfected with plasmids expressing dCas9 tagged with a LgBiT, plasmids expressing TFs tagged with a Small BiT (SmBiT), and single guide RNA (sgRNA). In the assay, synthetic sgRNA will guide the dCas9-LgBiT to the region where the TF binds. If the dCas9-LgBiT and TF-SmBiT complexes get in proximity, an active enzyme will form, producing a bright luminescent signal that can be quantified. Additionally, the TFs can be treated with various compounds, and changes in the signal can be detected. The project was limited and focused on investigating two disease implicated TFs, androgen receptor (AR) and signal transducers and activators of transcription 3 (STAT3). To create relevant disease models for each TF, two different cell lines were used: LNCaP for AR and HEK-293 for STAT3. Well-characterized TFs were selected for this project to establish PoC, optimize the assay for a 384-well plate format and assess the ability of small molecules to modulate TF-DNA interactions.

1.5 Disposition of the Thesis

The thesis comprises six themed chapters. After the introduction, chapter two initiates by delineating the theoretical framework of the research. Thereafter, the third chapter delves into the methodology employed throughout this project, followed by section four presenting the study's results. Chapter five, the discussion, deliberates and analyzes the projects findings. Finally, the sixth chapter encapsulates the thesis conclusion. Additional materials, supplementing the thesis, are available in the appendices.

1.6 Author Contribution

This project was conducted in the Cell Engineering Team within the Cell Sciences Department at AstraZeneca. The project was supervised by associate principal scientists Anna-Lina Cavallo and Sandra Stefanovic-Barrett. Before the project start and the authors engagement in the project, the dCas9-LgBiT plasmids had been designed by Anna-Lina Cavallo. These plasmids were then generated in-house by the Protein Expression and Molecular Biology Team. Furthermore, the different TF-SmBiT plasmids were designed by Anna-Lina Cavallo and outsourced for generation by the company VectorBuilder. The cell lines used throughout the project were provided by the Cell Banking Teams at AstraZeneca. Additionally, ranges of sgRNAs for each TF were designed with support from the Target Identification and Validation Informatics Team within AstraZeneca.

The experiments throughout this project were collaboratively planned by the author of this thesis and the supervisors. The laboratory work was conducted by the author who played a key role in producing project reagents, involving the expansion and purification of plasmids, cell culturing, and generating genetically engineered cell lines. Additionally, the author conducted tests to validate the correctness and functionality of the constructs. In the second part of the project, the author's focus shifted towards developing and optimizing an assay to measure TF-DNA binding. The projects results were analyzed and interpreted by the author, with support from the supervisors.

2

Theory

In this chapter, a theoretical background to the master's thesis project is presented. An extensive background to TFs is described followed by the concepts and methods underpinning this study to facilitate the comprehension of the results. The chapter is divided into two parts where the theoretical background will be presented first, followed by the experimental theory.

2.1 Theoretical Background

2.1.1 TFs Function in Eukaryotes

Transcription and translation constitute fundamental processes in which cells transcribe genetic information from DNA into mRNA, followed by the translation of mRNA into proteins. The regulation of transcription serves as a crucial mechanism, enabling cells to respond to signals, establish and sustain cell identity during development, and coordinate overall cellular activity [12]. In this context, TFs have a pivotal role and represent one of the most significant classes of proteins [1, 7]. One key function of TFs is to selectively engage the transcriptional regulatory machinery at specific genomic locations. The targeted binding of TFs to DNA allows for the precise control and modulation of the expression of nearby genes [1]. By recruiting various domains including co-activators/repressors, chromatin remodelers, histone acetylases, deacetylases, kinases, and methylases, and RNA polymerase II to target genes, TFs can regulate transcription [1, 2, 7, 13]. Depending on the recruited domains, TFs can function as either transcriptional activators or repressors [1]. Moreover, TFs play extensive roles in various signalling pathways, both intrinsically and extrinsically, thereby influencing cellular pathways such as growth, proliferation, metabolism, apoptosis, immune responses, and differentiation. [1, 3, 7]

TFs can be categorized into two groups: general and specific TFs. General TFs form the pre-initiation complex (PIC), initiating mRNA synthesis [14]. The PIC involves collaboration between RNA polymerase II and the general TFs: TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH [12, 14]. Specific TFs participate in the transcription elongation process and represent the majority of TFs, with their DBSs being more diverse. While general TFs exhibit highly conserved expression across all cell types, specific TFs show more variability in expression across different cell types [14]. A central paradigm asserts that TFs interact with regulatory elements to selectively govern the activity states of target promoters based on cell type [15]. Furthermore, TFs have been demonstrated to play a crucial role in differentiating stem cells into various cell types and directly reprogramming somatic cells into pluripotent stem cells. [2, 16]

The TF-DNA interaction can be either specific or non-specific. Specific interactions involve non-covalent bonds, mainly hydrogen bonds or Van der Waals forces, between TFs and specific DNA sites. The non-specific interactions are mediated by electrostatic interactions between TFs and the negatively charged DNA phosphate backbone [17]. Furthermore, TFs typically function as homodimers or multimers rather than in a monomeric state, and besides interacting with DNA independently, TFs commonly interact through conformational changes of DNA and as heterodimers with other proteins through PPIs. [2]

2.1.2 Anatomy of a TF

A TF comprises two distinctive components: a DNA-binding domain (DBD) and an effector domain, as illustrated in Figure 1. The DBDs role is to identify a specific genomic locus and subsequently bind to the DNA. This interaction between TFs and DNA is fundamental for regulation, gene organization, and the cells state [18]. The specific DNA sequences recognized by TFs are typically short, ranging between 4-20 base pairs in length [19]. Given the small and flexible nature of most TF DBSs, a human gene (typically exceeding 20 kb) may contain several potential binding sites [7]. A fundamental challenge for targeting TFs is locating DBSs within a vast array of non-specific DNA sequences [17]. The compaction of DNA into nucleosomes and chromatin allows the two-meter-long genome to fit into the nucleus while maintaining accessibility of genes and their regulatory elements within the chromatin landscape [18]. Gaining an understanding of how motifs are accessed in chromatin and the resulting transcriptional responses is crucial [20, 21]. Moreover, determining the TF DNA binding motifs is a crucial step in providing information about the TFs function.

The effector domain of TFs exhibits transcriptional activation potential by recruiting various collaborators [1, 2, 13]. The domain recruits various entities to form the transcriptional complex, including chromatin remodeling enzymes, co-activators/repressors, and the general transcriptional machinery. Transcriptional activators utilize a transactivation domain to recruit chromatin remodeling enzymes, histone-modifying enzymes, transcriptional co-activators, and other general TFs. These components synergistically enhance the accessibility of target genes, epigenetically marking them as active, and facilitating the recruitment of RNA polymerase II. Conversely, TFs acting as repressors employ a transrepression domain to recruit chromatin remodeling and epigenic enzymes, thereby attenuating the accessibility of target genes and marking them as inactive [1]. In addition, transrepression can involve more diverse mechanisms such as steric hindrance, roadblock, deformation, anti-activation, and promoter clearance [2]. Furthermore, TFs can contain regulatory domains, encompassing dimerization, nuclear transport, and autoinhibition domains, which further contributes to the complexity of the transcriptional complex. [1]

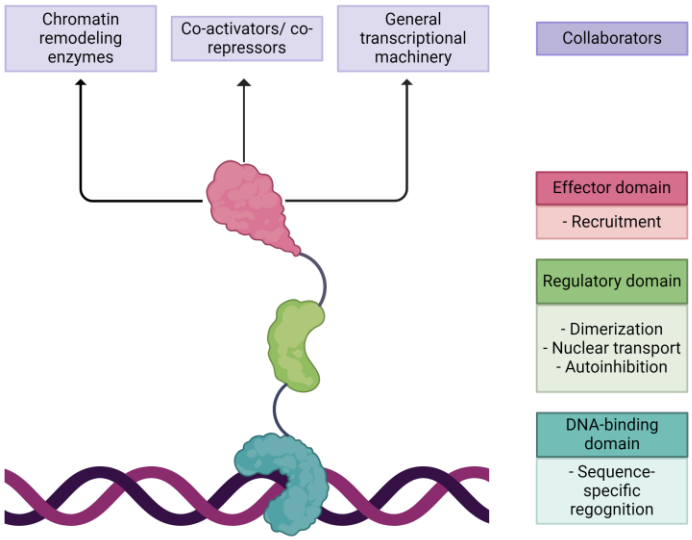


Figure 1: The structure of a TF comprises a DBD and an effector domain, enabling sequence-specific recognition and recruitment of various collaborative molecules. Additionally, many TFs also feature regulatory domains, which *e.g.*, regulates functional activity and the localization of TFs. Adapted from [1]. Created with BioRender.com.

TFs in eukaryotes exhibit diverse characteristics and functions, leading to their classification into several classes. A few primary classes include C2H2 zinc fingers, homeodomains, basic-helix-loop-helixes, basic leucine zippers, and nuclear hormone receptors [2, 7, 22]. Among these, nuclear receptors play a pivotal role in embryonic development, the maintenance of differentiated cellular phenotypes, metabolic processes, and apoptotic pathways [23]. Notably, nuclear receptors are considered the most druggable group of TFs due to their possession of a ligand binding domain (LBD), enabling them to enlist the transcriptional apparatus upon binding to a small-molecule ligand. Examples of nuclear receptors include oestrogen receptor (ER) implicated in breast cancer and AR linked to prostate cancer [24]. TFs including *e.g.*, nuclear receptors and the STAT family possess nuclear localization signals (NLS). These signals are recognized by importins, which are proteins involved in nuclear import, and facilitate the transport of the TFs into the nucleus. For instance, STAT3, associated with various cancers, inflammatory, and autoimmune diseases, features a domain known as SH2 [25]. This SH2 domain plays a crucial role in localizing the TF into the nucleus by regulating homo- and heterodimerization with other STAT TFs. [1]

2.1.3 TFs Regulatory Role in Gene Expression

Regulating transcription is a fundamental process for controlling and modulating gene expression [26]. TF-mediated gene regulation in eukaryotic cells is a highly intricate mechanism, given the transcription of numerous genes into protein-coding and noncoding RNA. The human genome, for instance, is estimated to contain 20,000 protein-coding genes and a similar number of non-coding genes. While protein-coding genes are well understood, the functions of non-coding genes, many of which modulate transcriptional or post-transcriptional processes, remain less known [12]. There are indications suggesting deep conservation of TFs, with similar

regulatory regions, and physiological roles, despite a high turnover in individual sequences. Moreover, a single TF can regulate several genes in different cell types, highlighting the dynamic nature of regulatory networks within the same organisms. [7]

The functionality of a TF at a specific DBS depends on numerous factors, including thermodynamic stability, DNA/chromatin accessibility, TF interaction with co-activators, and the kinetics of TF-DNA binding [1]. The TFs activity can be regulated by modulating its abundance in the active form or the accessibility of its binding sites [27]. Additionally, gene expression can be regulated through various cis-regulatory components such as core promoters, promoter-proximal elements, and elements located at a greater distance from the transcription start site (TSS), including enhancers, silencers, insulators, and tethering elements [28]. A TFs promoter region typically contains a TSS and a TATA-box, facilitating the formation of the PIC and activating RNA polymerase II [1]. In contrast, enhancers lack a TSS and can be situated several kilobase pairs away from a TSSs. Conversely, they bind to TFs and activate transcription when positioned in proximity to a promoter. Due to this distance disparity, understanding how gene regulation is influenced by enhancers is an area of intense study [1]. Moreover, DNA loop formation and maintenance is crucial for proper gene control, occurring during transcription initiation between enhancers and core promoter elements, stabilized by cohesin. The recruitment of chromatin remodelers is vital for mobilizing nucleosomes and facilitating the access of the transcription apparatus to DNA [4]. Figure 2 below provides an illustration of transcriptional regulation in eukaryotic cells.

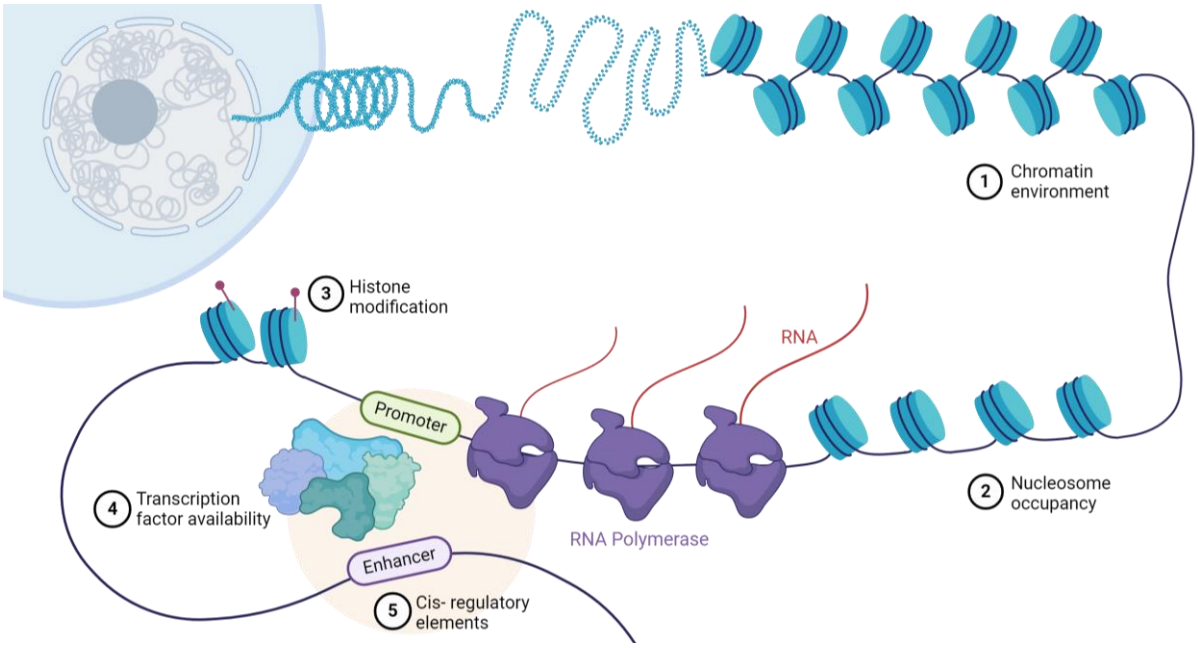


Figure 2: Illustration of the regulation process of transcription in eukaryotic cells. The figure captures how the process is influenced by the 1. chromatin environment, 2. nucleosome occupancy, 3. Histone modifications, 4. transcription factor availability and 5. cis-regulatory elements, such as enhancers. Adapted from "Regulation of Transcription in Eukaryotic Cells", by BioRender.com (2024). Retrieved from https://app.biorender.com/biorender-templates.

Regulation of TF activity can also be achieved through post-translational modifications (PTMs). Similar to other proteins, TFs are affected by PTMs, influencing their activity, cellular location, and interaction with other proteins. Several types of PTMs, such as phosphorylation, ubiquitination, sumoylation, and acetylation, have been recognized to affect various aspects of TF functionality. The impact of these modifications heavily relies on the protein sequence of TFs, contributing to an extra level of complexity to the control of gene expression by TFs [8]. Furthermore, TFs consists of both structured DBDs and long IDRs [29]. These IDRs are embedded with motifs controlling transcription. Although these regions were previously overlooked in standard structure-function paradigms, recent insights suggest that they play a crucial role in transcriptional regulation. Another behaviour identified in gene regulation is liquid-liquid phase separation (LLPS). The hypothesis of LLPS proposes that substantial levels of TF binding at super-enhancers may result in elevated concentrations of cofactors and other constituents of the transcriptional machinery, culminating in the formation of transcriptional condensates. A single proteins transitions into and out of a condensate can be facilitated by PTMs altering IDR properties. The formation of these transcriptional condensates has shown to influence and regulate gene expression, although exactly how is not fully understood and it has become an area of intense research. [1]

2.1.4 Dysregulation of TFs

Dysregulation of gene expression mediated by TFs, that lead to abnormal gene expression, are increasingly being linked to various human diseases [4, 5, 7, 30]. The occurrence of mutations affecting TFs and their DBSs can result in the misregulation of gene expression, contributing to the onset of various diseases [2]. Several types of gene expression dysregulations exist, including upregulation, downregulation, and alterations in target genes, all of which can contribute to disease pathology. Upregulation may involve amplification, gain of function, and pathway overactivation. Downregulation can result from loss of function and overactivation of repressors. Additionally, changes in target genes may encompass shifts in chromatin architecture, gene translocations, and fusion TFs [1]. Given the direct involvement of TFs in transcriptional dysregulation and disease causation, they are considered appealing therapeutic targets for disrupting disease processes. [1, 3]

Aberrant TF activity, often resulting from elevated levels of translocations, gene amplification or deletion, and point mutations, plays a significant role in several types of cancers [3, 5]. The dysregulation of transcription is considered one of the hallmarks of cancer, with TFs acting as fundamental catalysts of oncogenic transformation, proliferation, and survival [1, 5]. Although chemotherapies have been longstanding cornerstones in anticancer treatments, they can be highly toxic, leading to numerous side effects and there is a risk of relapse with resistant tumours. This development of resistance against drugs is a major challenge in todays cancer treatments [9]. Given that certain TFs are more active in human cancer cells, TFs have become appealing targets in the development of anticancer drugs. A small set of TFs with wide

expression is indicated to drive tumour growth, with tumour-driving master TFs typically expressed in a limited number of cell types [31]. Compared to the numerous oncogenes in different signalling pathways, the number of oncogenic TFs is limited, making them potentially easier to target. Moreover, modulating oncogenic TFs activity with small molecules has shown immense potential for some TFs. [5]

Apart from TFs roles in various cancers, transcriptional dysregulation is implicated as a contributing factor in many other diseases, including autoimmune and inflammatory diseases, diabetes, and cardiovascular diseases, among others. In the context of the immune response to infection, TFs are often activated at the end of signalling pathways, and their involvement is crucial in mediating immune responses. In the context of diabetes, the condition typically involves concurrent alterations in numerous genes. Nonetheless, certain types of monogenic diabetes have been associated with mutations in individual TFs. For instance, activating STAT3 mutations have been identified as linked to early-onset diabetes. Similarly, cardiovascular diseases encompass a spectrum of conditions influenced by multiple interconnected genetic risk factors, rather than being solely driven by singular proteins. TFs are indispensable for both the formation and maintenance of the cardiovascular system, thus playing crucial roles in specific cardiovascular conditions. Furthermore, loss-of-function mutations in particular TFs have been recognized as causative factors in various cardiovascular diseases. [1, 4]

2.1.5 Progress in Targeting TFs

Various approaches to modulate TF activity have been demonstrated in preclinical studies, and in certain instances, in clinical settings as well. These include inhibiting TF-cofactor PPIs, inhibiting TF-DNA binding, and modulating TF activity by altering ubiquitylation levels and subsequent proteasome degradation or inhibiting regulators of TF expression. Emerging approaches include modulation of auto-inhibition, PROTACs, use of cysteine-reactive inhibitors, targeting IDRs of TFs, and combining TF inhibitors with kinase inhibitors to block resistance development. [5]

The potential of TFs as targets for cancer therapy was first highlighted by James Darnell in an article published in 2002 [32]. The most successful small molecules that have been developed to target oncogenic TFs are the AR and ER antagonists. Several second-generation AR antagonists have been clinically approved since 2018, including enzalutamide, apalutamide and darolutamide, to treat patients with castration-resistant prostate cancer (CRPC). Furthermore, ER modulators including tamoxifen, raloxifene and toremifene are well-established endocrine therapies. Utilizing small molecules to treat diseases caused by TF abnormalities in expression levels is appealing due to TFs crucial importance in various diseases [33]. The aforementioned targeted TFs, AR and ER are nuclear receptors which possess a LBDs and can bind to a small-molecule modulator, altering its activity to regulate gene expression. The advantage of nuclear receptors lies in having a LBD where a small molecule already binds, facilitating agent

development. By far, this have been the most successfully group of TFs to target [5]. When a ligand binds to the LBD, it usually triggers activation of the receptor. This process exposes a hydrophobic groove for co-activator binding, thereby increasing the receptors affinity for nuclear localization factors and other nuclear receptor molecules. This activation allows multiple forms of modulation: agonism, antagonism, and inverse agonism for individual receptors, providing diverse approaches to modulate aberrant transcriptional programs in drug discovery efforts. [1]

Additionally, degraders molecules targeting TFs have been developed including PROTACs, molecular glues, and monomeric degraders [1]. The intracellular ubiquitin-proteasome system is a highly controlled process in cells where misfolded proteins can be destructed. PROTACs were developed by hijacking E3 ligase, involved in this ubiquitin-proteasome machinery, to degrade a protein of interest (POI) [33]. PROTACs are ligands binding to the POI, linked to a ligand recruiting ubiquitin E3 ligase. This initiates a neo-interaction between the target protein and E3 ligase, ultimately resulting in the degradation of the target protein [1]. The potential of PROTACs lies in their ability to bind to a protein and modulate its function through degradation [33]. Moreover, molecular glues and monomeric degraders represent a similar approach, in which all are based on degradation of targeting proteins. [1]

PTMs such as ubiquitination, hydroxylation, methylation, acetylation, and phosphorylation offer additional avenues for targeting TFs [34]. Targeting TF PTMs could effectively alter DNA binding activity and cofactor interactions. PTMs include phosphorylation of serine, threonine, and tyrosine, methylation of lysine and arginine, acetylation of lysine, and ubiquitylation, sumoylation, and ADP ribosylation. Another potential strategy involves targeting IDRs of TFs. Although traditionally ruled out due to lack of structure and binding pockets, recent analysis suggests that folded IDRs bound to partners have a higher proportion of potential cavities for small molecules to bind. Their inherent flexibility may enable conformational adjustments to complement small molecules more easily [5]. Furthermore, phase-separated condensates, which compartmentalize biochemical reactions within cells, could offer valuable insights into targeting TFs [35]. Studies show that drug candidates targeting TFs are more likely to enter transcriptional phase-separated condensates. Understanding how partitioning into condensates affects the ligands ability to reach the target could be helpful when developing novel drugs targeting TFs.

2.1.6 Advancements in Studying TFs

Studying TFs has generated significant interest, driven by the desire to explore potential applications for this group of proteins. The development of novel technologies for identifying and characterizing TFs is crucial, and various cross-cutting areas should be explored to fully understand the versatility and potential of TFs as drug targets [2]. To fully comprehend the function of a TF, determining the DBSs is usually the first step, as it provides a gateway for

further examination [7]. Even though TFs can regulate transcription of genes through various mechanisms, most of them relies on their interactions with their DBSs [2]. As a single TF can converge multiple signalling pathways with few possible targets, understanding the DNA binding of TFs is dependent on factors such as the DNA sequence, TF concentration, chromatin accessibility, and cofactors [36]. To comprehensively study TFs and target all classes of TFs, it is essential to develop methods for measuring the binding of TFs and DNA in cells. [8]

A plethora of innovative techniques for studying and characterizing TF-DNA interactions have been developed. These include both *in vivo*, *in vitro*, and *in silico* techniques, and no single universal solution may address all challenges. An approach considering trade-offs based on specific requirements may be the most pragmatic. *In vivo* technologies focus on studying TF-DNA interactions in their natural context but have limitations in achieving substantial enrichment of binding fragments and conducting biophysical quantification and characterization of protein-DNA interactions. *In vivo* assays encapsulate the collective impacts of variations in both TF and DNA attributes, influenced by factors like non-coding RNA. Furthermore, *in vivo* methodologies offer distinct advantages, as it provides information directly related to binding occurrences in cells, contributing to a better understanding of transcription regulation mechanisms. In contrast, *in vitro* methods offer valuable insights but might not fully reflect the true regulatory role of TFs when studied solely *in vitro* or in heterologous systems. *In vitro* systems involve simplified settings and study TF-DNA interactions between TFs and synthesized DNA molecules. [2]

ChIP is a promising and useful *in vivo* method for studying TF-DNA interactions, allowing the identification of TF DBSs throughout the genome [37]. This technique is based on a cross-linking reaction, forming covalent bonds between TF and bound DNA. Next, the chromatin is sheared into protein-DNA fragments of certain lengths. A TF specific antibody is then added and incubated along with the sheared chromatin. Then the cross-linkage is reversed, and the DNA is purified [2]. The enriched fragments of DNA can then be detected using *e.g.,* qPCR microarray (ChIP-chip) or sequencing (ChIP-seq) [7]. ChIP allows for studying TF-DNA binding in the endogenous environment. For example, the technique has been used to identify many TFs DBSs in yeast and the STAT3 DBSs in human B cells. [2]

The aforementioned *in vivo* technique comes with several advantages. However, there exist challenges of achieving substantial enrichment of binding fragments and capturing the quantitative nature of TF-DNA binding events. To quantitatively assess TF-DNA binding, relevant *in vitro*-based assays can offer valuable insight [2, 8]. PBMs, EMSA, and SELEX, can all be useful to study TFs binding to DNA [7, 8]. PBM is an experimental method where a purified and epitope-tagged TF can bind directly to a double-stranded DNA microarray. The protein-bound array is then labelled with a fluorophore-conjugated antibody that is specific to the epitope. The technique provides signals that can be quantified, and which are relative to the amounts of protein bound to each of the probe sequences on the array [38]. EMSA is a technique where the binding of a target protein to DNA can be evaluated through observing shifts in the

electrophoretic migration of DNA [7]. Lastly, SELEX-based methods are based on the incubation of a purified TF with a DNA pool which contains a multitude of randomized sequences. The oligos that bound to the TF are selected and amplified using PCR. The oligos are then re-incubated repeatedly in order to identify high-affinity binders or determine the TF DBSs. [7, 8]

In the era of increasing biodata, *in silico* methods based on machine learning and artificial intelligence are gaining prominence for solving biological problems. These methods could be useful in predicting TF DBSs and lead to increasing hit rates for PPIs such as TF-DNA interactions. [9, 39]

2.1.7 Current Limitations Associated with Targeting TFs

TFs absence of enzymatic activity and easily identifiable binding pockets for small molecules, make the development of chemical probes challenging [3]. Additionally, many TFs contain a significant degree of IDRs, complicating targeting efforts that heavily rely on structural information for ligand optimization. Modulating TF activity with small molecules is an appealing therapeutic goal. Yet, outside nuclear receptors, there are few drugs or thoroughly validated chemical probes that directly target TFs [1]. Furthermore, while some small-molecule degraders can efficiently target TFs, certain proteins may be resistant to this approach. [40]

Despite extensive studies of the DBDs of TFs, predicting functional TF binding sites remains challenging due to difficulties in understanding three-dimensional chromatin architecture and the abundance of non-specific binding sites. Studying TFs is challenging due to this chromatin structures covering a larger fraction of DNA, hindering protein binding to regulatory sequences. Additionally, when purifying these proteins, they are intrinsically disordered, making them hard to study by X-ray crystallography and cryogenic electron microscopy. [1, 21]

Despite some advances, there is still a gap in measuring the direct binding of TFs to DNA, in high-throughput, and understanding how they translocate and bind to DNA in live cells [1]. To address these challenges, there is a pressing demand for cellular assays that assess direct target engagement within disease-relevant contexts, particularly in high-throughput settings. Furthermore, this could be important in facilitating a more predictive translation into clinical efficacy. [11]

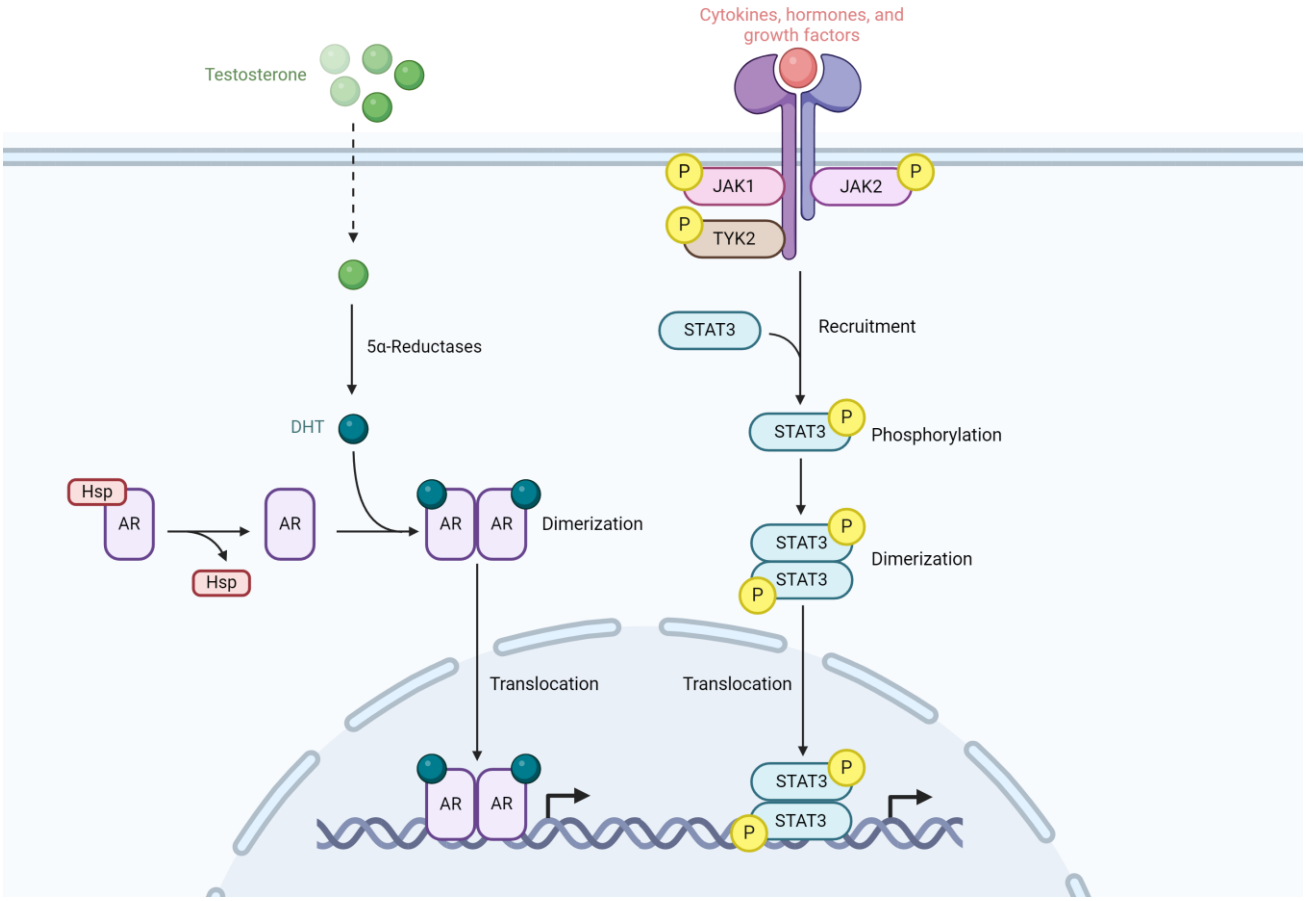
2.1.8 AR and STAT3

AR is a ligand-binding TF belonging to the family of steroid hormone nuclear receptors [41]. Upon ligand binding, a NLS escorts AR into the nucleus, where it exerts its transcriptional regulatory functions. Conversely, upon ligand withdrawal, a nuclear export signal facilitates the export of AR to the cytoplasm [42]. AR orchestrates pivotal roles across diverse cellular systems, governing the development and maintenance of the immune, cardiovascular, neural, reproductive, musculoskeletal, and hematopoietic systems. Particularly significant is its

mediation of male sex androgens effects, such as testosterone and dihydrotestosterone, through AR signalling, critical for the development of the male reproductive system and secondary sexual characteristics.

Maintaining precise AR signalling is imperative, as aberrations in AR function may underlie tumorigenesis in various organs, including the prostate, bladder, liver, kidneys, and lungs [41]. Notably, AR stands out as a key driver in prostate cancer, underscoring the importance of unravelling its mechanisms and devising targeted therapies to advance prostate cancer treatment [11]. Prostate cancer stands as the most prevalent cancer in men, with AR playing a pivotal role in its pathogenesis. Current therapeutic strategies often revolve around depriving AR of testosterone, supplemented with AR antagonists to achieve AR blockade. However, a formidable challenge in prostate cancer management lies in the emergence of resistance to antiandrogen therapies [42, 43]. Notably, in 2019, enzalutamide, an AR antagonist effective in CRPC treatment, received FDA approval. Enzalutamide competitively inhibits androgen binding to the TF receptor, offering a promising avenue in combating advanced prostate cancer. [2]

STAT TFs are activated through tyrosine phosphorylation. This phosphorylation event is succeeded by dimerization through reciprocal SH2 phosphotyrosine interaction. Upon dimerization, STAT proteins bind to importin and translocate into the nucleus, where they induce the transcription of specific genes. The STAT family have crucial roles in mediating cellular responses to extracellular signals, particularly cytokines and growth factors [25]. These TFs also play pivotal roles in various physiological processes and are implicated in numerous diseases, including autoimmune disorders, inflammatory conditions, cancer, and diabetes. Overactivation of STAT activity is often associated with autoimmune diseases, while activating mutations in STAT3 have been linked to early-onset type 1 diabetes, Crohn's disease, psoriasis, and multiple sclerosis. Conversely, inactivating STAT members can lead to immunodeficiencies and heightened susceptibility to infections. Targeting STAT proteins has emerged as a promising therapeutic strategy for intervening in diseases characterized by dysregulated STAT activity [1]. Figure 3 illustrates examples of signalling pathways of the two TFs, AR and STAT3.



DHT: Dihydrotestosterone, **Hsp**: Heat-shock proteins, **JAK:** Janus kinase, **TYK:** Tyrosine kinase, **P:** Phosphorylation

Figure 3: Examples of signalling pathways of the two TFs: STAT3 and AR. The TFs undergo dimerization and expose NLS on the protein complex, facilitating their translocation into the nucleus. Adapted from [44, 45]. Created with BioRender.com.

2.2 Experimental Theory

2.2.1 Plasmid Design

Figure 4 illustrate the different vectors designed for this project. More detailed plasmid maps are provided in Appendix A. The first vector contains sequences encoding the CMV (cytomegalovirus) promoter and dCas9 in connection to either N- or C-terminus of the LgBiT. These plasmids also include a neomycin resistance cassette, to enable later selection of the cells expressing the plasmid. Furthermore, the figure shows the design of two vectors expressing the elongation factor 1-alpha (EF-1α) promoter and distinct TFs tagged with a SmBiT. These plasmids included a marker encoding puromycin resistance.

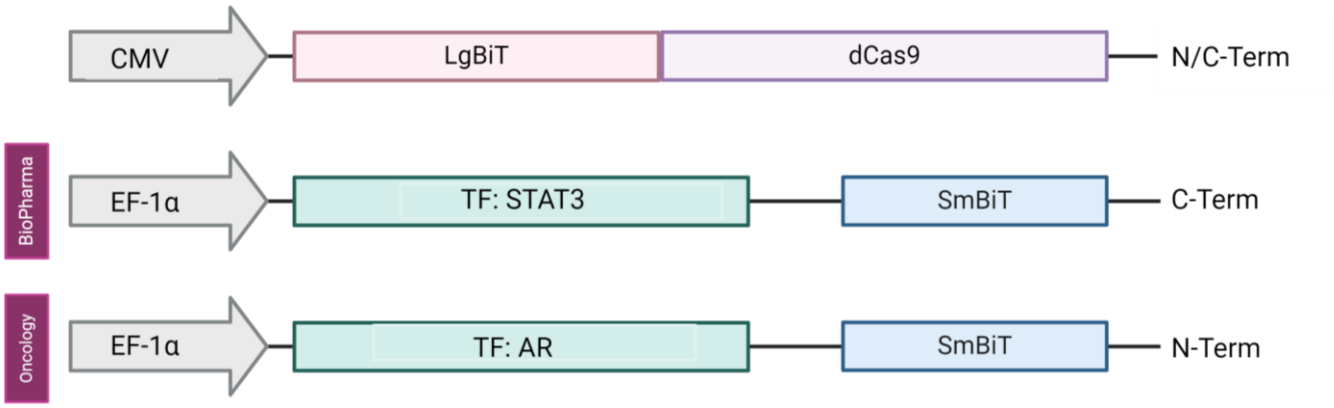


Figure 4: In total 4 different plasmids were designed for this project. The first plasmids encode the CMV promoter and the dCas9 in connection to either N- or C-terminus of the LgBiT. The other two plasmids encode the EF-1alpha promoter, distinct TFs and a SmBiT. Created with BioRender.com.

2.2.2 *In Vitro* Cell Models

Limited options exist for directly measuring the binding of TFs to DNA within a complex, *in vitro,* and cellular environment. Such systems are crucial for accurately representing diseases in relevant settings [11]. In this project, *in vitro* cell models will be employed because they provide natural DNA context and cofactors, and the transcriptional regulatory machinery are available. More specifically, immortalized cell lines will be used. Immortalized cell lines are derived from either tumorous cells or genetically modified cells, allowing for continuous propagation in culture and indefinite proliferation.

The Lymph Node Carcinoma of the Prostate (LNCaP) cell line was initially documented by Horoszewicz *et.al.* It originates from a metastatic lymph node lesion of human prostate cancer and is characterized by its adherence and epithelial nature, along with being AR-positive, indicating it exhibit AR-sensitive growth. Few well-characterized cell models are available for preclinical studies of prostate cancer that effectively recapitulate the disease. Nonetheless, LNCaP stands out as a commonly used *in vitro* model in prostate cancer research. [46]

Established in 1973, Human Embryonic Kidney 293 cells (HEK-293) are adherent and epithelial cells derived from primary embryonic kidney cell cultures of an aborted embryo. They were created by introducing DNA fragments of adenovirus type 5 into these cultures. HEK-293 cells have since become widely recognized and extensively utilized in cell biology and biotechnology research. [47]

2.2.3 Stable and Transient Transfection

Transfection of eukaryotic cells serves as a potent analytical tool for the incorporation of foreign DNA. This technique allows for the modification of the cells genetic content, facilitating the study of cellular processes, disease mechanisms, and gene function and expression in cells [48]. There exist three main types of transfections: biological, chemical, and physical, as shown in Figure 5. Additionally, transfection can be categorized as stable or transient. Stable transfection refers to the continuous and long-term expression of a transgene, which is accomplished by either integrating the DNA into the host genome or keeping it within an episomal vector in the host nucleus. This allows the transgene to be constitutively expressed as the cells proliferate. Conversely, transient transfection does not integrate into the host genome and is expressed for a limited time, eventually being lost as the cells proliferate [49]. In this project, cells will be transfected with the dCas9-LgBiT-N/C-terminus plasmids to generate stable expression of the plasmids. The stable transfection is preferable over the transient transfection since they provide a higher run-to-run reproducibility which could be critical when developing a HTS assay [50]. For the TF-SmBiT-N/C-terminus constructs and sgRNAs, transient transfection of the stable cell lines will be employed.

Electroporation, a widely employed gene edit methodology for both transient and stable transfections, involves the use of high-voltage electric shocks, creating temporary pores in cell membranes, enabling random integration of DNA and other substances into the cells. This physical transfection method is compatible with various cell types and is valued for its simplicity, requiring few steps [51]. Electroporation will be utilized in this project to perform transfections and generate both stably and transiently transfected cells.

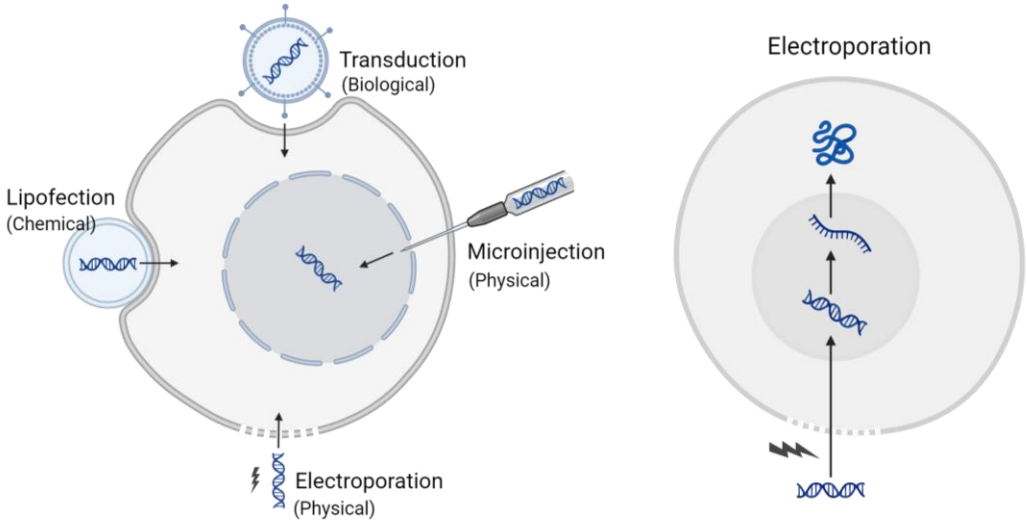


Figure 5: Image illustrating different transfection approaches: biological, chemical, and physical. Created with BioRender.com.

2.2.4 Western Blotting

Western blotting is a widely used technique for identifying proteins within complex mixtures. It involves three main steps: separation by size through gel electrophoresis, transfer to a solid support, and labelling the POI using primary and secondary antibodies. During gel electrophoresis, proteins are separated based on their molecular weight. The separated proteins are then transferred to a membrane, generating individual bands for each protein. Next, the membrane is incubated with antibodies specific to the POI, and any unbound antibodies are washed away. The antibodies bound to the POI are detected by developing a film. Ideally, only one band representing the POI should be visible, and the size of the protein can be determined using a protein standard. [52]

2.2.5 Cell-Based Assay

2.2.5.1 dCas9 and sgRNAs

The field of genome engineering has undergone a revolution with the advent of the clustered regularly interspaced short palindromic repeats (CRISPR) and the Cas9 endonuclease system. Originating from the bacteria *Streptococcus pyogenes*, where it serves as a defence mechanism against infections caused by viruses, this system can be programmed by single guide RNA (sgRNA) to induce a double-stranded break in the DNA at a genomic locus following a protospacer adjacent motif (PAM) sequence [53]. Extensively utilized, the CRISPR/Cas9 system has been adapted and engineered for various applications. A mutated form of the wild-type Cas9 protein, known as dead Cas9 or dCas9, lacks endonuclease activity, rendering it incapable of cleaving DNA. Unlike its active counterpart, dCas9 is employed for binding to DNA and modulating gene expression without introducing a double-stranded break. Guided by sgRNA and

recognizing the PAM sequence, dCas9 can be harnessed for various applications [54]. CRISPR activation is one such method in which dCas9-sgRNA complexes bind to DNA to initiate transactivation of genes. On the contrary, CRISPR interference, can cause steric hindrance that halts transcription elongation and represses the expression of the target gene [55]. In this project, dCas9 was not used to modulate gene expression but was utilized to bring the LgBiT subunit in proximity of the TFBSs.

Within the CRISPR/Cas9 system, sgRNA plays a pivotal role as it guides the Cas9 protein to specific sequences of the DNA. sgRNA is a chimeric molecule composed of crRNA (CRISPR RNA) and tracrRNA (Trans-activating CRISPR RNA), linked together by a loop. The spacer region, a 17-20 nucleotide sequence, is complementary to the promoter region of the target gene. The scaffold region, housing tracrRNA, acts as a binding scaffold for the dCas9 protein. By selecting or synthesizing appropriate sgRNAs, the dCas9-sgRNA complexes can be guided to the target gene, facilitating the regulation of gene expression. For activation, sgRNAs are typically designed to target regions approximately -400 to -50 base pairs upstream of the TSS, while for repression, they are designed for regions -50 to +300 base pairs from the TSS [54]. In this assay, ranges of synthetic sgRNAs will be designed and synthesized to guide the dCas9-LgBiT in proximity to the TFBSs, while ensuring that the dCas9 protein does not block and interfere with the TF-DNA binding.

2.2.5.2 Split-Reporter System NanoBiT®

NanoBiT® Promega employs a split-reporter system to tag proteins and monitor molecular interactions within living cells. This system utilizes NanoLuc®, which is an engineered luciferase derived from deep-sea luminous shrimp. The luciferase comprises two subunits: the Large BiT (LgBiT) and the Small BiT (SmBiT). The SmBiT is a 11 amino acid subunit, which together with the 17.6 kDa LgBiT subunit, forms an active enzyme. When two target proteins tagged with these subunits are in close proximity, and a substrate is added, the subunits form the active enzyme, producing a bright luminescent signal, that can be detected. The high dissociation constant, K_D=190 μM, and low catalytic activity of the NanoBiT® complementation reporter system subunits, results in that the subunits need to be brought into close proximity to reassemble the full-length NanoLuc® active enzyme [56]. The split-reporter systems schematic view and its application in this study is illustrated in Figure 6. Detection of TF-DNA binding in this assay will be enabled by pairing split-nanoluciferase with various TFs and dCas9, guided by synthetic sgRNAs. The dCas9, incapable of cleaving DNA, will bind specifically to the DNA along with the sgRNA. The TFs tagged with SmBiTs will bind to specific TFBSs and when the subunits get in proximity, and a substrate is provided, a bright luminescent signal will be produced. If a compound is added to the TF, changes in the TF-DNA binding and the subunits proximity may occur, leading to a detectable change in signal.

Luciferase enzyme split reporter systems offer advantages such as reversible reporter formation, low background luminescence due to inactive subunits, and weak affinity between split luciferase subunits. However, certain fusion partner orientations may be limited due to steric effects [50]. Therefore, testing multiple expression construct combinations is recommended to achieve optimal fusion orientations. Spontaneous association between split luciferase subunits can reduce the assay window for compound treatment or lead to false positives. Without proper controls, signals resulting from spontaneous association of LgBiT-SmBiT could be misinterpreted as specific protein interaction. In this assay various negative controls will be implemented including parental untransfected cells, cells only expressing the dCas9-LgBiT, cells transfected with dCas9-LgBiT and TF-SmBiT without sgRNAs, and the use of scrambled sgRNA. Moreover, tool compounds, which modulate PPIs in a dose-response manner, can be used during assay development. These compounds may act as direct inhibitors of PPIs or alter PPI status directly or indirectly. Their use as controls can validate assay functionality and enhance its performance. Specific interaction can be confirmed if the expected change in luminescence occurs upon treatment with a tool compound, whether it leads to an increase or decrease in signal. [57]

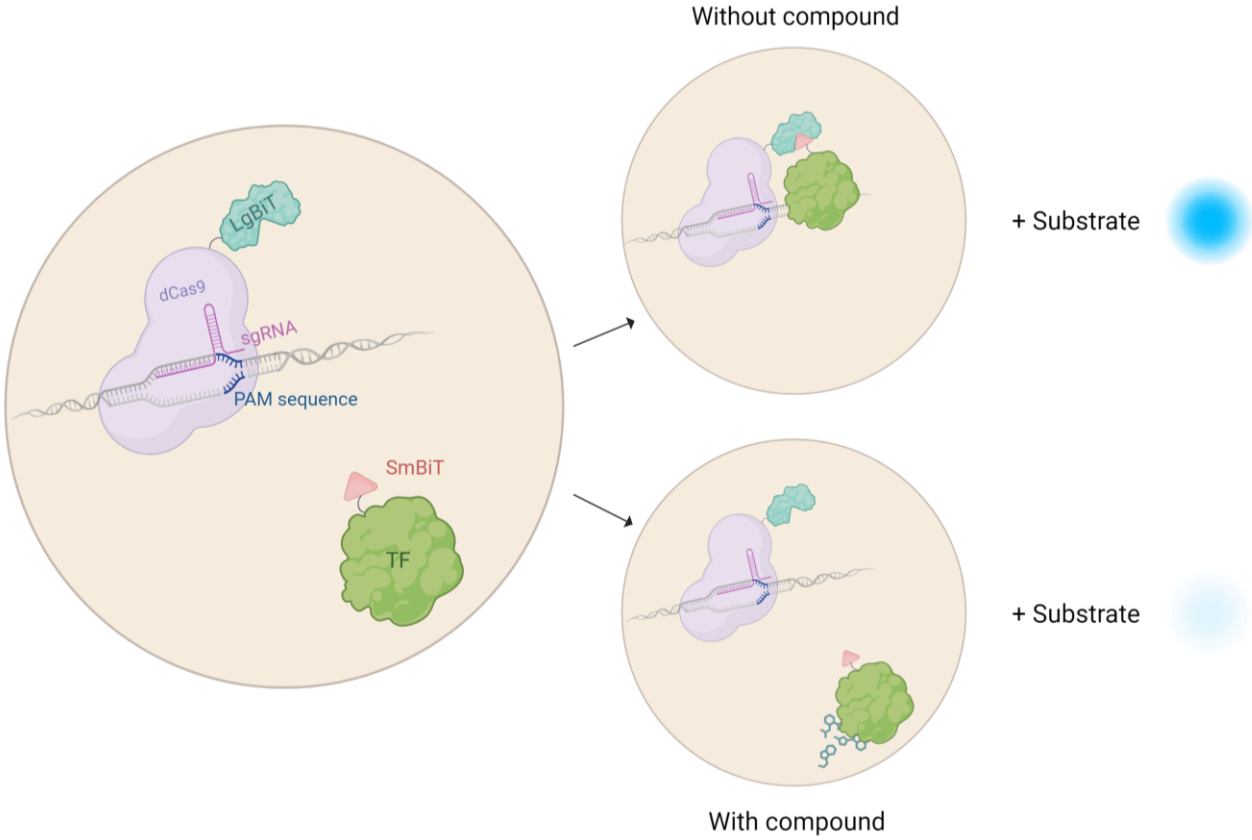


Figure 6: An overview of how the split-reporter system, NanoBiT®, employed in this project functions. The figure manifests how the system will allow for detection of TF-DNA interaction both with and without compound treatment. The blue signal display the luminescence. Here, compound treatment is shown to have an inhibitory effect on TF-DNA binding. A high detected signal upon substrate addition will indicate strong interaction between the TF and DNA. Conversely, a low signal will indicate weak TF-DNA interaction. Created with BioRender.co

3

Methodology

In the following section, the methodology of the thesis project is summarized. The chapter is divided into two parts: reagent generation, and assay development and optimization. Supplementary methodology and more detailed standard protocols are found in Appendix B. Moreover, the safety-, ethical- and environmental disclosures of this study are found in Appendix C.

3.1 Reagent Generation

3.1.1 Design, Generation and Amplification of Plasmids

The dCas9-LgBiT-N-terminus and C-terminus plasmids for this project were designed by Anna-Lina Cavallo. The plasmids were designed containing coding sequences of the CVM promoter and dCas9, in connection to either N- or C-terminus of the LgBiT. These constructs also included a marker encoding neomycin resistance. Furthermore, different plasmids expressing the EF-1α promoter, the TF of interest, connected to either N- or C-terminus of the SmBiT were designed by Anna-Lina Cavallo. These plasmids included a puromycin resistance cassette. The generation of these TF-SmBiT plasmids were outsourced to the company VectorBuilder. In total, 4 different vectors were created for this project: dCas9-LgBiT-N-terminus, dCas9-LgBiT-C-terminus, STAT3-SmBiT-C-terminus, and AR-SmBiT-N-terminus. Vector maps displaying positions of major elements of the sequence including promoters, POI, terminators, selectable markers, and origins of replication, together with the plasmid lengths can be found in Appendix A.

The dCas9-LgBiT-N-terminus and C-terminus plasmids were generated and provided in-house by the Protein Expression and Molecular Biology Team. In brief, the amino acid sequence of the constructs was backtranslated to DNA sequence and codon optimized to *Homo Sapiens* using GeneArt's algorithm. The outcomes were then fragmented into shorter fragments, ranging between 0.5-2 kb. Golden gate cloning was conducted in a thermal cycler by mixing the fragments, a pIRESneo backbone, the restriction enzyme BsaI, and T4 ligase. The assembled plasmids were then transformed into competent bacterial cells. Minipreps of the plasmids were conducted and after that correct clones could be confirmed by Sanger sequencing.

To amplify the plasmids, they were first transformed into One Shot™ TOP10 Chemically Competent *E. coli* (Invitrogen™, cat#C404003), according to the protocol in Appendix B section B.2.1. Next, the transformed bacteria were grown on LB-agar plates containing 1000X carbenicillin (stock 50 mg/mL). The following day, a single colony of bacteria was picked and inoculated overnight in Lysogeny broth (LB) media containing 1,000X carbenicillin (stock 50 mg/mL) to expand the plasmids. After expansion the plasmids could be purified using the GenScript AmMag™ Quatro system. This is an automated system for plasmid purification. Alternatively, the QIAGEN® Plasmid *Plus* Maxi Kit (QIAGEN, cat#12963) was used. To quantify the yield of the purified plasmid DNA, the Implen NanoPhotometer® instrument was used. Appendix B.2 provide a more detailed protocol of the bacterial transformation and plasmid purification. Furthermore, the plasmid concentrations, together with the optical density (OD) ratios, $OD_{260/280}$ and $OD_{260/230}$, can be found in Appendix D.1.

3.1.2 Culturing and Maintenance of Cell Lines

For each TF that was studied, an appropriate cell line was selected with consideration to disease to fit the biological model. The HEK-293 cell line was chosen to study STAT3 and LNCaP was chosen to study AR. The two cell lines HEK-293 and LNCaP utilized in this project were provided internally by the Cell Banking Teams at AstraZeneca. Prior to usage, the cell banks had been quality controlled, including confirmed to be mycoplasma free and had passed STR (short tandem repeat) fingerprint analysis. The HEK-293 cell line was cultured and maintained in DMEM/F-12, GlutaMAX™ (Gibco™, cat#10565018) supplemented with 10% fetal bovine serum (FBS). Furthermore, the LNCaP cell line was cultured and maintained in RPMI-1640 medium (Gibco™, cat#21875034) supplemented with 10% FBS. A detailed protocol of how the mammalian cell lines were handled *in vitro* can be found in Appendix B.1. This includes protocols for general maintenance, thawing, passaging, and cryopreservation of cells. The cells were routinely cultured in a humidified incubator at 37°C, 5% CO_2, and 95% relative humidity (RH).

3.1.3 Transfection of Cell Lines with dCas9-LgBiT-N/C-Terminus Constructs Using MaxCyte®

To genetically engineer the cell lines and incorporate the plasmids into the different cell models, electroporations were conducted. One day prior to the transfection, cells were plated in an appropriate flask to reach ~70% confluency on the day of the transfection. First, the media was aspirated from the flask and the cells were washed with phosphate buffered saline (PBS) pH 7.4 (Gibco™, cat#10010015). Next, the PBS was aspirated and StemPro™ Accutase™ Cell Dissociation Reagent (Gibco™, cat#10565018) was added, and the flask was placed in an incubator at 37°C for 5-10 min to detach the cells from the wall of the flask. Media, specific for the cell type, was added to the cells and the cell suspension was transferred to a conical tube. An aliquot was taken to measure the cell density using the ChemoMetec NucleoCounter® NC-3000™. The required volume of cell suspension to get 1,000,000 cells for each transfection was

calculated and transferred to conical tubes for centrifugation at 250 × g for 5 min, to harvest the cells. The supernatants were discarded, and the cells were washed with 5 mL of PBS by centrifugation at 250 × g for 5 min. Next, the cell pellets were resuspended in 200 µL of Electroporation Buffer (MaxCyte®, #cat EPB-1) for each transfection. Thereafter, 2.5 µg of plasmid DNA was prepared in 1.5 mL microcentrifuge tubes. The cell suspensions were added to the tubes containing plasmid DNA, and the mixtures were transferred to OC-400 RUO™ processing assembly (MaxCyte®, cat#SOC-4) cuvettes. The electroporations were conducted using the MaxCyte® ExPERT STx™ and selecting a program specific for the cell type. Next, the suspensions of transfected cells were transferred to small plates and incubated for 15 min at 37°C. After incubation, an appropriate volume of media was added to the cells and the cells were transferred to a suitable flask format. The flasks were then placed in the incubator at 37°C, 5% CO_2, and 95% RH.

3.1.4 Generation of Stable Cell Lines

After the cells were transfected and had been growing for a couple of days to recover from the electroporations, the work of generating stable cell lines was initiated. The plasmids contain a sequence encoding antibiotic resistance of neomycin, so a selection pressure in the form of Geneticin™ Selective Antibiotic (G418 Sulfate) (50 mg/mL) (Gibco™, cat#10131027), was applied to the cells to inhibit the growth of untransfected cells. The selection pressure was kept by continuously adding 700 µg/mL of G418, when maintaining the cells in culture. This enabled selection of the cells, that stably expressed the plasmid. Untransfected cells died off and were removed when changing the media. After a couple of weeks, when visible colonies of transfected cells had start to form in the flask, the cells were dissociated and cultured together in a flask to generate a stable pool of transfected cells. For the LNCaP cell line, a dose response experiment was conducted prior to treating the cells with antibiotics to determine the optimal concentration for selection. This experiment is described in Appendix B.3.

3.1.4.1 Expansion, Banking and Quality Control of Cell Lines

A frozen cell stock is important and gives an opportunity to go back to early passaged cells and thaw more if necessary. Therefore, the stable cell lines generated in this project were first expanded prior to cryopreservation. Expansion of the cells involved serial passages of the cell lines from smaller volumes and flasks to larger ones to ultimately reach a high number of cells. Then the cells were cryopreserved according to the protocol in section B.1.3 in Appendix B. Three different types of banks were established for the HEK-293-dCas9-LgBiT-N- and C-terminus cell lines: archival-, master-, and working banks. After banking, frozen cell pellets of the genetically engineered cell lines were sent to IDEXX for quality control. The quality control involved mycoplasma testing and STR profiling for cell line authentication.

3.1.5 Troubleshooting of Transfecting LNCaP Cells

LNCaP cells were first transfected using the same protocol as described in 3.1.3. However, no cells survived the later antibiotic selection. Once repeated, using 1 µg of plasmid DNA, selection failed again. Troubleshooting of transfecting the LNCaP cells was done, and the transfection efficiency was evaluated by using a plasmid encoding green fluorescent protein (GFP), serving as a control. After determining the transfection efficiency, some adjustments were made in the protocol. Firstly, the number of cells to transfect were increased. Secondly, the amount of DNA was decreased to lower DNA toxicity. The final adjustment was that the culturing flasks were exchanged to culturing flasks with Poly-D-Lysine coating, to enhance cell attachment. The MaxCyte® electroporations were then repeated with these adjustments for the LNCaP cells as described in Appendix B section B.4.1. In parallel to this, another electroporation method, the Neon™, was also tested for the LNCaP cells. The method of this electroporation method is described in the next paragraph and a more detailed protocol of the experiment can be found in Appendix B section B.4.2.

3.1.5.1 Neon™ Electroporation

LNCaP cells were plated the day before electroporation to reach a confluency of approximately 70% on the day of transfection. The cells were detached according to the dissociation protocol in Appendix B.1.2, diluted in RPMI 1640 Medium + 10% FBS (Gibco™, cat#21875034), and the cell density was measured. The required volume of cell suspension was calculated and transferred to a conical tube for centrifugation at 250 × g for 5 min, to harvest the cells. The supernatant was discarded, and the cells were washed with PBS (cat#10010015, Gibco™) by centrifugation at 250 × g for 5 min. The supernatant was discarded. Next, a Nunc™ Poly-D-Lysine Coated 24-well plate (Thermo Scientific™, cat#152025) was prepared by adding 490 µL of media to each well, and it was placed in the incubator. The Neon™ Transfection System 10 µL Kit (Invitrogen™, cat#MPK1025) was used to conduct the electroporations. Tubes with 3 mL of electrolytic buffer E was prepared and one was placed in the electroporator (use for a maximum of five reactions). The pellets were resuspended in 80 µL of buffer R. Calculated amounts of DNA was prepared in Eppendorf tubes. Then, 20 µL of cell suspension and varying volumes of buffer R was added to the tubes with DNA. The 10 µL pipette tips were used to electroporate the samples using the following protocol: Pulse voltage: 950V, pulse width: 30 ms and pulse number: 2 pulses. After successful electroporation the cells were immediately seeded into the corresponding wells of the Poly-D-Lysine plate.

3.1.5.2 Stop Decision of Generating Stable LNCaP Cell Lines Expressing dCas9-LgBiT

The transfected LNCaP cells were not successfully selected after implementing these adjustments, or when testing a different electroporation method. Therefore, it was decided to only continue with the HEK-293 genetically engineered cell lines. Further optimization is needed in the electroporation protocol of LNCaP cells in order to select and generate stable cell lines. With consideration of the time frame of this project and the aim being to establish PoC of a system, the project was proceeded without LNCaP cell lines stably expressing dCas9-LgBiT.

3.1.6 Western Blotting of HEK-293-dCas9-LgBiT-N/C-Terminus

A western blot was conducted in this project to confirm that the dCas9-LgBiT-N and C-terminus constructs had been correctly integrated into the transfected cells, and that the fused dCas9-LgBiT was being expressed. In this western blot the primary antibody, Anti-LgBiT Monoclonal Antibody (Promega, cat#N710A), was used to detect the LgBiT subunit. Furthermore, an antibody of the housekeeping protein, tubulin, was used as an additional primary antibody, as an internal positive control. Parental HEK-293 cells were used as a negative control. The secondary antibody used in the western blot was IRDye® 800CW Goat anti-Mouse IgG Secondary Antibody (LI-COR, cat#926-32210). Furthermore, a protein standard was included in the analysis as a reference.

To prepare the samples for western blot analysis, cells in culture were first detached according to the dissociation protocol in Appendix B.1.2. Next, the cells were diluted in media and the cell density was measured. The amount of cell suspension to get 5M cells was transferred to a conical tube. The tube was the centrifuged at 250 × g for 5 min. The supernatant was discarded, and the pellet was resuspended in 1 mL of PBS. The tube was centrifuged at 250 × g for 5 min, and once again the pellet was resuspended in 1 mL of PBS. The suspension was transferred to a 1.5 mL Eppendorf tube and centrifuged at 250 × g for 5 min. The supernatant was removed, and the pellet was stored in -20°C until further usage.

Day one

The pellets were kept on ice and were suspended in 80 µL of Laemmli lysis buffer. The lysates were then incubated on ice for 15 min and was vortexed from time to time. Next, the tubes were centrifuged in a tabletop microcentrifuge at 4°C and 13.3 rpm (maximum speed) for 20 min. The supernatants were transferred to new 1.5 mL Eppendorf tubes. 12 µL of NuPAGE™ Sample Reducing Agent (10X) (Invitrogen™, cat#NP0009) and 30 µL of NuPAGE™ LDS Sample Buffer (4X) (Invitrogen™, cat#NP0007) was mixed in a tube. 8 µL of the mixture was added to 30 µL of each lysate, in new tubes. The samples were incubated at 95°C for 8 min on a termoblock. Next, the electrophoresis system was assembled, the NuPAGE™ 4 to 12%, Bis-Tris, 1.0–1.5 mm, Mini Protein Gel (Invitrogen™, cat#NP0322BOX) was inserted, and MOPS SDS running buffer was

added. 10 µL of the samples, together with 3 µL of the Novex™ Sharp Pre-stained Protein Standard (Invitrogen™, cat#LC5800) were loaded. The gel was run at 200V for 40 min. After gel electrophoresis, the gel was gently removed, and the Nitrocellulose/Filter Paper Sandwich, 0.45 µm, 8.3 x 7.3 cm (Invitrogen™, cat#LC2001) was assembled. Next, the membrane transfer was run in 1L of transfer buffer overnight at 30V and at 4°C.

Day two

After the protein transfer, the membrane was removed from the transfer apparatus. To block nonspecific sites, Intercept® (PBS) Blocking Buffer (LI-COR, cat#927-70001) was added and the membranes were incubated with gentle rocking for 1 hour at room temperature. Next, the primary antibodies, anti-LgBiT monoclonal antibody and anti- tubulin monoclonal antibody, were diluted to 1 µg/mL in blocking buffer. The buffer was removed and replaced with the solution containing blocking buffer and the primary antibodies. Incubation occurred overnight at 4°C with gentle rocking.

Day three

The membrane was washed three times, 10 min per wash, in PBS containing 0.1% tween. The secondary antibody, goat anti-mouse IgG, was diluted in blocking buffer and was incubated for 1 hour at room temperature with gentle rocking. The membrane was washed once again, three times, 10 min per wash, in PBS with 0.1% tween. Lastly, the membrane was imaged using the LI-COR Odyssey® CLx Imaging System.

3.1.7 Design and Synthesis of sgRNAs

The sgRNAs synthesized for this project were designed with support from the Target Identification and Validation Informatics Team at AstraZeneca. For STAT3, two downstream target genes were identified: IL-23A and NFKB1. For these downstream target genes, ranges of sgRNAs were designed with respect of targeting the promoter sequence and dCas9 recognizing the PAM sequence 5'-NGG-3'. Moreover, the design of these sgRNAs were in the format $(N)_{20}$. The purpose of the analysis was to provide guides matching the PAM sequence within the target region window of 2 kb upstream from a TSS. In addition, the sgRNAs were designed targeting downstream genes of a TF while preserving the TFs binding activity.

The sgRNAs were designed using bioinformatic tools, in this case by using R. First, a region of 2 kb upstream of the TSS of the target gene was selected. Then the JASPAR2024 R package was used to select the binding profile of the TF in human. TFBSTools R package was used to search and identify TFBSs within the 2 kb region. Next, all TFBSs that were <50 nucleotides away from the guides were filtered away. This was done to select sgRNAs targeting sites that were non-overlapping with the identified TFBSs. Hence, not affecting the TFs ability to bind to its regulatory elements. The p-value and relative score for each TFBS was calculated using

TFBSTools. The p-values were then adjusted for multiple testing using the Benjamini-Hochberg method. The set of obtained P_{adjust}-values were then ranked from lowest to highest P_{adjust}-value, and the relative scores from highest to lowest. Then, ~5 sgRNAs were selected for each downstream gene associated with the TFs, with consideration of the P_{adjust}-value and the relative score. Furthermore, a randomized sgRNA sequence, scrambled sgRNA, was also designed, to serve as a control in the assay. The synthesis of the sgRNAs were outsourced to the company Sigma-Aldrich. The sequences of the designed sgRNA are provided in Appendix D.2.

To prepare the sgRNAs upon arrival, RNaseZap™ RNase Decontamination Solution (Invitrogen™, cat#AM9780) was used to wipe the pipette and other surfaces to destroy any RNases. The sgRNAs were then suspended in 30 µL Tris-EDTA Buffer 10 mM Tris, 1 mM EDTA, pH 8.0 (Synthego) to a concentration of 100 µM and stored in -20°C until further usage.

3.1.8 Transient Co-Transfection with TF-SmBiT Constructs and sgRNAs

Transient co-transfections of the dCas9-LgBiT cell lines with TF-SmBiT plasmid and various sgRNAs was done to enable measurement of the luminescence. Cells were seeded one day before transfection to reach ~70% confluency on the day of transfection. The cells were dissociated according to the protocol in Appendix B.1.2. Next, the cells were diluted to a volume of 10 mL in media. The cell suspension was transferred to a conical tube and an aliquot was taken to measure the cell density. The volume to get 500,000 cells for each transfection was calculated. The volumes of cells were transferred to 15 mL conical tubes. The tubes were centrifuged for 5 min at 250 × g to harvest cells. The supernatant was discarded and washed with ~5 mL of PBS by centrifugation for 5 min at 250 × g. The supernatants were discarded. Subsequent, the cell pellets were resuspended in 20 µL MaxCyte® reagent. The calculated volumes of DNA and sgRNA were prepared in Eppendorf tubes. When working with the sgRNAs, they were kept on ice to prevent degradation. RNaseZap™ RNase Decontamination Solution was also used to wipe surfaces and eliminate RNAses. For the STAT3-SmBiT plasmid with a concentration of 1193 ng/µL, amounts of 500 ng and 1 µg was prepared. Furthermore, 2 µL of the various sgRNAs (100 mM) was used. The resuspended cells were then added to the tubes with the reagents. A control transfected with the STAT3-SmBiT plasmid but without sgRNA was also included. The cell mixtures were then transferred to cuvettes (OC-25 x 3), and the transfections were conducted using the MaxCyte® electroporator. After electroporation, the transfected cells were added to a 24-well plate and were incubated for 15 min at 37 °C. Approximately 1 mL of media was added to the transfected cells. An aliquot was taken, and the cell concentration was determined. Each of the transfected cell suspensions was diluted to a concentration of 100,000 cells/mL, to get 4,000 cells per well when plating 40 µL cell suspension/well. The cells were then plated in CELL CULTURE F-BOTTOM, WHITE, TC, STERILE (Greiner, cat#781073) plates. Negative controls, including parental HEK-293 cells and HEK-293 cells stably expressing dCas9-LgBiT were also included. The plates were centrifuged for 30 sec at 100 × g, to spin down all reagents in the wells. The plates were then incubated at 37°C, 95% RH for 24-48 h.

3.2 Assay Development and Optimization

3.2.1 NanoBiT® Assay with dCas9-LgBiT, sgRNA and TF-SmBiT Reagents

The assay was performed within 24-48 h after the transient co-transfection. The Nano-Glo® Live Cell Assay System (Promega, cat#N2011) was conducted following the vendor's instructions [57]. Nano-Glo® LCS Dilution Buffer was equilibrated to ambient temperature if using it for the first time. The Nano-Glo Live Cell Substrate was removed from the freezer and was mixed. The assay reagent was prepared by diluting 1 volume of Nano-Glo® Live Cell Substrate with 19 volumes of the Nano-Glo® LCS Dilution Buffer to make the Nano-Glo® Live Cell Reagent. The Nano-Glo® Live Cell Reagent was made fresh for each experiment with accordance to the manufacturer's instructions. 10 µL of Nano-Glo Live Cell Reagent was added to each well of a 384-well plate. The plate was spun down in a centrifuge at 100 × g for 30 sec. Immediately after, the luminescence was measured using the PHERAstar® *FSX* HTS microplate reader (BMG Labtech).

3.2.2 NanoBiT® Assay of a Previously Developed System

A NanoBiT® assay with HEK-293 cells co-transfected with a LgBiT and SmBiT 64/66 plasmids, previously tested by Anna-Lina Cavallo, was conducted. One of these two SmBiT plasmids served as a negative control, having a point mutation in the gene expressing the SmBiT, hindering complementation of the LgBiT-SmBiT subunits. Furthermore, untransfected parental HEK-293 cells and HEK-293 cells only transfected with LgBiT also served as negative controls in the assay.

3.2.3 NanoBiT® Assay with IL-6 Stimulation

Cell lines expressing dCas9 fused to either the N-terminus or C-terminus of LgBiT were transfected with 1 µg of STAT3-SmBiT plasmid. Additionally, different sgRNAs (g1-5) targeting the downstream target gene IL-23A were included in the transfections. To aid in interpreting specific interactions, HEK-293-dCas9-LgBiT-N/C terminus cells transfected with STAT3-SmBiT but without sgRNA were included. Negative controls, comprising LgBiT-N and C-terminus cells, as well as parental HEK-293 cells, were also included in the assay. The assay was conducted where cells were stimulated with 100 ng/mL interleukin 6 (IL-6) (stock 100 µg/mL). The IL-6 used in this experiment was kindly provided by Mei Ding. 30 min prior to running the assay the cells were treated with IL-6. The controls and transfected cells were plated in duplicate, with one well left unstimulated and one well stimulated with IL-6. Then the luminescence was measured at different timepoints: 0, 1.5, and 3 h.

4

Results

In this paragraph, the findings of the thesis project are presented. Supplementary results of this master's thesis project can be found in Appendix D. The supplementary results are not necessary for the conclusion of the paper but can provide a comprehensive understanding of the study.

4.1 G418 Titration of the LNCaP Cells

In Figure 7 below, the results obtained from the geneticin (G418) titration experiment of the LNCaP cells are presented, as described in Appendix B.3. The curves in Figure 7a were generated by plotting the confluence against the time after antibiotic treatment. The deviance in confluency that can be observed at the timepoint of three days, is due to a media replacement. It can be observed that untreated cells and cells treated with concentrations of antibiotic ranging from 50-300 µg/mL are showing similar profiles in confluency. Concentrations of G418 in the span 400-600 µg/mL respectively 700-900 µg/mL, also exhibit similar profiles. At the concentration of 1,000 µg/mL, the end point was similar to the profiles of concentrations ranging from 700-900 µg/mL G418, however, the confluency decreases more rapidly. In Figure 7b the confluency was plotted against the ranging concentrations of G418, at the timepoint 5 days and 18 hours after starting antibiotic treatment. It can be observed that almost all LNCaP cells are dead within 5 days and 18 hours of treating the cells with G418, with concentrations ranging from 700-1,000 µg/mL.

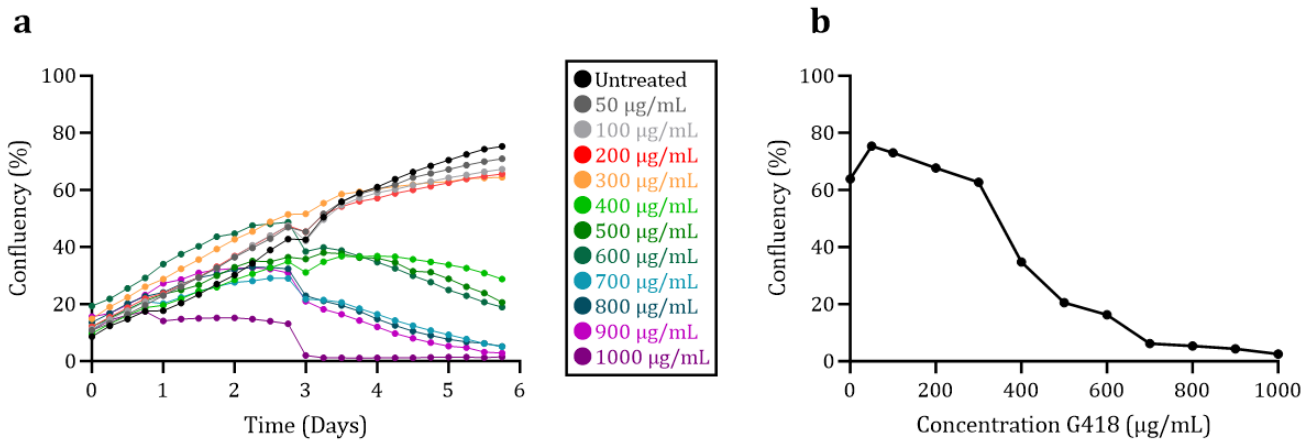


Figure 7: Titration of G418 for LNCaP cells. **a,** Time-kill curves of the LNCaP cells. The growth rate confluence was quantified every 6 hours for 5 days and 18 hours. Each curve represents different concentrations of G418, spanning from 50-1,000 µg/mL. After three days the media was replaced. **b,** Confluency versus concentration of G418, 5 days and 18 hours after antibiotic treatment. The graphs were created in GraphPad Prism version 10.01.2.

Next, imaging of the cells, 5 days, and 18 hours after treatment of G418, is presented in Figure 8. The figure display microscopy images a-f with increasing amounts of added G418 spanning from 0-1,000 µg/mL. It can be observed that between 400 and 800 µg/mL of G418, the confluency changes exceedingly. In the microscopy images of the cells, the profiles of confluency over time in Figure 7a, and the curve in Figure 7b, the antibiotic concentrations ranging from 700-1,000 µg/mL exhibit the highest decline in confluency after G418 treatment.

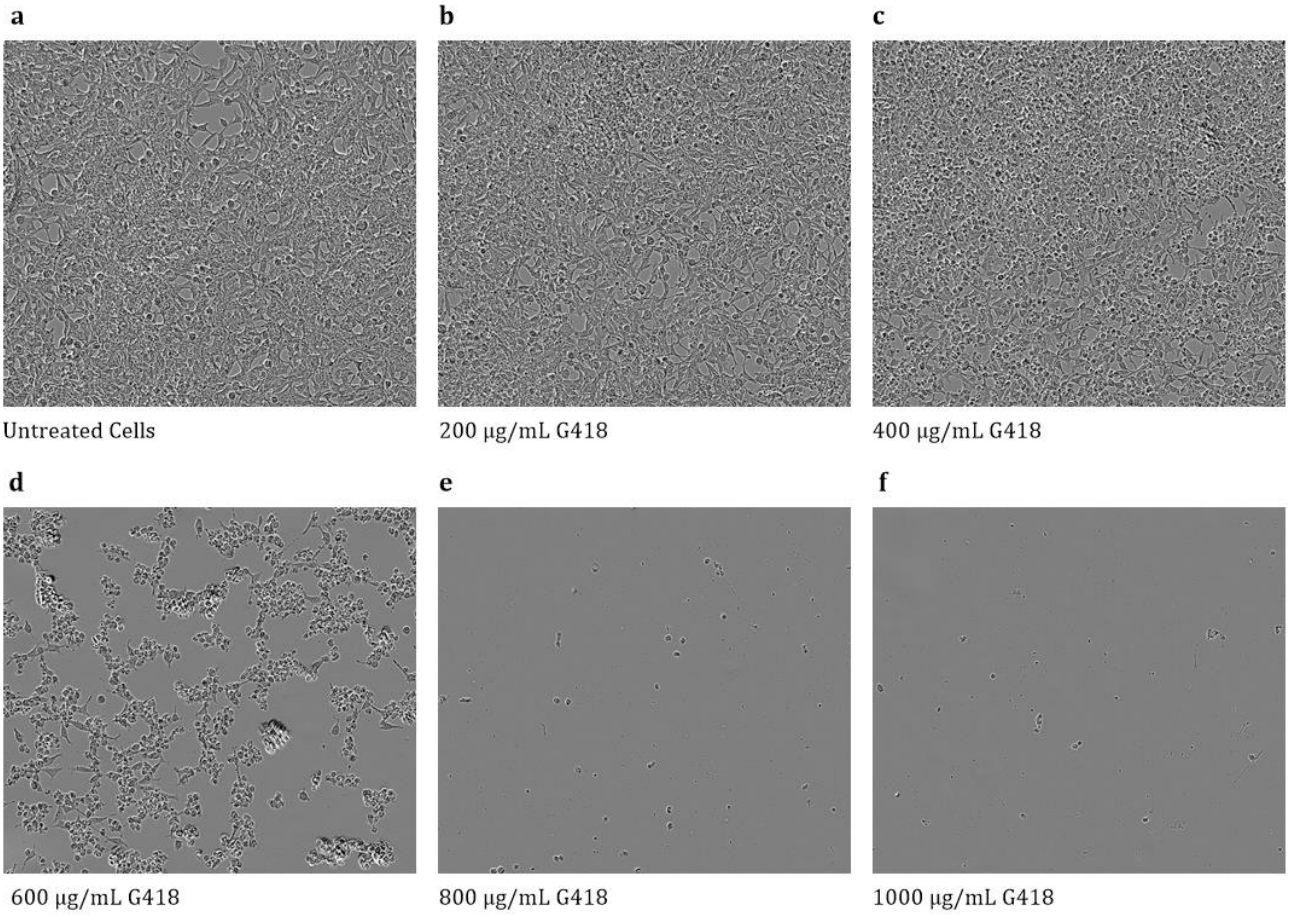


Figure 8: Monitoring of the LNCaP cells in the Incucyte® S3 instrument, 5 days, and 18 hours post antibiotic treatment. **a,** The untreated cells. **b-f,** Treatment of cells with increasing amounts of G418, ranging from 200-1,000 µg/mL. The microscopy images were obtained using the Incucyte® S3 Software (v2022B).

From these results presented of the G418 titration of the LNCaP cells, a concentration of 700 µg/mL G418 was selected to be an appropriate and effective concentration for later antibiotic selection of transfected LNCaP cells. The concentration was chosen to prevent any cell growth of untransfected cells after applying the selection pressure and enable generation of stable pools expressing the dCas9-LgBiT plasmids.

4.2 Evaluation of Transfection Efficiency of LNCaP Cells

4.2.1 MaxCyte®

Herein, the results from transfecting LNCaP cells with a GFP plasmid using the MaxCyte® transfection system is presented. The experiment was conducted to evaluate the transfection efficiency in LNCaP cells. The protocol of the experiment is found in Appendix B.4.1. Varying concentrations of plasmid DNA was tested to assess the transfection efficiency and potential DNA induced toxicity of the LNCaP cells. The presented results were obtained by monitoring the LNCaP cells in the IncuCyte instrument for 48h post transfection. In Figure 9 and 11, different graphs displaying transfection efficiency and confluency of the LNCaP cells and microscopy images of the cells are presented. After transfection the cells failed to be evenly distributed. Therefore, the number of cells in each well was varying, and generally exhibited a higher density at the centre of each well. For this reason, the mean confluency only reached about 30%. This was not ideal; however, this was similar for all the wells.

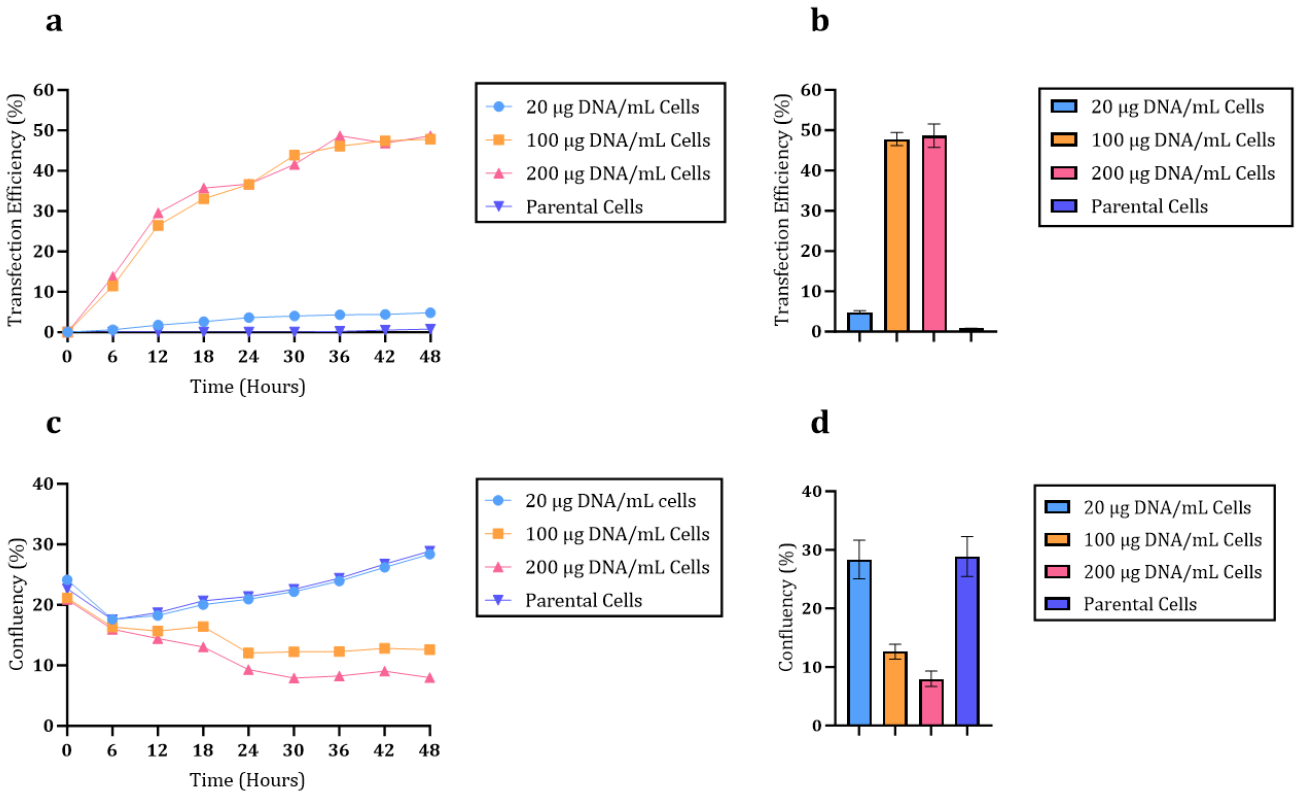


Figure 9: **a,** Graph of the GFP transfection efficiency at each concentration from 0-48 hours after transfection. **b,** The GFP transfection efficiency for each concentration at the timepoint 48 hours post transfection. **c,** Graph of the confluency at each concentration from 0-48 hours after transfection. **d,** The confluency at each concentration at the timepoint 48 hours after transfection. Graphs were created in GraphPad Prism version 10.01.2.

It could be observed that the highest concentration of plasmid DNA resulted in the highest transfection efficiency (Fig. 9). However, 48 hours post transfection, a concentration of 100 µg DNA/mL showed higher viability. At the lowest concentration of DNA, the efficiency was very low, although the viability was much higher 48h post transfection. At the 100 µg DNA/mL concentration the efficiency was high; however, the cell death was also high. These results suggest that a concentration between 20 µg DNA/mL cells and 100 µg DNA/mL would be the optimal concentration to get a high efficiency while still reducing the toxicity causing cell death.

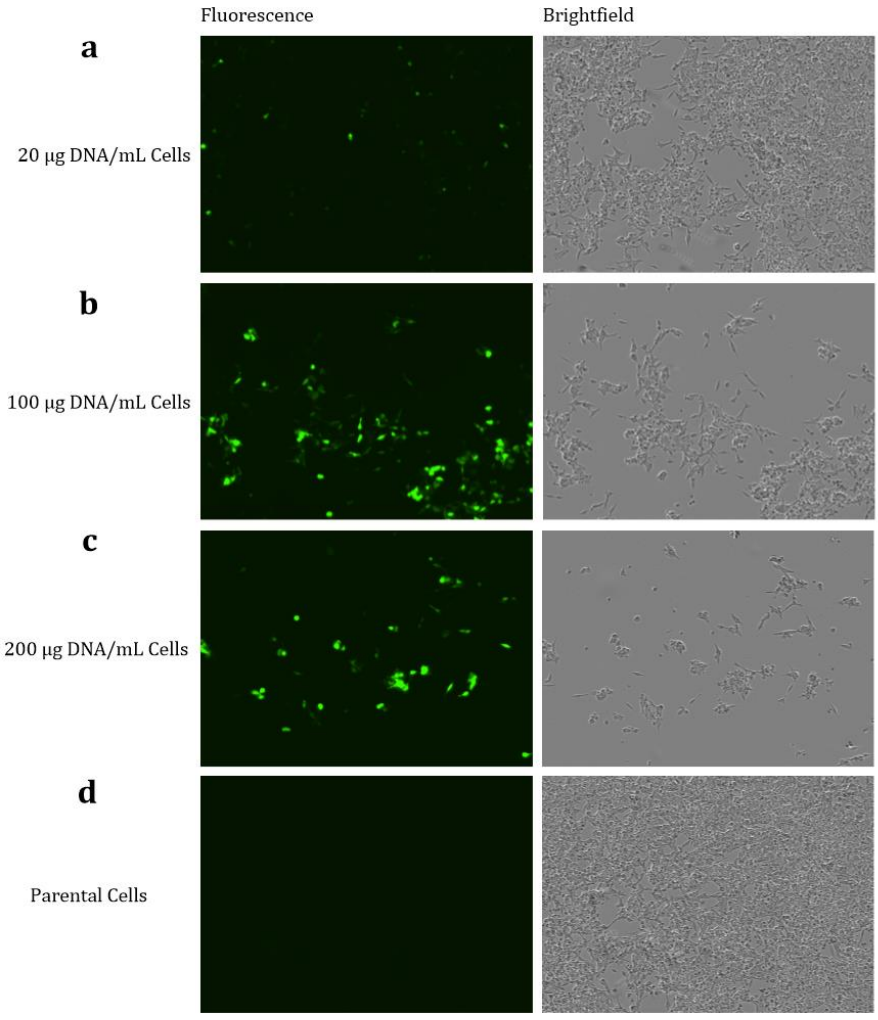


Figure 10: Transfection efficiency analysis of LNCaP cells 48 hours following transfection. **a-c,** Fluorescence and brightfield microscopy images of the cells at the different concentration of DNA. **d,** Untransfected control of parental cells. The microscopy images were obtained using the Incucyte® S3 Software (v2022B).

From the microscopy images in Figure 10 it can be observed that the 100 and 200 µg DNA/mL concentrations exhibit similar transfection efficieny of GFP plasmid, while the 20 µg DNA/mL concentration is exeedingly lower and that the untransfected negative control of parental cells do not express any GFP.

4.2.2 Neon™

Presented below are the results from the experiment described in Appendix B.4.2, where LNCaP cells were transfected using the Neon™ Transfection System (cat#MPK5000, Invitrogen™). The LNCaP cells were transfected with the dCas9-LgBiT plasmids. A GFP control and an untransfected control was also included in the experiment to evaluate the transfection efficiency.

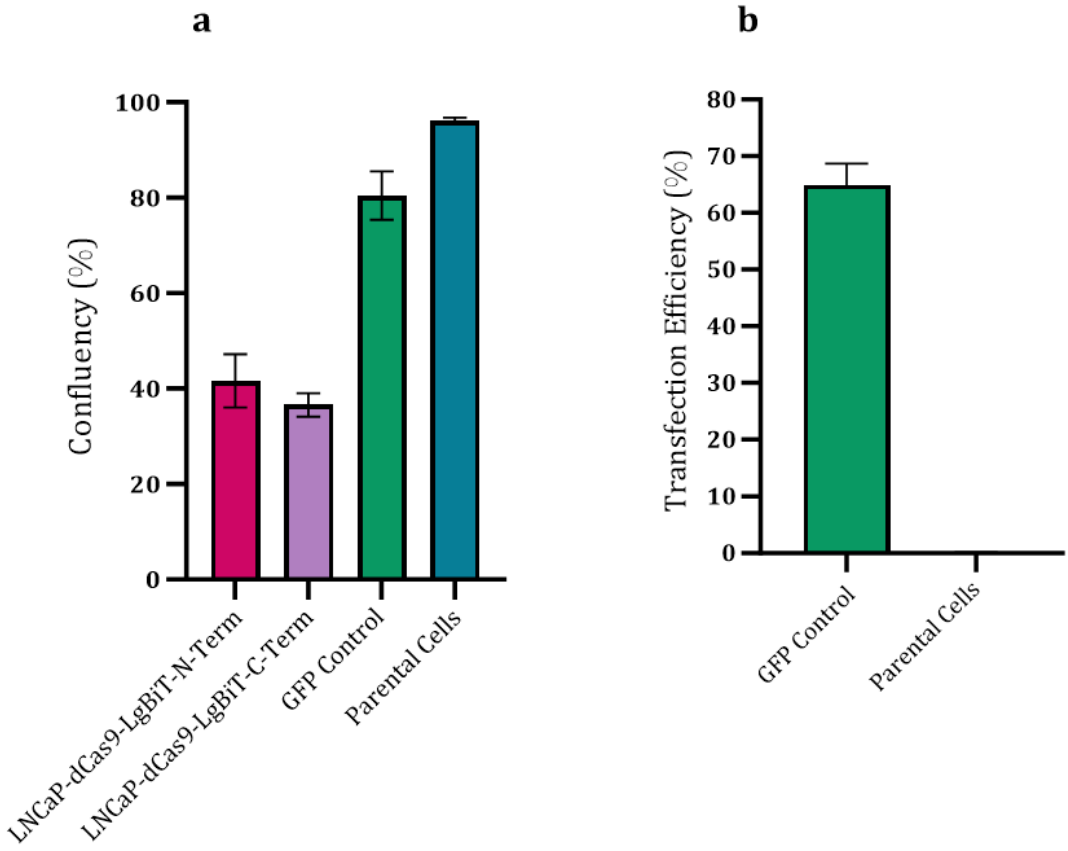


Figure 11: **a,** Charts of the confluency for the transfected LNCaP cells with the dCas9-LgBiT plasmids and the two controls. **b,** Graph of the transfection efficiency of the GFP control. Graphs created in GraphPad Prism version 10.01.2.

The results in Figures 11 and 12 indicate that the two plasmids are toxic to the LNCaP cells, which can be observed from the decreased confluency and amount of dead cells in the microscopy images. The dead cells had deatached from the surface of the flasks and were floating around in the media. Compared to the GFP control with a confluency of ~80% the dCas9-LgBiT-N-terminus and C-terminus exhibit a confluency of ~40% (Fig. 11). So, in comparison to the GFP control, transfections with the two dCas9-LgBiT plasmids lead to a high cytotoxicity on the LNCaP cells. Therefore, the evaluation of the transfection efficiency of the two dCas9-LgBiT plasmids is difficult to estimate. It is believed to be much lower considering the experienced difficulties of selecting cells stably expressing the constructs.

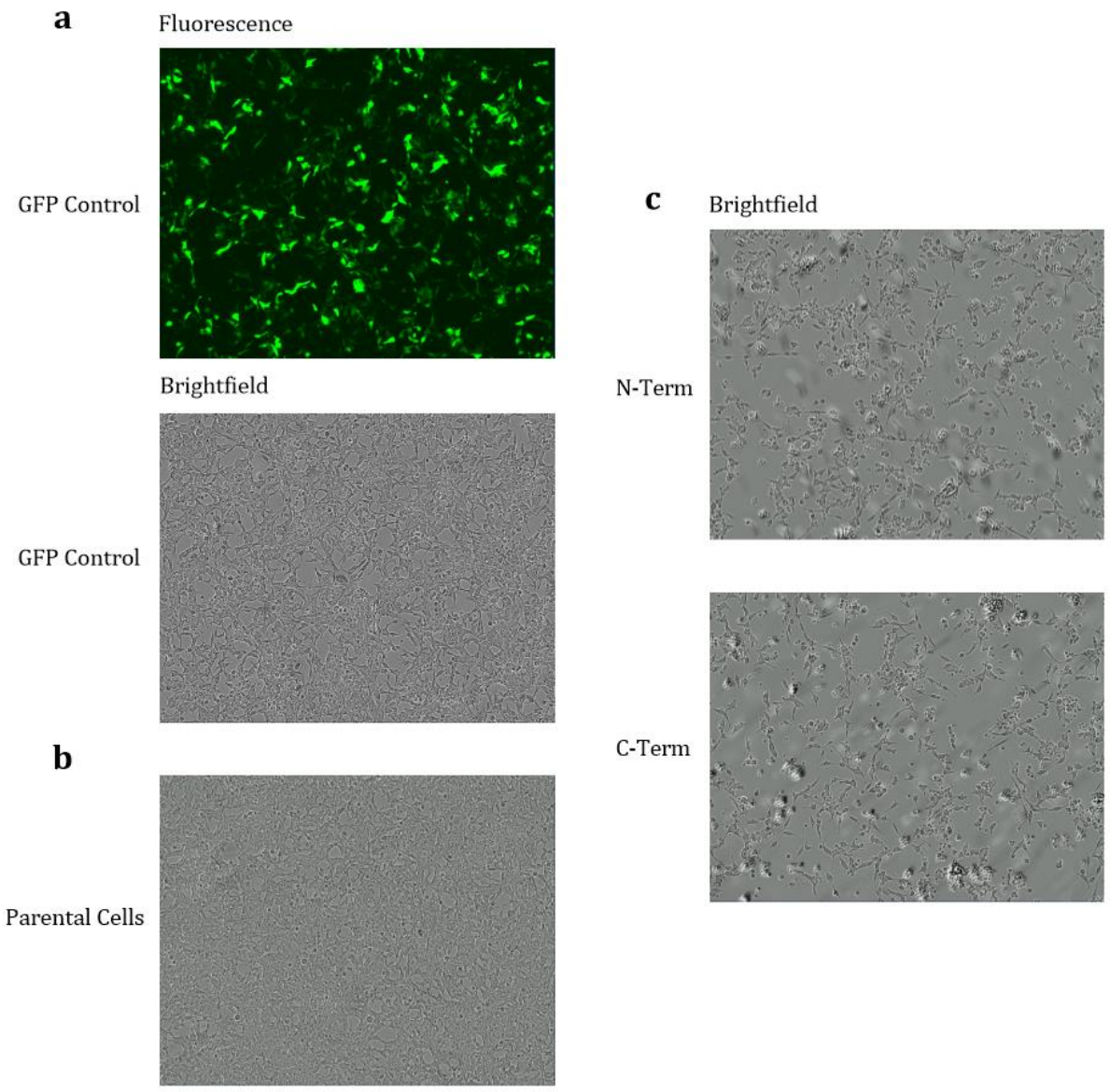


Figure 12: 72 hours post transfection analysis of LNCaP cells. **a-b,** Microscopy image of the GFP control with fluorescence and phase contrast and the untransfected control in brightfield. **c,** Brightfield microscopy images of the transfected LNCaP-LgBiT-dCas9-N-Terminus and LNCaP-LgBiT-dCas9-C-Terminus. The microscopy images were obtained using the Incucyte® S3 Software (v2022B).

From the microscopy images, it becomes evident that the confluency of the cells transfected with dCas9-LgBiT-N or C-Terminus have decreased excedingly 72 hours post transfection compared to the control of parental cells and the GFP control (Fig.12). With the transfection conditions described in Appendix B.4, the electroporations using the Neon™ transfection system showed higher efficiency in comparison to the MaxCyte® when comparing the GFP controls. For the Neon™ transfections, the efficiency was around 65% and for the MaxCyte® transfections, the efficiency was approximately 50%.

4.3 Generation of Cell Lines Stably Expressing dCas9-LgBiT

The generated HEK-293 cell lines stably expressing dCas9-LgBiT in connection to N- or C-terminus were negative for mycoplasma, and both showed >80% identity match for the STR profiling. An identity matching score above 80% indicate that the tested sample is consistent with the cell line of origin.

In Figure 13 below, the results from the western blot of the untransfected HEK-293 cells and the HEK-293 cell lines stably expressing dCas9 fused to either N- or C-terminus of the LgBiT is presented.

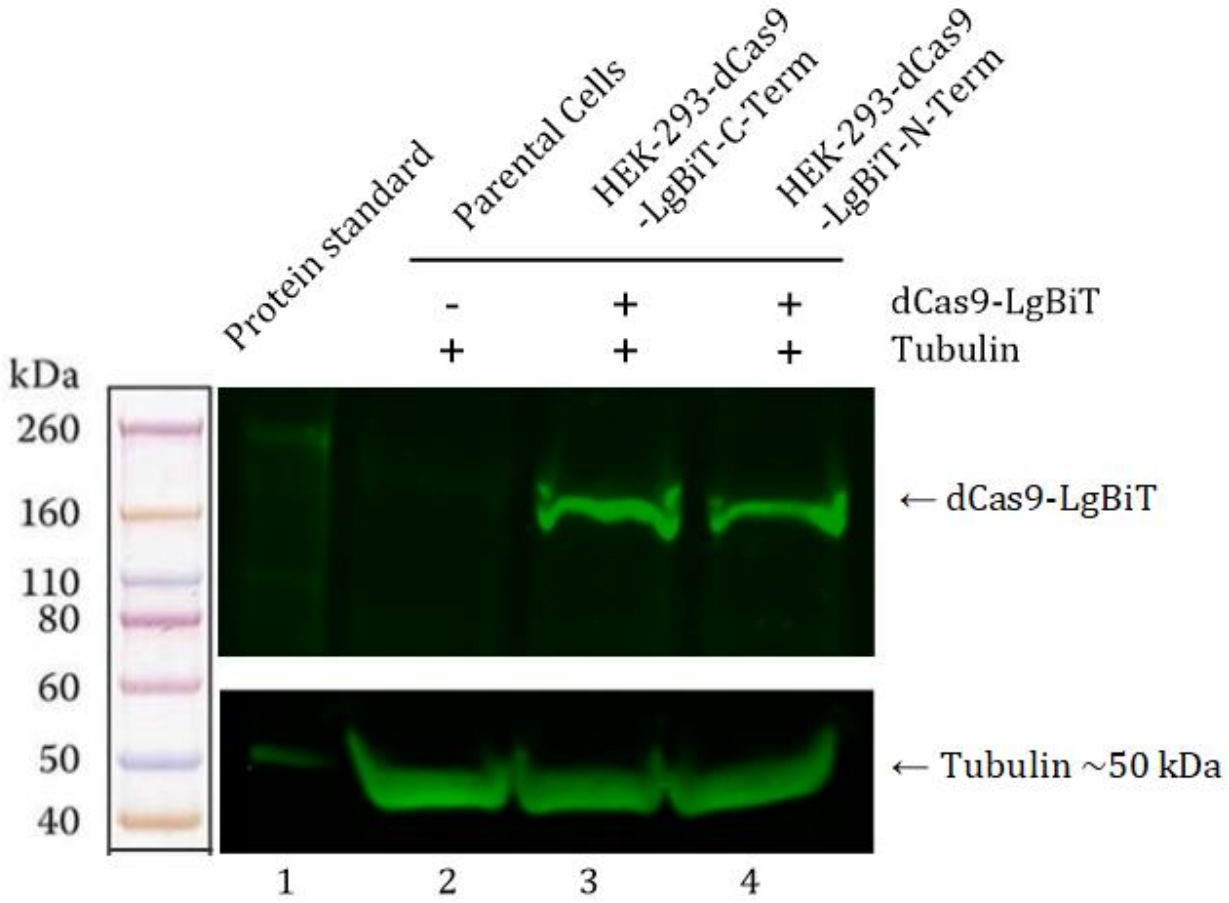
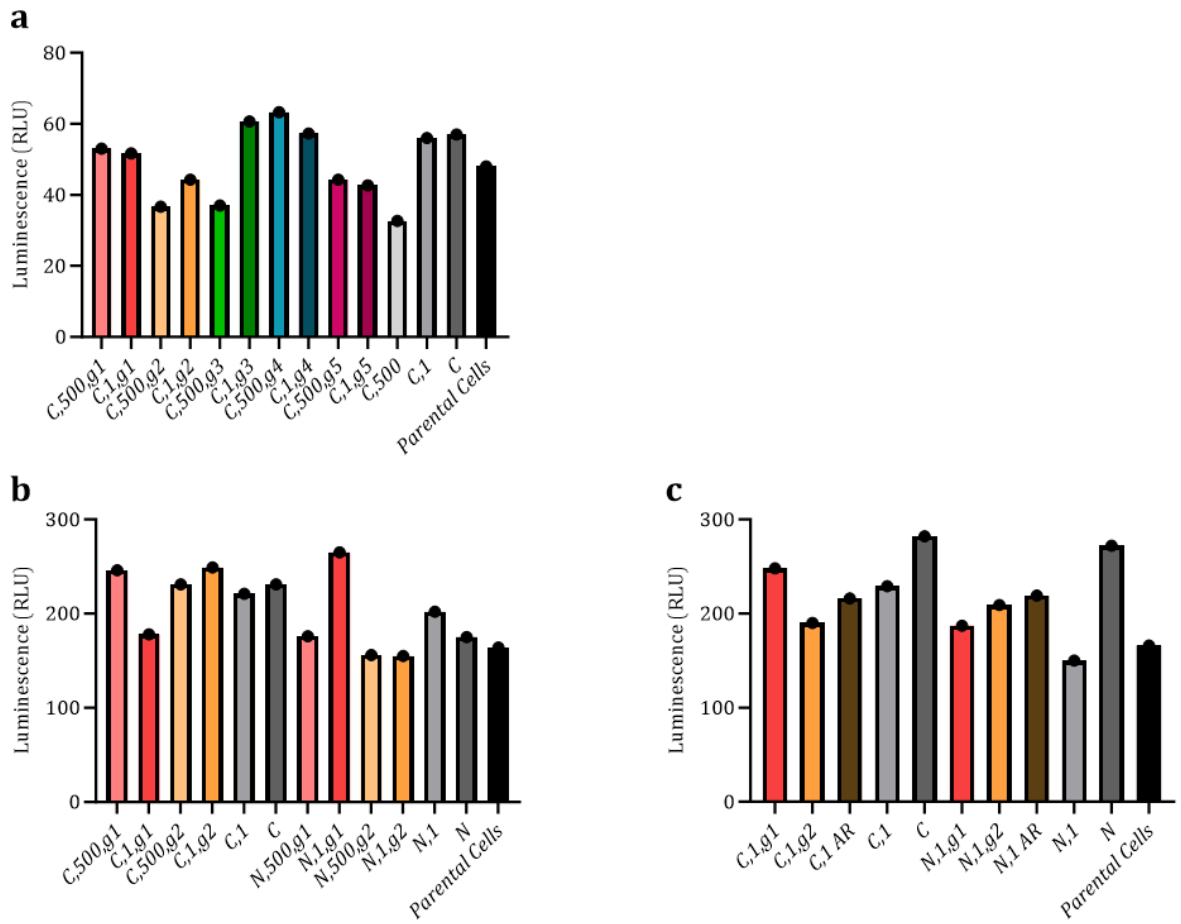


Figure 13: Western blot of the HEK-293, HEK-293-dCas9-LgBiT-N and C-terminus cell lines using anti-LgBiT and anti-Tubulin primary antibodies. The first lane contains the protein standard. The second lane contains parental HEK-293 cell lysate. The third and fourth lane contains cell lysates of HEK-293-LgBiT-dCas9 connected to C and N-terminus, respectively. The first range of bands around 50 kDa represents the tubulin protein and the second range of bands show the LgBiT fused dCas9. The figure was created in GraphPad Prism version 10.01.2.

The result from the western blot confirmed that the LgBiT fused dCas9 is being stably expressed in the genetically engineered HEK-293 cell lines (Fig. 13). Furthermore, the housekeeping protein tubulin, serving as a positive control, is being expressed in both the genetically engineered and the parental samples.

4.4 Initial Tests of the Developed NanoBiT® System

Figure 14 a-c displays the outcomes of assays conducted using various combinations of reagents for the transient transfections. Cell lines where dCas9 was fused to either N- or C-terminus of the LgBiT were transfected using different amounts of STAT3-SmBiT plasmid, either 500 ng or 1 µg. Furthermore, different sgRNAs g1-5 designed for the downstream target gene IL-23A was also included in the transfections. HEK-293-dCas9-LgBiT-N/C terminus cells transfected with STAT3-SmBiT without sgRNA were included to aid in the interpretation of specific interactions. Negative controls including LgBiT-N and C-terminus cells and parental HEK-293 cells were included in all assays. In assay c, all transfections were made in transient to assess if the cell lines stably expressing dCas9-LgBiT-N/C-teminus had any effect on the signal. This assay also included transfections with the AR-SmBiT plasmid to determine the level of nonspecific background noise.



C= HEK-293-dCas9-LgBiT-C-terminus cells, **N**= HEK-293-dCas9-LgBiT-N-terminus cells, **500**= 500 ng STAT3-SmBiT plasmid, **1**=1 µg STAT3-SmBiT plasmid, **1 AR**= 1 µg AR-SmBiT plasmid, **g1-5**= sgRNAs designed for the downstream target gene IL-23A, **Parental cells**= HEK-293 cells. **RLU**: Relative light units.

Figure 14: The measured luminescence of three NanoBiT® assays a-c. **a,** Measurements of the HEK-293-dCa9-LgBiT-C-terminus cell transfections with a measurement interval time of 0.1 s. **b,** Assay including the HEK-293-dCas9-LgBiT-N-terminus cells to see if any interaction could be detected using the N-terminus tagging strategy. A measurement interval time of 0.5 s. **c,** All transfections were made in transient. A measurement interval time of 0.5 s. Graphs created in GraphPad Prism version 10.01.2.

All measurements obtained during the initial system test are considered as background signal, indicating the absence of specific PPI interactions and LgBiT-SmBiT complementation. The lack of discernible signals suggests that TF-SmBiT did not bind in close proximity to or interact with dCas9-LgBiT. However, it remains uncertain whether a signal higher than background is anticipated when transfecting with both LgBiT and SmBiT subunits due to the fact that spontanous association of subunits can occur.

4.5 Transient Transfection Efficiency, 0-48h

For further exploration into the absence of detected interaction or complementation of LgBiT-SmBiT during testing of the developed NanoBiT® system, a GFP control was incorporated to assess transfection efficiency. The outcomes of these tests are presented in Figures 15 and 16.

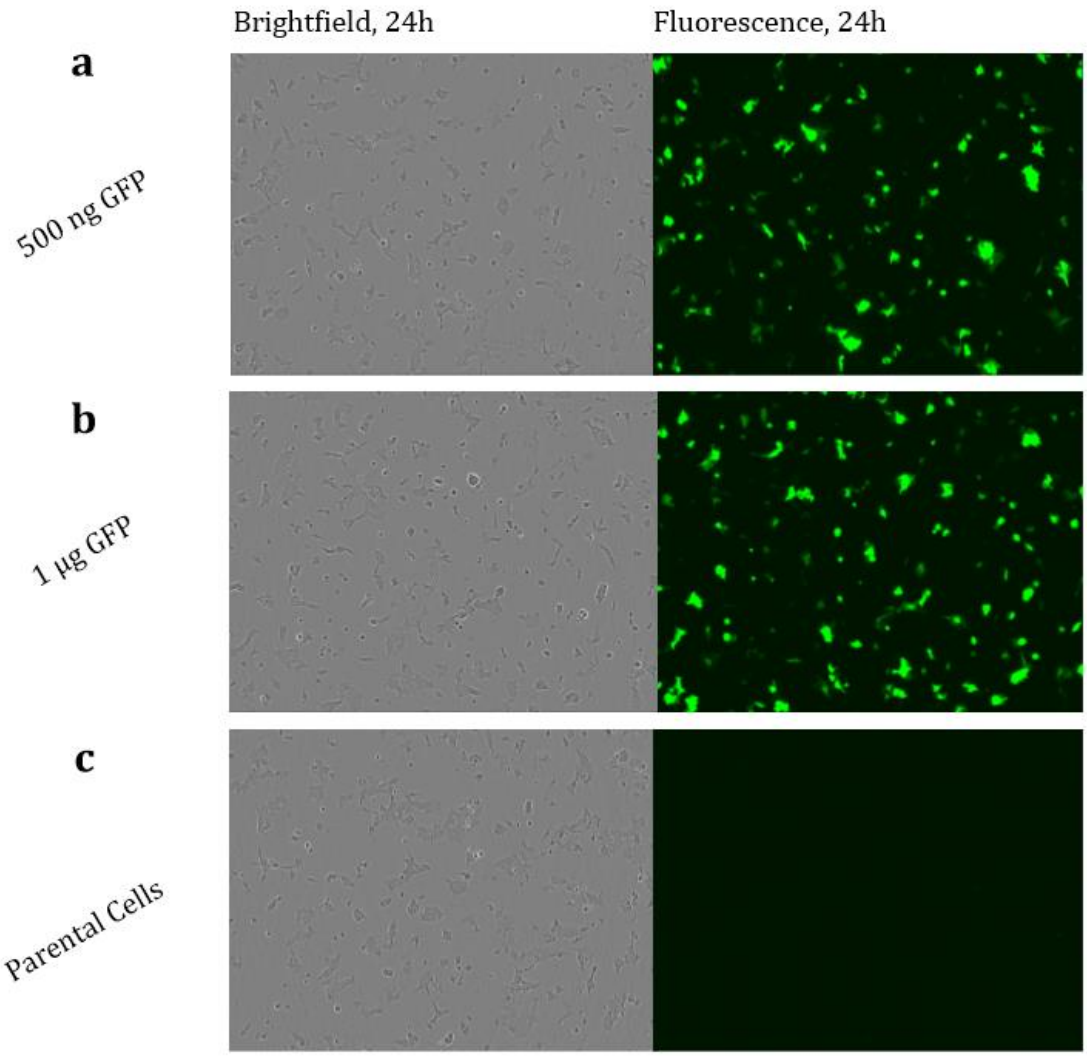


Figure 15: Microscopy images of HEK-293 cells transiently transfected with GFP plasmid. **a,** 500 ng GFP plasmid **b,** 1 µg GFP plasmid. **c,** Untransfected control of parental HEK-293 cells. The microscopy images were obtained using the Incucyte® S3 Software (v2022B).

From the microscopy images in Figure 15, the cells transfected with 500 ng respectively 1 µg of GFP plasmid show similar expression of fluorescence, while the negative control of parental cells show no fluorescent expression.

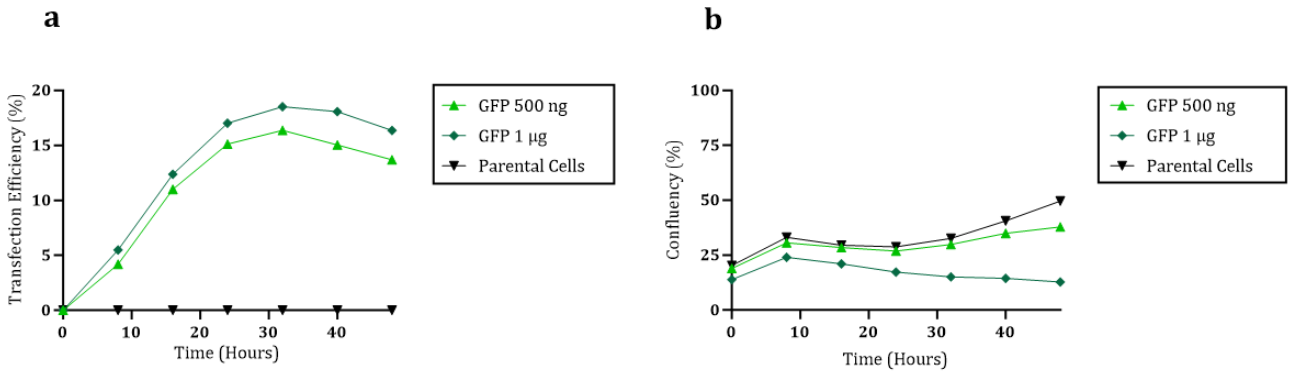
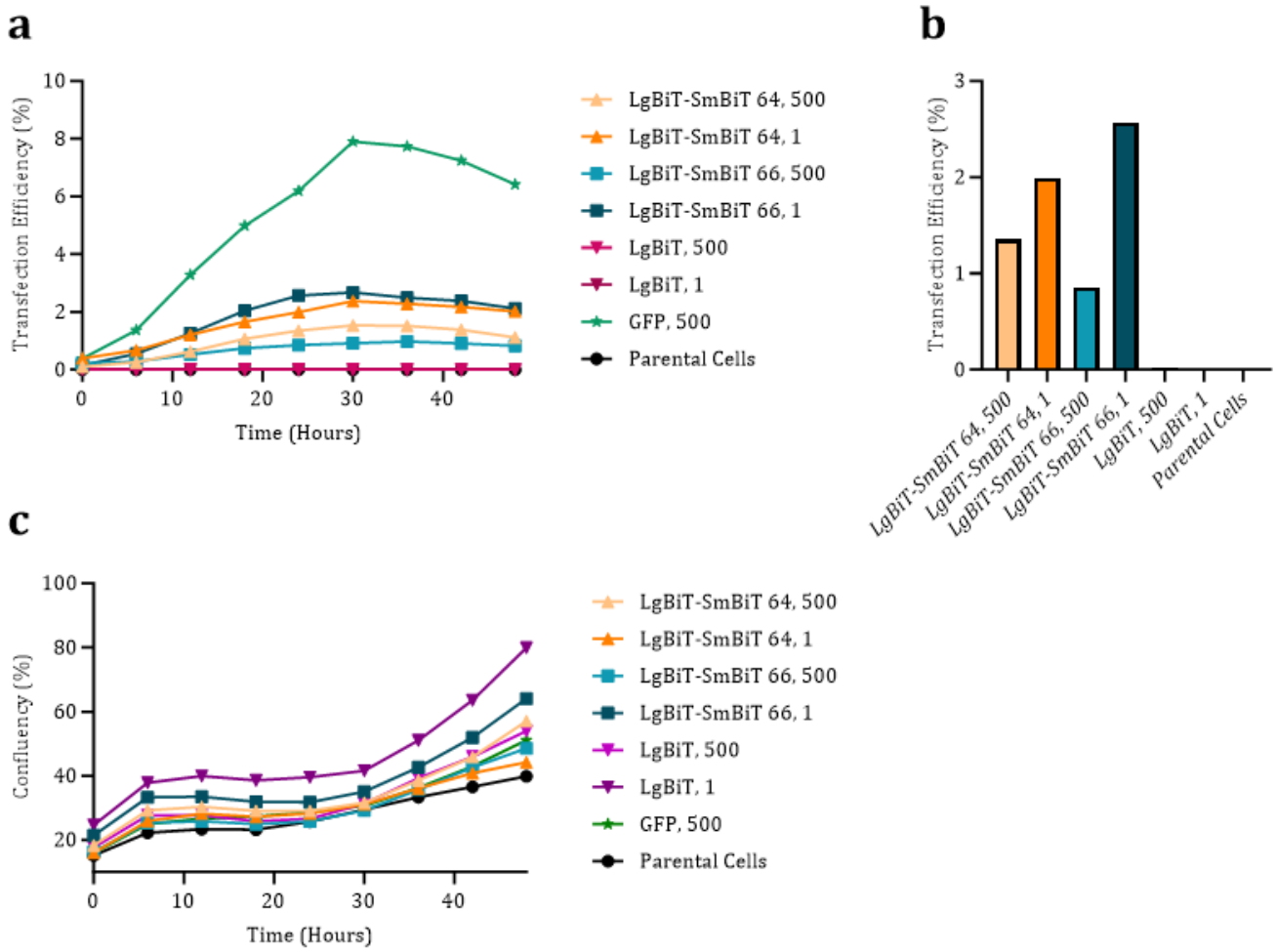


Figure 16: **a,** Graph of the transfection efficiency plotted against time after transfection. **b,** Graph of the confluency against hours post transfection. Graphs created in GraphPad Prism version 10.01.2.

After transfecting with 500 ng and 1 µg of GFP plasmid, the transfections demonstrate relatively low efficiency, ranging from 15% to 20% (Fig. 16). Further refinement of the transfection protocol is advisable. In Figure 16a, both the 500 ng and 1 µg quantities exhibit similar transfection efficiencies. In Figure 16a it can also be observed that the cells reach their peak transfection efficiency around 24 to 32 hours after transfection, suggesting this timeframe as optimal for conducting the NanoBiT® assay.

4.6 Test of a Previously Developed NanoBiT® System

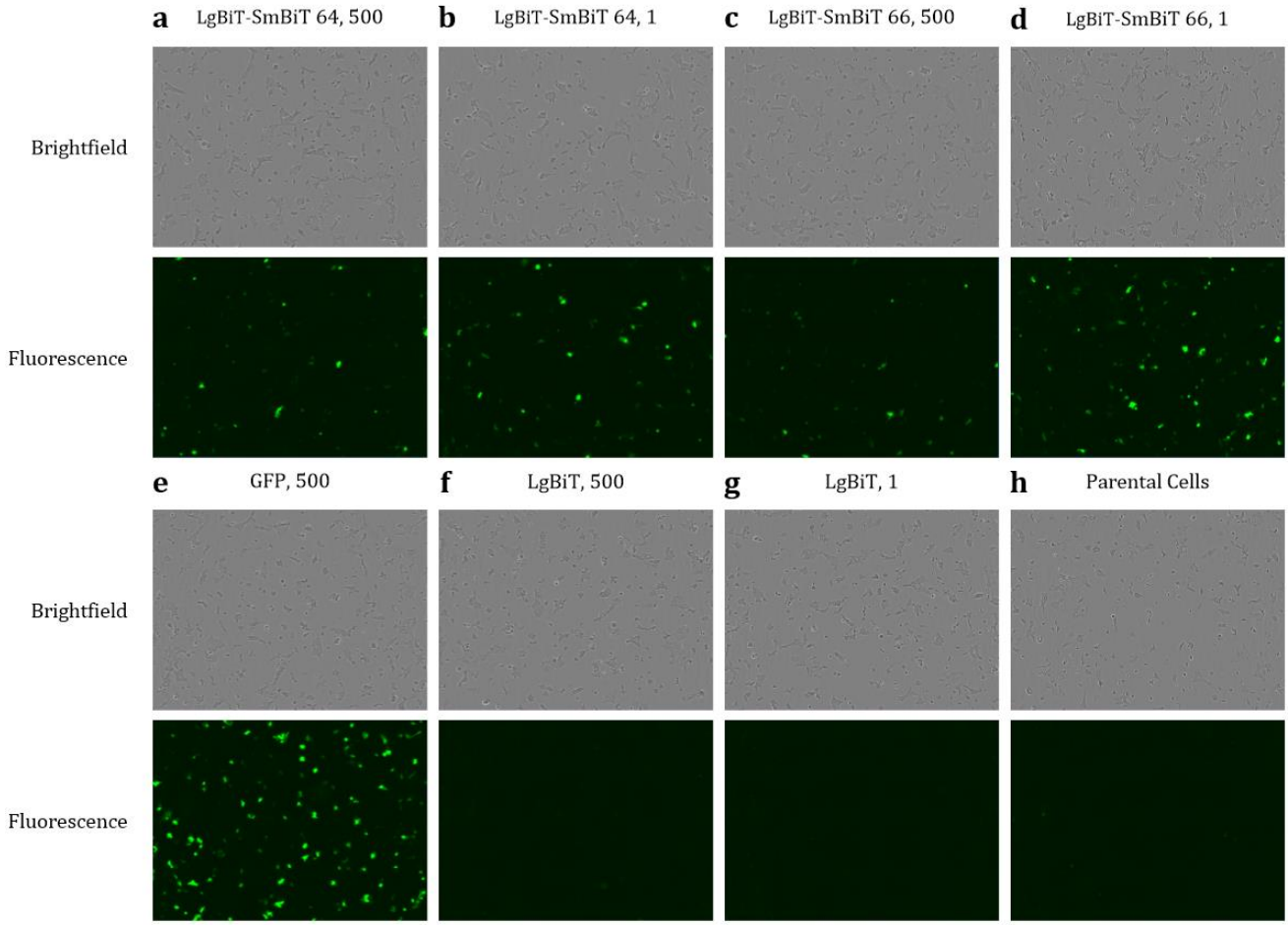
It was decided to evaluate a previously established NanoBiT® system to determine if any specific signal could be detected. This system comprised a LgBiT plasmid and two SmBiT plasmids: 64 and 66. One of these plasmids harbours a mutation in the SmBiT gene, which hinders the complementation of LgBiT-SmBiT and serves as a negative control. Additionally, these SmBiT plasmids included a reporter gene encoding GFP, allowing for the monitoring of transfection efficiency. Figures 17 and 18 illustrate the monitoring of the SmBiT transfection efficiency.



500= HEK-293 cells transfected with 500 ng LgBiT and 64/66-SmBiT, or GFP plasmid, **1**= HEK-293 cells transfected with 1 µg LgBiT and 64/66-SmBiT plasmid, **Parental cells**= HEK-293 cells.

Figure 17: Monitoring of the HEK-293 cells transiently transfected with the LgBiT, SmBiT 64 and 64 constructs. **a,** Graph over the transfection efficiency **b**, Transfection efficiency at the timepoint 24h. **c**, The confluency monitored over time. Graphs created in GraphPad Prism version 10.01.2.

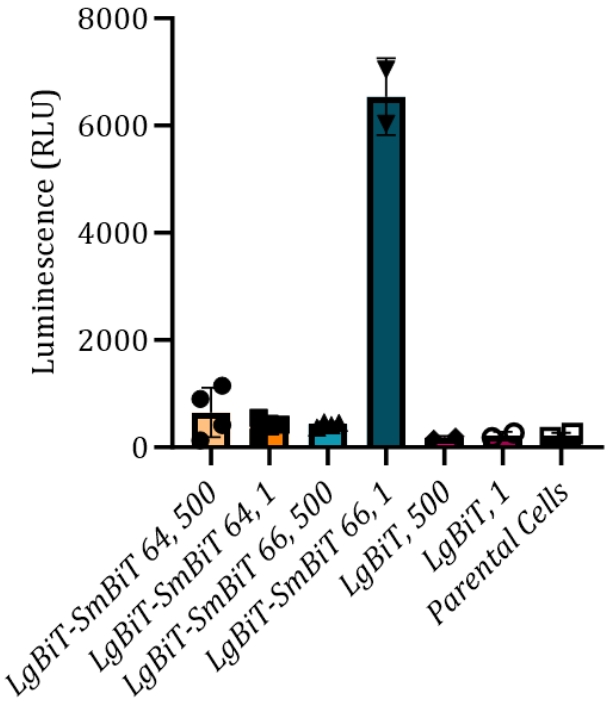
The results presented in Figure 17 demonstrate that the peak transfection efficiency occurs between 24 to 32 hours post-transfection, suggesting that a measurement of the NanoBiT® assay at 24 hours post-transfection is appropriate. This corroborates the findings presented in section 4.5.



500= HEK-293 cells transfected with 500 ng LgBiT and 64/66-SmBiT, or GFP plasmid, **1**= HEK-293 cells transfected with 1 µg LgBiT and 64/66-SmBiT plasmid, **Parental cells**= HEK-293 cells.

Figure 18: Monitoring of the transfection efficiency of the SmBiT 64 and 66 plasmids. Transfection with a GFP plasmid was also included. **a-d**, SmBiT 64 and 66 transfection efficiency. **e**, GFP transfection efficiency. **f-g**, Negative controls including HEK-293 cells only transfected with the LgBiT and parental cells. The microscopy images were obtained using the Incucyte® S3 Software (v2022B).

In Figure 18, the three negative controls show no fluoresent expression. The GFP control exhibit the highest fluorescent expression, while all the combinations of LgBiT-SmBiT shows similar, but a bit lower expression of fluorescence. The decreased transfection efficiency observed, in contrast to the previous experiment described in section 4.5, may be partially due to the cells being plated three days prior to transfection, instead of the typical one-day interval. This deviation from the standard protocol was implemented due to time constraints associated with the project, emphasizing the testing, and troubleshooting of the system to confirm the detection of LgBiT-SmBiT complementation signals. Moreover, the GFP plasmid demonstrates greater transfection efficiency in comparison to the SmBiT plasmids. This is likely due to the larger size of the SmBiT plasmids, which typically results in reduced transfection efficiency. Next, Figure 19 illustrates the luminescent readout of the assay.



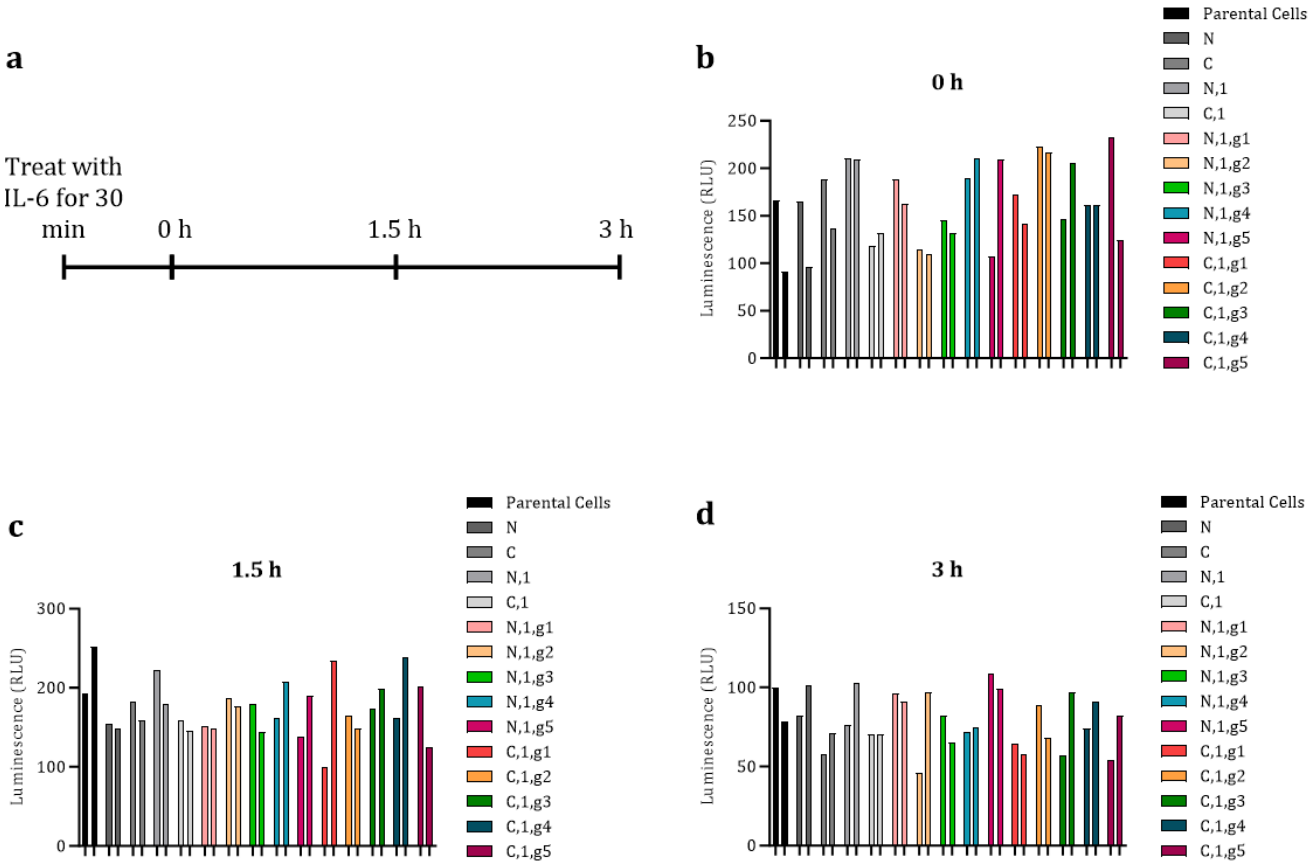
500= HEK-293 cells transfected with 500 ng LgBiT and 64/66-SmBiT plasmid, **1**= HEK-293 cells transfected with 1 µg LgBiT and 64/66-SmBiT plasmid, **Parental cells**= HEK-293 cells. **RLU**: Relative light units.

Figure 19: Luminescence readout of the assay. Graphs created in GraphPad Prism version 10.01.2.

Despite the low transfection efficiency, a luminescent signal around 6,000-7,000 RLU was detected in cells transfected with 1 µg of the LgBiT and SmBiT 66 plasmids. These results indicate an interaction and complementation of the LgBiT and SmBiT subunits.

4.7 NanoBiT® Assay After Stimulating the Cells with IL-6

Herein the results from the assay where cells were stimulated with 100 ng/mL of IL-6 (stock 100 µg/mL) are presented. The cells were treated with IL-6 for 30 min prior to measuring the luminescence. The controls and transfected cells were plated in duplicate, with one well left unstimulated and one stimulated with IL-6. Next, the luminescence was measured at three different timepoints: 0h, 1.5h and 3h which can be observed in Figure 20.



C= HEK-293-dCas9-LgBiT-C-terminus cells, **N**= HEK-293-dCas9-LgBiT-N-terminus cells, **1**=1 µg STAT3-SmBiT plasmid, **g1-5**= sgRNAs designed for the downstream target gene IL-23A, **Parental cells**= HEK-293 cells. **RLU**: Relative light units.

Figure 20: The cells were treated with IL-6 100 ng/mL. The measured luminescence at three different time points b-d. **a,** Overview of the timelines for treatment and measurements. **b,** Assay at the timepoint 0 h. Each sample have measured luminescence for unstimulated cells followed by cells stimulated with IL-6. A measurement interval time of 0.5 s. **c,** Assay at the timepoint 1.5 h. Each sample have measured luminescence for unstimulated cells followed by cells stimulated with IL-6. A measurement interval time of 0.5 s. **d,** Assay at the timepoint 3 h. Each sample have measured luminescence for unstimulated cells followed by cells stimulated with IL-6. A measurement interval time of 0.5 s. Graphs created in GraphPad Prism version 10.01.2.

The IL-6 stimulation did not show any notable effect on the signal. The measurements are interpreted as background signal, indicating the absence of specific protein-protein interactions and LgBiT-SmBiT complementation.

5

Discussion

5.1 Interpretation of Results

5.1.1 Generation of Assay Reagents

After unsuccessful antibiotic selections of LNCaP cells that stably express the dCas9-LgBiT, the transfection conditions and efficiency were investigated. The results from the transfection efficiency of LNCaP cells using the Neon™ electroporation method showed that the prostate cancer cells experienced higher DNA-induced toxicity from the dCas9-LgBiT plasmids than of the GFP plasmid. Electrotransfer of DNA into cells is an efficient method, but integration of larger plasmids can result in low viability and low transfection efficiency of the cells. However, the lengths of the GFP and dCas9-LgBiT plasmids are similar, so the toxicity issue is believed to be with the specific plasmids expressing dCas9-LgBiT. It is possible that the expression of the transgene is toxic to cell type. Overexpression of certain proteins can disrupt cellular processes and cause cytotoxicity. Studies in *E. coli* have shown that overexpression of the protein dCas9 can be toxic and can cause morphological defects of the cells. [58, 59]

Some adjustments in the transfection protocol were made and two different electroporation methods were evaluated. Some cells survived the selection pressure and eventually a few colonies formed. However, this was not enough to expand the LNCaP cells. Once the cells were replated, and grew sparsely, they became unhealthier the longer they were cultured. Once selection of edited cells failed again, it was decided to continue the project using the HEK-293-dCas9-LgBiT stable cell lines. The transfection conditions, such as amount of DNA and transfection reagents, need to be further optimized for the prostate cancer cell line LNCaP. Furthermore, different approaches of integrating the DNA into the cells could be assessed. While some cell types are more sensitive to electroporations, other methods might be more efficient. The following review underscores the importance of considering cell type when choosing transfection method, as variations in cellullar properties directly influence transfection efficiency [48]. Biological methods such as viral transduction or chemical transfection such as lipofectamine could be evaluated. However, due to the time limitation of this project, it was decided to not continue with generating LNCaP cell lines stably expressing dCas9-LgBiT. In addition, the NanoBiT® assay could be assessed by transiently transfecting the LNCaP cells with all the constructs.

HEK-293 cell lines stably expressing the dCas9-LgBiT-N/C-Terminus was successfully established. By western blot, it could be confirmed that cell lysates of these stable pools express LgBiT fused dCas9. These lines were also consistent with the cell line of origin and were negative for mycoplasma. Having these cell lines authenticated and validated is valuable when proceeding with developing the assay.

5.1.2 Development of the NanoBiT® Assay

Following several cell-based assays testing different combinations of dCas9-LgBiT plasmid, TF-SmBiT plasmid, and sgRNAs, no specific luminescent signal was detected. All measurements in these assays were interpreted as background signal. The absence of a luminescent signal suggest that the TF does not bind in close proximity to where the sgRNA guides the dCas9-LgBiT, hence no interaction was detected.

Since this is development of a new system, which lack positive controls, troubleshooting is challenging. Therefore, a previously established system featuring LgBiT and two SmBiTs, serving as a positive and negative controls, was examined to understand what type of signals could be anticipated for specific interactions. One of the SmBiT plasmids contains a point mutation in the gene expressing SmBiT, hindering LgBiT-SmBiT complementation and thereby serving as a negative control. The outcome from this assay was measured luminescent signals around 6,000-7,000 RLU in cells transfected with 1 µg of the LgBiT and SmBiT 66 plasmids. This implies an interaction and complementation between the LgBiT and SmBiT subunits. Based on these results, it is inferred that the initial tests of the developed system did not demonstrate any specific interactions, suggesting the need for additional testing with more sgRNAs and/or sgRNAs targeting promoter regions of other downstream target genes.

It was decided to evaluate the assay after pre-stimulating the cells with IL-6, using a similar approach as described in the following study [60]. In this study they found that when IL-6 binds to its receptor, the IL-6R forms a complex with the epidermal growth factor receptor. This interaction prolongs the activation of STAT3. They quatified phosphorylated STAT3 levels using western blotting to demonstrate that the IL-6R and epidermal growth factor receptor interaction leads to sustained STAT3 activation. So, upon IL-6 binding to its receptor, it triggers a signalling cascade that leads to the phosphorylation and activation of STAT3. Exposing the cells to IL-6 results in STAT3 activation, facilitating its dimerization and translocation into the nucleus. This ensures that the TF of interest is in its active state, ready to bind to its target DNA sequences. Therefore, stimulating the cells was anticipated to increase the likelihood of detecting STAT3-DNA interactions. However, this assay did not show any specific interactions based on the assay readout. These results indicate that the TF does not bind close to the sgRNAs designed targeting the promoter region of the downstream target gene IL-23A.

5.2 Limitations

While cell-based assays are powerful in drug discovery and in understanding cellular mechanisms, there are several challenges associated with these types of assays [61]. One such challenge is cell line variability, where cells may respond differently to the same stimulus, leading to variability in assay results. This variability can arise from factors such as the passage number. Additionally, varying cell culture conditions such as temperature, humidity, and nutrient availability when maintaining cells in culture can affect cell behaviour and assay outcomes. The assay-to-assay reproducibility, where achieving consistent and reproducible results across different experiments can be challenging due to inherent variability in cell culture, reagents, and experimental conditions. Ensuring the purity of cell lines is also crucial for reliable assay results and contaminations of cell lines can compromise assay outcomes. Developing cell-based assays can be resource-intensive in terms of time, labour, and cost.

Some cell-based assays are intricate, making them technically challenging to develop and interpret. In such instances, the inclusion of positive and negative controls is valuable and can assist in the result interpretation. Controls are essential for distinguishing specific interactions from background signals in assays. Several negative controls including parental cells, cells only expressing the dCas9-LgBiT construct, and cells transfected with dCas9-LgBiT and TF-SmBiT but without sgRNA, were included in the assays. However, the assays lacked positive controls. Without them, it becomes challenging to assess whether the assay is functioning correctly and is detecting the intended signal.

The parameters for genetically modifying cell lines typically vary, and transferring the same conditions between different cell lines could be challenging. In certain cell types, it might be necessary to amplify the promoter driving the expression of the constructs. In addition, cell lines like LNCaP used in this project, may pose challenges for transfection. Achieving successful transfection in such cell models often requires careful optimization of transfection conditions and reagents. Preparing these necessary reagents for the assay can be a time-consuming and labour-intensive process, particularly when dealing with cell lines that exhibit high sensitivity to transfection. These considerations highlight the importance of tailoring transfection protocols to suit the cell line being studied. As a result, some cell lines may prove to be more difficult to study using the developed system. Furthermore, developing an assay based on overexpression of TFs pose a risk regarding preservation of the biological function. Representative cell lines with natural levels of TFs rather than engineered cell lines with overexpression of protein might be crucial in understanding and modelling disease.

Another limitation of the developed system is that the design of sgRNAs targeting downstream target genes of the TFs can be challenging, especially if the promoter regions are not well-known. The endogenous promoter region of downstream targets associated with the TFs needs to be identified. Furthermore, TFBSs are often located in regions of the genome with high sequence

complexity, making it difficult to identify suitable sgRNA target sites with high specificity. Designing sgRNAs targeting TFBS requires an understanding of both the TF binding motifs and the genomic context, along with the use of computational tools to predict sgRNA efficacy and specificity. Today there exist several predictive tools for designing sgRNAs, ensuring high specificity and efficiency, while minimizing off-target effects [62-64]. Nevertheless, testing of only 5 guides per downstream target gene within a rather narrow range might not be successful. Additional testing of guides could be required to find and identify regions close to where the TF bind. Ideally, an identified promoter region of a downstream target gene could be scanned with sgRNAs covering this whole region.

Utilizing luciferase enzyme split reporter systems poses challenges due to susceptibility to false positives and false negatives when testing compounds effect. False positives may arise from interference caused by luciferase inhibitors or enzymatic activity, while false negatives can result from compound autofluorescence. Also, spontaneous association between luciferase subunits can diminish the assay window for compound treatment, potentially leading to false positives. Additionally, once measuring the luminescent signal, it is crucial to recognize that RLU are not standardized across instruments, requiring caution when comparing RLU values between different setups. To ensure high accuracy, subtracting background from raw RLU values is recommended upon identification of specific interactions.

5.3 Outlook

The NanoBiT® assay could enable live cell measurements of TF-DNA interactions in real-time, providing dynamic insights into the kinetics and dynamics of transcriptional regulation. This split complementation assay is also fully reversible. In contrast, methods like ChIP typically provide snapshots of TF binding events. ChIP assays, notably ChIP-seq, provide valuable insights into TF interactions with DNA, offering moderate to high throughput. However, their practical throughput is constrained by the complexity of protocols and the need for sequencing. Despite their effectiveness, ChIP assays are costly due to reagents, antibodies, and labour-intensive procedures [37]. In contrast, split-reporter system assays, like split-luciferase, are highly adaptable to HTS, accommodating hundreds to thousands of compounds simultaneously. These assays are also often compatible with automated liquid handling systems, making them efficient and less labour-intensive. The costs associated with split-reporter assays are comparatively lower, primarily covering reporter constructs, cell culture, and detection reagents. These advantages could make split reporter systems a valuable complement to traditional methods like ChIP, SELEX, EMSA and PBMs in the study of TF-DNA interactions.

The NanoBiT® technology has been applied in numerous studies, demonstrating its versatility and sensitivity across various biological applications. One study employs a CRISPR-Cas9-based split luciferase biosensor with NanoBiT® to image unique DNA sequences in individual cells, enabling precise, real-time visualization of genetic elements [65]. Another study uses NanoBiT®

to track membrane protein trafficking, essential for drug discovery and development, by monitoring protein movement and localization in response to drug candidates [66]. Additionally, the technology has been used to study the localization and interaction of AR and its splice variant AR-V7. Tagging these proteins with split luciferase fragments allows researchers to observe homo- and heterodimer formation, providing insights into their physiological and pathological roles [67]. Split NanoLuc technology has also been employed in research that quantifies interactions between the PII protein and its partners, revealing transient interactions with unprecedented sensitivity [68]. Another study demonstrated that optimization of NanoBiT® enhances its ability to measure protein interactions within cells, detecting even transient and weak interactions with high sensitivity and quantitative accuracy [69]. The NanoBiT® technology can significantly advance biological research and therapeutic development and holds potential for additional applications that have yet to be explored.

In the development of the system in this study the next step should be to test more sgRNAs and sgRNAs targeting promoter regions of different downstream target genes. Subsequently, the TF-SmBiT plasmids should be verified. By using *e.g.*, sanger sequencing it can be confirmed that the constructs are correct and that the assay reagents are accurate.

Once specific interactions of TF-DNA binding are identified, it is advisable to incorporate tool compounds that modulate PPIs in a dose-response fashion during assay development. Employing these compounds as controls can confirm the functionality of the assay and improve its efficacy. Confirmation of a specific interaction can be achieved by observing the anticipated change in luminescence following treatment with a tool compound.

The developed assay is anticipated to be applicable using different cell models and TFs. Further cell models and TFs should be evaluated to confirm that the assay can be adapted for different cell lines and TFs. In addition, any DNA binding protein, like scaffolding proteins, could also be compatible with this assay which can be explored in future studies.

6

Conclusion

This project was undertaken to develop a novel assay for quantifying TF-DNA interactions levering a split-reporter system and luminescence measurement. Stable HEK-293 cell lines expressing dCas9 in connection to either the N- or C-Terminus of the LgBiT were established and confirmed through western blots. After all assay reagents were generated the NanoBiT® assay was developed. Specific interactions were not demonstrated during initial testing with the first range of sgRNAs. However, despite its exploratory nature, this study provides some insight into the potential applications of the developed system. By designing and testing more sgRNAs and sgRNAs targeting other downstream target genes, the developed system holds promise for efficiently assessing TF-DNA binding in high throughput, identifying TFBSs, and evaluating compound effects on TF-DNA interactions. Nevertheless, additional validation is crucial to ensure the assays reliability and suitability for future studies. Future research should focus on exploring the assays versatility and compatibility with various cell models and TFs. Furthermore, testing tool compounds is essential to confirm the systems functionality and sensitivity in identifying compound effects on TF-DNA binding. These steps will contribute to enhancing the understanding and utility of the developed assay.

Bibliography

[1] M. J. Henley and A. N. Koehler, "Advances in targeting 'undruggable' transcription factors with small molecules," *Nature Reviews Drug Discovery,* vol. 20, no. 9, pp. 669-688, 2021/09/01 2021, doi: 10.1038/s41573-021-00199-0.

[2] H. He *et al.*, "Mechanisms and biotechnological applications of transcription factors," *Synthetic and Systems Biotechnology,* vol. 8, no. 4, pp. 565-577, 2023/12/01/ 2023, doi: https://doi.org/10.1016/j.synbio.2023.08.006.

[3] A. Chen and A. N. Koehler, "Transcription Factor Inhibition: Lessons Learned and Emerging Targets," (in eng), *Trends Mol Med,* vol. 26, no. 5, pp. 508-518, May 2020, doi: 10.1016/j.molmed.2020.01.004.

[4] T. I. Lee and R. A. Young, "Transcriptional regulation and its misregulation in disease," (in eng), *Cell,* vol. 152, no. 6, pp. 1237-51, Mar 14 2013, doi: 10.1016/j.cell.2013.02.014.

[5] J. H. Bushweller, "Targeting transcription factors in cancer — from undruggable to reality," *Nature Reviews Cancer,* vol. 19, no. 11, pp. 611-624, 2019/11/01 2019, doi: 10.1038/s41568-019-0196-7.

[6] R. Mundade, H. G. Ozer, H. Wei, L. Prabhu, and T. Lu, "Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond," (in eng), *Cell Cycle,* vol. 13, no. 18, pp. 2847-52, 2014, doi: 10.4161/15384101.2014.949201.

[7] S. A. Lambert *et al.*, "The Human Transcription Factors," (in eng), *Cell,* vol. 172, no. 4, pp. 650-665, Feb 8 2018, doi: 10.1016/j.cell.2018.01.029.

[8] M. Geertz and S. J. Maerkl, "Experimental strategies for studying transcription factor-DNA binding specificities," (in eng), *Brief Funct Genomics,* vol. 9, no. 5-6, pp. 362-73, Dec 2010, doi: 10.1093/bfgp/elq023.

[9] F. Fontaine, J. Overman, and M. François, "Pharmacological manipulation of transcription factor protein-protein interactions: opportunities and obstacles," (in eng), *Cell Regen,* vol. 4, no. 1, p. 2, 2015, doi: 10.1186/s13619-015-0015-x.

[10] Y. S. Ma *et al.*, "Paving the way for small-molecule drug discovery," (in eng), *Am J Transl Res,* vol. 13, no. 3, pp. 853-870, 2021.

[11] J. Shaw *et al.*, "Determining direct binders of the Androgen Receptor using a high-throughput Cellular Thermal Shift Assay," *Scientific Reports,* vol. 8, no. 1, p. 163, 2018/01/09 2018, doi: 10.1038/s41598-017-18650-x.

[12] A. Casamassimi and A. Ciccodicola, "Transcriptional Regulation: Molecules, Involved Mechanisms, and Misregulation," (in eng), *Int J Mol Sci,* vol. 20, no. 6, Mar 14 2019, doi: 10.3390/ijms20061281.

[13] A. H. Brivanlou and J. E. Darnell, Jr., "Signal transduction and the control of gene expression," (in eng), *Science,* vol. 295, no. 5556, pp. 813-8, Feb 1 2002, doi: 10.1126/science.1066355.

[14] B. J. Greber and E. Nogales, "The Structures of Eukaryotic Transcription Pre-initiation Complexes and Their Functional Implications," (in eng), *Subcell Biochem,* vol. 93, pp. 143-192, 2019, doi: 10.1007/978-3-030-28151-9_5.

[15] T. C. Voss and G. L. Hager, "Dynamic regulation of transcriptional states by chromatin and transcription factors," *Nature Reviews Genetics,* vol. 15, no. 2, pp. 69-81, 2014/02/01 2014, doi: 10.1038/nrg3623.

[16] J. Joung *et al.*, "A transcription factor atlas of directed differentiation," *Cell,* vol. 186, no. 1, pp. 209-229.e26, 2023/01/05/ 2023, doi: https://doi.org/10.1016/j.cell.2022.11.026.
[17] D. M. Suter, "Transcription Factors and DNA Play Hide and Seek," (in eng), *Trends Cell Biol,* vol. 30, no. 6, pp. 491-500, Jun 2020, doi: 10.1016/j.tcb.2020.03.003.
[18] R. D'Oliveira Albanus *et al.*, "Chromatin information content landscapes inform transcription factor and DNA interactions," *Nature Communications,* vol. 12, no. 1, p. 1307, 2021/02/26 2021, doi: 10.1038/s41467-021-21534-4.
[19] A. Jolma and J. Taipale, "Methods for Analysis of Transcription Factor DNA-Binding Specificity In Vitro," (in eng), *Subcell Biochem,* vol. 52, pp. 155-73, 2011, doi: 10.1007/978-90-481-9069-0_7.
[20] L. Isbel, R. S. Grand, and D. Schübeler, "Generating specificity in genome regulation through transcription factor sensitivity to chromatin," *Nature Reviews Genetics,* vol. 23, no. 12, pp. 728-740, 2022/12/01 2022, doi: 10.1038/s41576-022-00512-6.
[21] C. Felipe, J. Shin, and A. B. Kolomeisky, "How Pioneer Transcription Factors Search for Target Sites on Nucleosomal DNA," *The Journal of Physical Chemistry B,* vol. 126, no. 22, pp. 4061-4068, 2022/06/09 2022, doi: 10.1021/acs.jpcb.2c01931.
[22] O. Oksuz *et al.*, "Transcription factors interact with RNA to regulate genes," (in eng), *Mol Cell,* vol. 83, no. 14, pp. 2449-2463.e13, Jul 20 2023, doi: 10.1016/j.molcel.2023.06.012.
[23] H. Gronemeyer, J. A. Gustafsson, and V. Laudet, "Principles for modulation of the nuclear receptor superfamily," (in eng), *Nat Rev Drug Discov,* vol. 3, no. 11, pp. 950-64, Nov 2004, doi: 10.1038/nrd1551.
[24] B. A. Porter, M. A. Ortiz, G. Bratslavsky, and L. Kotula, "Structure and Function of the Nuclear Receptor Superfamily and Current Targeted Therapies of Prostate Cancer," (in eng), *Cancers (Basel),* vol. 11, no. 12, Nov 23 2019, doi: 10.3390/cancers11121852.
[25] D. E. Levy and C. K. Lee, "What does Stat3 do?," (in eng), *J Clin Invest,* vol. 109, no. 9, pp. 1143-8, May 2002, doi: 10.1172/jci15650.
[26] S. Naqvi *et al.*, "Precise modulation of transcription factor levels identifies features underlying dosage sensitivity," *Nature Genetics,* vol. 55, no. 5, pp. 841-851, 2023/05/01 2023, doi: 10.1038/s41588-023-01366-2.
[27] P. Weidemüller, M. Kholmatov, E. Petsalaki, and J. B. Zaugg, "Transcription factors: Bridge between cell signaling and gene regulation," (in eng), *Proteomics,* vol. 21, no. 23-24, p. e2000034, Dec 2021, doi: 10.1002/pmic.202000034.
[28] F. Spitz and E. E. M. Furlong, "Transcription factors: from enhancer binding to developmental control," *Nature Reviews Genetics,* vol. 13, no. 9, pp. 613-626, 2012/09/01 2012, doi: 10.1038/nrg3207.
[29] M. Már, K. Nitsenko, and P. O. Heidarsson, "Multifunctional Intrinsically Disordered Regions in Transcription Factors," (in eng), *Chemistry,* vol. 29, no. 21, p. e202203369, Apr 13 2023, doi: 10.1002/chem.202203369.
[30] C. Rastogi *et al.*, "Accurate and sensitive quantification of protein-DNA binding affinity," *Proceedings of the National Academy of Sciences,* vol. 115, no. 16, pp. E3692-E3701, 2018/04/17 2018, doi: 10.1073/pnas.1714376115.
[31] J. Reddy *et al.*, "Predicting master transcription factors from pan-cancer expression data," *Science Advances,* vol. 7, no. 48, p. eabf6123, 2021, doi: doi:10.1126/sciadv.abf6123.
[32] J. E. Darnell, "Transcription factors as targets for cancer therapy," *Nature Reviews Cancer,* vol. 2, no. 10, pp. 740-749, 2002/10/01 2002, doi: 10.1038/nrc906.

[33] Y. Li, J. Song, P. Zhou, J. Zhou, and S. Xie, "Targeting Undruggable Transcription Factors with PROTACs: Advances and Perspectives," (in eng), *J Med Chem,* vol. 65, no. 15, pp. 10183-10194, Aug 11 2022, doi: 10.1021/acs.jmedchem.2c00691.

[34] J.-j. Zhuang, Q. Liu, D.-l. Wu, and L. Tie, "Current strategies and progress for targeting the "undruggable" transcription factors," *Acta Pharmacologica Sinica,* vol. 43, no. 10, pp. 2474-2481, 2022/10/01 2022, doi: 10.1038/s41401-021-00852-9.

[35] B. R. Sabari *et al.*, "Coactivator condensation at super-enhancers links phase separation and gene control," *Science,* vol. 361, no. 6400, p. eaar3958, 2018, doi: doi:10.1126/science.aar3958.

[36] R. T. Pop *et al.*, "Identification of mammalian transcription factors that bind to inaccessible chromatin," *Nucleic Acids Research,* vol. 51, no. 16, pp. 8480-8495, 2023, doi: 10.1093/nar/gkad614.

[37] P. J. Park, "ChIP–seq: advantages and challenges of a maturing technology," *Nature Reviews Genetics,* vol. 10, no. 10, pp. 669-680, 2009/10/01 2009, doi: 10.1038/nrg2641.

[38] M. F. Berger and M. L. Bulyk, "Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors," *Nature Protocols,* vol. 4, no. 3, pp. 393-411, 2009/03/01 2009, doi: 10.1038/nprot.2008.195.

[39] G. B. Kim, Y. Gao, B. O. Palsson, and S. Y. Lee, "DeepTFactor: A deep learning-based tool for the prediction of transcription factors," *Proceedings of the National Academy of Sciences,* vol. 118, no. 2, p. e2021171118, 2021/01/12 2021, doi: 10.1073/pnas.2021171118.

[40] M. Słabicki *et al.*, "Small-molecule-induced polymerization triggers degradation of BCL6," *Nature,* vol. 588, no. 7836, pp. 164-168, 2020/12/01 2020, doi: 10.1038/s41586-020-2925-1.

[41] R. A. Davey and M. Grossmann, "Androgen Receptor Structure, Function and Biology: From Bench to Bedside," (in eng), *Clin Biochem Rev,* vol. 37, no. 1, pp. 3-15, Feb 2016.

[42] J. E. Bradner, D. Hnisz, and R. A. Young, "Transcriptional Addiction in Cancer," (in eng), *Cell,* vol. 168, no. 4, pp. 629-643, Feb 9 2017, doi: 10.1016/j.cell.2016.12.013.

[43] J. Xie *et al.*, "Targeting androgen receptor phase separation to overcome antiandrogen resistance," *Nature Chemical Biology,* vol. 18, no. 12, pp. 1341-1350, 2022/12/01 2022, doi: 10.1038/s41589-022-01151-y.

[44] A. R. Michmerhuizen, D. E. Spratt, L. J. Pierce, and C. W. Speers, "ARe we there yet? Understanding androgen receptor signaling in breast cancer," *npj Breast Cancer,* vol. 6, no. 1, p. 47, 2020/09/25 2020, doi: 10.1038/s41523-020-00190-9.

[45] A. Salas *et al.*, "JAK–STAT pathway targeting for the treatment of inflammatory bowel disease," *Nature Reviews Gastroenterology & Hepatology,* vol. 17, no. 6, pp. 323-337, 2020/06/01 2020, doi: 10.1038/s41575-020-0273-0.

[46] M. A. Castanares *et al.*, "Characterization of a novel metastatic prostate cancer cell line of LNCaP origin," (in eng), *Prostate,* vol. 76, no. 2, pp. 215-25, Feb 2016, doi: 10.1002/pros.23115.

[47] A. A. Stepanenko and V. V. Dmitrenko, "HEK293 in cell biology and cancer research: phenotype, karyotype, tumorigenicity, and stress-induced genome-phenotype evolution," *Gene,* vol. 569, no. 2, pp. 182-190, 2015/09/15/ 2015, doi: https://doi.org/10.1016/j.gene.2015.05.065.

[48] Z. X. Chong, S. K. Yeap, and W. Y. Ho, "Transfection types, methods and strategies: a technical review," (in eng), *PeerJ,* vol. 9, p. e11165, 2021, doi: 10.7717/peerj.11165.

[49] T. K. Kim and J. H. Eberwine, "Mammalian cell transfection: the present and the future," (in eng), *Anal Bioanal Chem,* vol. 397, no. 8, pp. 3173-8, Aug 2010, doi: 10.1007/s00216-010-3821-6.

[50] M. Wade, J. Méndez, N. P. Coussens, M. R. Arkin, and M. A. Glicksman, "Inhibition of Protein-Protein Interactions: Cell-Based Assays," in *Assay Guidance Manual*, S. Markossian *et al.* Eds. Bethesda (MD): Eli Lilly & Company and the National Center for Advancing Translational Sciences, 2004.

[51] H. Potter, "Transfection by electroporation," (in eng), *Curr Protoc Mol Biol,* vol. Chapter 9, p. Unit 9.3, May 2003, doi: 10.1002/0471142727.mb0903s62.

[52] T. Mahmood and P. C. Yang, "Western blot: technique, theory, and trouble shooting," (in eng), *N Am J Med Sci,* vol. 4, no. 9, pp. 429-34, Sep 2012, doi: 10.4103/1947-2714.100998.

[53] D. H. Lackner *et al.*, "A generic strategy for CRISPR-Cas9-mediated gene tagging," *Nature Communications,* vol. 6, no. 1, p. 10237, 2015/12/17 2015, doi: 10.1038/ncomms10237.

[54] C. K. S. Karlson, S. N. Mohd-Noor, N. Nolte, and B. C. Tan, "CRISPR/dCas9-Based Systems: Mechanisms and Applications in Plant Sciences," (in eng), *Plants (Basel),* vol. 10, no. 10, Sep 29 2021, doi: 10.3390/plants10102055.

[55] M. H. Larson, L. A. Gilbert, X. Wang, W. A. Lim, J. S. Weissman, and L. S. Qi, "CRISPR interference (CRISPRi) for sequence-specific control of gene expression," *Nature Protocols,* vol. 8, no. 11, pp. 2180-2196, 2013/11/01 2013, doi: 10.1038/nprot.2013.132.

[56] A. S. Dixon *et al.*, "NanoLuc Complementation Reporter Optimized for Accurate Measurement of Protein Interactions in Cells," (in eng), *ACS Chem Biol,* vol. 11, no. 2, pp. 400-8, Feb 19 2016, doi: 10.1021/acschembio.5b00753.

[57] *NanoBiT® Protein:Protein Interaction System Technical Manual*, Promega Corporation, USA, TM461, 2021, pp. 1-30.

[58] S. Cho, D. Choe, E. Lee, S. C. Kim, B. Palsson, and B.-K. Cho, "High-Level dCas9 Expression Induces Abnormal Cell Morphology in Escherichia coli," *ACS Synthetic Biology,* vol. 7, no. 4, pp. 1085-1094, 2018/04/20 2018, doi: 10.1021/acssynbio.7b00462.

[59] Y. J. Lee, A. Hoynes-O'Connor, M. C. Leong, and T. S. Moon, "Programmable control of bacterial gene expression with the combined CRISPR and antisense RNA system," *Nucleic Acids Research,* vol. 44, no. 5, pp. 2462-2473, 2016, doi: 10.1093/nar/gkw056.

[60] Y. Wang, A. H. H. van Boxel-Dezaire, H. Cheon, J. Yang, and G. R. Stark, "STAT3 activation in response to IL-6 is prolonged by the binding of IL-6 receptor to EGF receptor," *Proceedings of the National Academy of Sciences,* vol. 110, no. 42, pp. 16975-16980, 2013/10/15 2013, doi: 10.1073/pnas.1315862110.

[61] T. L. Riss, R. A. Moravec, S. J. Duellman, and A. L. Niles, "Treating Cells as Reagents to Design Reproducible Assays," *SLAS DISCOVERY: Advancing the Science of Drug Discovery,* vol. 26, no. 10, pp. 1256-1267, 2021/12/01 2021, doi: 10.1177/24725552211039754.

[62] S. Bae, J. Park, and J. S. Kim, "Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases," (in eng), *Bioinformatics,* vol. 30, no. 10, pp. 1473-5, May 15 2014, doi: 10.1093/bioinformatics/btu048.

[63] T. G. Montague, J. M. Cruz, J. A. Gagnon, G. M. Church, and E. Valen, "CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing," (in eng), *Nucleic Acids Res,* vol. 42, no. Web Server issue, pp. W401-7, Jul 2014, doi: 10.1093/nar/gku410.

[64] F. Heigwer, G. Kerr, and M. Boutros, "E-CRISP: fast CRISPR target site identification," *Nature Methods,* vol. 11, no. 2, pp. 122-123, 2014/02/01 2014, doi: 10.1038/nmeth.2812.

[65] N. G. Heath, H. O'Geen, N. B. Halmai, J. E. Corn, and D. J. Segal, "Imaging Unique DNA Sequences in Individual Cells Using a CRISPR-Cas9-Based, Split Luciferase Biosensor," (in eng), *Front Genome Ed,* vol. 4, p. 867390, 2022, doi: 10.3389/fgeed.2022.867390.

[66] A. Reyes-Alcaraz, E. Y. Lucero Garcia-Rojas, E. A. Merlinsky, J. Y. Seong, R. A. Bond, and B. K. McConnell, "A NanoBiT assay to monitor membrane proteins trafficking for drug discovery and drug development," *Communications Biology,* vol. 5, no. 1, p. 212, 2022/03/08 2022, doi: 10.1038/s42003-022-03163-9.

[67] J. Guzman *et al.*, "NanoLuc Binary Technology as a methodological approach: an important new tool for studying the localization of androgen receptor and androgen receptor splice variant V7 homo and heterodimers," *BMC Cancer,* vol. 24, no. 1, p. 346, 2024/03/19 2024, doi: 10.1186/s12885-024-12110-2.

[68] R. Rozbeh and K. Forchhammer, "Split NanoLuc technology allows quantitation of interactions between PII protein and its receptors with unprecedented sensitivity and reveals transient interactions," *Scientific Reports,* vol. 11, no. 1, p. 12535, 2021/06/15 2021, doi: 10.1038/s41598-021-91856-2.

[69] A. S. Dixon *et al.*, "NanoLuc Complementation Reporter Optimized for Accurate Measurement of Protein Interactions in Cells," *ACS Chemical Biology,* vol. 11, no. 2, pp. 400-408, 2016/02/19 2016, doi: 10.1021/acschembio.5b00753.

[70] *AmMagTM Quatro Plasmid Purification Kit Instruction Manual*, GenScript USA, Inc., 2022, pp. 1-7.

[71] *Quick-Start Protocol QIAGEN® Plasmid Mini, Midi and Maxi Kits*, 12162, 12163,12165, 2016.

A

Plasmid Maps

This appendix provides detailed maps of the vectors constructed for this project.

A.1 dCas9-LgBiT

A map of the 6,237 bp cloning vector used to generate the dCas9-LgBiT-N/C-terminus plasmids is provided in Figure A.1 below. The plasmid was constructed using the mammalian expression pIRES backbone. The final dCas9-LgBiT-N/C-terminus plasmids contain sequences encoding the CMV promoter, dCas9, the LgBiT and a neomycin resistance cassette. The plasmid also encodes motifs such as an ampicillin resistance cassette, terminators, the pUC origin of replication, and contain a Kozak sequence, initiating protein translation of mRNA transcripts.

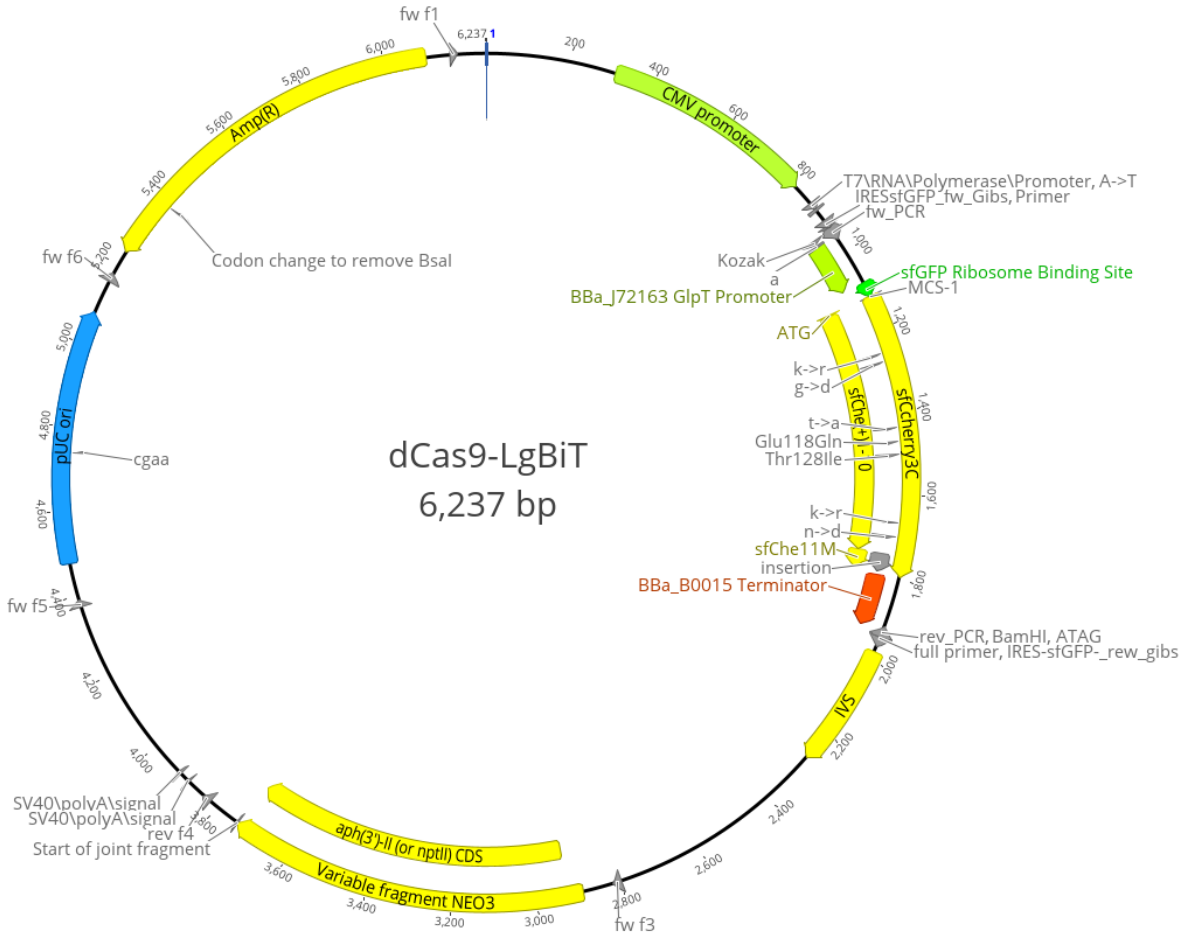


Figure A.1: Map over the dCas9-LgBiT cloning vector. The final plasmid contains coding sequences for the CMV promoter, dCas9, the LgBiT and a neomycin resistance. Vector created in Geneious version 2023.1 created by Biomatters. Available from https://www.geneious.com.

A.2 TF-SmBiT

In Figure A.2.1 the 7,864 bp AR-SmBiT vector with the pRP expression vector backbone is shown. It was designed to include a sequence encoding the EF-1α promoter which drives expression of the gene of interest, in this case the SmBiT connected to AR via a GGGS linker. The vector also includes a Kozak consensus sequence which facilitates translation initiation. Following the coding sequence is the SV40 late pA which terminates transcription of the upstream ORF (open reading frame). Then the CMV promoter drives expression of a downstream marker gene, in this case an antibiotic resistance gene for puromycin. Next is the BGH pA motif which terminates transcription of the upstream ORF. Lastly, the vector includes the pUC origin of replication enabling a high copy number and replication in *E. coli* and a sequence expressing the ampicillin resistance gene.

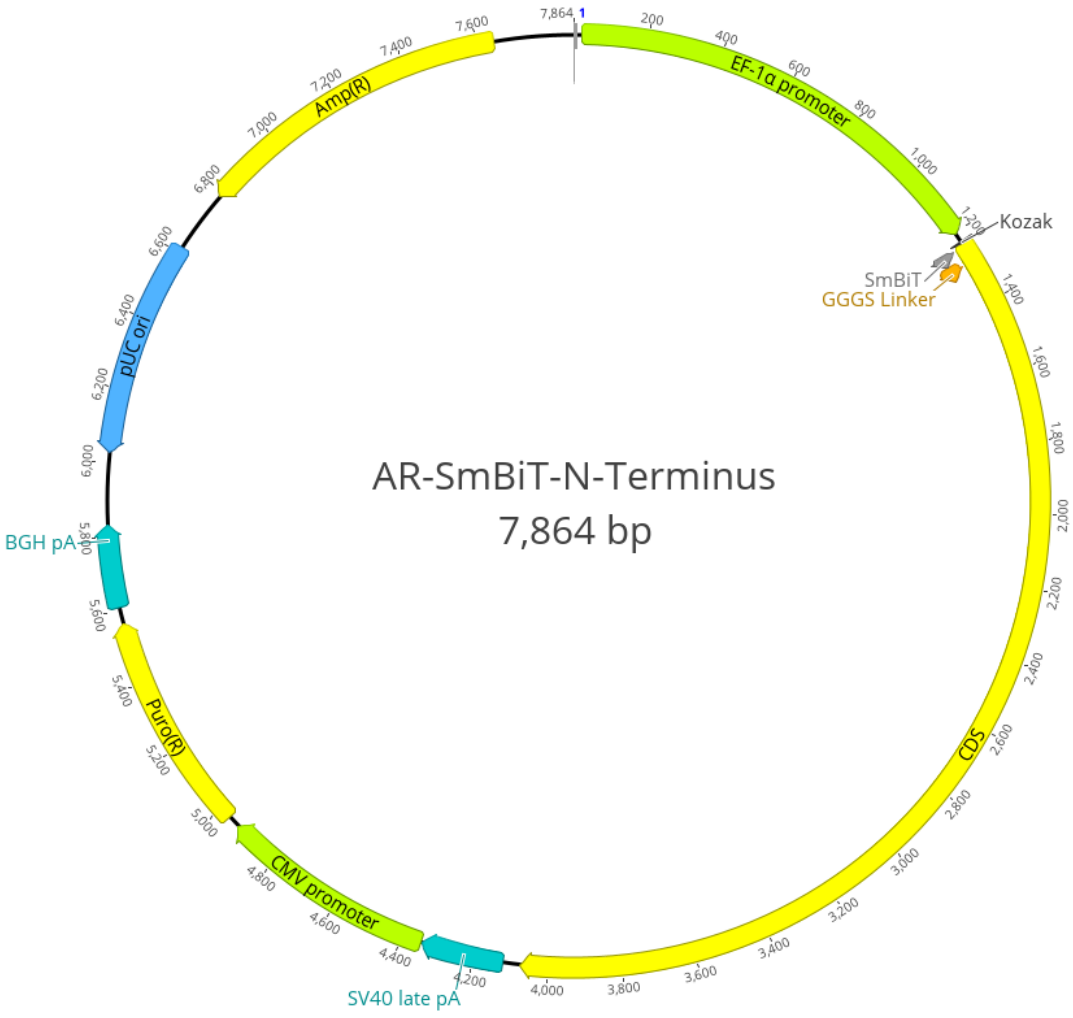


Figure A.2.1: Plasmid map overviewing the elements of the AR-SmBiT plasmid. The plasmid contains coding sequences for the EF-1α promoter, SmBiT, AR, terminators, CMV promoter, a puromycin resistance cassette, pUC origin of replication and an ampicillin resistance marker. Vector created in Geneious version 2023.1 created by Biomatters. Available from https://www.geneious.com. pRP[Exp]-Puro-EF1A>hAR, was constructed and packaged by VectorBuilder. The vector ID is VB231017-1379qcd, which can be used to retrieve detailed information about the vector on https://www.vectorbuilder.com.

In Figure A.2.2, an overview of the STAT3-SmBiT-C-Terminus with a pRP expression vector backbone can be observed. The plasmid is 7,411 bp and contains sequences expressing the EF-1α and CMV promoters, puromycin and ampicillin resistance cassettes, a Kozak sequence, and the pUC origin of replication. The vector also contains coding sequences of STAT3, a linker and the SmBiT. Further, the vector includes expression of SV40 late pA and BGH pA, responsible for termination of transcription and addition of a PolyA tail to the mRNA transcripts.

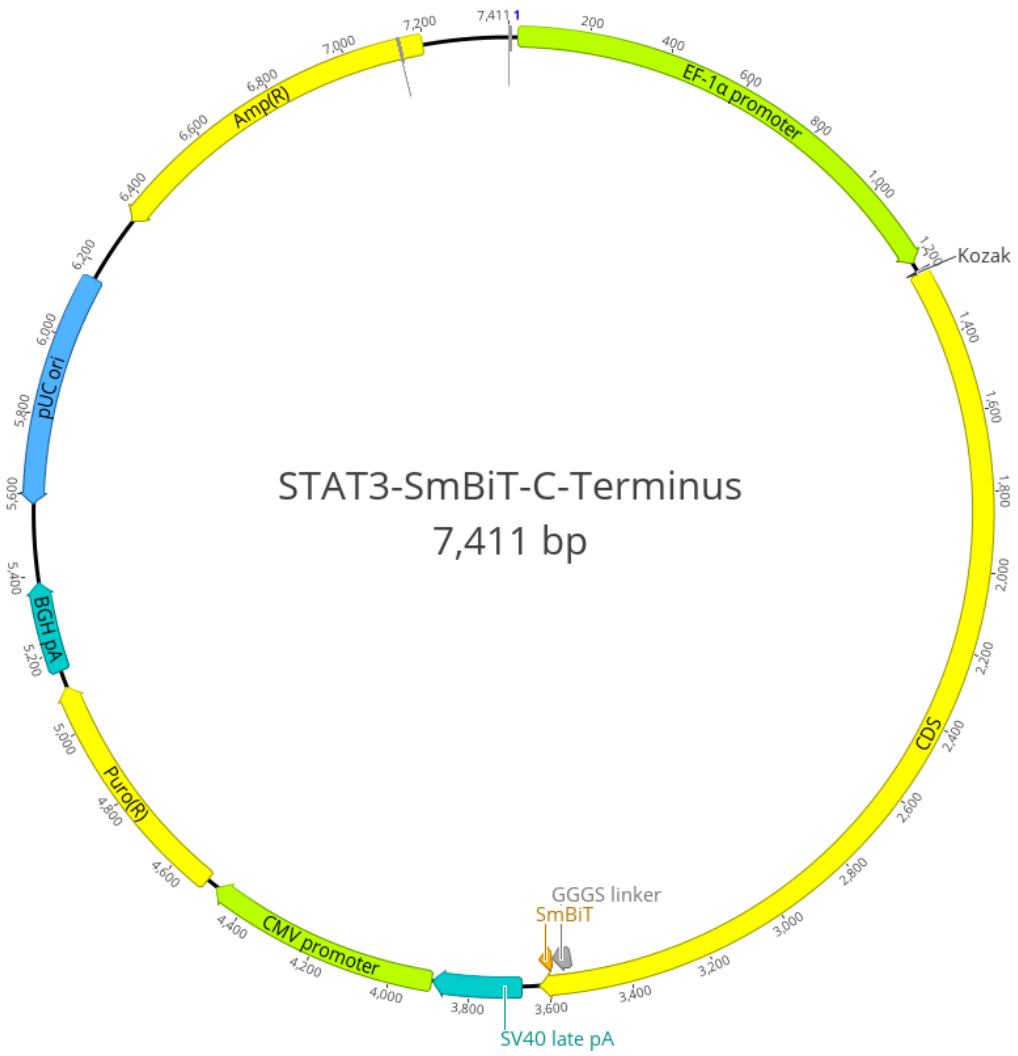


Figure A.2.2: Plasmid map overviewing the elements of the STAT3-SmBiT plasmid. The plasmid contains coding sequences for the EF-1α promoter, SmBiT, STAT3, terminators, CMV promoter, a puromycin resistance cassette, the pUC origin of replication, and an ampicillin resistance marker. Vector created in Geneious version 2023.1 created by Biomatters. Available from https://www.geneious.com. pRP[Exp]-Puro-EF1A>hSTAT3, was constructed and packaged by VectorBuilder. The vector ID is VB231024-1317vkw, which can be used to retrieve detailed information about the vector on https://www.vectorbuilder.com.

B

Additional Laboratory Protocols and Reagents

In this appendix, general protocols, and detailed laboratory procedures for different experiments throughout this project are presented.

B.1 Techniques for Culturing and Maintaining Mammalian Cells *in Vitro*

Herein, general guidelines that were used for maintaining and handling mammalian cells *in vitro,* are presented. This includes thawing, dissociation, passage, and cryopreservation of cells. All laboratory work for growing mammalian cells was according to aseptic techniques to prevent any contaminations. Personal protective equipment (PPE) was used, which includes usage of gloves, safety goggles, lab coats, and shoe covers. The experimental work was conducted in laminar-flow (LAF) hoods. These are biohazard cabinets certified for Level II handling of biological material. All reagents and equipment that might encounter cells were wiped with 70% ethanol prior to bringing them into the sterile environment in the LAF-bench. In the LAF-bench, a vacuum pump with an aspiration inlet was used to aspirate reagents and cell suspensions. To monitor the cells, an inverted microscope with a magnitude of 10-100X was used. Moreover, to measure the cells density, viability, diameter, and number of cells in aggregates, the automatic cell counter ChemoMetec NucleoCounter® NC-3000™ was used. For long-time cryopreservation, cells were stored in -150°C liquid nitrogen freezer. For storing sensitive reagents, a -80°C freezer was used and for storing media and other reagents, a refrigerator 2-8°C was used.

When growing mammalian cells, a complex media for the cell type is required to achieve an optimal combination of nutrients, growth factors, hormones, and attachment factors. For the HEK-293 cells, DMEM/F-12, GlutaMAX™ (Gibco™, cat#10565018) supplemented with 10% FBS was used. For the LNCaP cells, RPMI 1640 Medium (Gibco™, cat#21875034) supplemented with 10% FBS was used. The cells were grown in Corning®Costar® Cell Culture Flasks (T-25-225) (Costar®, cat#430639, 430641U, 431080, 431082) for optimal cell attachment. This adherence to a surface provides optimal cell division and growth. The flasks come in a variety of sizes and some general guidelines for flask volumes of different reagents and flask sizes can be found in Table B.1. The cells were grown in incubators, which provided a sterile environment. The incubators were set to 37°C, 5% CO_2, and 95% RH for optimal pH levels and growth conditions. To pre-heat media and other reagents water baths filled with Lab Armor Beads at 37°C were used. A media change of cultivated cells is done to replenish nutrients and keep the correct pH level. Generally, the media was changed every 2-3 day for cells in cultivation.

Table B.1: List of reagents and the optimal volumes used for culturing and maintaining mammalian cells in different flask formats, varying from 25-225 cm^2.

Flask size (cm^2)	PBS volume (mL)	Accutase volume (mL)	Media volume (mL)	Culturing volume (mL)
T-25	5	0.5	4.5	5
T-75	10	1	9	15
T-175	15	3	12	40
T-225	20	4	16	50

B.1.1 Thawing of Cells

Medium, specific for the cell type, supplemented with 10% FBS and a 50 mL conical tube with Milli-Q® water were pre-heated at 37°C. Next, the cryo-vial with cells were taken from the -150°C freezer and was instantly placed in the conical tube with warmed Milli-Q® water for about 1 minute until about 80% thawed. The vial was wiped with 70% ethanol and brought into the LAF-bench. 1 mL of cells were transferred to 9 mL of warm medium in a conical tube. The cryo-vial was rinsed with medium to ensure all cells were transferred. Next, the cell density was measured using the ChemoMetec NucleoCounter® NC-3000™. The conical tube was centrifuged at 250 × g for 5 min, to remove any dimethyl sulfoxide (DMSO). The supernatant was discarded, and the pellet was resuspended in the appropriate amount of medium according to the flask type and seeding density.

B.1.2 Dissociation and Passaging of Cells

When the cells have reached a confluency of approximately 70%, the cells should still be in the log phase of growth and need to be passaged to maintain health and propagation. To detach the cells from the adherence of the flask enzymatic methods can be utilized. First, the media was aspirated, and the cells were washed with PBS pH 7.4 (Gibco™, cat#10010015) to remove any access FBS. The PBS was aspirated and StemPro™ Accutase™ Cell Dissociation Reagent (Gibco™, cat#10565018) was added. The flask was rolled to ensure all cells were covered, and then incubated at 37°C, under 5% CO_2, and 95% RH for 5-10 min. A microscope was used to visualize if all cells had properly detached. The flask was gently tapped on a solid surface to further detach the cells If needed. The dissociation reagent was neutralized by adding media containing FBS. After cell dissociation the cells were transferred to a conical tube. An aliquot was taken to measure the cell density using the ChemoMetec NucleoCounter® NC-3000™. Based on the measured number of cells, the volume of cells required to get the appropriate seeding density for the flask size was calculated. A new flask was seeded with the number of cells and fresh media. The passage number was written on the flask to keep track of the numbers of sub-cultures the cells have gone through.

B.1.3 Cryopreservation of Cells

Freezing media was prepared according to the following ratio: 8% DMSO (Sigma-Aldrich, cat#D2438), 10% FBS and 82% media (supplemented with 10% FBS) and was placed in the fridge. Labels containing the name of the cell line, number of cells in each vial, passage number and the date of cryopreservation were also prepared. The cells were detached according to the cell dissociation protocol B.1.2. Next, the cell density was determined using the ChemoMetec NucleoCounter® NC-3000™. The volume of cells to be cryopreserved was calculated and added to a conical tube. Next, the conical tube was centrifuged at 250 × g for 5 min to harvest the cells. The supernatant was aspirated, and the pellet was resuspended in the calculated volume of chilled freezing media. 1 mL of cells were aliquoted in each Nunc™ Biobanking and Cell Culture Cryogenic Tube (Thermo Scientific™, cat#374081). Next, the cryovials were stored in a freezing container, ensuring a controlled freezing rate of approximately -1°C/min, and placed in the -80°C freezer overnight. The next day, the cryovials were transferred to the liquid nitrogen -150°C freezer, for long-time storage.

B.2 Plasmid Preparation

B.2.1 Bacterial Transformation and Plasmid Expansion

One Shot™ TOP10 Chemically Competent *E. coli* (Invitrogen™, cat#C404003) was transferred from the -80°C freezer to ice for 15 min until thawed. 50-100 ng of plasmid DNA was added directly to the cells, and they were incubated for 30 min on ice. Next, the cells were heat chocked for 30 sec at 42°C. The cells were transferred back on ice for a 2 min incubation. 200 µL of LB media was added to the transformed cells and the starter culture was placed in a shaking heating bock at 37°C and 225 rpm for 60 min. After incubation, the starter culture was added to a pre-heated LB agar plate containing carbenicillin. The cells suspension was spread using ColiRollers™ Plating Beads (Novagen®, cat#71013). The plate was incubated over night at 37°C, facing downwards to avoid desiccation of the agar. Next morning, a single colony was picked and inoculated in 5 mL LB media supplemented with 1,000X carbenicillin (stock 50 mg/mL) to make a mini culture. After 8 hours, 1 mL of mini culture and LB media supplemented with 1,000X carbenicillin was used to prepare a 130 mL culture. The flask was placed in a shaking incubator overnight at 37°C and 225 rpm. The next day, the plasmid DNA had been expanded and was ready for isolation.

B.2.2 Plasmid Purification

B.2.2.1 AmMag™ Quatro System

Some of the plasmids were purified using the automated AmMag™ Quatro System (GenScript, cat#D00019) and the AmMag™ Quatro Plasmid Purification Kit Maxi 24 prep (GenScript, cat#L00943-24). First, the GeneScript Quatro tube was pre-weighed. The overnight *E. coli* culture was transferred to the GeneScript Quatro tube and was centrifuged at 4000 × g for 15-20 min. The supernatant was discarded, and the tube was weighed again to get the weight of the pellet, that should be within the range of 1-1.5 g. 1 mL of Milli-Q® was added to the tube, and it was vortexed to fully resuspend the bacterial pellet. The tube was then loaded into the automated AmMag™ Quatro system together with the other reagents, following the vendors manual [70]. When the program was finished, the collecting tube was centrifuged at 4000 × g for 20 min. The column inside of the collecting tubes was discarded and the purified plasmid was transferred to a 1.5 mL microcentrifuge tube and stored in the fridge prior to measuring the plasmid concentration.

B.2.2.2 QIAGEN® Plasmid Plus Maxi Kit

The QIAGEN® Plasmid Plus Maxi Kit (QIAGEN, cat#12963) was used for purification of the plasmids. The purification was conducted according to the protocol QIAGEN® Quick-Start Protocol [71]. First, the cells were harvested by centrifugation for 15 min at 6000 × g and 4 °C. Subsequently, the bacterial pellet was resuspended in 5 mL of Buffer P1. 5 mL of Buffer P2 was

added and mixed by vigorously inverting the tubes 4-6 times. The tubes were incubated at room temperature for 3 min. 5 mL of Buffer S3 was added to lysate the cells and the tubes was mixed immediately by vigorously inverting the tubes 4-6 times. Next, the lysate was transferred to the QIAfilter cartridge, and was incubated at room temperature for 10 min. During incubation, the QIAGEN Plasmid Plus spin columns were placed into the QIAvac 24 Plus (QIAGEN, cat#19413). Then the tube extenders were inserted into each of the columns. The plunger was gently inserted into the QIAfilter cartridge, and the cell lysate was filtered into a new tube, that allowed for the addition of Buffer BB. 5 mL of Buffer BB was added to the tube to clear the lysate and the tube was mixed by vigorously inverting it 4-6 times. Next, the lysate was transferred to a QIAGEN Plasmid Plus spin column with a tube extender attached on the QIAvac 24 Plus. Approximately 300 mbar of vacuum was applied until the liquid had been drawn through the columns. The vacuum was then switched off. 700 µL of Buffer ETR was added to wash the DNA and vacuum was applied until all liquid had been drawn through. 700 µL of Buffer PE was then added to the columns and vacuum was applied until all liquid has been drawn through. The vacuum was then switched off. To completely remove the residual wash buffer, the QIAGEN Plasmid Plus Maxi spin column was placed into a 2 mL collection tube and was centrifuged at 10,000 × g for 1 min in a tabletop microcentrifuge. Lastly, to elute the DNA, the QIAGEN Plasmid spin column was placed into a clean 2 mL tube. 50-400 µL of Nuclease-Free Water (Ambion™, cat#AM9937) was added to the centre of the spin column, it was set to stand for >1 min, and was then centrifuged at 10,000 × g for 1 min.

B.2.2.3 Plasmid Quantification

The day after the plasmid purification, the plasmid concentration was quantified using the Implen NanoPhotometer® N50. First a blank sample was measured using 1 µL of Nuclease-Free Water (Ambion™, cat#AM9937). Next, the plasmid concentration was quantified by adding 1 µL of DNA onto the pedestal.

B.3 Geneticin™ Titration of LNCaP Cells

Mammalian cell lines vary in sensitivity of antibiotics. Therefore, it is always preferred to conduct a dose response experiment in which cells are subjected to increasing amounts of antibiotics, when working with new cell lines and different types of antibiotics. The dose response experiment is performed to determine the minimum concentration of antibiotics needed to kill off untransfected cells efficiently and generate stable cell lines. In this project an antibiotic killer curve was conducted for the LNCaP cells to determine the concentration of antibiotic for later selection of the transfected cells.

The cells were transfected with a vector containing a marker for neomycin resistance, so the antibiotic G418 was used in this experiment. LNCaP cells were harvested at ~70% confluency, according to the dissociation protocol B.1.2, a day prior to antibiotic addition. The cell density was determined using the ChemoMetec NucleoCounter® NC-3000™. Next, the cells were plated to 30,000 cells/cm^2 in a 24-well (1.9 cm^2/well) Corning® Costar® TC-Treated Multiple Well Plate (Corning®, cat#3524). The cells were diluted in RPMI 1640 medium (Gibco™, cat#21875034) supplemented with 10% FBS, to plate 0.5 mL of cell suspension per well. The plate was placed in the incubator at 37°C, under 5% CO_2, and 95% RH. The next day, RPMI 1640 medium supplemented with 10% FBS, containing increasing amounts of Geneticin™ Selective Antibiotic (G418 Sulfate) (50 mg/mL) (Gibco™, cat#10131027), was prepared in 1.5 mL microcentrifuge tubes according to Table B.3.

Table B.3: Presented is the dilution of the G418 stock-solution (50 mg/mL) with RPMI 1640 medium. From this 1,000 µg/mL solution, further dilutions were made to generate different concentrations of G418, ranging from 50-1,000 µg/mL.

G418 Stock (50 mg/mL) (µL)	**Media (µL)**	**Final Volume (µL)**	**Final G418 concentration (µg/mL)**
200	9,800	10,000	1,000
G418 (1 mg/mL) (µL)	**Media (µL)**	**Final Volume (µL)**	**Final G418 concentration (µg/mL)**
1,200	0	1,200	1,000
1,080	120	1,200	900
960	240	1,200	800
840	360	1,200	700
720	480	1,200	600
600	600	1,200	500
480	720	1,200	400
360	840	1,200	300
240	960	1,200	200
120	1,080	1,200	100
60	1,140	1,200	50
0	1,200	1,200	0

The media from the wells were aspirated and 0.5 mL of fresh RPMI 1640 media containing increasing amounts of G418 was added. The range of concentrations were plated in duplicates and a non-antibiotic control containing only RPMI 1640 media and cells was also plated. The 24-well plate was placed in the Sartorius Incucyte® S3 instrument for live-cell monitoring. After 3 days the media was replaced with new fresh media with increasing concentrations of antibiotic with accordance to Table B.2. After a few more days, approximately a week after antibiotics treatment, the antibiotic kill-curve was terminated using the Incucyte® S3 Software (v2022B).

B.4 Transfection Efficiency Evaluation of LNCaP Cells

The transfected LNCaP cells was not successfully selected after G418 treatment, when using the same transfection protocol as for the HEK-293 cells. Once repeated, using lower amounts of plasmid DNA, the selection failed once again. Therefore, some experiments were conducted to determine the transfection efficiency. A GFP plasmid was used at varying concentrations and electroporations were performed using the MaxCyte®. In parallel, another electroporation method, the Neon™, was also tested in order to generate stable LNCaP-dCas9-LgBiT-N/C-Terminus cell lines. The GFP plasmid used in these experiments were kindly gifted by Susanna Engberg.

B.4.1 MaxCyte® Electroporation

The MaxCyte® transfections with the GFP plasmids were conducted following the protocol described in section 3.1.3.1 of the thesis, with some minor adjustments. In this experiment, smaller electroporation cuvettes, OC-100, were used. The cells were therefore resuspended in 50 µL of MaxCyte® buffer. Furthermore, different concentration of GFP plasmid (1.6 µg/mL) was used for the transfections: 20, 100 and 200 µg/mL of cell suspension. An untransfected control was also included in the analysis.

After the 15 min incubation of the transfected cells and dilution in media, 0.5 mL cell suspension were plated in each well of a 24-well plate according to Table B.4.1. The plate was then placed in the IncuCyte for live-cell analysis and the fluorescence was measured every sixth hour for 72 hours.

Table B.4.1: Plating of the cells in a 24-well plate to evaluate the transfection efficiency of the LNCaP cells using the MaxCyte®. Electroporator

	1	**2**	**3**	**4**	**5**	**6**
A	20 µg/mL cells	20 µg/mL cells	20 µg/mL cells	20 µg/mL cells	20 µg/mL cells	20 µg/mL cells
B	100 µg/mL cells	100 µg/mL cells	100 µg/mL cells	100 µg/mL cells	100 µg/mL cells	100 µg/mL cells
C	200 µg/mL cells	200 µg/mL cells	200 µg/mL cells	200 µg/mL cells	200 µg/mL cells	200 µg/mL cells
D	Untransfected control	Untransfected control	Untransfected control	Untransfected control	Untransfected control	Untransfected control

B.4.2 Neon™ Electroporation

Prior to transfection, the LNCaP cells were plated to reach ~70% confluency on the day of transfection. The cells were detached according to the dissociation protocol, diluted in RPMI 1640 Medium + 10% FBS (Gibco™, cat#21875034), and the cell density was measured. The required volume of cell suspension was calculated and transferred to a conical tube for centrifugation at 250 × g for 5 min, to harvest the cells. 400,000 cells were prepared for each transfection. The supernatant was discarded, and the cells were washed with PBS (Gibco™, cat#10010015) and once again centrifuged at 250 × g for 5 min. A Nunc™ Poly-D-Lysine Coated 24-well plate (Thermo Scientific™, cat#152025) was prepared by adding 490 µL of media to each well, and it was placed in the incubator. The Neon™ Transfection System 10 µL Kit (Invitrogen™, cat#MPK1025) was used to conduct the electroporations. The pellets were resuspended in 80 µL of buffer R. Next, DNA, Buffer R, and cell suspension was prepared in tubes according to Table B.4.2.1 below. The tubes were prepared with reagents for four electroporations: Three reactions to plate triplicates, including one extra reaction to account for dead volumes.

Table B.4.2.1: Calculated volumes of reagents for four reactions using the Neon™ electroporator.

Cells	DNA (µL)	Buffer R (µL)	Cell suspension (µL)	Total volume (µL)
Untransfected Control	0	20	20	40
dCas9-LgBiT-N-Term	5.4	14.6	20	40
dCas9-LgBiT-C-Term	4.2	15.8	20	40
GFP Control	1.2	18.8	20	40

Next, 3 mL of electrolytic buffer E was prepared in a tube and placed in the electroporator. The 10 µL pipette tips were used to electroporate the controls and samples using the following protocol: Pulse voltage: 950V, pulse width: 30 ms and pulse number: 2 pulses. After successful electroporation the cells were immediately seeded into the corresponding wells of the Poly-D-Lysine plate according to Table B.4.2.2.

Table B.4.2.2: Plating of the cells in a 24-well plate to evaluate the transfection efficiency of the LNCaP cells using the Neon™ electroporator.

	1	**2**	**3**	**4**	**5**	**6**
A				dCas9-LgBiT-N-Term	dCas9-LgBiT-N-Term	dCas9-LgBiT-N-Term
B				Untransfected Control	Untransfected Control	Untransfected Control
C	GFP Control	GFP Control	GFP Control			
D	dCas9-LgBiT-C-Term	dCas9-LgBiT-C-Term	dCas9-LgBiT-C-Term			

C

Safety-, Ethical- and Environmental Disclosures

In this appendix, the safety-, ethical- and environmental aspects of this master's thesis project are addressed. AstraZeneca have identified internal policies, standards and procedures for safety, health, and environment, called SHE, which have been used as guidelines throughout this thesis project.

C.1 Safety Aspects

Safety routines at AstraZeneca are updated with accordance to the Swedish legislation, EU directives and regulations from the Swedish Work Environment Authority. All researchers and other personal working in the labs at AstraZeneca must undergo mandatory trainings prior to working in the labs with biosafety levels 1 and 2. Personal are also obligated to undergo additional individual courses depending on the laboratory work that will be performed.

In all labs where experimental work was performed throughout this project, PPE were obligatory, which protects all personal from the potentially hazardous agents. The safety equipment of the laboratory work included a combination of primary barriers such as LAF benches, PPE, and enclosed containers for potentially contagious waste. The liquid biological waste was disposed by using the vacuum system in the LAF-benches. This vacuum system in the LAF benches is connected to a liquid waste bin. When assembling new liquid waste bin, 20 mL of Pursept® (Schülke) is added. Additionally, 20 mL of Pursept® is added prior to disposing of the liquid waste, to eliminate any biological materials. All biological waste from the cell labs is disposed of as risk class 2, and all solid waste that have been in contact with cells must be enclosed inside a LAF-bench prior to disposing of it in the biological waste bins. Furthermore, all wastebins are labelled as contagious material.

A general risk assessment for the laboratory work in this project was outlined. The risk assessment covers handling of hazardous chemicals and biological agents, and handling of laboratory equipment and instruments. In Table C.1, potentially hazardous chemical agents used in the experimental work throughout this project are listed. Furthermore, risk assessments were outlined in this project for the established genetically engineered cell lines.

Table C.1: List of potentially hazardous chemical agents used in the laboratory work throughout this project. The table lists the reagents followed by risk declaration, reference, and additional safety precautions.

Reagent	Risk Declaration	Reference	Additional Safety Precautions
Pursept®	Pursept® can cause irritation to the skin, eyes, and respiratory system upon contact or inhalation.	SDS Sigma-Aldrich	Wear PPIs, including lab coat, gloves, and safety goggles. Avoid inhalation and contact with skin and eyes. Work with Pursept® in a well-ventilated area or under a fume hood to minimize exposure to vapours.
DMSO	DMSO can cause irritation to the skin, eyes, and respiratory system upon contact or inhalation.	SDS Sigma-Aldrich	Wear PPIs, including lab coat, gloves, and safety goggles. Avoid inhalation and contact with skin and eyes. Work with DMSO in a well-ventilated area or under a fume hood to minimize exposure to vapours.

C.2 Ethical Aspects

The immortalized cell lines used in this project are commonly used and no ethical approval was required for working with them. The cell lines were genetically engineered; however, this did not add any additional aspects regarding ethical approval.

C.3 Environmental Aspects

In this project, single-use plastic was frequently used in the laboratory work and was one of the major contributors affecting the environment. This included the usage of *e.g.*, plastic tubes, pipette tips, and culturing flasks/plates. It is not possible to eliminate this usage of plastics in the laboratory work. However, careful planning of experiments in advance helps to minimize the single-use plastics. Further environmental aspects are the high energy- and water consumptions associated with the experimental work. One way to decrease this consumption is to, if possible, turn off instruments when they are not in use. Another important environmental aspect is the waste handling and disposal of material that have been in contact with hazardous biological and chemical agents, to ensure these are not released into the environment.

D

Supplementary Results

In this appendix, supplementary data of the thesis project is presented, providing additional results, which were not included in the main text.

D.1 Plasmid Concentrations and OD Ratios

High-quality DNA is characterized as having an $OD_{260/280}$ ratio between 1.8-2.2, and an $OD_{260/230}$ ratio of 1.8–2.2. In Table D1 below, the measured value of plasmid concentrations and different OD ratios of the plasmids generated in this project are presented.

Table D.1: The measured plasmid concentrations (ng/µL), and $OD_{260/280}$ and $OD_{260/230}$ ratios of the different plasmids.

Plasmid	Concentration (ng/µL)	$OD_{260/280}$	$OD_{260/230}$
dCas9-LgBiT-N-Term	369.2	1.868	1.719*
dCas9-LgBiT-C-Term	470.8	1.900	1.786*
STAT3-SmBiT-C-Term	1193	1.947	2.363*
AR-SmBiT-N-Term	675.3	1.947	2.411*

All the $OD_{260/230}$ ratios were outside the range to be considered as high-quality DNA. An $OD_{260/230}$ below 1.8 could suggests contamination with organic compounds such as phenol, carbohydrates, or salts. Furthermore, a ratio above 2.2 could indicate the presence of RNA contamination or other substances like guanidine thiocyanate. Another potential reason for the ratios falling out of the range could be residual contaminants on the surface of the pedestal, leading to inaccurate measurements.

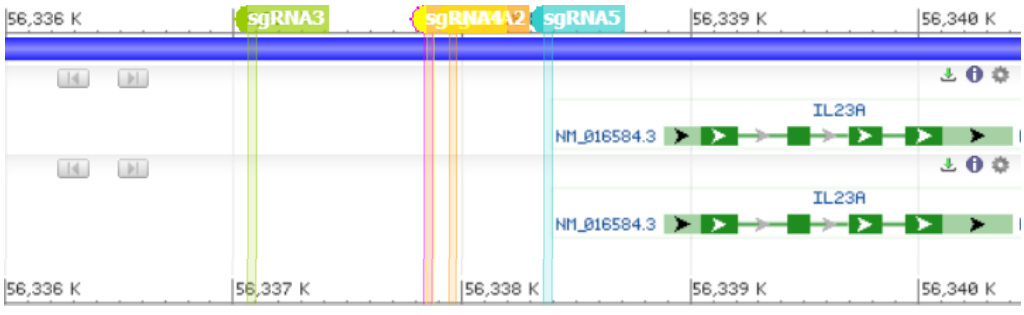
D.2 sgRNA Design

After identifying the downstream target genes and conducting the bioinformatical analysis in R, the following sgRNAs presented in Table D.2 were chosen for this project. The two identified downstream target genes of STAT3 were: IL-23A and NFKB1. For each downstream target gene, 5 sgRNAs were designed. The sgRNAs were selected based on the relative scores and adjusted p-values. These sgRNAs, together with a scrambled sgRNA were then ordered to be synthesized by Sigma-Aldrich.

Table D.2: Identified downstream target genes associated with STAT3, retrieved sgRNA sequence, followed by the PAM sequence.

Guide	Guide sequence	PAM
IL23A:1	5'-CACUCGAUGUUGCAUCAUGG-3'	5'-AGG-3'
IL23A:2	5'-UUCGACUACCCAGAACUCCU-3'	5'-GGG-3'
IL23A:3	5'-GGAGACGUUGUACACAAGUA-3'	5'-AGG-3'
IL23A:4	5'-ACUCGAUGUUGCAUCAUGGA-3'	5'-GGG-3'
IL23A:5	5'-GCCCGUUUAUUCCUAAUAGG-3'	5'-GGG-3'
NFKB1:1	5'-GUAUGCUCUCUCGACGUCAG-3'	5'-UGG-3'
NFKB1:2	5'-UGCGUUCCCCGACCAUUGAU-3'	5'-UGG-3'
NFKB1:3	5'-CCCUUAGGGGCUAUGGACCG-3'	5'-CAU-3'
NFKB1:4	5'-CCCGTTGTAAGCTGTTCAGT-3'	5'-CGG-3'
NFKB1:5	5'-UAUGCUCUCUCGACGUCAGU-3'	5'-GGG-3'
Scramble sgRNA	5'-GAUCCGCUGGGCGGUGUUCAGCU-3'	-

In Figure D.2 the localizations in the genome of the sgRNAs designed targeting the promotor region of the downstream gene IL-23A are presented. The sgRNAs are located within a 1.5 kb region, with three of the guides situated close in the genome. Two of the guides, sgRNA1 and sgRNA4, are also overlapping and only differs in one nucleotide.



Pink: sgRNA1, **Orange:** sgRNA2, **Green:** sgRNA3, **Yellow:** sgRNA4, **Blue:** sgRNA5

Figure D.2: An overview of the sgRNAs 1-5 localization in the genome. sgRNA 1 and 4 are overlapping. These guides were designed targeting the promoter region of the STAT3 downstream target gene IL-23A. NCBI Reference Sequence: NC_000012.12.

DEPARTMENT OF LIFE SCIENCES
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2024
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY