Understanding the dynamic regulation of SOCS3

Nikoletta Kalenderoglou

This dissertation is submitted for the degree of Doctor of Philosophy

November 2019

School of Biomedical and Life Sciences

"Ἓν οἶδα ὅτι οὐδὲν οἶδα"

Σωκράτης

"All I know is that I know nothing"

Socrates

ACKNOWLEDGEMENT

Special thanks go to Sarita, my long-lasting friend and made of honour, for help, support and always being there for a cigarette break during long lab incubation times.

To my friends that I left at home, Sevi, Christina, Georgia, Eirini and Amalia for always having time for coffee and endless PhD ranting. Thank you for being there, in spite of the distance. Your encouragement and friendship led me to where I am today. Yeia Mas!

Most of all, I am extremely grateful for my wife Jacqui. Thank you for your love and companionship, you are my rock. Your faith in my abilities gave me the confidence and determination to get me through the dark times. Your car rides to University and back, last minute cell movie analysis advice, bringing me coffee and my favourite snacks, I will be forever in your debt.

For last I reserve my deepest expression of gratitude to my parents and my sister. Especially to my mum and dad, who always encouraged me to pursue my dreams and never doubted my abilities.

It is for my family that I have completed this thesis.

Declaration

This thesis has not been submitted in support of an application for another degree at this or any other university. It is the result of my own work and includes nothing that is the outcome of work done in collaboration except where specifically indicated. Many of the ideas in this thesis were the product of discussion with Dr Richard Mort (chapter 4) and my supervisors Dr Karen Wright and Dr Rachael Rigby (chapters 1-3).

Abstract

In the past two decades, it has become evident that signal transduction pathways are more than two dimensional pathways consisted of proteins that are just activated or supressed in response to distinct cues. Instead, the dynamic nature of key proteins regulates the strength and quality of the signal. Several key signal transduction pathways are controlled by negative feedback loops that are highly dynamic and demonstrate oscillatory behaviours. Negative feedback regulation of the JAK/STAT pathway by Suppressors of Cytokine Signalling (SOCS) is an example of oscillatory signalling.

We sought to investigate the oscillatory capacity of the tumour suppressor protein SOCS3 and its role in important cellular functions using whole-cell population and single-cell analysis.

An important aspect of cell biology using experimental cell-population techniques is to produce a synchronized cell culture. Serum starvation and subsequent shock is able to capture the oscillatory behaviour of SOCS3 protein to some extent. However, the average response in whole-cell population systems demonstrated to be 'noisy' leading to establishment of a single-cell analysis system.

To investigate SOCS3 oscillation at the single cell level, we first attempted to generate cell clones stably expressing SOCS3 C-terminal GFPSpark fusion protein from its respective endogenous promoter to monitor its expression in real time with confocal microscopy. Despite careful optimization of each step of CRISPR/Cas9 strategy, the generation of GFPSpark knockin cell line was not successful.

Finally, we utilised the tandem fluorescent protein timer (tFT) strategy to investigate localisation and trafficking of SOCS3 protein and monitor its promoter activity in response to different stimuli. The use of tFT provided us the ability to analyse SOCS3 dynamics across spatial and temporal dimensions under either normal culture conditions or different treatments that are known to influence on SOCS3 half-life and degradation rates.

Contents

4 *IN VIVO* ANALYSIS OF SOCS3 PROTEIN DYNAMICS AND PROMOTER

List of Tables

List of Figures

List of Appendices

1 Literature Review

1.1 Inflammatory Bowel Disease (IBD)

Inflammatory Bowel Disease (IBD) comprises a group of idiopathic, chronic and recurrent inflammatory conditions of the gastrointestinal tract. Traditionally, the highest occurrence of IBD has been reported in norther Europe, United Kingdom and North America. Its incidence and prevalence has raised significantly in the last six decades in the Western industrialized countries and there is growing evidence that the incidence is steadily rising in Easter Europe and developing countries, including China and India, attributed largely to the rapid modernization and Westernization of the population [1, 2].

Patients diagnosed with IBD colitis typically present with gastrointestinal symptoms of abdominal pain, severe diarrhoea, rectal bleeding, fever and weight loss. Crohn's disease (CD) and Ulcerative Colitis (UC) include the two major clinic pathological phenotypes of IBD [3-5]. CD and UC are most frequently diagnosed in late adolescence or early adulthood, but the diagnosis may occur at any age. Mean and median ages at

the time of diagnosis for patients presented with CD are commonly 5-10 years earlier than those associated with UC [6-9].

A proper diagnosis between Crohn's disease (CD) and Ulcerative Colitis (UC) can be of pivotal importance in clinical practice, especially in the context of tailoring clinical therapy, as each disease classification often includes distinct therapeutic management and prognosis. Though, up to 10% of patients presenting with IBD colitis cannot be classified in either category and is thus termed IBD as yet unclassified (IBDU) [10].

Ulcerative Colitis is characterised by a diffuse mucosal inflammation that is limited to the colon: it extends proximately from the rectum, spreading in a continuous manner and frequently involves the periappendiceal region (figure 1.1). Histopathological features involve the presence of an extensive neutrophil infiltration limited to the mucosa and submucosa with the formation of scattered micro-abscesses and the depletion of goblet cell mucin is commonly seen. By contrast, Crohn's disease can cause transmural inflammation and it involves any site of the gastrointestinal tract with the terminal ileum and the perianal region is the most commonly affected by the disease (figure 1.1). The microscopic features of CD include aggregation of macrophages that frequently result in the formation of non-caseating granulomas, segmental and transmural inflammation, thickened submucosa and epithelial ulceration [11-13].
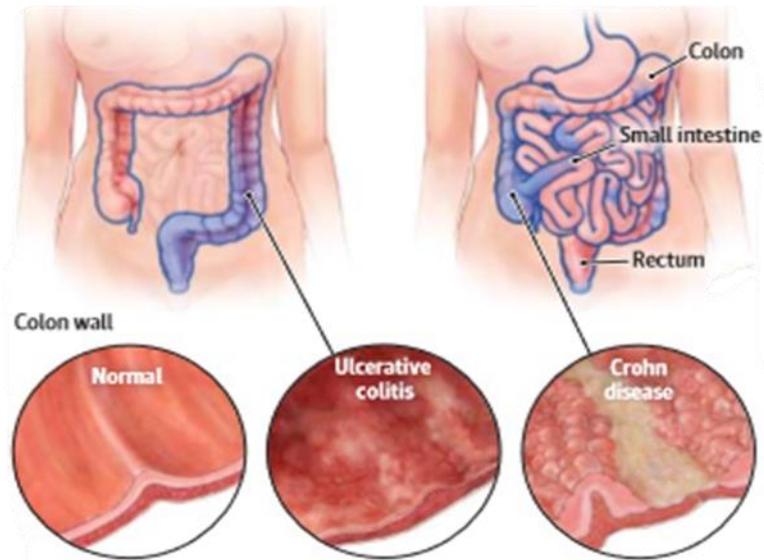
Figure 1.1: **Diagram of Inflammatory Bowel Disease (IBD).** Healthy colon wall (left), Ulcerative Colitis (middle) typically affects only the inner colon wall, Crohn's disease (right) affects any site of the GI tract with transmural inflammation. Adapted from [5, 14].

Currently, the exact aetiology of IBD is only partially understood but it is well documented that it is a multifactorial disorder. A variety of factors and triggers influence the pathogenesis of IBD (figure 1.2). These triggers involve a series of interactions between host genetics, dysregulated immune responses of the gut-associated lymphoid tissue (GALT), dysbiosis between the gut microbiota and the host intestinal cells as well as environmental factors that include a vast array of triggers such as smoking, diet, Western life style, sanitation, hygiene, antibiotics and other drugs [15-17].
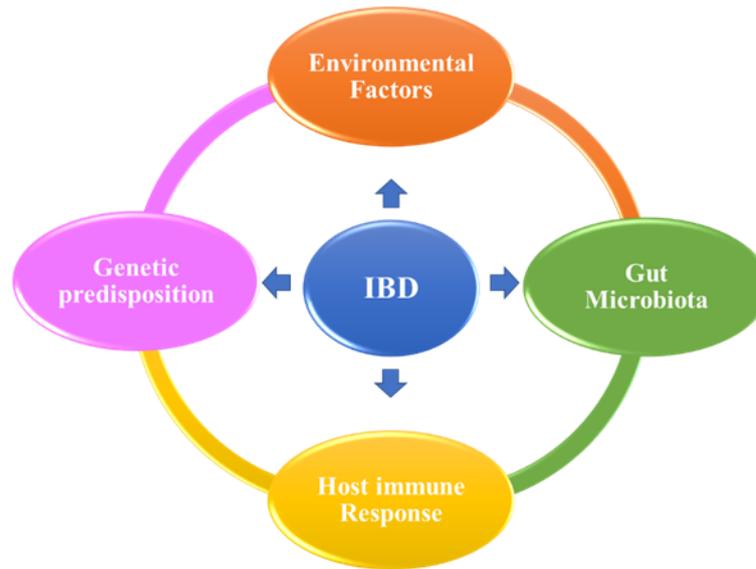
Figure 1.2: **The etiopathogenesis of IBD.** The interplay between host microbiota dysbiosis, genetic predisposition, host immune response and environmental triggers contribute to the pathogenesis of IBD.

Since the exact aetiology of IBD is currently unknown, treatment is currently symptomatic and targets general inflammatory mechanisms rather than specific pathophysiological changes. Treatment options of patients with IBD include intense and long-term therapy that frequently prolongs to lifelong treatment, which is often accompanied with side effects associated with high dose intake. Therefore, a better understanding of the pathogenesis of these complex diseases is needed to expand and improve treatment modalities for IBD [18-20].

Nowadays, IBD therapeutic regimens mainly involve anti-inflammatory compounds, such as corticosteroids, 5-aminosalicates (5-ASA), such as mesalazine, and immunosuppressant modulators, including methotrexate, thiopurines, tacrolimus and cyclosporine [21, 22]. Mesalazine, the active moiety of sulfasalazine, acts as an anti-inflammatory compound, and it is used as a first-line treatment of active UC of mild to moderate severity. It is demonstrated to be a highly effective treatment option with less

side effects compared to other therapeutic compounds [23]. Furthermore, anti-inflammatory anti-TNFα biological agents, such as infliximab, adalimumab, certolizumab and golimumab, have been primarily used as a treatment option to patients with moderate to severe CD and UC as well as to those not responding to steroids and/or immunomodulators [24, 25]. Only recently, biological agents such as vedolizumab, that have been found to supress the migration of leucocytes to the mucosa, have been added to the armamentarium [26]. Alternative therapeutic strategies, such as faecal microbiota transplantation (FMT) are also increasingly considered as treatment option for IBD [27]. Over the course of the past decade, genome-wide association studies (GWAS) have identified a large number of genes associated with the development of both CD and UC. Recent studies have indicated that the number of IBD susceptibility loci raised to 163 in European population [28]. Moreover, 110 of those IBD susceptibility genes are linked with both CD and UC, such as IL23R, IL12B, HLA, NKX2-3 and MST1), whereas 50 of which represent an equal effect for both phenotypes, 23 demonstrate risk effects that are UC-specific and the remaining 30 are considered as CD-only genes [28, 29].

Nevertheless, meta-analysis of Genome-wide association studies (GWAS) have implicated key pathways associated with intestinal homeostasis in IBD pathogenesis. Paramount among these are related to barrier function, epithelial restitution, microbial pathogen sensing, innate immune regulation, reactive oxygen species (ROS) generation, regulation of adaptive immunity, endoplasmic reticulum (ER) stress and autophagy [13].

Several of the loci identified in GWAS for CD and UC are known to encode and regulate cytokines and receptors that signal via the janus kinase (JAK)/signal transducer and

activator of transcription (STAT) pathway, e.g., IL-23R, IL-10, IFN-γ or IL-12B [30, 31]. Furthermore, multiple single nucleotide polymorphisms have been detected in several JAK-STAT pathway genes including Jak2, Tyk2, STAT1, STAT3 and STAT4, which have been described to entail an elevated risk for the progression of IBD [28, 32].

1.2 Colorectal Cancer (CRC)

Colorectal cancer (CRC), also known as bowel cancer, is the second leading cause of cancer-related deaths for both women and men. Despite advances in diagnosis and treatment, it is the fourth most common cancer in the UK, with nearly 42,000 new cases diagnosed in 2015 and an estimated 14,000 deaths in 2014 [33, 34]. Approximately, 70-80% of cases of CRC occur sporadic with the presence of somatic mutations [35] whereas the remaining 20-30% have an inherited component [36]. The five-year survival rate of early stage diagnosed CRC can reach as high as 90%, whereas this survival rate is dramatically decreased to 10% for CRC cases diagnosed at later stages. Hence, early diagnosis is of greatest significance [37].
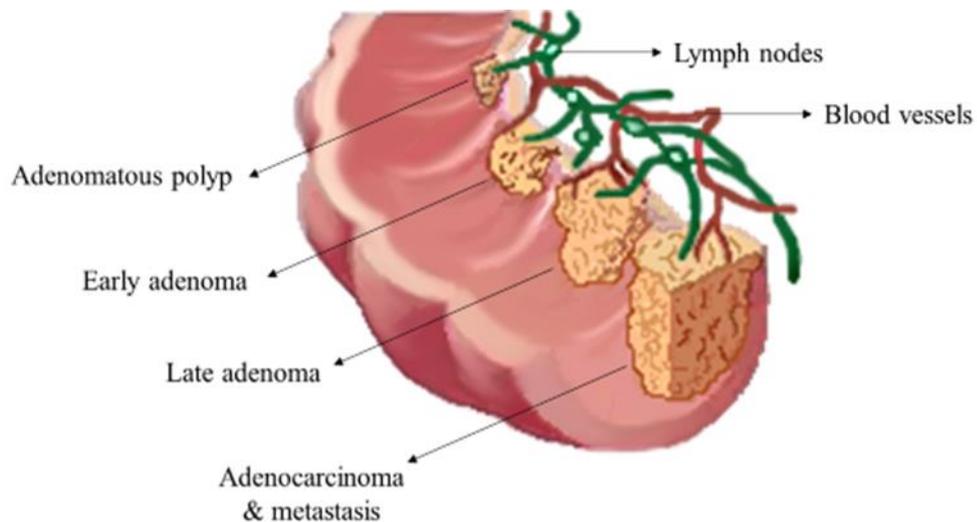
Figure 1.3: **The developmental phases of Colorectal adenocarcinoma.** Adapted from [38].

The progression of CRC from normal epithelial cells to adenocarcinoma and finally to invasive carcinoma typically follows several consecutive steps (figure 1.3). The clear majority of CRC cases initiate from an adenomatous polyp (adenoma). Adenomatous polyps are well-demarcated masses of epithelial dysplasia, with uncontrolled crypt cell proliferation. An adenoma can be considered malignant when neoplastic cells invade through the muscularis mucosae and infiltrate the submucosa [39, 40].

The development of sporadic CRC is mainly influenced by diet, lifestyle, environmental factors and host genetics. Moreover, these tumours are known to affect different molecular pathways and arise from the accumulation of several genetic and epigenetic alterations in a sequential order. Specifically, three distinct pathogenetic pathways are found to be involved in the evolution of this malignancy: (1) chromosomal instability (CIN), (2) microsatellite instability (MSI) and (3) CpG island methylator phenotype (CIMP) [40-42].

Interestingly, methylation of the SOCS3 promoter was found to play a crucial regulatory role in colonic SOCS3 expression. It was found that IL-6 signalling can induce an increase in methylation of SOCS3 by stimulating increased expression of DNA (cytosine-5-)-methyltransferase (DNMT1), which in turn induced signalling through STAT3. In line with this observation, overexpression of DMNT1 protein was detected in UC and its expression was strongly correlated in areas with high IL-6 signalling and those presented with cancer. The SOCS3 promoter STAT3-binding module has a vital role for stimulating SOCS3 expression in response to IL-6 signalling, and its methylation appears to demonstrate a rate-limiting step in CRC development [43].

Chromosomal instability is characterised as an elevated propensity to acquire chromosome aberration such as gene deletions, duplications and chromosomal

rearrangements. CIN occurs approximately 70%-85% of colon cancers and are characterised by the accumulation of genetic aberrations in distinct oncogenes and tumour suppressor genes including adenomatous polyposis coli (APC), KRAS, phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha (PIK3CA), B-Raf proto-oncogene, serine/threonine kinase (BRAF), SMAD4 and TP53, hence influencing pathways important for carcinogenesis [44, 45].

Furthermore, these specific pathogenetic pathways are thought in fact to play a vital role in the diversity of gene expression, phenotypic changes and the intra-heterogeneity within these CRC tumours. This heterogeneity serves a major clinical challenge for devising appropriate therapeutic treatments. Further understanding of the heterogeneity in CRC molecular pathways is needed to elucidate the diverse morphologic features of Colon malignancy and customize treatment strategies tailored to the need of an individual patient [37, 46].

1.3 IBD-associated colorectal cancer (IBD-CRC)

Inflammatory bowel disease (IBD)-related colorectal cancer (IBD-CRC) accounts for about 1% to 2% of all colorectal cancers and counts for approximately 10% to 15% of all deaths among IBD patients [47]. Furthermore, patients with IBD colitis are at 6 times higher the risk to develop colorectal cancer than the general population and this risk is related with the extend and duration of the disease, early age of onset and the severity of inflammation [48].

The development of the disease has been shown to occur through a multistep process of increasing grades of epithelial dysplasia, from no dysplasia progressing to indefinite dysplasia, low-grade dysplasia, high-grade dysplasia and finally to invasive adenocarcinoma. However, colorectal cancer can develop without progressing through

each of these steps [49]. Ulcerative colitis-associated colon cancer tends to affect the rectum and sigmoid colon, while colorectal cancer associated with CD is evenly distributed between the different colon segments [47, 50].

It is now clear that chronic inflammation plays a crucial role in inflammatory bowel disease (IBD) and inflammation-promoted colorectal cancer. Even though the molecular mechanisms of inflammation is being studied extensively, its role in the development of IBD-CRC is not fully understood yet [51].

Strong line of evidence suggests that when commensal enteric bacteria breach the intestinal epithelial barrier, they induce a chronic, immune-mediated inflammation, which is an indispensable participant in the neoplastic development of the overlying intestinal epithelium [52, 53].

At the cellular level, a variety of soluble mediators that are known to critically be involved in IBD pathogenesis have also been shown to mediate interactions between cancer cells and stromal cells in the tumour microenvironment of CAC. Elevated production of cytokines, chemokines, growth factors matrix-degrading enzymes as well as reactive oxygen and nitrogen species favours the emergence of tumorigenesis by establishing a microenvironment promoting intestinal epithelial cell proliferation, survival and migration [52-55]. Previous clinical reports and animal experiments have established a significant role of proinflammatory pathways, particularly nuclear factor kappa B (NF-κB), cyclooxygenase (COX-2)/PGE$_2$, IL-23/Th17 and IL-6/STAT3 signalling processes, in the pathogenesis of colitis-associated CRC. In addition, these signalling pathways control the expression of a vast array of inflammatory regulators and orchestrate a cancer-promoting microenvironment. The malignant-supporting effects of these processes include the upregulation of antiapoptotic molecules and the

increased proliferation of intestinal epithelial cells at the inflammatory location, which initiate cancer progression and growth in CRC [56]. Moreover, these signalling processes also promote the expression of several growth factors, including VEGF as well as chemokines such as IL-8 leading to angiogenesis [57].

Importantly, these physiological processes governing inflammation show a wide range of dynamic behaviours. In the past two decades, these signalling systems have attracted the attention by system biologists, who have studied their dynamic behaviour of these pathways using mathematical modelling. Specifically, in response to consistent cytokine stimuli, the NF-κB transcription factor complex exhibits a repeated oscillatory pattern between the nucleus and the cytoplasm with a period of ~100 min. The dynamic behaviour of NF-κB appears to regulate the cellular response to inflammation [58, 59].

Moreover, research studies analysing the dynamics JAK-STAT signalling cascade have proven beneficial to gain insights into the molecular basis of this complex signalling process, and its mechanisms in a variety of physiological conditions [60]. Negative feedback regulation of the JAK/STAT pathway by Suppressors of Cytokine Signalling (SOCS) also compromises an example of oscillatory signalling. Yoshiura et al. [61] identified serum-induced ultradian oscillations of STAT3 protein and its inhibitor SOCS3 in mouse fibroblast cells. It is yet unknown as to the functional consequences of perturbation of SOCS3 oscillation into a variety of physiological conditions, including inflammation.

Notably, autophagy machinery has also been recognized as a critical regulator of intestinal homeostasis and adaptation to environmental stress conditions, including nutrient or growth factor deprivation, pathogen invasion, endoplasmic reticulum (ER) stress, oxidative stress and hypoxia [62]. The first evidence of an existing crosstalk

between autophagy and IBD emerged from the identification of a gene variant (ATG16L1) in CD patients. ATG16L1 proteins are known to act as molecular scaffolds, regulating protein-protein interactions responsible for the LC3 lipidation during autophagosome formation upon classical and pathogen-related autophagy stimuli [63]. When the expression of ATG16L1 was reduced in hypomorphic mice, it exhibited prominent abnormalities in the granule exocytosis pathway. Notably, restoring the levels of ATG1L protein into the system reversed this reaction [64]. Recently, Luan et al. have shown that SOCS3 is regulated partly through an autophagy-based mechanism [65]. However, there is no research undertaken to investigate the role of autophagy in regulation of SOCS3 oscillatory signalling.

Conceptually, the connection between colitis and CAC has been well established both clinically and on a molecular basis. Concerning the latter, strong evidence from previous research has indicated that Il-6 plays a major role in the progress of inflammation-associated carcinogenesis, and its signalling essentially depends on STAT3 [66]. Hence, members of JAK/STAT pathway are also involved in neoplastic transformation in long-standing colitis [54].

1.4 Cytokine Signalling in Intestinal Epithelial Cells (IECs)

The Janus Kinase/signal transducers and activators of transcription (JAK/STAT) pathway represents an evolutionarily conserved signalling cascade that transduces signals from the cell membrane to the nucleus in response to a wide range of extracellular cytokine stimuli to orchestrate the expression of the appropriate set of genes. The activation of JAK/STAT pathway is utilized by a large number of cytokines, interferons and growth factors receptor. It critically regulates several cellular pathways important for haematopoiesis, immune response, and allelotaxy, including innate and

adaptive immune functions, development, cell proliferation, differentiation, survival, apoptosis and inflammation [67, 68]. Interaction of cytokines to their corresponding transmembrane receptors leads to a receptor complex formation that subsequently induces receptor dimerization of its subunits and association with JAK tyrosines (figure 1.4.A). In mammals, the JAK family includes four members: JAK1, JAK2, JAK3 and Tyk2 (non-receptor Protein Tyrosine Kinase-2). The association of the two intracellular kinase domains results in their phosphorylation, either via autophosphorylation and/or cross phosphorylation by other JAK members or other families of tyrosine kinases. This activation leads to an elevated JAK kinase activity. The phosphorylated tyrosine residues then serve as high-affinity binding sites that allow the binding of other Src homology 2 (SH2) domain-containing signalling molecules such as STATs, Src kinases, protein phosphatases, and other adaptor signalling proteins such as Shc, Grb2, and PI-3 kinase [69, 70]. STAT proteins are latent transcription factors located in the cytoplasm in an inactive form until activated via a JAKs mediated phosphorylation at a conserved single tyrosine residue at the carboxyl terminus. This phosphotyrosine allows the dimerization of STATs via reciprocal interaction with their conserved SH2 domain [71]. Following dimerization, phosphorylated STATs translocate to the nucleus by a mechanism that is dependent on importin α-5 (also known as nucleoprotein interactor 1) and the Ran nuclear import process [72, 73]. In the nucleus, dimerized STATs interact with specific regulatory sequences to induce or inhibit transcription of target genes. Therefore, the JAK/STAT pathway offers a direct machinery to translate an extracellular signal into transcriptional response [74].
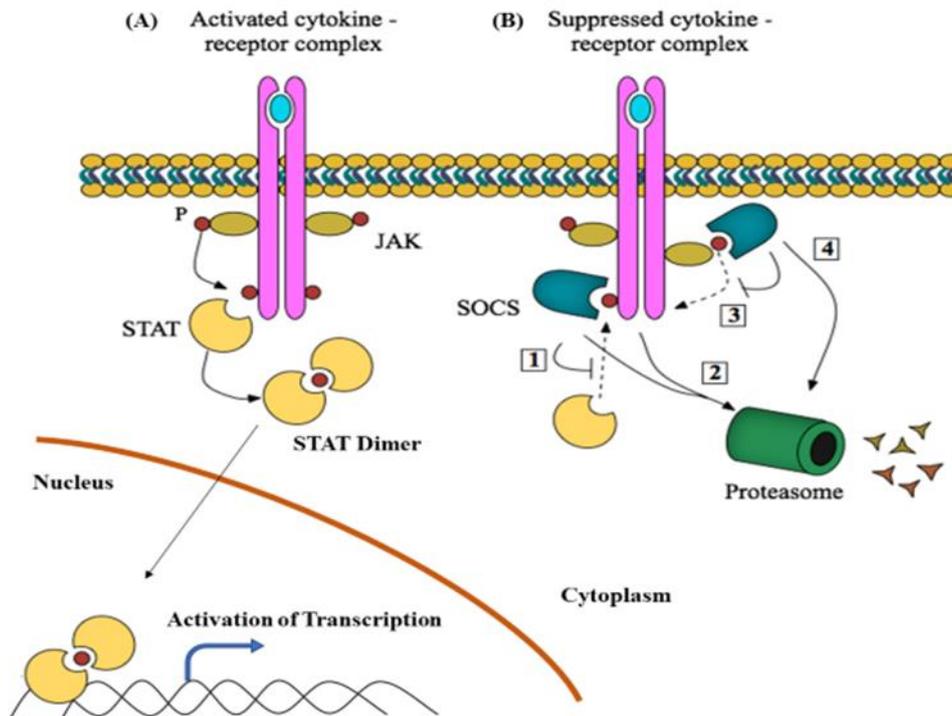
Figure 1.4: Principles of JAK-STAT signalling pathway. (A) Upon interaction of a cytokine to its receptor, conformational changes cause JAK auto or transphosphorylation and consequently phosphorylation of the receptor by JAKs. This leads to the formation of docking sites for STATs, which in turn they become phosphorylated by JAKs, they dimerize and translocate to the nucleus, where they ac transcription factor. **(B)** The four mechanisms by which SOCS inhibit cytokine signalling: (1) inhibiting STAT proteins to bind to the cytokine receptor; (2) targeting the receptor proteasomal degradation; (3) interacting with JAKs and inhibiting their kinase activity; (4) targeting JAKs for proteasomal degradation. Adapted from [75-77].

Seven STATs with different functions have been identified so far in mammalian cells, STAT1 to 6, including STAT5a and STAT5b, with over 40 different polypeptides to be known to induce one or more STATs [78]. STAT3 was initially identified as a DNA-binding protein induced by epidermal growth factor (EGF) and interleukin-6 (Il-6), capable of binding with an enhancer module in the promoters of acute phase genes [79]. Further research articles showed that STAT3 is induced in response to a vast array of

cytokines and growth factors, including interferon (IFN) and leptin. In healthy cells, the induction of STAT3 is transient like other STAT proteins, however, in a great number of primary tumours and cancer-derived cell lines, it remains constitutively induced which may be triggered by impairment of negative regulation or mutation of STAT3 itself [80, 81]. In IBD, induction of STAT3 that is controlled by the interaction of IL-6/sIL6R complexes and gp130 increases the expression of antiapoptotic proteins BCL-2 and BCL-X$_1$, which are proven to promote apoptotic resistance in CD4$^+$ cells and enhance the progression of chronic intestinal inflammation [82]. Additional evidence demonstrates that persistent activation of STAT3 may lead to cellular transformation, contributes to tumour-induced angiogenesis, inhibits anti-tumour immune responses and subsequently further enhances cancer progression [83, 84]. Thus, regulation of STAT3 expression is emerging as a novel target for cancer treatment [83-85].

The JAK/STAT pathway is one of the several regulatory signalling pathways that play an important role in the maintenance of intestinal homeostasis. Recent studies demonstrated that activation of STAT3 protects against the development of colitis and promotes intestinal repair following mucosal injury [86-88]. Research articles using mouse models of colitis-associated cancer treated with AOM and DSS demonstrated that IL-6 was mainly expressed in infiltrating macrophages, T cells and dendritic cells during carcinogenesis [89, 90]. Interestingly, deletion of IL-6 decreased tumour numbers, tumour size and tumour multiplicity. Furthermore, deletion of IL-6 in mice resulted in a more sever DSS-induced colitis, increased cell death and reduced IEC proliferation [89]. Similarly, mice with specific STAT3-deficiency were more prone to develop chemical-induced experimental colitis causing epithelial erosions and mucosal inflammation [91].

Collectively, it is demonstrated that IL-6 plays an important role as a tumour promoter in the progression of colitis-associated cancer and STAT3 acts as a crucial modulator for the transduction of malignant-promoting signals from IL-6.

1.5 Suppressor of cytokine signalling (SOCS)

Key regulators of the JAKs and STATs include members of Suppressor of cytokine signalling (SOCS) family proteins [92]. Suppressor of cytokine signalling (SOCS) proteins and cytokine inducible SRC homology 2 (SH2)-domain-containing proteins (CISs or CISH) belong to a family of intracellular proteins, several of which have profound roles in modulating the responses of immune cells to cytokines [77, 93].

The CIS/SOCS proteins constitute a family of eight members (CIS, SOCS1, SOCS2, SOCS3, SOCS4, SOCS5, SOCS6 and SOCS7) and all share a common modular organization of a central SH2 domain, an N-terminal domain of variable length and divergent sequence, and a carboxyl-terminal 40-amino acid module referred to as the SOCS-box (figure 1.5) [77, 93, 94].
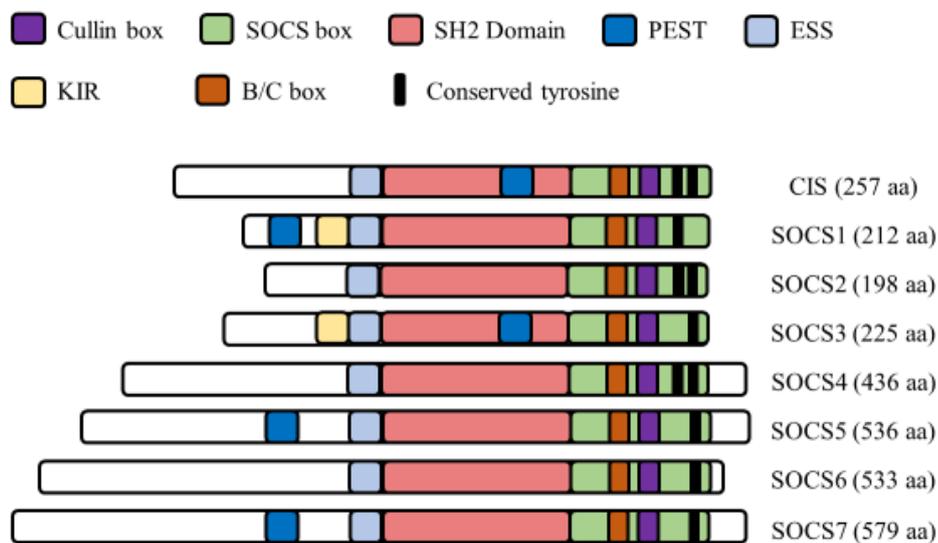
Figure 1.5: **Schematic representation of the structural domains of the SOCS proteins.**

The SOCS proteins are characterised by a variable NH2-terminal region including an extended SH2 subdomain (ESS), an SH2 domain and a SOCS box. Only SOCS1 and SOCS3 contain a KIR domain, whereas SOCS4-7 comprise a C-terminal extension. Adapted from [95].

Once SOCS3 is recruited to an induced cytokine receptor, such as gp130 (the signal transducing module of the interleukin-6 (IL-6) signalling process), SOCS3 develops an interaction motif upon phosphorylation of specific Tyr residues by cytokine-induced JAKs. Furthermore, SOCS3 supress signalling from gp130 by binding via their central SH2 domain to PTyr759. This allows it to interact with and block adjacent receptor-bound JAKs via its Kinase inhibitory region (KIR), located immediately upstream of the central SH2 domain, thereby inhibiting the recruitment and tyrosine phosphorylation of STATs [96]. Only the two most potent members of the SOCS family, SOCS1 and SOCS3, comprise KIR motif and is their dominant mode-of-action *in vivo* [97-99]. In addition to their ability to downregulate signalling through inhibiting the recruitment of pSTATs to JAKs, SOCS3 also suppress signalling via a second mechanism. This mechanism utilizes the action of the SOCS box at their C terminus. The SOCS box is responsible for binding with a protein complex consisting elongin B, elongin C, cullin-5, RING-box-2 (RBX2) and E2 ubiquitin transferase (figure 1.6).
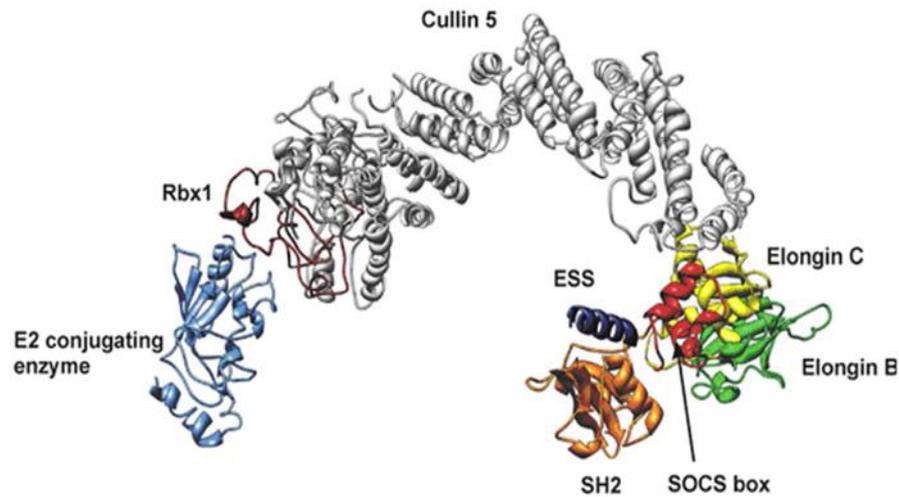
Figure 1.6: **Homology modelling of SOCS3 E3 complex.** SOCS3 protein interacts with a substrate via its SH2 domain. The SOCS box binds with Elongin C and B. The E3 ligase complex formation completes when Cullin5 and Rbx1 is also recruited. RBx1 interacts with E2 ubiquitin conjugating enzyme, permitting ubiquitin conjugation to the substrate. Image from [95].

In turn, the resulting complex acts as a putatitative E3 ubiquitin ligase that facilitates the degradation of proteins that they associate with their amino-terminal regions. This allows the SOCS proteins to function as an adaptor to trigger ubiquitination and proteasomal degradation of proteins involved in the entire cytokine-receptor complex, including Janus kinase (JAK) proteins and SOCS protein themselves (figure 1.4) [94, 100, 101].

While stimulation of the SOCS3 transcription and translation by a vast array of cytokines and growth factors has been well established, much less is known about post-translation regulation of SOCS3 turnover. What is known is that SOCS3 can be rapidly polyubiquitinated and subsequently targeted for degradation by the proteasome, explaining its short biological half-life. Sazaki et al. [102] investigated SOCS3 turnover in Ba/F3 pro-B cells and demonstrated that SOCS3-Lys6 residue may play an important

role for SOCS3 protein ubiquitination and degradation. However, it has since been established that the PEST sequence within the SH2 domain is also a crucial mediator of SOCS3 stability and functions independently from the proteasome [103].

CIS, SOCS1, SOCS2 and SOCS3 are generally expressed at low levels in unstimulated cells but mRNA and protein levels are rapidly upregulated in response to stimulation by a great variety of cytokines, growth factors and hormones [77].

SOCS3 expression was identified in a wide array of murine and human tissues. In murine, SOCS3 expression was found in the spleen, thymus, and lung, whereas in human tissues, SOCS3 was widely expressed in the lung, adipose tissue, ovary and aorta, and in a lower degree of expression in the colon, spleen, bladder, peripheral blood leukocytes, trachea and placenta [104, 105].

Gene knockouts studies in mice have demonstrated essential roles of SOCS3 protein. For instance, mice lacking SOCS3 displayed fatal placental defects during embryonic development, and although they did not show any anatomical abnormalities, they did not survive past 13 days of gestation. Furthermore, these embryos exhibited extended numbers of giant trophoblast cells in the placenta, and defects in the spongiotrophoblast and labyrinth placental layers [106].

Nevertheless, despite improvements in our molecular understanding of how SOCS3 interacts with cytokine receptors and JAKs, the extent to which other cellular mechanisms control SOCS3 function is uncertain. Recently, it has been shown that CUE domain-containing 2 (CUEDC2) is a novel SOCS3-interacting protein. Zhang et al. demonstrated that CUEDC2 functions collaboratively with SOCS3 to inhibit JAK1/STAT3 signalling by elevating SOCS3 stability via enhancing its interaction with Elongin C [68]. Furthermore, Williams et al. identified a novel functionally significant

mechanism linking SOCS3-controled inhibition of cytokine signalling to localisation at the plasma membrane via interaction with and stabilisation of cavin-1, a component of caveolae [5]. These recent findings increase the possibility that additional mechanisms may be needed to maximise the ability of SOCS3 to maintain cytokine signalling. The vast array of effects of SOCS3 in different cell lines and experimental models lay the grounds to call for thorough investigations to clarify its key mechanisms and targets.

1.6 SOCS3 in IBD and IBD-CRC

Members of the SOCS family have a relatively short biological half-life (SOCS3 half-life ranges from 40-120 minutes depending on the cell type), consequently their negative regulatory effects are mostly transient [107]. Nevertheless, persistent stimulation of STAT signalling pathway induced by chronic inflammation can promote constitutive activation of SOCS expression, leading in turn to persistent silencing of important signalling pathways and pre-disposition to progression of organ-specific illnesses [108].

Recently, several groups extensively investigate the role of SOCS3 protein in intestinal homeostasis and inflammation and sought to understand how dysregulation of colon homeostasis leads to transformed cutaneous immune responses during inflammatory pathogenesis in the intestine. Indeed, SOCS3 expression was found to be elevated in lamina propria and epithelia cells in the colon of mice with IBD as well as in patients suffering from ulcerative colitis and Crohn's disease, again demonstrated to be due to increase IL-6 signalling via STAT3 [109].

Moreover, intense immunostaining of SOCS3 was detected in the majority of leucocytes, particularly neutrophils, macrophages and lymphocytes in IBD mucosa [110]. In addition to IL-6, both LPS and TNF-α have been demonstrated to increase the

levels of SOCS3 protein in macrophages. Since both these elements are found in IBD, they could be mechanisms to explain the enhanced SOCS3 expression in IBD [111].

Furthermore, peroxisome proliferator-activated receptor-γ (PPAR γ) expression was found to be upregulate in M2 but not M1 macrophages in ameliorating intestinal inflammation associated with IBD. A research study showed that in macrophages deficient to PPARγ, a crucial upregulation of SOCS3 was also noted [112]. Interestingly, recent research articles have also investigated the dynamic behaviour of JAK-STAT pathway in different cells of innate and adaptive immunity, and they have successfully established a link between the members of this pathway to the responsiveness of upstream cytokines, such as IFN type 1 [113, 114]. However, it is still unclear if this dynamic behaviour of JAK-STAT pathway or even its regulator, SOCS3, exists or if it is altered in the pathogenesis of IBD.

Conversely, research reports using mouse models has demonstrated that the absence of glycoprotein 130 (gp130)-IL-6-STAT3 pathway in intestinal epithelial cells induces susceptibility to luminal irritant Dextran Sulphate Sodium (DSS)-induced colitis. This appears to be due to a reduced epithelial cell survival and proliferation. Therefore, the high persistent expression of SOCS3 in epithelial cells in inactive UC may diminish normal IL-6-STAT3-regulated epithelial homeostasis, leading to high susceptibility of these epithelial cells to inflammatory impairment [115]. Furthermore, Li et al. [116] showed substantial silencing of SOCS3 expression in the inflammation-dysplasia-carcinoma sequence in patients diagnosed with UC, but absent in patients with Crohn's disease (CD)-associated dysplasia and patients with CD-associated CRC.

On this line of evidence, accumulating reports suggest that SOCS3 may act as a tumour suppressor in the intestine. This has been best demonstrated by the observation that mice

with IEC-specific deficiency in SOCS3, were highly susceptible to IEC injury in the colon following administration of DSS [91, 117].

Reciprocally, *in vitro*, elevated expression of SOCS3 limited cell proliferation of colon cancer cell lines [118]. Rigby et al. [118] also showed that in AOM/DSS model of colitis-associated cancer, IEC-specific deletion of SOCS3 increased crypt proliferation and promoted tumour growth. These mice exhibited increases in both number and size of tumours, when compared to SOCS3-sufficient mice. Finally, recently data implicate elevated SOCS3 expression, observed in IBD, potentially with characteristic reduced epithelial restitution and repair [119].

Together these data provide strong evidence that SOCS3 normally play an important role in the progression of inflammation-associated colorectal cancer. It is evident that IEC-targeted deletion of SOCS3 induces tumorigenesis by affecting multiple signalling pathways linked to intestinal cancer. However, the exact role of SOCS3 in IBD pathogenesis and IBD-CRC remains unclear [120].

In conclusion, this thesis investigated the dynamic behaviour of the tumour suppressor protein SOCS3 and its role in crucial cellular functions using whole-cell population and single-cell analysis in different epithelial and fibroblast cell lines.

1.7 AIMS

The JAK/STAT pathway, and SOCS3 in particular, are known to be implicated in intestinal homeostasis. SOCS3 has long been associated with cancer cell proliferation as well as cancer-associated inflammation. However, the role of SOCS3 in various types of malignancies is controversial as SOCS3 has been demonstrated to possess both tumour suppressor and tumour promoting properties [121]. The multiple effects of SOCS3 in different tissues hint that the proper regulation of JAK/STAT signalling in a complex environment like the intestinal mucosa is accomplished in a highly dynamic fashion.

While some research studies have investigated the dynamic behaviour of JAK-STAT pathway in various tissues by establishing a link between the members of this pathway to the responsiveness of upstream cytokines, its functional role during health and disease is still unclear. In this thesis we focus on the dynamic expression of SOCS3 and pSTAT3 proteins.

In chapter 2, we investigate the oscillatory capacity of these proteins using whole-cell population analysis. The aim of this chapter is to investigate whether SOCS3 and pSTAT3 oscillate in human intestinal epithelial cells and colon cancer cells. In addition, we assess different methods for synchronizing cells to elucidate SOCS3 and pSTAT3 oscillation.

In chapter 3, the development of a single-cell based assay using the CRISPR/Cas9 system is undertaken. We assess the effectiveness of CRISPR-Cas9 system to engineer cell clones stably expressing SOCS3 C-terminal GFP fusion protein from its respective endogenous promoter.

In chapter 4, we utilize the Flp-In system and the tandem fluorescent protein timer (tFT) strategy to investigate localisation and trafficking of SOCS3 protein and monitor its promoter activity in response to different stimuli. In addition, we assess whether SOCS3 dynamics change across spatial and temporal dimensions under normal culture conditions or different treatments that are known to influence on SOCS3 half-life and degradation rates.

2 Whole-culture synchronization method: caveat emptor

2.1 Introduction

The recent tremendous breakthroughs towards identifying proteins that are activated or repressed in response to specific inputs, assembling them into signal transduction networks, and establishing interactions between them, has led to a growing appreciation of the complexity of these networks.

In the past two decades, it has become clear that signal transduction pathways are more than two-dimensional pathways composed of proteins just activated or repressed in response to specific inputs. In contrast, the dynamics of key proteins such as the regulation of the strength and quality of the signal is a largely unexplored area of research but the existing evidence indicates that this is a fruitful area for further study [122].

Physiological processes in living cells show a wide range of dynamic behaviours. Oscillations are commonly observed in cellular behaviour and span a wide range of

timescales, from fractions of a second/minute (calcium signalling) to hours/days (circadian clock and cell cycle). In between lie oscillations, known as 'ultradian', with time periods of the order of hours and they have been found in several systems of interacting proteins including the Hes1 regulatory protein which controls neuronal differentiation, the p53-Mdm2 system that mediates the DNA damage response and the NF-kB signalling network that governs the immune response and inflammation [123-128].

These signal transduction networks (STN) play an important role in regulating molecular communications and lead to the dynamic spatio-temporal pattern of gene expression. Furthermore, gene expression is largely controlled by transcription factors. However, the activities and levels of these transcription factors are controlled by different types of molecules, including kinases and phosphatases. The abundance of these regulatory molecules are also controlled by other gene regulatory networks, illustrating the complexity of different networks functionally interacting with each other [129]. Signal transduction systems typically comprise three types of network wirings: the feedback loops, feed forward loops and autoregulatory circuits (figure 2.1A-C). These network motifs are known to play a key role in biochemical systems in a broad range of species and produce gene expression patterns in a spatio-temporal manner [130].
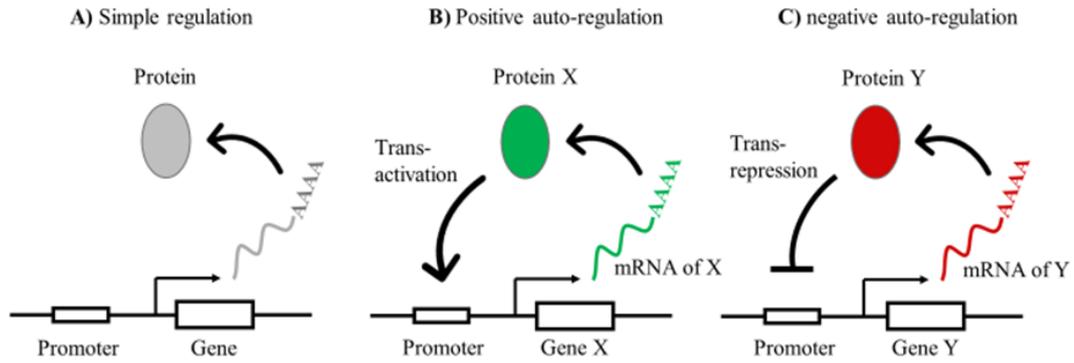
Figure 2.1: **The simplest forms of signal transduction network motifs.** A) During a simple regulation, a gene is transcribed into a mRNA and subsequently translated into a protein, which eventually turns over. B) Positive auto-regulation occurs when the transcription activator increases its own expression. C) During negative auto-regulation, the transcription factor supresses the transcription of its own gene. Adapted from [129].

Positive feedback loops are known to amplify signals and maintain the chosen state for several generations by epigenetic memory, whereas negative feedback loops are found to attenuate signals. Both types of network motifs exhibit dynamic behaviour in a broad range of kinetic parameters. Moreover, the interplay of multiple feedback loops are found to impact in a better robustness of the biological system [131]. The simplest form a positive feedback loop comprises the positive autoregulation (figure 2.1.B), where it occurs when a transcription factor (protein X) enhances its own expression. On the other hand, during negative autoregulation (figure 2.1.C), a transcription factor (protein Y) interacts with its own promoter site and supresses its own gene expression [130].

However, in intracellular molecular systems, feedback loops are often compromised by multiple different components of diverse characteristics, such as nucleic acids, proteins, and small molecules that are connected in different ways to carry key complex cellular

functions. Notably, oscillatory gene expression in eukaryotes and prokaryotes is primarily driven by negative feedback loops. A biological network lacking these motifs can only establish a unique fixed point, regardless of the initial condition [132].

Typically, the structure of a negative feedback loop is composed of a transcription factor and a repressor. This type of network motif can give rise to a number of distinct oscillations in respect to the relative values of feedback constraints including the time at which the gene encoding the inhibitor is expressed, or the rate at which the inhibitor and the transcription factor form an inactive system. A loop with the same logical structure can give rise to a monotonic increase protein levels in response to a stimulus, damped oscillations or undamped (sustained) oscillations (figure 2.2). If the time delay between the increase of the transcription factor and the increase of the repressor is long, oscillatory behaviour is assumed. However, if the rate of basal degradation of these proteins is great, this would result in the formation of damped oscillations [124, 129, 133].
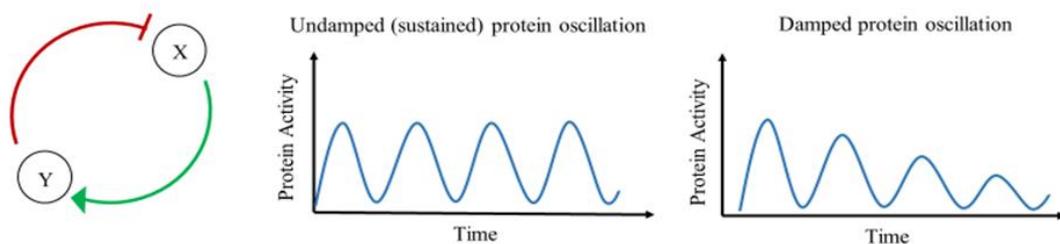
Figure 2.2: A typical model of a negative feedback loop and its different behaviour shifts from steady state, to sustained or damped oscillation. Adapted from [123, 124].

One such example is the oscillatory capacity of the tumour suppressor protein p53 [134].
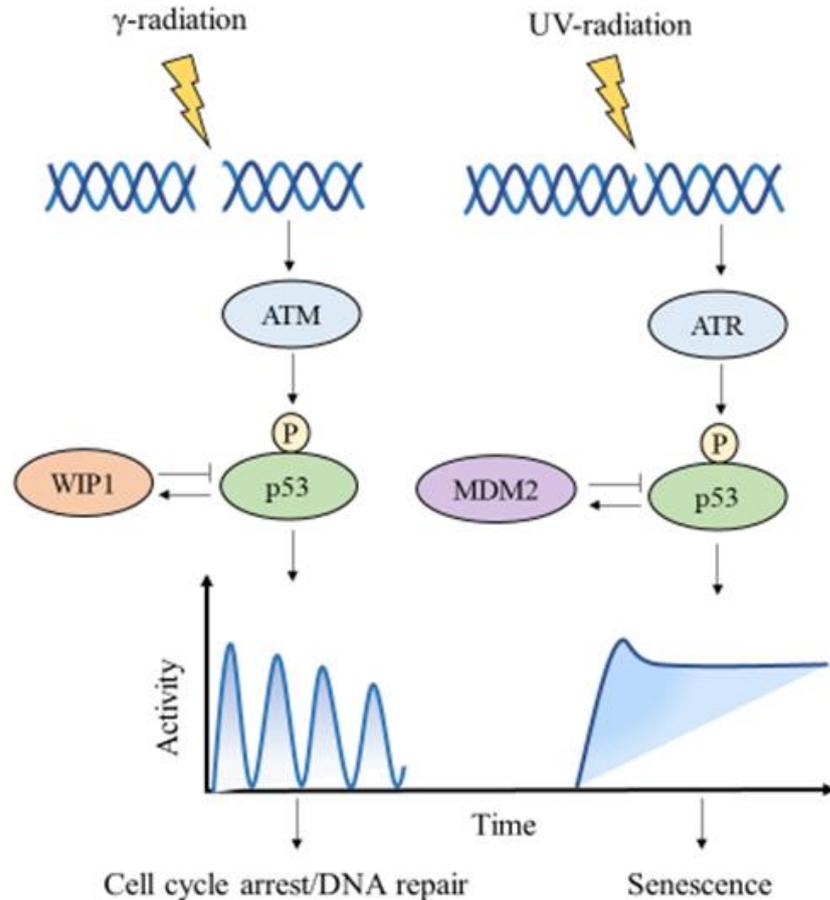
Figure 2.3: Dynamic behaviour of p53 in response to γ-radiation and UV-radiation. DNA damage by γ-radiation or UV radiation induces the activation of two kinases, ataxia telangiectasia mutated (ATM) or ataxia telangiectasia and Rad3-related (ATR) respectively. These kinases, in turn, cause the phosphorylation of p53. Phosphorylated p53 is then transferred to the nucleus and triggers the transcription of a number of target genes, including its regulators. Two key regulators of p53 include the ubiquitin ligase MDM2, which leads p53 protein to proteasomal degradation and phosphatase WIP1, which decreases p53 stability through dephosphorylation [130]. Adapted from [130, 135].

Computational and experimental analysis of p53 oscillatory behaviour have demonstrated that in response to double strand breaks (DSBs) induced by γ-irradiation, the expression levels of p53 protein is elevated dramatically and then reduced in a series of damped oscillation, in which the amplitude of the oscillations is declined in time

(figure 2.3) [136]. On the other hand, single-stranded DNA damage generated by ultraviolet (UV) induced a graded single pulse of p53 activation rather than pulsatile oscillation [137]. Additionally, when an inhibitor of Mdm2 (Nutlin-3) was added to the system to artificially stimulate the sustained induction of p53 activity, cells selected to undergo the fate of senescence rather than apoptosis, demonstrating that p53 oscillation acts as a cell fate determinant. These results suggest that p53 pulses can induce different cellular responses (cell cycle arrest/DNA repair, senescence and apoptosis [137, 138].

Detailed analysis of these dynamic behaviours in diverse biological networks and under various conditions underlies its potential to improve our understanding of the core mechanisms that drive these signalling pathways and control cell function. Ultimately, visualizing these dynamic behaviours and identifying how they differ, might also provide insights into varying behaviours within cell populations or different cell types between health and disease [139].

This level of understanding can be only achieved with quantitative measurement on the dynamics of key proteins under different conditions and perturbations to the normal network structure and study how these alterations influence the dynamic behaviour [140].

2.1.1 Dynamics of JAK-STAT signaling pathway

In the past decades, great research effort has been devoted to investigating the molecular interactions of the Janus kinase (JAKs) and signal transducers and activators of transcription (STATs) signalling pathway. Despite these important efforts, the understanding of the detailed mechanisms responsible for this pathway remains limited. Furthermore, the high complexity of JAK-STAT signalling pathway, with its extensive

list of genes encoding ligands, receptors and downstream regulators, and the vast degree of crosstalk between pathways, ensures that this task remains an enormous challenge [60, 141].

In the past two decades, JAK-STAT signalling system has also attracted the attention by system biologists, who have investigated the dynamics of the pathway using mathematical modelling.

Particularly, systems biology in JAK-STAT signalling cascade has proven beneficial to develop a better understanding into molecular basis of this complex signalling process, and its mechanisms into different physiological conditions [60, 113].

Regarding the link between JAK-STAT signalling and malignancy, Raia et al 2011 developed a quantitative mathematical modelling to study the dynamics of JAK-STAT pathway in two lymphoma-derived cell lines, including the primary mediastinal B-cell lymphoma (PMBL) and classical Hodgkin lymphoma (cHL) [142]. Specifically, the dynamic model demonstrated the interleukin-13 (IL13)-induced activation of the pathway integrated by IL4 receptor α, JAK2, and STAT5, and its inducible negative regulators CISH and SOCS3. Sensitive analysis of the optimized model identified six possible therapeutic targets and experimentally validate that one of their target dynamic predictions, the inhibition of STAT5 phosphorylation, induced decrease in target gene expression [142].

In a second research article, Rateitschak et al. 2010 investigated the induction of the STAT1 cascade in pancreatic stellate cells by combining quantitative experimental data and computational simulations. Their research focused on the aspects of antifibrotic IFNγ induction of the signalling pathway, the dynamic behaviour of STAT1

phosphorylation/dephosphorylation and the role of the negative regulator SOCS1 in the signalling regulation [143].

Furthermore, Christine A. Biron's research group has demonstrated that dynamic levels of STAT1 and STAT4 dictate responsiveness to upstream cytokines such as IFN type 1 in a variety of cells including CD8[+] T cells and NK cells [113, 114, 144, 145].

In a latter work, Pertsovskaya et al. 2013 set up a mathematical kinetic model accounting for the IFNβ activation of the JAK-STAT pathway and analysed the effects of the feedback loop regulation exerted by the protein inhibitor SOCS1, whose transcription is triggered by STAT1 [146]. In their work, phosphorylation of the STAT1 (pSTAT1) protein and mRNA levels of SOCS1 demonstrated an attenuated dynamic behaviour in response to IFNβ in the macrophage cell line RAW 264.7. Furthermore, mRNA levels of the interferon regulatory factor IRF1 increased rapidly in the initial 50-90 minutes after cytokine induction and begin to decline gradually about 200-250 minutes. The analysis of the mathematical model demonstrated a crucial role of the negative feedback from SOCS1 in controlling the detected attenuated oscillatory behaviour, and of the positive feedback from IRF1 in elevating STAT1 basal expression. Moreover, their study concluded that the system operates as a biological damped relaxation oscillator grounded on a phosphorylation-dephosphorylation network focused on STAT1 [146].

To date, only two studies, we know of, have successfully visualised and quantitively analysed an ultradian oscillatory behaviour of SOCS3 protein. Firstly, Yoshiura *et al.* (2007) [61] identified serum-induced ultradian oscillations of STAT3 protein and its inhibitor SOCS3 in C3H10T1/2 mouse fibroblast cells. In particularly, serum-stimulation induced an oscillation pattern of SOCS3 mRNA with peaks at 50 and 170

minutes. SOCS3 protein also displayed an oscillatory behaviour with peaks 2 hours and 4 hours. It is yet unknown as to the functional consequences of perturbation of SOCS3 oscillation [61]. Interestingly, p-STAT3 levels were oscillating with peaks at 1, 3 and 5 hours. They were also successful in establishing a link between p-STAT3 and SOCS3 oscillation by either chemically inhibiting JAK2-dependent phosphorylation of STAT3 that resulted in abolishing SOCS3 dynamic behaviour or knocking-down SOCS3 expression with siRNA which resulted in a steady upregulation of pSTAT3 levels [61].

A second, more recent study carried out by Khalili M. and her colleagues (2018) investigated the effects of oxidative stress on Hes1 oscillator and its upstream regulators using NIH3T3 mouse fibroblast cells [147]. Notably, Khalili *et al.* [147] followed the same research model of SOCS3 oscillation previously done by Yoshiura *et al*. [61], where mouse fibroblast cells were serum starved for 24 hours and then treated with a single dose of serum to achieve a synchronized cohort of cells. Upon serum stimulation, SOCS3 mRNA expression displayed an oscillation pattern with a high- and a low-intensity peaks at 120 minutes and 270 minutes respectively. On the other hand, the research article proposed that SOCS3 protein also followed an oscillatory behaviour with an oscillatory peak occurring about 4 hours after serum treatment. The protein abundance of p-STAT3 oscillated with only one peak at 120 minutes [147]. It is yet unknown as to the functional consequences of perturbation of SOCS3 oscillation [61].

2.1.2 Serum shock as method to synchronize ultradian oscillations through the circadian clock

Circadian clock is a cellular timekeeping mechanism that orchestrates diverse rhythmic functions in organisms from microscopic cyanobacteria and fungi to plants and animals [148].

Circadian clocks are endogenous molecular feedback loops with a period length of 24 hours that typically underlies a wide variety of biochemical, metabolic, physiological and behavioral cycles to facilitate adaptation to the changing light and temperature driven by the rotation of the earth [149]. In mammals, the central pacemaker encoding the circadian clock is located at the hypothalamic suprachiasmatic nucleus (SNC), and it communicates with tissues via bidirectional neuronal and humoral pathways [150].

Several authors have suggested that at least a number of circadian oscillators comprise coupled ultradian ones [151, 152]. A recent study using bioluminescence reporters investigated the expression of core clock genes in the hypothalamic suprachiasmatic nucleus (SNC) of freely moving mice for three weeks. Results suggested that the circadian rhythms were superimposed by episodic bursts, which displayed ultradian periods of approximately 3 hours and concluded that the clock gene expression in the SCN in vivo is maintained by the circadian pacemaker and ultradian rhythms of unknown origin [153].

In mammals, the core factors of the molecular circadian clock machinery, compromises an interlocked autoregulatory transcription-translation feedback loop, in which Circadian locomotor output cycles kaput (*Clock*), Brain and muscle ARNT-like protein 1 (*Bmal1*) heterodimerize to form an active transcription complex that interacts with E-box enhancer elements in the promoter region of various clock-controlled or clock genes and increase their transcription (figure 2.4) [154-156].

Activation of the clock-controlled genes initiate a series of transcriptional and translational events leading to the elevated expression of important transcription factors that mediate the expression of various output genes, some of which are crucial in

regulating metabolic and physiological functions and some can also be feed back to the circadian machinery (figure 2.4) [156-158].
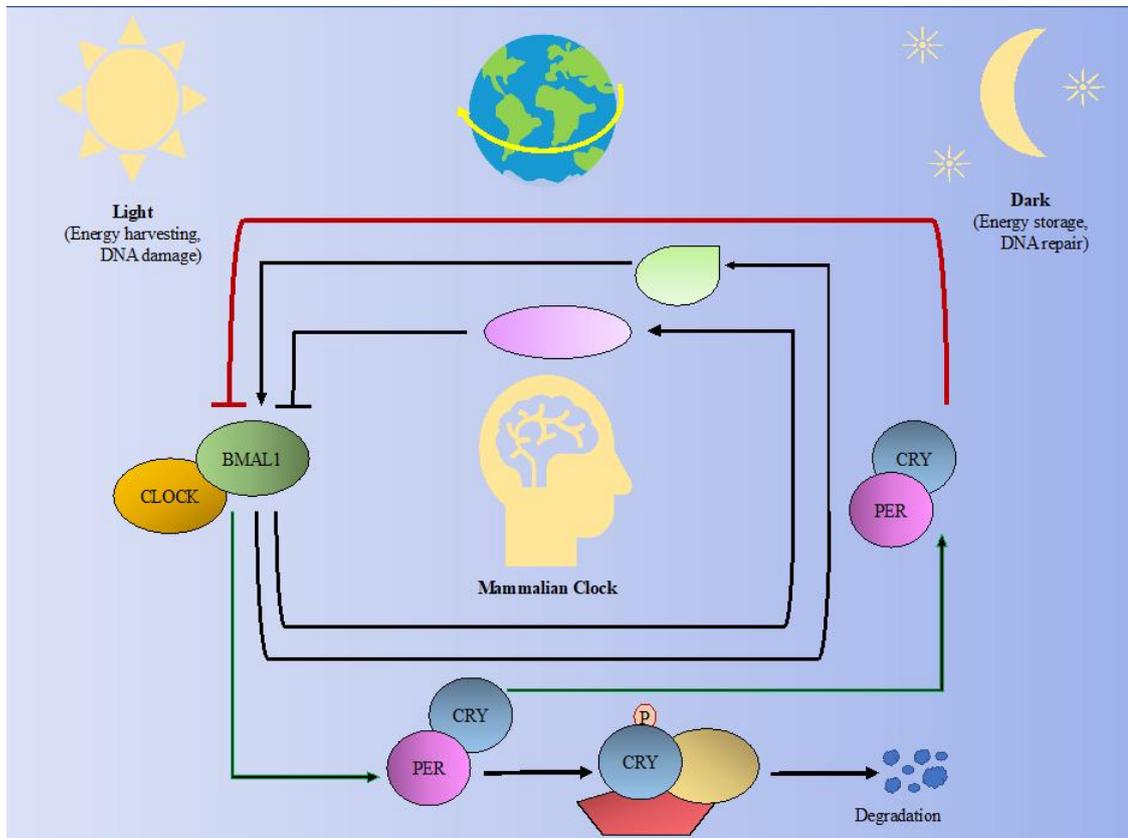
Figure 2.4: The core machinery of mammalian circadian clock pathway. The mammalian circadian clock system compromises a series of interacting transcription-translation feedback loops of circadian genes. Figure adapted from [151].

The cryptochrome (*CRY1* and *CRY2*) and period (*PER1, PER2* and *PER3*) families compromise a set of clock-controlled genes that negatively mediate the clock machinery. When Per and Cry protein levels are elevated in the cytoplasm, they form heterodimers that enter into the cell nucleus and downregulate CLOCK–BMAL1-mediated transcription, consequently terminating the negative feedback loop [124, 152]. A cyclic ubiquitin-proteasome proteolytic degradation of the repressor complex allows the feed-forward pattern of gene activation to resume [149, 159].

The circadian system orchestrates the control of virtually all aspects of human physiology such as sleep and wakefulness, body temperature, blood pressure, hormone production, digestive secretion and immune response [160]. Furthermore, it has been demonstrated that disruption of the central pacemaker is linked to many disorders including insomnia, jet lag, coronary heart attacks and depression [161]. Importantly, it has also commonly been detected among cancer patients. There has been a long history of research investigating the association between dysregulation of circadian rhythms and cancer. Dysregulation of the circadian clock machinery have been associated with the progression of many different malignancies such as breast [162], prostate [163], endometrial [164], colon [165], lung [166], ovarian [167] and hepatocellular carcinoma [168]. Moreover, studies have demonstrated that the circadian regulation of cell fate influences the response of cancer treatment: the efficacy and/or toxicity of radiotherapy and anti-malignant therapeutics have been indicate to be dependent on the timing of dose administration [169]. Therefore, a better understanding of the intimate connection between circadian clock and tumorigenesis is needed to improve development of new therapeutics, and in the optimization of current therapies.

Interestingly, the roles of the core circadian components have been reviewed in several malignant versus healthy tissues via overexpression or deletion experiments and it has been reported that circadian rhythms affect key signaling pathways. For instance, a recent research study using lung cancer and glioma cells, showed that knockdown of BMAL1 increased cancer cell proliferation, invasion, and tumor growth, while its overexpression decreased cellular invasiveness by suppressing the PI3K-Akt-MMP-2 pathway [170].

Another recent study conducted by Luo and Sehgal [171] reported that miR-279 controls behavioral output of circadian clock through the JAK/STAT pathway. Alterations in miR-279 expression levels attenuated the rest:activity rhythms in Drosophila. They also identified a ligand of JAK/STAT pathway, *Upd*, that acted as a target of miR-279, and showed that knockdown of *Upd* rescued the behavioral phenotype of miR-279 mutants. Finally, manipulations of the JAK/STAT signaling pathway dysregulated behavioral rhythms and found that the expression of STAT93E exhibited a rhythmic behavior derived from circadian regulation at the level of *Upd*. The involvement of JAK/STAT pathway in the circadian clock is a new area of research but the existing evidence indicates that this is a fruitful area for further investigation. It would be interesting to study the role of other components of JAK/STAT pathway, including SOCS3 in circadian rhythms.

2.1.3 Aims

An important aspect of cell biology using experimental cell-population techniques is to produce a synchronized cell culture. Traditionally, *in vitro* cell biology studies have performed serum starvation to produce a more homogenous population of proliferating cells since they withdraw from the cell cycle to enter the quiescent G0/G1 phase. However, recent evidence suggests that despite cell cycle synchrony, molecular processes of several proteins are not dependent on cell cycle status.

Furthermore, serum-starvation-induced synchronization, followed by serum shock (e.g., 50% serum) has been routinely used in cell cycle and circadian rhythm research. Even though, the precise relation of ultradian to peripheral circadian rhythms remain elusive, it has been shown that CLOCK protein synchrony can be achieved by serum shock. This chapter investigated the most effective method of synchronising intestine

cells in vitro culture, in order to fully dissect the putative functional outcomes of oscillatory proteins such as SOCS3 in health and disease.

2.2 Material and Methods

2.2.1 Cell Culture

To date, the development of a long-term primary culture of normal small intestinal and colon cells has not been possible despite the tremendous advances in isolation methodologies and manipulation of the cell microenvironment. This has led to the establishment of a large panel of cells lines derived from gastrointestinal tumours that have been widely used in *in vitro* and *in vivo* studies to provide insight into the biological and physiological mechanisms underlying carcinogenesis as well as the identification of factors able to generate a more benign or differentiated phenotype in neoplastic cells [172, 173]. In this respect, the human colon adenocarcinoma cell line Caco-2 has been extensively used to illustrate physiologic and cell biologic intestinal processes *in vitro* [174]. Furthermore, Caco-2 cells show many similar properties of the small intestine epithelium when grown in culture as they form a well-differentiated polarized monolayer of columnar absorptive cells in long-term culture after post-confluence.

The differentiation process in the culture of CaCo-2 cells is similar to the differentiation process of the mature enterocytes residing along the crypt-villus axis of the small intestine [175]. The differentiation process of Caco-2 cells is growth related as exponentially-dividing cells are undifferentiated, and differentiation is only induced once confluence has been reached [176]. Fully differentiated Caco-2 cells are described by cell polarization, presence of intercellular tight junctions and typical brush border microvilli projecting perpendicularly to the surface. The expression and activity of brush border membrane-associated hydrolases are elevated during cellular

differentiation. Among these is the hydrolase alkaline phosphatase (alkP), which is widely used as a marker of differentiation in colon cancer cells [177, 178].

2.2.2 Cell-line maintenance and subculturing protocol

The human colon adenocarcinoma cell line Caco-2 was obtained from the European Collection of cell cultures (ECACC, Porton Down, UK). In this report, human adenocarcinoma (Caco-2) cells were cultured routinely in T75 flasks with Minimum Essential Medium-MEM supplemented with 10% (v/v) fetal bovine serum (FBS) and 1% of MEM Non-Essential Amino Acid (NEAA). This will be assigned as "complete medium".

Human Intestinal Epithelial Cells (HIECs) compromise a human, foetal-derived cells, a gift from Jean-Francois Beaulieu (Université de Sherbrooke). HIECs were cultured in Opti-MEM + Glutamax medium, supplemented with 5% Foetal bovine serum (FBS), 1 % 0.01M HEPES and 10 µl of 0.1 mg/ml epidermal growth factor (EGF).

All cells were maintained at 37°C in the 5% CO_2 atmospheric condition. Medium was changed every 2 to 3 days. When cells reached the subculturing density of 70% confluence, they would either be passaged into new T75 culture flask for further culture or into Petri dishes or 6/12-well plates for experiments.

When subculturing cells, the media was discarded prior rinsing the cells twice gently with Dulbecco's Phosphate-Buffered Salines (DPBS), no calcium, no magnesium (Thermo Fisher Scientific, Loughborough UK). Adherent cells were detached from the bottom of cell culture flasks with Trypsin supplemented with EDTA (0.25%) (Thermo Fisher Scientific). Cells were incubated for 5-10 minutes in a incubator at 37°C in the 5% CO_2 atmospheric condition and then resuspended with fresh media and reseeded.

2.2.3 Alkaline Phosphatase (alkP) cell staining

The process of CaCo2 cell differentiation from day 0 to day 21 was investigated via alkP cell staining using the VECTOR Red Alkaline Phosphatase (Red AP) Substrate Kit (VECTOR LABORATORIES, Cat No. SK-5100) according to the manufacturer's suggested protocol. Briefly, Caco-2 cells were seeded into cover slips at a density of 10^5 cells/ ml in a 6 well plate and maintained for 3, 7, 11, 15, 18 and 21 days in complete medium; the medium was changed every 2 to 3 days. At indicated times, Caco2 cells were rinsed with pre-warmed DPBS without calcium and magnesium and a substrate working solution was prepared with the following components added in a 5ml of 150 mM Tris-HCL, pH 8.2 buffer:

- 2 drops (80 µl) of Reagent 1

- 2 drops (80 µl) of Reagent 2

- 2 drops (80 µl) of Reagent 3

The final working solution was mixed well and added to Caco2 cells and incubated with the substrate working solution for 30 minutes. Following incubation, the cell monolayer was rinsed twice with buffer and then with DPBS without Calcium and Magnesium. Following the washing step, 5 µl of of Hoechst 3342 nuclear stain reagent was added in 5 ml of buffer, mixed well and added to the cells for another 30 minutes at 37°C. After the incubation, cells were carefully rinsed with DPBS without Calcium and Magnesium to remove and dead cells in the well and fixed with freshly made 4% paraformaldehyde (PFA) for 20 minutes at room temperature. PFA was then washed off by buffer for 5 minutes. Following the washing steps, cover slips were dehydrated, cleared and mounted into microscopic slides using a permanent mounting medium, such as VectaMount (VECTOR LABORATORIES, Cat. No. H-500) before imaging with a

Zeiss confocal microscope by using mCherry and DAPI filter sets and ZEN (blue edition) Image Software.

2.2.4 Cell differentiation protocol

For the differentiation experiments, Caco-2 cells were seeded into 12-well plates at a density of 10^5 cells/ ml and maintained for 14-16 days in complete medium; the medium was changed every 2 to 3 days.

2.2.5 Cell cycle synchronization procedure

2.2.5.1 Cell cycle synchronization

For cell cycle synchronization, HIECS and Caco-2 cells were subjected to serum deprivation for 24 hours or 48 hours and thereafter resupplied with complete medium.

2.2.5.2 Cell population synchronization based on the circadian clock system

The integrative physiology of circadian and metabolic systems has emerged through a combination of biochemical and experimental genetic research involving *in vivo*, *ex vivo* and *in vitro* studies carried out in many different laboratories [179]. In relation to *in vivo* and *ex vivo* studies for investigating clock rhythms, cell based *in vitro* assays are more experimentally tractable for phenotypic characterization and rapid discovery of the core circadian system [180].

Considering that circadian oscillations are dynamic and longitudinal measurements with high temporal resolution, the systematic analysis of this oscillatory behaviour in cell-based studies strategically requires to produce a synchronized cohort of cells using

a wide range of intervening conditions, including several serum concentrations [181], different temperatures [182, 183], and treatment with pharmacological compounds [180]. On this report, population synchronization in cultured cells was entrained through treatment with high concentrations of serum (50%) for 2 hours to serum starved cells.

2.2.6 Western Immunoblot

2.2.6.1 Cell lysis

For cell cycle synchronization time series, cells were treated with serum at t = 0, and they were rinsed twice with DPBS with Calcium and Magnesium and harvested with Pierce IP lysis buffer (Thermo Scientific) every 30 minutes up to two hours and every 1 hour afterwards up to 8 hours. For serum shock synchronization time series, cells were treated wish a serum shock for 2 hours and they were harvested every 10 minutes up to two hours. Cell lysis was carried out on ice to prevent protein degradation. Insoluble cell debris was removed by centrifugation at 12,000 rpm for 2 minutes. Cell lysates were stored at -20°C.

2.2.6.2 SDS-PAGE

Protein Samples were mixed with 4x NuPGE LDS sample buffer (Life Technologies) or 4x Laemmli Sample buffer (Biorad) supplemented with 5% of 2-mercaptoethanol and boiled at 95°C for 2 minutes. Equal amounts of protein resolved on AnyKD™ Mini-PROTEAN® TGX Stain-Free™ Precast Gels. 5 μl of BioRad precision plus protein standards were also loaded as a molecular weight reference. The precision plus protein standards contain a mixture of 10 recombinant proteins (10-250kD), including three reference bands (25, 50, and 75kD). Running buffer consisting of 25 mM Tris, 192 mM

Glycine, and 0.1% SDS was used. The gel was run at a constant running voltage of 180V until protein bands reached at the bottom of the gel. Proteins were then transferred onto a nitrocellulose membrane by using a Trans-Blot Turbo System (BioRad) set at 1.3 A/25 V/7 min.

2.2.6.3 Antibody Staining

Next, the nitrocellulose membrane is blocked in 2% Bovine Serum Albumin (BSA) in Tris-buffered saline with 0.1% Tween (TBS/T) for 1 hour at room temperature to prevent any nonspecific binding of antibodies to the surface of the membrane. After the blocking step, membrane was incubated with primary antibody diluted in 1-3% BSA with 0.1% Tween (TBS/T) Blocking buffer. Incubation was performed overnight at 4°C. Membrane was then washed with 1X TBS supplemented with 0.1% Tween for 3 times, 10min each. This was followed by incubation with the appropriate horseradish peroxidase-conjugated anti-mouse monoclonal antibody (1:10,000) or anti-rabbit monoclonal antibody (1:10,000) for 1 hour in room temperature. Washing steps were repeated after the incubation. This is followed by blot developing with chemiluminescence (ECL Plus) and visualized using the BioRad ChemiDoxTM XRS+ System. Densitometry was performed via Image Lab™ software to measure the expression of the protein of interest.

2.2.6.4 Stain-free total protein staining

AnyKD™ Mini-PROTEAN® TGX Stain-Free™ Precast Gels were used because they enable a quick and easy way to visualize proteins in the gel prior and post transfer. Stain-Free (SF) gels incorporate the proprietary trihalo compound in the gel formulation

that reacts with proteins to enable total protein visualization by CCD (charge-coupled device) imagers [184]. This methodology is based on the ability of the tryptophan residues within proteins to undergo an ultraviolet light-induced reaction with trihalo compounds to produce fluorescence [185]. Western blot was carried by loading equal amounts of protein. Gels were activated by ultraviolet (UV) exposure for 2 min using the Bio-Rad ChemiDoc MP imager and subsequently transferred to nitrocellulose membranes (Bio-Rad) using a Trans-Blot Turbo apparatus (Bio-Rad). After protein transfer, membranes were imaged for Stain-Free staining and total protein was quantified using ImageLab 4.1 software (Bio-Rad). Membranes were then blocked with 2% (w/v) BSA in TBST for 1 h and then stained with primary antibodies.

2.2.7 Mathematical Modeling

Gaussian Process regression with short-range and periodic correlation components was used to quantify periodicity of SOCS3 and pSTAT3 oscillation as a function of the serum shock (stimulus). The advantage the Gaussian process regression modelling approach is that it permits the detection and quantification of periodic components of a noisy timeseries dataset of a limited number of measurements where assumptions about the shape of the periodicity (e.g. sinusoidal) are not desired. The Gaussian Process regression analysis with short range and periodic correlation components is essentially a smoothing process; the intuition is that measurements that are close together in time should be more similar than measurements far apart in time, and that measurements that occur at the same time with respect to a 120-minute period should also be similar. The output of the analysis is a smoothed representation of the data, from which evidence for a consistent variation in the measurement data over time can be drawn.

Consider a time series regression task where we have the y_{it} outcome (i.e. independent variable) for the ith experimental replicate at time t.

Each observation y_{it} can be related to the underlying function of the temporally-correlated random effect $u(t)$ and an overall average (mean) value α through a Gaussian noise model

$$y_{it} = \alpha + u(t) + \varepsilon_{it}$$

where ε_{it} represents measurement noise.

The random effect $u(t)$ is assumed to have a Gaussian distribution with zero mean, and the correlation $p(t, t')$ between observations is assumed to be comprised of a Matérn

structure with effective range of 15 minutes as well as a 120-minute periodic structure. Therefore

$$u(t) \sim GP(0, p(t, t')).$$

Having fitted the Gaussian process regression model to the measurement data, the predictive mean value of the outcome measure $\hat{y}_t$ with respect to time is plotted, together with 95% confidence intervals. Evidence for a temporal signal is taken to be where the 95% confidence interval does not include *y=0* for all timepoints.

2.3 Results

2.3.1 Effects of serum deprivation on SOCS3 dynamic behaviour

SOCS3 has been proven to behave as an anti-proliferative agent in several human intestinal cell lines as well as colorectal cancer cell lines, such as the Caco-2 cell line [118, 186]. However, only recently has been indicated that the turnover of SOCS3 is controlled partly through an autophagy-based mechanism [65]. Thus, it is important to establish and study the oscillatory behaviour of the basal expression of SOCS3 and its regulatory pattern in proliferating Caco-2 cell line.

Furthermore, serum-deprived cells are synchronized at a point in the G1 phase or in the G0 phase and when a stimulus is applied, the stimulated cells grow and respond as a synchronized cohort through the cell cycle [187]. There is a number of research articles that have used this method to synchronize the Caco-2 cell population for their experimental designs. In most experiments, confluent monolayers of Caco-2 cells were synchronized by serum deprivation for overnight, 1 or even 2 days [65, 118, 188, 189].

To identify the dynamic profile of SOCS3 basal expression, this report initially investigated the ultradian oscillation by performing protein expression profiling of Caco2 cells after serum stimulation. In the first set of experiments, 2.5×10^5 cells were seeded in petri dishes containing complete medium supplemented with 10% FBS for 24 hours. This was to allow Caco-2 cells to adhere but prevent them to undergo instant differentiation since it only achieved 60% of cell density. Caco-2 cells were serum starved in serum free medium overnight to produce a synchronized cohort of cell population and arrest of cell cycle at a G0 or G1 state. Caco-2 cells were subjected to serum stimulation time course as described in Methods.

The results from the time course experiment are presented in figure 2.5.
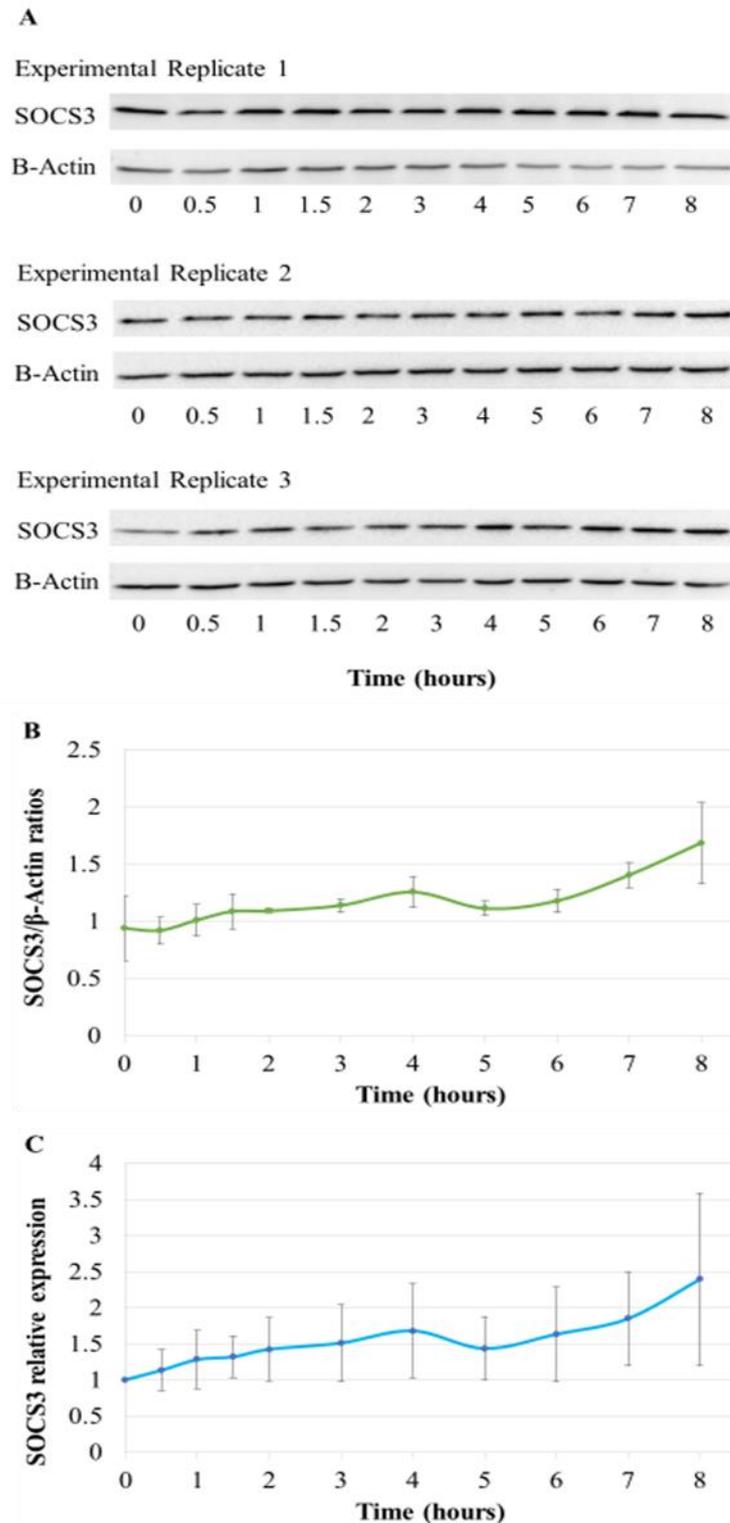
Figure 2.5: Relative expression of SOCS3 to β-actin protein levels. **(A)** Immunoblots of β-actin and SOCS3 in serum starved proliferative Caco2 cells (N=3). **(B)** Quantification of SOCS3 and levels normalized by β-actin, in units relative to peak intensity **(C)** Relative

abundance of SOCS3 protein expressed as a fold change in relation to t=0. Means with standard error (SE) of three independent experiments are shown in all graphs.

The data from the western immunoblot image in figure 2.5 does not show any visible fluctuation in the expression of SOCS3 during the 8-hour time course. Even though, the bands of β-actin protein show an equal protein loading across the blot, the above results fail to show the presence of any type of dynamic behaviour of SOCS3 after serum stimulation. The analysis of the western blot assay also demonstrated that the expression of SOCS3 protein is steadily increased during the time course to serum stimulation indicating no oscillation behaviour.

Furthermore, low standard errors (SE) in the graph illustrating the ratios levels of SOCS3 and β-actin (figure 2.5B) demonstrates that the uncertainty of the reported measurements during the 8-hour course is low. However, the value of the error bars of the first-time point (t=0) is higher than the rest of the time points, excluding the last time point. Consequently, the fold change of SOCS3 protein in relation to t=0 (figure 2.5C) generates higher error bars in the rest of the time course. Moreover, it is also evident that the values of the error bars on the same graph are increased in relation to time point indicating that the cell population is losing the ability to behave as a synchronized cohort during time.

The absence of the oscillatory capacity in the Caco-2 cell model system indicated that the above cell culture conditions prior to serum stimulation failed to show the oscillatory behaviour of SOCS3 protein. It appears that the degree of synchronization of the cell population was not adequate to elicit a dynamic behaviour of SOCS3 protein. Therefore, to further attempt to establish an oscillatory pattern of SOCS3 protein, the human colorectal Caco-2 cell line was seeded in complete medium overnight to allow cells to

adhere. The culture medium was then changed, and cells were incubated with serum free medium for 48 hours to achieve a higher degree of cell synchronization. The extracted proteins for each time point was assessed with western blot assay and the analysis of the results are presented in figure 2.6.
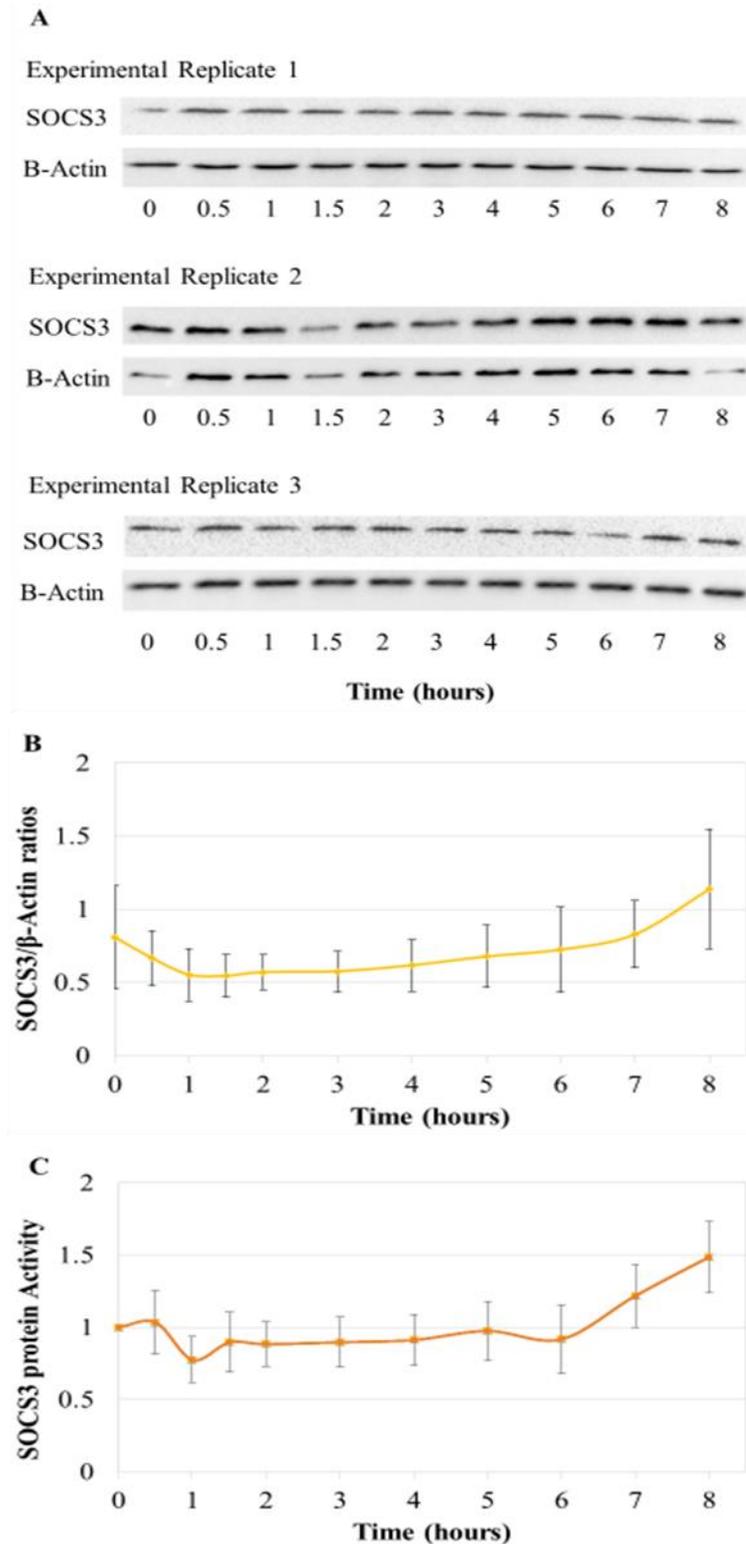
Figure 2.6: **Analysis of SOCS3 basal expression after 48-hour serum starvation and serum stimulation.** Immunoblots of β-actin and SOCS3 protein profiles in proliferating Caco2 cells treatment with complete medium overnight and serum starved for 48 hours prior to serum stimulation (N=3). **(B)** Quantification of SOCS3 and levels normalized by β-actin, in

units relative to peak intensity. **(C)** Relative abundance of SOCS3 protein expressed as a fold change in relation to t=0. Means with standard error (SE) of three independent experiments are shown in all graphs.

The results from the western immunoblot image in figure 2.6A does not present any clear dynamic fluctuation in the abundance of SOCS3 during the 8-hour time course, indicating that these specific culture conditions also failed to display a distinct protein oscillating behaviour of SOCS3. Moreover, the bands of β-actin protein failed to show an equal protein loading across the three Western blot experimental replicates.

When the synchronized cell population was exposed to serum stimulation in respect to a series of time intervals, the Western blot analysis (figure 2.6 B and C) also demonstrated that the abundance of SOCS3 protein is initially low in the first 5-6 hours and it is steadily increased during the time course to serum stimulation. According to the analysis of the present time series, it is apparent that upon serum starvation of 48 hours, autophagy is highly induced in response to a prolonged nutrient deprivation, reducing SOCS3 protein abundance. This observation supports the findings of Luan K. et al. 2014 [65] that the abundance of SOCS3 is regulated partly through an autophagy-based mechanism.

Additionally, the data analysis presenting the fold change of SOCS3 protein in relation to t=0 (figure 2.6 C) showed that the abundance of SOCS3 protein showed a minor fluctuation in the time period of 2 hours. However, the presence of the high values of the errors bars and failing to generate peaks with values 1-1.5x fold higher than the lowest signal values during the 8-hour time course, the generated data can't be considered to produce statistical significant observations.

The population of Caco2 cells under study did not appear to be homogenous in their proliferating cancerous state at any given time point or culture condition to elucidate a SOCS3 dynamic behaviour. This could be beneficial for producing cell type diversity during cancer development. In this context, in a healthy intestinal tissue, various members of different signalling networks are synchronized to be activated or repressed in an oscillatory manner in response to a stimulus. If this highly regulated system is perturbed by genetic alterations or other consistent stimuli, cancer may potentially hijack this system and use it for its own benefit. This population robustness that is generated from this disturbed cellular heterogeneity could possibly be a mechanism that intestinal cancer utilizes to maintain its proliferative potential against the human immune system or anticancer therapies. Therefore, it would be interesting to investigate if there is any dynamic basal levels of SOCS3 protein in a cell model that does not appear to have a carcinogenic phenotype.

However, the establishment of a long-term primary culture of healthy small intestinal and colon cells has not been possible regardless of the great advances in isolation techniques and manipulation of the cell microenvironment. Currently, Caco2 cell line represents the best characterised gut epithelial monolayer system available for *in vitro* study because it shows shared characteristics as enterocytes residing in the human intestinal epithelium. This cell line model offers the ability to differentiate upon reaching cell confluency and subsequently the formation of a monolayer of polarised cells that share functional and structural similarities to a normal small intestinal epithelium. Thus, the following section of this report will utilize this differentiated Caco2 model system to investigate and analyse the dynamic behaviour of SOCS3 protein.

2.3.2 Dynamic behavior of SOCS3 protein in Differentiated Caco2

The colon adenocarcinoma Caco-2 cell-culture model is a well-established model system to investigate cellular differentiation of mature enterocytes in human gut epithelium [175]. In contrast to proliferating Caco-2 cells, the fully differentiated Caco-2 cells have lost their tumorigenic phenotype during the differentiation process. These phenotypic changes are associated with downregulation of various proteins involved in the cell cycle progression, DNA synthesis/repair and RNA processing as well as a series of distinct biochemical pathways that are involved in protein folding, cytoskeleton formation and maintenance and nucleotide metabolism [190, 191].

Importantly, in the literature, there is a great number of discrepancies in the findings with the use of Caco2 cell model across different laboratories and these inconsistencies might be due to several factors such as the source of cells, the culturing conditions and maintenance and the associated cell passage numbers [192].

Thus, the process of CaCo2 cell differentiation from day 0 to day 21 was investigated via cell staining for a brush border enzyme alkaline phosphatase (AlkP). In addition, it has been shown that ALkP acts as an important indicator during the different phases of enterocyte differentiation as it is highly expressed in fully differentiated Caco2 cells [177]. In agreement with the literature, confocal results (figure 2.7 and 2.8) demonstrated that the expression levels of AlkP is increased in the CaCo2 cells when they were in culture for 14 to 21 days, in relation to the first and third day of culturing, verifying that the CaCo2 cells portray differentiated phenotype from day 14 and onwards.
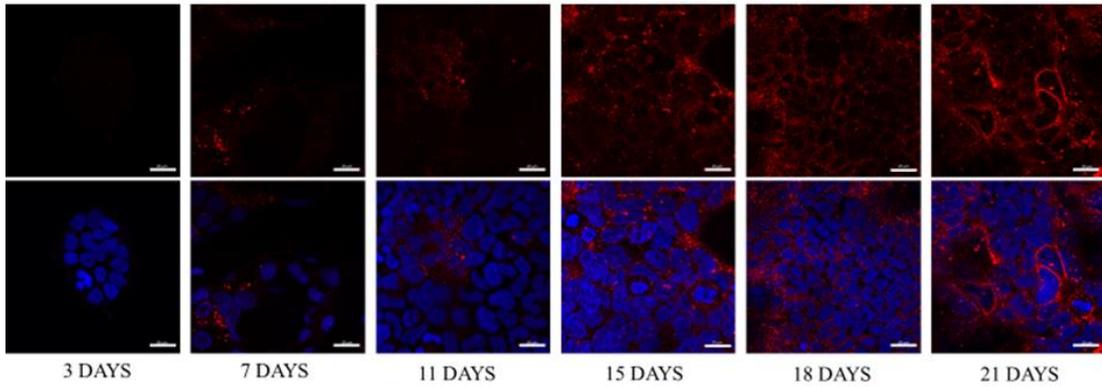
Figure 2.7: CaCo2 cell differentiation from day 0 to day 21 was investigated via cell staining for alkaline phosphatase (AlkP). Top panel shows the red staining fluorescence of Caco2 cells indicating the distribution of AlkP levels whereas the bottom panel is an overlay of the DAPI (blue) nuclear staining and AlkP staining. Images were captured with a Zeiss confocal microscope by using mCherry and DAPI filter sets and analysed with Zen (blue edition) Image software. Scale bar at 20 µm, n=2 per time point. Images taken at 40x magnification.
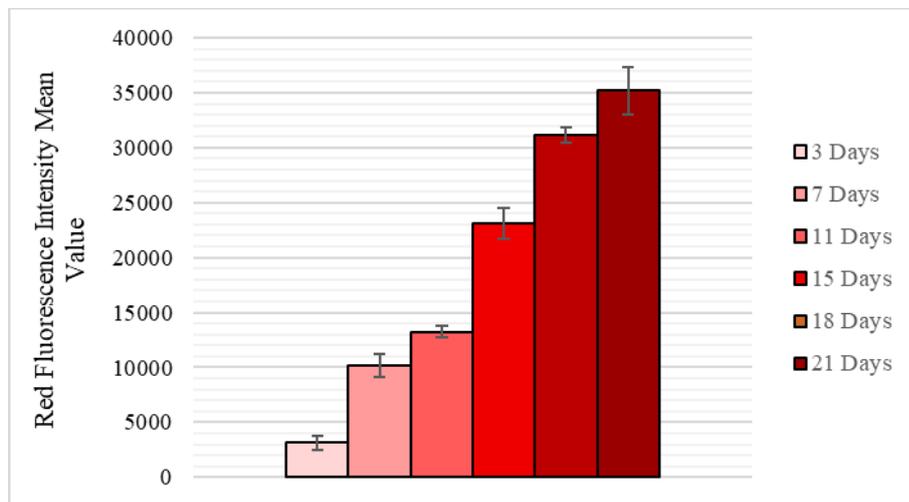
Figure 2.8: An approximate quantification of the red fluorescence intensity during CaCo2 cell differentiation from day 0 to day 21. Region of Interest (ROI) were drawn to all the images with their respective number of days during the differentiation process and the red fluorescence intensity mean values were obtained for n=2. Background correction was performed by subtracting the mean fluorescence intensity from a control sample where cells

were not stained. Images were analysed with Zen (blue edition) Image software. Means of n=2 with standard error (SE) are shown in the graph.

Additionally, an approximate measure of the mean red fluorescence signal was obtained with the use of Zen (blue edition) software and displayed a steady increase of red fluorescence from day 0 to day 21 (figure 2.8). Moreover, from confocal imaging analysis (figure 2.7), it became apparent that the differentiated Caco2 monolayer started to lose the ability to retain the tight junction formed between the cells after the 15th day of cell culture. This was evident by observing empty spaces formed at the bottom of the well where a uniform cell monolayer existed previously. Therefore, it was decided that caco2 cells were allowed to differentiate for a period of 14 days prior to start the serum starvation and subsequent serum stimulation time series experiments.

Differentiated Caco-2 cells have been widely used in several studies to investigate important cellular processes that have altered in human colorectal cancer [65, 193, 194]. However, there is no published research article in the literature demonstrating the use of differentiated Caco-2 cells to investigate the protein oscillatory capacity in human intestinal epithelium systems.
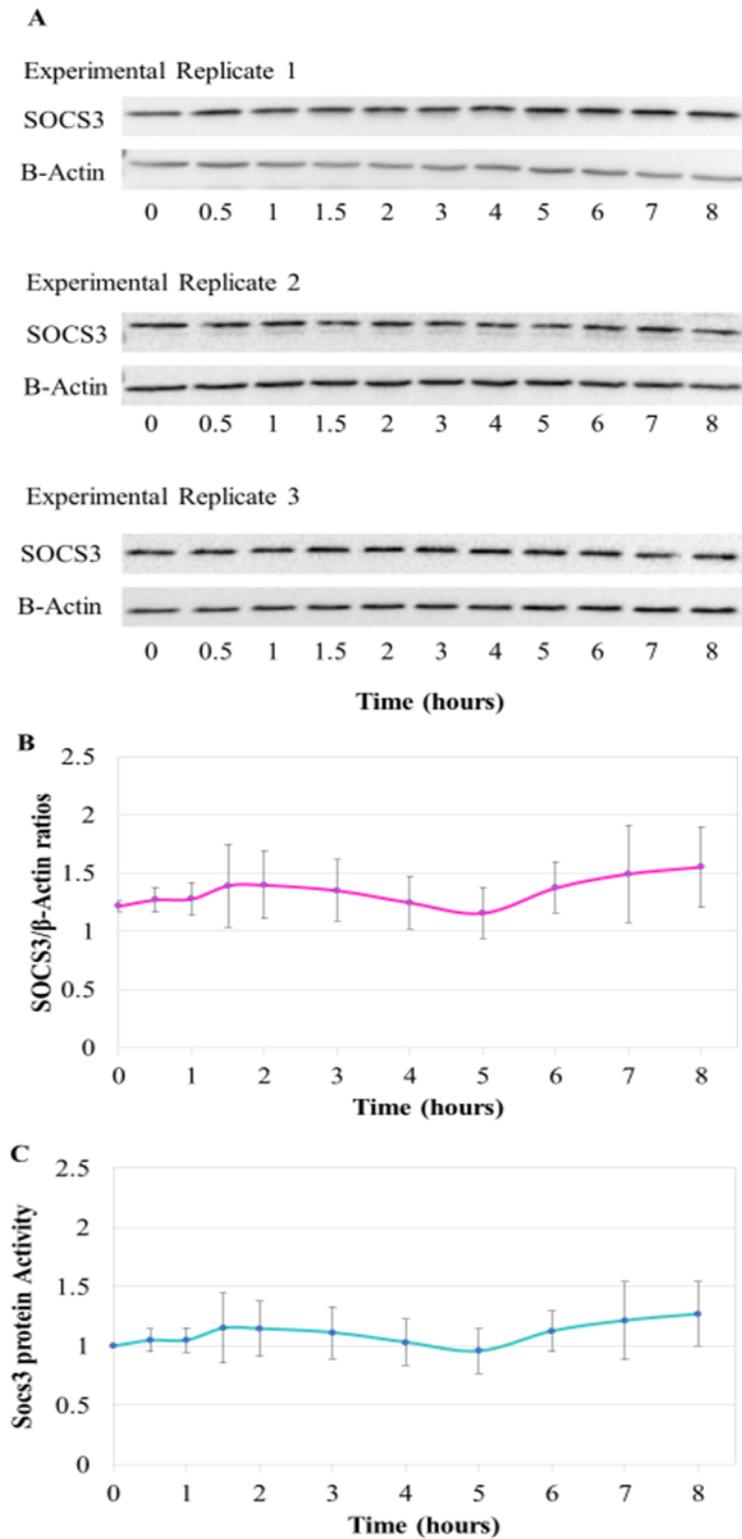
Figure 2.9: SOCS3 protein profiles in differentiated Caco-2 cells from overnight serum free medium. (A) Immunoblots of β-actin and SOCS3 kinetics in differentiated CaCo2 cells starved overnight prior to serum stimuli. **(B)** Protein quantification of the presented blots normalized by β-actin, in units relative to peak intensity. **(C)** Relative levels of SOCS3 protein

expressed as a fold change in relation to t=0. Means with standard error (SE) are displayed for each time point. Blots and graphs are representative of 3 independent experimental replicates (n=3).

For the first set of time course experiments, fully differentiated Caco-2 cell cultures were incubated with serum free medium overnight to allow cell synchronization prior to serum stimulation. The control time point 0 was not subjected to serum stimulation during the time course. The results from the time course experiment in fully differentiated Caco-2 cells are presented in figure 2.9.

The immunoblot images from the western blotting experiment (figure 2.9A) did not display any visible clear oscillatory pattern in the abundance of SOCS3 protein during the 8-hour time course in differentiated Caco2 cells. Even though, the bands of β-actin protein show an equal protein loading across the experimental replicates, the above results fail to establish the presence of any type of dynamic behaviour of SOCS3 after serum stimulation. Furthermore, the western blot analysis displaying the fold change of SOCS3 levels in relation to t=0 (figure 2.9 C) demonstrated that the abundance of SOCS3 protein showed a minor fluctuation in the time period of 8 hours. However, the data failed to generate peaks with values 1-1.5x fold higher than the lowest signal values during the time course, consequently the results can't be considered as a statistically significant.

The absence of the oscillatory capacity in the differentiated Caco-2 cell model system demonstrates that the above cell culture conditions prior to serum stimulation was inadequate to display any meaningful dynamic behaviour of SOCS3 protein. Therefore, to further attempt to establish an oscillatory pattern of SOCS3 protein, the serum starvation of the differentiated colorectal Caco-2 cell line was further optimized.

Considering carcinogenesis as a robust system, which retains its robustness capacity by redundancy through the genetic heterogeneity of cancer cells and through feedback loops that include intracellular and host-tumour interactions [195], it is important to maintain cellular stress at a minimal level in order to avoid false findings resulting from the starvation process and achieve the maximum cell synchronization in culture. Thus, the differentiated Caco-2 cell line was serum starved for 36 hours prior to serum stimulation. The data are presented in figure 2.10.
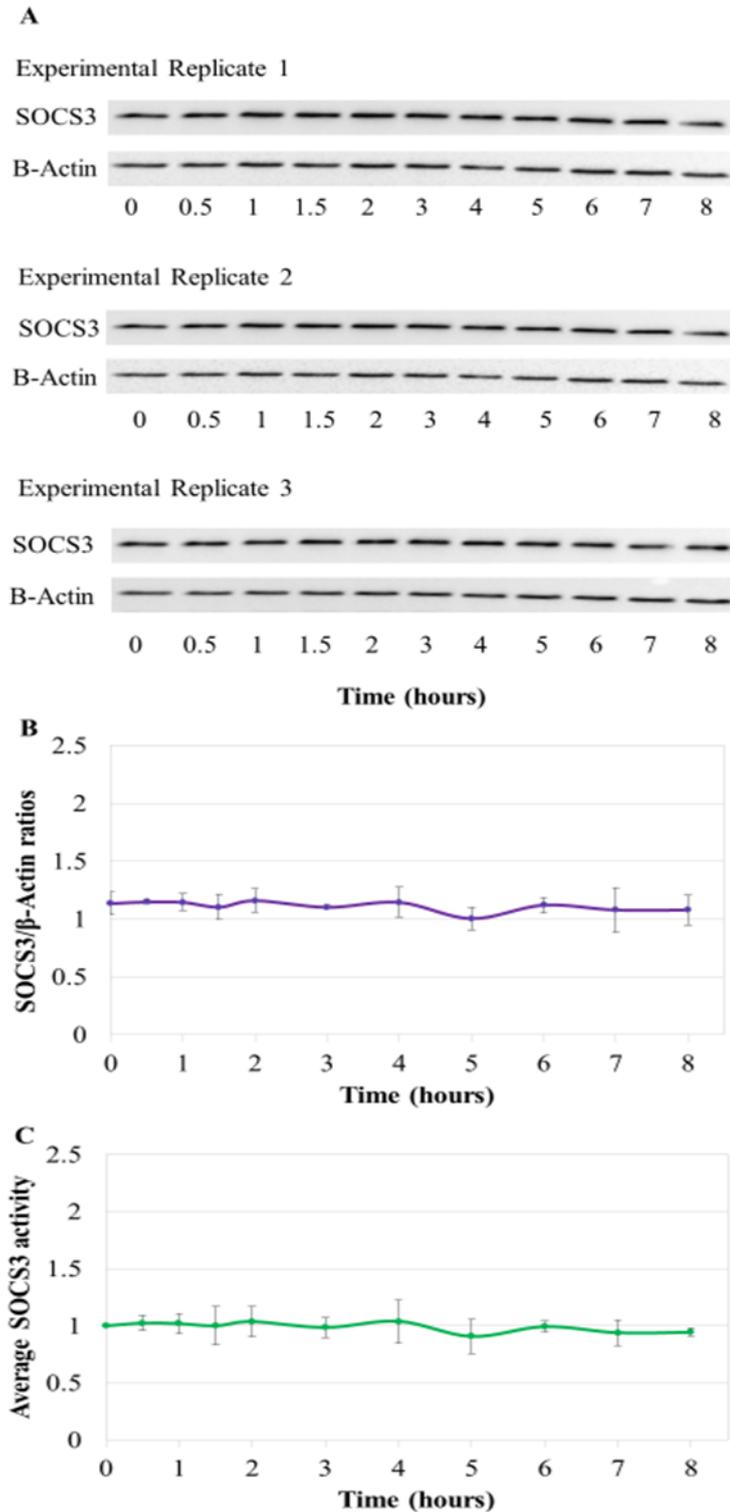
Figure 2.10: SOCS3 expression profiles in differentiated Caco2 cells from 36 hours serum free medium. (A) Western blot analysis of SOCS3 protein levels (n=3). **(B)** and **(C)** Quantification of SOCS3 levels in fully differentiated Caco-2 cells was quantitated, normalized and graphed against a linear time scale (n=3). Means with standard error (SE) are displayed for each time point.

The immunoblot figures from the western blotting experiment (figure 2.10A) demonstrate clearly that there is no oscillatory pattern in the abundance of SOCS3 protein during the 8-hour time course in differentiated Caco2 cells. It is clear from both the immunoblots and graphs that quantified SOCS3 levels against β-actin levels and normalized to the first-time point (t=0), the period of serum starvation and subsequently serum stimulation does not play an important factor in elucidating an oscillatory behaviour of SOCS3. Additionally, this concluding observation is strengthened by the low values of errors bars across the time points in the time course.

However, this report failed to detect any consistent dynamic behaviour of SOCS3 with serum stimulation. This behaviour could be explained by a weak auto-inhibition in this specific cell line, or a quick desynchronization of cell population after the addition of complete medium.

The initial objective was to establish a synchronous cell population. Despite cell cycle synchrony, molecular processes of several proteins are not dependent on cell cycle status. Peripheral circadian rhythms are thought to impact on cellular ultradian cycles and these can be entrained by serum shock [181].

Therefore, to further characterize this system, this report attempts to find other methods to synchronize the cell population and elicit the rhythmicity of SOCS3 protein abundance and study its functionality.

2.3.3 Serum shock as method to achieve whole population synchrony

A vast array of physiological and biological pathways are controlled by endogenous circadian dynamics regulated by two systems: a master clock, which operates at the systemic level and individual cellular clocks, which operates at the single cell level. The cellular clock relies on a complex system of core clock genes that control the circadian expression of non-clock genes involved in various cellular processes [196].

Serum depletion-induced synchronization, followed by serum shock (i.e. 50% v/v) has been used extensively in protein dynamic research, especially in the investigation of circadian oscillators.

To this end, this report took advantage of a serum shock strategy, initially described by Balsalobre et al. [181], through which mammalian derived fibroblasts could be stimulated to synchronously express clock genes in a circadian dynamic pattern in response to a short dose of high serum concentrations. In turn, these cellular circadian rhythms are thought to impact on cellular ultradian cycles. However, this report slightly modified this technique to incorporate a serum depletion step prior to serum shock. Although addition of serum to cell medium offers optimal conditions for cell growth, its poorly defined complex and above all inconsistent composition signifies a crucial and undesirable confounding role while performing time series [197]. Thus, depletion of serum from culture medium decreases known and unknown unknowns, eliminates analytical interference, and it offers more reproducible experimental conditions [198]. Furthermore, it has been reported to decrease basal cellular activity and noise and enforces the population of proliferating cells to act as a more homogenous cohort [199].

Therefore, this report will use this synchronization technique to attempt demonstrate, first, that the ultradian oscillatory system of JAK/STAT network can be entrained in culture, and second, that the pattern of SOCS3 and pSTAT3 protein are controlled on a cellular clock system.

In addition, it has reported that SOCS3 protein has a relatively short biological half-life that ranges from 40-120 minutes depending on the type of cell line under investigation [107]. Thus, it will be important to consider the fact that there might hidden dynamic protein fluctuations during the 30 minutes and 1-hour intervals from previous experiments. Consequently, the sample harvesting time intervals from 30 minutes and 1 hour has been decreased to 10 minutes during the time course.

Briefly, for the new optimized experimental conditions for serum shock entrainment, cells were seeded in 35 mm^2 petri dishes and allowed to grow overnight. The following day, the culture media was removed, and cells were serum starved overnight with FBS-free basal medium. Cells were then serum shocked with their respective growth complete media supplemented with 50% v/v horse serum and incubated for 2 hours. Following this, the cells were washed twice with PBS and basal culture medium was placed back into the petri dishes. The first-time point was taken after the serum shock (t=0) and then the cells were harvested every 10 minutes for a period of 2 hours.

The extracted proteins for each time point was assessed with western blot assay and immunoblot were analysed using the Stain free gel quantitative analysis described in Methods. A reference protein sample loaded on all western blot gels was used to normalize all time points and the analysis of the results from time series of proliferating Caco2 cells are presented in figure 2.11.
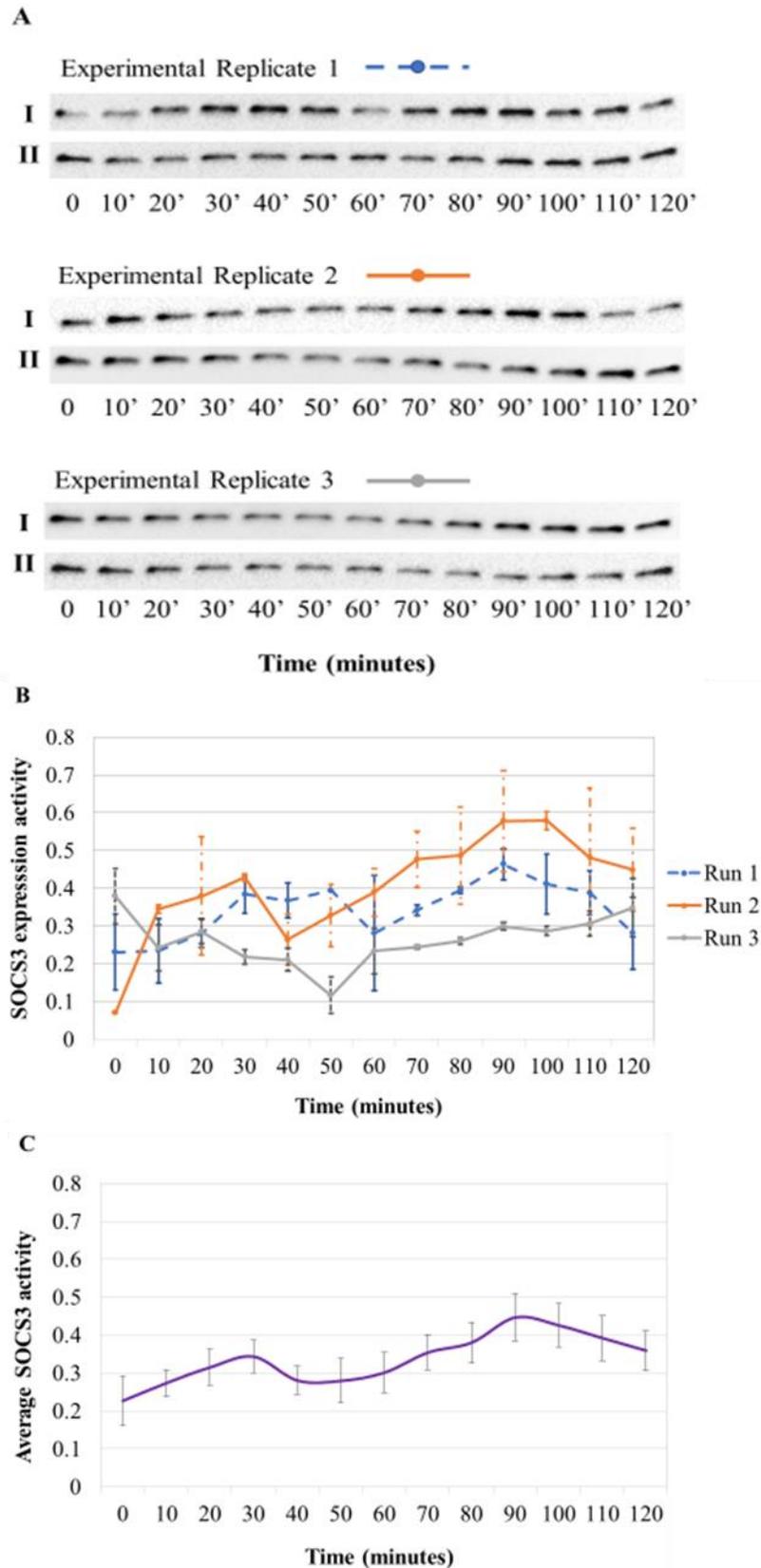
Figure 2.11: Expression profiles of SOCS3 protein in Caco2 cells after shock treatment.

(A) Western blot analysis of SOCS3 protein levels (n=3). Each Experimental replicate compromised two biological replicates. **(B)** and **(C)** Abundance of SOCS3 protein in Caco-2 cells was quantitated, normalized and graphed against time. **(B)** graph shows the dynamics of each experimental run whereas **(C)** shows the average (n=3). Means with standard error (SE) are displayed for each time point.

Serum shock allowed the observation that the ultradian SOCS3 gene was rhythmically expressed and degraded in proliferating colorectal cancer cells (figure 2.11).

According to this experimental set up, every experimental time series, there were two biological replicates for every time point (figure 2.11 A). Inspecting visually the immunoblots, it is evident that there is a high degree of biological variation within each time course. For instance, comparing the biological time series during the initial time course, there is delay in the increase of SOCS3 protein expression observed in the first 80 minutes. SOCS3 protein in the first biological series is increased and reaches its peak between 30 to 40 minutes, following its decrease to its lowest peak in 60 minutes. However, according to the second biological time series, the dynamic expression of SOCS3 is increased and reaches its peak between 50 to 60 minutes and begins to decrease in 70 minutes. The same pattern of biological variation between time series is also observed during the other two time courses.

Averages of the biological replicates for each experimental run are presented in figure 2.13 B. The graph (figure 2.11 B) displays the dynamic activity of SOCS3 expression. It also shows that there is an experimental variation between the time points for each time series. During the first time series, the SOCS3 expression fluctuated with two peaks at 30-50 minutes and 90 minutes. During the second time course, SOCS3 protein activity was oscillated with peaks at 30 minutes and 90-100 minutes, whereas according

to the third time course, SOCS3 protein exhibited a dynamic pattern with a low peak at 50 minutes. Although, the experimental runs do not appear to replicate each other exactly, it is evident that they do share the same overall dynamic pattern. This could be explained by the oscillatory sequences initiating at slightly different time points. Indeed, if there was a shift in the dynamic pattern of the first run (run 1) with the oscillation beginning 10-20 minutes earlier (figure 2.11 B), SOCS3 protein of Run 1 will cycle with a similar period and amplitude to run 2. This suggests that the SOCS3 protein does have a reproducible oscillatory pattern, but this is hard to demonstrate with western blot analysis as each run has a different start point. Since it is not clear at this stage how the clock synchrony regulates SOCS3 expression, it is difficult to show the robustness of the oscillatory pattern through immunoblot analysis, as it is not feasible to control the time at which the oscillation begins and therefore different runs will initiate at different times making direct comparisons between different runs very difficult.

Additionally, the average for all the experimental replicates are shown in figure 2.11 C. Despite the timing issues discussed above, it does seem that there is a small degree of oscillation when the data is averaged for each time point (figure 2.11 C). Perhaps, if there is a method of making sure that SOCS3 protein activity begun its dynamic cycling at the same time point, this averaged data will display a clearer oscillation with a larger amplitude.

The data were analysed using a Gaussian Process regression with short-range and periodic correlation components. The data were fitted in three models by Dr James Hensman and presented in figure 2.12. The first graph represents a periodic model, the second model represents a smooth Model and the final is a noise model. Each model compromises a hierarchical structure to demonstrate the variance between the

experimental replicates. In the periodic model, 30% of the variance can be explained by the period trend, whereas 50% by between-experiment variance, and the remaining time points by noise.
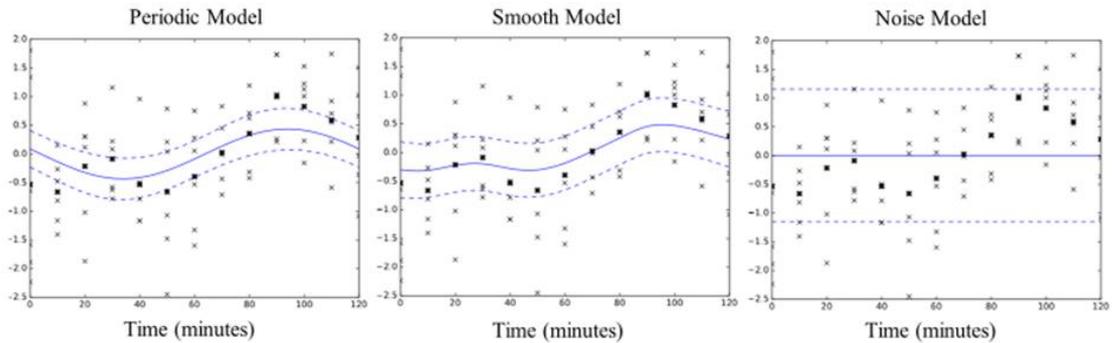
Figure 2.12: Average Expression SOSCS3 levels in Caco2 cells showed evidence of periodicity. Results were analyzed using the version 0.2 of the python Gaussian process toolbox Gpy. Each time point is represented with an 'x' symbol. The average for each time point is marked with a bold '**x**'. (pers. comm. Dr James Hensman 2016, 2017).

According to the noise model, which can be considered as the ''null'' model, the variance obtained by the data is assumed to be explained only by noise. The log likelihood ratios generated by the three models were calculated by Dr James Hensman and showed that the periodic model was preferred over the noise model. Dr James Hensman also demonstrated that the log likelihood for the periodic model in relation to the smooth model was calculated as 0.00626. The conclusion is that SOCS3 protein activity displayed a dynamic component according to his analysis, however it was difficult to differentiate the periodic model from the smooth model (0.00626 periodic vs. smooth).

Therefore, by achieving both cell cycle and clock synchrony, Caco-2 cells were able to exhibit a 2-hour oscillatory period of SOCS3 expression with peaks at 30 minutes and 90 minutes.

Nevertheless, this dynamic activity of SOCS3 protein is a resulting product of the JAK/STAT signalling process, where the phosphorylation of STAT3 (at Tyr-705) stimulates the expression of SOCS3, which in turn blocks JAK-dependent phosphorylation of STAT3, closing this negative feedback loop [61].

Therefore, it will be important to investigate if pSTAT3 protein also exhibits a dynamic behaviour in proliferating Caco2 cells. The immunoblots that were initially stained with SOCS3 antibody were also stained with pSTAT3 antibody to unravel its activity. The resulting bands for pSTAT3 were quantified using the Stain free gel analysis and normalized to a standard reference sample. The immunoblots for each experimental run are displayed in figure 2. 13.

Inspecting visually the immunoblots (figure 2.13 A), it is noticeable that the biological variation within each time course that was initially observed for SOCS3 activity, is also persisted for the expression for pSTAT3 protein.
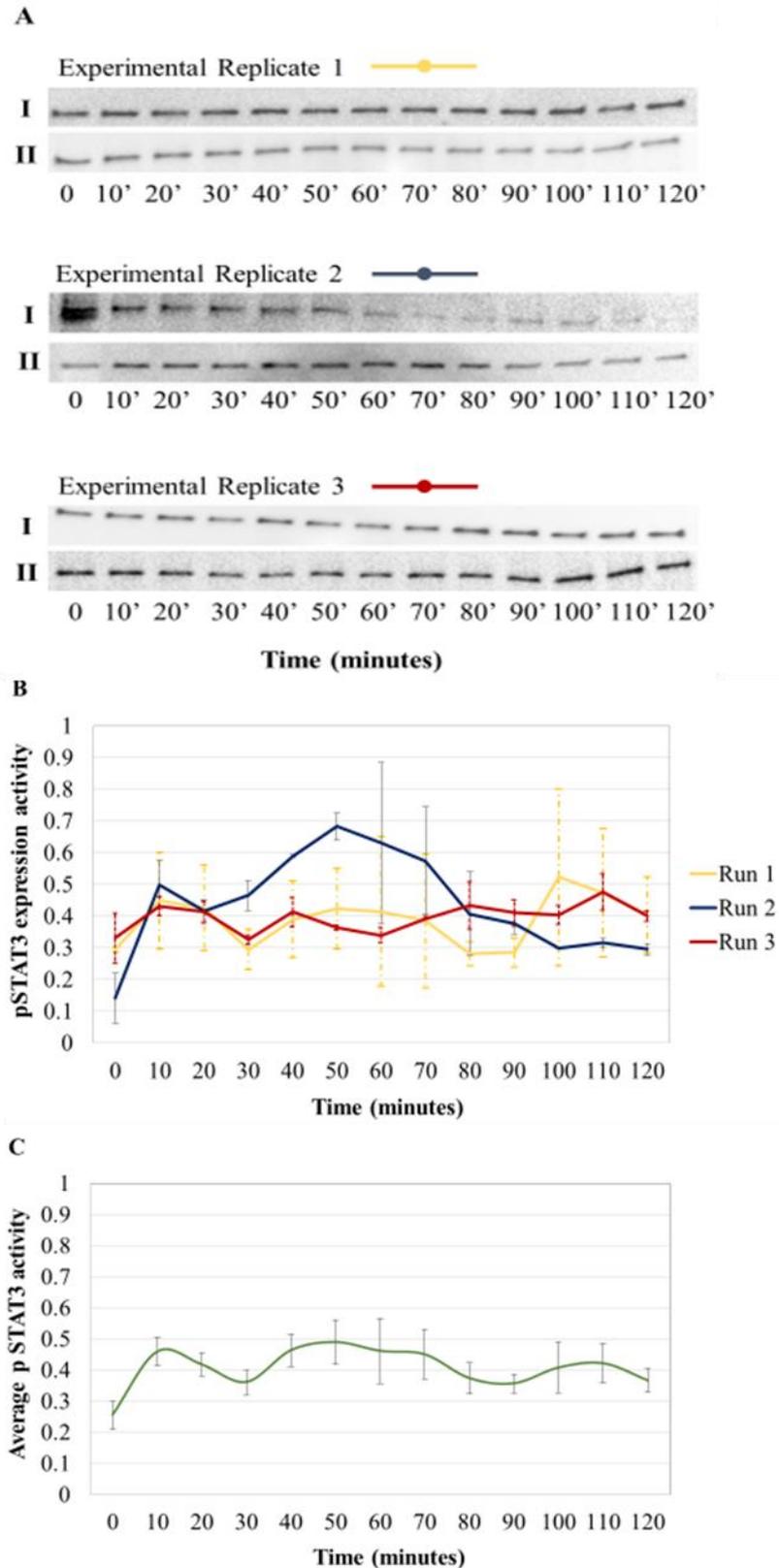
Figure 2.13: **Expression profiles of pSTAT3 protein in Caco2 cells analysed by Western Blotting.** **(A)** Western blot analysis of pSTAT3 protein levels (n=3). **(B)** Quantification of pSTAT3 protein for each experimental replicate based in units relative to peak intensity. **(C)**

Analysis of pSTAT3 dynamic activity (n=3). Means with standard error (SE) are displayed for each time point.

This observation is displayed clearly during the second time course, where the first biological time series illustrates the levels of pSTAT3 protein dropping from 50 minutes onwards, whereas the second biological time series shows a steady increase of pSTAT3 activity (figure 2.13 A and B). Averages of the biological replicates for each experimental run are presented in figure 2.13 B. The graph (figure 2.13 B) also illustrates that there is a high degree of variation in the dynamic activity of pSTAT3 activity between the different experimental replicates.

Additionally, the average for all the experimental replicates are shown in figure 2.13 C. Despite the high degree of variation between time courses, it does seem apparent that there is a small degree of oscillation when the data is averaged for each time point (figure 2.13 C). From the graph, it shows that the phosphorylation of STAT3 displays an oscillatory pattern with three minor fluctuations at the time points of 10 minutes, 50-60 minutes and 110 minutes. However, the data failed to produce peaks with values 1-1.5x fold higher than the lowest signal values during the time course, consequently the results can't be considered as a statistically significant. This is also supported by the high values of error bars presented for each time point (figure 2.13 C). Dr Christopher Jewell analysed these data to test if there was any level of periodicity for the phosphorylation of STAT3 activity. Results are presented in 2.14. The conclusion from the Gaussian Process regression analysis is that pSTAT3 protein expression did not show a dynamic component.
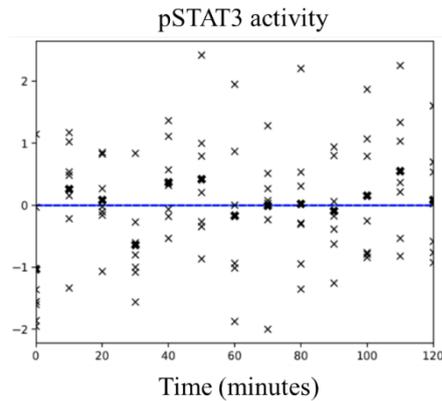
Figure 2.14: **Average Expression of pSTAT3 levels in Caco2 cells analysed with Gaussian process toolbox Gpy.** Results were analyzed using a Gaussian Process regression analysis with short range and periodic correlation components. Each time point is represented with an 'x' symbol. The average for each time point is marked with a bold 'x'. The temporally-correlated random effect *u(t)* is represented with a blue line. (pers. comm. Dr James Hensman and Dr Christopher Jewell).

The population of Caco2 cells under study did not appear to be homogenous in their proliferating state to elucidate a pSTAT3 dynamic behaviour that could mirror SOCS3 dynamic behaviour. This could be explained by the fact that the oscillatory sequences of SOCS3 protein initiated at slightly different time points.

Consequently, it seemed appropriate to also investigate if there is any dynamic behaviour of SOCS3 protein and the phosphorylation of STAT3 activity in a non-cancerous cell model. Proliferating Caco2 cells were seeded in 35 mm^2 petri dishes and allowed to differentiate for 14 days. On the 14th day, the culture media was removed,

and differentiated Caco2 cells were serum starved overnight with FBS-free basal medium.
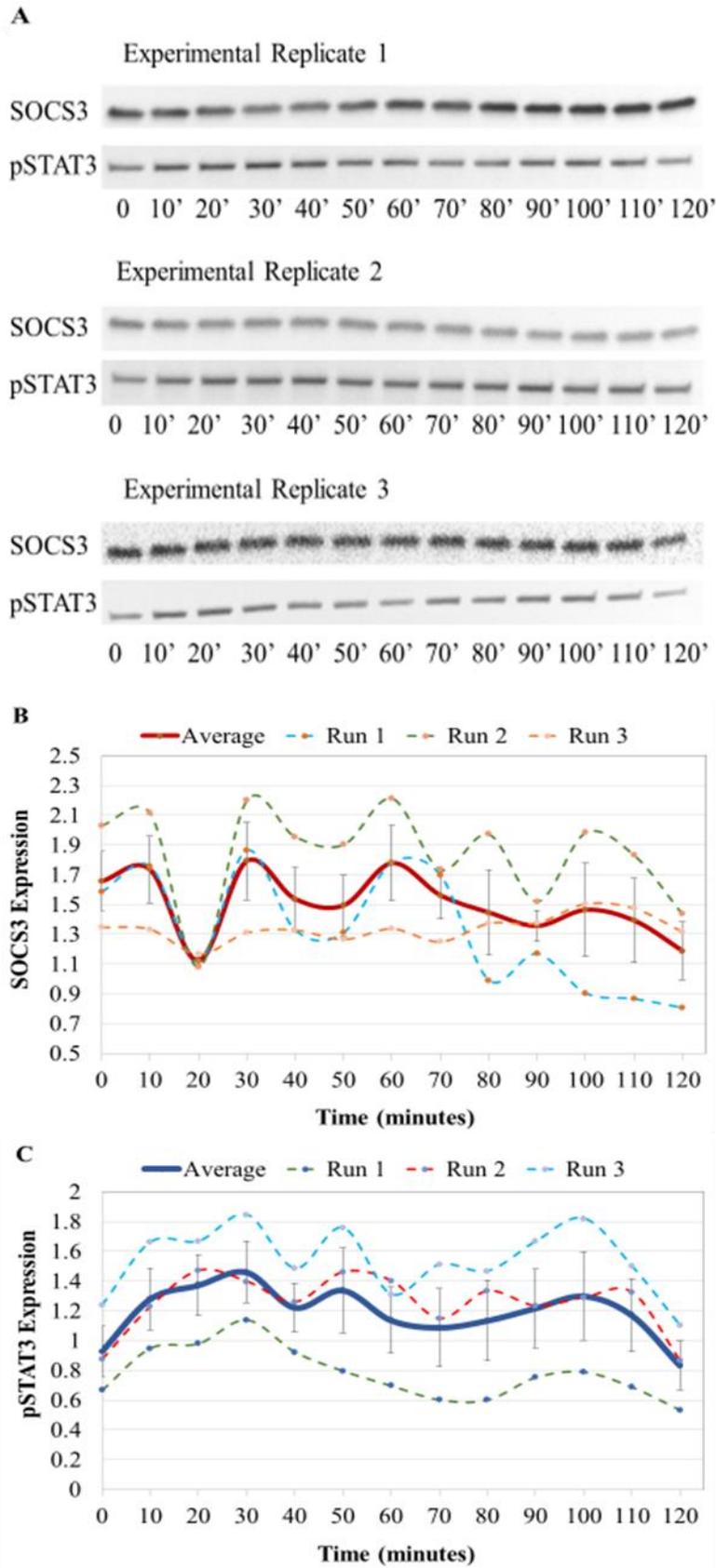
Figure 2.15: Cell cycle and Clock synchrony in differentiated Caco2 cells.

(A) Expression profiles of SOCS3 and pSTAT3 proteins in differentiated Caco2 cells were analysed by Western Blotting (n=3). **(B)** Quantification and normalization of SOCS3 protein for each experimental run based in units relative to peak intensity. Data was graphed against time Average line for n=3 is also displayed (thick red line). **(C)** Quantification and normalization of pSTAT3 activity for each experimental replicate based in units relative to peak intensity. Average line for n=3 is also displayed (thick blue line). Means with standard error (SE) are displayed for each time point.

Cells were then serum shocked with 50% v/v horse serum and incubated for 2 hours. The first-time point was taken after the serum shock (t=0) and then the cells were harvested every 10 minutes for a period of 2 hours. The extracted proteins for each time point was assessed with western blot assay and displayed in figure 2.14.

The immunoblot images from the western blotting experiment (figure 2.14 A) for SOCS3 protein did not display clear fluctuations. According to the ChemiDoc analysis, the quantification and normalization of SOCS3 illustrated numerous minor fluctuations during the time course (figure 2.14B).

However, pSTAT3 activity showed an increase during the first 20-40 minutes of the time course before its levels got dropped and then increased for the rest of the time course. This was also evident from the graph that displayed the average activity of pSTAT3 in figure 2.14 C.

The data were also fitted in a statistical modelling system by Dr James Hensman and presented in figure 2.15.

Dr James Hensman fitted the data in a periodic model and he concluded that the relative expression of pSTAT3 protein in differentiated Caco-2 exhibited a level of periodicity.

However, the combined experimental replicates of SOCS3 protein did not generate a distinctive oscillatory pattern.
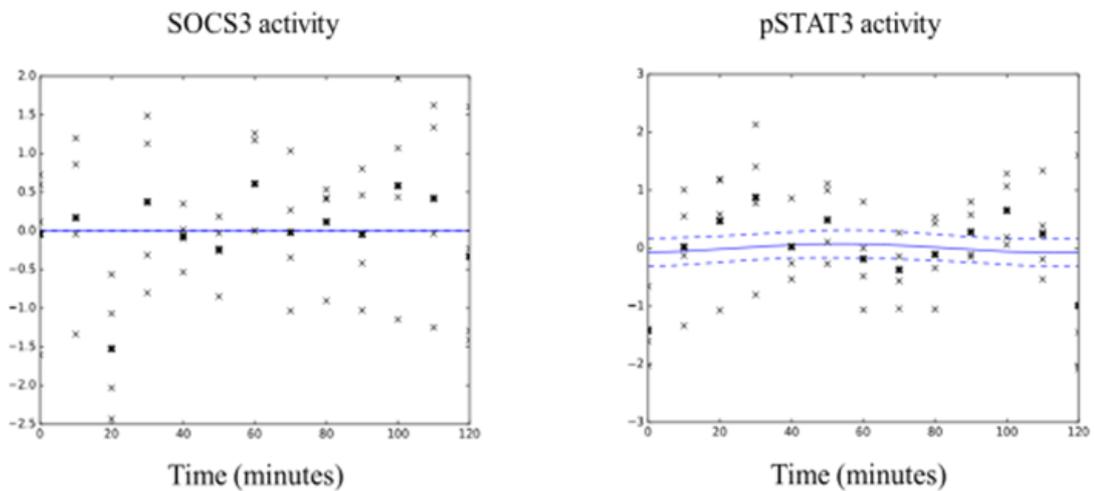
Figure 2.16: **Average Expression of SOCS3 and pSTAT3 levels in differentiated Caco2 cells analysed with Gaussian process toolbox Gpy**. Results were analyzed using the version 0.2 of the python Gaussian process toolbox Gpy. Each time point is represented with an 'x' symbol. The average for each time point is marked with a bold 'x'. (pers. comm. Dr James Hensman 2016, 2017).

Despite evidence that synchrony is occurring, noise from a whole-cell population detract from the oscillatory signal. Notably, the observation that proliferating Caco2 cells were able to illustrate a dynamic pattern of SOCS3 protein but fully differentiated Caco-2 cells did not display a SOCS3 oscillatory behaviour could be explained by the fact that differentiated cells and nondividing cells are presented as different from dividing and growing cells due to their particular biochemical differentiation characteristics and their inability to divide [187].

Therefore, it will be important to investigate if there is a dynamic behaviour of SOCS3 protein and pSTAT3 activity in a non-cancerous cell line at its proliferating state.

Primary Human Intestinal Epithelial Cells (HIECs) were isolated from foetal small intestine with typical crypt cell proliferative characteristics. HIECS represent a cell line model that has the ability to generate continuously growing human intestinal epithelial cell cultures, and simultaneously retain the ability to express a number of normal functional intestinal cell markers over a number of passages [200]. To this end, this report took advantage of this cell line and investigate the existence of oscillatory behaviour of the negative feedback loop between SOCS3 and pSTAT3. Results for SOCS3 protein is presented in figure 2.16.
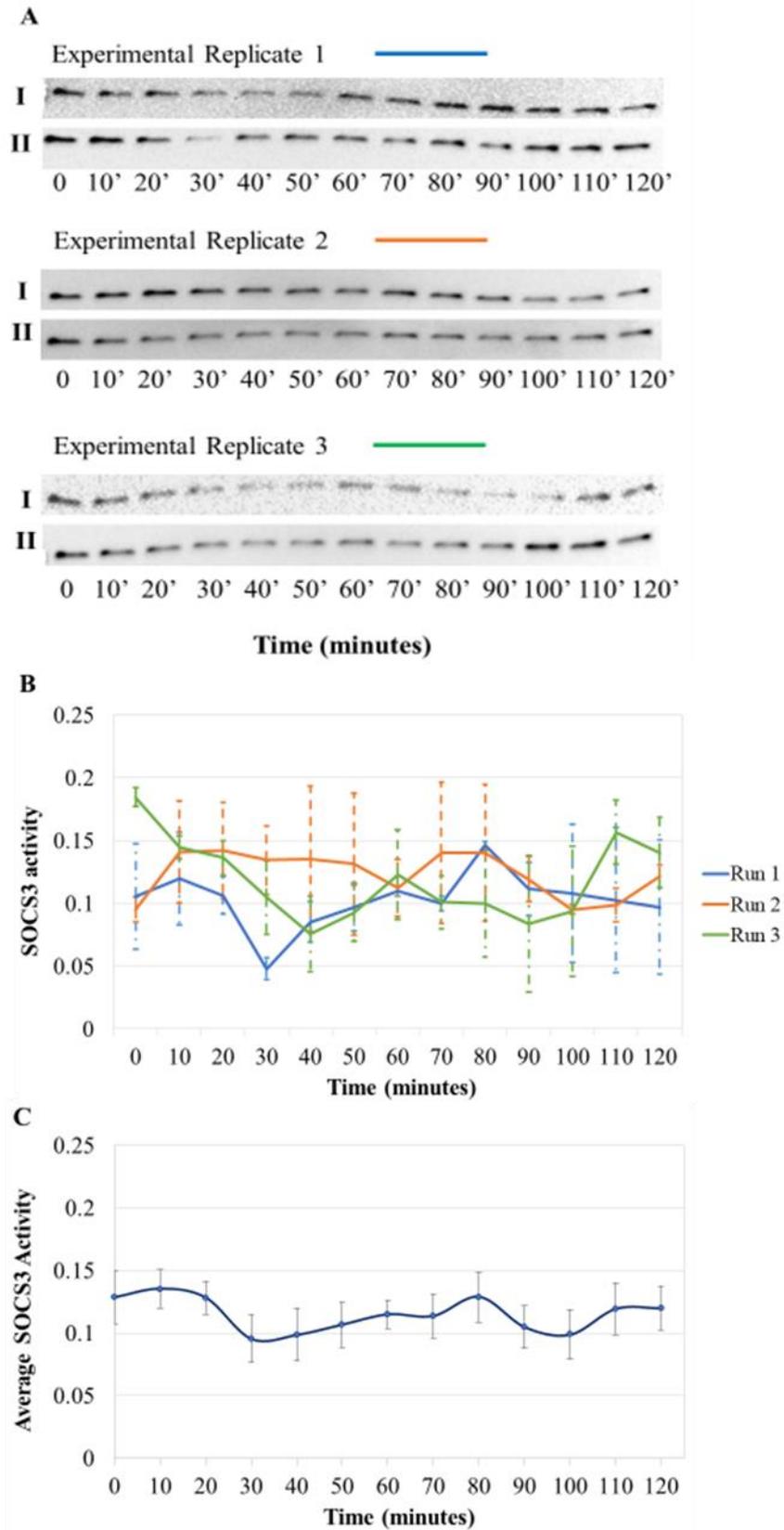
Figure 2.17: SOCS3 activity after cell cycle and clock synchrony in HIECs. (A) Western

blot analysis of SOCS3 protein levels (n=3). **(B)** Quantification of SOCS3 protein for each

experimental replicate plotted against time. **(C)** Average response of SOCS3 dynamic activity (n=3). Means with standard error (SE) are displayed for each time point.

The use of Western blotting to explore the dynamic profile of JAK/STAT pathway has proven noisy. To tackle this issue, every experimental time course was compromised two biological replicates for every time point (figure 2.16 A). Inspecting visually the immunoblots, it is apparent that each time course provides a different dynamic behaviour of SOCS3 protein.

Furthermore, the report also obtained average values from each individual time course and presented in figure 2.16 B. This dynamic variation between time courses is also supported by the graph (figure 2.16 B). The dynamic variation for each biological replicate within a time course is also evident by the big values of error bars.

Despite this variation and noise of the system, it is important to note that SOCS3 protein shows a high degree of oscillation in proliferating HIECs compared to the SOCS3 expression profiles in differentiated Caco2 cells. This supports the hypothesis that cells in their proliferative state display a more dynamic behaviour in the regulation of protein turn over than nondividing cells. However, as it is depicted in the results from figure 2.16, the increased rate of oscillatory behaviour might also lead to a noisier system, making direct comparisons between different runs very challenging.

Regardless the high degree of SOCS3 protein oscillations in each time course, this dynamic behaviour is abolished when the data is averaged for each time point, as shown in figure 2.16 C. Additionally, the averaged data failed to retain significant fluctuations in the levels of SOCS3 during the time course, consequently the results can't be considered as a statistically significant.

Dr James Hensman also fitted the data in a periodic model and he concluded that the relative expression of SOCS3 protein in HIECS did not generate any level of periodicity (figure 2.18). This could be due to the fact that SOCS3 protein failed to show a consistent oscillatory behaviour between the experimental replicates.
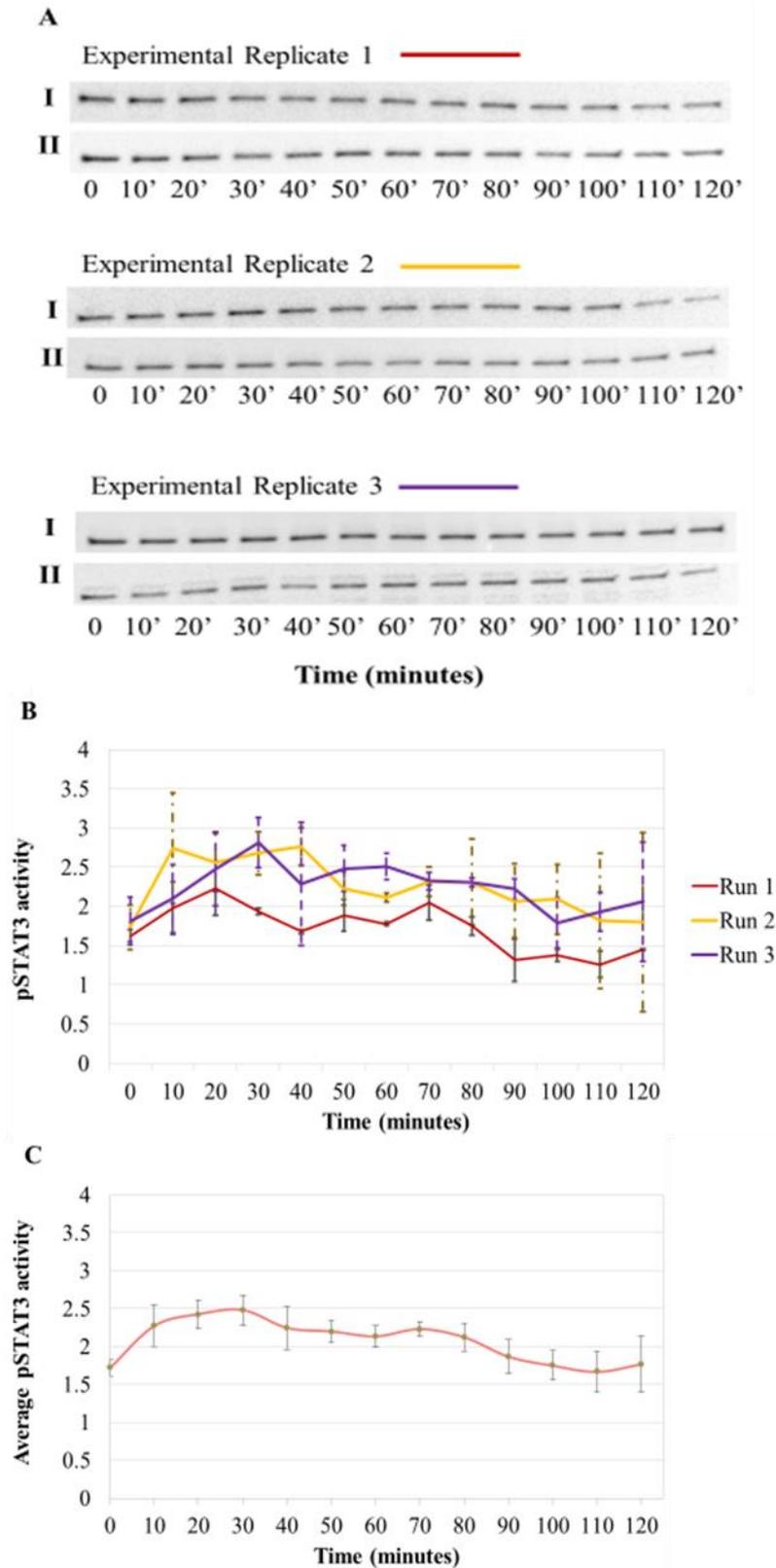
Figure 2.18: pSTAT3 activity after cell cycle and clock synchrony in HIECs. (A) Western

blot analysis of pSTAT3 protein levels (n=3). **(B)** Quantification of pSTAT3 protein for each

experimental replicate plotted against time. **(C)** Average response of pSTAT3 dynamic activity (n=3). Means with standard error (SE) are displayed for each time point.

The report also explored the abundance of pSTAT3 protein in HIECs. The immunoblots that were initially stained with SOCS3 antibody were also stained with pSTAT3 antibody to unravel its activity. The resulting bands for pSTAT3 were quantified and analysed (described in Methods). The immunoblots for each experimental run are displayed in figure 2.17 A. The biological variation in every time course was also found for pSTAT3 protein as depicted in figure 2.17 A. Furthermore, the obtained average values from each individual time course in figure 2.17 B show that pSTAT3 levels in Run 2 and Run 3 follow a similar pattern where it is increased up to 30-40 minutes, following its steady decrease. On the other hand, pSTAT3 protein in Run 1 only display minor fluctuations during the 2 hour time period.

Average response of pSTAT3 dynamic activity (n=3) is presented in figure 2.17 C. Averaging the response of pSTAT3 in all three time courses demonstrates an increase of pSTAT3 up to 30 minutes, following its steady decline.

Dr James Hensman's analysis on pSTAT3 protein behaviour demonstrated that expression profiles of pSTAT3 protein in HIECs exhibited a strong evidence of periodicity. However, it was not feasible to make any direct or statistical comparison of pSTAT3 dynamic behaviour with the expression of SOCS3 protein.
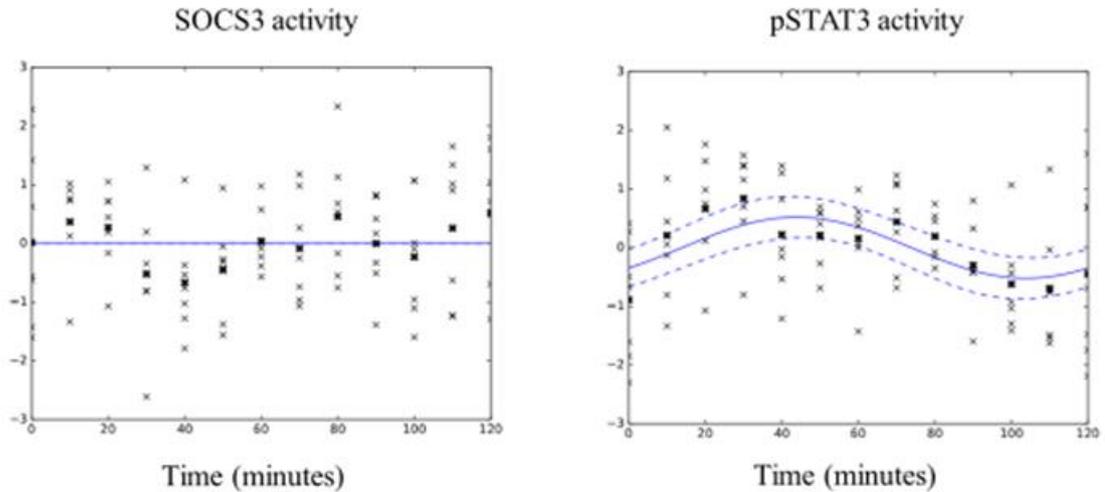
Figure 2.19: **Average Expression of SOCS3 and pSTAT3 levels in HIECs analysed with Gaussian process toolbox Gpy.** Results were analyzed using the version 0.2 of the python Gaussian process toolbox Gpy. Each time point is represented with an 'x' symbol. The average for each time point is marked with a bold '**x**'. (pers. comm. Dr James Hensman 2016, 2017).

The quantitative measurement of dynamic behaviours in biological systems requires whole population synchrony. Serum starvation and subsequent shock is able to achieve this to some extent. However, the average response in a population of cells proved to be 'noisy' in the vast majority of time courses, leading to the development of other sensitive quantitative techniques that could minimize the variation between experimental replications.

2.4 Discussion

Cellular behaviour in a complex environment requires to continuously detect, process and appropriately respond to a vast array of changing stimuli. Consequently, the whole cellular system processing is dynamic in nature. The temporal coordination of key proteins within dynamic networks is important for the precise regulation of cellular activity and function. Physiological processes in living cells display a wide series of dynamic patterns. Oscillations with an ultradian time period of 1 to several hours are ubiquitous in biology and are found to be involved in diverse key processes such as developmental pathways, cell proliferation and division, DNA damage responses and immune responses. These ultradian oscillation within cells belong to well-conserved molecular networks that are found in various species [129].

There are several factors that influence the temporal oscillations of a particular biological system, including multiple feedback loops, time delays and noise, which results from processes of a random nature, e.g. expression and turnover of gene products and interactions that involve collision of diffusible molecules [201].

A large proportion of *in vitro* studies have concentrated on investigating the oscillatory dynamics of the expression or activity of three transcription factors (Hes1, p53 and NF-κB) that all belong to short negative feedback loops. In fact, transient stimulation of the cells induced oscillatory gene expression with a duration of 2 to 3 hours [59, 136, 202, 203]. Detailed analysis of these dynamic behaviours was established with quantitative measurement on the dynamics of these key proteins under different conditions and perturbations on the normal system structure and investigate how these changes affect the oscillatory pattern [140]. This can be achieved by using different experimental cell-population techniques such as Western Blot and RT-PCR as well as single-cell

techniques including immunofluorescence, flow cytometry and fluorescence microscopy [204].

The quantitative measurement of dynamic behaviours in the vast majority of biological systems requires analysis at the single-cell level, as measurements of average response to the population of cells can be misleading [139].

To tackle this problem, the experimental strategy of this report attempted to serum starve proliferating and fully differentiated Caco-2 cells in the G1 or G0 phase of the cell cycle to produce a synchronized culture. It is well known that the serum-deprived cells are synchronized at a point in the G1 phase or in the G0 phase and when a stimulus is applied, the stimulated cells respond as a synchronized cohort through the cell cycle [187]. Serum stimulation has been commonly used to study signal transduction pathways in several biological systems including cell cycle and growth, neuronal excitation and immune response. There are several research articles that have used this mode of cell population synchronization in intestinal Caco-2 cell line for their experimental designs. In most experiments, confluent monolayers of Caco-2 cells were synchronized by serum deprivation for overnight, 1 or even 2 days [65, 118, 188, 189].

This report attempted to find the optimal duration of serum-starvation of the human colon adenocarcinoma Caco-2 to achieve the highest degree of cell culture synchronization in order to establish a distinct oscillatory pattern of SOCS3. However, when proliferating Caco2 cells were serum starved 24 hours, they did not establish distinct dynamic pulses for SOCS3. Moreover, when fully differentiated Caco2 cells were serum starved overnight (16 hours), they also exhibited a non-significant minor fluctuation of SOCS3 protein. Furthermore, when proliferating Caco2 cells were

incubated with serum free medium for 48 hours to achieve a higher degree of cell synchronization, SOCS3 protein demonstrated a reduction of its protein abundance in relation to t=0 and it was slowly increased after serum stimulation. This could be explained by the fact that upon serum starvation of 48 hours, autophagy is highly induced in response to a prolonged nutrient deprivation, decreasing SOCS3 protein levels. Consequently, this could also perhaps abolish any dynamic behaviour since cells are trying to restore SOCS3 levels back to normal. Moreover, when fully differentiated Caco-2 cells were starved for 36 hours, they did not establish any clear dynamic pulses for SOCS3 protein. Notably, in comparison to 32 hours starvation period, the reduced duration of starvation for 16 hours displayed at least a low degree of protein oscillatory behaviour. This is supported by the fact that differentiated cells and nondividing cells are presented as different from dividing and growing cells due to their particular biochemical differentiation characteristics and their inability to divide [187].

The analysis of cellular dynamics at the cell population level with Western blotting demonstrated that p53 and Mdm2 undergo damped oscillation after strong DNA damage [136]. Unsustained pulses in TNF-α stimulated cell populations were also established in the NF-κB pathway [205]. Typically, in these biological systems, the oscillatory pattern established in immunoblots is explained by the level of stimulus applied as the stronger the stimulus, the stronger the amplitude of the cell population response [206]. In support of this notion, the absence of an appropriate stimuli that could lead to the cycling phenomenon of SOCS3 could perhaps explain why the results obtained from serum starvation experiments demonstrated a non-oscillatory pattern of SOCS3 in both proliferating and fully differentiated Caco-2 cells.

Despite cell cycle synchrony, various proteins belonging to different molecular pathways are found to be independent from the cell cycle status. It has been further reported that peripheral circadian pulses are thought to impact on cellular ultradian rhythms and these can be entrained by serum shock [181]. Thus, to further characterize this dynamic network, this report utilized a serum shock methodology to synchronize the cell population and elicit the rhythmicity of SOCS3 protein abundance and investigate its functionality.

Serum starvation and subsequent shock was able to achieve this to some extent. Specifically, Caco2 in their proliferative state exhibited an oscillatory behaviour of SOCS3 protein with peaks at 30 minutes and 100 minutes. This was also examined with statistical modelling and showed that there is a dynamic behaviour of SOCS3 that cannot be explained by the noise of the system. These results demonstrate that there is a relatively fast SOCS3 protein oscillation in Caco2 cells. As the negative regulatory effect of SOCS3 protein has been characterised as transient with a short biological half-life, we were expecting to detect a periodic protein expression within the 2-hour time period [103, 107]. However, as the majority of research articles have reported a half-life ranging from 36 minutes to 120 minutes, we were expecting to detect a single peak within the 2-hour time series [102, 207, 208]. In support of this notion, Yoshiura et al. [61] reported the first peak of SOCS3 protein oscillation at 2 hours in their time series. However, it is important to mention that this inconsistency with the previously reported SOCS3 half-lives can be attributed to the specific cell line's preferred method of SOCS3 clearance from the cell. Consequently, it would be of interest to measure SOCS3 half-life by treating Caco2 cells with cycloheximide [209] and perform a time series. The results of this experiment could give us the chance to experimentally determine the half-

life of SOCS3 protein in Caco2 cells and verify the fast SOCS3 protein fluctuations observed in Caco2 cells.

In addition, pSTAT3 levels in proliferating Caco2 cells also illustrated a dynamic pattern with more frequent oscillatory proteins peaks at 10 minutes, 50 minutes and 110 minutes. The fast cycling nature of pSTAT3 protein has also been described in the literature. Specifically, previous articles have demonstrated that transient STAT3 activation and its rapid return to the basal state is crucial to maintain cell homeostasis [210]. Furthermore, Yoshiura et al. [61] also demonstrated a faster pSTAT3 protein oscillation compared to SOCS3 protein oscillation with peaks at 1,3 and 5 hours.

However, it is evident that Caco2 cells started to lose the ability to behave as a synchronized cohort after the 1 hours of serum shock and this could perhaps explain the long amplitude of the second peak. Interestingly, according to the results from the western blotting and the quantitative analysis of the protein bands, the dynamic protein peaks of SOCS3 always followed after the protein peaks of pSTAT3.

Differentiated Caco2 cells did not exhibit a consistent dynamic behaviour of SOCS3 protein and it is difficult to draw any certain conclusions of the dynamic protein changes by inspecting visually the immunoblots. Furthermore, the statistical analysis of SOCS3 in differentiated Caco2 also did not identify any oscillatory protein sequence that was not explained by the noise of the system. However, pSTAT3 proteins were subjected to a dynamic cycling over the period of 2 hours. It was not as highly dynamic as seen in proliferating Caco2 cells, but the statistical analysis proved that this protein cycling pattern could be explained above the experimental noise.

On the other hand, the most dynamic cycling of SOCS3 protein was observed in HIECS. However, the system proved to be very noisy and results drawn from each individual

time course showed high variations when compared to one another. This was also supported by the statistical modelling system of Dr James Hensman. Despite this variation, pSTAT3 protein was proven statistically to behave in a dynamic fashion.

It has become increasingly clear that JAK/STAT pathway is a high dynamic complex system that exhibits various dynamic patterns in response to different stimuli. However, there are also several factors that could influence the temporal oscillations of this particular biological system, such as the experimental design and the limitations of the method detecting and quantifying these protein changes. All these factors will be discussed in the next following pages in detail.

2.4.1 Experimental design of horse serum shock

Mammalian cell cultures have the ability to demonstrate periodicity in protein expression. This was first observed when high levels of serum induced Clock-controlled gene expression in RAT-1 fibroblasts and H35 Hepatoma Cells [9]. However, the mechanism by which serum affects dynamic protein expression is still not fully understood. Notably, we were only successful in demonstrating SOCS3 and pSTAT3 protein oscillations when we treated cells with a high dose of horse serum. Furthermore, Balsalobre et al. [181] demonstrated that serum from different species were also able to efficiently elicit a circadian oscillation. Therefore, further work is needed to test whether serum from different species (i.e. rabbit, cow, big rat) would generate the same dynamic expression. This work could clarify whether the nature of a rich medium generated the observed SOCS3 and pSTAT3 protein oscillation.

Secondly, in our serum shock experimental set up, cells were initially serum starved prior to their treatment with horse serum. Conceivably, this experimental set up was not

able to specify whether the dynamic protein expression was generated from either a synchronization of already existing cycles in desynchronized cells, or an induction of oscillations in arrhythmic cells. Therefore, it would be interesting to test whether the observed protein oscillation could only be generated as a result of the serum stimulation of growth-arrested cells.

Thirdly, since the dynamic behaviour of SOCS3 and pSTAT3 occurred immediately after the treatment of the horse serum-rich media, we wonder whether this would suffice to trigger this oscillatory expression without following the addition of complete medium (medium containing 10% FBS). Therefore, the experiment should be repeated by replacing the serum-rich medium after the first 2-hour period with serum-free medium. Moreover, it would also be interesting to examine where this dynamic protein expression also occurs immediately after the addition of the serum-rich media. These results could demonstrate whether the shock with only the serum-rich medium is efficient to stimulate an ultradian oscillation and only the serum factors included in the horse serum are essential for the observed dynamic behaviour.

Lastly, animal serum is composed of a complex mixture of a large number and variety of components. When serum is used with cell culture media, it serves as a source for amino acids, proteins, vitamins (especially fat-soluble vitamins including A, D, E and K), carbohydrates, lipids, glucocorticoid hormones, growth factors, minerals, and trace elements. These components have shown to be involved in a vast array of signalling pathways [211]. Consequently, it would be difficult to investigate the role of each component to the oscillatory behaviour of SOCS3 and pSTAT3 protein. However, since horse serum shock has been mainly used in circadian research, several of these studies have investigated which signalling pathways could probably activate the expression of

circadian clock-controlled genes. So far, these have involved forskolin, an activator of adenylate cyclase, and dibutyryl cAMP; photnol-12-mytistate-13-acetate (PMA), an activator of protein kinase C (PKC); calcimycin (A23187), a calcium ionophore; and epidermal growth factor (EGF), insulin and fibroblast growth factor (FGF), three inducers of receptor kinases (RTKs) [212]. Notably, many of these components are also known to be involved in the induction of JAK/STAT pathway, such as the inducer of receptor kinases and activator of protein kinase C [213]. Therefore, it would be interesting to test if shocking the cells for a short period of time with any of these components/signalling factors individually could induce the same dynamic behaviour of SOCS3 or pSTAT3. In our experiments, we generated a large concentration difference in these growth factors and nutrients simply by increasing the serum concentration. Conceivably, it may be the difference in the levels of these signalling factors, rather than their absolute concentration, that synchronizes ultradian oscillations in JAK/STAT pathway. Collectively, further analysis is needed to determine whether the experimental set of serum shock or the content of horse serum is causing the induction of SOCS3 and pSTAT3 dynamic behaviour.

2.4.2 Limitations of Western Blotting methodology

Quantitative Western blotting has been widely used in life science labs to semi- quantify the relative abundance of target proteins. However, there are many potential stumbling blocks in this multistep method that can produce unreliable results. These may include sample preparation, normalization, SDS-PAGE gel loading, protein transfer, blocking buffers, primary and secondary antibody selection, incubations, washes and quantification of densitometric data [214]. The traditional and most frequently used method to correct for possible artifacts emerging from pipetting inaccuracy, imprecise

protein estimation, or uneven transfer of proteins, a loading control is employed. Loading control methods involve the reprobing of membranes with an antibody against a target protein that exhibit high-level, constitutive expression in the cell type examined or staining the membrane with a total protein stain. The most commonly used loading control is the housekeeping protein β-actin [184].

Importantly, during the experimental design, the selection of a protein to act as the loading control needs to be ensured that it does not alter as a consequence of the experimental variable under test and does not saturate the detection system. For instance, β-actin is shown to play a role in several intracellular processes and can be altered by different processes such as changes in cell cycle and during differentiation of neuronal cells [215, 216]. Furthermore, several reports have demonstrated the clinical implications of β-actin in cancer development due to its differential expression in cancer. This is evident by a study that analysed β-actin expression and distribution in normal and tumour tissue samples of gastric adenocarcinoma patients using

immunohistochemistry. The results from this study demonstrated that β-actin exhibited a higher expression predominantly contributed by inflammatory or tumour immune cells of the tissue microenvironment and associates with tumour grade [217]. This study showed consistent β-actin in fully differentiated but not in proliferating Caco-2 cells during the time courses were cells were subjected to serum starvation and subsequently to medium supplemented with 10% FBS.

To tackle this issue, this report employed the stain-free total protein staining methodology for quantifying the data from serum shock time courses. According to this method, membranes can be imaged for Stain-free staining and total protein abundance can be quantified for each lane. These values obtained from total protein quantification

can be used then to normalise the obtained results from the protein of interest. This method could perhaps demonstrate a more representative process to quantify and obtain real values of the protein of the interest. Furthermore, a same standard control protein was always loaded in every western blotting gel. A standard control was used to make sure that every time point for every time course was normalised equally to the same protein band. This was employed to eliminate or reduce any intrinsic heterogeneity that could be caused by all the pitfalls of every step of the western blot technique. Taking together all these optimization settings for the serum shock experiments, there was an improvement in visualising and analysing the dynamic behaviour of SOCS3 and pSTAT3 protein. However, the biological variation between experimental replicates make it difficult to draw a conclusive observation of the behaviour of these proteins. Possible these variations could have been further studies with the use of other quantifying techniques. For instance, it would be interesting to investigate if serum shock also leads to a strong transcriptional activation of both SOCS3 and STAT3 mRNA. Specifically, if there was also a high degree of variation between experimental replications in SOCS3 mRNA and it did not follow any logical biological sequence to the results obtained with the western blotting, it could point out that the complexity of JAK/STAT system might cause this high degree of variation.

2.4.3 Complexity of JAK-STAT pathway

In healthy tissues, signalling via the JAK/STAT pathway is tightly monitored at several levels within the cell. One important controlling module includes a long (and growing) group of phosphatases, such as SHP1, SHP2, CD45, PTPRD, PTPRT, TCPTP, PTP1B and DUSP2 that dephosphorylate tyrosine residues on JAKs, STATs and upstream receptors and thereby limiting downstream signalling [218, 219]. A second regulating

module involves members of the PIAS (protein inhibitor of activated STAT) family. These proteins negatively regulate STAT signalling through a variety of mechanisms that typically involve interfering with the DNA-binding activity of STATs or by a SUMO (small ubiquitin-like modifier) E3 ligase that supresses STAT signalling by promoting proteolysis of STATs [220]. A third prominent module includes members of the SOCS family. These proteins behave as negative feedback circuits in which each is transcriptionally activated by the STAT they are intended to supress. However, there is also promiscuity as one STAT can activate multiple SOCS and, in turn, a single SOCS can control several STATs [77]. Furthermore, similarly to STATs, SOCS members comprise SH2 domains that enable receptor binding and competing with STATs for binding as well as interfering with the activity of JAK kinases [92]. Accumulative research evidence has shown that regulating the intensity and/or duration of JAK activity, SOCS proteins dictate both quantitative and qualitative aspects of cytokine signalling. For instance, IL-6 switches from a STAT3-driven transcriptional signature to a STAT1-driven transcriptional signature in the lack of SOCS3 [221, 222].

The human intestinal Caco-2 cell line is known to exhibit high basal SOCS3 levels [223]. A recent study conducted by Koay et al. [65] demonstrated that upregulation of autophagy is associated with a mechanism contributing to cyclic expression of proteins, including SOCS3. More specifically, the finding that cannabinoids decreased SOCS3 abundance, partly associated with autophagic degradation of SOCS3 protein, represents a possible mechanism by which cyclic SOCS3 expression could be maintained. Koay et al. [65] showed that with the administration of cannabinoids in differentiated intestinal Caco-2 cell line induced autophagy that led to the downregulation of SOCS3, which was reversed by blocking early and late autophagy. However, the data from this

report failed to establish a correlation between the oscillatory patterns of LC3 protein with SOCS3 due to failure of optimizing the system.

Collectively, all these research studies highlight the flexibility of intrinsic STAT availability within cells to alter intrinsic immune cellular responsiveness to specific cytokines and how relative concentrations of different combinations of STAT proteins can be dynamical controlled to alter the behaviour of distinct immune cell subsets to cytokine exposure. This flexibility allows dynamic differential control of specific STATs expression to utilize limited genetic material for shaping complex immune responses [113, 224].

2.4.4 Whole cell population versus single cell analysis

Population-averaged studies are essential tools in biology, enabling the identification of key components and complex component interactions spanning signal transductions, transcriptional regulations, and metabolism. Such quantitative read outs of bulk populations can succinctly capture population state, and readily report how these states alter under different conditions and perturbations. Although informative and essential, the basis of population studies mainly relies on the assumption that the assay averages represent the dominant biological mechanism operating within an entire population [225].

However, all these population studies face a common limitation: as such, a mixture of cells under investigation are considered homogenous for only some parameters, including cell size or shape, tissue localization, protein expression, cell cycle or cell type subset. However, it has become evident that for any single cell response, there are

diverse dynamics that could give rise to the same underlying outcome of gene expression (figure 2.21) [226].
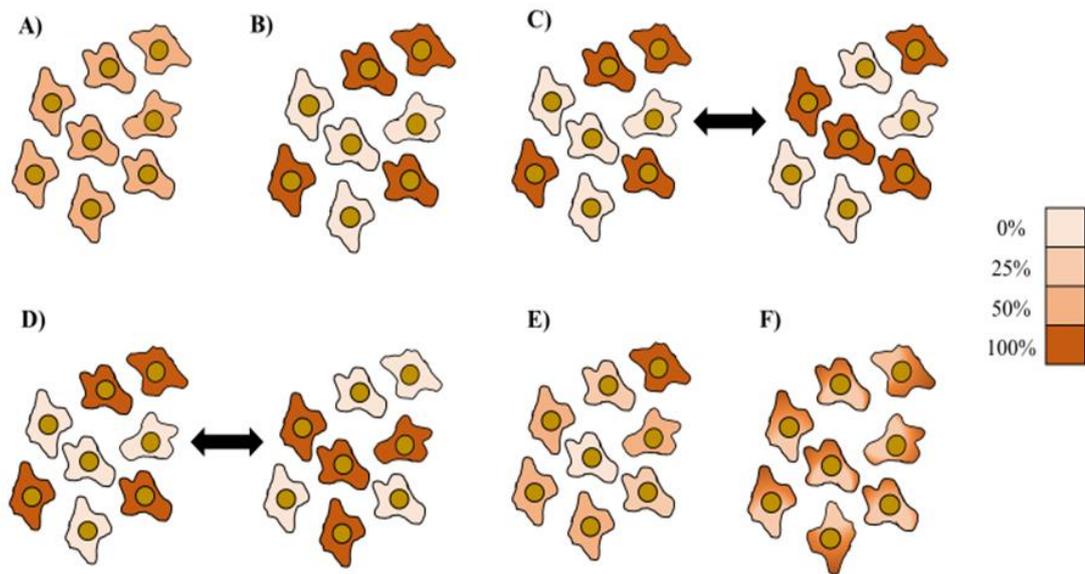
Figure 2.20: Different single cell measurements that give rise to the same average response in the cell population. Individual cell's signaling activity is indicated by its color. The color range is shown in a gradient bar that displays the spectrum between maximum (100%) and minimum (0%) color value representing cell's high and low signaling activity respectively. Every panel **(A-F)** demonstrates how individual cells with varying signaling activity can display the same cell population outcome. The population average of signaling activity for every example is 50%. **A)** every cell signals at 50%. **B)** half of the cells exhibit the highest (100%) signaling activity whereas the rest exhibit 0% activity. **C)** During the third example, the initial state of cells is as illustrated in the example **(B)** but single cells may change states in time without influencing the population average. This switching may arise stochastically in individual cells **(C)** or emerge from oscillatory signaling activity with varying frequency and phase in individual cells **(D). E)** According to the fifth example, cells' signaling activity can be dispersed between 0% and 100%, resulting in a population average of 50% signal activity. **F)** signaling activity within an individual cell is spatially

heterogeneous causing the average expression within a single cell to be 50%. Adapted from

[135].

The mechanisms controlling gene expression introduce remarkable cell-to-cell heterogeneity within genetically identical cells. This heterogeneity can arise, in part, from intrinsic fluctuations (due to cell cycle or transcription bursting) and extrinsic variations (due to exposure to stimuli or interactions with neighbouring cells) [227, 228]. Specifically, factors that contribute to extrinsic noise involve differences in the concentrations, states and locations of several molecules, such as regulatory proteins and RNA polymerases, involved in gene expression [229]. On the other hand, intrinsic noise usually emerges from infrequent synthesis of mRNAs which in turn lead to bursts of protein expression. This intrinsic stochasticity can give rise to substantial randomness in terms of the order and timing of events, even in a (hypothetical) population of genetically identical cells that share identical concentrations of cellular components [230, 231].

A second source for cell-to-cell heterogeneity can arise from cellular response variability evolved from stochastic differences in post-translational alterations as well as in the concentrations, activities and localization of different states of proteins involved in signalling pathways within cells [135].

Finally, these sources of heterogeneity highlight the difficulty in unravelling important dynamic behaviours and mask true dynamic responses in population-averaged analysis. Research reports with single-cell analysis are therefore important because immunoblots have limited resolution as they demonstrate averages of cell populations [206].

Therefore, to further characterize and study this dynamic system, the report needs to implement a single cell approach based on a fluorescent reporter and time-lapse confocal microscopy.

2.5 Conclusion and Future Steps

The JAK-STAT signalling pathway and its regulators do not only play an important role in physiological processes involved in growth control, cell differentiation and maintenance of tissue homeostasis, but they are also found to be involved in carcinogenesis. Various studies have demonstrated that distinctive perturbations of JAK-STAT pathway in tumour cells involve constitutive activation of JAKs by mutations, increased expression of STAT proteins (such as STAT3), and decreased levels of negative regulators of JAK-STAT pathway. System biology has already been employed to study the dynamic behaviour of JAK-STAT pathway under various physiological settings and is increasingly used to the investigate the pathway in cancer and healthy model states. The use of system biology approaches may assist to analyse the efficiency or inefficiency of antitumor agents applied to JAK-STAT pathway at the cellular level [232].

This study attempted to demonstrate the presence of SOCS3 and pSTAT3 oscillatory behaviour in the Human Intestinal epithelial cells (HIECS) as well as in the human colon carcinoma cell line (Caco-2) during the proliferating and differentiated state. However, the oscillation analysis in both model systems needs to be repeated with a more sensitive technique that could obtain results faster and provide a bigger data set in order to obtain more conclusive results. The average response in a population of cells demonstrated to be 'noisy' leading to establishment of a single-cell analysis system. Analysis of the oscillatory capacity of SOCS3 protein at the single cell level, using a

type of a SOCS3 engineered protein and confocal microscopy will elucidate the validity of this phenomenon. At present, the application of mathematical modelling in complex biological pathways, is hindered by several factors, nevertheless the inadequate availability of quantitative biological data suitable for statistical modelling is certainly considered one of the most crucial aspects for optimizing the experimental design.

Future results could identify protein fluctuations that are differentially controlled in intestinal cancer cells and intestinal non-cancerous cells, and these differences can be used to offer treatment mechanisms that attack the cells at the time points where dysregulation will be more likely to happen.

Understanding and characterising the down-stream proteins of the molecular clock rhythmicity in gut tissues will play an essential role in both optimizing the time of treatment administration and the development of new therapeutics.

To our knowledge, this is the first report studying the impact of the molecular circadian synchronization in JAK/STAT pathway in human cancer intestinal cells and non-cancerous intestinal cells.

3 CRISPR/Cas9-mediated generation of knockin cell line expressing endogenous fluorescently tagged SOCS3 protein

3.1 Introduction

3.1.1 Single cell analysis

Cancer is a complex and heterogeneous disease that involves the upregulation of cell proliferation and suppression of apoptosis, resulting from alterations in gene expression that allow the cell to bypass normal growth of inhibitory signals [233]. Over the past decade, it has become evident that cancer maintains its proliferative potential against a wide range of anticancer therapies by viewing it as a robust system that is composed of tumour cells [195]. In general, the concept of robustness (or fragility) is defined as the ability of a system to maintain a performance in the presence of various perturbations or conditions of uncertainty. The mechanisms and the components in the biological system that are responsible for this robustness in the presence of molecular noise and environment perturbations is a very complicated topic [234, 235].

Currently, it is regarded as a highly conserved organizing principle in biology, as established by several systems-level research of bacterial chemotaxis [235], the cell

cycle , circadian rhythms [236], Drosophila segmentation [237], tolerance of stochastic perturbations in fundamental biochemical processes including transcription and protein interactions [238] as well as large-scale biochemical networks [239].

Typically, cancer acquires its robustness by functional redundancy, which is achieved by cellular heterogeneity, and feedback-control systems that are utilized to facilitate survival in the presence of several environmental perturbations, such as anti-cancer therapy [195]. Negative feedback loops are known to give rise to many distinct oscillating patterns to maintain the robustness of the system which can play key roles during cell fate decisions [129]. The methodical investigation of protein dynamics in most biological networks depends heavily on the analysis at the single cell level, as results of average response from the population of cells can be masked by random noise generated by the system [139].

The emerging field of systematic single-cell analysis has revealed significant and unprecedented insights into cellular function and heterogeneity. Recent technological advances coupled with mathematical modelling and computer science have revolutionized the expanding repertoire of novel and sensitive tools that can be used for single cell studies, enabling research to begun addressing long-standing questions such as the basis of cellular heterogeneity and how two whole genome sequences differ between two single cancer cells [240].

Single-cell analysis using fluorescence or bioluminescence are employed in several detecting methods including conventional FACS, microscopic imaging as well as immunochemistry. Furthermore, they allow quantitative measurements of translational activity in single cells and often with subcellular resolution [241].

Currently, the majority live single-cell microscopy studies rely on a common strategy where a fluorescent epitope is fused to the N-terminus or C-terminus of the protein under investigation, and the fusion construct is then stably or transiently expressed in mammalian cells [242]. There is an extensive collection of DNA vectors commercially available for tagging proteins. These can range from small peptide epitopes, such as the Hemagglutinin (HA), FLAG, and myc tags, to small protein, such as the green fluorescent protein (GFP) [243]. Transient overexpression of an epitope-tagged protein in cells has been widely used in the past two decades to study the subcellular localization and the dynamic behaviour of proteins by confocal microscopy [242, 244, 245]. However, it is unclear whether genome insertion context or the uncontrolled expression of exogenous proteins play an important role in the protein dynamics observed. Furthermore, strategies using transiently or constitutively overexpressed epitope-tagged constructs produce a multitude of artifacts that frequently disrupt normal cellular functions [246, 247]. These overexpression induced limitations most often include mislocalizations and protein aggregation [246], aberrant organelle morphology [248, 249], violated balanced gene dosage [250] and others [247].

Emerging genome-editing strategies which enable the introduction of an epitope tag at genomic loci by homologous recombination largely avoid these limitations and has been applied to culture cells [251], but its extremely low homologous recombination efficiency stand as a major problem to the *in vitro* and *in vivo* of application of these techniques [252, 253].

The recent breakthrough of the type II bacterial clustered regularly interspaced short palindromic repeat (CRISPR) and CRISPR-associated protein 9 (CRISPR/Cas9) system

has revolutionized the research field of genome editing in a wide variety of cells and organisms of different species [254, 255].

3.1.2 CRISPR-Cas9 system

The Clustered regularly interspaced short palindromic repeats (CRISPR), and associated proteins (Cas) constitute the CRISPR/Cas system and makes a crucial contribution to the adaptive immunity of many prokaryotic microbes against exogenic elements such as viruses and plasmids[256, 257]. The CRISPR/Cas systems are highly diverse and are classified into three major lineages (Type I, II and III), each of which is further arranged into sublineages based on their genetic composition as well as structural and functional differences[258-260]. Despite their high diversity, CRISPR-Cas systems share a core defining feature, the *cas* genes and the proteins they encode[257, 258].

Mechanistically, the activity of CRISPR-Cas immune system hinges on three major stages, termed as adaptation, CRISPR RNA (crRNA) biogenesis and target interference (figure 3.1). During the adaptation stage, CRISPR immunization involves the endonucleases Cas 1 and Cas 2, which carry the acquisition and polarized integration of invasive DNA sequence as a novel CRISPR spacer into the CRISPR locus. In the second stage, the CRISPR cassette is transcribed into a full pre-CRISPR RNA (pre-crRNA) transcript. The pre-crRNA is subsequently processed into small, mature, interfering crRNAs harbouring partial CRISPR spacer sequences attached to partial CRISPR repeats, forming CRISPR guide RNAs.
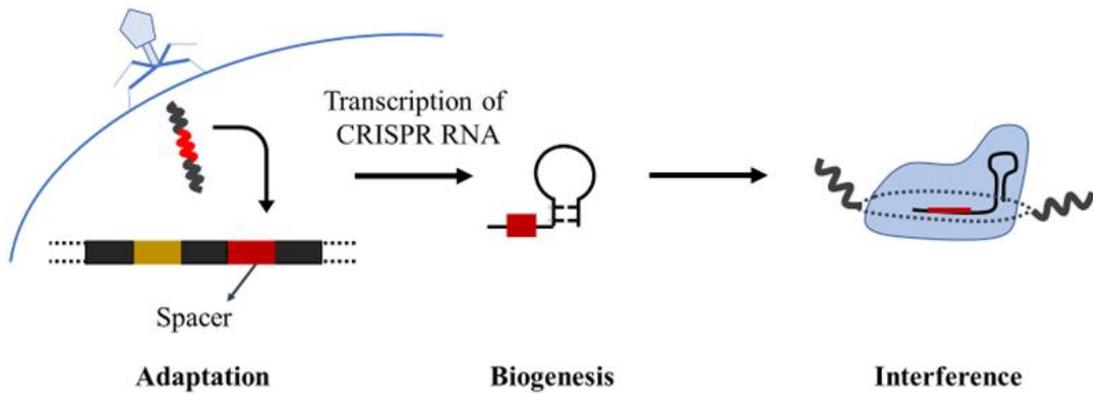
Figure 3.1: The three important stages of CRISPR-Cas immunity. Adaptation: integration of new spacers into the CRISPR locus. 2) Biogenesis: transcription of the CRISPR locus and expression of CRISPR RNA. 3) Interference: target degradation of mobile genetic elements by CRISPR RNA and Cas9 machinery [261]. Figure adapted from [261].

In the final stage, these cRNAs interact with the Cas protein machinery and guides it to target homologous foreign dsDNA (double-stranded DNA) or ssRNA (single stranded RNA) for nucleolytic degradation during CRISPR interference [257, 261-265].

Intense research efforts into the understanding of the mechanisms underlying the Type II CRISPR/Cas system spurred the development of novel programmable CRISPR/Cas9-based platforms revolutionizing the genome editing technology [266-270]. CRISPR/Cas9-based tools, thus far, have been applied very successfully in a wide range of organisms and cell lines and hold a great potential in realizing efficient genome editing as well as regulating gene expression without host dependence [271].

Engineered CRISPR-Cas9 system mainly compromises two components: a 'guide' RNA (gRNA) and a non-specific CRISPR-associated endonuclease (Cas9) [272]. The gRNA compromises a short synthetic RNA sequence consisting a 'scaffold' sequence required for Cas9-binding and a user-defined ~20 nt (nucleotide) targeting sequence which defines the genomic target to be altered. Furthermore, the target specificity

between the gRNA and the complementary genomic sequence requires the presence of Protospacer Adjacent Motif (PAM) located downstream of the target site, but it is not a part of the gRNA sequence [273].
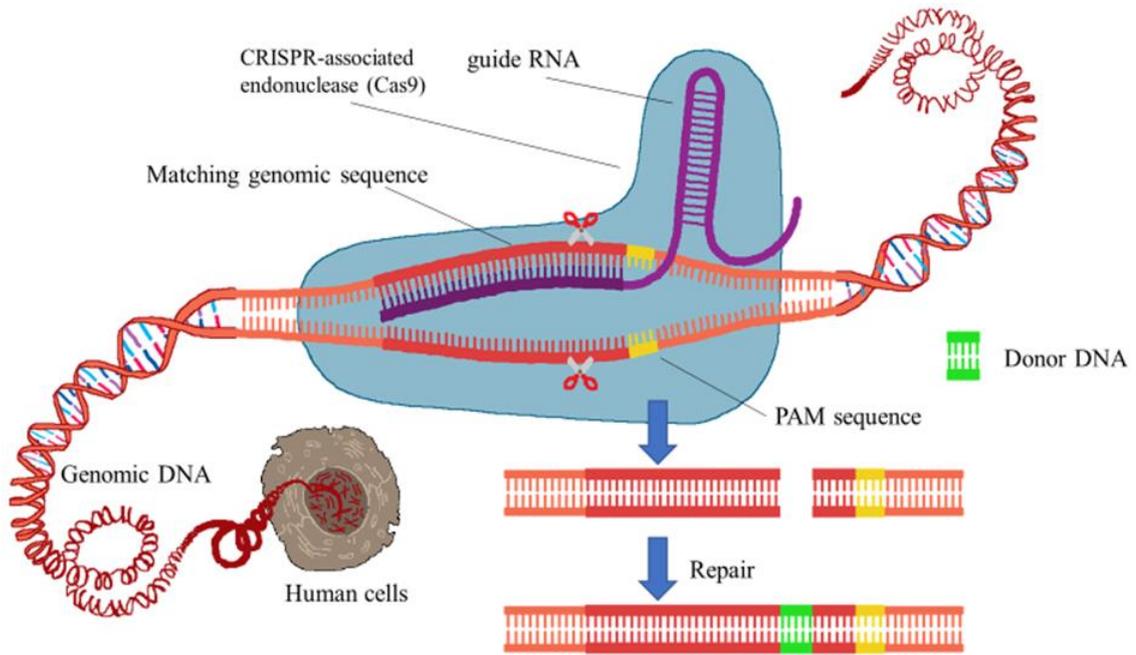
Figure 3.2: Schematic of the RNA-guided Cas9 nuclease. The cas9 nuclease is targeted to genomic DNA by an sgRNA consisting of a 20-nt guide sequence and a scaffold (purple). The guide RNA sequence matches with the DNA target directly upstream of a 5—NGG adjacent motif (PAM; yellow). Cas9 protein causes a double strand break ~3 bp upstream of the PAM (scissors icon) which is then repaired and the donor DNA template (green) can be incorporated into the genome. Adapted from [272].

A foundational initial step for achieving targeted genome editing primarily relies on delivering a DNA double-stranded break (DSB) at target locations in the genome. Two highly conserved DNA-repair machinery pathways typically repair DSBs: error-prone non-homologous end-joining (NHEJ) and highly precise homology-directed repair (HDR), both of which are active in nearly all cell types and organisms [274, 275].

The NHEJ repair pathway, as an error-prone process, can trigger the efficient introduction of undefined small insertion/deletion mutations (indels) at the DBS sites, which can result in the destruction of the translational reading frame of a coding sequence or the binding sites of *trans*-acting factors in promoters or enhancers. Subsequently, stimulation of NHEJ by site-specific DSBs can induce a robust transcription inactivation or activation of downstream target genes [256, 261, 276].

By contrast, homologous-directed repair (HDR) pathway triggers a precise recombination event where a homologous DNA molecule is used to provide the homology necessary to repair the double strand break in a template-dependent manner. Higher eukaryotic cells utilize the sister chromatid as a homologous template. Consequently, HDR pathway is only active in the S and G2 phase of the cell cycle but repressed in M and early G1 phase to prevent telomere fusion [277-279]. The repair machinery by HR has been characterized as a highly evolutionary conserved process present in all organisms. HDR mostly operates in an error free manner and repair DSBs with slower kinetics than NHEJ [278, 280].

In contrast to the unpredictable mutations achieved with NHEJ, specific DSBs can induce precise HDR-based gene editing by stimulating HR of the targeted locus and an exogenously supplied donor template. By providing an exogenous supplied DNA "repair template", it can generate defined deletions, insertions or nucleotide substitutions [276, 281].

Armed with the ability to harness the cell's endogenous DNA repair pathway, it is now feasible to introduce any type of genomic insertion, deletion or change in a target-specific manner [276].

Owing to its error-prone nature, NHEJ repair system is employed as the simplest form of genome editing tool to efficiently introduce small alterations at target regions. Classical NHEJ drives the direct ligation of the broken DNA ends which frequently causes undefined small insertions or deletions (indels) at the DSB site, presumably resulting in amino acid deletions, insertions, or frameshift mutations leading to a loss-of-function mutation within the targeted gene. However, the degree of the knock-out phenotype for a given mutant clone is ultimately decided by the extent of residual gene function [271, 277, 282, 283].

So far, the CRISPR/Cas9 engineering tool has been successfully utilized for the establishment of gene knockouts and knock-ins in cells derived from a vast array of species such as human, rats, mice, zebrafish, drosophila melanogaster, nematodes and the parasite Plasmodium *yoelii* [273, 284-288].

3.1.3 Genomic Engineering by Homologous Recombination and CRISPR/Cas9-Induced DSBs

In the short time that has passed since the emergence of CRISPR/cas9 technique, Cas9 has already been used for endogenously inserting epitope tags and fluorescent proteins in proteins and generating gene expression reporters in flies, mice, rats and human cells [242, 266, 286, 289].

In Drosophila melanogaster, a histone acetyltransferase protein encoded by the gene *chameau* was fused with GFP at the C-terminus and an uncharacterized gene *CG4221* was fused with myc [290].

In mice, the *Sox2* gene was fused with the V5 epitope. Moreover, Yang et al. 2013 [266] also generated mice carrying two different fluorescent reporter constructs in the genes

Nanog and *Oct4*. These fluorescent constructs utilized either a viral 2A peptide or an internal ribosome entry site (IRES) to exhibit the same expression level of fluorescent proteins as the endogenous gene without being attached to the protein product [266].

Ratz and his colleagues [242] were the first research goop to use the CRISPR/Cas9 technology to achieve locus-specific fluorescent protein insertion into human cells. The research group successfully managed to tag the nuclear DNA-binding non-histone high mobility group protein HMG-I (HMGA1), the cytoskeletal class-III intermediate filament protein Vimentin (VIM), and the focal adhesions plaque protein Zyxin (ZYX) [242]. They established different heterozygous and homozygous human U2OS knockin cell lines expressing the reversibly switchable fluorescent protein rsEGFP2 from their respective native genomic loci.
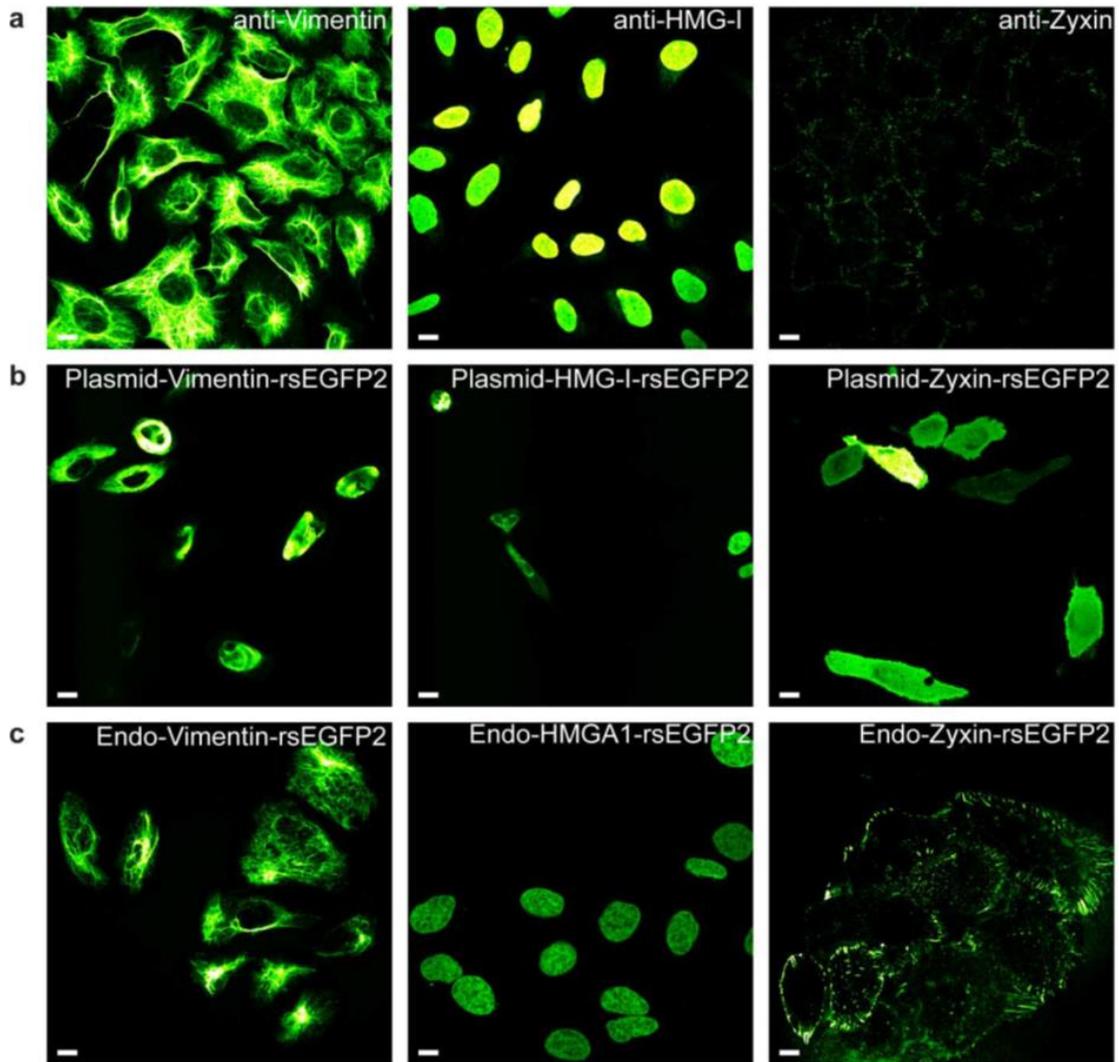
Figure 3.3: Fluorescence confocal analysis of endogenous rsEGFP2 tagged U2OS cells.

(a) Immunostaining analysis of U2OS cells indicating the expression levels of Vimentin, HMG-I and Zyxin. (b) Plasmid transfected U2OS cells exhibit heterogenous abundance of Vimentin-, HMG-I- and zyxin-eaEGFP2 fusion proteins under the control of a cytomegalovirus (CMV) promoter. (c) Live cell imaging of engineered U2OS cells showing the expression of rsEGFP2 tagged proteins from the endogenous genomic site. Scale bar, 10 μm. Figure from [242].

They further exhibited the advantages of endogenous protein abundance in relation to overexpression and indicated that typical overexpression-induced artifacts were circumvented in CRISPR-engineered cells [242] (figure 3.3).

Amongst the tags for monitoring protein expression, there are also tags to control other properties of protein function. The CRISPR/Cas9 editing tool have also been used to insert a small destabilization domain tag into the essential gene Treacher Collins-Franceschetti syndrome 1 (TCOF1) in human 293T cells [291]. This FKBP protein-based destabilization domain-tagging approach enables the protein to which it is attached, susceptible to degradation in the absence of an exogenously supplied compound (Shield-1), enabling accurate post-transcriptional regulation over protein expression [292].

3.1.4 Aim

Our hypothesis is that the cell population dynamically behaves differently from single cells in response to the same stimuli. Therefore, we used the CRISPR-Cas9 system to generate cells stably expressing a SOCS3 C-terminal GFP fusion protein and analyse the dynamic expression of SOCS3 at the single cell level. The aim of this chapter was to assess the feasibility of using the CRISPR-Cas9 system to endogenously tag SOCS3 protein. We also investigate different methods to increase the efficiency of homologous recombination by comparing different methods of cell transfections and chemical compounds that are known to stimulated HDR events at Cas9-generated DSB sites.

3.2 Experimental Design

3.2.1 Design and Cloning of CRISPR/Cas9 sgRNA vectors

The GeneArt® CRISPR Nuclease Vector with CD4 Enrichment Kit (ThermoFisher Scientific, C# A21177) was used to construct a plasmid vector for expressing all the functional elements needed for CRISPR/Cas9 genome editing tool. The kit includes a linearized GeneArt® CRISPR Nuclease Vector that consists of a Cas9 nuclease expression cassette and a guide RNA (gRNA) cloning cassette that enables for a rapid cloning of DNA that encodes target-specific CRISPR RNA (crRNA) (Appendix 1).

3.2.2 Design of CRISPR/Cas9 gRNA sequences

The design of gRNA oligos to successfully target the DNA region of interest must follow certain considerations. There are several online tools developed for designing sgRNA targets, with the main consideration to avoid off targets in the genome [293-295]. These online platforms usually consider an input sequence, a genomic site, or a gene and compute potential target/guide sequences with predicted reduced off-target effects [296]. These searching tools differ in their scheme for scoring potential guides and off-targets, as they input data from previous systematic mutagenic studies [297] or user-incorporated parameters [293] to individually score off-targets based on location and number of mismatched to the gRNA.

Common design considerations to obtain appropriate gRNAs specific to the target genomic region are summarized in table 3.1

Table 3. 1: **CRISPR-Specific parameters**

CRISPR-Specific considerations		References
sgRNA length	The standard sgRNA length for Cas9 is 20 nt	[298, 299]
sgRNA PAM sequence	The CRISPR-Cas9 3' PAM default is NGG	[298]
Self-complementarity	New evidence suggests that self-complementarity within the sgRNA or between the sgRNA and RNA backbone can hinder sgRNA efficiency. sgRNAs are most effective with a GC-content between 40 and 70%.	[300-302]
off-targets in the genome	Single-base mismatches up to 11 bp 5' of the PAM hinder cleavage by Cas9. However, mutations further upstream of the PAM maintain cleavage activity.	[254, 268, 297]
5' end of the gRNA	5' end dinucleotides is advised to be 5' GN- (for the U6 promoter) or 5' GG- (for T7 polymerase)	[303]
Distance between the modification site and DSB	DSB in close proximity to the insertion site enhances HDR event	[304, 305]

3.2.3 Design homology-directed repair (HDR) donor template

Typically, targeted DNA alterations require the construction of a plasmid-based donor repair template harbouring homology arms flanking the site of modification. The homology arms on each side can vary in length, but are typically longer than 500 bp [281, 306]. This method is employed to achieve large DNA modifications that involve insertion of reporter genes such as fluorescent proteins or antibiotic resistance genes.

Double stranded DNA (DsDNA) repair templates can be inexpensively produced at kilobase lengths by *de novo* synthesis, by PCR or through the generation of plasmids [307]. The vast majority of studies that have achieved to knock-in a large DNA fragment using CRISPR-Cas9 system have used 700-900-bp homology arms in both sides around the DSB [308, 309]. Therefore, the homology tails for inserting a fluorescent protein tag to SOCS3 CDS were designed to cover ~800 bp on each side of the CRISPR-Cas9 induced DSB. To ensure the integrity of the repair template, the sequence of the right arm was designed to begin 4 base pairs after the stop codon in order to disrupt the homology between the gRNA sequences and the repair template.

Importantly, the human SOCS3 gene (NCBI Reference Sequence: NG_016851.1) is located at the chromosome 17q25.3 and comprises two exons spanning 2729 nucleotides. SOCS3 is expressed in a long and a short isoform, with the short isoform being more resistant to proteasomal degradation. The expression of the two distinct SOCS3 isoforms is resulted from two alternative translational initiation sites, separated by 30 nucleotides within SOCS3 mRNA [310]. Furthermore, the SOCS3 gene includes a single 566 bp intron within the 985 bp long 5' UTR and a 1.6 kbp long 3' UTR (figure 3.5). The region within the 3'UTR of SOCS3 mRNA has shown to have a regulatory effect on SOCS3 mRNA half-life [102, 311]. The 5' UTR has shown to play a role in the ratio of SOCS3-isoform expression [310]. With the first exon of SOCS3 being an untranslated region, the coding sequence is concentrated specifically in exon 2 (total of 2401 nucleotides) consisting 678 nucleotides [312, 313].

Thus, the left homology arm and the reporter gene can be constructed with the help of a commercially SOCS3 fluorescent plasmid-based donor template that will include the coding sequence of SOCS3, a linker and the fluorescent tag sequence. The commercially available plasmid pCMV3-SOCS3-C-GFPSpark® tag (Sino Biological,

Cat. # HG11315-ACG) was used to construct the donor repair template. The donor plasmid was first subjected into a double restriction enzyme digestion using KpnI and PstI enzymes.

Homology tails were then designed and constructed to the linearized pUC19 vector using the Gibson assembly kit (NEB Cat. #E5510S). An overview of Gibson assembly technique is briefly described in figure 3.8.

DNA sequence for left homology arm (LHA) was amplified from SOCS3, C-GFPSpark® tag plasmid whereas right homology arm (RHA – 890 bp) was amplified from extracted human genomic DNA (HIECS) using the primer pairs listed in table 3.2.

Primers for Gibson Assembly were designed using the online NEBuilder Assembly tool (http://nebuilder.neb.com). PCR products were purified using the QIAquick PCR Purification kit (QIAGEN, UK, C# 28106) and cloned into a pUC19 vector using one-step isothermal assembly reaction [314]. The desiring end product is shown in table 3.3.

Table 3. 2: Primers used for constructing the homology repair template

Primer Name	Sequence (5' to 3')	Annealing Temp.
pUC19-Left-F1	AATTCGAGCTCGGTACTATGGTCACCCACAGCAAGTTTC	65.1 ℃
Left-Right-R1	GCCCTTTATTACTTGTACAGCTCGTCCATG	65.1 ℃
Right-Left-F2	GTACAAGTAATAAAGGGCGCAAAGGGCATG	68.3 ℃
pUC19-Left-R2	AGCTTGCATGCCTGCAAACAGGGAGCATTTAAGGCGAATC	68.3 ℃
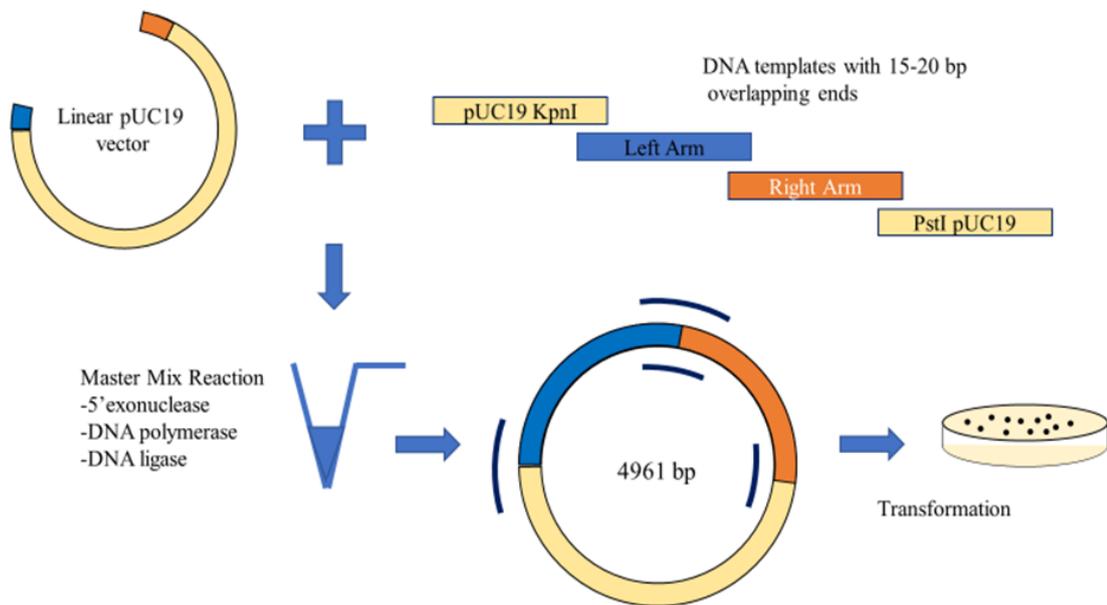
Figure 3.4: Gibson assembly Technique. Gibson assembly is a useful tool that includes three enzymatic activities in a single-tube reaction: a T5 exonuclease, that cleaves nucleotides at the 5' end DNA template and exposes the complementary sequence for annealing, a 3' extensions activity of a DNA polymerase, which fills in the annealed single-stranded gaps and finally a DNA ligase that ligates the nick and assembles the DNA fragments together [314].

The homology repair template from the assembled vector was then amplified with PCR and used for transfection.

Table 3. 3: PCR primers for repair DNA amplification and GFPSpark gene knock-in assembly with homology arms. ATG start codon for SOCS3 gene is indicated in green, primers are assigned with underline, stop condon is indicated in red, linker is shown in yellow block.

Homology left arm (SOCS3 CDS) (693bp)	TCTCGCGAATGCATCTAGAATGGTCACCCACAGCAAGTTTCCCGCCGCCGGGATGAGCCGCCCCCTGGACACCAGCCTGCGCCTCAAGACCTTCAGCTCCAAGAGCGAGTACCAGCTGGTGGTGAACGCAGTGCGCAAGCTGCAGGAGAGCGGCTTCTACTGGAGCGCAGTGACCGGCGGCGAGGCGAACCTGCTGCTCAGTGCCGAGCCCGCCGGCACCTTTCTGATCCGCGACAGCTCGGACCAGCGCCACTTCTTCACGCTCAGCGTCAAGACCCAGTCTGGGACCAAGAACCTGCGCATCCAGTGTGAGGGGGGCAGCTTCTCTCTGCAGAGCGATCCCCGGAGCACGCAGCCCGTGCCCCGCTTCGACTGCGTGCTCAAGCTGGTGCACCACTACATGCCGCCCCCTGGAGCCCCTCCTTCCCCTCGCCACCTACTGAACCCTCCTCCGAGGTGCCCGAGCAGCCGTCTGCCCAGCCACTCCCTGGGAGTCCCCCCAGAAGAGCCTATTACATCTACTCCGGGGGGCGAGAAGATCCCCCTGGTGTTGAGCCGGCCCCTCTCCTCCAACGTGGCCACTCTTCAGCATCTCTGTCGGAAGACCGTCAACGGCCACCTGGACTCCTATGAGAAAGTCACCCAGCTGCCGGGGCCCATTCGGGAGTTCCTGGACCAGTACGATGCCCCGCT
T2A linker – GFPSpark (732 bp)	==GGGGGTGGAGGCTCT==GTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGAAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGCTAACTTCAAGGTTCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCA==TGGACGAGCTGTACAAG====TAA==
Homology right arm (890 bp)	==TAAAGGGCGCAAAGGGCATG==GGTCGGGAGAGGGACGCAGGCCCCTCTCCTCCGTGGCACATGGCACAAGCACAAGAAGCCAACCAGGAGAGAGTCCTGTAGCTCTGGGGGGGAAAGAGGGCGGACAGGCCCCTCCCTCTGCCCTCTCCCTGCAGAATGTGGCAGGCGGACCTGGAATGTGTTGGAGGGAAGGGGGAGTACCACCTGAGTCTCCAGCTTCTCCGGAGGAGCCAGCTGTCCTGGTGGGACGATAGCAACCACAAGTGGATTCTCCTTCAATTCCTCAGCTTCCCCTCTGCCTCCAAACAGGGGACACTTCGGGAATGCTGAACTAATGAGAACTGCCAGGGAATCTTCAAACTTTCCAACGGAACTTGTTTGCTCTTTGATTTGGTTTAAACCTGAGCTGGTTGTGGAGCCTGGGAAAGGTGGAAGAGAGAGAGGTCCTGAGGGCCCCAGGGCTGCGGGCTGGCGAAGGGAAATGGTCACACCCCCCGCCCACCCCAGGCGGAGGATCCTGGTGACATGCTCCTCTCCCTGGCTCCGGGGAGAGTCACCGAAAAACACAGGTTCCAAAGTCTACCTGGTGCCTGAGAGCCCAGGGCCCTTCCTCCGTTTTAAGGGGGAAGCAACATTTGGAGGGGATGGATGGGCTGGTCAGCTGGTCTCCTTTTCCTACTCATACTATACCTTCCTGTACCTGGGTGGATGGAGCGGGAGGATGGAGGAGACGGGACATCTTTCACCTCAGGCTCCTGGTAGAGAAGACAGGGGATTCTACTCTGTGCCTCCTGACTATGTCTGGCTAAGAGAT==TCGCCTTAAATGCTCCCTGT==

3.3 Methods

3.3.1 Design of CRISPR sgRNAs

CRISPR sgRNAs were designed using the Optimized CRISPR Design tools (http://crispr.mit.edu) and (http://chopchop.cbu.uib.no). Since the aim of this thesis is to endogenously insert a fluorescent tag into the C-terminal region of SOCS3, the gRNA oligos were designed specifically to target the genomic site adjacent to the stop codon of SOCS3 coding sequence (figure 3.4). Analysis of gRNA sequences was computed according to the suggested parameters of the two CRISPR website tools. Off-target hit-scores are established by taking into consideration the sum of all mismatches, mismatch absolute position in the gRNA sequence (to account for the relatively high disturbance of mismatches in close proximity to the PAM site and mean pairwise distance between mismatches (to take into consideration the steric outcome of closely neighboring mismatches in altering gRNA-DNA template interaction). gRNA sequences with quality scores of more than 50% are considered candidate targeting sequence if no high scoring off targets fall in gene sites. Prior to ordering the oligo templates, overhangs are needed to be added in each side of the gRNA sequences. GTTTT is inserted to the 3′ end of the oligonucleotide. The GTTTT is complementary to the overhang sequence, CAAAA, in the linearized CRISPR Nuclease Vector and compromises the first 5 bases of the tracrRNA. Furthermore, CGGTG is added to the 3′ end of the oligonucleotide. This sequence is complementary to the overhang sequence, CACCG, in the linearized GeneArt® CRISPR Nuclease Vector and compromises the last 4 bases of U6 promoter and the first base required for PolIII transcription start site.

The designed single-stranded oligonucleotides containing the guide sequence of the gRNA were ordered from Sigma-Aldrich (UK).
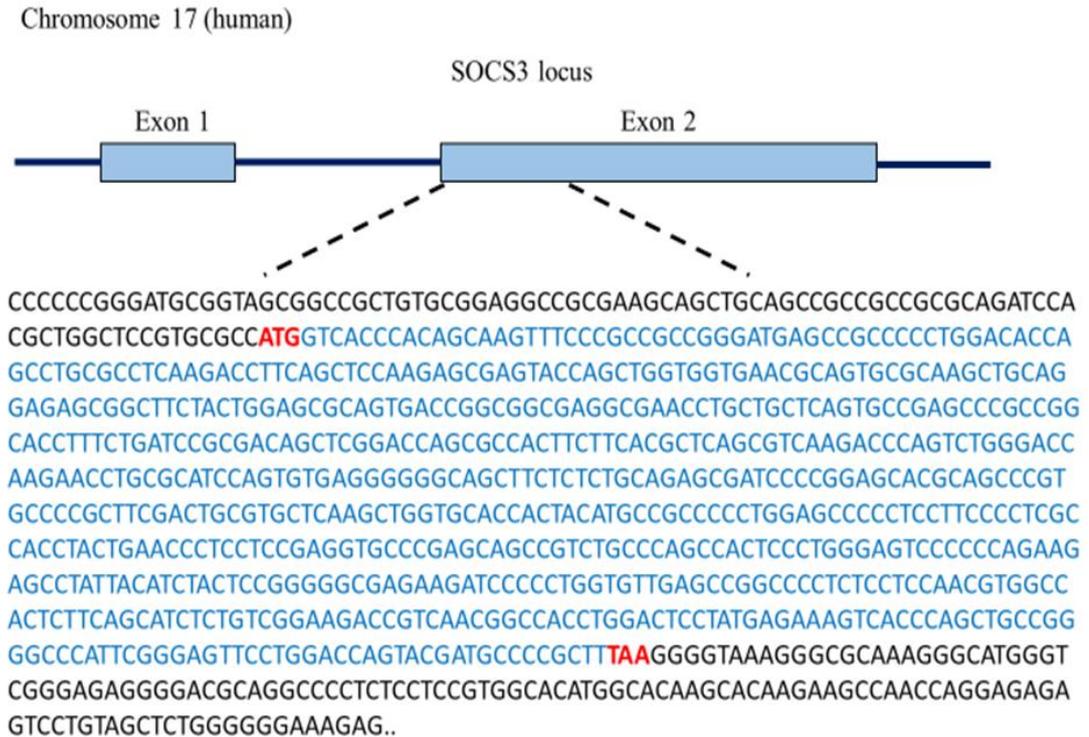
Figure 3.5: Design of gRNA sequence targeting SOCS3 CDS. An outline of the strategy used to generate CRISPR-Cas9 harbouring gRNA targeting the human SOCS3 gene locus (Gene ID: 9021). SOCS3 gene compromises two exons. The coding sequence is located at the start of exon 2. The start and stop codons are indicated in Red. The coding sequence within the start and stop codons is indicated in blue.

3.3.2 Annealing Single-Stranded oligonucleotides to generate a double-stranded oligonucleotide

For synthesis of double-stranded oligonucleotides a mix was prepared with the following components in 0.5 ml Individual Tube:

Forward strand oligonucleotide (200 μM)	5 μl
Reverse strand oligonucleotide (200 μM)	5 μl
10x Oligonucleotide Annealing Buffer	2 μl
DNase/RNase-Free Water	8 μl
Total Volume	20 μl

The tube was then incubated at 95ºC for 4 minutes in a heat block. Following the incubation, the reaction mixture was left at room temperature to cool down to 25 ºC for 10 minutes. The mixture was centrifuged briefly for 5 seconds and then mixed gently. The resulting 50 μM ds oligonucleotide solution was either stored at -20 ºC or further diluted and used to downstream applications.

Once the single-stranded oligonucleotides are annealed, the resulting 50 μM ds oligonucleotide mixture is required to be further diluted to a final working concentration of 5 nM.

A 500 nM ds oligonucleotide stock solution was prepared by adding the following components in a sterile 0.5 ml tube:

50 μM ds oligonucleotide stock	1 μl
DNase/RNase-Free Water	99 μl
Total Volume	100 μl

The final solution was mixed thoroughly and stored at -20 ºC.

A 5 nM ds oligonucleotide stock solution was prepared by adding the following components in a sterile 0.5 ml tube:

500 nM ds oligonucleotide solution	1 μl
10x Oligonucleotide Annealing Buffer	10 μl
DNase/RNase-Free Water	89 μl
Total Volume	100 μl

The resulting mixture was mixed and kept on ice for downstream applications.

3.3.3 Cloning double-stranded oligonucleotide into CRISPR Nuclease Vector

Once the ds oligonucleotides have been generated and prepared at the correct stock solutions, they are cloned into a linearized GeneArt®CRISPR Nuclease Vector provided by the kit (Appendix 1).

For preparation of a ligation reaction, a mixture is prepared at room temperature by adding the following elements in the order shown:

5x Ligation Buffer	4 µl
Linearized GeneArt®CRISPR Nuclease Vector	2 µl
Ds oligonucleotide (5 nM)	2 µl
DNase/RNase-Free Water	11 µl
T4 DNA Ligase	1 µl
Total Volume	20 µl

The mixture is then mixed thoroughly by pipetting up and down and then incubated for 2 hours at room temperature (25-27 °C). Once the incubation time is finished, 3 µl of the ligation was used for bacterial transformation (see details below).

3.3.4 Agarose gel Electrophoresis

Agarose gel electrophoresis was used to separate DNA fragments by size. Agarose solution was prepared at concentrations varying from 1-2%, depending on the size of DNA fragments, in 1xTAE buffer (1 L of 10x TAE stock: 48.4g Tris base [tris(hydroxymethyl)aminmethane], 11.4 ml glacial acetic acid (17.4M), 20ml of 0.5M EDTA and deionized water). The mixture was heated in a microwave until all the agarose had completely dissolved. The mixture was left to cool down to 55-60°C and GelRed DNA stain was added to a final concentration 1:50000. The solution was poured

into a gel casting tray and allowed to solidify. DNA samples were mixed with 6x gel loading dye (NEB, C# B7021S) and added to the wells. In addition, 5µl of a molecular weight ladder was loaded onto each gel to allow the size of DNA bands to be identified by comparing them against bands of known size.

For this thesis, either 1 kb or 100 bp molecular weight standards were loaded on each gel, thus allowing the identification of DNA samples in a ranger of 0.5-10 kb and 100-1517 bp respectively.
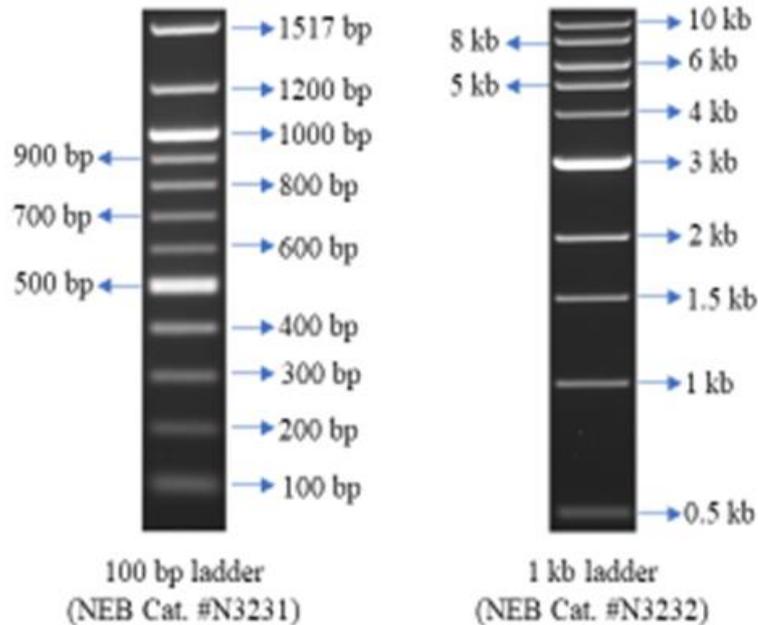
Figure 3.6: **Molecular weight Ladders.** 100 bp and 1 kb molecular weight ladder used during agarose gel electrophoresis. Figure adapted from NEB website (www.neb.com).

Voltage was set between 70-120 V depending on the size of the gel. Gels were imaged using the BioRad ChemiDox[TM] XRS+ System. Densitometry was performed via Image LabTM software.

3.3.5 Bacterial transformation

NEB 5-alpha Competent E. coli cells (High Effciency, NEB Cat. #C2987) were removed from the -80ºC freezer and thawed on ice. Either 2 µl of Gibson assembled product, or 10 ng plasmid DNA was added to the competent cells and mixed gently by flicking the tube 4-5 times. Cells were incubated on ice for 30 minutes and then heat-shocked at 42ºC for 45 seconds. The mixture was then placed on ice for 2 minutes and 950 µl of room temperature SOC media was added to the cells. The solution was then placed in a shaking incubator, set to 37ºC and 250rpm for 1 hour. For identification and recovery of successful transformants, 50-200 µl of this cell mixture was spread on pre-warmed agar plates, containing appropriate antibiotics, using a sterile spreader. Plates were incubated overnight at 37ºC in an inverted position. The following day, the culture plates were examined for colony formation and 3-4 single colonies were picked using a sterile 200 µl pipette tip and further cultured in 50 ml LB-Broth media in the presence of appropriate antibiotics in a shaking incubator, set to 37ºC and 250rpm. Successfully amplified colonies were employed in downstream applications such as plasmid isolation, restriction digest, sequencing and transfection. Agar plates were wrapped in parafilm and stored in a refrigerator.

3.3.6 Glycerol Stocks

For long term storage, transformed bacteria were stored at -80ºC in glycerol. Bacterial cell cultures were grown in LB liquid media supplemented with appropriate antibiotic overnight at 37ºC and 250rpm. The following day, 0.5 ml of this bacterial culture was mixed with 0.5 ml of filtered-sterilized 70% glycerol solution in cryogenic vials (1.8 ml Cryogenic Vials, Starlab) for storage.

3.3.7 Plasmid isolation

Plasmid DNA was isolated from successfully transformed bacterial cells using GeneJET Plasmid Miniprep Kit (ThermoFisher Scientific, C# K0502) according to the manufacturer's directions. Briefly, cells were harvested by centrifugation of 5 ml culture at 10000 rpm for 3 minutes. Supernatant was carefully removed without disrupting the cell pellet. The pelleted cells were resuspended with 250 µl of the Resuspension Solution (supplemented with RNase A solution) mixed thoroughly by vortexing and the cell suspension was transferred to a new sterile microcentrifuge tube. For lysis, 250 µl of the Lysis solution was added and mixed gently by inverting the tube 6 times. The cell lysate was then incubated for 2 minutes at room temperature but not more than 5 minutes to avoid denaturation of supercoiled plasmid DNA. For neutralization of the cell lysate, 350 µl of the Neutralization Solution was added and mix immediately and thoroughly by inverting the tube 6 times. Separation of cell debris and chromosomal DNA from plasmid DNA was performed by centrifugation at 14000 rpm for 5 minutes. After centrifugation, the upper phase was transferred into a GeneJET spin column provided by the kit. A second centrifugation at 14000 rpm was repeated for 1 minute and the flow-through was discarded. Plasmid DNA was washed by adding 500 µl of Wash Solution (supplemented with 100% EtOH) and the column was centrifuged at 14000 rpm for 45 seconds. The flow-through was discarded and the washing step was repeated once again. Once the flow-through was discarded from the second wash step, an additional centrifugation was performed to remove any traces of Wash solution. Finally, the GeneJet spin column was transferred into a new sterile 1.5 ml microcentrifuge tube and 50 µl of Elution Buffer was added to the centre of the spin

column membrane. The column was incubated at RT for 5 minutes and then centrifuged at full speed for 2 minutes. Following centrifugation, mini-preps were stored at -20 ℃.

3.3.8 Gibson Assembly reaction

DNA templates with sharing terminal overlaps were derived from PCR reactions using the primers listed table 3.2. and were assembled in a 20 µl reaction mix. Volumes for each DNA fragment and the double digested pUC19 plasmid vector were adjusted according to the manufacturer's recommendations. Based on Gibson optimized protocol, the highest optimized cloning efficiency can be achieved if 50-100 ng of vector is used with 2-3-fold of excess inserts. Furthermore, the total volume of amplified PCR fragments should not exceed 20% of the total reaction volume.

Recommended protocol used for assembling 3 DNA Fragments		
DNA Ratio	Vector : insert = 1:2	
Total Amount of pUC19 vector plasmid	50-100 ng	73.5 ng or 0.0425 pmoles or 5 µl
Total Amount of Left-HA	2-3 fold excess	0.085 pmoles or 78.6 ng or 0.36 µl
Total Amount of Right-HA	2-3 fold excess	0.085 pmoles or 50.66 ng or 0.30 µl
Assembly Master Mix (2x)	10µl	10µl
DNase/RNase-Free Water	10 – X µl	4.34µl
Total Volume	20 µl	20 µl

The number of pmoles of each DNA template for optimal assembly, according to fragment length and weight, the following formula was used:

$$Number\ of\ pmoles = \frac{(weight\ in\ ng)\ x\ 1000}{base\ pairs\ x\ 650\ daltons}$$

Once all components are combined into a single master mix, the sample is then incubated in a thermocycler at 50 °C for 15 minutes. Following incubation, the sample is then placed on ice for 30 minutes. 2 µl of assembled product then is used for bacterial transformation.

3.3.9 Genomic DNA extraction

Adherent culture cells were rinsed once with 5 ml of Dulbecco's Phosphate Buffered Saline (DPBS) without calcium and magnesium to remove residual medium. An appropriate volume of TrypLE^{TM} Express Enzyme (1X) (ThermoFisher Scientific, C# 12604013) was added to flask and incubated at 37 °C until cells had detached. Once cells were detached, 5-10 ml of pre-warmed complete medium was added, and the cell suspension was then centrifuged for 5 minutes at 250 x g. The supernatant was removed, and the cell pellet was resuspended in 200 µl of DPBS. Genomic DNA was then isolated using the GeneJET Genomic DNA Purification Kit (ThermoFisher Scientific, C# K0721) according to the manufacturer's protocol. Briefly, cells were transferred to a microcentrifuge tube and centrifuged at 200 x g for 5 minutes. Supernatant was removed, and cells were resuspended in 200 µl of deionized sterile water. Moreover, 200 µl of Lysis Solution and 20 µl of Proteinase K Solution was added to the cell pellet. The mixture was mixed thoroughly by pipetting up and down to obtain a uniform suspension. The sample was then incubated at 56°C for 10 minutes. Following

incubation, 20 µl of RNase A solution was added, and the mixture was mixed by vortexing and incubated at room temperature for further 10 minutes. Next, 400 µl of 50% ethanol was added, and the final sample was mixed by vortexing. Furthermore, a GeneJET column was inserted in a collection tube and the prepared lysate was transferred into the column. The column with the sample was then centrifuged for 1 minute at 6000 x g. Once the solution was flown to the collection tube, the tube was replaced with a new one and inserted to the column. Next, 500 µl of Wash buffer I supplemented with 100% ethanol was added to the column and centrifuged for 1 minute at 8000 x g. The flow-through was discarded and further 500 µl of Wash buffer II supplemented with 100% ethanol was added to the column. The column was then centrifuged at maximum speed 16000 x g for 3 minutes. The collection tube with the flow-through was again discarded and the column was inserted to a sterile 1.5 ml microcentrifuge tube. Finally, 20 µl of Elution buffer was added to the centre of the column membrane to elute genomic DNA and incubated for 5 minutes at room temperature prior to centrifuging for 2 minutes at 8000 x g. The purified was then immediately stored at -20°C for downstream applications.

3.3.10 Polymerase Chain Reaction (PCR) amplification

All primer sets used in this report were designed using either the NEBuilder Assembly tool (http://nebuilder.neb.com) or the online primer blast tool provided on the homepage of the National Center for Biotechnology Information and purchased from Sigma Aldrich, UK. Moreover, all PCRs were catalysed using Q5® High-Fidelity DNA Polymerase (C# M0491, New England BioLabs). PCR reactions were carried out in clean, nuclease-free 0.2 ml thin-walled microcentrifuge tubes, where each reaction

containing a PCR master mix with a final volume of 50 µl**.** The PCR master mix contained 10 µl of 5X Q5 Reaction buffer, 1 µl of 10mM dNTPs, 2.5 µl of 10 µM of each Forward and Reverse primer, 10 ng of Genomic DNA or 1 ng of plasmid DNA, 0.5 µl of Q5 High-Fidelity DNA polymerase, 10 µl of 5X Q5 High GC Enhancer and finally Nuclease-Free water to a final volume of 50 µl. A negative control containing Nuclease-Free water instead of DNA template was always included in every reaction. The PCR master mix was mixed gently and spun briefly to collect all liquid to the bottom of the tube. Thermocycling was performed in a MJ Research PTC-100 Programmable Thermal controller according to the parameters listed in the table below.

Step	Temperature	Time
Initial Denaturation	98ºC	30 seconds
35 cycles	98 ºC	10 seconds
	Depending on the primer set used	30 seconds
	72 ºC	30 seconds/kb
Final Extension	72 ºC	2 minutes
Hold	4 ºC	15 minutes

3.3.11 PCR Purification

PCR products were purified using the QIAquick PCR Purification kit (QIAGEN, UK, C# 28106) according to the manufacturer's recommendations. To this end, 5 volumes of Buffer PB (supplemented with 1:250 volume pH indicator I) was mixed with 1 volume of the PCR reaction. Furthermore, 10 µl of 3 M Sodium Acetate, pH 5.0 was added to the reaction and mixed thoroughly. Meanwhile, a QIAquick column is inserted into a provided 2 ml collection tube. The reaction mix is then applied into the QIAquick column and centrifuged for 45 seconds at 17,900 x g. Once centrifugation was finished,

the flow-through was discarded and 750 µl of Buffer PE (supplemented with 100% Ethanol) was added to wash DNA. The mixture was centrifuged again for 1 minute at 17,900 x g. Flow-through was discarded once again and the column was centrifuged for 1 minute at maximum speed. The QIAquick column was then placed into a new sterile 1.5 ml microcentrifuge tube. For DNA elusion, 20 µl of Buffer EB (10 mM Tris•Cl, pH 8.5) was added to the centre of the QIAquick membrane and incubated at room temperature for 5 minutes. The column was then centrifuged for 2 minutes at maximum speed.

3.3.12 Measurement of nucleic acid concentrations

Nucleic acid concentration was measured using *NanoDrop* 2000c Spectrophotometer. OD measurement at a wavelength of 280 nm and 230 nm provide information concerning the purity of the analysed/purified samples. Accordingly, contamination with proteins usually give an $E_{260/280} < 1.8$ for DNA. Impurity due to salt or solvents usually show ratios as $E_{260/230} < 2$ for DNA samples. Typically, DNA templates used in cloning and PCR amplification experiments displayed an $E_{260/280} > 1.8$ and $E_{260/230} > 2$.

3.3.13 DNA extraction from agarose gels

Typically, DNA purification was important during cloning experiments either in order to clean up templates after enzyme digestion or PCR or following agarose gel electrophoresis so as to extract DNA templates from the respective agarose gel. In the latter case, extraction and purification of DNA from agarose gels was performed by using the GeneJET Gel Extraction kit (ThermoFisher Scientific, C# K0692) according to the manufacturer's instruction. Briefly, DNA fragments were excised from the

agarose gel using a clean scalpel and its weight was determined. 100 µl of Binding buffer was added per each 100 mg of gel slice and the samples were incubated at 50ºC until the agarose was completely dissolved. Subsequently, the individual samples were loaded onto columns and the contained DNA bound to the silica membrane by centrifugation at 12000g for 1 minute. Following centrifugation step, the flow throw was discarded and 100 µl was added to the column. The centrifugation step was repeated. Samples were washed with 700 µl of Wash Buffer and again centrifuged at 12000 g for 1 minute, followed by a third centrifugation step at 12000 for 2 minutes in order to dry the silica membrane. Purified DNA was afterwards eluted in 15-50 µl of Elution Buffer by incubation at RT for 3 minutes and subsequent centrifugation at maximum speed for 1 minute.

3.3.14 **Restriction Enzyme Double Digestion**

The preparative specific cleavage of plasmid vectors by restriction enzymes was used during cloning procedures in order to linearize plasmids for homology recombination events. In addition, analytical restriction digests were performed to control the outcome of cloning procedures.

A double restriction digestion was used to linearize pUC19 plasmid vector prior its use in Gibson Assembly reaction.

The following elements were added in a microcentrifuge:

Component	50 μl Reaction
DNA	1 μg
10x CutSmart Buffer	5 μl (1x)
KpnI-HF	1 μl (or 10 units)
PstI-HF	1 μl (or 10 units)
DNase/RNase-Free Water	to 50 μl

Plasmids were also subjected to a single enzyme digestion using Sca-I to obtain linearized templates for either agarose gel analysis or for preparation to transfection to cells.

Component	25 μl Reaction
DNA	0.5 μg
10x CutSmart Buffer	2.5 μl (1x)
ScaI-HF	0.5 μl (or 5 units)
DNase/RNase-Free Water	to 25 μl

Once all components were added, the mix reaction was centrifuged briefly (5 seconds) and then incubated at 37 ºC for 15 minutes.

Sca-I-HF cleavage reaction was stopped by carrying out a heat inactivation of the enzyme at 80ºC for 20 minutes if necessary.

3.3.15 Transfection

3.3.15.1 Lipid-mediated cell transfection

Cell passage was not higher than 10 and the chosen confluency at the time of transfection was 60-70%. All the parameters for lipid-mediated transfection were according the FuGENE® HD Protocol online database (https://www.promega.com/techserv/tools/FugeneHdTool/) with minor adjustments.

One day prior to transfection, the cells were plated in a 6-well plate at a cell density of $1\text{-}2 \times 10^5$ cells per well in 3 ml of complete growth medium (MEM + 10% Foetal bovine serum). For plasmid DNA transfection, 3.3 µg of DNA was mixed with 77.5 µl of sterile deionized water, followed by addition of 87.4 µl of sterile deionized water containing 9.9 µl of FuGENE® HD reagent. The solution was mixed thoroughly and incubated for 15 minutes at room temperature prior to addition to the cells. The entire solution was added to the cells and mixed by gently swirling the plate. The place was then incubated at 37ºC for 48 hours in a 5% CO_2 incubator.

3.3.15.2 Electroporation

Electroporation for HIECS cells was done in a Nucleofector® I device (Lonza) using the Amaxa® Cell Line Nucleofector® Kit L (Lonza, C# VCA-1005) and the T-020 program according to the manufacturer's instructions. On the day of transfection, the cells were washed once in DPBS (warmed at 37ºC) to remove any medium containing FBS. Cells were detached with Gibco TrypLE™ Express Enzyme (1X) reagent (Thermofisher Scientific, C# 12604013), resuspended in complete media and counted. It was important to avoid temperature fluctuations, thus all solution used during the transfection protocol were either warmed at 37ºC or at ambient temperature. 10^6 HIECS cells were centrifuged at 100xg for 10 minutes at room temperature. Following centrifugation, supernatant was removed completely. The cell pellet was resuspended carefully in 100 µl room-temperature Nucleofector® Solution L. 2 µg of DNA was then mixed gently in the cell suspension and the cell/DNA solution was then transferred carefully into a certified cuvette. The cuvette was placed into the Nucleofector® Cuvette Holder and the desired program was selected. Immediately after pulse application, the electroporated cells were then incubated for 10 minutes at room temperature. Following

incubation, 500 µl of pre-equilibrated culture medium was added to the cuvette and the sample was transferred into a prepared 12 or 6 well plate. The prepared culture plate contained pre-warmed culture media and the electroporated cells were added to the media in a dropwise fashion using a kit supplied pipette. Once the cells were added into the wells, the plate was incubated at 37ºC in a 5% CO_2 incubator.

3.3.16 Antibody-free magnetic cell sorting

Transfected cells with CRISPR-Cas9 plasmid were screened and sorted for CD4 positive cells using the Dynabeads® CD4 Positive Isolation kit (Invitrogen-Life Technologies, C#1131D) according to the manufacturer's directions with minor adjustments. Firstly, Dynabeads (Invitrogen-Life Technologies, super-paramagnetic polystyrene beads 4.5 µm in diameter and coated with primary monoclonal anti-CD4 antibody) were resuspended thoroughly by vortexing for 30 seconds. Once Dynabeads were mixed, 25 µl of beads was transferred into a sterile microcentrifuge and 1 ml of Buffer 1 (DPBS without Ca^{2+} and Mg^{2+} supplemented with 0.1 % FBS and 2 mM EDTA, pH 7.4) was added. The final mixture was resuspended, and the tube was then placed into a DynaMag^{TM} – Spin Magnet (ThermoFisher Scientific, C# 12320D) for 1 minute and the supernatant was discarded carefully without disturbing the beads. The tube was then removed from the magnet and the beads were resuspended in 250 µl of Buffer 1.

Three days after CRISPR-Cas9 plasmid transfection, cells were washed and detached from the culture plates, and they were resuspended in 5 ml of pre-warmed complete medium. Cell solution was then centrifuged at 250 x g for 10 minutes at room temperature and the supernatant was removed without disturbing the cell pellet. The cell pellet was resuspended in 750 µl of Buffer 1. To this end, the prepared sample was

mixed with the prepared mix of 25 µl Dynabeads and the final solution was incubated for 20 minutes at 4 °C with gentle rotation. Following incubation, the tube was placed into a magnet for 2 minutes. While the tube was still in the magnet, the supernatant was carefully removed without detaching the cells. The tube is then removed from the magnet and 1 ml of Buffer is added into the cells and mixed by pipetting up and down for 2-3 times. The tube is then returned back to the magnet for 2 minutes and repeat the same washing procedure for 2 more times. Once the three washes are completed, the cell pellet is then resuspended in 100 µl Buffer 2 (Cell culture media supplemented with 1% FBS). To release CD4$^+$ positive cells from the beads, 10 µl DETACHaBEAD$^®$ solution is added to the cell suspension and mixed thoroughly by pipetting. The final cell mixture was incubated for 45 minutes at room temperature with gentle rotation. After the incubation, the tube is place back to the magnet for 1 minute. While the tube is on the magnet, the supernatant containing the released cells from the beads are transferred into a new sterile tube. The tube was removed from the magnet and 500 µl of Buffer 2 was added into to the beads and mixed well. The tube was returned back to the magnet for 1 minute and supernatant was transferred to the second tube. This washing and collecting step was performed 3 more times in order to obtain any residual cells. The final collected cell suspension was washed thoroughly by mixing the cells with 4 ml of Buffer 2 and centrifuged for 6 minutes at 400 x g to remove DETACHaBEAD$^®$ reagent. The isolated cells then were either transferred in culture plates or flasks supplemented with pre-warmed complete medium or studied by flow cytometry.

3.3.17 Flow Cytometry

For determining the percent of GFP-positive cells, cells were centrifuged to remove culture media, washed in DPBS without Ca^{+2} and Mg^{+2}, and then suspended in FACS buffer (1mM EDTA, 25 mM HEPES pH. 7.4, 1% FBS and 1 x DPBS without Ca^{+2} and Mg^{+2} prior analysing them by flow cytometry using CytoFlex S (Beckman Coulter). Cells were initially gated for the intact cell population using forward scatter versus side scatter plots, and then gated for transfected cells based on the presence of the transfected control. Transfected cells were gated for GFP-positive cells on the side scatter versus GFP plots. Pure cells (Caco2 or HIECS) without treated for CRISPR/Cas9 were used as a negative control and cells transiently transfected with SOCS3, C-GFPSpark® tag plasmid was used as positive controls for FACS.

The expression of CD4 cell surface marker was examined in HIECS using flow cytometry. Briefly, cells were washed twice in DPBS without Ca^{+2} and Mg^{+2}, counted and resuspended in 100 µl of cell staining buffer, which contained DPBS without Ca^{+2} and Mg^{+2}, 0.5 % FBS, 0.04 EDTA and 0.05% Sodium Azide (NaN_3). Cells were then treated with 5 µl of Human TruStain FcX™ (Fc Receptor Blocking Solution, Biolegend, C# 422302) for 10 minutes at room temperature to block nonspecific Fc binding sites. Following incubation, cell were washed with DPBS and then resuspended in 100 µl of cell staining buffer. Cells were incubated with 5 µl of PE anti-human CD4 antibody (Biolegend, C# 300508) or with 5 µl of PE mouse IgG1, κ Isotype Control Antibody (Biolegend, C# 400112) on ice for 20 minutes in the dark. Then, cells were washed twice with 2 ml of cell staining buffer by centrifugation at 350 g for 5 minutes. Cell pellet was resuspended in 500 µl of cell staining buffer and analysed them by flow cytometry using CytoFlex S (Beckman Counter). Data analysis for transfection

efficiency and CD4 cell surface marker staining was performed by using the CytExpert Software (Beckman Coulter).

The flow cytometer was routinely calibrated with CytoFlex Daily QC setup and CytoFlex Daily QC Fluorospheres to verify its optical alignment and fluidics system.

3.3.18 GeneArt® Genomic Cleavage Detection Kit

The percentage of locus-specific indel formation was analysed in HIECs by GeneArt® Genomic Cleavage Detection Kit (GCD) according to the manufacturer's protocol. Specifically, 10^6 HIECs were electroporated with 2 µg of bicistronic Cas9/gRNA1 or gRNA2 plasmid. 48 hours post electroporation, 10^6 HIECs were trypsinized and mixed with 50 µl of a lysis solution that contained 50 µl Cell Lysis Buffer and 2 µl Protein Degrader. Thermocycling was performed in a MJ Research PTC-100 Programmable Thermal controller according to the parameters listed in the table below.

Temperature	Time
68ºC	15 minutes
95 ºC	10 minutes
4 ºC	5 minutes

Following thermal cycling completion, the cell lysate was briefly mixed and 2 µl of the lysate was used in a PCR reaction that contained:

Component	Sample	Control
Cell lysate	2 µl	-
10 µM Forward/Reverse	1 µl	-
Control Template & Primers	-	1 µl
AmpliTaq Gold® 360 Master	25 µl	25 µl

Nuclease-Free H$_2$O	22 µl	24 µl
Total	50 µl	50 µl

The forward and reverse primers were designed to yield an amplicon with the length of 529 bps and the potential cleavage site to potential produce two distinct product bands with lengths of 195 bp and 334 bp for gRNA1, and 201 bp and 328 bp for gRNA2.

Primers Name	DNA sequence (5'→3')
GCD forward	CCAGAAGAGCCTATTACATC
GCD reverse	GACACTTCGGGAATGCTG

A positive control was provided by the manufacturer and it included a set of primers and template. A negative control was also included, and nuclease water was used instead of cell lysate.The PCR mixture was placed on a thermal cycler and subjected to the following parameters listed in the below table:

Stage	Temperature	Time	Cycles
Enzyme activation	95ºC	10 minutes	1 x
Denature	95ºC	30 seconds	40 x
Anneal	58 ºC	30 seconds	
Extend	72 ºC	30 seconds	
Final extension	72 ºC	7 minutes	1 x
Cooling	4 ºC	2 minutes	1 x

For the cleavage assay, PCR fragments were denatured and randomly re-annealed to generate heterogenous DNA duplexes. A master mix was formed by combining 8 µl of PCR product and 1 µl Detection Reaction Buffer. The mixture was placed on a thermal cycler and subjected to the following conditions:

Stage	Temperature	Time	Temperature/time
1	95ºC	5 minutes	-
2	95-85ºC	-	- 2ºC/seconds
3	85-25 ºC	-	- 0.1 ºC/seconds
4	4 ºC	3 minutes	-

For the enzyme digestion assay, 1 µl of Detection Enzyme was added to all the test samples and mixed thoroughly. 1 µl of Nuclease-free water was added to the negative control sample. The final mixture was then incubated at 37 ºC for 1 hour. Following incubation, the sample was mixed well and analysed on a 2% agarose gel.

3.3.19 Sequencing

Purified PCR products and plasmids were stored at 4ºC before shipment for Sanger sequencing by Source BioScience Sequencing (Nottingham, UK). Briefly, 5 µl of primers at concentration of 3.2 pmol/µl, aliquots of plasmids at concentration of 10 ng/µl and PCR templates at concentration of 100 ng/µl. DNA alignments were achieved by using the online tool ClustalW2 from EMBL-EBI website (https://www.ebi.ac.uk/Tools/msa/clustalw2/).

Primers Name	DNA sequence (5'→3')
CRISPR-Cas9 U6 Forward	GGACTATCATATGCTTACCG
HDR template- Forward	ATGGTCACCCACAGCAAGTT
GDC sequencing primer	CTACTGAACCCTCCTCCGAG

3.3.20 PrestoBlue Viability Essay

The assay was carried out in a 96-well plate. Each well contained 5000 cells and cultured medium. Once cell were treated with the corresponding chemical compound and incubated for 48 hours, 10 µl of PrestoBlue (Invitrogen) was added per 100 µl media to each well and 4 extra wells were filled with 100 µl media and 10 µl PrestoBlue without any cells as a background control. Plate was incubated at 37 ℃ for 4 hours and then the fluorescence was measured using a multi-well plate reader with the excitation/emission wavelengths set at 560/590 nm for PrestoBlue. Relative fluorescence was calculated by subtracting the average of the background controls from each of the well reading, then comparing the average of the treated wells to the average of the wells from cells that were only treated with DPBS.

3.4 Results

3.4.1 CRISPR/Cas9-mediated knockin of GFPSpark at SOCS3 loci.

A general workflow was established to generate HIECS and Caco2 cell lines stably expressing SOCS3 C-terminal GFPSpark fusion proteins from its respective endogenous promoter utilizing the CRISPR/Cas9 engineering system (figure 3.7).
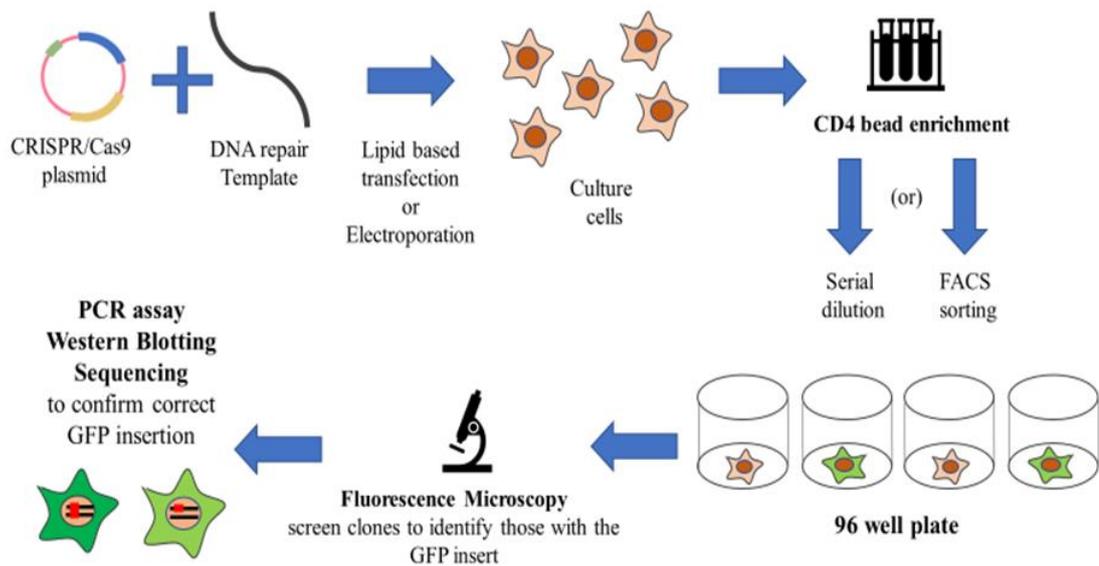
Figure 3.7: Workflow for generation fluorescent tag knockin cell lines with CRISPR-Cas9. A general workflow to establish cell lines stably expressing C-terminal SOCS3 - GFPSpark fusion protein from its respective endogenous promoter using the CRISPR/Cas9 technique.

Specifically, the workflow to successfully generate monoclonal human knockin cell lines for dynamic live-imaging analysis involved 5 critical stages: 1) the design of donor plasmid and gRNA, which also included its cloning step to the CRISPR-Cas9 plasmid; 2) transfection to human cells, 3) CD4 bead enrichment and single cell serial cloning step to 96 well plate; 3) fluorescent confocal microscopy to screen through cell clones that have successfully accepted the GFPSpark to the C terminal of SOCS3 CDS; 4)

PCR, Western blotting and DNA sequencing to confirm correct GFPSpark insertion without any mutation generated from the CRISPR-Cas9 system.

3.4.2 Generation of CRISPR-Cas9 plasmid harbouring gRNA

The vast majority of genes that have been endogenously tagged with a reporter marker using the CRISPR-Cas9 technology, have used the C-terminus of the gene of interest to insert the reporter gene [242, 315, 316]. Thus, this thesis attempted to endogenously insert a fluorescent tag into the C-terminal region of SOCS3.

For this thesis, CRISPR sgRNAs were designed using the Optimized CRISPR Design tools (http://crispr.mit.edu) and (http://chopchop.cbu.uib.no). Analysis of the C-terminal coding region of SOCS3 revealed a number of potential Cas9 guide sites in the exon 2, in close proximity to the stop codon. Analysis of the C-terminal coding region of SOCS3 revealed a number of potential Cas9 guide sites in the exon 2, in close proximity of the stop codon. They were identified as demonstrating high target affinity, high quality scores and low off-target scores, especially low to none off-target sites within other genes. The sgRNA sequences that were closest to the stop codon, but had minimal mismatches to the human genome, were chosen. The target sites and gRNA sequences used in this study are shown in figure 3.8.
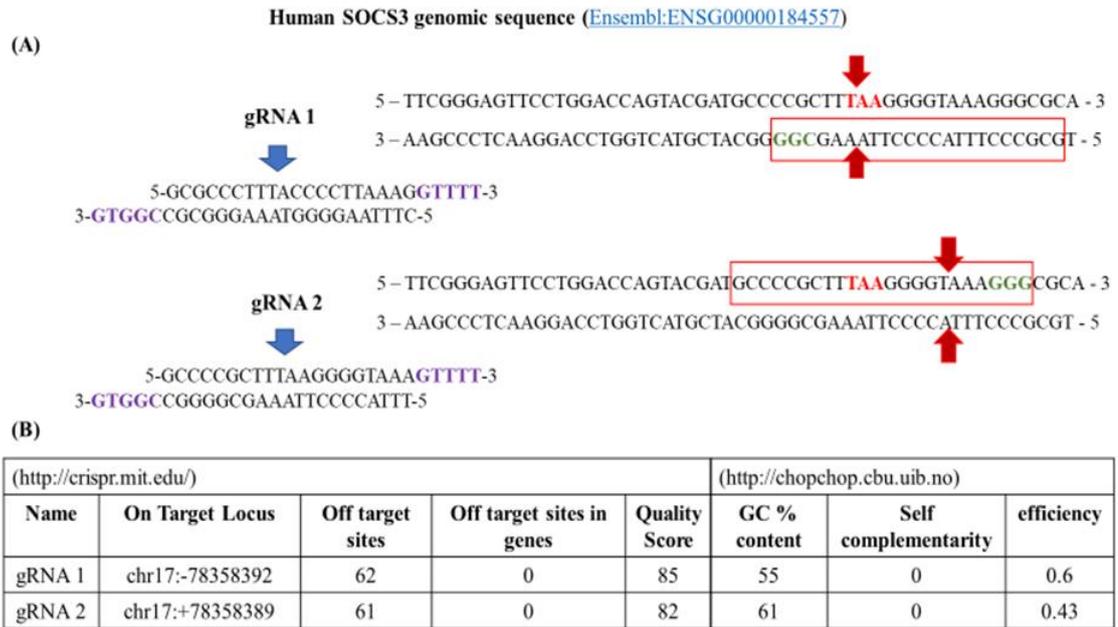
Figure 3.8: Synthesis of gRNA template. (A) Schematic representation of the chosen gRNA sequences located at the C-terminus of SOCS3 coding sequence. The stop codon (TAA) is indicated in red bold letters. The pam sequences are indicated in bold green letters. The position of genomic cuts is shown with a red arrow. **(B)** Design of gRNA sequences was computed according to the suggested parameters of two CRISPR website tools (http://crispr.mit.edu/) and (http://chopchop.cbu.uib.no). Quality score is demonstrated by the faithfulness of on-target activity calculated as 100% minus a weighted sum of off-target hit-scores in the genome. The efficiency score computed by the website tool (http://chopchop.cbu.uib.no) is based on a PAM proximal position 20th nucleotide-dependent influence on CRISPR-Cas9 cutting efficiency [301]. This efficiency score is normalized from 0 to 1 intervals.

Double stranded gRNA oligos were ligated in the linearized CRISPR-Cas9 vector and transformed into competent cells. The following day, few colonies were picked and expanded further. Purified Plasmids with the sequencing primer (U6 Forward) were sent to be analysed with Sanger sequencing by Source BioScience Sequencing (Nottingham,

UK). Plasmids that have successfully inserted gRNA oligos in the right direction are presented in figure 3.9.
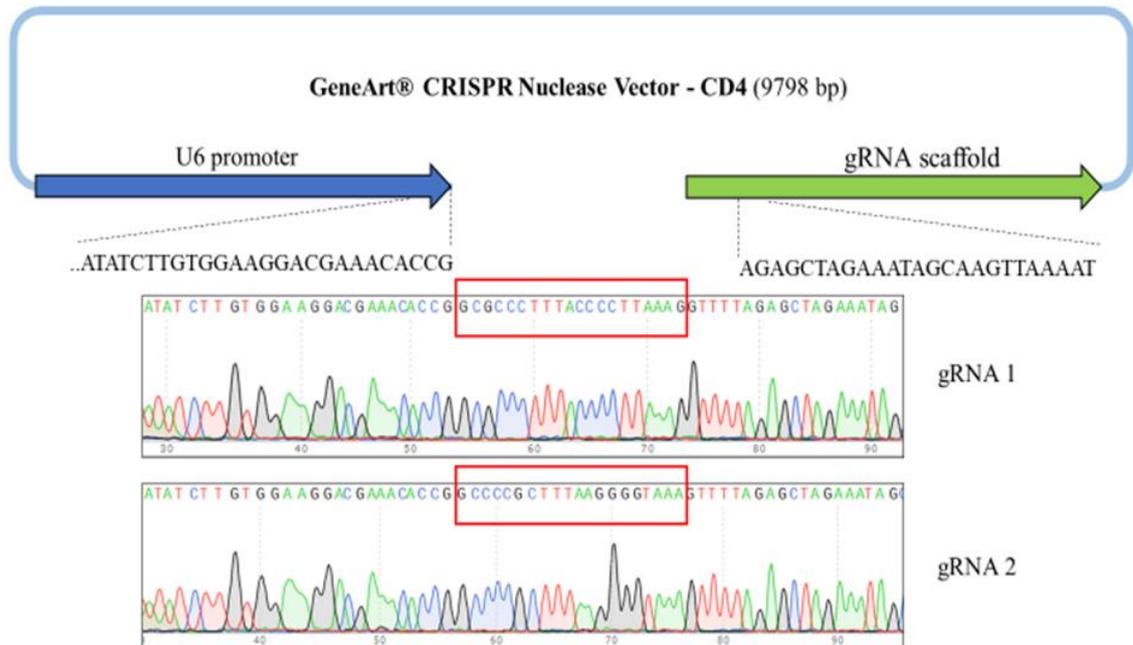
Figure 3.9: **DNA Chromatogram Analysis of CRISPR-Cas9 clone plasmids harbouring gRNA 1 or gRNA 2.** Results from Sanger BioScience Sequencing (Nottingham, UK).

3.4.3 Generation of homology-directed repair (HDR) donor repair template

Previous research articles that have generated an endogenous knock-in reporter gene using CRISPR-Cas9 system have used a minimum 700-900-bp homology arms in both sides around the DSB [308, 309]. Therefore, the homology tails for inserting a fluorescent protein tag to SOCS3 CDS were designed to cover ~800 bp on each side of the CRISPR-Cas9 induced DSB.

Homology arms were designed to facilitate the integration of the GFPSpark coding sequence at the 3'end of SOCS3 coding sequence, replacing the stop codon but leaving the genomic locus otherwise unchanged. The one half of the repair template that included the left homology arm linked to the fluorescent tag (GFPSpark) was amplified

from pCMV3-SOCS3-C-GFPSpark® plasmid. The PCR products were analysed in 1% agarose gel to verify its correct amplification. The other half of the repair template that included the right homology arm was amplified from extracted genomic DNA from HIECS. The PCR products were also analysed in 1% agarose gel to verify its correct amplification (figure 3.10).
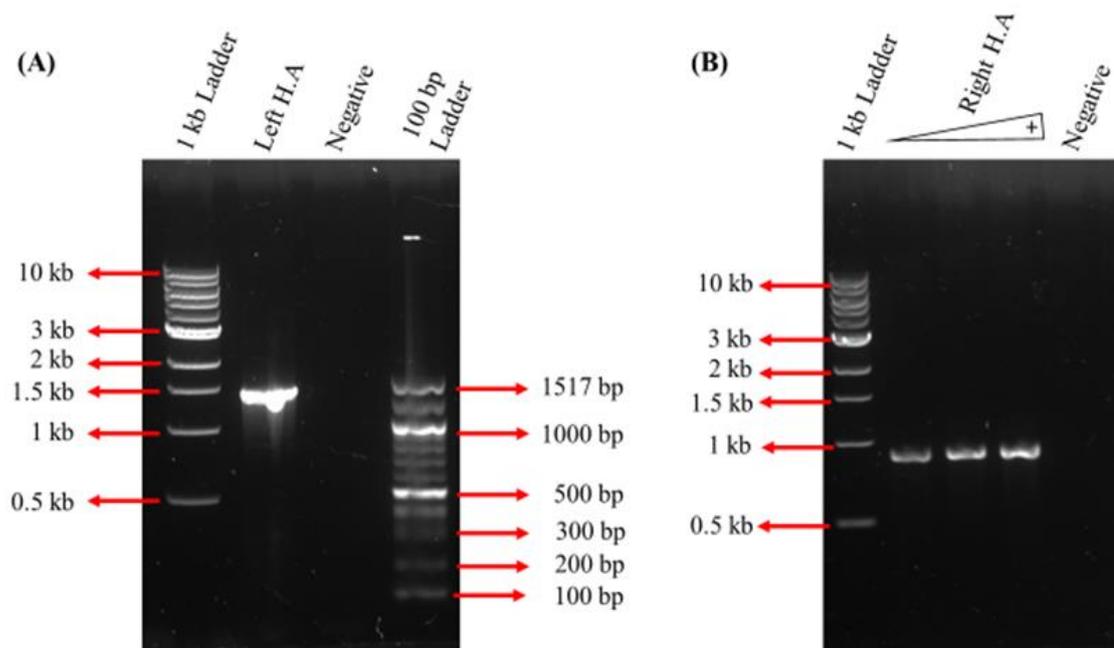
Figure 3.10: PCR amplification of Left and Right Homology tails. **(A)** Lane 1: 1kb Ladder NEB, lane 2: PCR amplification of the left homology arm (left H.A) linked to the fluorescent tag (GFPSpark), lane 3: Negative Control, Lane 4, 100bp Ladder (NEB). **(B)** Lane 1: 1kb Ladder NEB, lane 2,3,4: PCR amplification of the right homology arm, DNA amount loaded in an increasing manner, Lane 5: Negative Control

The expected DNA band sizes for the left homology tail linked to GFPSpark is 1406 bp and for the right homology arm is 890 bp. As indicated in figure 3.10, the amplified PCR products yielded correct band sizes in the agarose gels. Prior to Gibson assembly, the donor pUC19 plasmid was first subjected into a double restriction enzyme digestion using KpnI and PstI enzymes. For the construction of a complete SOCS3 HDR repair template, the two purified PCR products were assembled into the digested pUC19

plasmid using a one-step isothermal assembly reaction [314]. The assembled plasmids were transformed into competent cells, plated in agar plates supplemented with 100 µg/ml Ampicillin and incubated at 37ºC overnight to allow for bacterial colonies formation. The following day, 4 clones were picked and expanded further. These 4 bacterial colonies were then subjected to a miniprep purification step, linearized with an enzymatic digestion with Sca-I and analysed in 1% agarose gel. The results from the agarose gel is shown in figure 3.11.
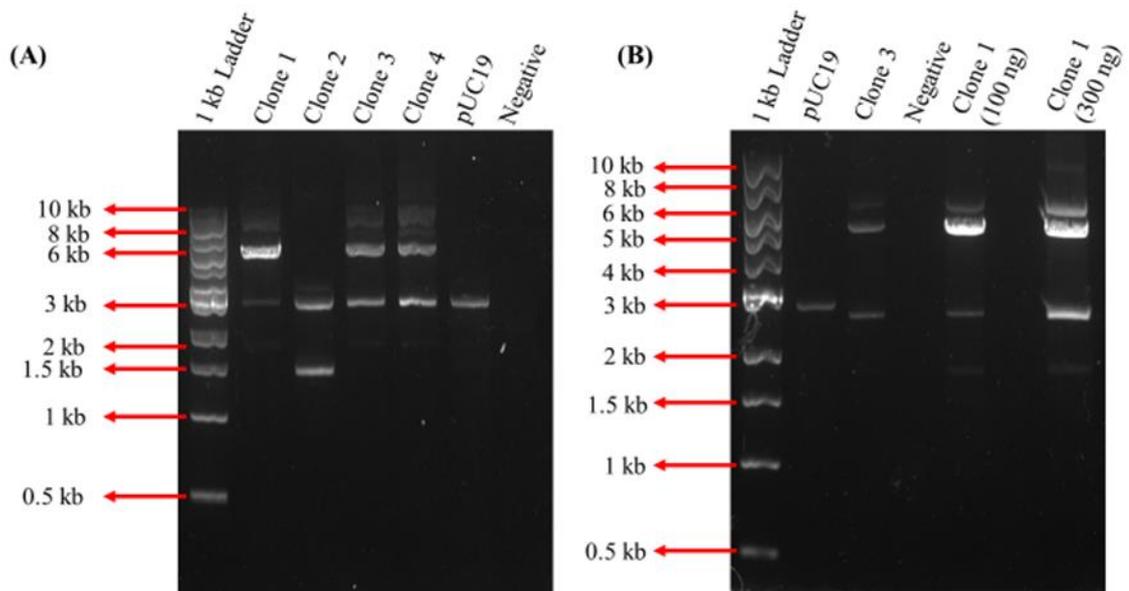
Figure 3.11: **Establishment of SOCS3 HDR repair template.** **(A)** Lane 1: 1 kb Ladder (NEB), lane 2-4: linearized assembled clone plasmids, lane 5: linearized pUC19 vector, lane 6: negative control. **(B)** 1 kb Ladder, lane 2: linearized pUC19 vector, lane 3: Clone 3, lane 3: negative control, lane 4: 100 ng of linearized clone 1, lane 5: empty, lane 6, 300 ng of linearize Clone 1 vector.

The expected band size for the assembled HDR donor plasmid is 4961 bp. According to the figure 3.11 (A), the only clone that created clearer results with one approximate band is clone 1. The other 3 clones created multiple bands. These bands could possible represent either linearized pUC19 plasmid or bands from incomplete pUC19 digestion.

Since the DNA ladder did not load correctly, the agarose gel analysis was repeated. The repeated analysis is shown in figure 3.11 B. Clone 1 was loaded in two different amounts (100 ng and 300 ng) in a 0.7% agarose gel. Even though, the ladder did not run correctly, it provided a better indication of the sizes of DNA bands in the gel. However, loading a bigger amount of linearized clone 1 plasmid, it also displayed multiple bands. Therefore, we attempted to gel extract and purify the band of interest using the GeneJet Gel Extraction kit (ThermoFisher Scientific). Absorbance measurements were made on a Nanodrop™ spectrophotometer. The gel purification step did not yield a considerable amount of DNA (< 3 ng/µl) and the ratio of absorbance at 260 nm and 280 nm was low (<0.3) where it should have been ~1.8. In addition, the DNA samples from the gel purification experiment were also analysed in an 1% agarose gel. The results from the agarose gel are presented in figure 3.12 (A). According to the agarose gel analysis, the DNA samples failed to show any clear DNA band, but instead they only exhibited a smearing effect. PUC19 vector acted as a control and it was loaded and gel purified along with clone 1. This was done to test if the protocol being used for gel purification was the reason that it was not working.

Instead, I proceeded with PCR amplifying the region of interest that included the left homology arm linked to GFPSpark and the right homology arm as a complete repair template using the primers pUC19-Left-F1 and pUC19-Left-R2. The PCR products were purified and analysed in 1% agarose gel to verify its correct amplification. The expected band size for the assembled HDR template is 2296 bp.

According to the agarose gel analysis presented in figure 3.12 (B), the PCR amplification yielded bands of the correct sizes.
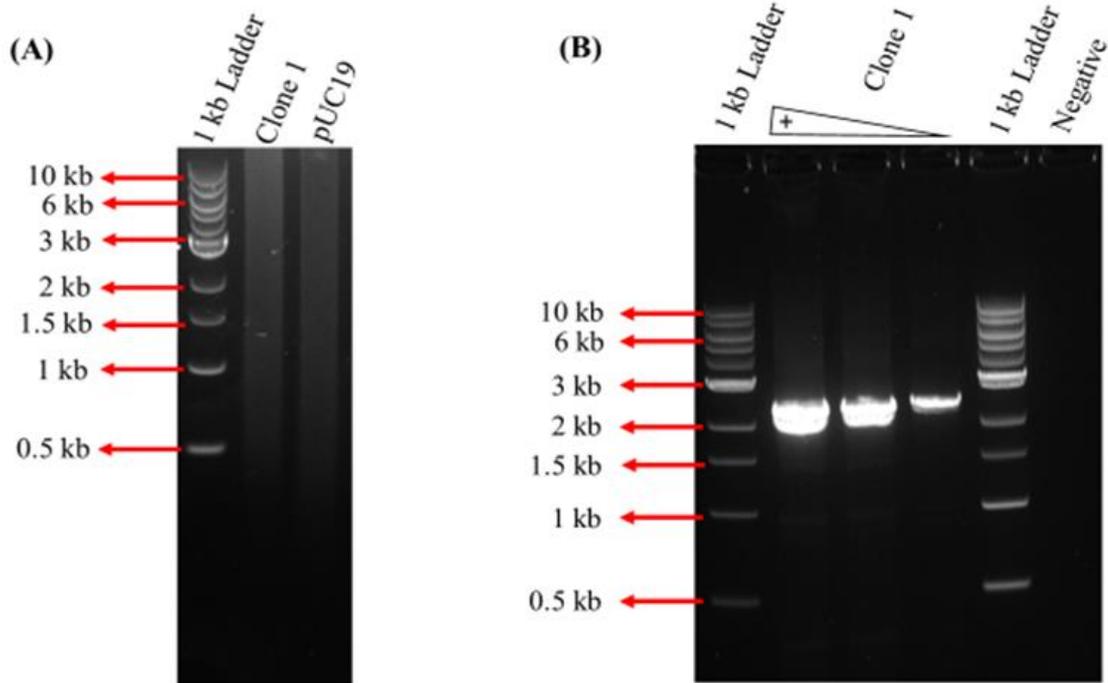
Figure 3.12: **Purifying SOCS3 HDR repair template (A)** Lane 1: 1 kb Ladder (NEB), lane 2: gel purified clone 1 plasmid, lane 3: gel purified pUC19 plasmid. **(B)** Lane 1: 1 kb Ladder (NEB), lane 2-4: PCR bands of HDR template, DNA amounts were loaded in a decreasing manner, lane 5: 1 kb Ladder (NEB), lane 6: negative control.

Purified PCR HDR templates with the sequencing primer (HDR-template Forward) were sent to be analysed with Sanger sequencing by Source BioScience Sequencing (Nottingham, UK).

The results obtained from Source Bioscience Sequencing were analysed by aligning the sequenced DNA sample with the expected DNA sequence from the HDR donor template. The DNA alignment was achieved by using the online tool ClustalW2 from EMBL-EBI website (https://www.ebi.ac.uk/Tools/msa/clustalw2/). Results are shown below.

```
HDR-template          75  GCCTGCGCCTCAAGACCTTCAGCTCCAAGAGCGAGTACCAGCTGGTGGTG  124
                          |.||||||||||||||||||||||||||||||||||||||||||||||||
Sequencing Analysis   30  GNCTGCGCCTCAAGACCTTCAGCTCCAAGAGCGAGTACCAGCTGGTGGTG  79

HDR-template         125  AACGCAGTGCGCAAGCTGCAGGAGAGCGGCTTCTACTGGAGCGCAGTGAC  174
                          ||||||||||||||||||||||||||||||||||||||||||||||||||
Sequencing Analysis   80  AACGCAGTGCGCAAGCTGCAGGAGAGCGGCTTCTACTGGAGCGCAGTGAC  129

HDR-template         175  CGGCGGCGAGGCGAACCTGCTGCTCAGTGCCGAGCCCGCCGGCACCTTTC  224
                          ||||||||||||||||||||||||||||||||||||||||||||||||||
Sequencing Analysis  130  CGGCGGCGAGGCGAACCTGCTGCTCAGTGCCGAGCCCGCCGGCACCTTTC  179

HDR-template         225  TGATCCGCGACAGCTCGGACCAGCGCCACTTCTTCACGCTCAGCGTCAAG  274
                          ||||||||||||||||||||||||||||||||||||||||||||||||||
Sequencing Analysis  180  TGATCCGCGACAGCTCGGACCAGCGCCACTTCTTCACGCTCAGCGTCAAG  229

HDR-template         275  ACCCAGTCTGGGACCAAGAACCTGCGCATCCAGTGTGAGGGGGGCAGCTT  324
                          ||||||||||||||||||||||||||||||||||||||||||||||||||
Sequencing Analysis  230  ACCCAGTCTGGGACCAAGAACCTGCGCATCCAGTGTGAGGGGGGCAGCTT  279

HDR-template         325  CTCTCTGCAGAGCGATCCCCGGAGCACGCAGCCCGTGCCCCGCTTCGACT  374
                          ||||||||||||||||||||||||||||||||||||||||||||||||||
Sequencing Analysis  280  CTCTCTGCAGAGCGATCCCCGGAGCACGCAGCCCGTGCCCCGCTTCGACT  329

HDR-template         375  GCGTGCTCAAGCTGGTGCACCACTACATGCCGCCCCCTGGAGCCCCCTCC  424
                          ||||||||||||||||||||||||||||||||||||||||||||||||||
Sequencing Analysis  330  GCGTGCTCAAGCTGGTGCACCACTACATGCCGCCCCCTGGAGCCCCCTCC  379

HDR-template         425  TTCCCCTCGCCACCTACTGAACCCTCCTCCGAGGTGCCCGAGCAGCCGTC  474
                          ||||||||||||||||||||||||||||||||||||||||||||||||||
Sequencing Analysis  380  TTCCCCTCGCCACCTACTGAACCCTCCTCCGAGGTGCCCGAGCAGCCGTC  429

HDR-template         475  TGCCCAGCCACTCCCTGGGAGTCCCCCCAGAAGAGCCTATTACATCTACT  524
                          ||||||||||||||||||||||||||||||||||||||||||||||||||
Sequencing Analysis  430  TGCCCAGCCACTCCCTGGGAGTCCCCCCAGAAGAGCCTATTACATCTACT  479

HDR-template         525  CCGGGGGCGAGAAGATCCCCCTGGTGTTGAGCCGGCCCCTCTCCTCCAAC  574
                          ||||||||||||||||||||||||||||||||||||||||||||||||||
Sequencing Analysis  480  CCGGGGGCGAGAAGATCCCCCTGGTGTTGAGCCGGCCCCTCTCCTCCAAC  529

HDR-template         575  GTGGCCACTCTTCAGCATCTCTGTCGGAAGACCGTCAACGGCCACCTGGA  624
                          ||||||||||||||||||||||||||||||||||||||||||||||||||
Sequencing Analysis  530  GTGGCCACTCTTCAGCATCTCTGTCGGAAGACCGTCAACGGCCACCTGGA  579

HDR-template         625  CTCCTATGAGAAAGTCACCCAGCTGCCGGGGCCCATTCGGGAGTTCCTGG  674
                          ||||||||||||||||||||||||||||||||||||||||||||||||||
Sequencing Analysis  580  CTCCTATGAGAAAGTCACCCAGCTGCCGGGGCCCATTCGGGAGTTCCTGG  629

HDR-template         675  ACCAGTACGATGCCCCGC-TGGGGGTGGAGGCTCTGTGAGCAAGGGCGAG  723
                          ||||||||||||||||||| |||||||||||||||||||||||||||||||
Sequencing Analysis  630  ACCAGTACGATGCCCCGCTTGGGGGTGGAGGCTCTGTGAGCAAGGGCGAG  679

HDR-template         724  GAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGT  773
                          ||||||||||||||||||||||||||||||||||||||||||||||||||
Sequencing Analysis  680  GAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGT  729

HDR-template         774  AAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCT  823
                          ||||||||||||||||||||||||||||||||||||||||||||||||||
Sequencing Analysis  730  AAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCT  779

HDR-template         824  ACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTG  873
                          ||||||||||||||.|||||||||||||||||||||||||||||||||||
Sequencing Analysis  780  ACGGCAAGCTGANCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTG  829
```

Nikoletta Kalenderoglou - November 2019

```
HDR-template      874  CCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAG  923
                       |||||..||||||||||||||||||||||||.||||||||||||||||||
Sequencing Analysis 830 CCCTGNNCCACCCTCGTGACCACCCTGACCTACNGCGTGCAGTGCTTCAG  879

HDR-template      924  CCGCTACCCCGACCACATGAAGAAGCACGACTTCTTCAAGTCCGCCATGC  973
                       ||||||||||||||||||||||.|||||||||||||||||||||||||||
Sequencing Analysis 880 CCGCTACCCCGACCACATGAAGAANCACGACTTCTTCAAGTCCGCCATGC  929

HDR-template      974  CCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAAC  1023
                       |||||.|||||||||||||||||||||||||||||| |.||||||||.||
Sequencing Analysis 930 CCGAANGCTACGTCCAGGAGCGCACCATCTTCTTC-ANGACGACGGCNAC  978

HDR-template      1024 TACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGA-CACCCTGG-TGAAC  1071
                       || ..||||||||||||||||||||||||||||||| ..||||| ..||.
Sequencing Analysis 979 TA-NNGACCCGCGCCGAGGTGAAGTTCGAGGGCGANNNCCCTGGNNNAAN  1027

HDR-template      1072 CGCATCGAGCTGAAGGGCATCGAC-TTCAAGGAGGACGGCAACATCCT--  1118
                       ||||||.||||| |....||||| ||||.||||...|||...|||.
Sequencing Analysis 1028 CGCATCNAGCTG-ANNNNATCGACTTTCANGGAGGNNNGCANNNTCCNNG  1076

HDR-template      1119 GGGGCACAAGCT-GGAGTACAACTACAACAGCCACAACGTCT--ATATCA  1165
                       ||||.|..|||| |||.|| ....|||||||||.| .||..|| ....|.
Sequencing Analysis 1077 GGGGNANNAGCTGGGANTA-NNNNACAACAGCNA-NACNNCTNNNNNNNCN  1124

HDR-template      1166 TGGCCGACAAGCAGAAGAACGGCATCAAGGC--TAACTTCAAGGTTCGCC  1213
                       ..|.......|.||.||.|..|..||..|||. .||.||..|||....||
Sequencing Analysis 1125 NNGNNNNANANCANAANANNGNNATNNAGGNNTNAANTTNNAGGNNNNCC  1174

HDR-template      1214 ACAACATCGAG     1224
                       .|.|.||..|
Sequencing Analysis 1175 -NANCNTCNNG     1184
```

Figure 3.13: Pairwise Sequence alignment. The length of the DNA sequence analysis derived from Sanger Sequencing is 1161 bp. The similarity between the expected DNA sequence from HDR template and the DNA sequence obtained from Sanger sequencing is approximately 91.9%.

The length of the DNA sequence analysis with one sequencing primer obtained from Sanger Sequencing was only 1161 bp. Even though the length of the sequencing analysis did not cover the whole length of the HDR repair template (2315 bp), it shows that the left homology arm linked to GFPSpark shows a high degree of homology (91.9%) between the expected DNA sequence from HDR-donor template and the results obtained from Sanger sequencing.

Following sequence analysis, both HDR-donor template and CRISPR-Cas9 harbouring gRNA1 and gRNA2 plasmids were used to be transfected into Caco2 and HIECS cells.

3.4.4 Optimization of CRISPR-Cas9 methodology

In the present study, important factor that govern the transfection efficiency were investigated by varying the method of transfection, dose of transfection reagent, the dose of CRISPR-Cas9 elements/homologous repair template, treatments with chemical compounds and cell density.

3.4.4.1 Lipid-mediated transfection

The ideal cell density is extremely dependent on cell type and requires to be optimized experimentally. The transfection efficiency was investigated by transfecting HIECS and Caco2 cells with pCMV3-SOCS3-C-GFPSpark® plasmid using a recently introduced liposomal (FuGENE HD) transfection agent. Cells were transfected with the SOCS3 fused to GFP plasmid and monitored the transient expression of GFPSpark protein with flow cytometry 48 hours post transfection.

The ideal cell density plays an important role in the efficiency of lipid-based transfection and is extremely dependent on cell type and requires to be optimized experimentally. The chosen cell confluency for HIECS at the time of transfection was 50-60%. Firstly, the ratio of reagent-to-DNA in the transfection complex was investigated. 25,000 cells were plated in a 24 well plate and on the following day, transfection was carried out using a reagent to DNA ratio of 2 to 1, 2.5 to 1, 3 to 1, 4 to 1 and finally 5 to 1. Each well was transfected with a total of 275 ng of DNA. 48 hours

post transfection, cells were analysed with a fluorescent microscope and flow cytometry.
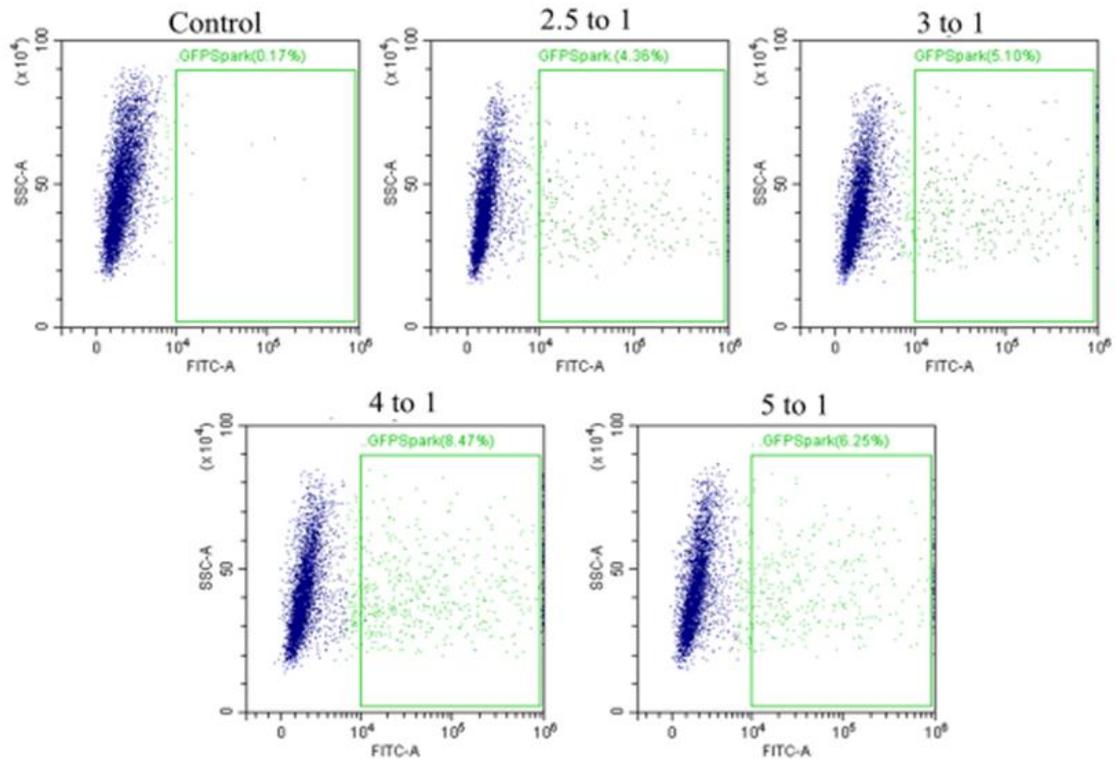
Figure 3.14: Optimization of reagent to DNA ratio. GFPSpark positive HIECS cells were determined by flow cytometry analysis approximately 48 hours after transfection. On the day of transfection, the transfection was performed using the reagent to DNA ratio of 2 to 1, 2.5 to 1, 3 to 1, 4 to 1 and 5 to 1. Control included nuclease-free water instead of DNA. Cells transiently expressing the SOCS3-GFPSpark fusion protein were examined and transfection efficiency was meassured using Flow cytometry (n=2).

Results from the flow cytometry are presented in figure 3.14. The experimental procedure was repeated twice producing similar data. According the FACS results, the transfection efficiency increases as reagent to DNA ratio increases. However, 5-1 ratio showed a decreased transfection efficiency compared to 4-1. Thus, it was determined that the reagent to DNA ratio of 4 to 1 had the most cells expressing GFPSpark.

Furthermore, since the optimal number of cells seeded on a well was found to be less than the manufacturer's recommendations (50000 cells per well), the amount of DNA complex added to the cells had to be optimized. On the day of transfection, the transfection was performed using the reagent to DNA ratio of 4 to 1 and a total DNA amount of 1.7 μg. Once the transfection complex was made, varying amounts of 25%, 50%, 75% and 100% of the complex was added to each well. Furthermore, the diluent used to form the transfection complex was also examined. The transformation complex was either formed in Opti-MEM™|Reduced Serum Medium (ThermoFisher Scientific, C# 31985062) or sterile nuclease-free water.
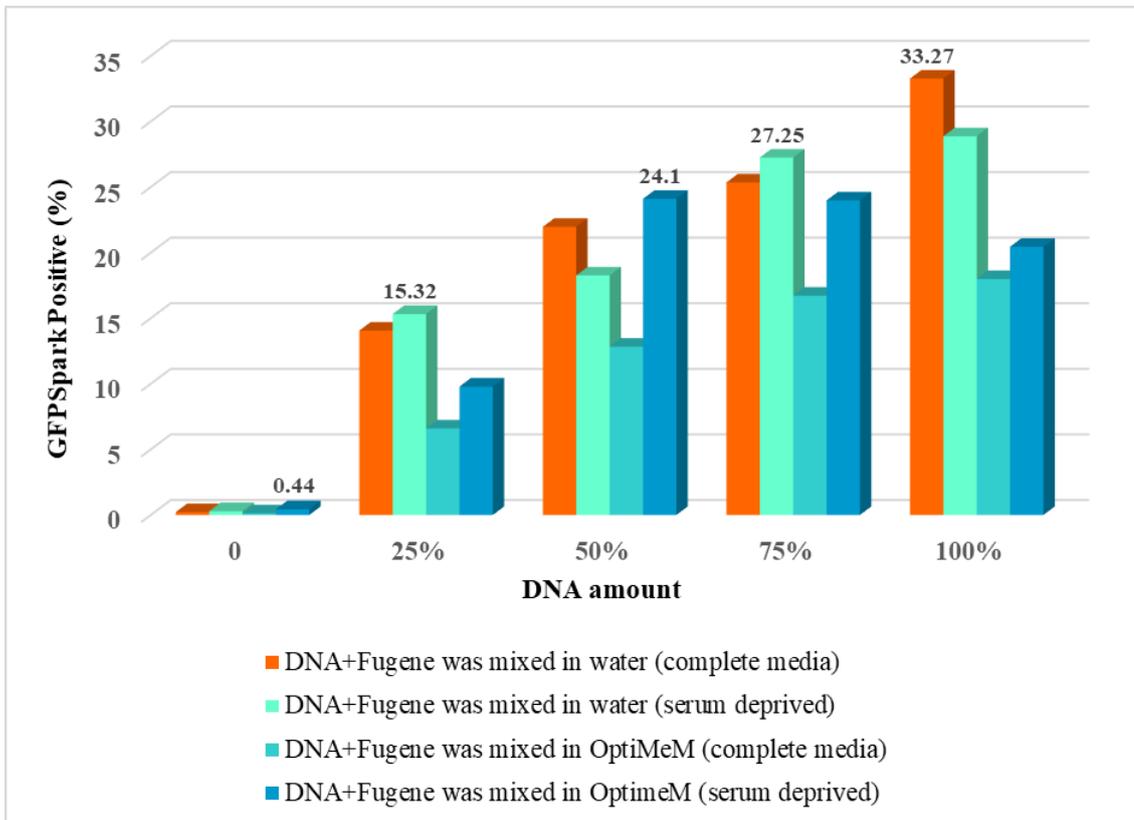
Figure 3.15: **Optimization on different parameters to enhance transfection efficiency.** GFPSpark positive HIECS cells were determined by flow cytometry analysis approximately 48 hours after transfection. On the day of transfection, a transfection complex was made and different amounts of the complex (25%, 50%, 75% and 100%) was added to cells that were either cultured in the presence of serum (complete medium) or absence of serum.

Furthermore, the transformation complex was either formed in Opti-MEM™|Reduced Serum Medium or sterile water. Cells transiently expressing the SOCS3-GFPSpark fusion protein were analysed by Flow cytometry.

Furthermore, cell synchronization by serum starvation was also tested to determine if it plays any role in influencing the transfection efficiency. Once all these parameters were set appropriately, the cells were incubated 37ºC in a 5% CO_2 for 48 hours. Cells transiently expressing the SOCS3-GFPSpark fusion protein were analysed by Flow cytometry. Results are shown in figure 3.15. As shown in figure 3.15, the highest transfection efficiency was estimated to be 33.27% positive GFP HIECS cells. This transfection efficiency was achieved by using water as a diluent to form transfection complex and cells were cultured in complete media during the transfection 48 hours period.

Additionally, according to figure 3.15, it is evident that as the amount of DNA added to the cells increase as the number of fluorescent cells also increases. Thus, it appears that the amount of transformation complex did not require adjustment even if the cell density was lower than that of manufacturer's directions (50000 cells per well, 12 well plate). Furthermore, according to the manufacturer's protocol, it is highly suggested that transfection mixture should be formed in OptiMeM solution. However, the lowest transfection efficiencies were achieved by using Opti-MEM media as a diluent and cells were synchronized with serum starvation. Additionally, according to the figure 3.15, Fugene-HD lipid-based method was not very efficient in facilitating transfection of DNA in HIECS, even when higher concentration of DNA was used.

These parameters were also optimized for Caco2 cell line. The optimal cell confluency for Caco2 cells at the time of transfection was 60-70%. To this end, 50,000 cells were

seeded in a 12 well plate prior to transfection. Since the optimal number of cells seeded on a well was found to be less than the manufacturer's protocol (100000 cells per well), the amount of DNA complex added to the cells was further optimized. On the day of transfection, the transfection was performed using the reagent to DNA ratio of 3 to 1 as it is recommended by Fugene HD online protocol for Caco2 cell line (https://www.promega.com/techserv/tools/FugeneHdTool/) and a total DNA amount of 2.5 μg. Once the transfection complex was made, varying amounts of 25%, 50%, 75% and 100% of the complex was added to each well. Furthermore, the diluent used to form the transfection complex was nuclease-fee water as it showed the highest transfection efficiency in HIECS. 48 hours post transfection, the cells were studied by a confocal microscope or subjected to flow cytometric analysis to determine the percentage of Green fluorescent positive cells. GFPSpark plasmid uptake into Caco2 cells was determined by Flow cytometric analysis of DNA FITC positive cells and expressed as % of all living cells post-transfection. Data analysis from flow cytometry. Results form the flow cytometric analysis showing the transfection efficiency in Caco2 cells are exhibited in figure 3.16.
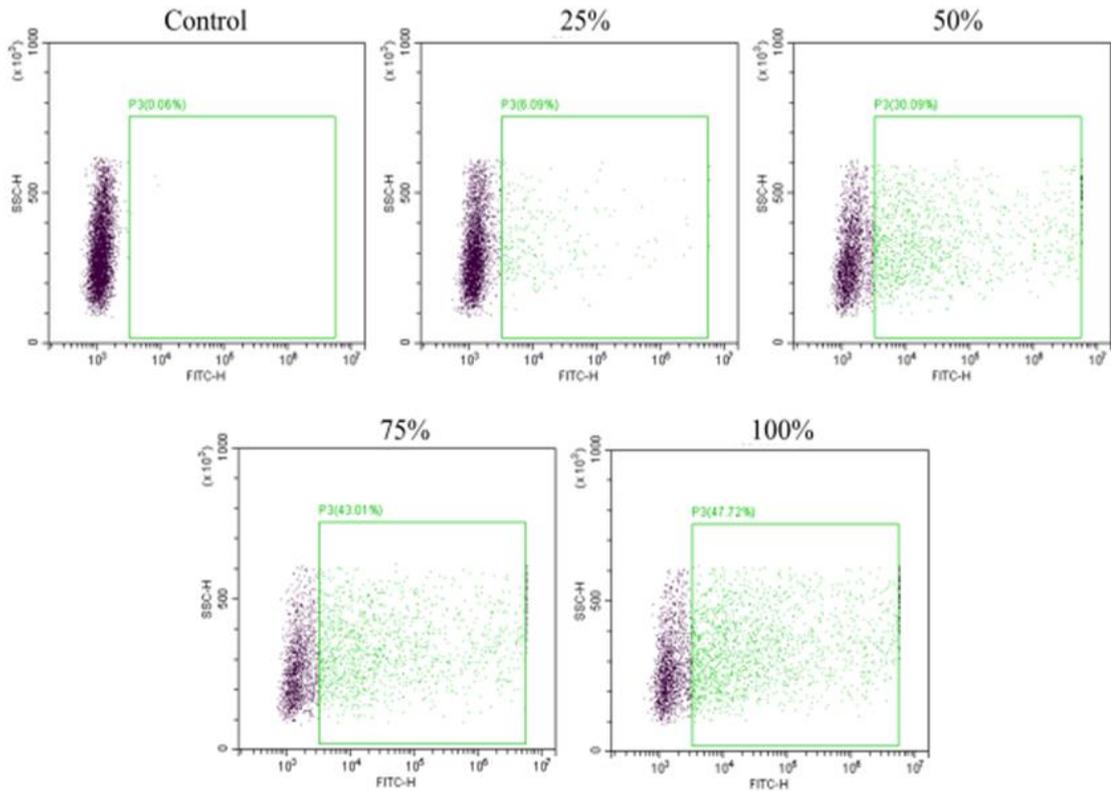
Figure 3.16 Optimization on DNA amount transfected into Caco2 cells. GFPSpark

positive Caco2 cells were determined by flow cytometry analysis approximately 48 hours

after transfection. On the day of transfection, a transfection complex was made and different

amounts of the complex (25%, 50%, 75% and 100%) was added to cells. Non-transfected

Caco2 cells served as reference to set the gate. The green boxes indicate the gate that define

GFP positive cells. (n=2)

According to the flow cytometric results presented in figure 3.16, when Caco2 cells

were transfected with 25% of the transfection complex, cells exhibited a transfection

efficiency of 5.92% $\pm$ 0.18%, whereas when cells were transfected with 50% of the

complex, they showed a transfection efficiency of 30.25% $\pm$ 0.16%. Furthermore, when

cells were transfected with 75% of the complex presented an efficiency of 44.08% $\pm$

1.07%, whereas when 100 % of the complex was used for transfection, it established

the highest cell transfection efficiency reported as 47.35% $\pm$ 0.37%.

To assess the frequency of integration events in Caco2 cells, cells were transfected with equal amounts of the donor Cas9 plasmid with the gRNA1 and the donor repair template. Cells were kept in complete medium for 72 hours at 37ºC in a 5% CO_2 incubator. Following incubation, cells were studied by an epifluorescent microscope or subjected to flow cytometric analysis to determine the percentage of Green fluorescent positive cells. Furthermore, the CRISPR plasmid encoding the Cas9 plasmid also incorporates a CD4 reporter, which allows bead-based enrichment and tracking the transfection efficiency using anti-CD4 fluorescent antibodies. Thus, the transient expression of CD4 reporter at the cell surface was also monitored 48 hours after transfection. Cells were gated and counted according to their FITC or PE positivity and negativity for GFP or CD4 respectively. Firstly, the inspection of the cells treated with CRISPR-Cas9 under the epifluorescent microscope did not show any significant GFP fluorescence that could not be explained by cell autofluorescence. Furthermore, cells appeared morphologically normal, but there was evidence of a small percentage of cell death and the cell population did not appear to grow in the same rates to cells that were not transfected with CRISPR-Cas9. Secondly, flow cytometric analysis monitoring the expression of CD4 and GFP are presented in figure 3.17. CD4 expression has not been reported in Caco2 cells. However, nonspecific CD4 binding was blocked with FC reporter blocking solution and it was also monitored by treating the cells with IgG1 isotope control. According to the data from figure 3.17, the isolated cell population in both the control cells, which were not subjected to CRISPR-Cas9 treatment and the tested samples showed a low expression of CD4 surface marker. Specifically, control cells exhibited a 2.33% positivity and CRISPR treated cells showed 2.83% of CD4 positivity.
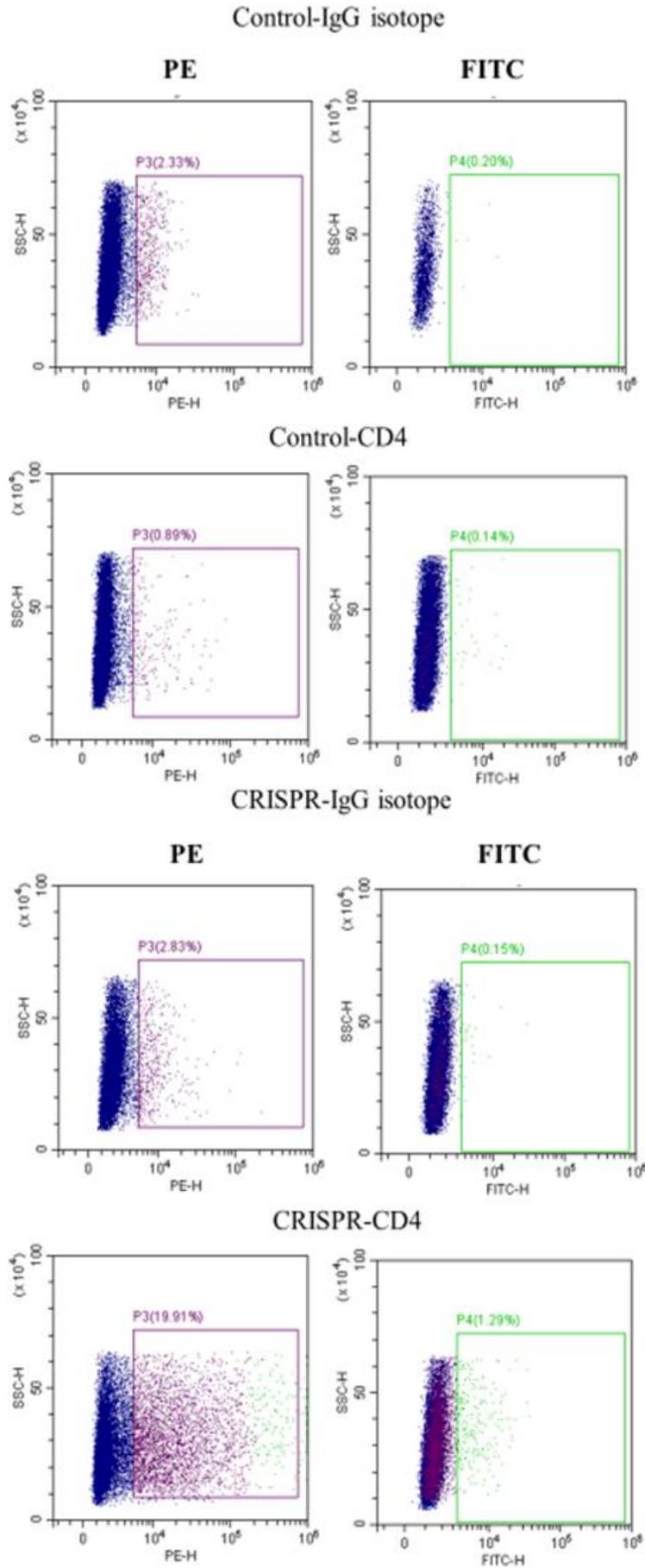
Figure 3.17: GFPSpark integration events monitored by FACS in Caco2 cells. Expression of GFPSpark and CD4 at the cell surface was tested 72 hours post transfection with CRISPR-Cas9 plasmid by staining with PE-CD4 antibody. A side scatter height (SSC-H) vs. forward scatter height (FSC-H) dot plot was used to exclude debris and dead cells and the gated cells were then analysed in SSC-H vs PE-H for the expression of CD4 reporter and SSC-H vs FITC-H for the expression of GFPSpark. Nonspecific CD4 binding was monitored by treating the cells with IgG1 isotope control.

Caco2 cells were only transfected with donor template and the CRISPR-cas9 harbouring gRNA1 to test the transfection efficiency and efficiency of integration events. Expression of CD4 at the cell surface of Caco2 cells showed an approximate 20% of the population transfected with CRISPR-Cas9 plasmid and the donor repair template. The low percentage of cells expressing the CD4 surface marker could be explained by the fact that only half of the transfection complex solution contained the CRISPR plasmid whereas the other half of the complex included the donor repair template. Approximately, 1.3% of the CRISPR-gRNA1 and donor template transfected cell population exhibited GFPSpark expression, indicating possible integration of the GFPSpark cassette into SOCS3 CDS.

Importantly, it is worth noting here that low to very low-abundance FP tagged proteins that are driven by their endogenous promoters can be challenging to distinguish with the use of flow cytometry. Previous research articles have successfully established high transfection efficiency of expression Cas9 vectors and the donor template but they have only managed to achieve successful recombination events in a very small cell populations with reporting efficiencies as low as 0.1% [242, 317]. Moreover, cell gating under these circumstances is difficult because placing the FITC gate at close proximity to the whole population increases the detection of false positive FITC cells. This is

evident from the results reported from previous research articles as they demonstrated that post FACS sorting, only a small proportion of positive sorted cells actually exhibited true FP expression.

Cells transfected with CRISPR complex were subjected to CD4 bead-based enrichment and single cells were transferred onto 96-well plates. Cells were allowed to proliferate for 10 days prior to their investigation under confocal microscope. From the single cell cloning step, only 50% of cell clones achieved to proliferate and generate single cell clonal populations. It was apparent that bead-based enrichment step was toxic to Caco2 cells and they were unable to survive and proliferate. The inspection of the single cell clones under the confocal microscope did not display any significant GFP fluorescence that was differentiated from cell autofluorescence. This could probably be explained by the fact that confocal microscope not being able to detect endogenous driven GFPSpark protein. Furthermore, the percentage of positive cells detected by FACS could be the result of spill over of fluorescent signal from PE channel.

Therefore, the CRISPR complex transfection was repeated in Caco2 cells but the step of flow cytometry analysis was avoided to minimize the time that cells stayed out of the optimum cell culture conditions. After CD4 bead-based enrichment, single cells were seeded onto 96 well black polystyrene Microplates with clear flat bottom. Cells were allowed to proliferate for 10 days prior to their investigation under confocal microscope. From the single cell cloning step, it was apparent that there was a slight increase of cell survival after bead-based enrichment, but confocal microscope did not demonstrate any significant GFP fluorescence.

Expression of GFPSpark and CD4 at the cell surface was also studied in HIECS. However, since the transfection efficiency of Caco2 cells did not yield any successful

cell clone to express SOCS3 endogenously GFPSpark protein, a more efficient transfection method had to be used for HIECS other than lipid-based method.

3.4.4.2 Optimizing transfection efficiency in HIECS

Several biologically and physiologically relevant cell lines, including progenitor cells, do not demonstrate high transfection efficiency using lipid-based methods. The transfection efficiency for Caco2 was quantified to be approximately 48%. According to manufacturer's protocol, the transfection efficiency to be expected for Caco2 cells was approximately 55-65%. However, the transfection efficiency for HIECS was measured to be 33%. This low transfection efficiency could be the reason that CRISPR assay did not yield a detectable number of cells endogenously expressing GFPSpark. Thus, any improvement in the efficiency of DNA transfection would facilitate the increase of likelihood to achieve large DNA modification. Further optimizing of transfection efficiency in HIECS was necessary. An alternative method used for DNA uptake into cultured cells is electroporation. Electroporation has been used with varying success in different cell lines, but it allows the efficient transfer of exogenous DNA to a large number of cells. To this end, uptake of pCMV3-SOCS3-C-GFPSpark® plasmid into HIECS was used as a means to quantify the electroporation efficiency. Furthermore, the kit used for DNA electroporation in HIECS was Amaxa® Cell Line Nucleofector® Kit L (Lonza, C# VCA-1005). The kit also provided a pmaxGFP® Vector to monitor transfection efficiency. This vector drives the expression of an enhanced green fluorescent protein (maxGFP) under the control of the CMV promoter. In order to establish if electroporation is a better method than FuGene HD transfection, both plasmids were electroporated into HIECS as detailed in methods. Forty-eight hours after electroporation, cells were subjected to flow cytometric analysis to determine the

percentage of Green fluorescent positive cells. As visible in figure 3.18, electroporation yielded a transfection efficiency higher than that one obtained by the use of FuGene HD reagent. The method exhibited a moderate reproducibility, as established by comparing efficiencies from several replicate experiments.
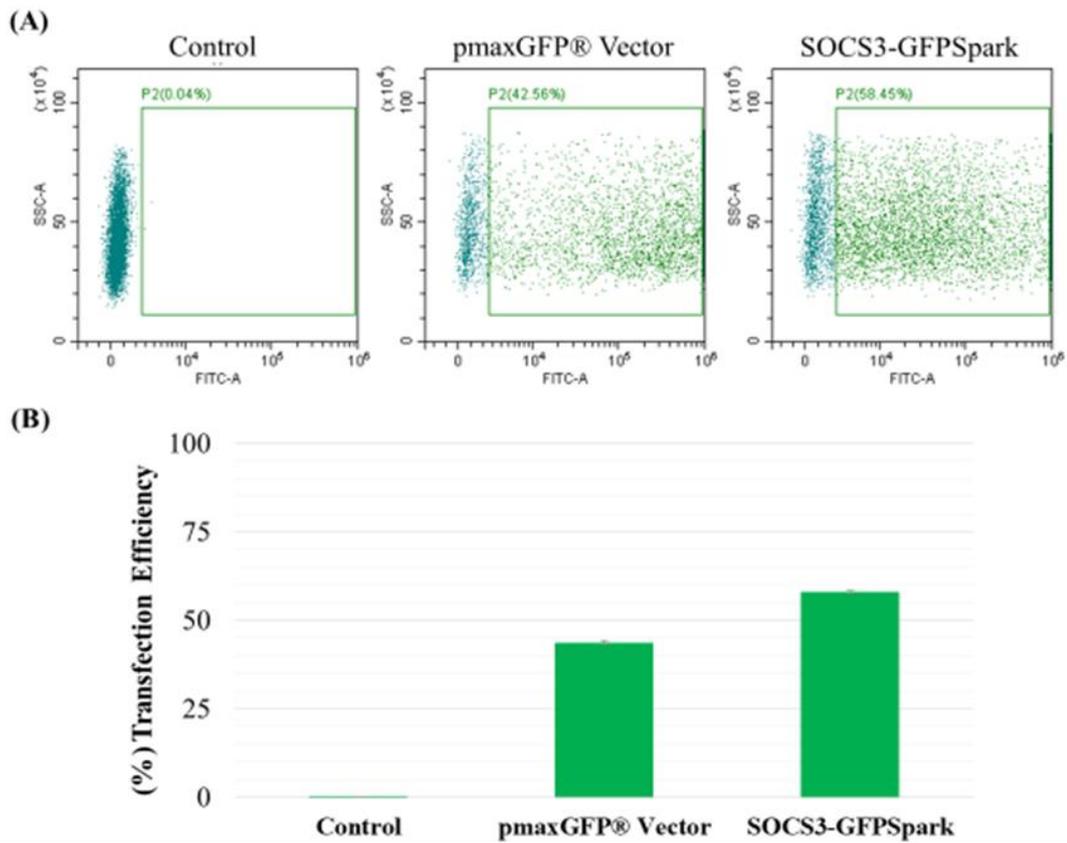
Figure 3.18: Optimization of HIECS transfection by electroporation. A side scatter area (SSC-A) vs. forward scatter area (FSC-A) dot plot was used to exclude debris and dead cells and the gated cells were then analysed in SSC-A vs FITC-A for the expression of green fluorescence. **(A)** Representative FACS dot plots showing transfection efficiency for control, pmaxGFP® Vector and SOCS3-GFP Spark vector. **(B)** Column chart showing the electroporation efficiency of control, pmaxGFP® Vector and SOCS3-GFP Spark vector measured from Flow cytometry (n=3). Data represent mean ± SEM obtained from triplicates.

Specifically, flow cytometric analysis showed that HIECS exhibited a transfection efficiency of 43.62% ± 0.54% when transfected with pmaxGFP® Vector and 58.01% ± 0.34% when transfected with SOCS3-GFPSpark. Even though pmaxGFP® Vector is a relatively small plasmid (3486 bp) compared to SOCS3-GFPSpark (7525 bp), it exhibited a lower transfection efficiency than SOCS3-GFPSpark vector. This could probably be explained that SOCS3-GFPSpark plasmid exhibits a more supercoiled structure than pmaxGFP® plasmid, as plasmids with a supercoiled structure are associated with high transfection efficiencies.

To assess the frequency of mediated homologous recombination in HIECs, cells were transfected with equal amounts of the donor Cas9 plasmid harbouring gRNA1 or gRNA21 and the repair molecule. Cells were kept in complete medium for 7 days at 37ºC in a 5% CO_2 incubator. However, since it has reported by many research laboratories that HDR efficiency following CRISPR/Cas9-mediated DSB is very inefficient (as low as 0.1%) in relation to the higher efficiency of NHEJ in mammalian cells (which can generate up to 100%) [317], we also attempted to test alternative DNA methods to increase this efficiency. Previously, Liang et al. [305] observed that sequential delivery of Cas9 RNPs and donor template increased HDR events. This observation could possibly be explained by the fact that when Cas9 complex is paired with the donor template results in a decreased transfection efficiency. Thus, we attempted to test this approach and investigated whether a sequential delivery could increase the efficiency of homologous recombination. For sequential delivery of the bicistronic Cas9/gRNA1 plasmid and repair template, 10^6 HIECS cells and 2 µg of bicistronic Cas9/gRNA1 or gRNA2 plasmid were resuspended carefully in 100 µl room-temperature Nucleofector® Solution L and electroporated using the T-020 program. The electroporated cells were then transferred to 500 µl of DPBS. Upon

centrifugation at 2000 x g for 5 minutes, the supernatant was carefully removed, and the cell pellet was resuspended in 100 µl room-temperature Nucleofector® Solution L, followed by the addition of 2 µg of donor repair molecule. The final cell mixture was used for electroporation using the same instrument settings. Upon electroporation, the cells were seeded into 3 wells of a 12-well plate containing 1 ml culture media. For sequential delivery, the viability of HIECs was below 50%. Cells were allowed to recover for 7 days with fresh medium replaced every two days, prior to their investigation with flow cytometry. Results from the flow cytometry analysis are shown in figure 3.22.
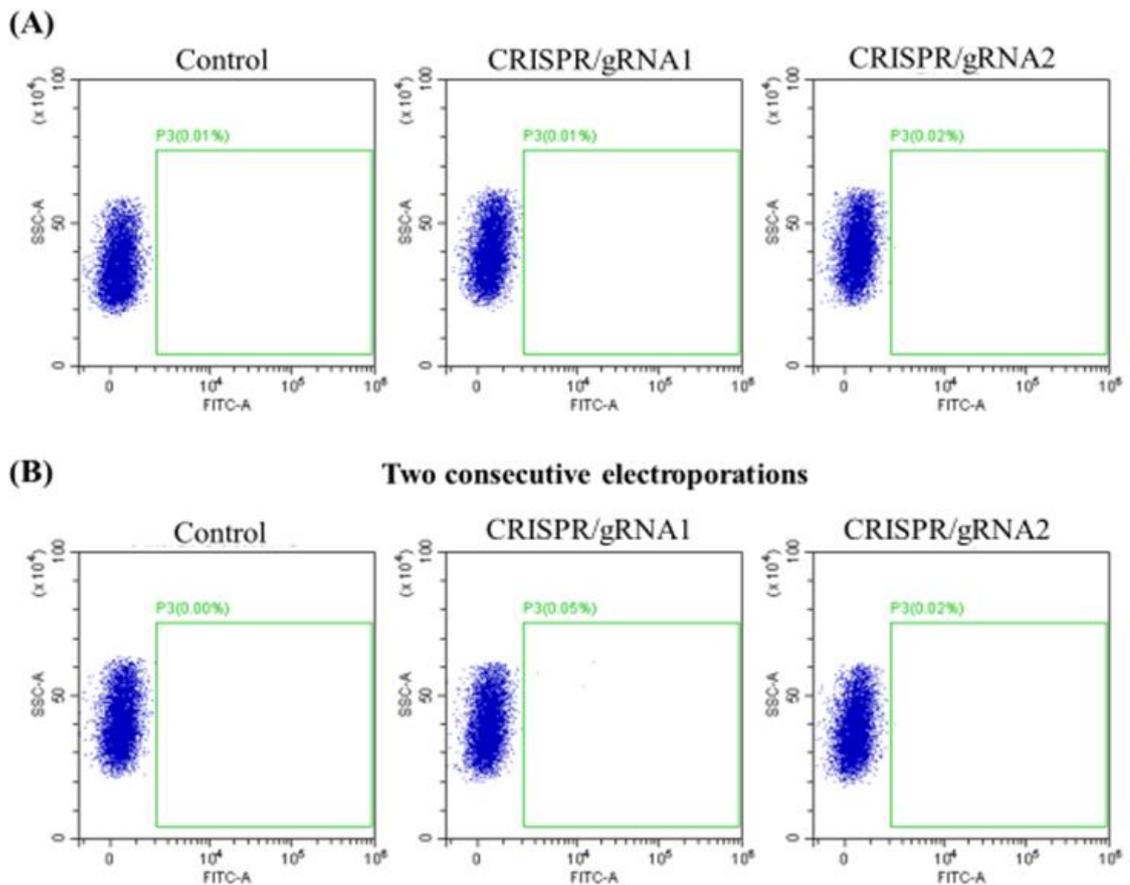
Figure 3.19: **Assessment of CRISPR mediated-HDR efficiency by flow cytometry. (A)** HIECS were transfected with bicistronic CRISPR/gRNA1 or CRISPR/gRNA2 construct and the donor repair template (CRISPR). Seven days after transfection the amount of FITC-GFPSpark positive cells (given in %) was determined using FACS. Threshold was determined

based on FACS of untransfected HIECs (Control). **(B)** Effects of sequential delivery of bicistronic construct and the donor molecule for HDR efficiency. SSC-A vs FSC-A dot plots were used to exclude debris and dead cells and the gated cells were then analysed in SSC-A vs FITC-A for the expression of GFPSpark. The green boxes indicate the gate that define GFPSpark positive cells (n=3).

According to the data from figure 3.19 untransfected HIECs (control) did not exhibit any green fluorescence in both set of experiments. Moreover, this percentage was not changed in HIECs transfected with the bicistronic CRISPR/gRNA1 or CRISPR/gRNA 2 plasmid (0.01% and 0.02%). Additionally, sequential delivery did not facilitate any improvement in HDR efficiency. The absence of positive GFPSpark cells as assessed by flow cytometry could be due to multiple factors. For instance, even though the transfection efficiency in HIECs was assessed by the transient expression of green fluorescent protein from two plasmids, that doesn't necessarily represent the true transfection efficiency of the bicistronic Cas9/gRNA plasmid. However, Cas9/gRNA vector used for this thesis incorporates a CD4 marker that allows to track the transfection efficiency using CD4 fluorescent antibodies.

Therefore, a repeated transfection on the same cells following identical process was performed.

Specifically, HIECs were transfected with equal amounts of the donor Cas9 plasmid with the gRNA1 and the donor repair template. Cells were kept in complete medium for 72 hours at 37ºC in a 5% CO_2 incubator. Following incubation, cells were trypsinized, washed twice with DPBS and a third of the cell sample was kept in ice to investigate the expression of CD4 and GFPSpark prior to bead-based enrichment (figure 3.20). The rest of the cells were subjected to CD4 enrichment. During the CD4

enrichment procedure, cells that did not bind to CD4 beads were collected and cultured for further study. The rest of the cells were seeded in a 6 well plate and allowed to proliferate. Results showing the expression of CD4 and GFPSpark (presented in %) prior to CD4 enrichment are presented in figure 3.20. Nonspecific CD4 binding was blocked with FC reporter blocking solution and it was also monitored by treating the cells with IgG1 isotope control. According to the data from figure 3.20, untransfected HIECS demonstrated a low IgG isotope binding affinity (0.3%). However, this percentage was slightly increased in HIECs transfected with the bicistronic Cas9/gRNA1 plasmid and repair template (1.17%). Furthermore, untransfected HIECs stained with CD4 antibody also exhibited a low percentage of cells detected as CD4 positive (0.11%). On the other hand, HIECs transfected with Cas9/gRNA1 and repair template showed an increased level of CD4 positive cells (12.84%). This low percentage of cells expressing the CD4 reporter could be explained by the fact that only half of the transfection complex solution contained the Cas9/gRNA vector while the other half of the complex included the repair template.

Control-IgG isotope

Control-CD4

CRISPR-IgG isotope

CRISPR-CD4
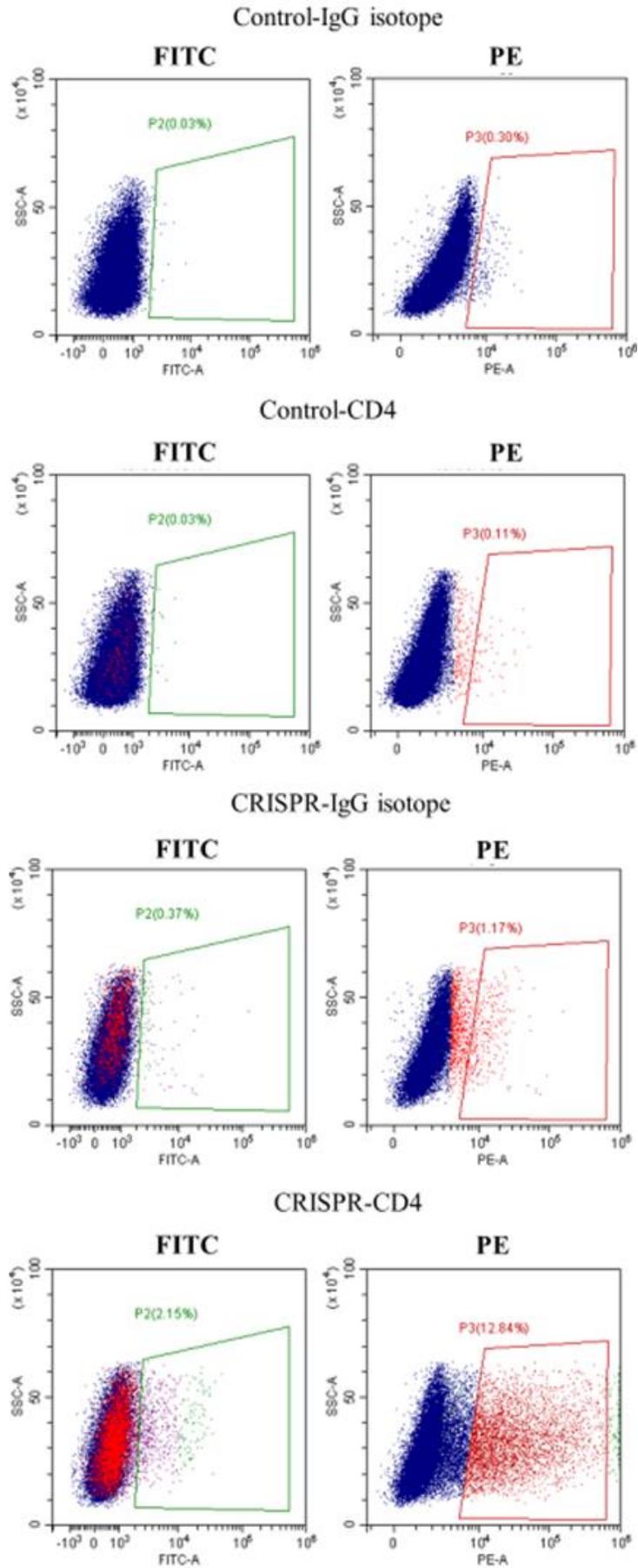
Figure 3.20: GFPSpark positivity as assessed by flow cytometry in HIECS after

transfection. HIECS were transfected with bicistronic Cas9/gRNA1 construct and the donor repair template (CRISPR). Three days after transfection the amount of FITC-GFPSpark and PE-CD4 positive cells (given in %) was determined using FACS. Threshold was determined based on FACS of untransfected HIECs (Control). For each test sample, 30,000 cells were analysed. SSC-A vs FSC-A dot plots were used to exclude debris and dead cells and the gated cells were then analysed in SSC-A vs PE-A for the expression of CD4 reporter and SSC-A vs FITC-A for the expression of GFPSpark.

Approximately, 2.15% of HIECs transfected with CRISPR-gRNA1 and donor template appeared to show GFPSpark expression, indicating possible integration of the GFPSpark cassette into SOCS3 CDS.

To assess the effects of CD4 bead enrichment on detecting more efficiently positive GFPSpark cells, cell aliquots collected during the CD4 enrichment experiment were left to proliferate for 6 days. Following incubation, cells were trypsinized, washed twice in DPBS and half of the cells were analysed by Flow cytometry and the other half was further subcultured into T75 flasks.

(A) Negative control cells

(B) Cells that did not bind to the CD4 beads

(C) CRISPR cells post CD4 enrichment
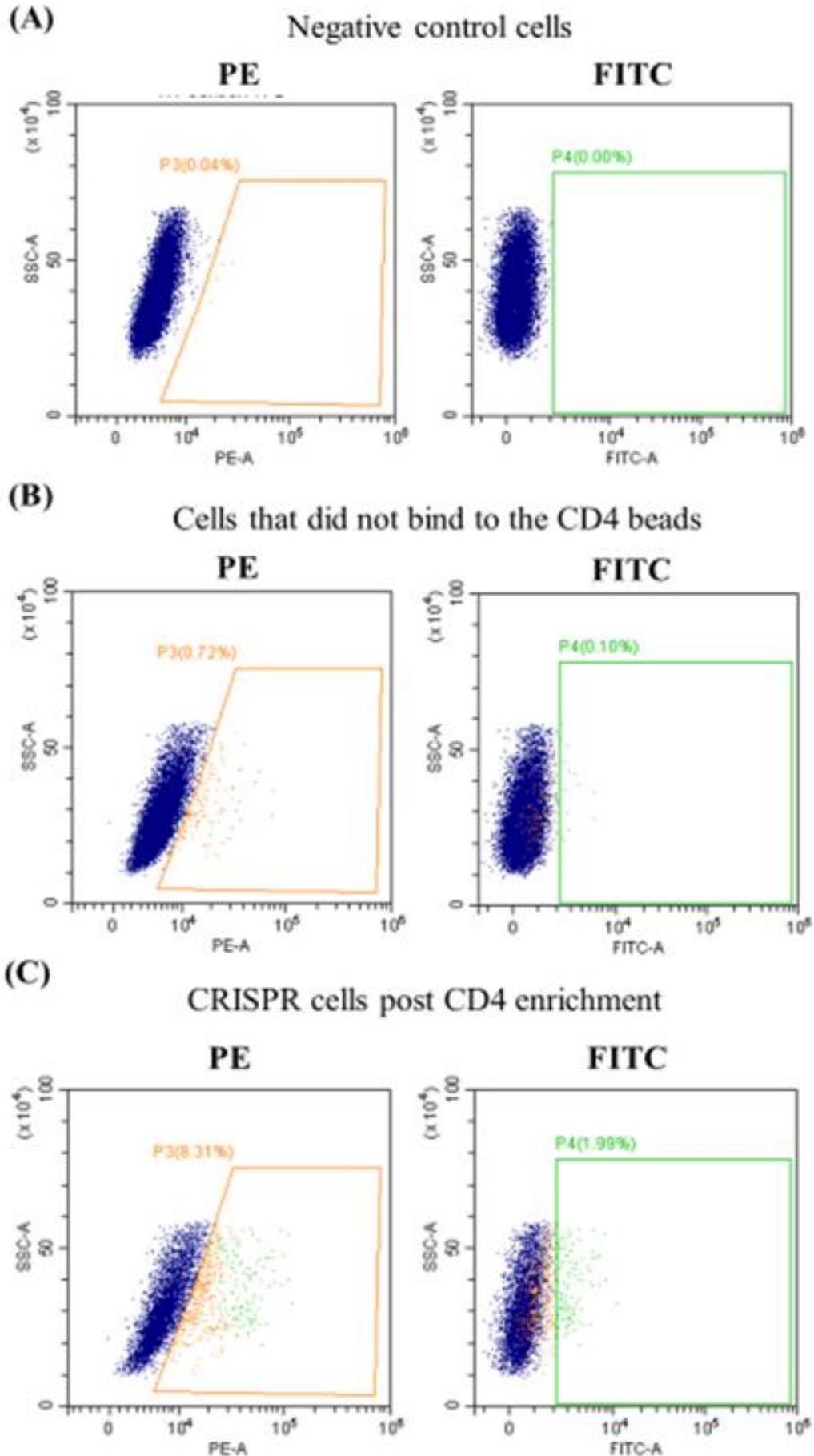
Figure 3.21: Effects of CD4 bead enrichment on isolating positive GFPSpark cells. To assess the effects of CD4 bead enrichment on detecting positive GFPSpark cells, cell aliquots were collected post CD4 purification experiment and used for flow cytometry analysis. The amount of FITC-GFPSpark and PE-CD4 positive cells (given in %) was determined using FACS. Threshold was established based on FACS of untransfected HIECs (Negative control). For each test sample, 30,000 cells were analysed. Population of cells were gated for FSC-W (width) vs. FSC-A to exclude debris and nonproliferating cells. Gated cells were analysed based on their expression of CD4 with the PE channel and GFPSpark with the FITC channel. (A) untransfected HIECs (Negative control), (B) CRISPR transfected cells that did not bind to CD4 beads, (C) CRISPR transfected cells that were enriched for CD4 expression at their cell surface.

Results from the flow cytometry analysis are shown in figure 3.21. According to the results from figure 3.21, untransfected HIECS demonstrated a very low expression of CD4 reporter and none GFPSpark protein. Moreover, cells that did not bind to CD4 beads during the enrichment procedure also demonstrated a low expression of CD4 marker (0.72%) and a low expression of GFPSpark protein (0.10%). On the other hand, HIECs that were purified from CD4-bead experiment maintained a GFPSpark expression as previously detected by FACS (figure 3.21). Specifically, 1.99% of the enriched CRISPR cell population exhibited GFPSpark expression. Additionally, the bead-purified CRISRPR cells demonstrated a lower CD4 expression at their cell surface (8.31%) compared to the CRISPR cells analysed immediately after the CD4 bead purification experiment (12.84%). This could be possibly explained by the fact that nucleases within CRISPR cells begun to degrade the Cas9/gRNA1 plasmid or diluted by cell division.

Since cell cloning by serial dilution did not succeed to distinguish a cell endogenously expressing GFPSpark protein, CRISPR-HIECs were cultured for 14 days and then Dr John Worthington analysed the cells by FACS sorting depended on their GFP positivity at the Flow Cytometry Facilities, University of Manchester (UK). Results from FACS sorting analysis are shown in figure 3.22.
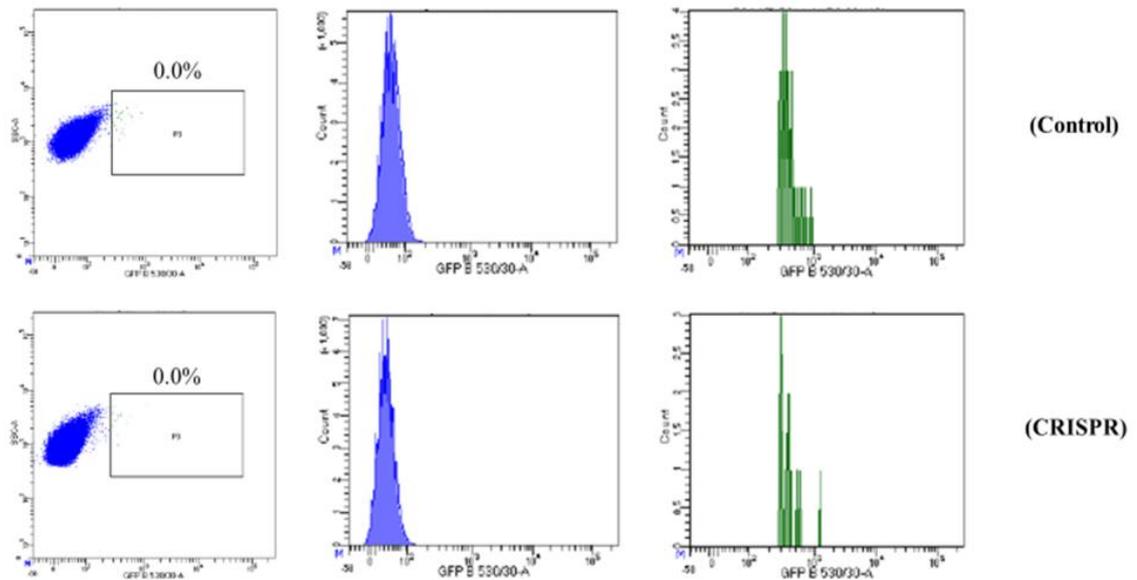
Figure 3.22: Fluorescence activated cell soring (FACS) dot plots and histograms for HIECs. 25 days after transfection and 14 days after CD4 bead-based purification, the amount of GFPSpark positive cells (presented in %) was analysed using FACS. Threshold was chosen based on FACS sort of untransfected HIECs (Control). Population of cells were gated for FSC-W (width) vs. FSC-A to exclude debris and nonproliferating cells. Gated cells were then analysed based on their expression GFPSpark with the GFP channel. Blue dot: no fluorescent cells, green dots: green fluorescent cells. FACS analysis was performed by Dr John Worthington.

FACS sorting analysis did not yield any successful HIEC clone that endogenously expressed detectable GFPSpark signal. This could possibly be explained by the fact that the low percentage of HIECS that initially presented green fluorescent signal got diluted by cell division during the culturing period of two weeks.

Furthermore, at the same time that CRISPR-HIECs were prepared for FACS sorting, Caco2 cells were also transfected with the bicistronic Cas9/gRNA1 construct and the donor repair template (CRISPR). After 72 hours, Caco2 cells were purified by CD4 bead-based protocol and the purified Caco2 cells were left to proliferate for a week. Dr John Worthington analysed the cells by FACS sorting depended on their GFP positivity at the Flow Cytometry Facilities, University of Manchester (UK). Results from FACS sorting analysis are shown in figure 3.23.
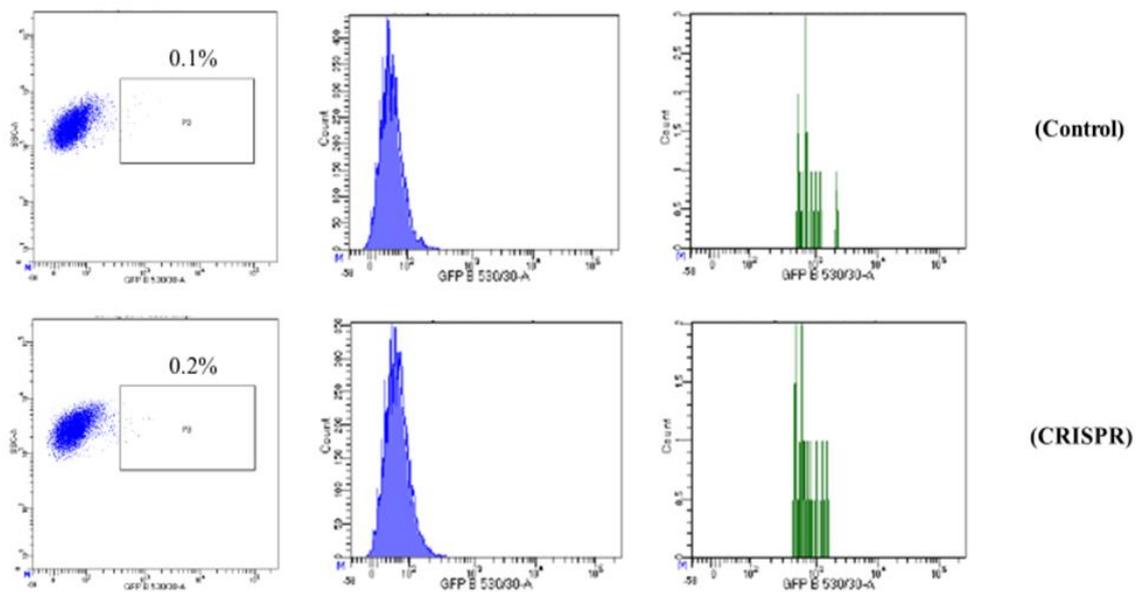
Figure 3.23: **Fluorescence activated cell soring (FACS) dot plots and histograms for Caco2 cells.** The amount of GFPSpark positive Caco2 cells (presented in %) was analysed using FACS. Threshold was chosen based on FACS sort of untransfected Caco2 (Control). Population of cells were gated as described in figure 3.22. Gated cells were then analysed based on their expression GFPSpark with the GFP channel. Blue dot: no fluorescent cells, green dots: green fluorescent cells. FACS analysis was performed by Dr John Worthington.

According to the FACS sorting analysis, there was only 0.1% of CRISPR-Caco2 cells that exhibited a positive GFP signal. Those cells were collected, and single cells were seeded onto 96 well black polystyrene Microplates with clear flat bottom. Cells were

allowed to proliferate for 14 days prior to their investigation under confocal microscope. Following cell culturing for 2 weeks, confocal microscope did not demonstrate any significant GFP fluorescence from any FACS sorted Caco2 cell clone. Collectively, the outcome of the FACS sorting and subsequent the confocal imaging strengthened the possibility that the percentage of positive cells detected by FACS prior to FACS sorting could be the result of spill over of fluorescent signal from PE channel to the FITC channel.

In agreement with previous research articles, the efficiency of repairing Cas9 induced DSBs by HDR was probably relatively low [281]. However, during the past decade, various approaches have been explored to improve the repair efficiency catalysed by HDR for genome engineering. These approaches mostly involved the manipulation of the cell cycle and the regulation of various protein expressions that belong to important repair processes. Therefore, we attempted to increase HDR events by utilizing these approaches and further optimizing the CRISPR editing system on HIECs and Caco2 cells.

3.4.5 Chemical compounds stimulating HDR events at Cas9-generated DSB sites

3.4.5.1 Nocodazole treatment

Another strategy to stimulate HDR events takes advantage of the fact that cell repair system is dependent on the cell cycle state. Although cells demonstrate varying abilities to repair DSBs using NHEJ or HDR, the stage of the cell cycle highly directs the option of repair pathway. Specifically, NHEJ is known to dominate DNA repair pathways during G1, S and G2 phases of the cell cycle, while HDR activity is restricted to late S and G2 phase, when the process of genomic duplication is completed, and sister

chromatids are accessible to serve as repair templates [318]. Previous studies have used various well-established cell cycle synchronization approaches to enhance CRISPR editing efficiency by achieving controlled Cas9 nuclease action at the stage of the cell cycle best for HDR. These included cell treatments with Nocodazole, Thymidine, and Aphidicolin [319].

Thus, we attempted to test this approach and investigated whether cell cycle arrest at late S and G2 phase caused by Nocodazole treatment will increase HDR in this system. The ideal compound concentration that will not cause cell toxicity and produce maximal cell cycle arrest was initially investigated in HIECS. To this end, HIECS were seeded in 6 well plates so that cell confluency was 70% at the day of treatment. The following day, cells were treated with Nocodazole (Sigma-Aldrich, C# SML1665-1ML) in varying concentrations and cells were incubated for 24 hours at 37 °C. The following day, cells were fixed in ice cold ethanol and cell cycle arrest was investigated using Propidium Iodide staining and flow cytometry. Results from Flow cytometry were analysed and presented in figure 3.24.

According to the data exhibited in figure 3.24, cell cycle synchronization confirmed by FACS analysis demonstrated that increasing concentration of Nocodazole also increase the percentage of cells arrested at G2/M phase.

(A)

Sample	G0/G1	S	G2/M	Count
Control	54.05	16.8	23.25	3371
200 ng/ml Nocodazole	7.43	3.52	77.8	4323
400 ng/ml Nocodazole	13.7	2.55	77	4384
800 ng/ml Nocodazole	6.42	2.85	80.8	3751
1200 ng/ml Nocodazole	1.85	1.65	87.4	3997

(B)
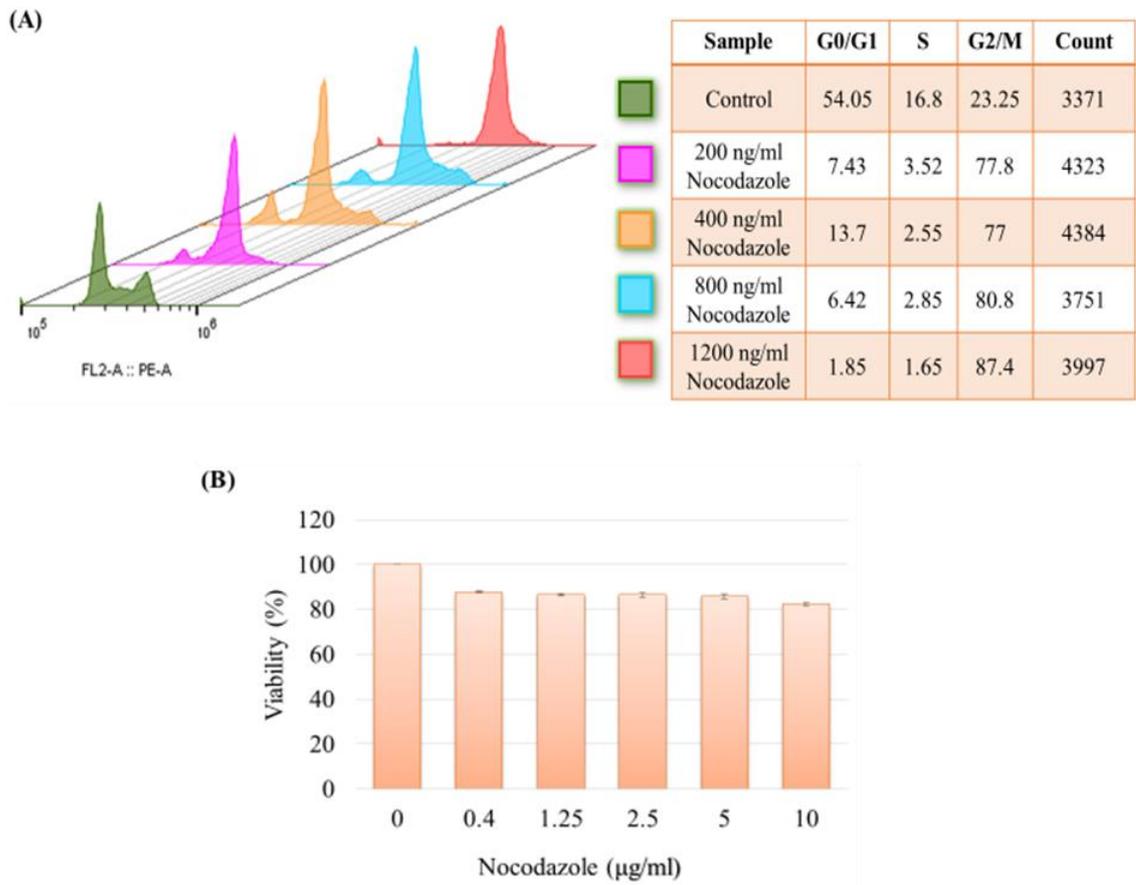
Figure 3.24: Analysis of cell cycle and cell viability in HIECS treated with Nocodazole.

(A) FACS analysis reveals that different Nocodazole concentrations (ng/ml) generated different percentages of cells that are arrested at G2/M phase. Overlaid histograms showing cell count (y-axis) versus PE fluorescent signal as detected on FL 1 channel (x-axis). Overlaid histograms with representative colours indicate the control (sample treated with only DPBS) and Nocodazole treated samples. On the right, the table indicates the cell percentages arrested at specific phases of the cell cycle. **(B)** The percentage of viable cells after treatment of different concentrations of Nocodazole was measure by PrestoBlue™ Cell viability assay. Data represent mean ± SEM (n=4).

Specifically, when cells were treated with the maximum concentration of 1200 ng/ml, 87% of cells were arrested at G2/M phase, 1.85% of cell at G0/G1 and 1.65% of cells at S phase. By contrast, cells that were treated with just DPBS showed that 54.04% of cells were found at G0/G1 phase, 16.8% of cells at S phase and 23.25% of cells at G2/M

phase. Furthermore, cells in all treatment conditions appeared morphologically normal. To further characterize the biological effects of Nocodazole at the cellular level, cell viability was measured by PrestoBlue cell viability assay (figure 3.24 (B)). Nocodazole had only minor effect on the growth or survival of HIECS under normal culture conditions after 48 hours of treatment.

To test whether synchronization of HIECs at G2/M phase is increasing HDR events, HIECs were treated with 1 μg/ml Nocodazole for 1 day, followed by release prior to electroporation with Cas9/gRNA and repair template. Cells were kept in complete medium for 72 hours at 37ºC in a 5% CO_2 incubator. Following incubation, cells were trypsinized, washed twice with DPBS and a third of the cell mixture was kept in ice to measure the expression of CD4 and green fluorescence prior to bead-based purification. The rest of the cells were subjected to CD4 enrichment. After CD4 based purification, cells were returned to the incubator and left for 48 hours to recover. Following incubation, cells were subjected to flow cytometry and results are presented in figure 3.25.

(A) CRISPR cells

Prior CD4 bead based enrichment
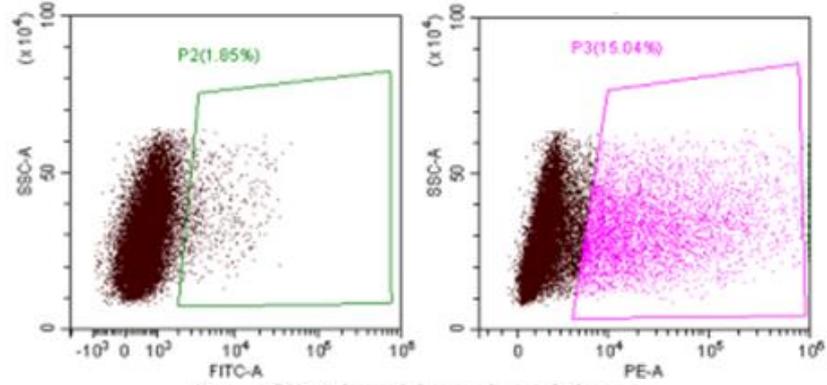
Post CD4 bead based enrichment
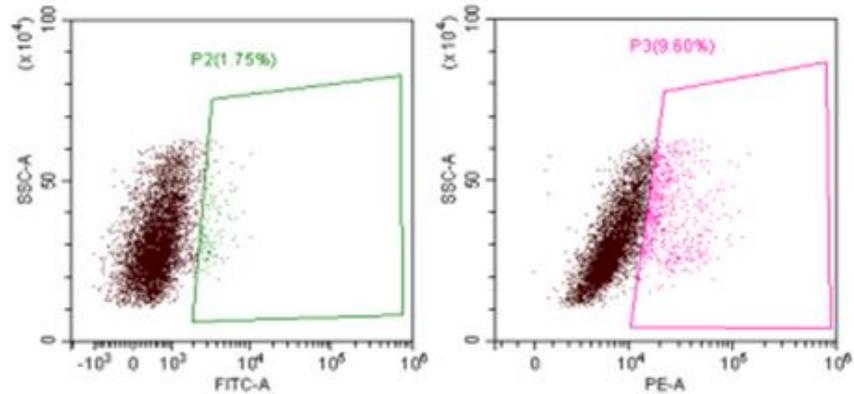
(B) Nocodazole treated CRISPR cells

Prior CD4 bead based enrichment
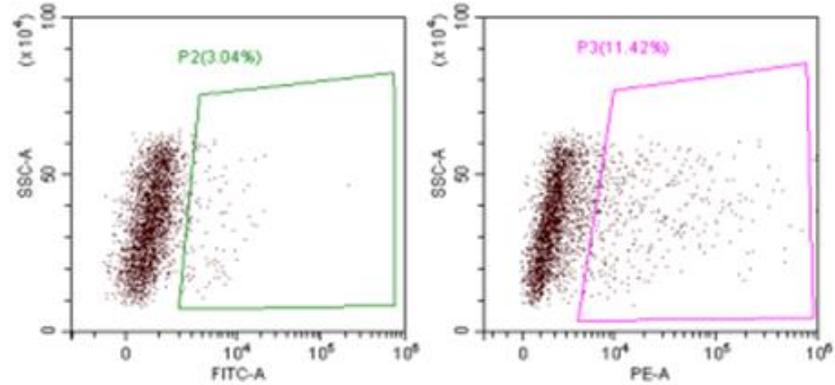
Post CD4 bead based enrichment
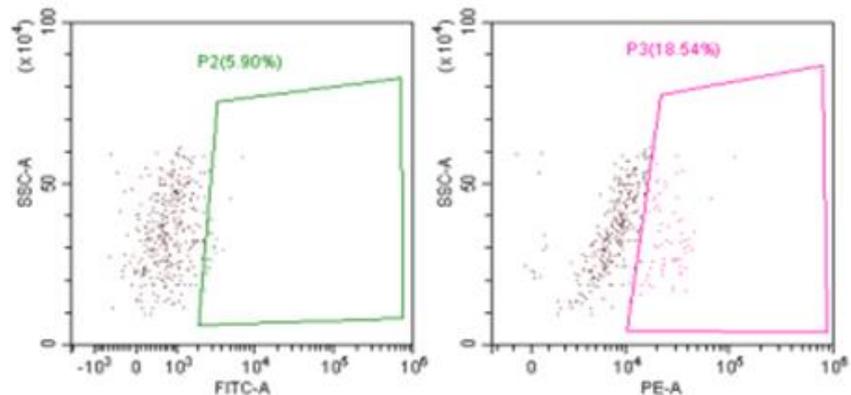
Figure 3.25: Effects of Nocodazole treatment on HDR efficiency in HIECS. To assess the effects of Nocodazole on HDR efficiency, cell aliquots were collected prior and 48 hours post CD4 purification and used for flow cytometry analysis. Comparison of FITC-GFPSpark and PE-CD4 positive cells (given in %) in HIECs treated with CRISPR alone or with Nocodazole. Threshold was chosen based on FACS sort of untransfected HIECs. For each test sample, 30,000 cells were analysed. SSC-A vs FSC-A dot plots were used to exclude debris and dead cells.

According to the results presented in figure 3.25, approximately 1.85% of the CRISPR transfected cell population exhibited a possible GFPSpark expression and 15.04% of cells showed a PE-CD4 expression. Following CD4 bead enrichment, the percentage of cells expressing green fluorescence decreased to 1.75% and the percentage of cells transiently expressing the CD4 marker decreased to 9.60%.

On the other hand, HIECs initially treated with Nocodazole and then subjected to electroporation and CRISPR showed a significant lower cellular viability than those cells that were only subjected to CRISPR. However, the portion of cells possibly expressing GFPSpark endogenously was measured to be approximately 3% which suggests 62% increase of HDR efficiency. Furthermore, the percentage of cells transiently expressing the CD4 marker was found to be 11.42%. This decrease of cells expressing CD4 observed in CRISPR cells treated with Nocodazole could be possibly explained by the high degree of cell death. FACS results post CD4 bead enrichment also demonstrated a high degree of cell death. Moreover, both percentages of cells indicating positivity in green and red fluorescence were increased post CD4 purification with 5.9% positivity detected in FITC channel and 18.54% positivity observed in PE channel. However, the cell viability was significantly decreased post CD4 purification step even though cells were allowed to recover for 48 hours prior to flow cytometry

analysis. This high degree of cell toxicity could have been possibly occurred from a combination of effects attributed to the cells being forcefully arrested at G2/M phase, subjected to DNA damage from CRISPR and the harsh cellular conditions from the CD4 bead-based purification step. Consequently, it is difficult to draw a definite conclusion from the results generated after CD4 enrichment. Nevertheless, single cells were transferred onto 96 well black polystyrene Microplates with clear flat bottom. Cells were allowed to proliferate for 10 days prior to their investigation under confocal microscope. From the single cell cloning step, only 20% of cell clones achieved to proliferate and generate single cell clonal populations. However, confocal microscope did not establish any significant GFP fluorescence.

3.4.5.2 L755505 treatment

Recent research reports have also attempted to improve HDR efficiency by biochemically altering the HDR or NHEJ pathways. Modulation of HDR within the context of CRISPR-genome engineering has been employed by many research investigators directed to the establishment small chemical compounds that increase HDR events of Cas9-generated DSB in cultured cells, specifically for use with single stranded oligonucleotide donors (ssODN) rather than dsDNA as donor DNA template. Particularly, a recent study performed a screen for small compounds and identified two chemical compounds (Brefeldin A and L755507) that established a 9-fold increase for insertion of a point mutation by ssODN, as well as a 2- and 3-fold increase in the frequency of insertion of GFP into the Nanog genomic site using a dsDNA template with large homology tails [315]. However, it remains unclear how Brefeldin A, an inhibitor of ER to Golgi transport, or L755505, a β3-adrenergic receptor agonist can exert this HDR enhancing outcome [307].

Therefore, this study attempted to investigate whether L755507 treatment in HIECS and Caco2 will increase the frequency of GFPSpark integration into SOCS3 CDS.

These small molecules have been demonstrated to be cell type specific and context dependent, with scientists demonstrating activity to varying degrees [315, 320]. Previous reports have used different concentrations of L755507 (Tocris Bioscience, C# 2197) ranging from 5-40 µM, with the majority of articles reporting an optimal concentration of 10 µM [315, 320, 321]. To characterize the effects of L755507 treatment on HIECs and Caco2 cells, cell viability after the addition of L755507 compound into cell culture was measured by PrestoBlue Staining 48 hours after treatment. L755507 was tested at concentrations ranging from 0.313 µM to 160 µM. Results from PrestoBlue staining assay are presented in figure 3.26.
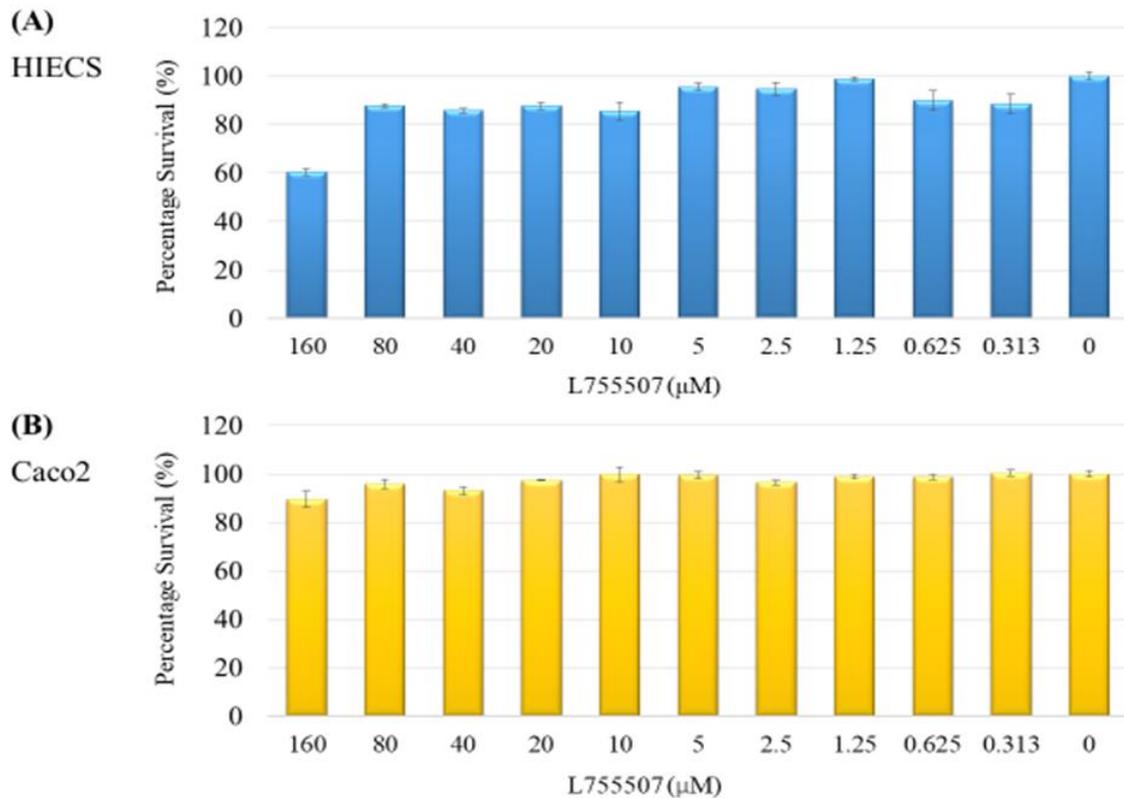
Figure 3.26: Cell survival after L755507 treatment determined by PrestoBlue assay. Cell viability after the addition of L755507 compound into cell culture was measured by PrestoBlue Staining 48 hours after treatment. L755507 was tested at concentrations ranging from 0.313 μM to 160 μM in HIECs **(A)** and Caco2 **(B).** Vehicle-treated control cells were considered as 100% viable against which treated cells were compared to. Data were expressed as mean ± SEM of quadruplicate experiments.

According to the results from PrestoBlue assay, HIECs exhibited varying percentages of cell survival upon LY755507 treatment. Specifically, when HIECs were treated with the lowest concentration (0.313 μM), they displayed 88.5 ± 4.18% cell viability, whereas when treated with 5 μM, cells exhibited with 95.6 ± 1.71% viability. The lowest percentage survival was achieved when HIECs were treated with 160 μM and exhibited 60.50 ± 1.48% viability. No significant decrease of cell viability was observed in Caco2 when treated with various concentrations of L755507.

To test whether L755507 treatment in HIECS has any effect in the frequency of GFPSpark integration into SOCS3 CDS, cells were co-electroporated with *SOCS3* template donor and CRISPR-Cas9 plasmid. After cell adherence, transfected cells were treated with L755507 at 10 μM concentration for 48 h and then with complete media for 24 hours. Following incubation, cells were trypsinized, washed twice with DPBS and prepared for CD4 enrichment and flow cytometry analysis as described above (Nocodazole treatment). HIECs that were subjected to CD4 enrichment are presented in figure 3.30.

(A) CRISPR - HIECs

Prior CD4 bead based enrichment
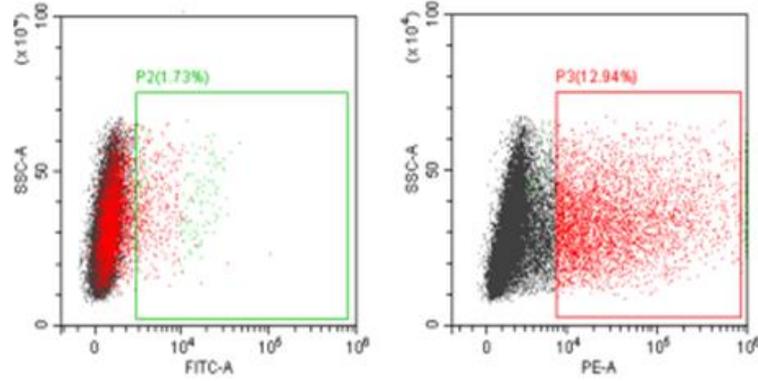
Post CD4 bead based enrichment
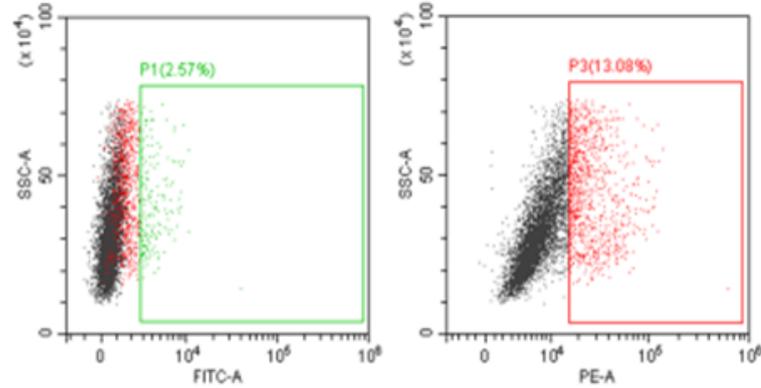
(B) L755507 treated CRISPR - HIECs

Prior CD4 bead based enrichment

Post CD4 bead based enrichment

Nikoletta Kalenderoglou - November 2019
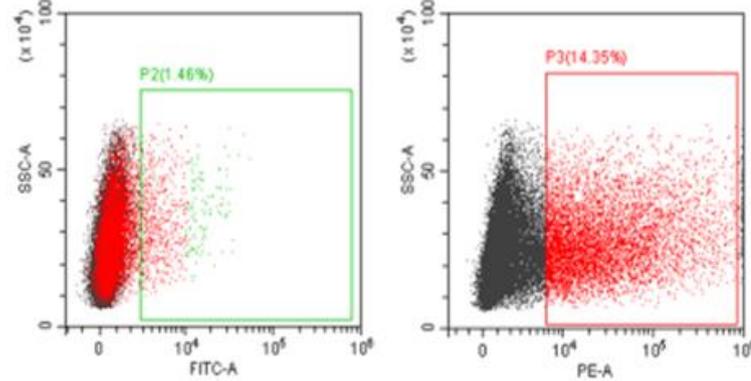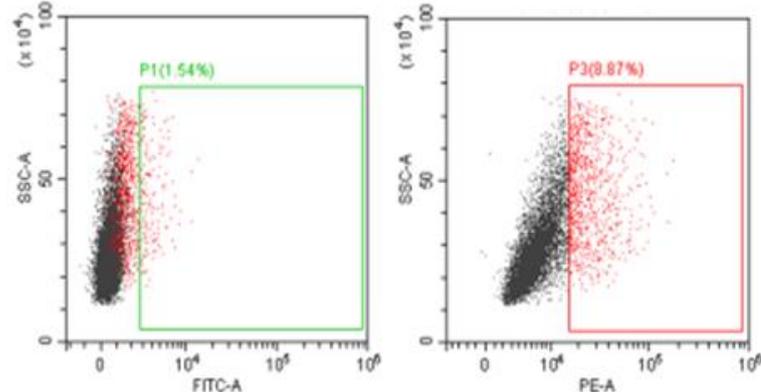
Figure 3.27: Effects of L755507 treatment on HDR efficiency in HIECS. To assess the effects of L755507 on HDR efficiency, cell aliquots were collected prior and 48 hours post CD4 purification and used for flow cytometry analysis. Comparison of FITC-GFPSpark and PE-CD4 positive cells (given in %) in HIECs treated with CRISPR alone or with L755507. Threshold was chosen based on FACS sort of untransfected HIECs. For each test sample, 30,000 cells were analysed. SSC-A vs FSC-A dot plots were used to exclude debris and dead cells.

After CD4 based purification, cells were returned to the incubator and left for 48 hours to recover. Following incubation, cells were subjected to flow cytometry and results are exhibited in figure 3.27.

According to the results presented in figure 3.27, prior to CD4 bead purification step, there was a reduction of GFPSpark of the CRISPR transfected cells treated with L755507 compared to the untreated CRISPR transfected cell population (from 1.73% to 1.48%). However, the percentage of PE-detected fluorescence in both cell populations showed similar percentages indicating that L755507 did not affect the expression of CD4 reporter. Post CD4 enrichment, the percentage of cells expressing green fluorescence increased to 2.57% in untreated CRISPR transfected cells whereas L755507 treated cells showed a minimal increase of green fluorescence (1.54% from 1.48%). The percentage of cells transiently expressing the CD4 marker decreased to 12.09% in untreated HIECs and 9.87% in treated HIECs.

Single CD4 sorted cells were seeded onto 96 well black polystyrene Microplates with clear flat bottom and allowed to proliferate for 10 days prior to their investigation under confocal microscope. From the single cell cloning step, 70% of cell clones achieved to

proliferate and generate single cell clonal populations. However, confocal microscope did not establish any significant GFP fluorescence.

The effects of L755507 compound on HDR efficiency was also examined in Caco2 cells. Caco2 cells, cells were transfected with equal amounts of the donor Cas9 plasmid with the gRNA1 and the donor repair template. Cells were kept in complete medium for 24 hours at 37ºC in a 5% CO_2 incubator. Following incubation, 10 μM of L755507 was added to the cells without changing the culture medium. Cells were further allowed to proliferate for 48 hours. After the incubation, cells were accessed by their fluorescence signal detected by PE-channel and FITC-channel using flow cytometry. Furthermore, Caco2 cells were not subjected to CD4 bead-based enrichment because it was shown to be very toxic to the cells. However, the expression of CD4 reporter at the cell surface was monitored by CD4 antibody. Results are presented in figure 3.28.
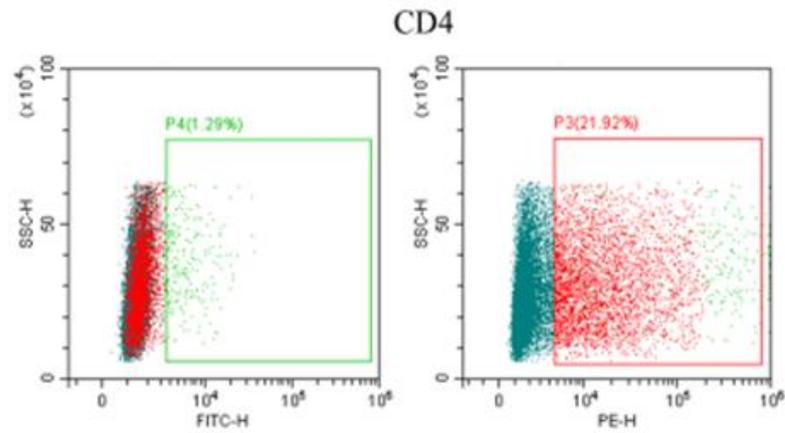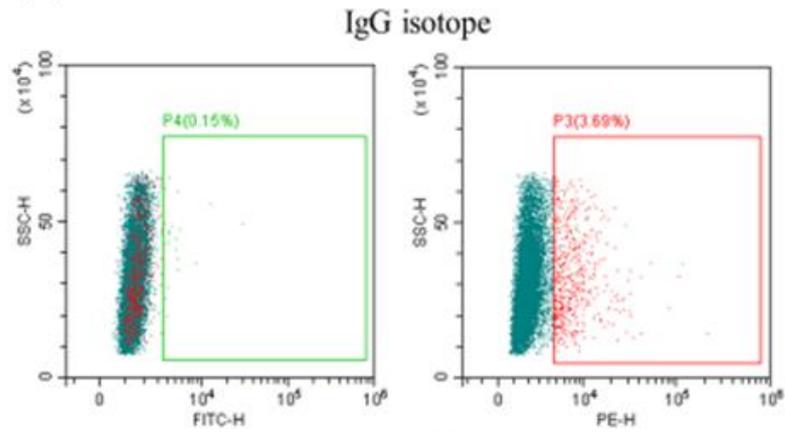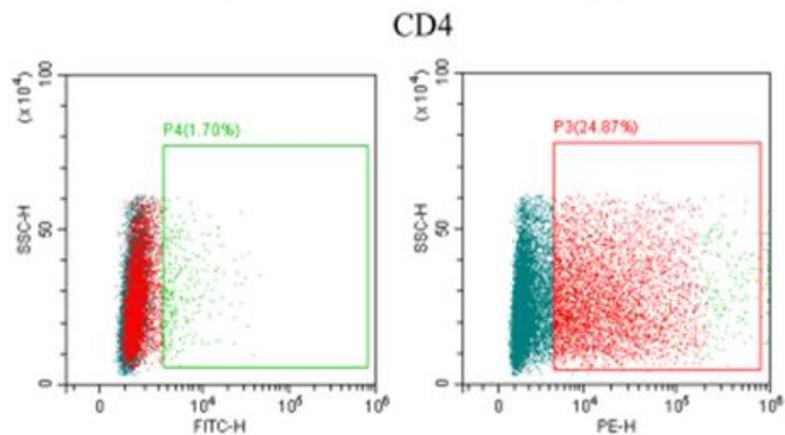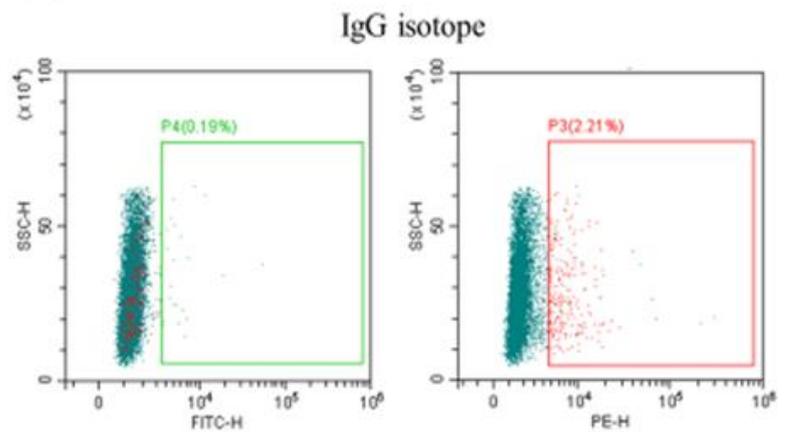
Figure 3.28: Effects of L755507 treatment on HDR efficiency in Caco2. To assess the effects of L755507 on HDR efficiency, expression of GFPSpark and CD4 was monitored 72 hours post transfection with CRISPR-Cas9 plasmid/gRNA1 and by staining with PE-CD4 antibody. **(A)** cells were not treated with L755507 and acted as a control whereas **(B)** cells were transfected with complete CRISPR complex, allowed to proliferate for 24 hours and then treated with L755507 for 48 hours prior to FACS. A side scatter height (SSC-H) vs. forward scatter height (FSC-H) dot plot was used to exclude debris and dead cells and the gated cells were then analysed in SSC-H vs PE-H for the expression of CD4 reporter and SSC-H vs FITC-H for the expression of GFPSpark. Nonspecific CD4 binding was monitored by treating the cells with IgG1 isotope control.

Results showing the expression of CD4 and GFPSpark are presented in figure 3.28. Nonspecific CD4 binding was blocked with FC reporter blocking solution and it was also checked by treating the cells with IgG1 isotope control. According to the data from figure 3.28, Caco2 cells transfected with CRISPR complex and treated with L755507 showed a lower IgG isotope binding affinity (2.21%) than the untreated transfected CRISPR Caco2 cell population (3.69%). Furthermore, the percentage of transfected untreated Caco2 cells representing FITC positivity demonstrated a low GFPSpark signal (0.15%) whereas the percentage of transfected and L755507 treated Caco2 cells showed a slightly increased green fluorescence (0.19%). However, this percentage was significantly increased in both cell populations when stained with CD4 antibody. This observation highly supports the notion that the percentage of positive cells detected by FITC could be the result of spill over of fluorescent signal from PE channel.

Nevertheless, single cells were transferred onto 96 well black polystyrene Microplates with clear flat bottom. The step of CD4 bead-based enrichment was avoided because it showed previously that is very toxic to Caco2 cells. Cells were cultured for 10 days

prior to their investigation under confocal microscope. From the single cell cloning step, confocal microscope did not establish any significant GFP fluorescence.

3.4.6 Validation of CRISPR genome editing

Another necessary step towards employing successfully genome editing tools, such as CRISPR, is to investigate the efficacy of the programmable nuclease, which can differ dramatically in relation to the sequence that is targeted. To date, various detection approaches to identify CRISPR/Cas9-inudced indels have been published. A strategy that is most commonly used is the mismatch cleavage assay. This method is easy to use and it uses enzymes that cleave DNA heteroduplexes at mismatch sites [281].

Although the Mismatch Cleavage assay is error prone due to subjective bias (manually choosing size of regions to estimate peak intensity for densitometry) and high background, gel band patterns produced by mixed cell pools can be used as an approximate indication for low or high performing gRNAs. For instance, features common to low-activity gRNAs produce 'uncut' DNA band intensities or cut products that may be hardly visible above background. On the other hand, a feature common to moderate and high activity gRNAs produces a low signal of the parental band intensity relative to that PCR product not treated with digestive enzyme, concomitant with increased band intensities in the cut products of expected sizes.

We delivered the bicistronic Cas9/gRNA1 or gRNA2 construct targeting the region surrounding the SOCS3 stop codon in HIECs using electroporation as described above. Upon 48-hour post transfection, the cells were harvested to analyse the genome modification efficiencies. This experiment was performed twice (n=2). The percentage of mismatch sites was determined using the GeneArt genomic cleavage detection assay.

As depicted in figure 3.29, the positive control only produced visible cleavage bands indicating that the experiment worked. However, both gRNA1 and gRNA2 treated cell pools failed to generate cleaved bands. These results strengthen the hypothesis that Cas9 nuclease was unable to access the target sequence and it was unable to cleave at target site.
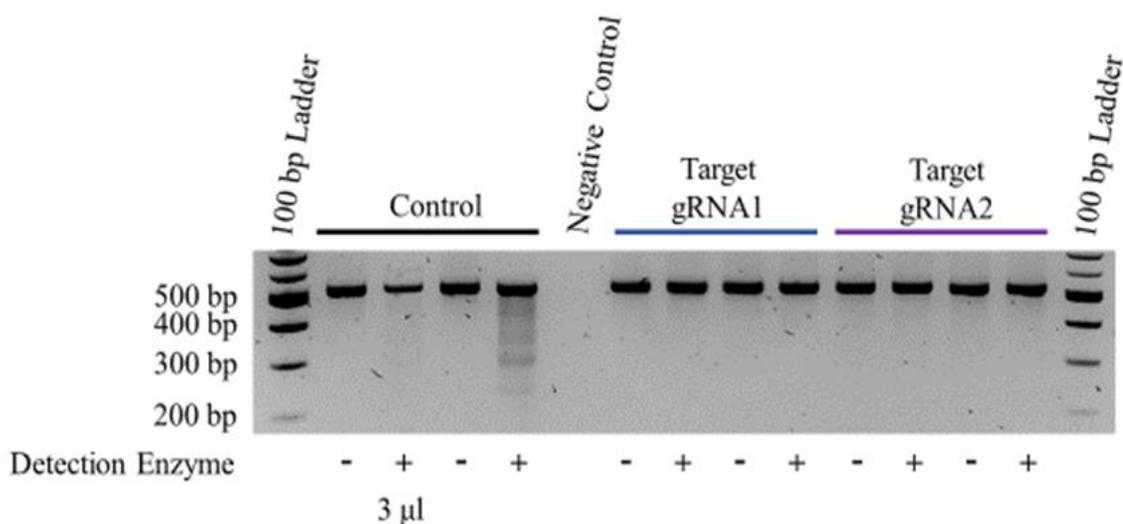
Figure 3.29: **Gel image of genomic cleavage detection assay.** HIECS were transfected with the bicistronic Cas9/gRNA1 or gRNA2 construct targeting SOCS3 gene (Target gRNA1 and Target gRNA2) (n=2). The samples were PCR amplified using the same set of primers flanking the region surrounding the intended DSBs. A positive control sample for gene modification was prepared by using control template and primers provided by the kit. A negative control was also prepared by excluding DNA in the mixture reaction. Following re-annealing step, samples were incubated with and without Detection Enzyme and run on a 2% agarose gel. Where not indicated, 5 μl of re-annealed and enzyme digested reaction was loaded in a well.

To further verify this lack of cleavage ability of the designed gRNA targeting SOCS3, a stretch of DNA template around the intended DSBs was PCR amplified for a control (cell pool that was not subjected to CRISPR) and a cell sample treated with Cas9/gRNA2. Since both gRNAs cut at close proximity of SOCS3 stop codon, it was

decided to only investigate gRNA2 cut efficiency. The PCR products were then sent to be sequenced (Sanger BioScience Sequencing, Nottingham, UK). The two resulting sequencing traces were analysed using a designed web tool known as TIDE (tracking of Indels by Decomposition) (available at http://tide.nki.nl).

This web tool utilizes the quantitative sequence pattern results from two standard Sanger sequence reactions to measure the editing efficiency and establish the major indels in the genomic sequence of a targeted cell pool [322]. According to this method, the DNA sequence pattern from the cells expressing the targeted nuclease exhibits a mixture of signals generated from unmodified DNA and DNA containing a mixture of indels (deletions and insertions) after the DSB site. An example of this method is depicted in the example shown in figure 3.30 (A).
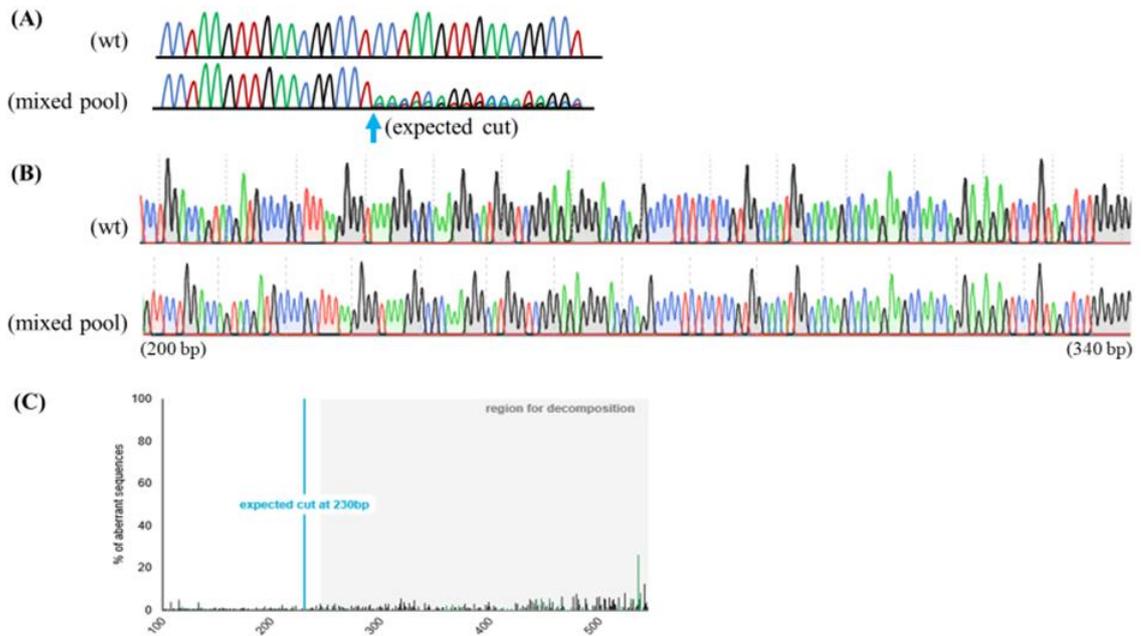
Figure 3.30: **Evaluation of Cas9/gRNA cleavage efficiency by sequence trace decomposition.** **(A)** A hypothetical example illustrating that an imperfect repair after cutting by Cas9 nuclease causes the DNA in the mixed pool to contain a mixture of indels, generating a composite DNA trace after the DSB. **(B)** Chromatogram analysis of control cells that were not subjected to the treatment of Cas9 and cells that were treated with Cas9 nuclease. **(C)** Outline of TIDE output, which comprises the visualization of aberrant DNA signal in control (black) and Cas9-treated sample (green), the expected break site (vertical blue line) and the region used for decomposition (gray bar).

According to the chromatogram analysis depicted in figure 3.30 (B), the DNA sequence trace generated from the pool of cells treated with Cas9/gRNA2 is very similar to the DNA sequence trace produced by the control pool. The TIDE algorithm has the ability to resolve indel size frequencies derived from Cas9 treated cells by comparing and decomposing Sanger sequences made from PCR templates of nuclease targeted sites from WT and edited regions [322]. TIDE quantitative analysis from the sequence trace data did not exhibit any significant aberrant base signal surrounding the expected cut

site, indicating the absence of any type of possible indel within the decomposition region. Furthermore, TIDE algorithm calculated that 1.2% of the Cas9/gRNA2 treated HIECs carried an indel which can also represent the overall cut efficiency of the system. These results further support that the designed gRNA targeting the stop codon of SOCS3 gene is unable to cut efficiently.

3.5 Discussion

Since the advent of recombinant DNA technology in the 1970s, molecular biologists have endeavoured to develop new approaches to manipulate or modify genomic information. These technological breakthroughs in genome editing have given the ability to directly edit the function of DNA in its endogenous context for essentially any organism and further more to explore the functional architecture of the genome at the systems level [254]. Several techniques have been established in the last two decades that enable the generation of targeted double-stranded DNA breaks which can be fixed by the error-prone NHEJ or via the HDR system provided a suitable donor is available [323].

These targeted genome editing methods involve the use of meganucleases (MNs) [324], zinc finger nucleases (ZFNs) [325], transcription activator-like effector nucleases (TALENS) [326], targetrons [327] and finally the newly recently developed clustered regularly interspaced short palindromic repeat (CRISPR)-associated nuclease Cas9 [262, 305].

Zinc finger nucleases (ZFNs), transcriptional activator-like effector nucleases (TALENs), and the recently developed two-component CRISPR/Cas9 tool can all be manipulated to produce targeted double-strand DNA breaks (DSBs) in genomic DNA and facilitate epitope and fluorescent tags to label endogenous proteins and establish gene expression reporters [328].

Scientists usually choose which genome-engineering tool to use based on the breadth and ease of use and the efficiency and specificity of editing. ZFNs and TALENs are characterized as dimers with each monomer involving a fusion protein-binding domain

that together define target specificity and cleavage is carried out by the attached FokI nuclease catalytic site [329, 330].

However, ZFNs system usually develops high frequencies of off-target effects and high toxicity due to its low affinity and low specificity. On the other hand, the methodology of TALENs relies on one-to-one recognition rules between the TALE DNA-binding domain which is composed of multiple repeats and the target nucleotide sequence, resulting in an enhanced editing efficiency and decreased off-target edits [331, 332]. Nonetheless, the development of the nuclease protein is challenging and a new TALEN protein must be produced for each genomic sequence, which consequently increases the cost and time for development [333]. Additionally, these approaches have proven to show high sensitivity to DNA methylation and consequently are not ideal for GC-rich regions, which can be often located at the UTR part of the genes [334].

Following the initial demonstrations of CRISPR/Cas9 system in 2012, CRSPR-based methods have rapidly become a favoured technique to perturb or endogenously tag genes to uncover their functions in virtually all commonly studied eukaryotes, ranging from yeasts to humans [261]. Moreover, Cas9 is simple to use and re-programme it to target different genomic locus or target several other genomic sites simultaneously. However, the system has few limitations. First of all, it requires the presence of a PAM sequence in close proximity to the target sequence [261, 296]. A crucial element that dictates the degree of editing efficiency is the proximity of the double-stranded break induced by Cas9 nuclease to the target site. Several reports have concluded that as just a few additional bases further upstream or downstream of the target site can generate an important difference in editing efficiency. Thus, the genomic site of a potential DSB and subsequently the efficiency of donor insertion into the genome, is governed by the

availability of PAM sites in proximity to the intended target site [305]. However, even if a gRNA target site is located in close proximity to the edit site, it is not definite to achieve high editing efficiency. The gRNA activity is highly dependent on the nature of gRNA sequence, chemical modification and the susceptibility of the DSB to be repaired efficiently, as well as accessibility of the targeted genomic site. Consequently, alternative engineering tools such as TALENs altered to lack the 5'T targeting element have acquired a fundamental advantage over CRISPR since they can be engineered to target virtually any genomic locus without the need of a PAM site [335].

The gRNA targeting sequences can be introduced into the cells as a plasmid construct that requires to be transcribed or as a synthetic RNA. The Cas9 nuclease can be delivered as a plasmid construct, mRNA, or purified protein. Liang et al. [336] investigated the kinetics of genome cleavage by transfection KE293 cells with either plasmid DNA, mRNA, or Cas9 protein, followed by cleavage detection assay and western blot analysis of cell lysates. They observed that Cas9 protein accumulated during the 72 hours period in plasmid DNA-transfected cells, whereas the Cas9 protein levels from mRNA- or purified Cas9-transfected cells peaked in the first 4-24 hours and it was not detectable after 48 hours. They argued that all three delivery methods showed similar cleavage kinetics on the intended genomic sites but Cas9 plasmid DNA produced higher ratios off-target indels [336]. Several articles that have endogenously inserted a fluorescent reporter to a targeted region have used the plasmid format to introduce Cas9 into the cells [242, 315, 316, 321], as plasmid DNA is persevered longer inside the cells, possibly leading to more complete modification but at the cost of off-targeting effects [337]. For this reason, we used a plasmid DNA format to introduce Cas9 into HIECs and Caco2 cells.

On a different note, it has demonstrated that CRISPR/Cas9 editing efficiencies differ dramatically for different gRNAs, even for those targeting the same genomic region [338]. At present, our knowledge of the link between sequence factors of gRNAs and their on-target cleavage efficiencies remains largely unknown, due to challenges in evaluating the cleavage efficacy of a large number of gRNAs. Nevertheless, recent studies have attempted to demonstrate various gRNA sequence determinants that may play an important role in CRISPR/Cas9 cut efficiency [301, 339]. For instance, guanines are preferred at -1 and -2 upstream of the PAM motif, and cytosine at -3 downstream, which is the Cas9 editing site [340]. In our strategy, only gRNA1 has a guanine at -1 upstream of the PAM motif, whereas both gRNAs contain an adenine at the Cas9 editing site. Additionally, gRNA secondary structure and sequence factors such as high GC-content may also be ascribed as factors lowering the cleavage efficiency of a particular gRNA [341, 342]. However, these parameters were optimized during gRNA design with the help of two web tools (http://crispr.mit.edu and http://chopchop.cbu.uib.no). Moreover, Doench et al. [343] observed that sgRNA targeting the 5' and 3' UTRs were highly ineffective and their activity quickly lowered as a number of few nucleotides (1-15 nts) away from the CDS. In our strategy, gRNA2 cuts 5 nucleotides downstream of SOCS3 stop codon. These factors could have probably contributed to their low cleavage efficiency. Thus, validation methodologies for measuring the modification frequencies of Cas9 nuclease are typically applied to evaluate activity. We utilised the GeneArt genomic cleavage detection assay to determine the activity of both designed gRNAs targeting SOCS3 stop codon. However, both gRNA sequences failed to generate any bands in the agarose gels indicating that both gRNAs are either unable to access the targeted region or unable to generate a cleaved site. This was also verified by the sequence trace data generated by Sanger sequence analysis. Therefore, it is important

to consider the design and development of new gRNA sequences targeting the surrounding area of the stop codon of SOCS3 gene.

Ratz et al. [242] used gRNA sequences that targeted genomic sites 8-125 bp away from the intended HDR site to endogenously tag HMGA1, VIM and ZYX genes. Thus, even though it is highly suggested that gRNA targeted sites should be in close proximity to HDR site, farther targeting regions from the HDR site should be considered when designing new gRNAs with higher cleavage efficiencies.

Fundamentally, eukaryotic gene expression is controlled at the epigenetic level by organising DNA into nucleosomes and forming histone octamers. These eukaryotic chromatin structures vary greatly from those found in bacteria. Consequently, it is conceivable that Cas9, as a prokaryotic derived enzyme, is not optimized to explore and engineer large, chromatin-bound eukaryotic genomes [344]. This has been supported by previous research publications demonstrating that chromatin structure affects Cas9 binding by restricting the accessibility of the intended editing site [339, 345]. Cas9 tends to favour more accessible euchromatin regions than target sites that are buried in heterochromatin regions [345]. Furthermore, multiple research articles have established a correlation between rates of DNA binding and cleavage in genomic regions of open chromatin as analysed by DNase I hypersensitivity [339, 346]. However, it is difficult to draw a definite conclusion for target accessibility in the case of SOCS3 gene since there is no published work that targeted the stop codon of SOCS3 using CRISPR technology and we only tested two gRNA sequences. Nonetheless, it has been observed that sgRNAs targeting the 3'end of a gene do not tend to be as efficacious as those targeting the 5'end [294]. These factors though cannot be further optimized unless if CRISPR/Cas9 is designed to target the 5' end of SOCS3 and endogenously tag the N-

terminus. Nevertheless, a high efficacious DNA delivery method can impact on the overall efficiency of CRISPR engineering method. Thus, we attempted to optimize the transfection efficiency in both Caco2 and HIECs.

For Caco2 cell line, the highest transfection efficiency achieved with FUGENE-HD transfection agent (approximately 48%) was by doubling the amount of DNA transfected to the cells than what the manufacturer's online protocol suggested. Nevertheless, finding transfection methods that enable efficient DNA uptake into Caco2 cells is relatively problematic. Additionally, it is crucial to find a good technique to transfect this cell line with high efficiency devoid of cellular alteration and maintaining their ability to differentiate. An important limitation with working with Caco2 cells is that it is difficult to produce single cells during cell culture and this is probably due to their ability to form tight junctions. They also tend to exhibit high sensitivity to physical stress and changes in temperature. Thus, handling and manipulation during transfection is a very important step. For HIECS, FUGENE-HD online tool did not provide any directions for an optimized transfection protocol. Therefore, this study attempted to optimize HIECS transfection efficiency with FUGENE-HD method through varying different transfection parameters, including cell density, FUGENE-HD to DNA ratio and amount of DNA transfected to cells. To this end, the highest transfection efficiency established by FuGene HD for HIECS was approximately 33% and it was achieved with a 4:1 DNA:FuGene HD agent (μl) and doubling the amount of DNA transfected to the cells than what the manufacturer's usually suggests for the corresponding starting cell density. However, the electroporation method gave much higher transfection efficiency for HIECs, while simultaneously retaining cell toxicity to the minimum compared with FUGENE-HD lipid method. Specifically, flow cytometric analysis showed that HIECS displayed a transfection efficiency of 43.62% $\pm$ 0.54% when transfected with

pmaxGFP® Vector and 58.01% ± 0.34% when transfected with SOCS3-GFPSpark. These results show that transfection efficiency depends on the type of plasmid used for transfection. Therefore, future work should also involve an additional method for monitoring Cas9 abundance in transfected cells. For instance, one method that could detect and monitor the activity of Cas9 is to lyse cells 4-8 hours post transfection and investigate the expression of Cas9 protein by western blotting using a Cas9 monoclonal antibody. Since Caco2 and HIECs showed a moderate level of transfection efficiency, future work should utilize a different type of cell line that would show a higher transfection efficiency to identify the most efficient SOCS3 gRNA. Once the most efficient SOCS3 gRNA is identified, then it can be applied to HIECs and Caco2 cells.

Recent research has also established that there is a strong positive correlation between the frequency of successful recombination events and the length of DNA repair template [11]. In general, targeted integration of large DNA fragments such as a reporter gene requires a repair template that harbours left and right homology arms of 700-900 bp in length [309]. We designed the length of the homology arms to be 693 bp for the right arm and 890 bp for left arm. However, we were unable to confirm if the length of the HA were long enough to achieve the intended integration since we did not have an efficient SOCS3 gRNA to generate DSBs. Shy et al. demonstrated that when an increase in homology tail length is increased from 25 bp to 1500 bp is associated with an equivalent linear increase in on-target DNA insertions [307]. Therefore, future work should also consider the possibility that longer homology arms could potential increase integration efficiency. Moreover, during the design of the repair template, the sequence of the right arm was designed to begin 4 base pairs after the stop codon to disrupt the homology between the gRNA sequences and the repair template. However, this could have also been achieved by disrupting the PAM with a silent mutation in the donor

repair template to avoid recutting of the engineered target by the active Cas9/gRNA protein complex.

Regarding the donor repair template, it is also worth noting that the use of linear DNA repair templates have shown to improve HDR efficiencies in relation to the circular plasmid templates, but the higher degradation frequency and undesirable spontaneous insertion could be crucial for genome stability [325].

Currently, the major limitation of CRISPR/Cas9-induced HDR is that it typically occurs at very low efficiencies and substantially more variable frequencies compared to NHEJ. Several research articles have successfully achieved high transfection efficiency of expression Cas9 vectors, the gRNA and the repair template but they have only managed to achieve successful recombination event in a small cell population (<10%) with the majority of articles reporting efficiencies as low as 0.1% [242, 266, 317, 321, 347, 348]. However, the efficiency of achieving deletions by NHEJ has been reported as high as 90% [266, 349]. Various methods have been adapted to attempt to increase this efficiency. For instance, Liang et al. [305] observed that sequential delivery of Cas9 RNPs and donor template increased HDR events. However, sequential delivery may not be appropriate to cell lines that are sensitive to several cycles of electroporation. As noted from our results, HIECs needed a short period of time to recover prior to their investigation with flow cytometry. It would be interesting to try and generate KI cells by using a two-step transfection sequencing on consecutive days or first one in the morning and second one in the evening.

Recent research reports have also attempted to improve HDR efficiency by biochemically changing the HDR or NHEJ processes. For instance, the treatment of several mammalian cell lines with Scr7, a DNA ligase IV inhibitor, demonstrated a 19-

fold increase in HDR efficiency [274]. Additionally, the simultaneous treatment of both KU70 and DNA ligase IV with siRNAs increased the HDR events by 4-5-fold in human and mouse cell lines [350].

Small-molecule compounds, including L755507, were recently demonstrated to notably increase CRISPR-mediated HDR [315]. Yu et al. observed that β3-adrenergic receptor agonist L755507 raised TNS efficiency in hiPSCs (from 0.35 to 3.13%) and knock-in efficiency in mouse ESCs (from 17.7 to 33.3%), although it still unclear how this compound alters the repair pathways [315]. Other reports have not identified any substantial influence on HDR by L755507 in HEK cells [321] or hiPSCs [309]. Furthermore, Pinder et al. [321] treated cells with SCR7, RS-1 and L755507 singly and together and observed that there was no additive effect when treating either SCR or L755507 together with RS-1 in relation to RS-1 treatment alone. These observations strengthen the argument that SCR7, L755507 and RS1 are not consistent between cell types. Furthermore, this strategy may possibly also lower HDR events due to reduced Cas9 protein levels caused by the compounds. Unfortunately, we are unable to draw a definite conclusion on the effects of L7777505 compound on HDR efficiency because at the time we were testing the L755507 compound in our system, we were using gRNA sequences that did not prove to generate DSBs efficiently.

Alternatively, different studies have utilized cell cycle synchronization approaches to improve CRISPR editing efficiency by achieving controlled Cas9 nuclease action at the stage of the cell cycle best for HDR. A well-established method that has been used by many laboratories is the cell treatment with Nocodazole as a means of cell cycle arrest. The treatment effect is due to Nocodazole synchronizing the cell cycle at the G2/M stage, leading to the accumulation of cells in the S/G2/M phases during which DSBs

are preferentially repaired by HDR instead of NHEJ [321].

Even though several studies have implemented various means to chemically or genetically interrupt the NHEJ process and increase HDR activity, such implementations can be challenging to employ, require a lot of optimization, they can be toxic to cells or all the above. Subsequently, high cleavage events of programmable Cas9 proteins is not necessarily associated with efficient HDR-mediated genome engineering. Additionally, these invasive alterations may be undesirable for therapeutic applications due to the fact that these treatments can change the cellular response to DNA damage at different off-target locations in the genome and result to tumour growth [351].

Integration of all these approaches and further optimization steps may lead to enhanced precise genome editing of SOCS3 gene in human cells and achieve the desired KI effect.

The CRISPR-Cas9 technology has taken the world of genome engineering by storm in the last decade enabling to target any genome sequences with unprecedented speed, precision, and accuracy [352]. Ideally, such genome-engineering tool would be 'scarless', excluding the need for a selection marker [11, 353, 354]. However, the absence of positive selectable cassettes integrated into the edited allele could also influence negatively on the efficiency of successful genome-editing and consequently the isolation of successful genome engineered cell clones [11]. Isolation of clonal cell lines with precise genomic alterations is highly desired. This can be performed after transfection by isolating single cells through either FACS or serial dilutions, followed by an expansion period to establish a new clonal cell line. It is worth noting that different cell lines can exhibit varying responses to single-cell isolation. For instance, Caco2 cells did not appear to grow and expand as efficiently as when they are seeded

confluently. On the other hand, serial dilution did not appear to negatively influence the culturing of HIECs. Moreover, as the analysis of any fluorescent molecule depends on the levels of expression, low to very low-abundance FP tagged proteins that are driven by their endogenous promoters can be laborious and difficult to detect even with the use of sophisticated instruments such as confocal microscopy or flow cytometry [348].

Thus, future work should consider using a selectable cassette along with the FP to be integrated at the targeting site, as it will provide an easier method to isolate the successful edited cell clones. As it can be seen in this chapter, the CRISPR-Cas9 technology includes a very laborious process that involves multiple optimization steps. However, there is no doubt that when this tool is careful optimized, it can deliver precise modifications in a genome and interrogate functionalities of any DNA regions.

3.6 Conclusion

A fundamental research goal of system biology is to investigate the molecular mechanisms of protein dynamics and relating it to function or dysfunction. Until recently, most measurements to investigate protein dynamics have been generated from population samples, where a great number of cells were homogenized and then analysed. However, such population measurements do not relieve information about the complex heterogeneity of cellular behaviour within the samples. To date, the systematic single-cell analysis has become a well-established method to exploit these system dynamics and reveal significant and unprecedented insights into cellular function and heterogeneity. Several research reports have used fluorescent proteins (FP) to study expression levels in single cells, either by fusing the FP to a protein of interest or by inserting the FP under the control of a promoter under investigation [355-357]. The transient or stable expression of a FP-tagged protein construct in cells is a simple, useful and rapid tool that enables the live-imagining of the direct subcellular localization and dynamic behaviour of proteins by confocal microscopy. However, FP-based imaging has several limitations in terms of resolution and sensitivity [243]. For instance, if the expression of the protein of interest is low or if the protein is localized in different compartments of the cell in low levels, the detection is difficult [358, 359]. Furthermore, in several signalling pathways, the precise timing of protein expression is tightly regulated since it can affect the output of the system [138, 360].

Therefore. fluorescent protein tagging a gene at its endogenous locus represents a particularly beneficial method for monitoring protein dynamics as it will preserve the normal expression level of the tagged protein [321].

Our current knowledge of dynamics of SOCS3 has been derived from results derived from whole cell population experiments [61]. Since members of SOCS family have a relatively short biological half-life and their negative regulatory effects are mostly transient [107], the ability to study SOCS3 dynamics without over-expressing SOCS3 or depleting endogenous SOCS3 may demonstrate a more accurate representation of SOCS3 dynamic behaviour.

The CRISPR/Cas9 system has been commonly used for the development of gene knock-ins in cells derived from a variety of species such as human, rats, mice, zebrafish, drosophila, nematodes and the parasite Plasmodium *yoelii* [11, 266, 285-287, 361-365]. In this chapter, we attempted to utilize this engineering tool to endogenously tag SOCS3 with a GFP at the C-terminus. Tagging SOCS3 at its endogenous locus will be particularly useful for live cell analysis of SOCS3 dynamics.

However, there are multiple determinants that could influence Cas9/sgRNA editing efficiency. Firstly, the intrinsic specificity encoded in the Cas9 protein, where gRNA activity is highly dependent on the nature of gRNA sequence and its accessibility to the targeted genomic site. Secondly, the susceptibility of the DSB to be repaired and facilitate the desiring knock-in of a large DNA fragment such as a reporter gene. Furthermore, maximum abundance and activity of the Cas9 protein to stimulate DSBs at the target genomic locus is essential for efficient genome engineering. We attempted to optimize the transfection efficiency of HIECs and Caco2 cell lines, but we only achieved a moderate level of efficiency.

Moreover, although we have provided evidence that both gRNA sequences were unable to produce DSBs at the stop codon of SOCS3, it is challenging to accurately predict the factors that inhibited this cleavage reaction. Our findings point to the need for

optimizing the design of gRNA cleavage efficiency further, particularly in the context of knocking-in efficiently a GFP at the C terminus or N-terminus of SOCS3.

In summary, this strategy can be used as a detailed guide for an improved SOCS3 editing method in a variety of human cells to insert a fluorescent protein at the stop codon and allow rapid selection of positive clones through FACS sorting or serial cloning.

4 *In vivo* analysis of SOCS3 protein dynamics and promoter activity

4.1 Introduction

Feedback loops represent arguably the most common regulatory mechanisms whereby the abundance of intracellular proteins are regulated. These mechanisms are utilised by a vast array of cellular networks, ranging from metabolic to transcriptional to signalling levels. Feedback loops provide proteins with self- tuning capability important for their survival in response to stochastic and environmental fluctuations. Homeostatic imbalance due to failure of appropriate feedback control may lead to a disease state of the cells [366]. So far, there is a large body of research that has investigated these feedback circuits and their implications at the molecular and cellular level. While positive feedback loops are known to give rise to different stable steady states, featuring switch-like behaviours and amplifying signals [367], negative feedback loops play a crucial role in maintaining homeostasis as well as developing oscillatory protein behaviours [366].

Despite the proliferation of several innovative methods for single-cell analysis, the vast majority of biological measurements include population-wide average measurements. However, cellular response heterogeneity regardless of isogeneity is a common phenomenon that has been described across a large array of cell types from bacteria to mammalian cells [133]. Cellular heterogeneity arises from the heterogeneity in the

abundance of transcripts and proteins, which in turn may be caused by a large array of factors including thermal fluctuations, noise in gene expression and minor differences in the cellular microenvironment. Therefore, quantitative analysis of protein oscillations requires analysis at the single-cell level, as measurements of average response to the population of cells can mask the underlying response heterogeneity [368].

A fascinating example of a such negative feedback system is the ultradian oscillation in the NF-κB signalling pathway. Activation of NF-κB is due to IKK-mediated phosphorylation-induced proteasomal degradation of IκB, enabling the NF-κB transcription factor subunits to be transferred to the nucleus and induce target gene expression (figure 4.1). As a result of the rapid degradation of IκB proteins and the time delay among their transcription and translation, NF-κB protein oscillations appear [59, 202].

These oscillations can be quantitatively evaluated at single-cell resolution using fluorescent fusion reporters, such as RelA-DsRed fusions in the case of NF-κB signaling. The expression of RelA-DsRed proteins acts as the nuclear-cytoplasmic (N-C) shuttling of endogenous RelA proteins, which allows the dynamic behaviour of NF-κB signaling to be visualized with respect to the N-C ratio of red fluorescence [127, 202]. Ashall et al. attempted to explain the functional role of NF-κB dynamic behaviour by treating RelA-DsRed-expressing cells with repeated short pulses of tumour necrosis factor-alpha (TNF-α) at various intervals to mimic pulsatile inflammatory signals. According to all dynamic intervals that were analysed, they observed that low frequency stimulations displayed repeated full-amplitude translocations, but higher frequency pulses showed reduced translocation, indicating that the dynamic behaviour regulates the type of cellular response [59]. Additionally, a second study showed that

Lipopolysaccharide (LPS) elicited a slow increase of NF-κB protein, which resulted to the activation of different set of genes than of those that were activated by TNF-α treatment. Therefore, these findings indicate that the oscillating capacity of NF-κB is critical for the induction of a stimulus-specific gene expression program [58].
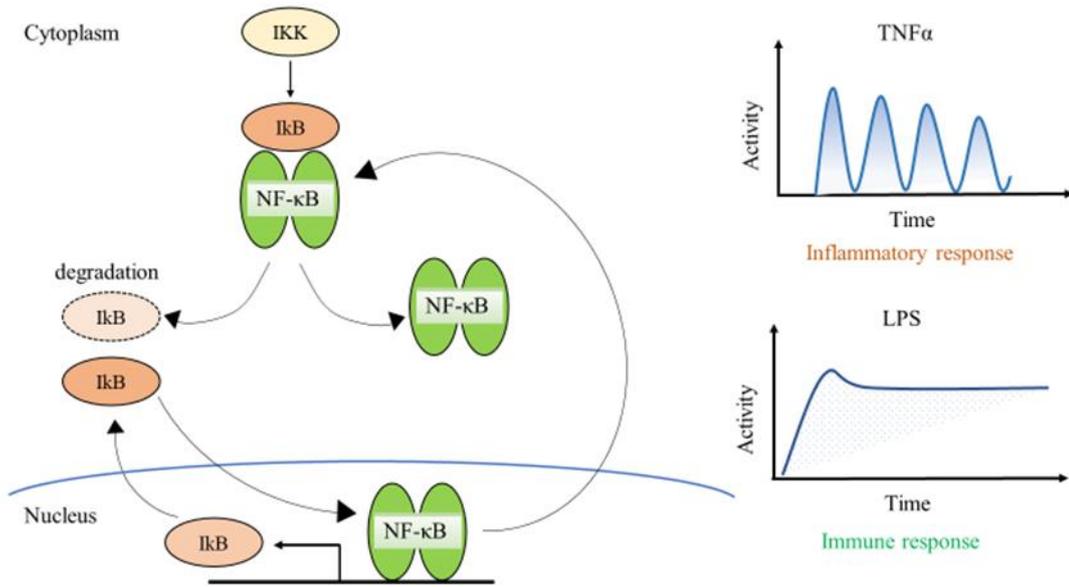
Figure 4.1: Ultradian oscillation of NF-κB protein. The negative feedback loop mechanism of NF-κB (left) and the relationship between the dynamic protein expression and the biological response (right) displayed. Adapted from [59].

Despite concrete progress in the field of system biology, it is still largely unknown what are the main mechanisms that are driving and regulating feedback circuits in several cellular systems. Significant development in molecular techniques has provided a greater understanding of the mechanistic picture of the cells, which have revealed that feedback loops often seem interconnected and act cooperatively rather than in isolation [369]. Dynamical network systems consisting of multiple negative-feedback loops that intertwine to regulate complex dynamics are ubiquitous in biology. For instance, the dynamic nucleo-cytoplasmic shuttling of the transcriptional nuclear factor kappaB (NF-

Nikoletta Kalenderoglou - November 2019

κB) is controlled by multiple negative feedback mechanisms, with the most prominent being those regulated by IκBα, IκBε and the ubiquitination enzyme A20 [370].

Similarly, the quick and short activation of JAK/STAT signalling pathway is tightly controlled by varied negative feedback loops within the cell. So far, there are three well-studied and prominent processes that efficiently supress the STAT signalling and its downstream functional proteins. These involve a long (and growing) group of phosphatases (such as SHP-1 and CD45), members of the PIAS (protein inhibitor of activated STAT) family proteins and finally the members of SOCS family [371]. However, to add to the intertwining structure of the JAK-STAT signalling pathway, it has become apparent that a single STAT can induce different SOCS, and, in turn, a single SOCS can regulate different STATs [77]. Moreover, SOCS1 can also negatively regulate the expression of SOCS3, and *vice versa* [372].

Therefore, under a mechanistic perspective, it has been very challenging to identify the proteins that belong to these feedback circuits and how they respond to various specific inputs, assembling them into signal transduction networks, and establishing interactions between them. Certainly, it has led to a growing appreciation of the complexity of these networks.

A possible solution to overcome this challenge is to gain a better insight into how members of these networks dictate the functional behaviour of these systems. Specifically, how the concentration, localisation, specific interactions and dynamic expression of these proteins regulate the processing of information within the cell [140].

Nevertheless, it is also worth noting that, although the feedback inhibition circuit is a significant parameter for oscillatory behaviour, it is insufficient for its emergence [133].

There are several other essential conditions, which rely on the kinetic parameters of molecular turnover and delays [133, 373].

Therefore, it is important to address these parameters to gain insight into the nature and mechanism of global changes in SOCS3 distribution and dynamics within a cell. Identifying these kinetic parameters of protein turnover and transcription delays as well as how they influence the dynamic expression of SOCS3, might lead to a better understanding of the apparent promiscuity of SOCS3 protein activity in response to different stimuli or drug treatment.

4.1.1 Protein turnover

As we have highlighted previously, molecular oscillation have been identified in a vast array of signal transduction pathways, and recent studies are beginning to demonstrate the significance of ultradian oscillations and their potential factors that drives them.

Since feedback loops are ubiquitously found in a variety of intercellular and intracellular signalling pathways, delay-driven oscillations could, theoretically, be also widespread. However, various mathematical modelling studies have demonstrated that delayed feedback drive dynamic behaviours only if the relevant mRNA and protein half-lives are sufficiently small relative to the delay [136, 373].

To highlight the importance of protein half-life in ultradian oscillations, Hirata et al. investigated a mutant clone of Hes7 that exerts a longer half-life (about 30 minutes) compared to the wild-type protein (approximately 22 minutes) but exhibits similar repressor activity [374]. Furthermore, knock-in mice expressing the mutant Hes7 showed that somite formation was severely disturbed following a few cycles of segmentation. Further analysis with mathematical simulations with altered conditions

for protein half-life was reproducing the Hes7 protein stabilization, concluding that instability of Hes7 is crucial for sustained oscillation and for its function as a segmentation clock [374].

The turnover rate of a protein (half-life) is a crucial feature of protein homeostasis (proteostasis). In dividing cells, proteins can be recycled by two general processes: dilution by cytokinesis and proteolytic degradation [375]. Proteins with half-lives longer than the cell cycle time are mainly recycled by cell division. Alternatively, short-lived proteins with half-lives shorter than the cellular growth are mostly removed by the cell's degradation pathways [376].

The harmony between protein production and proteolytic degradation, also known as turnover, is crucial for maintaining protein abundance and regulating proteostasis. Dysregulation or malfunction of the cellular proteome (proteostasis) can lead to increased susceptibility to various diseases including carcinogenesis and age related diseases [377].

Investigating protein turnover can assist in understanding the recycling mechanisms of various proteins within a cell and deciphering the molecular pathways involved in cancer and age-related disorders. Despite the importance of protein turnover, our knowledge of the dynamic mechanisms underlying protein turnover is very limited [378].

To gain insight into the nature and dynamics of the proteome in a living cell, protein turnover and protein mobility must both be investigated. Generally, this would entail more than one method. Several research reports have used fluorescent proteins (FP) to study expression levels in single cells, either by fusing the FP to a protein of interest or by inserting the FP under the control of a promoter under investigation [355-357].

The transient or stable expression of a FP-tagged protein construct in cells is a simple, useful and rapid tool that enables the live-imagining of the direct subcellular localization and dynamic behaviour of proteins by confocal microscopy. However, FP-based imaging has several limitations in terms of resolution and sensitivity [243]. For instance, a well-known limitation of GFP and its variants is that it tends to localize at the nucleus. Moreover, if the expression of the protein of interest is low or if the protein is localized in different compartments of the cell in low levels, the detection is difficult [358, 359]. Furthermore, the maturation time of FP can range from tens of minutes to an hour which makes the detection and analysis of protein dynamics even more challenging [379, 380]. In several signalling pathways, the precise timing of protein expression is tightly regulated since it can affect the output of the system [138, 360]. In addition, it has been shown that the stable fold of FP results in a very long half-life in the cell, which influences the regulation of oscillatory mechanisms and the transient protein synthesis [359]. In support of this notion, strong supporting evidence was derived from the first research study to investigate the oscillatory behaviour of SOCS3, where Yoshiura and his colleagues [61] revealed ultradian oscillations of SOCS3 and pSTAT3 expression upon serum treatment in nutrient deprived mice fibroblasts. One of their research questions was to study whether the dynamic activation of pSTAT3 relies on SOCS3 oscillation. When SOCS3 was cloned into a vector and expressed under the control of a CMV promoter in fibroblasts, the pSTAT3 dynamic behaviour was abolished, demonstrating that SOCS3 overexpression inhibited the formation of pSTAT3.

Furthermore, although the use of fluorescent reporter proteins represents an important tool for investigating protein localization, abundance and spatial dynamics, it is unable to delineate a protein's temporal dynamics. To tackle this issue, two main biological

methods have been designed. The first method utilizes photoactivable or photoswitchable FPs that are capable to alter their spectral properties in relation to irradiation with light of specific wavelengths [381]. The second method that has been gaining a lot of attention in the last few years, employs fluorescent timer proteins that alter their emission spectrum over time [382].

4.1.2 Timing protein turnover and degradation

Fluorescent timer proteins were initially introduced almost two decades ago; where Terskikh and colleagues introduced random mutations into drFP583 (red fluorescent protein) and discovered one mutant (named E5) that alters its fluorescence from green to red over the course of 18 hours in vitro. These colour changes rely on a fluorescence resonance energy transfer (FRET) among the early maturating green (5-6 hours after expression) and late-maturating red (more than 9 hours) monomers within a DsRed-E5 tetramer. The authors demonstrated that since E5 changes from green to red fluorescence over time, it can be utilized as a timer reporter for gene expression [383].

Until recently, fluorescent timers have received a low attention because the initial timer introduced, it exhibited a tetrameric state with a propensity to aggregate in cells; subsequently limiting its widespread use. Furthermore, the long maturation time of DsRed-E5 is not ideal for mammalian cell studies as it exceeds the period of a cell cycle and most proteins will be diluted by cytokinesis by the time DsRed-E5 fully matures [384].

Michael Knop in collaboration with Elmar Schiebel and their colleagues recently designed 'tandem' fluorescent timers, also known as tFTs, which involve a fusion of two single-colour fluorescent proteins [385].

Specifically, the function of a fluorescent timer depends on the tandem fusion of two fluorescent proteins that undergo maturation with different kinetics, thus referred to as fastFP (faster maturation) and slowFP (slower maturation). Consequently, this difference in maturation kinetics between the two fluorescent proteins drives a tFT to act as a timer and alter colour as a function of time [386]. Tandem fluorescent timers can be assembled with virtually any combination of FPs as long as their emission spectrum do not overlap extensively and demonstrate different maturation kinetics. Since there is a vast array of FP to choose for timer design, this raises the question of which fluorophore combination to select [387].

The Knop-Schiebel research group demonstrated the versatility of tandem fluorescent timers (tFTs) by designing a tFT composed of the fast-maturing monomeric green fluorescent protein super-fold GFP (sfGFP) [388] and the slower-maturing monomeric red fluorescent protein mCherry [389]. Khmelinskii et al. 2012 [385] described the maturation time of mCherry as a two-step process that involved maturation half-times of 16.91 ± 1.23 minutes (initial step) and 30.30 ± 1.88 minutes (second step), whereas sfGFP becomes fluorescent much faster with maturation half-time of 5.63 ± 0.82 minutes. Thus, the construct mostly exhibits a green fluorescence right after expression but gradually develops yellow fluorescence over time. This makes the mCherry-sfGFP construct particularly useful for investigating protein dynamics *in vitro* because the ratio of mCherry/sfGFP can be used as a descriptor of a protein pool's relative age [386].

Owing to the different maturation kinetics of the two fluorophores found in the system of tFT, a ratiometric analysis of the dual colour of a tFT can provide information about the average age of a protein and determine the rate of turnover. In a theoretical point of view, the slowFP/fastFP ratio is primarily low if a population of tFT tagged proteins

have been just synthesized, as only a low number of fastFP molecules had time to mature. This number monotonically increases in relation to time as the population ages till it reaches a plateau of complete maturation (figure 4.2) [386].
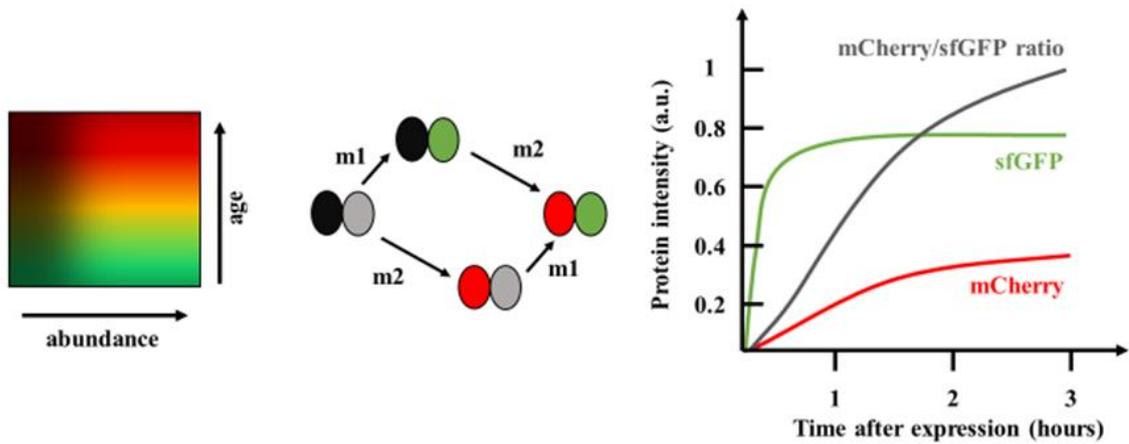
Figure 4.2: **Illustration of tandem fluorescent protein timer (tFT).** Due to the different maturation kinetics of the two fluorophores found in the system of tFT, they allow the easy visualisation of protein abundance and localisation, with a quantitative analysis of protein population age. A tFT is composed by a tandem fusion of a fastFP, defined by fast fluorophore maturation kinetics with rate constant m1, and a slowFP, represented by slow fluorophore maturation kinetics with rate constant m2. Fluorescence signal curves were measured according to published maturation kinetics for mCherry and sfGFP for a population of tFT tagged proteins initialized in the nonmature state in the absence of protein synthesis and degradation *[390]*. Image adapted from *[385]*.

However, on the subcellular level, it is evident that there are further conditions that will have a possible impact on the age of a protein, including the parameters of protein trafficking, or intracellular mobility. Khmelinskii et al. 2012 [385] performed an analysis of a system of ordinary differential equations characterising tFT dynamics and

demonstrated that in steady state the slowFP/fastFP ratio of signal intensities depends on the maturation kinetics of the tFT and relies on the degradation rate constant but it is independent on the production rate constant. Consequently, in the tFT protein system with independent rate constants of protein synthesis and degradation, the slowFP/fastFP signal ratio in steady state is measured only by the degradation rate. Thus, this tool can be utilized as a comparative measurement of protein degradation kinetics and apply in high-throughput screens for regulators of protein half-life [385].

So far, these fluorescent timers have proven to be very useful in the research field of genomics, proteomics and cell biology. These studies have included high throughput screening for regulators of protein degradation, mapping the inheritance of protein structures during cell cycle, analysing the up and downregulation of target promoters as well as investigating the relative age of target organelles or cell differentiation [383, 385, 391, 392].

Specifically, a recent research study system demonstrated the usefulness of tFTs by visualizing protein turnover in zebrafish [393]. Specifically, Dona et al. [393] tagged a G-protein-coupled receptor (GCPR) with a RFP-sfGFP tFT and demonstrated that the timer read-out can indicate receptor's age and therefore the turnover of the GPCR at the plasma membrane. Furthermore, the group showed that the GPCR-tFT behaved as a sensitive and reliable tool for tracking cytokine signalling activity in migrating cell collectives, which in this case exhibited a 'standing wave' that included a higher receptor activity at one end of the cell collective and lower signalling at the other. This standing wave of signalling provided the cell collective an intrinsic polarity that would have been very challenging to be visualised with other approaches [393].

4.1.3 Promoter Activity

A third important factor that regulates protein abundance in eukaryotic cells is the kinetics of transcription including the initiation of transcription, the level of promoter activity, pre-mRNA transcript splicing and regulation of the half-life of the processed mRNA. Theoretically, delays in these factors can provide the basis of an oscillatory behaviour of mRNA and therefore protein expression provided mRNA and protein half-lives are both sufficiently short-lived. Indeed, the coupled processes of transcript elongation, splicing, processing and export are complex and time consuming [373]. To highlight the importance of a delay in the negative-feedback loop, Takashima and his colleagues investigated whether the intronic delay, which represents the time necessary to transcribe and splice out intron sequences to generate mRNAs, plays a crucial role in the generation of stable oscillations of Hes7 in the presomitic mesoderm (PSM, the tissue that gives rise to the somites during embryonic development) [394]. The research group demonstrated that Hes7 oscillations were abolished by deletion of all three introns in mouse embryos, resulting in severe fusion of somites [394]. This observation was not only recapitulated by mathematical modelling in Hes7 dynamic expression but also in p53 and NF-κB [373].

Transcriptional regulation offers the ability to robustly control protein synthesis in living cells. The frequency of transcript synthesis is mainly dependent on the rate of initiation, which is directly associated with promoter strength [395]. Initiation and suppression of promoter induction is a critical method utilized by cells to control gene expression and results in different frequencies of protein production. High-frequency oscillations such as ultradian oscillations benefit from time delays originating from biological process that involve transcription, translation and synthesis, all of which are

dependent on the activity of promoter. Therefore, investigations of promoter sequence and its regulatory elements are essential for understanding global gene regulation and, importantly, studying how the promoter of an oscillatory protein behaves in response to different stimuli [396].

Reporter gene assays have been widely used to identify important regulatory sequences within a promoter and study gene initiation. The design of a reporter gene fusions conventionally involves the cloning and fusion of DNA fragments containing promoters or other regulatory sequences to a reporter gene [397]. Previous studies have used a wide array of reporter genes to study gene expression. The most frequently used assays include the use of β-galactosidase (*lacZ*), chloramphenicol acetyl transferase (*cat*), luciferase (*lus*), and green fluorescent protein (GFP) being utilized for multiple applications in living cells. The usual preferred method of incorporating the reporter is by using a knock in system such as CRISPR or BAC (large capacity of cloning vector) approach as it incorporates the reporter under the control of the endogenous genomic locus [395, 397].

Real-time gene expression in living cells cannot be measure with highly stable reporter protein due to the accumulation of residual reporters. Therefore, it is crucial that reporters are degraded when the gene is not being induced. The half-life of wild type GFP is measured to be 26-54 hours depending on the culturing conditions. On the contrary, the half-life of firefly luciferase is ranging between 3 to 4 hours and it is more sensitive than conventional GFP. Therefore, luciferase is typically the more preferred tool for studies investigating dynamic patterns, as it offers the ability report rapid changes of gene expression [398, 399].

Real-time transcription activity of SOCS3 expression has been studied in a variety of cell systems in response to a vast array of stimuli. Typically, SOCS3 protein is generally expressed at low levels in cells and its activity is highly induced transcriptionally in response to treatment of different cytokines, such as interleukin-2 (IL-2), IL-3, IL-4, IL-6, leukaemia inhibitor factor (LIF), growth hormones (GH) and erythropoietin [400, 401].

To date, a 1.7kb fragment of the human SOCS3 promoter has been cloned and studied. The first research report that successfully cloned the human SOCS3 promoter and studied its activity used a luciferase reporter, the authors identified an incomplete estrogen responsive element (ERE) motif located -1504 to -1484 upstream of the start site [402]. A more recent study conducted by Wiejak et al. [213] also investigated the same SOCS3 promoter fragment driving the expression of luciferase and identified the flavonoids naringenin and flavone as effective inducers of promoter activity. Subsequently a minimal functional SOCS3 promoter fragment was described that comprises two serum responsive elements (SREs), one proximal (-1069/-1060) and one distal (-1046/-1037), both sites are indicated in cyan colour in figure 4.3. The minimal functional SOCS3 promoter also contains a GC-rich region and a putative TATA box [403].

Interestingly, the new type of fluorescent protein timer now makes it possibly to follow promoter activity in real time. When Terskikh and colleagues developed the very first fluorescent protein timer (called E5) in 2000 [383], they examined the colour shift in spectrum over time under the control of inducible promoters, such as heat shock-dependent and tetracycline transactivator promoters (*Tet-On* and *Tet-Off*). They demonstrated that when the timer was under the control of a promoter that was just

induced, the signal of fluorescence was predominant green during the first hours of activity, whereas when the activity of promoter was constitutively active, it led to a mixture of green and red fluorescence. Only red fluorescence remained, few hours after the promoter became inactive [383]. To our knowledge, there has not been a research article that investigated where there is a link between promoter activity and SOCS3 protein oscillation.

Furthermore, the only evidence that establishes an implied link between SOCS3 promoter activity and its dynamic behaviour has been when various studies have treated cells with a certain type of cytokine, such as IL-6, and analysed the fluctuating protein behaviour over time. For instance, Yoshiura et al. [61] showed an oscillatory pattern of SOCS3 mRNA after stimulation with IL-6, demonstrating that the dynamic behaviour of pSTAT3-SOCS3 feedback system can be also induced by a single stimulator rather than a serum treatment that includes a mixture of stimulators. However, it will be interesting to test if the protein of SOCS3 can also exhibits a dynamic pattern by treating cells with IL-6 or other type of cytokine.

-1673

```
AGTGGGCCGA GGCTGGGTAG CTGCCACGAA GGGGCGTCCC CTGCAGCCCT GGGTCTCCCC TCTGGAATCT
GCCCGCAGGT GACTGTCGCA CGTCTCCAAC CTCCGGCTCC CGGGTCTGAC CCCAGCCCCG CTGCCTGGCT
GTGGGGTAGC CCCACTTCCC TTTCGGCTGG CCCGGGCCGC CTGACCCGCA GTTGGGCCCT CGGGGAGGCC
CGGGTGGCGG CGAGGGGGTG TCCCCGCGGG GTCCGGGAAA AGGGGAAGGG GAACCGGGAG
GCTCTCCAGG TCGGCCTCCT AGAACTGCCC GCTCTCCCGA AGCGGCGCCC CCGCCTCTGC CAGAAATCAG
CCTTCTTAGA AGGGAGGGGG GTGGGGGGAA AGTGTGAATG AGAAGTTGGG GGCGGAGCGC GCGGCGGGGA
GGGGCCGCTG CCAGGAACGC GCCGCCGGGG CTGGCGCCGC GCCCACCGGC CGCCTCGGCC GCCTCTCGTC
GCGCTTTGTC TCCGCGCGCG TCCCTCCCGG TCCCTGCCCC TGCTCGCGGC CCGCCCTCGG CGCCCGCGGC
CCCTCCCTCA CCCTCCGCGC TCAGCCTTTC TCTGCTGCGA GTAGTGACTA AACATTACAA GAAGGCCGGC
CGCGCAGTTC CAGGAATCGG GGGGCGGGGC GCGGCGGCCG CCTATATACCC GCGAGCGCGG
CCTCCGCGGC GGCTCCGACT TGGACTCCCT GCTCCGCTGC TGCCGCTTCG GCCCCGCACG CAGCCAGCCG
CCAGCCGCCC GCCCGGCCCA GCTCCCGCCG CGGCCCCTTG CCGCGGTCCC TCTCCTGGTC CCCTCCCGGT
TGGTCCGGGG GTGCGCAGGG GGCAGGGCGG GCGCCCAGGG GAAGCTCGAG GGACGCGCGC
GCGAAGGCTC CTTTGTGGAC TTCACGGCCG CCAACATCTG GGCGCAGCGC GGGCCACCGC TGGCCGTCTC
GCCGCCGCGT CGCCTTGGGG ACCCGAGGGG GCTCAGCCCC AAGGACGGAG ACTTCGATTC GGGACCAGGT
AGGAAGGAGG AGCGCGGCGT GGGGAGGGG TCTCGCTCAG TCCCGGGAGC TTTTCCCGGTT TCCCTCCCCT
TCCCGGGTCA TTCCCGGCAG GGAGGTGACG AGGTAGGGGC AGAGCGGATG GAAGCCGGAG ATCCCAGGTT
CCCGGAATAC TCCGGCTGGG GCCTTCGGGC TTCTCCTGTC CCCTCCCTAC CCCCGTGCCT CGGGTTTCTC
CCTCCGTCCA CACCGCCCGG GGCTACTGGA CTGAGCGGCG CCCAGGCAGT CCCGGGGGCC CTTCTCCTGC
TCCCCACCCG GCCACACTCC TGGAGACCTA ACTTCCGCGC GCGAGTTTCC CACGCTGCGC CCTTGCAGTG
CGCGCCTGGG AAGGGGCTGC CCGGGGCCAC CCTGCCGGCA GGGGCGGGAG CCGTGCGGGC TCCGTGAGGC
GCCTGGATCG GAGCGCGGGC CCAGGAGAGG GCCCCGGGG CAGTGGGTGC CCCAGTCGCT CGGCGAAGGC
AGGGGAGCCG GGGCGGGCCG GGCGCGCTGG AGGGTTCCGG GCACTCAACG CGCTCGCGCC
TTCCTCTCCGC AGCCCCCGG GATGCGGTAGC GGCCGCTGTG CGGAGGCCGC GAAGCAGCTG
CAGCCGCCGC CGCGCAGATC CACGCTGGCT CCGTGCGCCATG
                        +1
```

Figure 4.3: The 5' promoter region of human SOCS3. The start codon ATG of the SOCS3 gene is indicated in a grey box and defined as +1, whereas the 5' end of the promoter cloned is shown as -1673. The two putative STAT-binding regions are indicated in a cyan box. The G-rich region is defined in a yellow box and the putative TATA box is shown in a green box. An incomplete ERE motif is indicated in a pink box. This genomic sequence is identical to a region of the human genomic database sequence (GeneBank Accession No. AC061992).

4.1.4 Aims

The aim of this current chapter is to investigate whether the use of fluorescent timer proteins can shed light into the complexity of SOCS3 protein dynamic behaviour and how it relates protein turn over and degradation as well as critically assess their usefulness to study protein half-life and cellular localisation.

Furthermore, it aims to provide a detailed characterization of tFTs as a reporter of SOCS3 protein age and degradation. Furthermore, we will carefully assess their use to analyse SOCS3 dynamics across spatial and temporal dimensions under either normal culture conditions or different treatments that are known to impact on SOCS3 half-life and degradation rates.

Finally, this chapter will also examine whether the use of a tandem fluorescent timer under the control of a SOCS3 promoter can exhibit a dynamic readout of SOCS3 promoter activity in cells.

4.2 Material and Methods

4.2.1 Cell Culture

The mouse embryonic fibroblast Flp-In-3T3 cell line was obtained from Invitrogen™ (Thermo Fisher Scientific, C # R76107). The Flp-In-3T3 cell line was derived from NIH-3T3 cells by transfection with the vector pFRT/lacZeo. Therefore, this cell line utilizes a Flp recombination target (FRT) for the Flp recombinase-mediated integration of a Flp-In expression vector and zeocin resistance (figure 4.4). The pFRT/lacZeo construct allows the development of stable clonal cell lines [404].
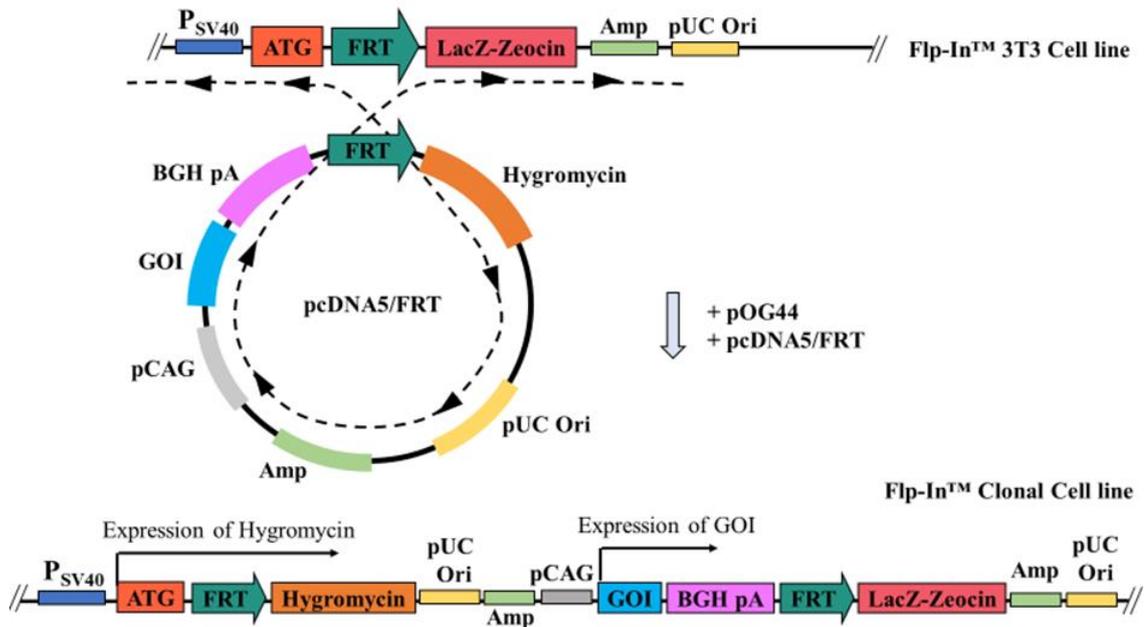
Figure 4.4: Schematic Illustration of the Flp-In System. The pOG44 vector and pcDNA5/FRT plasmid that also includes the gene of interest (GOI) are transfected into the FLP-In™ 3T3 cell line. Upon transfection, the FLP recombinase is expressed from pOG44 vector and drives the homologous recombination process between the two specific FRT sites on the pFRT/lacZeo construct and the pcDNA5/FRT expression vector. Integration of pcDNA5/FRT into the 3T3 genome at the FRT region places the SV40 promoter and the ATG start codon (from pFRT/lacZeo construct) into frame with the Hygromycin resistance gene, consequently inactivating the lacZ-Zeocin™ fusion gene. Therefore, the resulting stable clonal Flp-In™ cell line can be selected for hygromycin resistance and expression of the recombinant protein under investigation [404]. Image adapted from [404].

Flp-In-3T3 cell line was cultured and kept under standard cell culture conditions. In brief, cells were detached with TrypLE™ Express Enzyme (1X) (ThermoFisher Scientific, C# 12604013) and replaced in complete culture medium consisting of Dulbecco's Modified Eagle Medium (DMEM), high glucose, pyruvate (Thermo Fisher Scientific, C# 11995040) supplemented with 10 % (v/v) Foetal bovine serum (FBS), 1

% (v/v) HEPES 1M (Gibco™, Thermo Fisher Scientific, C# 15630080), 1 % (v/v) Penicillin-Streptomycin solution (Sigma-Aldrich, C# P4333) and 100 µg/ml of Zeocin™ Selection Reagent (Gibco™, Thermo Fisher Scientific, C# R25001).

Cells were maintained at 37°C in the 5% CO_2 atmospheric condition. Medium was changed every 2 to 3 days. When cells reached the subculturing density of 90% confluence, they would either be passaged into a new T75 culture flask for further culture or into µ-Plates 24 well plate with flat and clear bottom (Ibidi®, Planegg, Germany) for fluorescence imaging.

4.2.2 Vector design and construction

For the analysis of SOCS3 protein turn-over, two pCDN5/FRT vectors were designed to contain the SOCS3 coding sequence carrying the fluorescent timer (mCherry/sfGFP construct) either at the SOCS3 N-terminus or the C-terminus. An additional pCDN5/FRT vector that carried only the fluorescent timer sequence was used and served as a control (figure 4.5).

The pCR™-Blunt vector containing the SOCS3 sequence, the pCDN5/FRT empty vector, the pCDN5/FRT vector containing only the fluorescent timer, the pCDN5/FRT vector carrying the fluorescent timer at the SOCS3 N-terminus and the pOG44 plasmid were kindly provided by Dr Richard Mort.

(A)

(B)
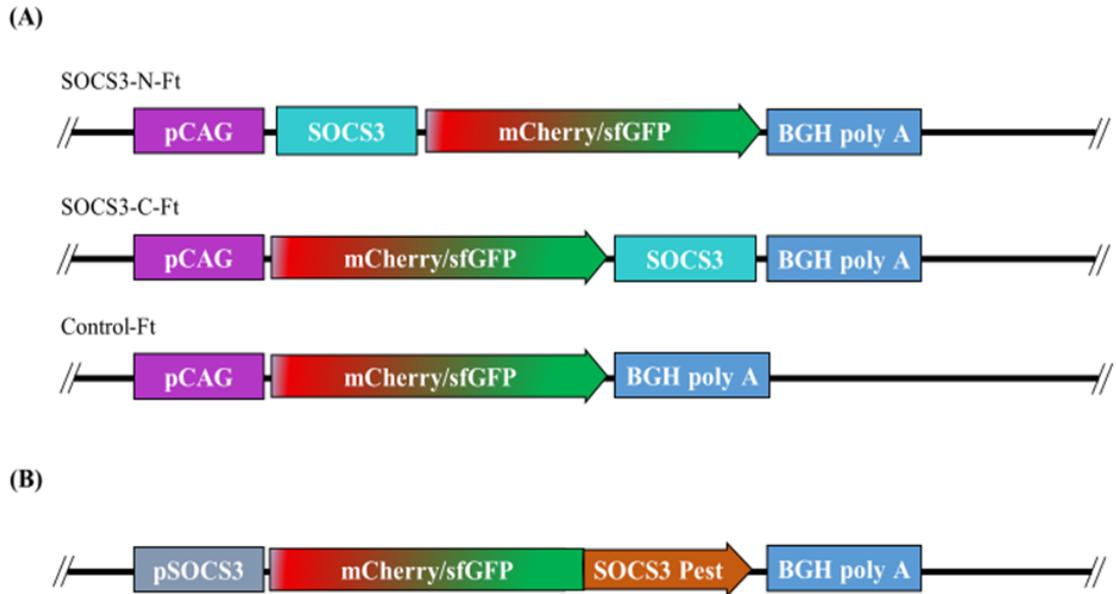
Figure 4.5: Plasmid constructs used to generate Flp-In™ 3T3 Cell lines. (A) Schematic representation of SOCS3-Ft N-terminal, C-terminal fusion and its control Ft vector. SOCS3 expression is driven by the hybrid CMV enhancer/chicken β-actin (CAG) promoter**. (B)** A construct plasmid was generated to report SOCS3 promoter activity. The expression of the Ft is driven by the 1.7 kb fragment of the 5-flanking region of SOCS3 promoter. SOCS3 pest sequence was fused to the C-terminus of the Ft.

The pCDN5/FRT vector carrying the fluorescent timer at the SOCS3 C-terminus was generated using the pCR™-Blunt vector containing the SOCS3 sequence as follows: the 675 bp coding sequence of SOCS3 was PCR amplified using a forward primer 5'-CAAAAAAGCTTATGGTCACCCACAGCAAGTTTC-3' and a reverse primer 5'-CAAAAAACGCGTAAGCGGGGCATCGTACTGG-3', where the underlined sequences represent restriction enzymes HindIII and MluI, respectively. The size of the amplified sequence was verified by running 5 μl of the PCR mixture on a 1% agarose gel. The PCR product was purified using the PureLink PCR Purification Kit (Thermo Fisher Scientific, C# K310001). Following PCR purification, the amplified SOCS3 fragment and the empty pCDN5/FRT vector were subjected into a double restriction

enzyme using Mlu1 and HindIII enzymes. The final mixtures were incubated at 37°C for 1 hour. Following incubation, the vector was mixed with 1 μl of FastAP Thermosensitive Alkaline Phosphatase (Thermo Fisher Scientific, C# EF0651) and incubated for 10 minutes at 37°C. This step was necessary to prevent the recircularization of the backbone in the ligation reaction. Following incubation, the mixtures were incubated at 75°C for 10 minutes to inactivate the restriction enzymes and FastAP Thermosensitive Alkaline Phosphatase. Both reactions were purified using the PureLink PCR Purification Kit.

The pCDN5/FRT vector carrying the fluorescent timer at the SOCS3 N-terminus was cloned using the same procedure with primers carrying the restriction enzymes HINDIII and AgeI.

For the analysis of SOCS3 promoter activity, the SOCS3-C-Ft plasmid was initially subjected to a double digestion with Spe1 and Mlu1 enzymes to excise the CAG promoter [405] and SOCS3 CDS. The digested DNA was then size separated by running it on 1% agarose gel in TBE buffer and the pCDN5/FRT backbone was gel purified using the Zymoclean™ Gel DNA Recovery Kit (ZYMO RESEARCH, C#D4007) as described below.

A 1.7 kb fragment of the human SOCS3 5' regulatory region was amplified from DNA extracted from Human Intestinal Epithelial Cells (HIECs) and using specific primers that carried suitable restriction enzyme sites (table 4.1). Isolation of genomic DNA was performed using the GeneJET Genomic DNA Purification Kit (ThermoFisher Scientific, C# K0721) according to the manufacturer's protocol (described in Chapter 3, Materials and Methods). Moreover, a short region of the C-terminus of SOCS3

known to contain the PEST sequence was also PCR amplified using the primers listed in table 4.1.

Table 4. 1: **Primers used for the construction of SOCS3 promoter-Ft fusion plasmid.**

Primers Name	DNA sequence (5'→3')
F-SOCS3-promoter	CAAAAAA<u>CTAGTA</u>GTGGGCCGAGGCTGGGTAG
R-SOCS3-promoter	CAAAAAA<u>ACGCGT</u>CATGGTGGCGGCGGTACCGGCGCACGGAGCCAGCGTGGATCTG
F-SOCS3-Pest	CAAAAAA<u>AAGCTT</u>CACTACATGCCGCCCCCTGGAG
R-SOCS3-Pest	CAAAAAA<u>ACCGGT</u>GGCTCTTCTGGGGGGACTCC

Spe1 (CTAGTA), Mlu1 (ACGCGT), HINDIII (AAGCTT), AgeI (ACCGGT)

Both PCR products were analysed using agarose gel electrophoresis to verify its correct amplification. Following agarose gel electrophoresis, the hSOCS3 promoter was subjected into a double restriction enzyme using Spe1 and Mlu1 enzymes, treated with FastAP Thermosensitive Alkaline Phosphatase and purified using the PureLink PCR Purification Kit.

Following purification, the hSOCS3 promoter was first ligated and cloned into the linearized pCDN5/FRT backbone and DNA sequencing was used as a method to confirm correct hSOCS3 promoter insertion. Then, the SOCS3-promoter-Ft plasmid and PCR-amplified SOCS3 pest sequence were digested with Age1 and Xho1 enzymes, purified and subsequently ligated to form a new plasmid.

All plasmids were sequenced to ensure that the desired fragment was cloned and to the right orientation.

4.2.3 Gel extraction

Gel extraction was performed using the Zymoclean™ Gel DNA Recovery Kit according to the manufacturer's protocol (ZYMO RESEARCH, C#D4007). Once the digested DNA run on 1% agarose gel in TBE buffer, the desired DNA fragment was extracted and mixed with 3 volumes of ADB solution. The mixture was then incubated at 55°C until the gel slice was completely dissolved. The resulting dissolved solution was transferred to a Zymo-Spin™ Column in a Collection Tube provided by the kit. The column containing the solution was then centrifuged at 14,000g for 60 seconds and the flow-through was discarded. Following centrifugation, 200 µl of DNA Wash Buffer was added to the column and centrifuged for a further 30 seconds at 14,000g. The wash step was repeated further two times. Following the last DNA wash, 6 µl of DNA Elution Buffer was added directly to the column matric. The column was then placed in a new sterile Eppendorf and centrifuged at full speed for 60 seconds to elute DNA. The purity of the gel excised DNA fragment was confirmed by running a second agarose gel.

4.2.4 Sequencing

Purified PCR products and plasmids were stored at 4°C before shipment for Sanger sequencing by Source BioScience Sequencing (Nottingham, UK). Primers used for sequencing are shown in Table 4.2.

Table 4. 2: Primers used for sequencing.

Primer Name	DNA sequence (5'→3')
F1-ChimTimerTag	GTTCGGCTTCTGGCGTGTG
F2-TimerTag	CAGACACAGCACAGGCGGCATG
F3-SOCS3	GCTGGTGCACCACTACATGCC
F3-TimerSOCS3	GAAGAGGGACCACATGGTGCTG
R1-Timer Tag	CCGTCCAGTTCCACCAGGATGG
R2-Timer Tag-socs3	CCTCGATCTCGAACTCGTGGC
R-bghpolyA	CGCTTAATGCGCCGCTACAGG
R-TimerSOCS3	GCTGAAGGTCTTGAGGCGCAG
F1-SOCS3-promoter-Seq	GGAGGTCGCTGAGTAGTGCG
R1-SOCS3-promoter-Seq	CTCGCGGGTATATAGGCGGC
F2-SOCS3-promoter-Seq	GCCTATATACCCGCGAGCGC
R2-SOCS3-promoter-Seq	CGCGTTGAGTGCCCGGAAC
F4-Pest	GCACCCAGAGCAGACTGAGC
R4-Pest	TAGAAGGCACAGTCGAGG

The results obtained from Source Bioscience Sequencing were analysed by aligning the sequenced DNA sample with the expected DNA sequence. The DNA alignment was achieved by using the online tool ClustalW2 from EMBL-EBI website (https://www.ebi.ac.uk/Tools/msa/clustalw2/). Results are shown in the Appendix.

4.2.5 Electroporation

Prior to electroporation, all plasmids were purified using PureLink™ HiPure Plasmid Filter Maxiprep Kit (Invitrogen™, C# K210017).

Electroporation was performed using the Neon® Transfection System 100 µl Kit (Thermo Fisher Scientific, UK) according to the manufacturer's instructions. Briefly, prior electroporation, cells were detached using TrypLE™ Express Enzyme (1X) (ThermoFisher Scientific, C# 12604013), washed with DPBS without Ca^{+2} and Mg^{+2} and resuspended in "R buffer" or "T buffer" from Neon kit to a concentration of 5 x 10^6

cells/100 µl and total volume of 200 µl. For each reaction, 18 µl of pOG44 DNA at a concentration of 1 µg/µl and 2 µl of donor vector at a concentration of 1 µg/µl was added to the cell mixture and mixed carefully. Extra care was taken to not allow the total volume of DNA to exceed 10% of the total volume of the mixture.

For each reaction, cells were loaded into a 100 µl tip and given 2 pulses of 1350 V for 20 ms. This step was performed twice to electroporate the whole 200 µl cell mixture using the same tip. For each clonal cell line, the electroporated cells were transferred immediately to a T75 flask containing 14 ml of pre-warmed Opti-MEM™ 1 Reduced Serum Medium (Gibco™, Thermo Fisher Scientific, C# 31985070) without antibiotics and then incubated for 2 days at 37°C in a 5% CO_2 incubator. 48 hours after transformation, the clonal cell lines were subjected to drug selection by changing the medium to DMEM supplemented with 10 % Foetal bovine serum (FBS), 1 % (v/v) HEPES, 1 % (v/v) Penicillin-Streptomycin solution and 100 µg/ml Hygromycin B solution (Thermo Fisher Scientific, C# 10687010). The flasks were maintained under these conditions for an additional 2 weeks and once the drug resistant colonies appeared, selection continued for an additional 2 weeks. The antibiotic resistant colonies were subsequently expanded and propagated in the presence of Hygromycin.

4.2.6 Confocal Microscopy

A Zeiss LSM880 laser scanning confocal microscope was used to visualized and acquire multi-channel data. During imaging session, the stage incubator of the confocal microscope was preheated and maintained to 37°C and 5% CO_2 was supplied to the culture plate.

Green and red fluorescence (40x objective, 5% laser power, green: excitation at 488 nm, emission at 489-550 nm; red: excitation at 561 nm, emission at 568-664 nm, A.U. = 1) were acquired sequentially by line using Transmitted light detector (T-PMT).

Furthermore, images from a well containing medium and a reference solution that was generated by adding same quantity of a green (GFP) and red (RFPs) dye solution was acquired.

Owing to the extreme difference in fluorescence intensities between CAG and SOCS3 promoter driving the expression of tFT, two different acquisition settings had to be used. SOCS3 promoter activity was acquired using 12% laser power.

4.2.7 Image analysis of lifetime ratios

Images were process with custom written macros for Fiji (ImageJ) software [406]. The complete image analysis code was generated by Dr Richard Mort. Cell parameters were acquired using manual segmentation in Fiji. Specifically, the green channel served as a reference channel due to the brightness of sfGFP and, from this, a binary mask was generated by smoothing and adaptive thresholding. From the binary mask, the cell parameters were collected and saved as regions of interest (ROI). These regions along with the cell nucleuses were then loaded to a custom written macro to obtain the perinuclear, periphery and intracellular regions as shown in figure 4.6.
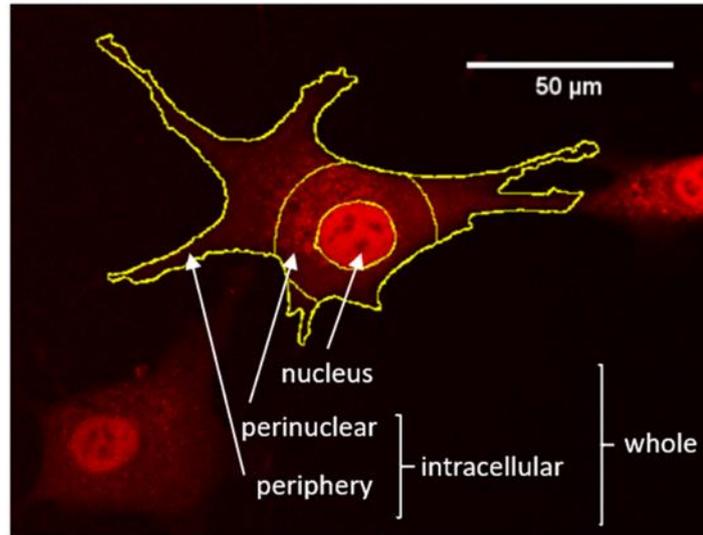
Figure 4.6: Cellular regions used to quantify mCherry/sfGFP intensity ratios. The mCherry/sfGFP signal ratio of each subcellular location within the cell of interest was quantified individually.

Flat field, cell autofluorescence and background corrections were applied to all images and mean fluorescence intensities of individual cells were quantified. A general workflow that outlines all the steps undertaking to process the images and quantify the ratios are shown in figure 4.7.
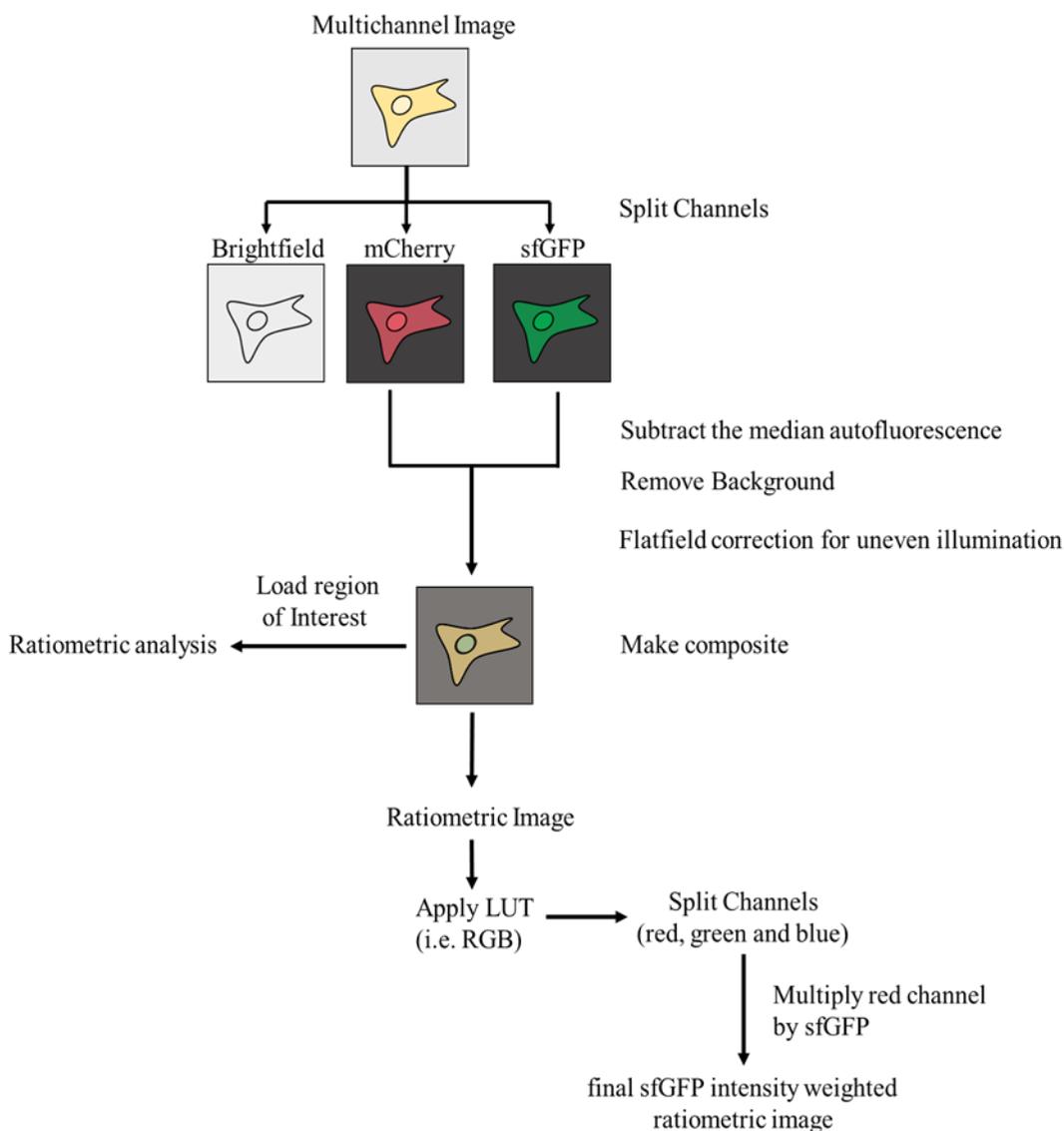
Figure 4.7: Image processing workflow. The chart demonstrates a pipeline for obtaining ratiometric data from cells expressing Ft and the method by which intensity weighted ratiometric images were obtained.

Red/green fluorescence signal ratios were estimated and normalized with the use of a homogeneous RFP-GFP reference solution acquired with the same acquisition parameters. Ratios were optimized by dividing by the ratio of the purified timer reference solution. This step was important to compensate for day-to-day variations of the microscope and allowed the direct assessment of images acquired on different

sessions. For visualization, images were colour-coded using the ratio lookup table in Fiji.

The acquired time-lapse fluorescence movies were obtained as stacks in tiff format and further analysed using Fiji software. Then each stack representing a single time point was analysed individually as described previously.

4.2.8 Statistical Analysis

The mean ratios in each group were compared by two-way analysis of variance (ANOVA). Where the two-way ANOVA significance level was < 0.05 pairwise comparisons were performed using a Tukey's HSD (honestly significant difference) test for multiple pairwise comparisons. Details of statistical tests are indicated in the figure and table legends. In box-and-whiskers plots the middle bar is the median, the hinges show the interquartile range (IQR). All analyses and graphs were performed in PRISM GraphPad version 8 (Software, San Diego, CA) and Gnumeric Portable (Software, GNOME Project & PortableApps.com).

4.3 Results

4.3.1 Strategy for tagging SOCS3 with a fluorescent tandem timer

Tandem fluorescent timers (tFT) incorporate a fast-maturing fluorescent protein and a slower maturing fluorescent protein fused to a protein of interest. The different maturation rates of the two fluorescent proteins are used to analyse protein age and stability by comparing the ration of the slow/fast fluorescent protein (figure 4.2). Degradation signals that are found in the protein of interest affect the tFT readout [385]. Thus, we utilised the tandem fluorescent protein timer (tFT) strategy to investigate the localisation of SOCS3 and examine how SOCS3 population age differs in distinct cellular compartments on manipulation of the proposed pathways that control its turnover.

Our chosen tFT tag consisted the fast-maturing fluorophore superfolder GFP (sfGFP) and the slower-maturing fluorophore mCherry because this combination has been shown to be sensitive over a temporal range of ~10 mins to ~3 hours [385, 390, 407] consistent with the reported oscillatory expression of SOCS3 with a period of about 2 hours [61].

Monomeric fluorescent proteins tend to be more suitable for generating fusion constructs as they appear to be the least disruptive to the function of the protein. However, we sought to investigate whether the position of the tFT at either the N- or C-terminus of SOCS3 influences the subcellular localization of the fusion protein and if it does, how it differs between the two fusion proteins. Therefore, we generated two constructs, where tFT was inserted at either the C- or the N-terminus of SOCS3 CDS using a PCR cloning method.

4.3.2 Expression of SOCS3-C-FT and N-FT-SOCS3 in stable NIH 3T3 cell lines

In order to produce a cell system stably expressing the SOCS3-C-Ft and N-Ft-SOCS3 constructs at moderate expression levels we used the Flp-in system (Thermofisher) to generate stable isogenic cell lines with a single integration of SOCS3-C-FT or N-FT-SOCS3 driven by the synthetic CAG promoter [404]. We chose the mouse embryonic fibroblast Flp-In™ 3T3 cell line as our host cell line because it has previously been used to investigate SOCS3 function [61, 311].

An additional cell line was generated by integrating a construct expressing the tFT alone under the control of CAG promoter. This cell line served as an internal control to distinguish the localization and degradation dynamics generated by the Ft from those driven by the SOCS3 fusion constructs.

4.3.3 Characterization of cell lines expressing SOCS3-fluorescent timer fusion constructs

To investigate the localisation and turnover of SOCS3 protein within cells, live cell imaging using confocal laser scanning microscopy was performed (see Methods). Images were processed as described in the materials and methods section. Briefly, for quantitative comparisons we compared the mean mCherry/sfGFP ratiometric value in a given subcellular region. For qualitative assessment we applied a Green-Red look up table (LUT) to the ratiometric image and then sfGFP weighted the resulting image (using the sfGFP intensity values as a multiplier) so that the oldest protein populations appeared red and the youngest green (figure 4.8, column 5).

Nikoletta Kalenderoglou - November 2019
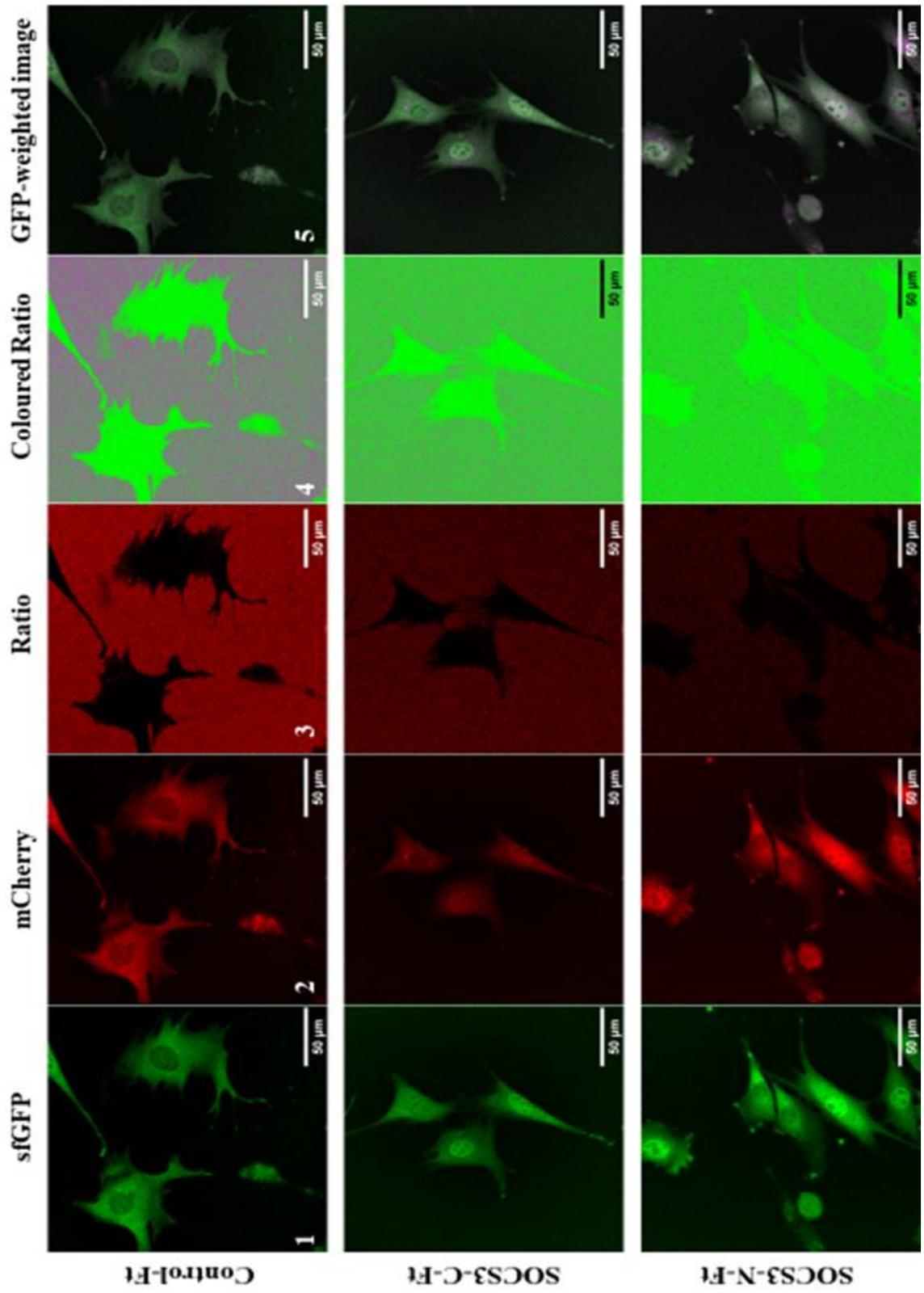
Figure 4.8: Live imaging of mouse embryonic fibroblast Flp-In™ 3T3 cell lines expressing fluorescently-labelled constructs. SfGFP (1) and mCherry (2) single-channel fluorescence microscopy images of cells expressing the Ft either alone, fused to the C- or N-terminus of SOCS3. The mCherry/sfGFP ratiometric image (3) was colour coded (4) and multiplied with the sfGFP image to generate the sfGFP intensity-weighted ratiometric image (5). Control-Ft exhibited an even distribution of sfGFP and mCherry fluorescence signal throughout the cell. SOCS3-C-Ft. SOCS3-C-Ft demonstrated a higher nuclear and perinuclear sfGFP abundance compared to the Control-FT. SOCS3-N-Ft cell line presented a nuclear sfGFP signal, but to a lesser extent, and an intracellular sfGFP/mCherry distribution pattern very similar to SOCS3-C-Ft. Scale bar 50 μm.

According to figure 4.8, row 1, the cell line expressing the Ft construct alone (Control-Ft) demonstrated that the fluorescence signal emitted from both sfGFP and mCherry was found throughout the cell, with nearly perfect co-localization of the two fluorescent proteins. It also showed a fine general punctuate pattern distributed over the intracellular region of the cell. As expected, the Ft alone did not establish any distinct localisation pattern in the cell.

On the other hand, when Ft was fused at the C terminus of SOCS3, we found that the localization pattern of fluorescence was quite different from that seen in Control-Ft. Specifically, the SOCS3-C-Ft cell line exhibited an enriched nuclear and perinuclear sfGFP signal. However, its intensity was significant lower in the periphery compared to the other cellular compartments. In the case of mCherry signal, the abundance of mCherry followed the same distribution pattern with sfGFP, but its intensity appeared to be lower in the nucleus. Moreover, mCherry showed a more distinctive punctuate pattern compared to Control-Ft and it was mainly concentrated at the perinuclear area of the cell.
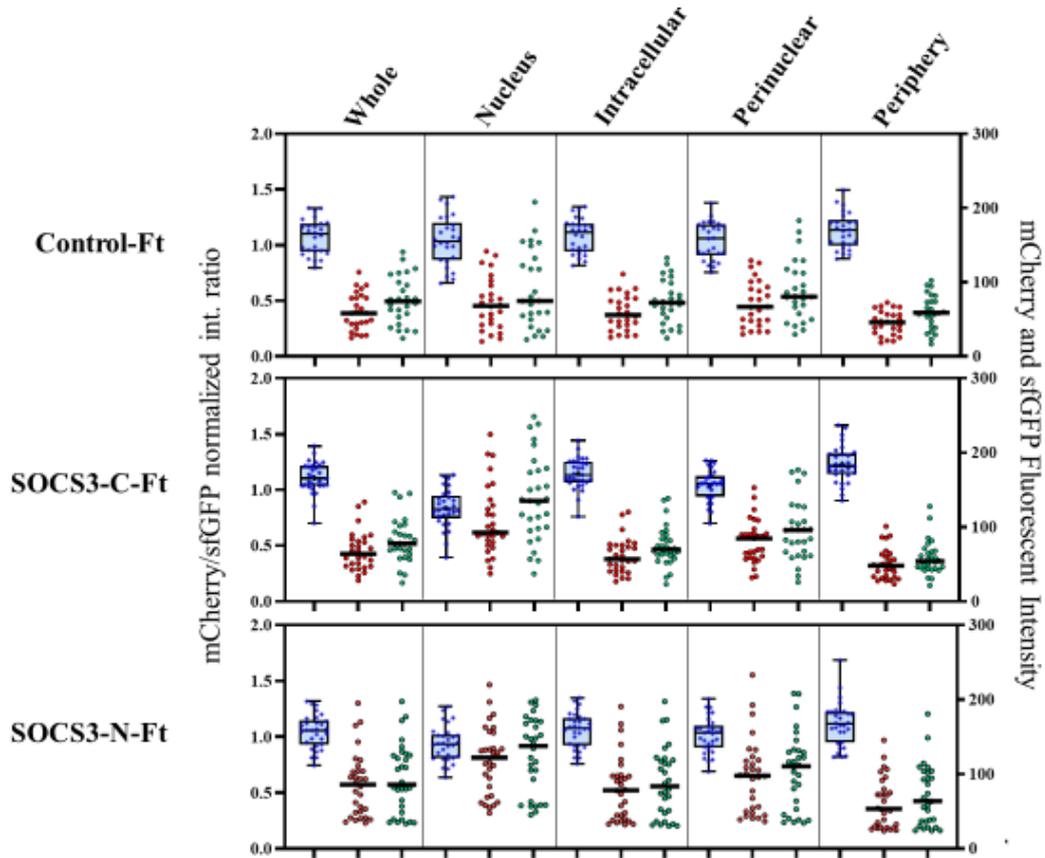
Figure 4.9: Ratiometric analysis of tandem fluorescent SOCS3 fusion constructs reveal distinct cellular and subcellular distribution patterns. For each Ft, the mCherry/sfGFP intensity ratios of individual cells were quantified and normalized to the median intensity ratio of a reference red/green solution. The mean values of nuclear and perinuclear fluorescent intensity ratios (FIR) are lower than the mean values generated from the peripheral FIR, suggesting that Ft-SOCS3 fusion protein is youngest in the nucleus and the perinucleus and with an older population in the periphery. Centrelines indicate the medians, box limits mark the 25th and 75th percentiles, and whiskers extend to the minimum and maximum values. Blue stars represent the individual cell ratios. The individual mCherry and sfGFP levels are displayed in red and green dots and are plotted against the right y-axis. Control-Ft, N=26, SOCS3-C-Ft, N=29 and SOCS3-N-Ft, N=30.

This is also clear from the sfGFP-weighted ratiometric images (figure 4.8 column 5), where tFT tagged SOCS3 protein is youngest in the nucleus with an older population in the perinuclear region consistent with a population that is undergoing degradation. This intracellular red fluorescence may also be from the endosomal/lysosomal compartments as this is one of the possible routes of SOCS3 protein recycling and turnover.

Interestingly, the cell line expressing the Ft construct fused to the N-terminus of SOCS3 did not demonstrate a consistent cellular SOCS3 localisation (figure 4.8, SOCS3-N-Ft, column 5). Specifically, during imaging, it was evident that there was a group of cells indicating a clear strong nuclear localisation of mCherry/sfGFP and a different pool of cells exhibiting a low intensity of mCherry/sfGFP fluorescence in the nucleus.

To gain a better insight into the cellular distribution of SOCS3-Ft fusion constructs, we quantified mCherry/sfGFP intensity ratios for each cell line. Results are depicted in figure 4.9. According to the figure 4.9, it is evident that the median sfGFP fluorescent intensities are higher than the median mCherry fluorescent intensities. The most profound difference in the median sfGFP and mCherry intensities was observed in the nuclear subcellular localisation in SOCS3-C-Ft cell line.

The mean fluorescent intensity ratios (FIRs) for each cell line were compared by two-way ANOVA. For two-way ANOVA, we first used the nuclear, perinuclear and peripheral cellular regions to compare the mCherry/sfGFP intensity ratios as the intracellular and whole compartments include other subcellular locations. Table 4.3 showed that there was a significant difference in the mean Ft ratios between the cell compartments ($p < 0.0001$). This variance is explained by the low Ft ratio values measured from the nuclear and perinuclear region of SOCS3-C-Ft cell lines compared to the mean Ft ratios measured from the other cellular compartments (figure 4.9). This

can be also seen in the nuclear region of SOCS3-N-Ft cell line but to a lesser extent. Furthermore, there was a meaningful variation of the mean Ft ratios across the cellular compartments measured in the different the cell lines ($p = 0.0001$). This effect is clearly seen in the nuclear mean FIR difference between Control-Ft and SOCS3-C-Ft (table 4.4, $p < 0.0001$ as calculated by Tukey's Honestly Significant Difference test)

Importantly, both cell lines demonstrated that newly synthesized proteins were mostly located in the nucleus whereas the older pool of SOCS3 protein was found in the intracellular region of the cell (Appendices, figure 6.2, significance values were <0.05 as calculated by Tukey's HSD test).

Additionally, these results demonstrate that the fusion of Ft to SOCS3 offered us insight into the dynamical subcellular trafficking of SOCS3 in vivo, can identify changes in SOCS3 abundance at a subcellular level and can be distinguished from that generated by the Ft alone.

Table 4. 3: The mean ratios in each cell line were compared by two-way analysis of variance (ANOVA).

ANOVA Table	Type III Sum of Squares	Degree of Freedom	Mean Square	F	Significance
Between cell compartments	2.081	2	1.041	35.05	$p < 0.0001$
Between Cell lines	0.09462	2	0.04731	1.593	$p = 0.2054$
Cell compartments by cell lines (interaction)	0.7064	4	0.1766	5.947	$p = 0.0001$
Error	7.305	246	0.02970		
computed using $\alpha = 0.05$, adj. p < 0.05					

Table 4. 4: Tukey's HSD (honestly significant difference) test to compare the means of Nuclear, Perinuclear and Peripheral cell compartment within each cell line.

	Mean Difference, Significance		
	Nucleus	**Perinuclear**	**Periphery**
Control-Ft vs. SOCS3-C-Ft	0.2005, $p < 0.0001$	-0.0008744, $p = 0.9998$	-0.1031, $p = 0.0705$
Control-Ft vs. SOCS3-N-Ft	0.1006, $p = 0.0769$	0.02634, $p = 0.8360$	0.01366, $p = 0.9529$
SOCS3-C-Ft vs. SOCS3-N-Ft	-0.09986, $p = 0.0689$	0.02721, $p = 0.8168$	0.1168, $p = 0.0265$
Tukey's multiple comparisons test computed using $\alpha = 0.05$ and adj. $p < 0.05$.			

4.3.4 Effects of proteasome inhibition on SOCS3 protein age and localization

Many studies reporting on SOCS3 transcription and translation in response to cytokine stimuli [408]. However, there is little information regarding post-translation dynamics of SOCS3 turnover in individual cells. SOCS3 is thought to undergo rapid ubiquitination and subsequent processing and degradation by the proteasome [101]. We therefore chose to explore the dynamics of SOCS3 protein turnover by examining SOCS3 in subcellular compartments upon proteasomal inhibition. MG132 is a well-documented proteasomal inhibitor that selectively inhibits the proteolytic activity of the 26S proteasome [409]. We inhibited proteasomal degradation by treating cells with 10 µM MG132 (Tocris Bioscience, C# 1748) for 6 hours and imaged on a Zeiss LSM 880 confocal microscope. Images obtained from the confocal microscope are depicted in figure 4.10.

We observed a profound increase on the average age in the Ft and SOCS3-Ft fusion protein pool upon MG132 treatment. All three cell lines showed an increased accumulation of red fluorescence inside the cells (figure 4.10, column 5). Control-Ft demonstrated a uniformly increase of sfGFP and mCherry intensity so that the ratiometric plots were dominated by older protein. Cells expressing SOCS3-Ft fusion constructs showed that the green fluorescence from sfGFP was profoundly localized at the perinuclear region of the cell, resembling a perinuclear necklace structure. Similar, overlapping perinuclear expression of mCherry was also found in the perinuclear region (figure 4.10, SOCS3-C-Ft, SOCS3-N-Ft, column 5). This punctuate pattern of fluorescence emitted from both sfGFP and mCherry seems to be in the endoplasmic reticulum (ER)/Golgi apparatus, the usual route of misfolded and short-lived proteins to the proteasome for degradation [410].
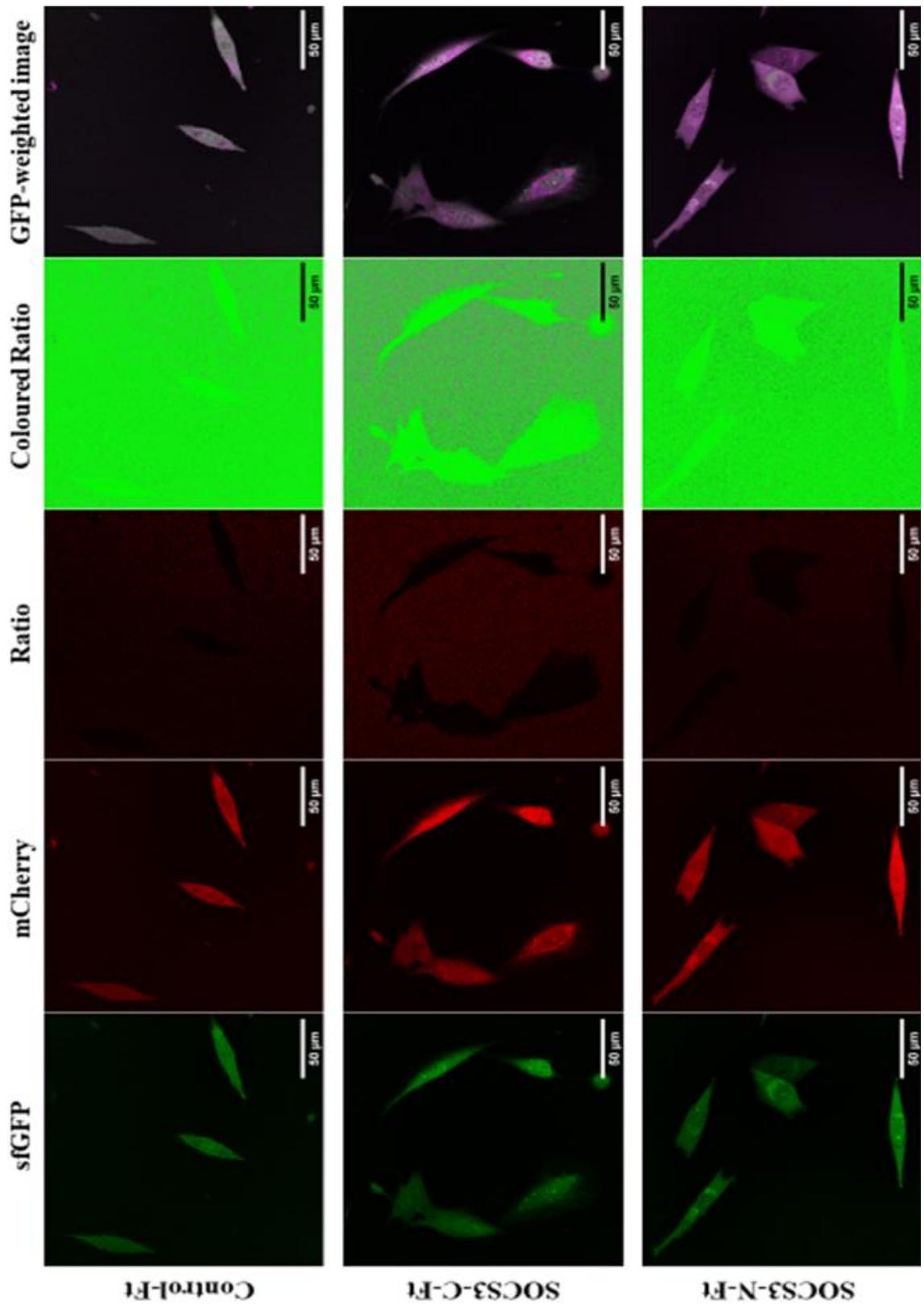
Figure 4.10: **Effects of proteasome inhibition on SOCS3 protein age and localization.**

Cells were treated with 10 μM MG132 for 6 hours before confocal microscopy. Control-Ft exhibited a uniformly increase of sfGFP and mCherry fluorescence signal throughout the cell. Treatment of the proteasome inhibitor MG132 in cells expressing SOCS3-Ft fusion constructs demonstrated an increase in the abundance and average age of SOCS3 protein.

Scale bar 50 μm.

Next, we performed a ratiometric analysis of mCherry/sfGFP intensity ratios of MG132 treated cell lines expressing Ft construct alone or tagged with SOCS3 protein. Results are depicted in figure 4.11. Interestingly, it is evident that the median ratios have dramatically increased compared to the results shown in figure 4.9. According to the results presented in figure 4.11, the Ft levels from the Control-Ft cell line were higher than those quantified from SOCS3-C-Ft and SOCS3-N-Ft cell lines (between cell lines, $p < 0.0001$, ANOVA table 4.5). Despite the increased accumulation of both fluorophores, the mean ratios (mCherry/sfGFP) quantified from both SOCS3-C-Ft and SOCS3-N-Ft cells indicated that they still retained the distinct subcellular pattern of SOCS3 abundance as seen in untreated cells (between cell compartments, $p < 0.0001$, ANOVA table 4.5). The most profound difference in the Ft levels was found in the nucleus and perinucleus of SOCS3-C-Ft and SOCS3-N-Ft cells, where the median ratios of those two subcellular locations are lower compared to those obtained from Control-Ft. Newly synthesized SOCS3 proteins are concentrated in the nucleus of the cell, while the older SOCS3 protein pool tends to be located at the intracellular region of the cell. Although the ANOVA analysis showed that there is a significant variation across cell lines and across subcellular localizations, the interaction between the two did not reach a statistical significance ($p = 0.8880$, cell compartments by cell line, ANOVA table 4.5).

Furthermore, the median sfGFP fluorescent intensities are lower than the median mCherry fluorescent intensities, demonstrating the increase on the average protein pool age. The most profound difference in the median sfGFP and mCherry intensities was found in the nuclear subcellular localisation in SOCS3-C-Ft cell line.
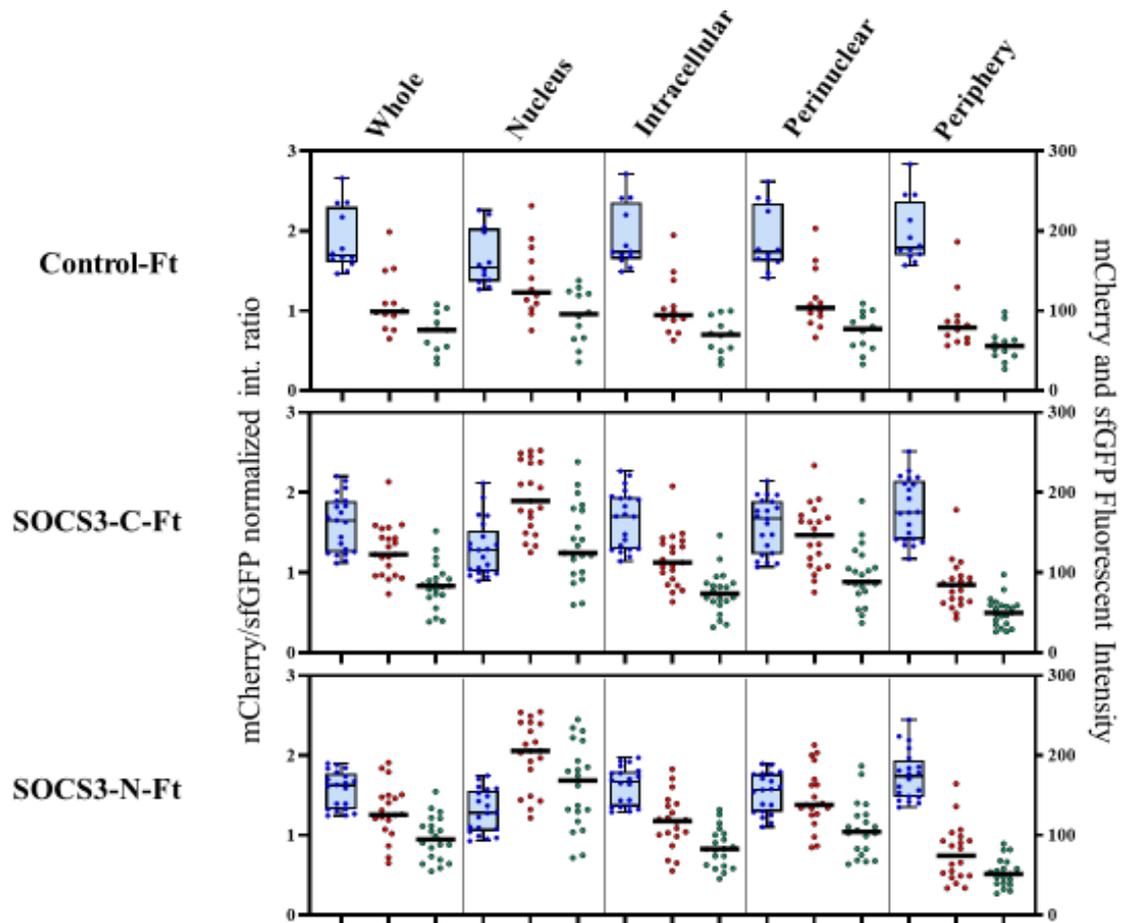
Figure 4.11: Ratiometric analysis of MG132 treated cell lines expressing Ft fusion constructs. For each Ft fusion construct, the fluorescent intensity ratio of individual cells was calculated and normalized to the median intensity ratio of a reference solution. Centrelines indicate the medians, box limits mark the 25th and 75th percentiles, and whiskers extend to the minimum and maximum values. Blue stars represent the individual cell ratios. The individual mCherry and sfGFP levels are displayed in red and green dots. Control-Ft (N=12), SOCS3-C-Ft (N=21), SOCS3-N-Ft (N=20).

Table 4. 5: **The mean FIR for each MG132 treated group was compared by two-way ANOVA.**

ANOVA Table	Type III Sum of Squares	Degree of Freedom	Mean Square	F	Significance
Between cell compartments	4.308	2	2.154	18.70	$p < 0.0001$
Between Cell lines	2.572	2	1.286	11.16	$p < 0.0001$
Cell compartments by cell lines (interaction)	0.1308	4	0.03271	0.2840	$p = 0.8880$
Error	17.28	150	0.1152		
computed using $\alpha = 0.05$, adj. $p < 0.05$					

Collectively, these results indicate that degradation of the Ft-tagged SOCS3 protein as well as the Ft alone is processed through the proteasome.

4.3.5 Autophagy regulates SOCS3 protein turn over and trafficking

A recent study showed that cannabinoid induced autophagy decreased SOCS3 protein expression and this was reversed by blocking early and late autophagy [65]. Therefore, we wanted to explore how SOCS3 whole population age is affected by blocking autophagy with Chloroquine Diphosphate. According to our previous finding, inhibition of proteasomal degradation did not establish a meaningful variation between groups. Thus, we wanted to test whether blocking autophagy influences SOCS3 turnover and how it differs from the turnover of the Ft alone.

Chloroquine Diphosphate is a lysosomotropic agent that inhibits endosomal acidification. Its activity depends upon its accumulation in endosomes and lysosomes, resulting in the inhibition of lysosomal enzymes that require an acidic pH, and subsequently blocking the fusion of endosome and lysosomes [411].

We blocked autophagy by treating cells with 50 μM Chloroquine Diphosphate (CQ) (Tocris Bioscience, C#4109) for 6 hours and imaged on a Zeiss LSM 880 confocal microscope. Images obtained from the confocal microscope are presented in figure 4.12.

Upon CQ treatment, SOCS3-C-Ft demonstrated a profound increase in the number of red fluorescent puncta per cell (figure 4.12, column 5), when compared to non-treated cells (figure 4.8, SOCS3-C-Ft, column 5). On the contrary, Control-Ft cells showed a relative diffuse red fluorescence in the cytosol. Interestingly, SOCS3-N-Ft cells did not display the same red punctuate fluorescent pattern to the same extent as observed with SOCS3-C-Ft cells (figure 4.12, SOCS3-N-Ft, column 5).

Taken together, these observations suggest that CQ treatment caused an increase to what it appears to be a pattern of vacuole formation in SOCS3-C-Ft cells. This pattern of vacuole formation was mainly concentrated at the perinuclear region, but it was also detected in the cytoplasm to a lesser degree. Moreover, this effect was also observed in cells expressing the N-Ft-SOCS3 construct, albeit to a lesser extent. This suggests that SOCS3 protein is probably processed through an endosome/lysosome-dependent trafficking route and it is only preserved when the N-terminus of SOCS3 is intact.
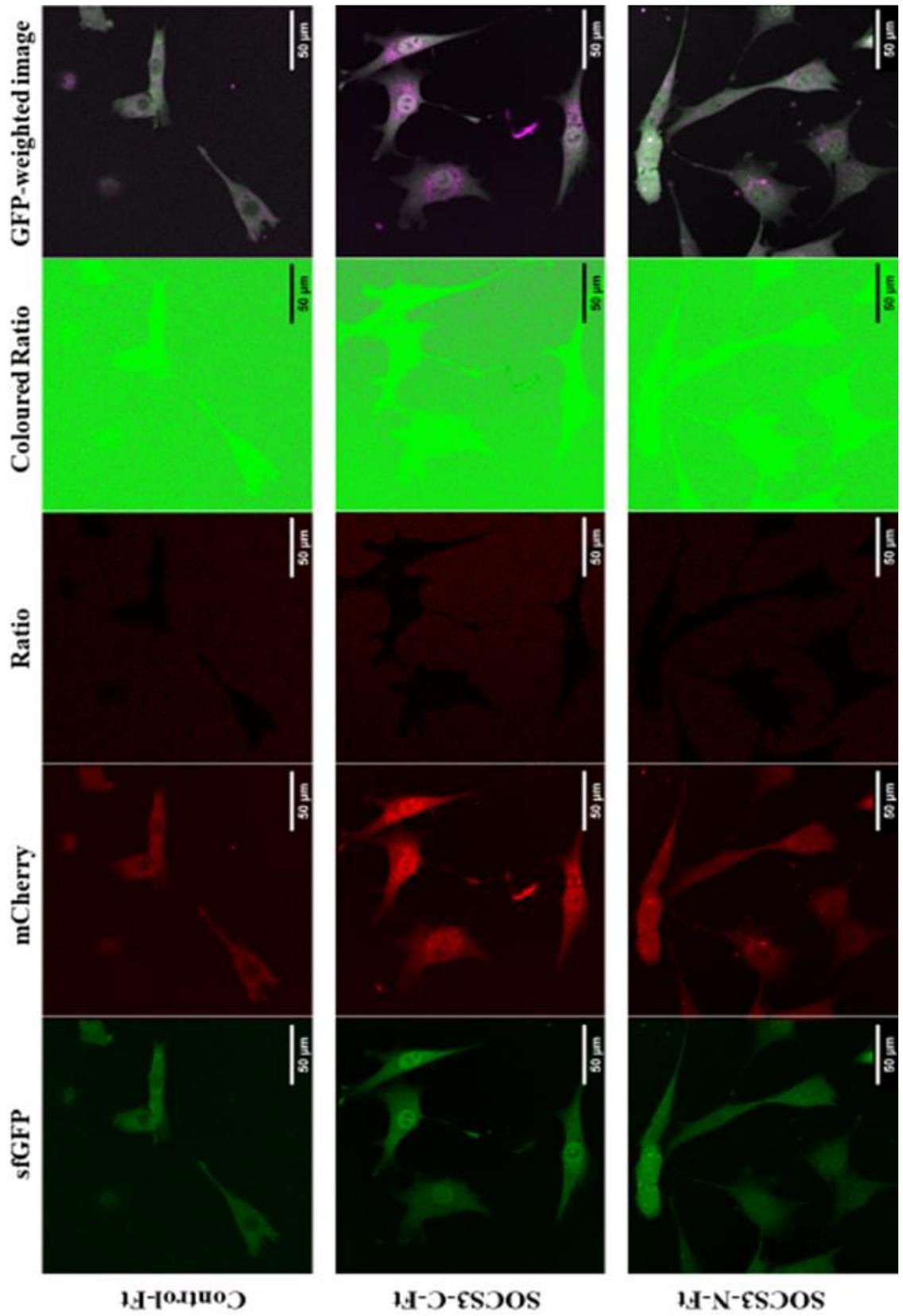
Figure 4.12: **Effects of Chloroquine Diphosphate treatment on SOCS3 protein age and localization.** Cells were treated with 50 μM Chloroquine Diphosphate (CQ) for 6 hours prior imaging. Control-Ft exhibited a slight increase of mCherry fluorescence signal throughout the cell. CQ treatment in cells expressing SOCS3-Ft tagged proteins also showed a settled increase in the amount of mCherry signal observed in the perinuclear region and a reduction in the average age of SOCS3 protein population, shown in the sfGFP weighted image. Scale bar 50 μm.

Next, we performed a ratiometric analysis of mCherry/sfGFP intensity ratios of CQ treated cell lines. Results from the ratiometric analysis are shown in figure 4.13. According to the box-and-whisker plots presented in figure 4.13, the majority of experimental groups demonstrate a slight increase in the average age of protein population compared to the cell lines that were not treated (figure 4.9). The mean ratios in each experimental group were compared by two-way analysis of variance (ANOVA). Results from the two-way ANOVA analysis are presented in table 4.6.

Furthermore, fluorescence intensity ratios quantified for SOCS3-C-Ft and SOCS3-N-Ft were not similar to those measured for the Control-Ft (between cell lines, $p < 0.0001$, ANOVA table 4.6). Both constructs expressing Ft-tagged SOCS3 appeared to show higher values in their ratios compared to Control-Ft.
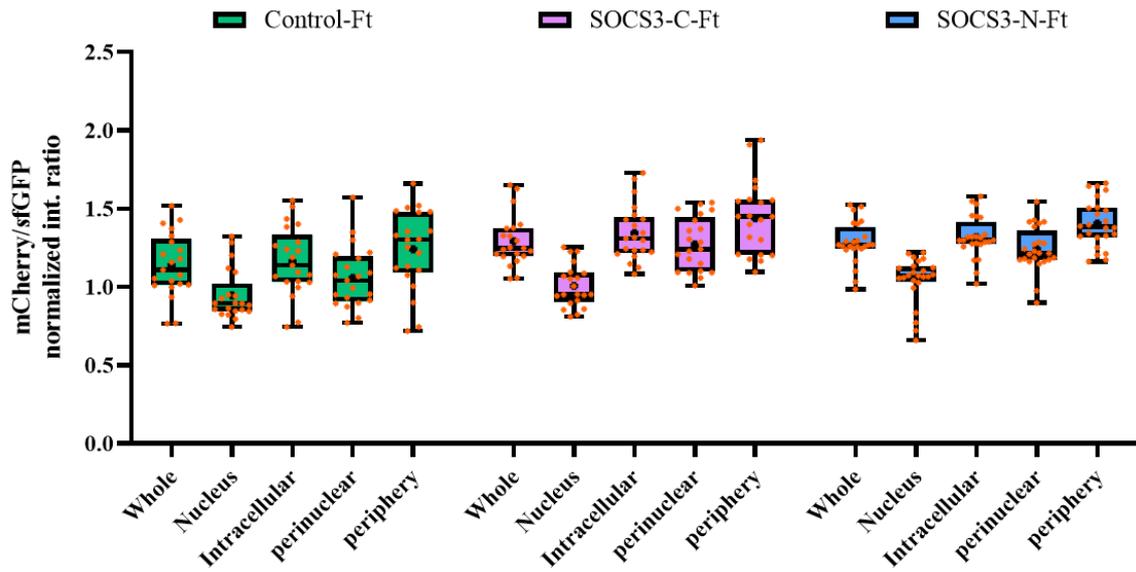
Figure 4.13: **Ratiometric analysis of Chloroquine Diphosphate treated cell lines expressing Ft fusion constructs.** Fluorescent intensity ratio of single cells was quantified and normalized to the median intensity ratio of a reference solution. Centrelines indicate the medians, means are indicated as '+', box limits mark the 25th and 75th percentiles, and whiskers extend to the minimum and maximum. Individual ratio values are displayed with orange stars. Control-Ft (N=21), SOCS3-C-Ft (N=21), SOCS3-N-Ft (N=24).

Table 4. 6: The mean FIR for each CQ treated cell line was analysed by two-way ANOVA.

ANOVA Table	Type III Sum of Squares	Degree of Freedom	Mean Square	F	Significance
Between cell compartments	4.329	2	2.164	66.53	$p < 0.0001$
Between Cell lines	1.006	2	0.5032	15.47	$p < 0.0001$
Cell compartments by cell lines (interaction)	0.1518	4	0.03795	1.167	$p = 0.3270$
Error	6.149	189	0.03253		
computed using $\alpha = 0.05$, adj. $p < 0.05$					

According to the figure 4.13it is evident that there is a significant variation across cellular compartments as the mean Ft ratios in the nucleus and perinucleus are significant lower compared to the rest cellular compartments ($p < 0.0001$, between cellular compartments, ANOVA table 4.6). Although the ANOVA analysis showed that there is a variation across cell lines and across cellular compartments, the interaction between the two did not reach a statistical significance ($p = 0.3270$, cell compartments by cell line, ANOVA table 4.6). As expected, these results suggest that cell line expressing Ft alone also shares the same endosome/lysosome-dependent trafficking route as observed with SOCS3.

4.3.6 Rapamycin induced autophagy increased the rate of SOCS3 turnover

Our previous results demonstrate that compromising the autophagosome-lysosome/endosome fusion process has a possible effect on SOCS3 turnover. Thus, we wanted to test whether inducing autophagy would also impact to the degradation rate of the tFT-fusion SOCS3 protein.

One of the important mediators of autophagy is the mechanistic target of rapamycin (mTOR kinase), which is a key of metabolic homeostasis that supresses autophagy in the abundance of growth factors and nutrients [412].

Rapamycin is a well-known inducer of autophagy and its activity depends on the inhibition of mTOR protein which mimics cellular starvation by preventing signal needed for cell growth and proliferation [413].

We induced autophagy by treating cells with 25nM and 50nM Rapamycin (Tocris Bioscience, C# 1292) for 18 hours and imaged on a Zeiss LSM 880 confocal microscope. Images corresponding to the 50 nM Rapamycin treatment are presented in figure 4.14.

According to figure 4.14, row 1, the cell line expressing the Ft construct alone (Control-Ft) showed a uniform distribution of the fluorescence signal detected from both sfGFP and mCherry in the cell, with nearly perfect co-localization of the two fluorescent proteins. Moreover, the GFP-weighted image of the Control-Ft (figure 4.14, column 5) also showed that the mCherry punctuate pattern concentrated at the perinuclear region of the cell as seem in non-treated cells, still persisted after the treatment with Rapamycin. As expected, the Ft alone did not establish any distinct localisation pattern in the cell.
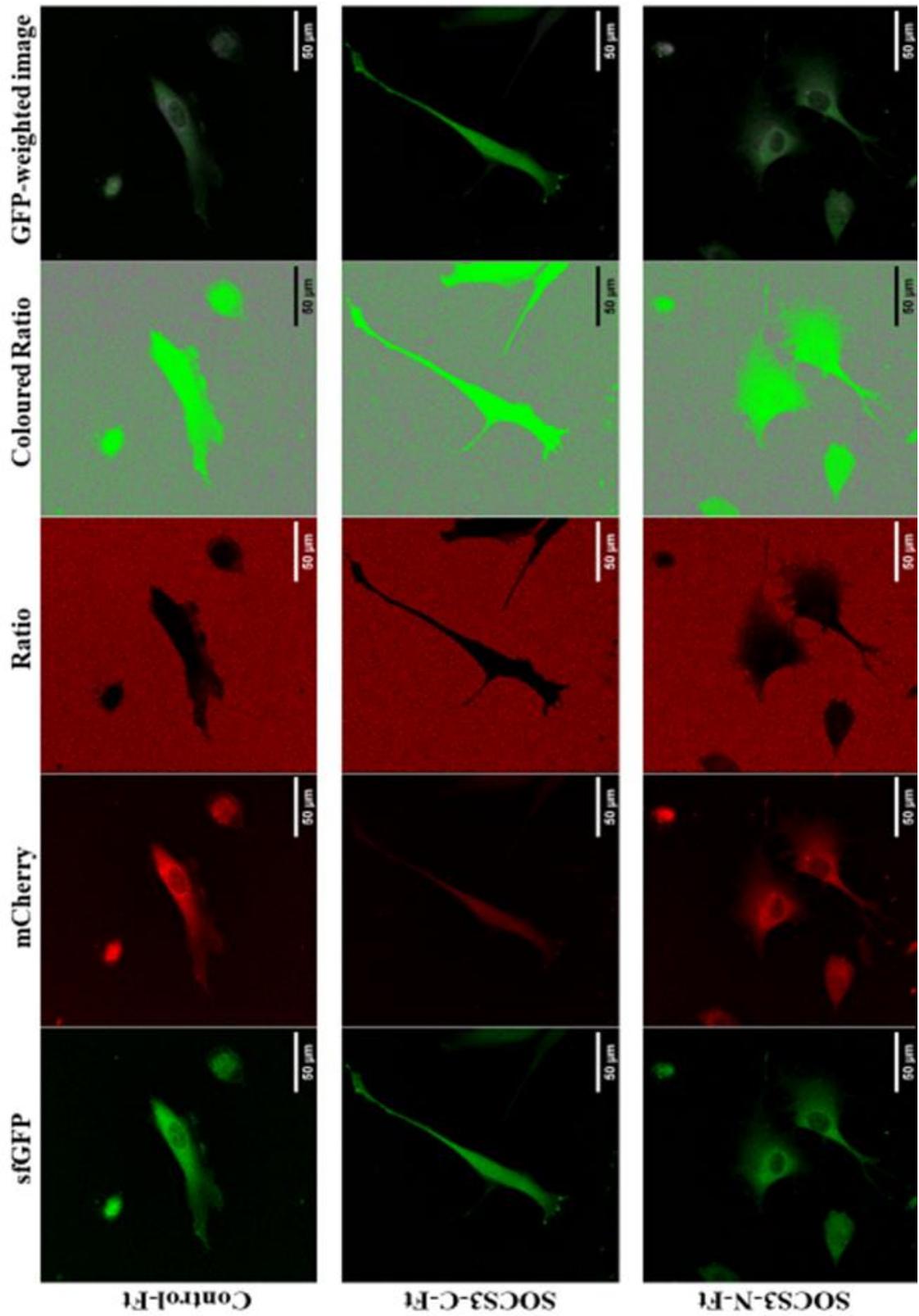
Figure 4.14: Effects of Rapamycin on SOCS3 turn over. Cells were treated with 50 nM Rapamycin for 18 hours prior imaging. Control-Ft exhibited a slight increase of mCherry fluorescence signal throughout the cell. SOCS3-N-Ft cells presented an even cytoplasmic distribution pattern of mCherry/sfGFP that was very similar to the tFT alone. Scale bar 50 μm.

On the other hand, SOCS3-C-Ft cells (figure 4.14, row 2) demonstrated a dramatic decrease in red fluorescence whereas the intensity of green fluorescence was similar to the sfGFP levels detected from Control-Ft cells (row 1). Finally, GFP-weighted image (column 5) of SOCS3-C-Ft suggests that there was an increase in the degradation rates of SOCS3 upon Rapamycin treatment.

Moreover, SOCS3-N-Ft cells (figure 3.14, row 3) presented similar levels of fluorescence emitted from sfGFP and mCherry to the levels detected from the cells expressing Ft construct alone.

Next, a ratiometric analysis was conducted to calculate the mCherry/sfGFP intensity ratios of rapamycin treated cell lines. Results from the ratiometric analysis are presented in box-and-whisker plots and shown in figure 4.15. The box plots in figure 4.15 summarizes the mCherry/sfGFP ratios quantified from each cell compartment in response to the 25 nM and 50 nM Rapamycin treatment.
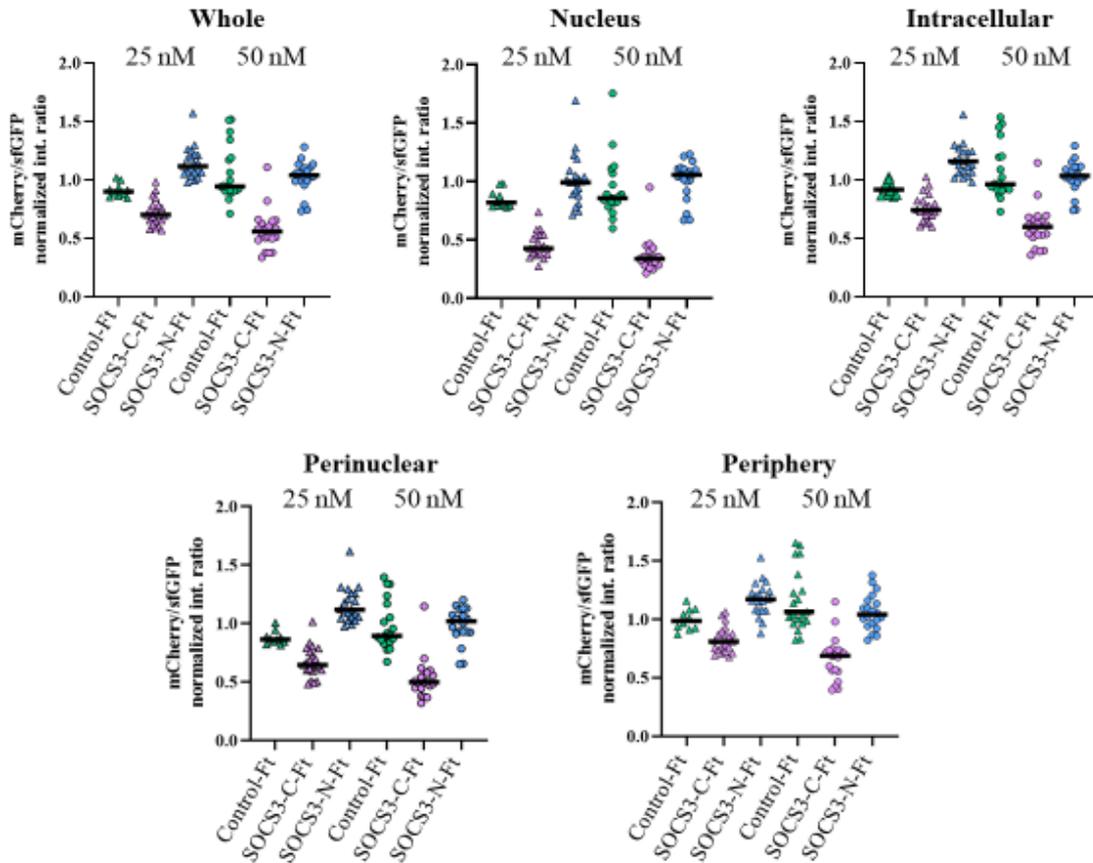
Figure 4.15: Ratiometric analysis of Rapamycin treated cell lines expressing Ft fusion constructs. Cells were treated with either 25 or 50 μM of Rapamycin for 18 hours prior to imaging. Fluorescent intensity ratio of single cells was quantified, normalized to the median intensity ratio of a reference solution and presented in dot plot format. Centrelines indicate the means. (Cells treated with 25 nM Rapamycin are indicated with '▲' whereas cells treated with 50 nM Rapamycin are shown with '●'). For Rapamycin 25 nM treatment: Control-Ft (N=10), SOCS3-C-Ft (N=21), SOCS3-N-Ft (N=21), Rapamycin 50 nM treatment: Control-Ft (N=20), SOCS3-C-Ft (N=20), SOCS3-N-Ft (N=20).

According to the figures 4.15, it is evident that the results of 25 nM and 50 nM Rapamycin treatment are qualitative similar. Thus, we only performed a two-way ANOVA analysis for the mean FIR obtained from the cell lines that were treated with 50 nM Rapamycin.

The fusion constructs expressed from Control-Ft and SOCS3-N-Ft did not appear to differ in their turnover in response to the Rapamycin treatment. In contrast, the fusion construct expressed from SOCS3-C-Ft cells exhibited the lowest mCherry/sfGFP ratio, demonstrating that this Ft-tagged protein was the most unstable upon Rapamycin treatment ($p < 0.0001$, between cell lines, ANOVA table 4.7). This decrease in Ft levels of SOCS3-C-Ft was found across all cellular compartments when compared to the other Ft constructs (($p < 0.0001$, computed by Tukey's multiple comparisons test, table 4.8). Interestingly, the mean of mCherry/sfGFP ratio decreased as the concentration of Rapamycin treatment increased from 25 nM to 50 nM, suggesting that SOCS3 turn over is increased in a Rapamycin dose dependent manner (figure 4.15). However, this was not observed with the mean FIR values measured from the Control-Ft cell line or SOCS3-N-Ft (Tukey's multiple comparisons test did not show significant difference between SOCS3-N-Ft and Control-Ft, table 4.8).

Table 4.7 also showed that there was a significant variation in the mean Ft ratios between the cell compartments ($p < 0.0001$). This variance is explained by the low Ft ratio values measured from the nuclear region of SOCS3-C-Ft cell lines compared to the mean Ft ratios measured from the other subcellular locations (figure 4.15). It is evident that there is a relationship between cell line and subcellular location ($p = 0.0398$, interaction, ANOVA table 4.7), which is contributed to the significant difference in the means obtained from the nuclear, perinuclear and peripheral regions across cell lines.

Table 4. 7: **The mean FIR for each Rapamycin (50 nM) treated cell line was analysed by two-way ANOVA.**

ANOVA Table	Type III Sum of Squares	Degree of Freedom	Mean Square	F	Significance
Between cell compartments	1.163	2	0.5813	15.45	$p < 0.0001$
Between Cell lines	9.991	2	4.996	132.7	$p < 0.0001$
Cell compartments by cell lines (interaction)	0.3868	4	0.09669	2.569	$p = 0.0398$
Error	6.435	171	0.03763		
computed using α = 0.05, adj. p < 0.05					

Table 4. 8: **Tukey's HSD test to compare the mean FIRs in each Rapamycin (50 nM) treated cell line.**

	Mean Difference, Significance		
	Nucleus	**Perinuclear**	**Periphery**
Control-Ft vs. SOCS3-C-Ft	0.5676, $p < 0.0001$	0.4475, $p < 0.0001$	0.4931, $p < 0.0001$
Control-Ft vs. SOCS3-N-Ft	-0.07708, $p = 0.4218$	-0.005790, $p = 0.9951$	0.1007, $p = 0.2308$
SOCS3-C-Ft vs. SOCS3-N-Ft	-0.6446, $p < 0.0001$	-0.4533, $p < 0.0001$	-0.3924, $p < 0.0001$
Tukey's multiple comparisons test computed using α = 0.05 and adj. $p < 0.05$.			

Taken together, cells expressing SOCS3-C-Ft construct exhibited lower mCherry/sfGFP fluorescence ratios in response to the Rapamycin treatment when compared to the cells expressing SOCS3-N-Ft as well as Ft construct alone. This indicates that the Rapamycin treatment destabilized the SOCS3-C-Ft levels of abundance. These results suggest that incorporating the Ft at the N-terminus of SOCS3 probably may induce protein misfolding and consequently masking important N-terminal amino acids that probably affect the destabilization of the protein [414].

4.3.7 Effects of MAPK kinase inhibition on SOCS3 turnover and localization

Mitogen-activated protein kinase (MAPK) signalling pathway is composed of a cascade of serine/threonine kinases that induced by a large array of extracellular stimuli that play crucial roles in regulating several cellular processes, from the cell surface to the nucleus [415]. Previous research have demonstrated that the MAPK and mTOR pathways have several cross-connections and are two signalling pathways regulating autophagy. Furthermore, the extracellular-signal-regulated kinase (ERK) pathway is one of the key signalling cassettes of MAPK signalling pathway [416]. Multiple studies have reported that autophagy is linked with the ERK1/2 signalling pathway [417]. Phospho-ERK1/2 was found to be activated by rapamycin [416]. Thus, we wanted to explore the effects of inhibiting MAPK pathway by treating cells with U0126. U0126 is a highly selective inhibitor of MAP kinase kinases, MEK1 and MEK2. Its action depends on the inhibition of the kinase activity of MEK1/2 and consequently ERK activation [418].

We inhibited MAPK signalling pathway by treating cells with 10 μM U0126 (Tocris Bioscience, C#1144) for 16 hours and imaged on a Zeiss LSM 880 confocal microscope. Images obtained from the confocal microscope are depicted in figure 4.16.

According to the confocal images in figure 16 (row 1), the levels of mCherry/sfGFP fluorescence of Control-Ft did not appear to have altered in response to MAPK inhibitor. The localisation and levels of fluorescence was distributed uniformly throughout the cell. On the other hand, the intensity of mCherry/sfGFP fluorescence from SOCS3-C-Ft and SOCS3-N-Ft is decreased compared to Control-Ft.

A ratiometric analysis was generated to calculate the mCherry/sfGFP intensity ratios of U0126 treated cell lines. Results from the ratiometric analysis are presented in figure 4.17.
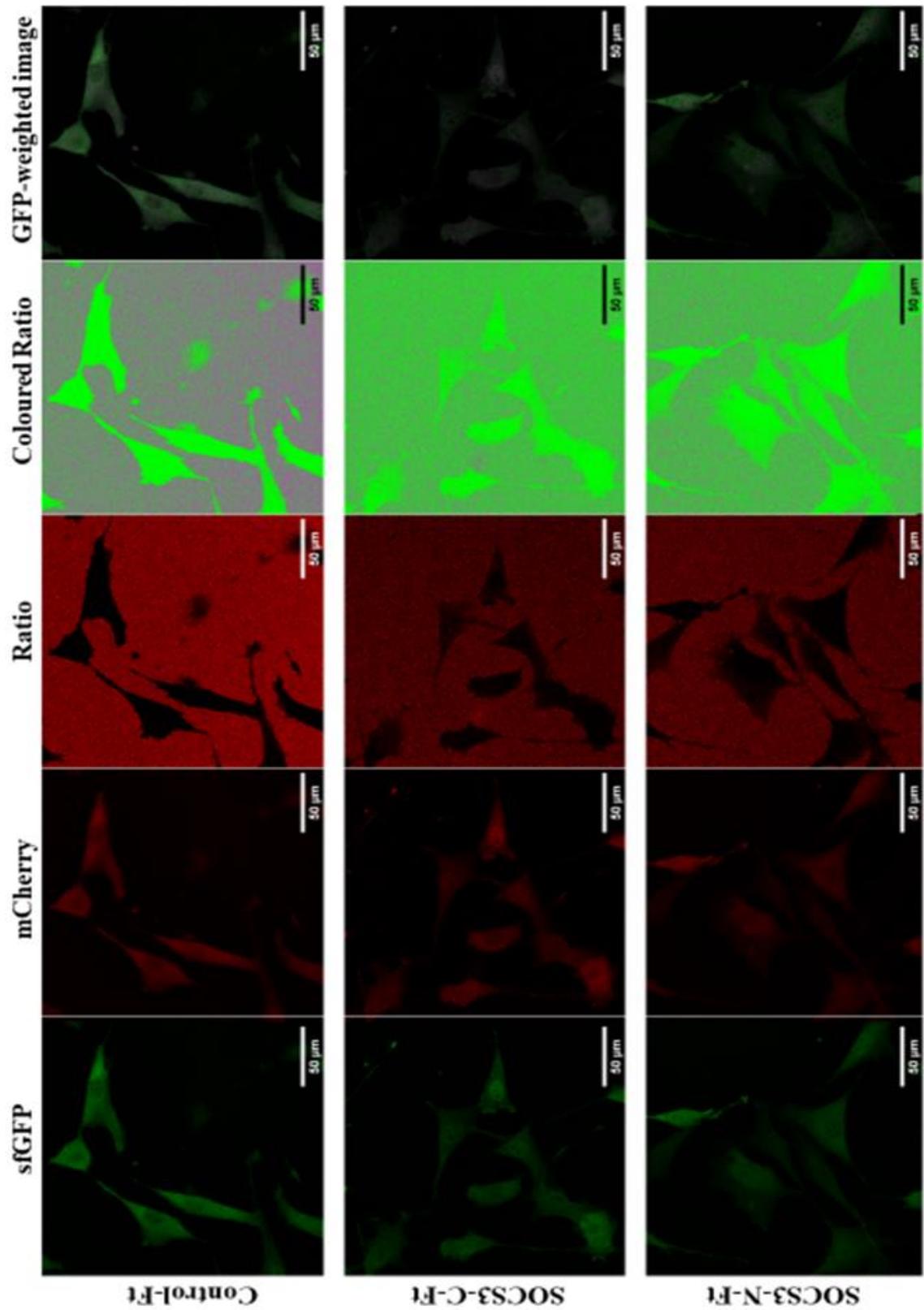
Figure 4.16: Effects of turn over and localisation upon inhibition of the kinase activity of MEK1/2. Cells were treated with 10 μM U0126 for 16 hours prior imaging. Control-Ft exhibited a presented an even cytoplasmic distribution pattern. SOCS3-C-Ft and SOCS3-N-Ft cells showed a decreased expression of mCherry/sfGFP compared to Control-Ft. Scale bar 50 μm.

The ratiometric analysis in figure 4.17 showed that Control-Ft demonstrated similar mean mCherry/sfGFP intensity ratios in each cell compartment. The same mean distribution was also seen in SOCS3-N-Ft. However, the mCherry/sfGFP fluorescence appears to be slightly increased in SOCS3-C-Ft in relation to the other two cell lines ($p < 0.0001$, between cell lines, ANOVA table 4.9). Interestingly, the previously observed SOCS3 localisation pattern, where tFT tagged SOCS3 protein is youngest in the nucleus with an older population in the cytoplasm, was also retained in response to U0126 inhibitor lines ($p = 0.0005$, between cell compartments, ANOVA table 4.9). Both SOCS3-C-Ft and SOCS3-N-Ft cell lines displayed lower values in their median Ft values when compared to the rest cellular compartments.
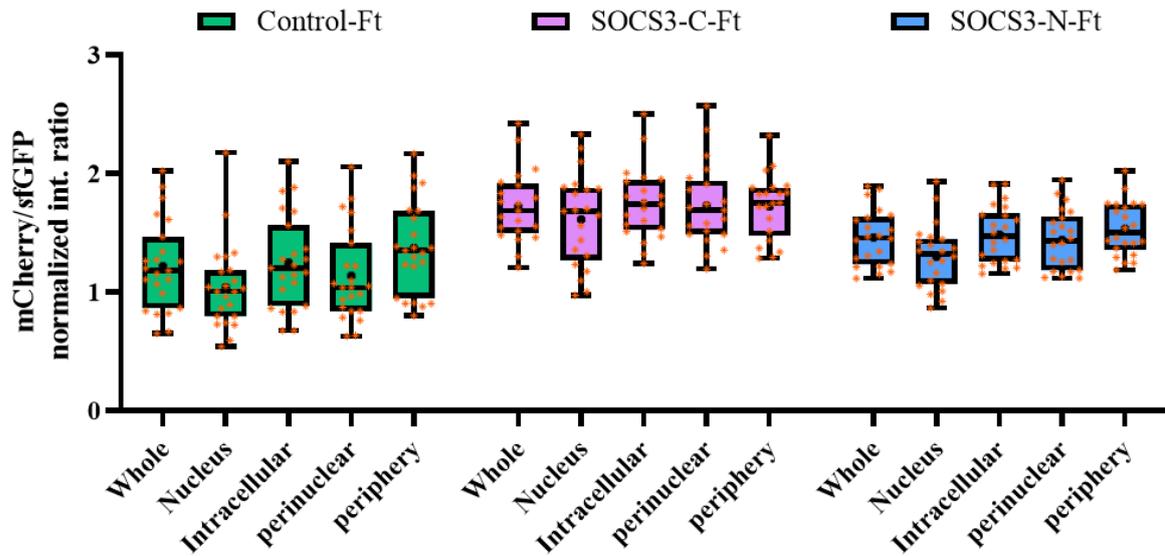
Figure 4.17: **Ratiometric analysis of U0126 treated cell lines expressing Ft fusion constructs.** Fluorescent intensity ratio of single cells was quantified and normalized to the median intensity ratio of a reference solution. Centrelines indicate the medians, means are indicated as '+', box limits mark the 25th and 75th percentiles, and whiskers extend to the minimum and maximum values. Individual ratio values are displayed with orange stars. Control-Ft (N=23), SOCS3-C-Ft (N=21), SOCS3-N-Ft (N=22).

Table 4. 9: Level of variance of the mean FIR for each cell line treated with U0126 MAPK inhibitor was computed by two-way ANOVA.

ANOVA Table	Type III Sum of Squares	Degree of Freedom	Mean Square	F	Significance
Between cell compartments	1.693	2	0.8464	7.896	$p = 0.0005$
Between Cell lines	8.381	2	4.190	38.99	$p < 0.0001$
Cell compartments by cell lines (interaction)	0.4047	4	0.1012	0.9414	$p = 0.4412$
Error	20.31	189	0.1075		
computed using $\alpha = 0.05$, adj. $p < 0.05$					

4.3.8 Descriptive analysis of SOCS3 promoter activity in response to TLR agonists

Chronic inflammation is associated with elevated levels of pro-inflammatory cytokines in the circulation. Consequently, regulation of pro-inflammatory signalling is crucial to prevent runaway inflammation which is associated with constitutive JAK/STAT activation and increased cancer risk [408]. Consistent with its role as a negative mediator of inflammatory signalling via JAK/STAT pathway inhibition, SOCS3 expression has been found increased at sites of chronic and acute inflammation [213]. In addition to SOCS3 regulatory activity on the JAK-STAT signalling pathway, SOCS3 protein has also been found to negatively regulate microbe-sensing Toll-like receptor (TLR)-induced inflammatory pathways, signalling from TNF receptor (TRAF) 2,3 and 6 as well as the downstream target NFκB [119, 408].

Furthermore, SOCS proteins are often present at low or undetectable levels in cells and their expression is induced transcriptionally upon stimulation of various cytokines such as interleukin-2 (IL-2), IL-3, IL-4, IL-6, leukaemia inhibitory factor (LIF) and growth hormones [419]. The transcription regulation of SOCS3 expression has been investigated in a vast array of cell systems in response to different stimuli using luciferase reporter assay [213].

Since tandem fluorescent timers have been previously used to report on gene expression dynamics [383], we asked whether the fluorescent timer mCherry/sfGFP has the required sensitivity to report on SOCS3 promoter activity in time-dependent manner.

To demonstrate the utility of the fluorescent timer as a tool for studying SOCS3 promoter activity, we used the Flp-in system (Thermofisher) to generate stable isogenic cell lines with a single integration of the mCherry/sfGFP fluorescent timer under the

control of a partial human SOCS3 promoter (1.7 kb fragment of the 5'-3' flanking sequence, figure 4.3). Furthermore, since the stable fold of FP has a half-life of several hours, we engineered the FT timer to also include the PEST sequence of SOCS3 at the C-terminus. Prior to electroporation, we sequenced the plasmid to verify the correct sequence of SOCS3 promoter and the correct insertion of the SOCS3 pest sequence at the C-terminus of the FT (see Appendix).

To investigate the promoter activity of SOCS3 within cells, live cell imaging using confocal laser scanning microscopy was performed. During confocal imaging, it was apparent that the partial human SOCS3 promoter was expressing the Ft at very low levels compared to the constructs where the expression of Ft was driven by CAG promoter (figure 4.18). Therefore, the laser power of the confocal microscope had to be adjusted in both sfGFP and mCherry channels for this set of experiments. Furthermore, the images obtained from SOCS3-promoter Ft cells (figure 4.20, row 2), demonstrated that the majority of cells presented a mixture of sfGFP and mCherry fluorescence, suggesting that the SOCS3 promoter is constitutively active but in low levels.
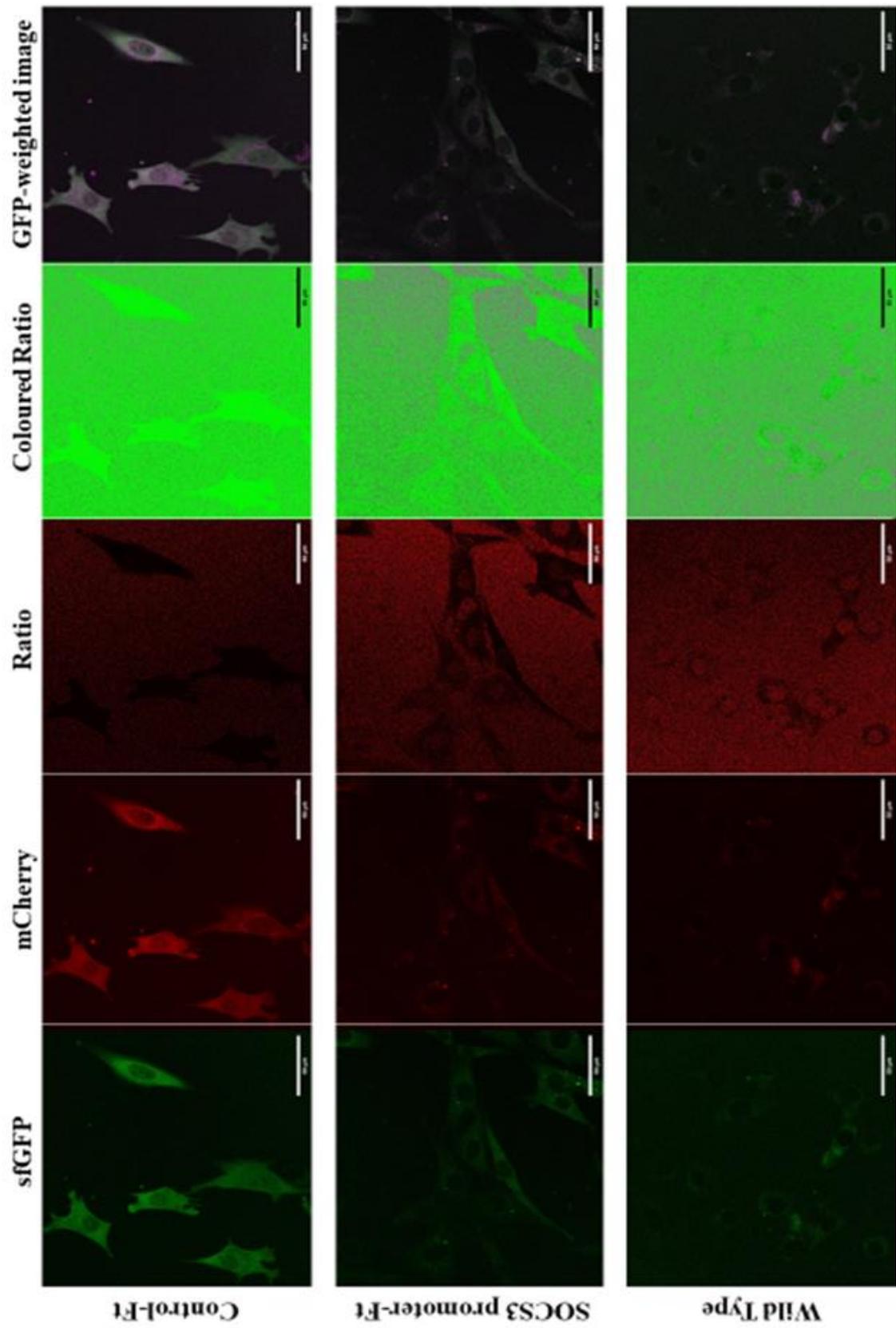
Figure 4.18: Live imaging of mouse embryonic fibroblast Flp-In™ 3T3 cell lines expressing tandem Ft under the control of SOCS3 promoter. Images obtained with the same acquisition settings show the difference in fluorescence levels between CAG, SOCS3 promoter and wild type driving the expression of tFT. Scale bar 50 μm

To identify the dynamic activity of SOCS3 promoter in real time, we treated SOCS3 promoter-Ft cells with 1 μg/ml Gram-negative bacterial outer membrane component lipopolysaccharide (LPS) (Sigma Aldrich, C# L4391-1MG) a well-known TLR4 agonist that induces SOCS3 expression. Following treatment, images were taken every 20 minutes up to 400 minutes using Zeiss LSM 880 confocal microscope. Ratiometric analysis was performed by analysing every frame individually as described in methods. Results from the ratiometric analysis are presented in figure 4.19.
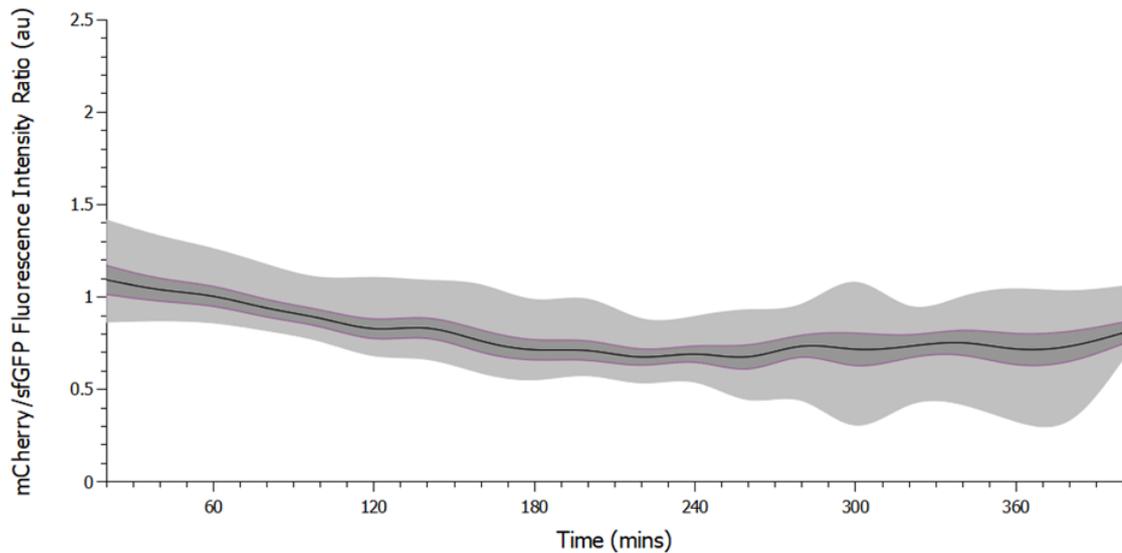
Figure 4.19: **Time course of SOCS3 promoter activity in response to LPS treatment.**
Cells were treated with 1 µg/ml of LPS prior to imaging. Images were obtained in 20 minute-intervals using confocal microscopy. The fluorescent intensity ratio of individual cells was calculated and normalized to the median intensity ratio of a reference solution. The solid black line indicate the mean FIR per frame (N=10). The upper and lower limits of the pale grey area show the maximum and minimum cell ratios. The dark grey area represents the data that falls within one standard error of the mean (movie 1).

According to figure 4.19, the mean FIR did not exhibit any visible fluctuation in the activity of SOCS3 promoter. Specifically, the data from the time course demonstrated that the mean FIR was calculated to be 1.014±0.077 au at t=0, it gradually decreased to 0.666±0.032 au at t=220 minutes, then slowly increased again and reached to 0.776±0.041 au at the final time frame (t=400 minutes). These results demonstrate that when the promoter was induced with LPS, the signal of fluorescence was predominant green during the first few hours, causing the mean FIR values to drop.

To further verify that the Ft can behave as a proxy for promoter activity, we treated the SOCS3 promoter-Ft cells with a different TLR agonist; flagellin (Sigma Aldrich, C# SRP8029-10UG) at a concentration of 200ng/ml. Following the addition of the TLR5 agonist, images were obtained at 20-minute intervals up to 400 minutes using Zeiss LSM 880 confocal microscope. Results from the ratiometric analysis are presented in figure 4.20.
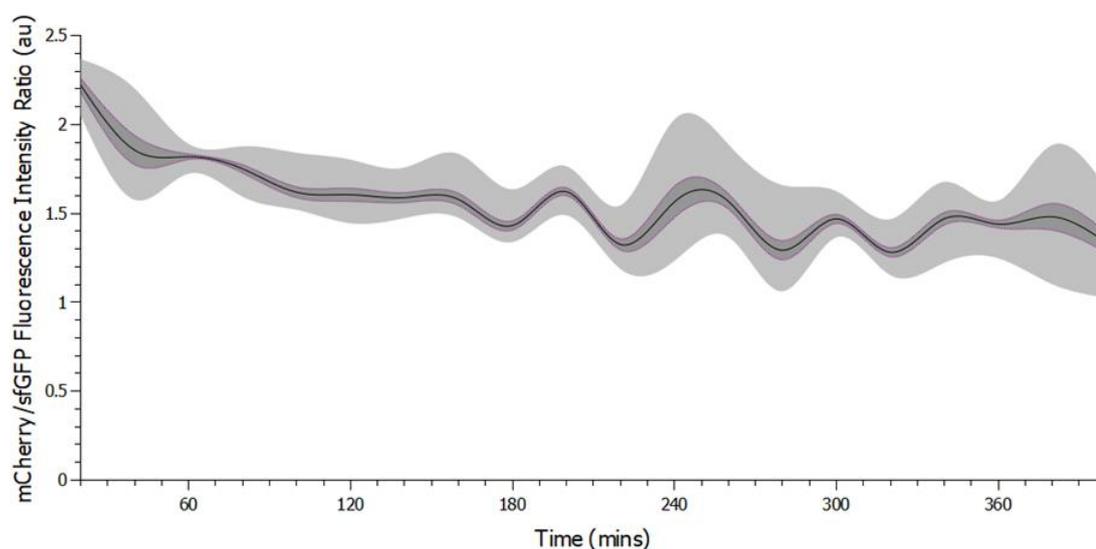
Figure 4.20: **Time course of SOCS3 promoter activity in response to Flagellin treatment.** Cells were treated with 200 ng/ml of Flagellin prior to imaging. Images were obtained in 20 minute-intervals using confocal microscopy. The FIR of individual cells was quantified and normalized to the median intensity ratio of a reference solution and plotted as a function of time. The solid black line indicate the mean FIR value based on the ratios of all the fully visible and non-overlapping cells (N>8). The upper and lower limits of the pale grey area indicate the maximum and minimum FIR. The dark grey area represents the data that falls within one standard error of the mean (Movie 2).

In response to flagellin treatment, the mean FIR demonstrated a small degree of dynamic behaviour in the activity of SOCS3 promoter. Particularly, the results from the time series showed that the mean FIR was quantified to be 2.221±0.039 au at t=0,

gradually decreasing to 1.4320±0.026 au at t=180 mins, and then an oscillatory pattern begun with first peak at t=200, second peak at t=260 and final peak at t=300 mins. From t=300 mins time point, the mean FIR values continued to gradually decrease with final mean value calculated to be 1.3365±0.056 au. Regardless of those small FIR fluctuations, the overall response of SOCS3 promoter activity in response to flagellin was steadily increasing during the time series.

4.3.9 Descriptive analysis of SOCS3 promoter activity in response to LIF

Leukaemia inhibitory factor (LIF) belongs to the gp130 cytokine family and signals through the receptor complex of gp130 and the LIF receptor to induce the JAK/STAT signalling pathway [106]. Previous have investigated the effects of LIF in the induction of SOCS3 promoter activity using luciferase as a reporter [420].

To investigate whether SOCS3 promoter activity exerts any dynamic behavior and follow its activity in real time, we treated SOCS3 promoter-Ft cells with 100 ng/ml LIF (PEPROTECH, C# 300-05). Immediately after the addition of LIF, images were acquired with an acquisition time of 10 minute-intervals and monitored its activity up to 500 minutes using a Zeiss LSM 880 confocal microscope. Ratiometric analysis was computed by analysing single time points separately. Results from the ratiometric analysis are presented in figure 4.21.

When cells were stimulated with LIF, they showed highly dynamic oscillatory behaviour, with the mCherry/sfGFP Fluorescence Intensity Ratio fluctuating multiple times before a significant drop at 500 minutes (figure 4.21), after which the activity of the cells reduced considerably.
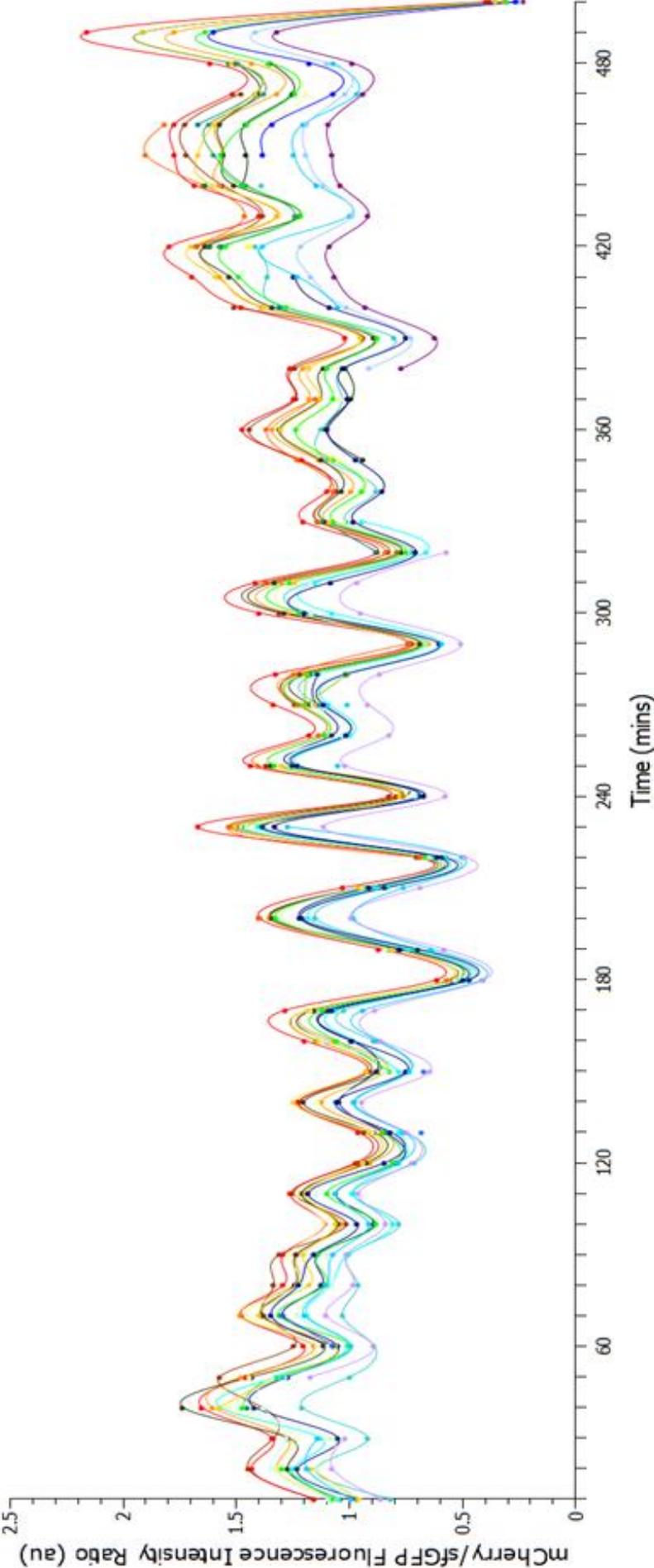
Figure 4.21: The mCherry/sfGFP Fluorescence Intensity Ratio in response to LIF plotted as a function of time. Data points are shown at 10-minute intervals which matches the recording frequency (average N=13 per frame). Interpolation was achieved with a natural cubic spline fit. Each colour represents a single cell which has been tracked through multiple time frames. Temporal gaps in data typically represent either cells disappearing out of frame or an overlap in cells, making measurements unreliable and therefore not considered (Movie 3).

All cells showed complementary oscillatory behaviour, with apparent differences between frames being highly consistent as can be seen by the continuation of the rainbow colouration across Figure A. The oscillations occur with a period of ~30 minutes, however the amplitude of the oscillations changes slightly; averaging ~0.2 au in the first three hours (phase 1), then increasing to 0.4 au for the second two hours (phase 2), then decreasing again for the last three hours (phase 3). The amplitude of the latter phase is harder to judge due to a decrease in regularity, although the amplitude seems to be at a value between that of the first and second phase. The final phase is not only more irregular in terms of the amplitude and period of the oscillations but also in the range of ratios observed. During the final phase the average range of ratios for any one given frame was 0.48 au, compared to 0.37 and 0.32 au for the first and second phase. The final observation based on Figure 4.21, is that the oscillations seem to be superimposed on a second trend which also shows a temporal change in ratios with time.
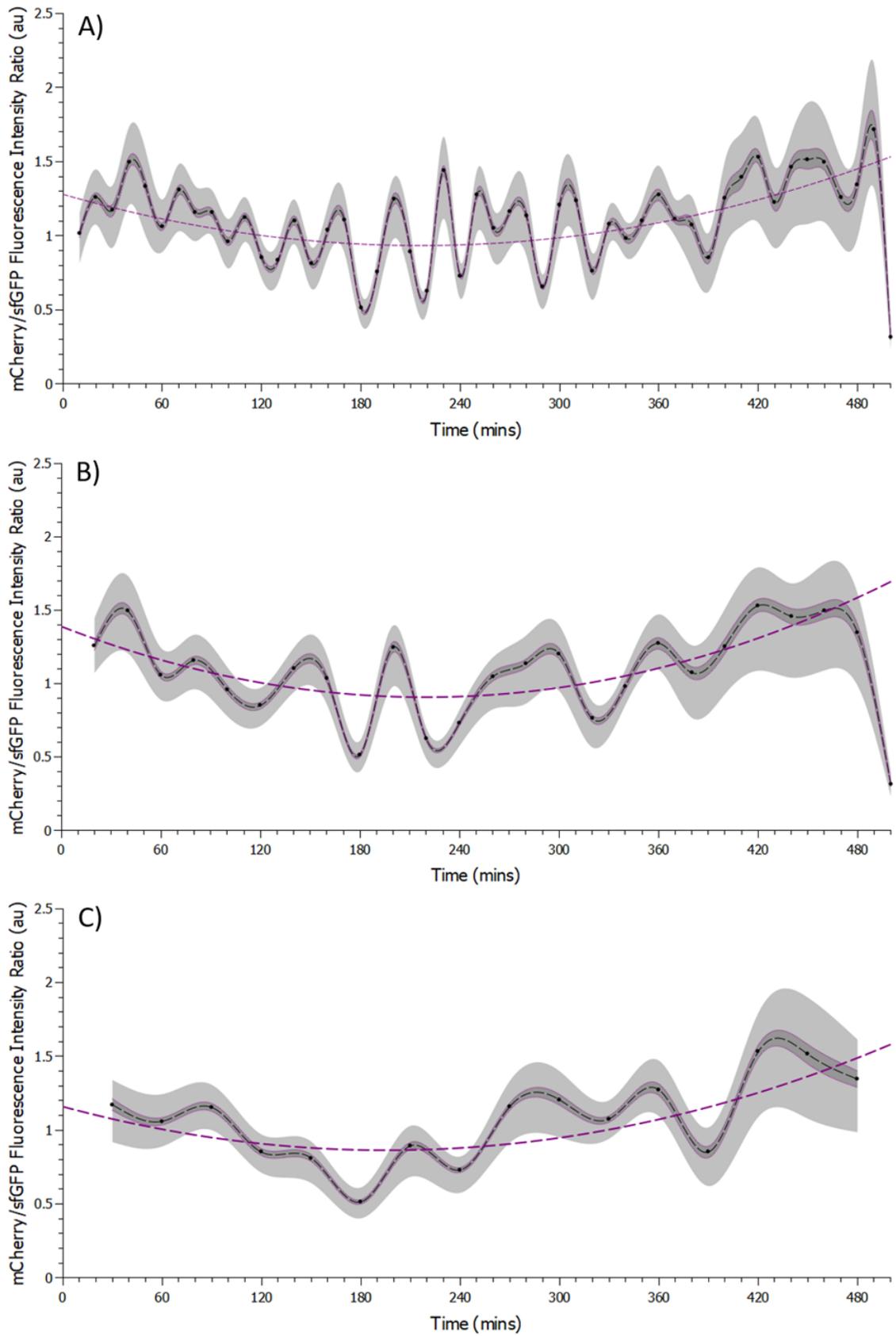
Figure 4.22: The mCherry/sfGFP Fluorescence Intensity Ratio in response to LIF plotted as a function of time with different temporal frequencies; A), B) and C) show data points every 10, 20 and 30 minutes respectively, although the original sampling frequency was at 10-minute intervals. The solid black line represents a mean value based on the ratios of all the fully visible and non-overlapping cells per frame (average N=13 per frame); the means have been interpolated using a natural cubic spline fit. The upper and lower limits of the pale grey area represent the maximum and minimum cell ratios, respectively, and thus the grey area (pale and dark combined) represents the range of values observed per frame. The dark grey area, outlined by a dotted purple line, represents the data that falls within one standard error of the mean. The dashed purple line represents a polynomial trendline through all the data minus the ratio recorded for the 500-minute frame as this was seen as anomalous to the trend.

Overall, the average ratios observed for the middle section of the plotted times are slightly lower than that for the start and end of the experiment which are comparable to each other, giving the overall impression of a smile.

It should be noted however, that any characterisation of oscillations is highly dependent on the frequency with which measurements are made (Figure 4.22). For example, when measurements are displayed with 20-minute intervals instead of 10-minute intervals the oscillations appear to occur every 55 minutes (Figure 4.22.B) rather than every 30 (Figure 4.22.A), whilst a 30-minute gap between measurements suggests a period of 65 minutes (figure 4.22.C). Nevertheless, there are also a number of similarities between these graphs (figure 4.22) representing different sampling frequencies. For example, despite the period being different, the amplitude of the oscillations in all three graphs (Figure 4.22) is comparable and show the same trend of having higher amplitudes during the middle phase of the experiments. Furthermore, for all three sampling

frequencies, the regularity of the oscillations decreases and the range of ratios increases in the latter hours of the experiment, there is the same consistency in how the ratios of individual cells change with time and a similar background trend of a smile (on average higher ratios at the start and end of the experiment, relative to the middle) exists for all sampling frequencies (figure 4.22), suggesting these are real differences and not experimental artefacts. Only the oscillation frequency, seems to be influenced on the sampling frequency, at least for the time periods considered.

However, these findings fail to demonstrate whether the observed oscillation represents a pulsatile activation of SOCS3 promoter or the degradation rate of the fluorescent reporter. Therefore, we analysed the temporal changes of total sfGFP and mCherry fluorescent signals of 4 cell trajectories over time. The analysis is presented in the figure 4.23.

It is clear from figure 4.23 that the behaviour of sfGFP and mCherry fluorescent signals are comparatively similar to the overall ratio behaviour, in that both signals are oscillating with a similar frequency to the ratio and in general, where there are peaks and troughs in the ratio timeline, there are also peak and troughs respectively, in both the sfGFP and mCherry signals. Also, in similarity to the overall ratio, both the sfGFP and mCherry signals show greater amplitudes in the middle of the time series (phase 2). This suggests that the ratio is oscillating due to changes in both the sfGFP and mCherry signals. That being said, it could be argued that in the first, third and fourth graph in figure 4.23, between ~50 and ~100 minutes (within phase 1), there is a period where the mCherry signal is relatively static but the sfGFP is distinctively oscillating. In contrast from ~290 minute onwards, the mCherry signal shows more dynamic behaviour than sfGFP.

Nikoletta Kalenderoglou - November 2019
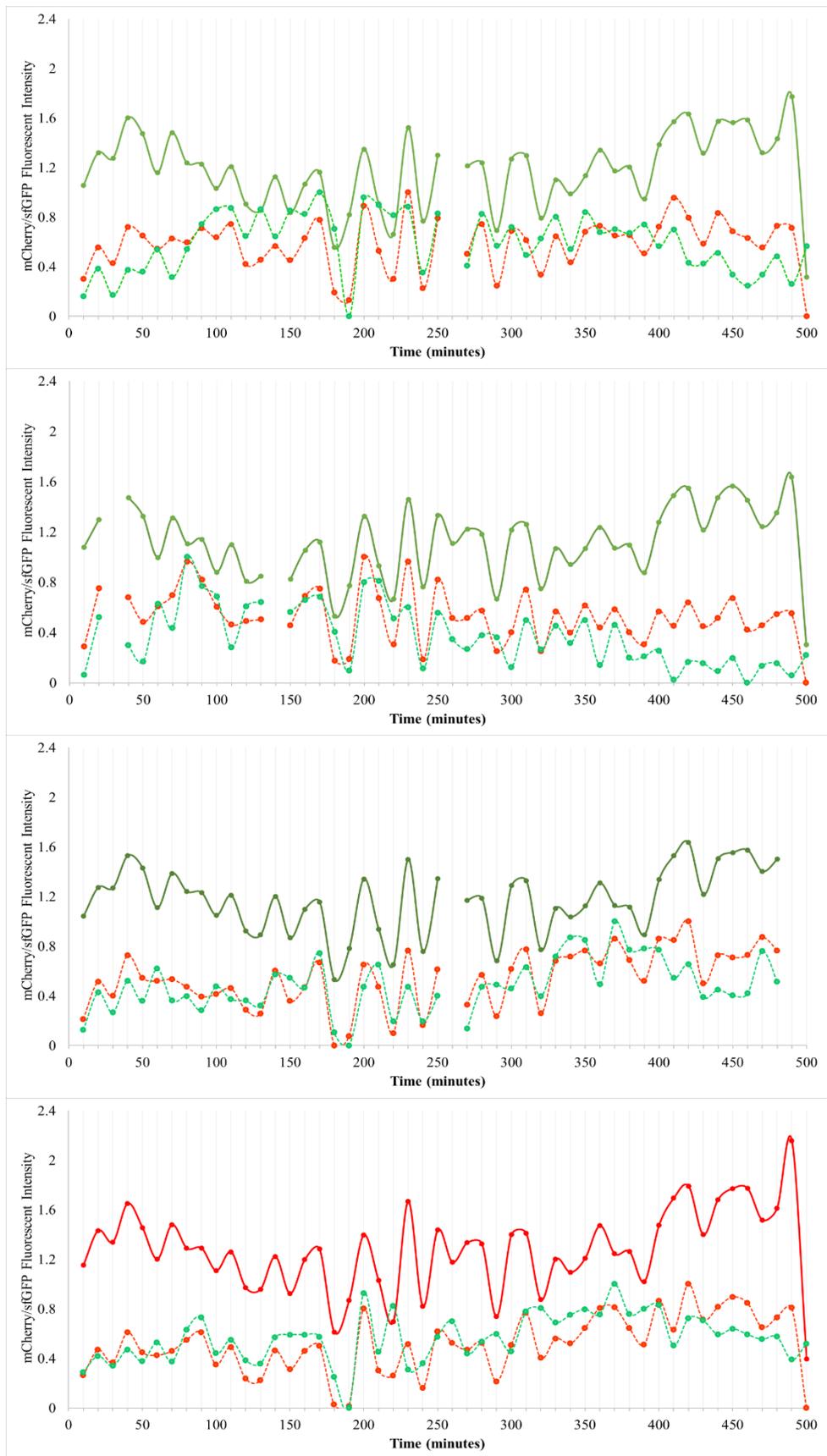
Figure 4.23: Individual mCherry and sfGFP Fluorescent signals for 4 cells plotted as a function of time. Data points are shown at 10-minute intervals which matches the recording frequency (average N=13 per frame). The solid coloured line represents the mCherry/sfGFP ratio values for a single cell trajectory. Each colour represents a single cell which has been analysed from the rainbow graph in figure 4.21 and tracked through multiple time frames. Temporal gaps in data typically represent either cells disappearing out of frame or an overlap in cells, making measurements unreliable and therefore not considered. The green and red dashed lines and markers represent the normalised sfGFP and mCherry cell trajectories from which the ratios were calculated.

Consequently, it could be argued that the oscillations in the ratio during phase 1 are primarily influenced by sfGFP, whereas in phase 3, mCherry becomes the more influential factor. During phase 2 both sfGFP and mCherry are highly dynamic.

Overall, the average intensity of mCherry remains relatively constant with time, however sfGFP tends to show an overall increase in intensity during the first ~180 minutes (phase 1) and an overall decline in intensity during the final ~130 minutes i.e. from ~370 minutes onwards (phase 3). This arched trend is most apparent in the first graph in figure 4.23 and is the most likely explanation for the overall 'smile' appearance of the ratio. i.e. the ratio has higher values during phase 1 and 3, relative to phase 2 because the sfGFP is lower at these periods whereas mCherry remains relatively constant; it is therefore sfGFP and not mCherry that is controlling the overall trend of a declining then increasing ratio.

It should also be noted that during phase 3 the sfGFP and mCherry intensities become more distinctly different from each other, owing largely to the decline of sfGFP signal.

It could be argued that this difference in signals causes the apparent slight deviance from regularity in phase 3 because when there is a large difference between two variables that make up a ratio only small changes in these variables (sfGFP and mCherry intensity) could result in large changes in the ratio.

Taken together, these data suggest that tFT exert the required sensitivity and specificity to follow the expression of weak promoters in real time and provide readouts of their activity in response to stimuli.

4.4 Discussion

Cells respond to a vast array of external signals utilizing only a limited number of signalling pathways by plasma membrane receptors, including G protein-couple receptors (GPCRs) and receptor tyrosine kinases (RTKs). These signalling processes do not only transduce the signal, but process, encode, integrate further signals and even generate the appropriate suppressors to regulate the response caused by the external cue. In the past decade, it has become evident that distinct spatio-temporal dynamic profiles of the same repertoire of proteins regulating these signalling pathways are associated in different gene activation patterns and distinct physiological responses [421].

Research reports on signalling dynamics *in vitro* and *in vivo* have been limited by a scarcity of fluorescent reporters. However, tandem fluorescent protein timers have the potential to revolutionise in this area of research by unravelling the dynamics of these highly complicated signalling systems. The usefulness of tandem fluorescent protein timers has been exploited in a series of research reports. It was initially developed to investigate the protein half-life in *S. cerevisiae* [385] as well as further explore chemokine signalling and cadherin stability along the zebrafish posterior lateral line primordium [393].

TFt protein timers has been characterised as a powerful tool for investigating changes in protein population age and thus provide single snapshot readouts for signalling events in relation to alterations in protein stability and location in the cell. Moreover, tFTs can provide qualitative measurements of intracellular protein dynamics and gain valuable quantitative information on the kinetics of signalling pathways under investigation [387].

Thus, we utilized fluorescent timers to investigate various factors that could influence SOCS3 oscillatory behaviour such as SOCS3 localisation, degradation dynamics and promoter activity.

4.4.1 Generation and verification of fluorescent timer fusion constructs

DNA sequencing was used as a method to confirm and verify the correct DNA sequencing for every plasmid used to generate the different Flp-In 3t3 cell lines. I attempted to verify the expression of SOCS3-Ft insertions in SOCS3-C-Ft and SOCS3-N-Ft cell line by western blotting (appendix), but the western blots only showed bands for the endogenous SOCS3 protein and did not present bands from the SOCS3-Ft insertions. This could be due to an issue with the antibody used to detect SOCS3 protein.

Correct integration of the fluorescent timer can be tested by Junction PCR using primers that bind outside the inserted construct and within the fluorescent timer. Therefore, validation for the correct insertion and expression should be carried out for these constructs.

For the purpose of this chapter successful integration of each fluorescent construct used was assumed.

4.4.2 Characterization of SOCS3-fluorescent timer cell lines

Tandem fluorescent timers have been characterised as a powerful tool to investigate how the of specific protein population changes in different cellular compartments [385]. Thus, our first aim was to study how SOCS3 localises in the cell and how its distribution in different cellular compartments is associated with the age of the protein. Since there is limited published information about the subcellular localisation of SOCS3 and how

it is mediated, we fused the fluorescent timer to both C and N termini of SOCS3 protein. We utilized the tFT itself (not fused with SOCS3) as an internal control because this fusion is highly stable within the cell [385]. This provided us an opportunity to experimentally determine whether the kinetics of SOCS3 turnover is similar between those cell subtypes.

According to the images obtained from confocal microscopy, both SOCS3-C-Ft and SOCS3-N-Ft showed similar SOCS3 localisation patterns. Specifically, the youngest pool of SOCS3 protein was found in the nucleus while the older population in the perinuclear region. Furthermore, we found that the fluorescence pattern of mCherry was quite different from that of sfGFP, as sfGFP was dispersed uniformly throughout the cytoplasm, whereas mCherry punctuate signal were relatively concentrated around the perinuclear region. This dotted pattern reminiscent of either endosome or lysosomes and appeared to be located in the endoplasmic reticulum/Golgi apparatus. If these vesicles represent endosomes, this imply the involvement of endosomal transport from the Golgi apparatus, if but if the vesicles are lysosomes, this implies that SOCS3 is targeted for lysosomal degradation.

One possible explanation of the increased accumulation of vesicular clumps may be caused by the improper folding of the protein in the endoplasmic reticulum, which results in the routing of misfolded protein to the proteasome for degradation.

To identify the nature of these vesicular compartments in which SOCS3-Ft was found, future work should apply staining methods to identify localisation of late endosome and lysosomes.

This artefact is frequently seen in studies using mCherry-derived FP when expressed for an extended period of time. This artefact could potential be explained by the fact

that mCherry displays a β-barrel structure making it resistant to lysosomal proteolysis and a low pKa. This difference of pKa values from both fluorescent proteins would also generate different protease sensitivity for both fluorophores. Consequently, when Ft-tagged proteins encounter an acidic environment the sfGFP fluorescence signal is diminished, whereas mCherry would remain fluorescent [422]. To improve imaging conditions when studying protein turnover and dynamics using Fluorescent timers would be to consider an alternative fluorescent protein such as Tag-red fluorescent protein (Tag-RFP), which has been shown to be less stable in acidic environments and consequently, present similar levels of proteasome sensitivity to that seen with sfGFP [393].

4.4.3 Subcellular Localization of SOCS3

Protein dynamics are closely associated with protein function and localisation. A better understanding to the precise localization of proteins are crucial for establishing a clear picture of their functions [244, 253, 423].

Immunohistochemical studies have been increasingly popular to investigate the distribution and localization of specific proteins at a cellular or even a subcellular scale [424]. However, the use of antibodies in immunohistochemical studies create a lot of uncertainties in the results that most often created by the specificity and sensitivity of antibodies. Moreover, approaches employing antibodies do not allow live cell imaging since they require fixation and permeabilization to penetrate the cells [253]. Since quantitative analysis of protein dynamics requires analysis across spatial and temporal dimensions, we used tandem fluorescent timers to analyse SOCS3 dynamic behaviour either in normal culture conditions or in response to treatments.

Images from the confocal microscope showed that there was a clear nuclear localisation of SOCS3-Ft protein. This was also seen in SOCS3-N-Ft cells but to a lesser extent. The inconsistency of nuclear translocation in SOCS3-N-Ft cells can probably be attributed to the fold of the tFT interfering the N-terminus of SOCS3. SOCS3-N-Ft cell line requires further characterization to investigate why SOCS3 nuclear translocation is not as evident as seen with SOCS3-C-Ft cell line.

Interestingly, a nuclear localization signal has been identified and characterized for SOCS1 but not for SOCS3 [425].

An apparent controversy is the intracellular localization of SOCS3 protein as the vast majority of research has been limited in a cytoplasmic localization of SOCS3, where it can exert its inhibitory roles by interacting with the cytoplasmic residues of transmembrane receptors and inhibit substrate phosphorylation via its SH2 and KIR domains, or by binding to several signaling intermediate kinases through the SOCS box and targeting them for ubiquitination and subsequent proteasomal degradation [110, 426].

A particular strong piece of evidence in favour of the nuclear localization came from the group of Lee and colleagues, who showed cytoplasmic and nuclear localisation [426]. Lee's group [426] observed nuclear localization of SOCS3 when they overexpressed SOCS3 in head and neck squamous cell carcinoma (HNSCC) cells.

There has been also one previous immunohistochemical report investigating SOCS3 expression in biopsy specimens from patients with ulcerative colitis (UC) and UC-related colorectal cancer (CRC) progression [116]. Although Li and her colleagues only reported a cytoplasmic localisation of SOCS3 protein [116], close examination of their immunohistochemical images presents a predominantly cytoplasmic SOCS3 expression

in epithelial cells, but also shows that SOCS3 protein is localised in the cytoplasm and nucleus in leucocytes with morphological appearances suggesting that they might be macrophages, DCs or lymphocytes in the large intestinal lamina propria.

Another piece of evidence produceed by White et al. [110] demonstrated by means of transfection, immunohistochemical studies and immunofluorescent staining of LPS-stimulated peripheral blood mononuclear cells, that SOCS3 protein may be present in both the nucleus and cytoplasm, in keeping with Lee's group [426] study using socs3 transfectants. It is still debatable what is the molecular mechanism underlying the tumor-specific accumulation of SOCS3 in the nucleus. Lee and colleagues [426] demonstrated that the elevated nuclear SOCS3 localisation was generated from the increased accumulation of SOCS3 protein in the cytoplasm. This increased accumulation of SOCS3 in the cytoplasm was achieved by treating cells with the proteasomal inhibitor MG132. When we treated the cells with MG132, we also observed an increased level of SOCS3 protein in the nucleus in agreement with Lee's [426] results. They also found that this nuclear localisation of SOCS3 caused the reduction of STAT3 protein. These results indicate that elevated SOCS3 cytoplasmic abundance is mediated through nuclear translocation [426]. Future work should address whether there is a nuclear localisation signal in SOCS3 that causes its nuclear translocation. Furthermore, it would be of interest to investigate whether there are specific proteins (i.e. importins) that are associated with SOCS3 protein translocation.

4.4.4 SOCS3 protein half-life

Notably, Khmelinskii and his colleagues proved that the fluorescence signal ratio did not relate to the rate of protein synthesis, as this was established by single-cell investigation of yeast strains with heterogenous protein expression levels driven by

plasmid copy number variation or by using promoters of varying strengths. Consequently, these results demonstrate that the degradation signals that are found in the protein of interest provide a direct readout of the protein degradation kinetics in steady state. [385].

We utilised fluorescent timers to explore the dynamics of SOCS3 protein turnover by examining SOCS3 in subcellular compartments upon different treatments. It has been previously shown that SOCS3 protein is mediated partly through an autophagy-based mechanism [65].

The process of autophagy can be pharmacologically activated or supressed through a vast array of chemical agents. We used rapamycin to activate the process of autophagy. Ratiometric analysis showed that SOCS3-C-Ft was the most unstable upon Rapamycin treatment. Control-Ft and SOCS3-N-Ft did not appear to be affected by the activation of autophagy. These results suggest that the fusion of the Ft at the N-terminus of SOCS3 may have probably caused SOCS3 protein to misfold and consequently mask important N-terminal amino acids, impacting the destabilization of the protein. This observation, agrees with finding from Sasaki et al. [102], who demonstrated that the deletion of a conserved ubiquitination residue (Lys-6) at the N terminus of SOCS3 showed a much longer half-life than the wild type SOCS3 in Ba/F3 hematopoietic cells.

The pharmacological inhibition of autophagy by either blocking lysosomal acidification or inhibiting MAPK pathway did not demonstrate significant variation in the mean FIRs between the three cell lines. Nevertheless, it should be noted that all of these pharmacological compounds lack the specificity to target solely the autophagy machinery, as they also known to disrupt other cellular processes. For instance, the inhibition of lysosomal degradation also influences mitosis and endocytosis [412, 427].

Furthermore, blocking the proteasomal degradation by MG132 can also increase autophagy by stabilizing Atg protein levels [428].

Nevertheless, we are unable to demonstrate categorically that SOCS3 is regulated through an autophagy-based mechanism. Therefore, future work should investigate the role of autophagy in the process of SOCS3 cycling by implementing genetic approaches to block autophagy.

The half-life of SOCS3 has been investigated in various cell lines from different research reports and it has been shown that there is a large discrepancy reported over its half-life that can vary from under 5 minutes to over 8 hours [102, 103]. However, the majority of research studies have established a short biological half-life for SOCS3 protein, ranging from 36 minutes to 120 minutes [102, 207, 208]. This discrepancy has been mainly attributed to the specific cell line's preferred method of SOCS3 clearance from the cell. Therefore, it would be of interest to measure SOCS3 half-life by treating SOCS3-C-Ft and SOCS3-N-Ft cells with cycloheximide (a well-known inhibitor of protein synthesis) [209] and perform a time series. This provided us an opportunity to experimentally determine the half-life of SOCS3 protein tagged with mCherry-sfGFP timer either at the C or the N terminus and thereby ask whether the kinetics of SOCS3 turnover is similar between those cell subtypes.

Furthermore, these research studies have also demonstrated that different deletions of different motifs in the SOCS3 coding sequence, such as the Pest motif or the SOCS box have resulted in an increase of its half-life [103, 208]. Future work should also consider implementing genetic approaches to alter the SOCS3 Pest motif or the SOCS3 box and investigate how this would affect the age population of SOCS3 in different cellular compartments.

Finally, it is important that a reporter gene has the potential to alter protein half-life by inducing protein aggregation, accumulating dominant degradation signals, masking degradation signals, impairing appropriate protein folding or impeding correct protein localization and/or assembly into complexes [385]. Nevertheless, we are unable to demonstrate categorically that the presence of the tFT reporter does not change other properties of the SOCS3 protein, including subcellular localisation.

4.4.5 Promoter Activity

Cytokine signalling in the gp130 family demonstrate functional redundancy as they signal through shared receptor components. Thus, the process of SOCS3 expression induction should be carefully timed and regulated in order to confer the appropriate feedback inhibition [429].

In this study we demonstrate the advantages of FTs for exploring the dynamics of promoter activity using as example the previously described partial fragment of SOCS3 promoter.

Due to the long half-life of stable fluorescent proteins (FPs) such GFP, that can extend up to 56 hours, the fluctuating behaviour of gene transcription cannot be effectively analysed using conventional FP expression as a reporter for promoter activity [430]. For this reason, real-time gene expression studies using luciferase reporter activity has been the most preferred method for investigating the temporal dynamics of gene transcription due to its short half-life [398].

Interestingly, a research study conducted by Babon et al., showed that the half-life of SOCS3 increased dramatically when they transfected 293T cells with a SOCS3 construct that lacked the PEST motif [103]. Therefore, we attempted to destabilize the

Ft by engineering the FT timer to also include the PEST sequence of SOCS3 at the C-terminus. Imaging of the new construct where the Ft expression was driven by SOCS3 promoter, mCherry/sfGFP was presented in a cytoplasmic distribution pattern that was very similar to the tFT alone.

To explore the dynamic activity of SOCS3 promoter in real time, we treated SOCS3 promoter-Ft cells with two different TLR ligands: LPS and flagellin. Both treatments reported a slow increase of Ft abundance. During single cell analysis, there was a significant level of variation observed in absolute mCherry/sfGFP signal ratio. However, this can be attributed to the stochastic nature of fluorescent protein maturation and specific cell histories contributing to this variability [431]. One the other hand, when SOCS3-promoter Ft cells were treated with LIF, the ratiometric analysis reported a high dynamic behaviour in the induction of Ft expression, with the mCherry/sfGFP Fluorescence Intensity Ratio fluctuating multiple times before a significant drop at 500 minutes.

Specifically, when measurements are presented with 10-minute intervals, the oscillations occur with a period of ~30 minutes. It is also evident that cell trajectories are oscillating with a high level of synchrony. This could be attributed by the low number of cells analysed to obtain this data. Thus, the experiment should be repeated to yield a more accurate reflection of SOCS3 promoter activity. Furthermore, the data acquisition parameters should also be changed by configuring the confocal microscope to acquire images from a culture well where cells are not subject to LIF treatment and from a culture well where cells are treated with LIF for every time point during the time course. This improvement will help to understand whether the oscillations we previously observed are either generated by the induced SOCS3 promoter or it is an

artefact of the confocal microscope. If the system retains its dynamic behaviour, then a Real-time PCR should be performed, where mRNA expression of tFT or SOCS3 is analysed in a time dependent manner in response to LIF treatment.

We also analysed the temporal changes of individual sfGFP and mCherry fluorescent intensities of 4 cells over time in order to investigate whether the observed oscillation in the ratios demonstrates a pulsatile activation of SOCS3 promoter or the degradation rate of the fluorescent reporter. It was evident from figure 4.23 that the behaviour of sfGFP and mCherry individual trajectories are comparatively similar to the overall ratio behaviour, in that both signals are oscillating with a similar frequency to the ratio. Interestingly, the period between ~50 and ~100 minutes (within phase 1), the mCherry signal is relatively static but the sfGFP is distinctively oscillating in an increasingly manner. This increase in sfGFP fluorescent intensity causes the fluorescent ratio to decrease in an oscillatory manner, demonstrating an activated SOCS3 promoter and a stimuli-induced Ft production. In contrast from ~290 minute onwards, the mCherry signal shows more dynamic behaviour than sfGFP. This could be explained by the promoter becoming inactive at the end of the time course. The mCherry signal appeared to remain overall at the same level but presented several short peaks, whereas the sfGFP fluorescent trajectories seemed to decrease in a steady fashion. According to this time period, it appears that the troughs and the peaks of the Ft ratios are mainly driven by the mCherry signal. Respectively, as the promoter becomes inactive, the oscillations in the Ft ratios appeared to be only influenced by the degradation rate of Ft. Collectively, it is evident that the troughs and the peaks observed in the Ft ratios could be regulated from both the LIF-induced production and the degradation rates of the Ft reporter. In our system, the degradation rates of the Ft reporter is regulated by the SOCS3 PEST sequence cloned at the end of mCherry cds. It would be interesting to investigate

whether LIF-treated cells retain this pulsatile activation of the truncated SOCS3 promoter when the Ft is expressed without the SOCS3 PEST sequence. In our knowledge, this is the first research study that used SOCS3 PEST sequence to decrease the long half-life of a fluorescent reporter. Thus, it would be also interesting to analyse and compare the half-lives of the Ft report with or without the SOCS3 PEST sequence by performing Cycloheximide chase assay.

Furthermore, it is important to note that the first major peak of the mean Ft ratio occurs at ~30 minutes in the time series, which coincides with the first peak of pSTAT3 protein oscillation in HIECs (figure 2.18). Although this time series was performed in a different cell line and the oscillation was induced by serum shock and not LIF treatment, it is interesting to see that the induction of SOCS3 promoter coincides with the increased expression of its transcription factor (i.e. pSTAT3). Future work could further investigate whether the pulsatile activation of SOCS3 promoter upon LIF treatment is corelated with the oscillatory expression of its transcription factors, for instance pSTAT3 or pSTAT1, since both are known to drive the transcription of SOCS3. Since STAT3 resides mainly in the cytoplasm in resting cells and upon ligand stimulation. it gets phosphorylated and translocases into the nucleus to induce the transcription of SOCS3 [432], it would be interesting to investigate whether pSTAT3 nuclear translocation coincide with the bursty promoter activity of SOCS3. This could be examined by tagging STAT3 with fluorescent protein and following its localisation using the confocal microscope. In theory, STAT3 should be largely located in cytoplasm and upon ligand stimulation pSTAT3 should be translocated into the nucleus. This could also be verified by performing a CHIP assay coupled with Quantitative PCR to confirm that pSTAT3 binds to SOCS3 promoter in cells within 30 minutes after LIF treatment.

Moreover, it should be noted that image acquisition always begun after 10-15 minutes of the ligand as the culture sterile room is in another floor from where the confocal microscope is located. Thus, it would be important to devise a way where data acquisition begun immediately after the addition of ligand without compromising culture sterility. This would help to investigate whether there is an immediate fluorescent response following the ligand addition.

Several studies aiming to unravel oscillatory behaviours of proteins (i.e. the activity of transcription factors) investigate target gene responses by a steady level of stimulation. However, steady inputs are rarely observed under natural conditions, and, at the cellular level, various signalling pathways have been evolved to give rise to a dynamic behaviour in transcription factor activity [433]. However, expression from weak promoter may need a prolonged accumulation of protein to detect and analyse efficiently the fluorescent signal *in vitro*. This extended stimulation period was chosen to ensure the investigation of steady-stated mRNA levels and subsequent the visualisation of reporter protein under the control of a single copy of promoter.

It is important to mention that imaging Ft expression under the control of a promoter that has not been imaged before, can be challenging to be calibrated appropriately to demonstrate equal and bright fluorescence in both channels and show the true temporal resolution for the promoter activity under study. We attempted to resolve that issue by using a reference solution that contained equal amounts of MitoTracker Red CMXRos (Thermofisher Scientific, C# M7512) and MitoTracker Green FM (Thermofisher Scientific, C# M7514) to optimize the imaging acquisition parameters prior to imaging the cells. However, it was evident that the brightness of these dyes were different from

that of mCherry/sfGFP. Future work should include a purified mCherry/sfGFP protein for a reference solution for imaging correction.

Finally, our data provide, to our knowledge for this first time, an *in vitro* analysis of SOCS3 promoter activity using Fluorescent timers.

4.5 Conclusion

So far, the use of Fts have been applied in different areas of cellular and organismal research.

Fluorescent timers (Fts) can be utilized as molecular genetically encoded tools to investigate trafficking of different cellular proteins and to gain knowledge into the timing of intracellular pathways [382]. The sequence of events during SOCS3 trafficking before it reaches its final compartment is largely unexplored.

The aim of this study was to investigate the spatial and temporal distribution of SOCS3 protein in live cells. Our work provided a descriptive analysis of tFTs as reporters of SOCS3 turn over and dynamic expression under normal culture conditions and in response to pharmacological agents. SOCS3 turnover was increased by rapamycin treatment and decreased by CQ and MG132 treatment. Interestingly, the fusion constructs differed broadly in their turnover, with C-SOCS3-Ft fusion protein being the most unstable. These results suggests that the fusion of the Ft at the N-terminus of SOCS3 stabilizes the protein, which in turn can be attributed to the fold of the tFT interfering with important degradation signals at the N-terminus of SOCS3. We provided a detailed characterization of SOCS3 pool's relative age in different cellular compartments and how this different when the fluorescent timer was fused either at the C or N- terminus of SOCS3.

Furthermore, our data provide, to our knowledge for this first time, an *in vitro* analysis of SOCS3 promoter activity using fluorescent timer. Future work should also investigate why LIF caused this dynamic behaviour of promoter activity.

Finally, it should be recognised that further work is needed for a clear characterization of the Ft properties to achieve an accurate interpretation of SOCS3 dynamics measurements across spatial and temporal dimensions.

5 Final Discussion

5.1 SOCS3 dynamics in a nutshell

Inflammation plays a crucial role to our protection against pathogens, nonetheless if it is not ordered and timely, the outcome of chronic inflammation can result in the development of various diseases. It is now widely accepted that inflammation plays a decisive role at multiple stages of carcinogenesis such as initiation, promotion, malignant conversion, invasion and metastasis [434]. Furthermore, chronic and non-resolving inflammation, today characterized as a hallmark of cancer [233], can increase the risk of multiple types of tumours, including hepatocellular carcinoma (HCC) and colon cancer [435]. The exact molecular pathways causing inflammation to promoter cancer are still not well understood. However, it is believed that cancer-stimulating inflammation, whether it anticipates or pursues tumorigenesis, belongs to a normal response to injury and infection that has been usurped by malignant cells to their own benefit [436].

Chronic intestinal inflammation in IBD has been shown to increase the risk of colitis-associated cancer (CAC) progression. Chronic inflammation in the intestine is often linked with high epithelial turnover, a disruptive mucosal barrier, aberrant activation of oncogenic signalling pathways and production of reactive oxygen species [437]. Subsequently, this results in a gradual accumulation of genetic and epigenetic alterations developing stepwise as low- and high-grade dysplasia and lastly as colorectal carcinoma [55].

Over the last three decades, it has been demonstrated that a vast array of cytokines, growth and hormonal factors released in response to inflammation transduce their signals via the JAKs and signal transducer and activator of transcriptions (STATs). Aberrant JAK/STAT signalling has been linked to malignancy and metastatic progression. Importantly, targeting of JAK/STAT pathways has become one of the most promising therapeutic interventions in prostate cancer, hematopoietic malignancies and sarcomas [438]. The CIS/suppressor of cytokine signalling (SOCS) proteins are well known regulators of the JAK-STAT pathway. So far, there are over 900 research articles demonstrating the link between SOCS proteins and cancer. Specifically, SOCS1 and SOCS3 proteins have been extensively studied using patient biopsies and gene-targeted mice and demonstrated the crucial roles of these proteins in different malignant processes, such as inflammation and tumorigenesis [438-440].

A research study conducted by Yoshiura et al. [61] illustrated a dynamic feedback regulation of the JAK/STAT pathway by SOCS3. Specifically, their findings of SOCS3 protein portraying an oscillatory behaviour with peaks 2 and 4 hours and pSTAT3 with peaks at 1, 3 and 5 hours laid the foundation for this PhD thesis. We hypothesized whether the perturbation or disruption of this dynamic feedback loop is a crucial factor driving the increased SOCS3 expression seen in IBD. Reciprocally, we also asked whether silencing of SOCS3 and disrupting the oscillatory feedback system is the leading factor to a constitutive JAK-STAT activation, which in turn is shown to drive the hyperproliferation of the epithelium and elevated tumour burden in colon cancer [118].

Thus, the main objective of this project focused on the oscillatory capacity of the tumour suppressor protein SOCS3 to investigate its role in cellular functions *using in vitro* cell lines.

Secondly, SOCS3 is also thought to be regulated through an autophagy-based mechanism and dysfunctional autophagy is a known causative mechanism of inflammatory bowel disease (IBD) [65]. Therefore, we also attempted to investigate the role of autophagy in regulation of oscillatory signaling.

Initially, we performed an extensive quantitative analysis of SOCS3 and LC3B-II proteins in serum starved Caco2 cells and differentiated Caco2 cells by performing a serum stimulation time course with time points every 30 minutes up to two hours and every 1 hours afterwards up to 8 hours. According to these results, SOCS3 expression did not illustrate any type of dynamic behaviour in response to serum treatment. Furthermore, the expression of LC3B-II protein was challenging to analyse with Western blotting as its antibody was generating high levels of noise and background on the blot. LC3B-II protein was mainly used to serve as an indicator of autophagy activity and its role would have been more useful at later studies where we would have been inducing or inhibiting autophagy to study its effects upon the oscillatory behaviour of SOCS3. Therefore, we halted LC3B-II protein analysis and only concentrated on establishing and studying the dynamic pattern of the basal expression of SOCS3 in proliferating Caco-2 cell line. In the following set of experiments, we attempted to find the optimal starvation duration and cell density so as to obtain a distinct and reproducible SOCS3 oscillatory pattern in proliferating Caco2 cells. However, none of these optimizing parameters seemed to play a decisive role in obtaining a SOCS3 oscillation. At this stage, we thought that the proliferating state of Caco-2 cells might

have lost the capacity to maintain a normal SOCS3 dynamic expression and results to a constant non-oscillatory pattern as illustrated by western blot analysis. Therefore, we also investigated whether the capacity of SOCS3 oscillation relies only on a non-malignant cell model system. Therefore, we studied whether the use of fully differentiated Caco-2 cells and HIECs can illustrate SOCS3 protein oscillatory expression. The following set of time series attempted to generate an optimized oscillatory model system in differentiated Caco-2 and HIECs by testing different serum starvation periods. However, none of the optimized serum starvation periods (overnight and 36 hours) did not exhibit a SOCS3 protein oscillation. We further attempted to extract a SOCS3 oscillation using a serum shock methodology to synchronize the cell population and elicit the rhythmicity of SOCS3 protein expression and investigate its functionality. Serum starvation and subsequent shock was able to achieve this to some extent. However, the consistency and level of the dynamic SOCS3 abundance was not sufficient to move forward and test autophagy inhibitors/inducers and determine how these changes affect the dynamic behavior.

Furthermore, we also concluded that western blotting analysis is not a sensitive and accurate method to study dynamic protein abundance as it appears to result in a population-wide average measurements and there are many potential stumbling blocks in this multistep method that can produce unreliable results [214].

Another method to study ultradian oscillations in a whole cell population setting, is by using Real-time PCR, where mRNA expression is analysed in a time dependent manner. However, this thesis did not examine SOCS3 mRNA expression as it was assumed that SOCS3 mRNA profiles would not represent true SOCS3 protein abundance in the cell. However, it is now known that analysis of SOCS3 mRNA was an important step but

was not carried out due to a lack of understand of its importance at the time of the experiments.

The extensive study of factors regulating the dynamics in biology has led to the observation that the rate of mRNA turnover is also a key element in the maintenance of ultradian oscillation. For instance, Hes1 is a bHLH repressor protein that was found to supress its own expression in a dynamic manner during somitogenesis. The dynamic expression of Hes1 has been investigated in a vast array of cell types, such as fibroblasts, neural progenitors and embryonic stem (ES) cells and demonstrated to oscillate with a period of 2 and 3 hours [441-443]. Furthermore, the half-life of Hes1 mRNA is estimated to be 20 minutes in several cell types, but in mouse ES cells it was found to be 40 minutes [442]. Interestingly, the period of Hes1 dynamic pattern was found to be 4 hours, longer than those seen in other cell lines. Consequently, this evidence supports the notion that mRNA half-life is probably affecting the oscillation period. Furthermore, the stabilization of Hes1 mRNA half-life by knockdown of micro-RNA 9 (miR-9), which also shares homology to the 3-UTR sequence of Hes1 mRNA, it disrupted oscillation in neural stem and progenitor cells [444, 445]. These observations further highlight the functional role of mRNA stability in the control of oscillatory dynamics.

Therefore, future work should analyse SOCS3 mRNA expression in response to serum stimulation and serum shock.

Although whole cell population studies analysing protein or gene expression have proven invaluable in dissecting signal transduction pathways and complex transcription networks, it is becoming increasingly recognised that single cell analysis represents a more accurately reflection of the molecular processes involved, especially when investigating transcriptional dynamics [206]. Recent advances in single cell analysis

has revealed surprisingly great intracellular heterogeneity in gene expression, which has brought forward many possible implications on a vast array of biological pathways which have not been considered previously [446].

To investigate SOCS3 dynamic behaviour at the single cell level, we first attempted to generate cell clones stably expressing SOCS3 C-terminal GFPSpark fusion protein from its respective endogenous promoter to monitor its expression in real time with confocal microscopy. Particularly, the fusion of a fluorescent protein to the coding sequence of SOCS3 at its endogenous locus represented an ideal strategy for monitoring protein dynamics as it would have preserved the normal expression level of the tagged protein [321]. We attempted to endogenously tag SOCS3 with a GFP by using the CRISPR/Cas9 system.

Other genome editing methods, such as zinc finger nuclease and TALENs, only require set of proteins to confer DNA specificity [326, 447]. However, CRISPR-Cas9 system confers targeting specificity to a particular site by using a guide RNA (gRNA), as well as the presence of a protospacer adjacent motif (PAM) sequence at the target DNA site [448]. At first, CRISPR/Cas9 strategy appeared simple to use and re-programme it to target different genomic locus or target several other genomic sites simultaneously. However, despite careful optimization of each step of CRISPR/Cas9 strategy, the generation of GFPSpark knockin cell line was not successful. There are multiple factors that potentially influenced Cas9/sgRNA editing efficiency. Therefore, this chapter should be reviewed as troubleshooting guide when future work attempts to endogenously tag SOCS3 or generally insert a reporter into a specific genomic locus.

Finally, we utilised the tandem fluorescent protein timer (tFT) strategy to gain a better insight into the dynamic the dynamics and mobility of SOCS3 protein and how this

relates to the different compartments within a cell. Using a single time-point measurement, the mCherry/sfGFP ratio of fluorescence intensities served as an indicator of the protein population age.

The use of tFT provided us the ability to study SOCS3 abundance across spatial and temporal dimensions under either normal culture conditions or different treatments that are known to impact on SOCS3 half-life and degradation rates. It is noteworthy to investigate whether the presence of the reporter gene affects the normal physiology and general characteristics of the transfected cells [449].

Further work and analysis on these Ft-fusion constructs is needed in order to gain a better understanding into the spatio-temporal SOCS3 dynamic in single cells.

6 Appendices

Supplementary material for chapter 3
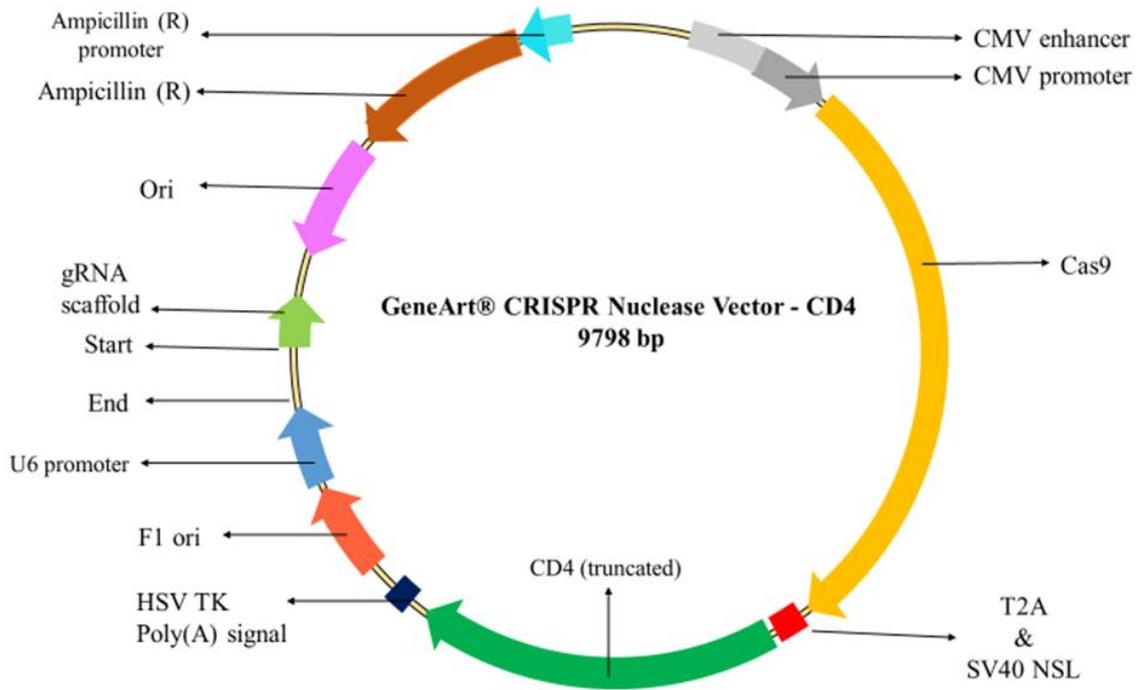
Figure 6.1: **GeneArt®CRISPR Nuclease Vector plasmid.** The GeneArt®CRISPR Vector

plasmid provided by the kit with a CD4 reporter for bead-based cell isolation.
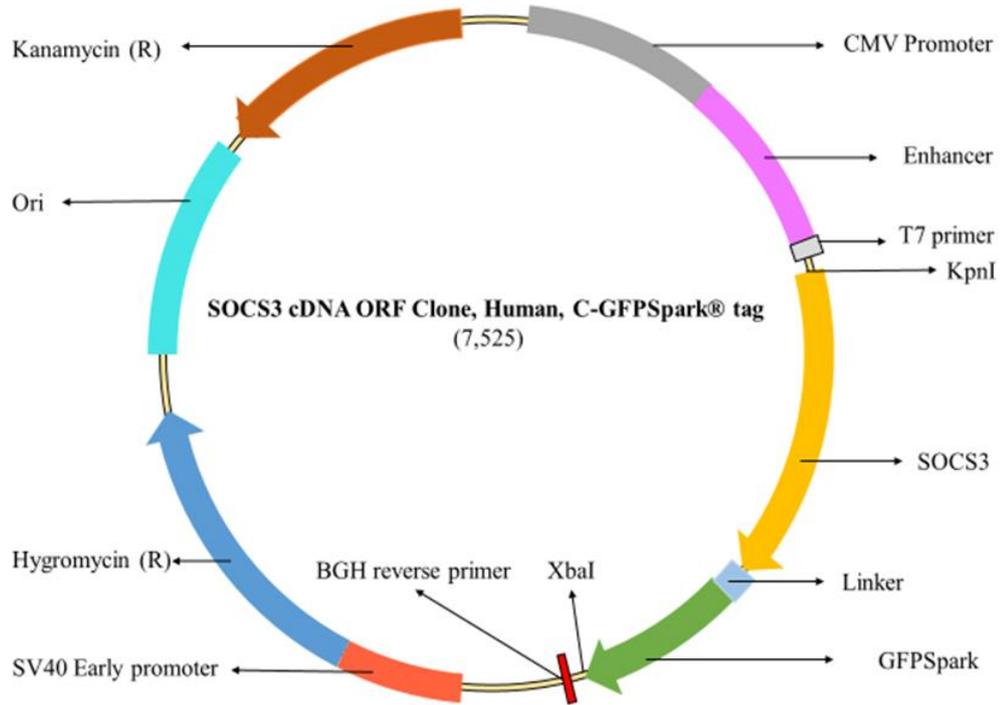
Figure 6.2: pCMV3-SOCS3-C-GFPSpark® tag vector plasmid (Sino Biological)
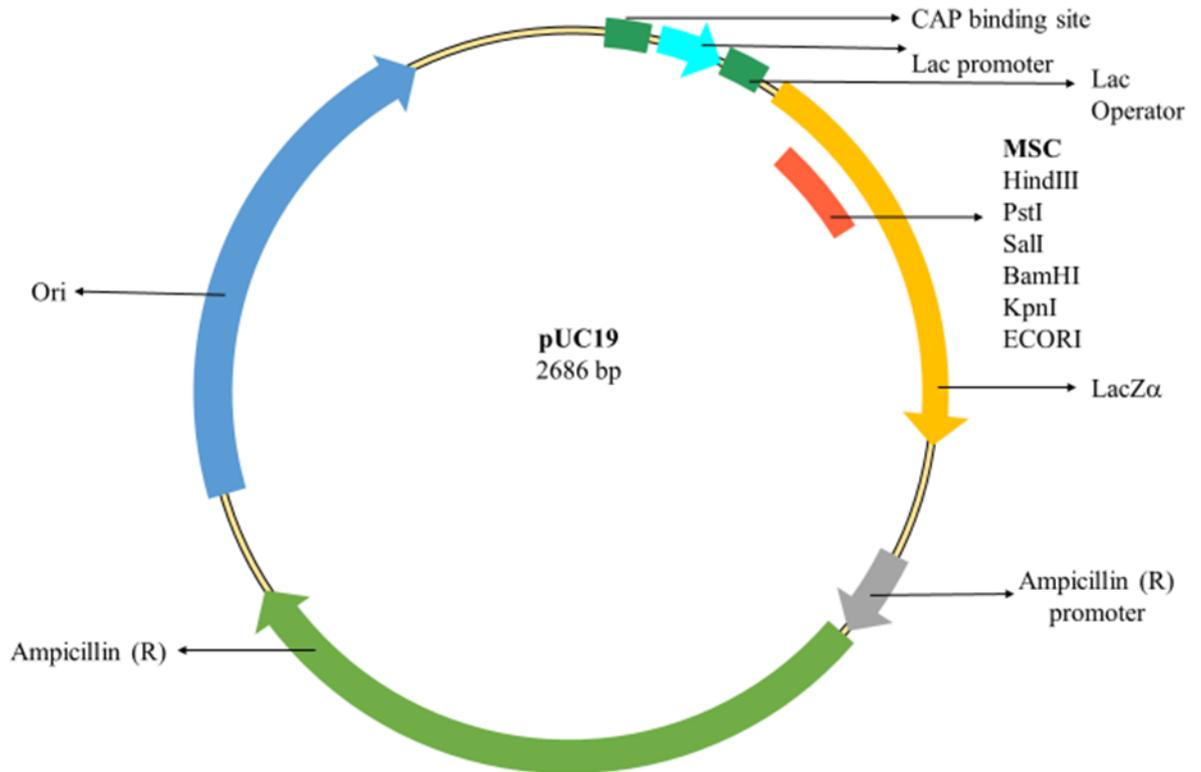
Figure 6.3: **Physical Map of pUC19 vector**. pUC19 vector plasmid was provided with the

Gibson Assembly Kit.

Supplementary material for chapter 4
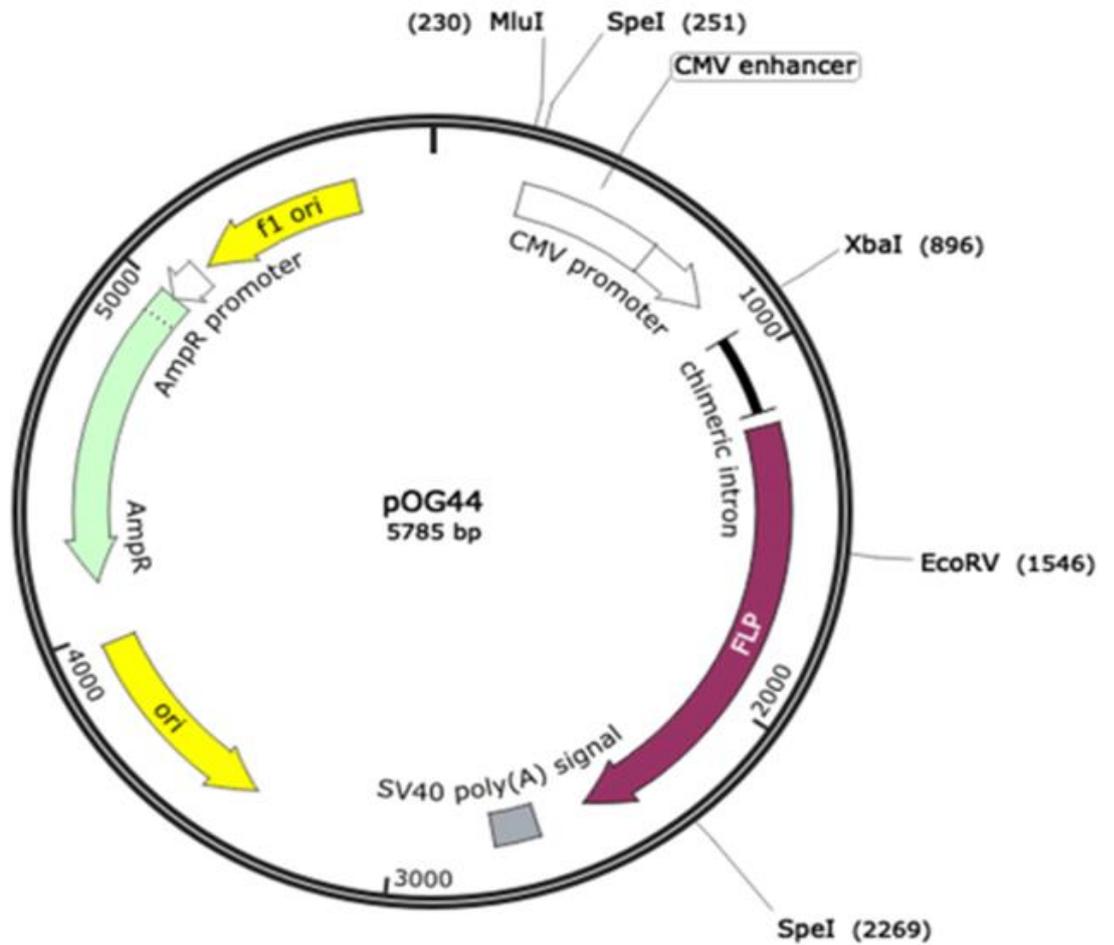
Figure 6.4: Map of pOG44 vector for constitutive expression of thermolabile Flp recombinase (flp-F70L) in mammalian cells.
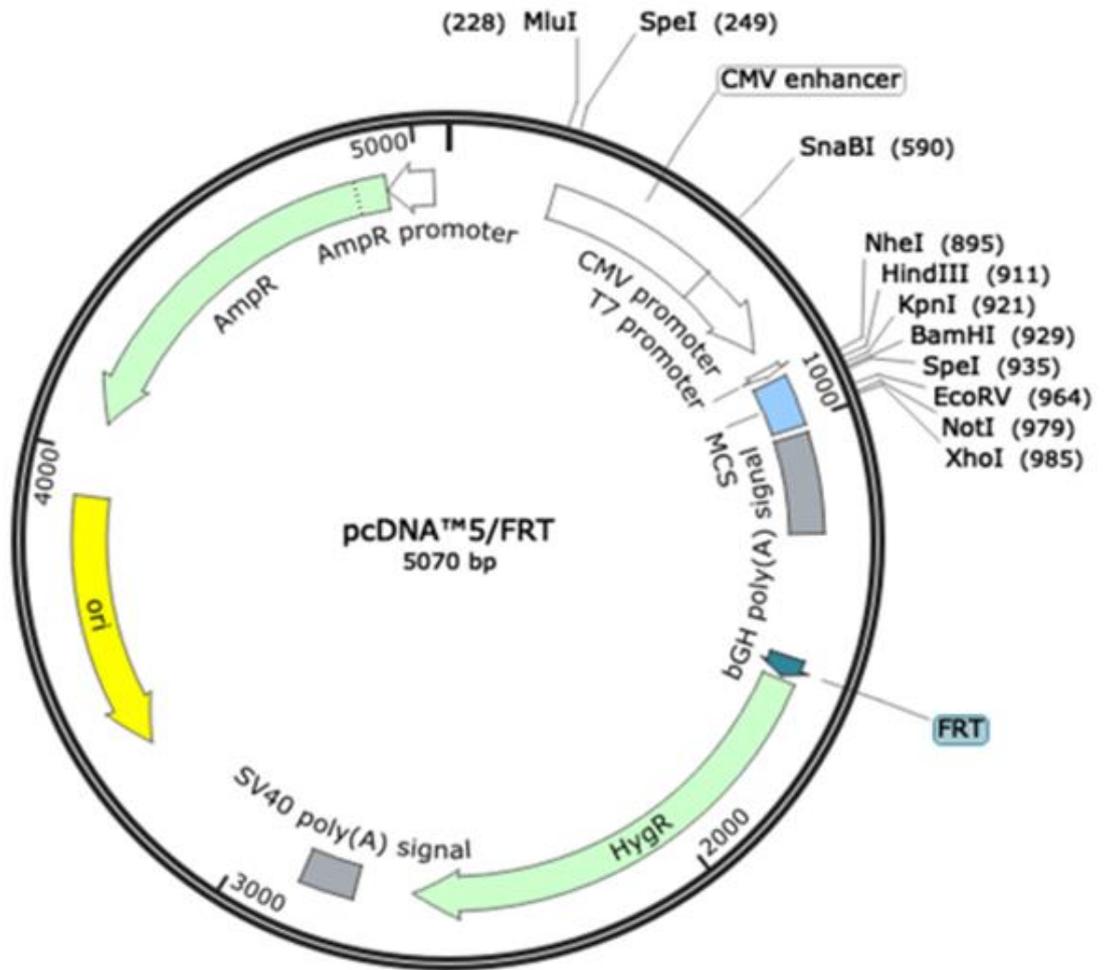
Figure 6.5: Map of pCDNA™5/FRT vector
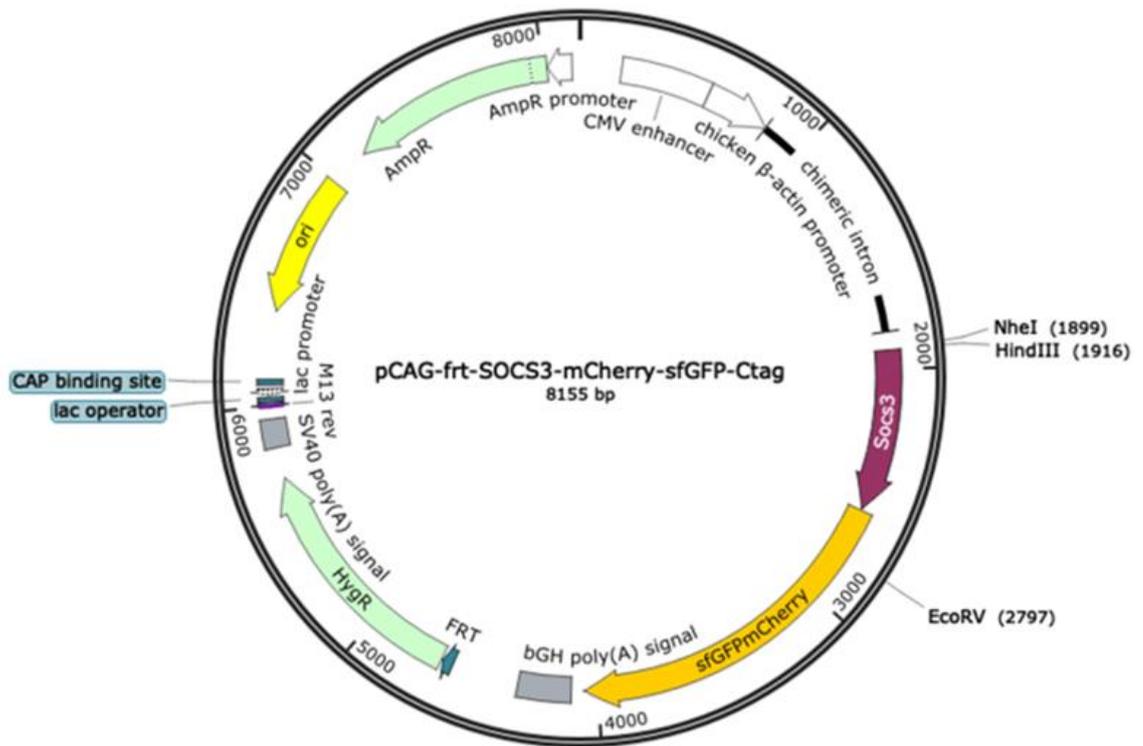
Figure 6.6: Vector Map for SOCS3-C-Ft construct
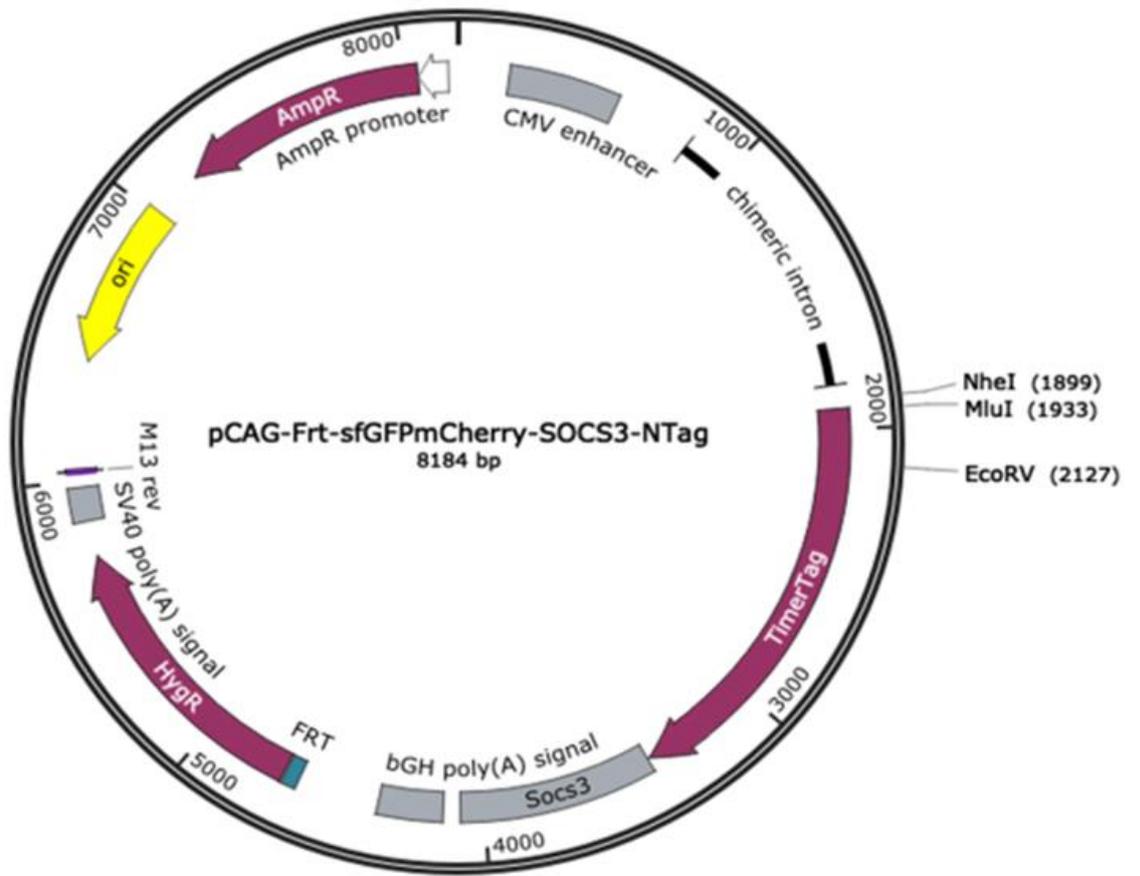
Figure 6.7: Vector Map for SOCS3-N-Ft construct
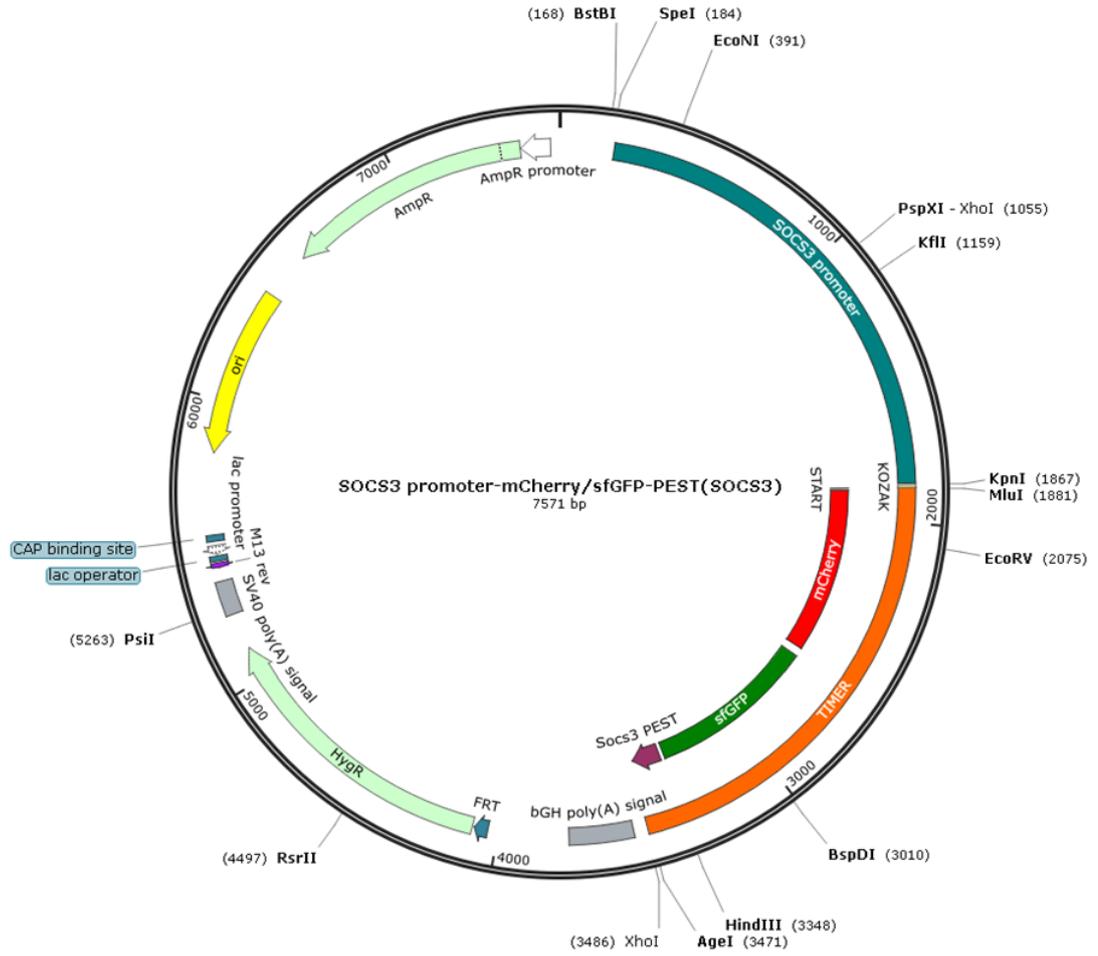
Figure 6.8: Vector map for SOCS3 promoter-mCherry/sfGFP-Pest(SOCS3)
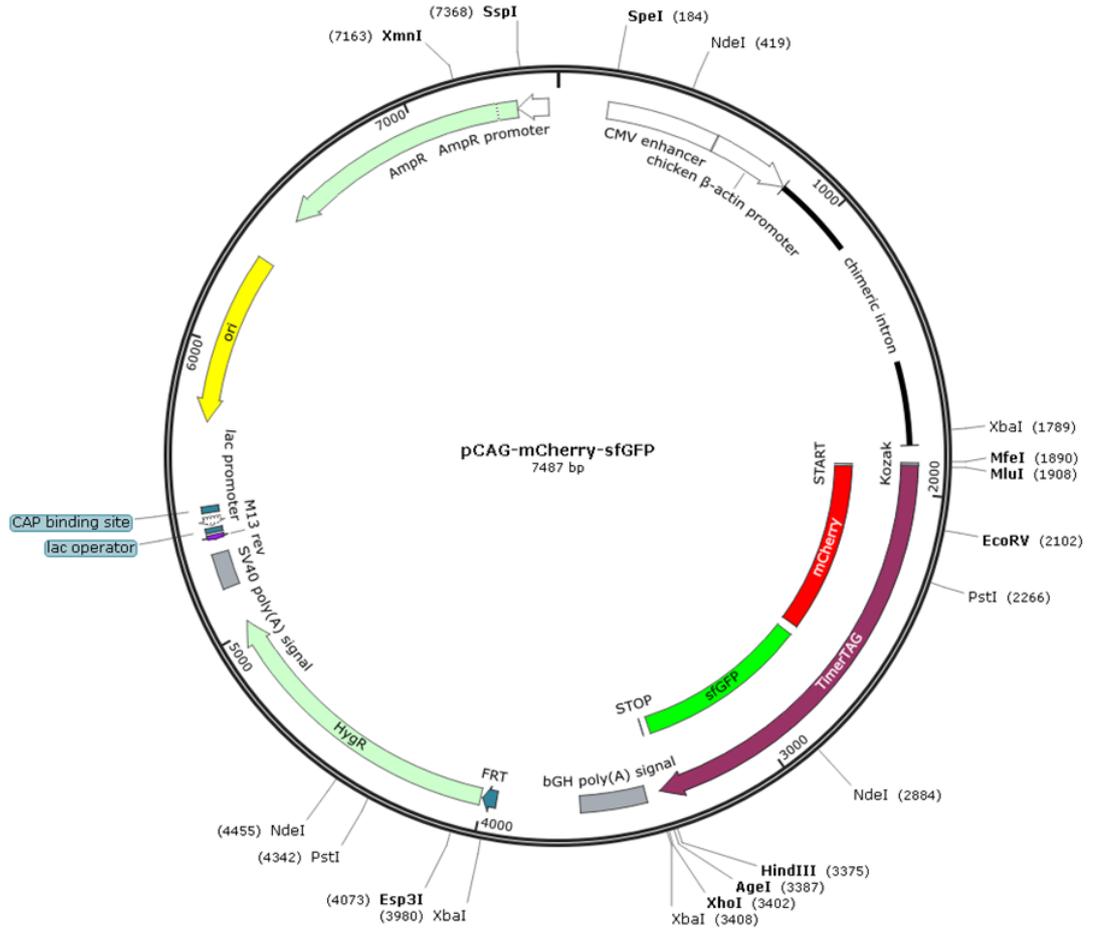
Figure 6.9: Vector Map for Control-Ft construct

Table 6.1 The mean ratios in each cell line for Nuclear and Intracellular Compartments were compared by two-way analysis of variance (ANOVA).

ANOVA Table	Type III Sum of Squares	Degree of Freedom	Mean Square	F	Significance
Between cell compartments	1.075	1	1.075	37.92	$p < 0.0001$
Between Cell lines	0.1495	2	0.07474	2.637	$p = 0.0746$
Cell compartments by cell lines (interaction)	0.5163	2	0.2581	9.106	$p = 0.0002$
Error	4.649	164	0.02835		
computed using $\alpha = 0.05$, adj. $p < 0.05$					

Table 6.2 Tukey's HSD (honestly significant difference) test to compare the means of Nucleus and Intracellular within each cell line.

	Mean Difference, Significance	
	Nucleus	**Intracellular**
Control-Ft vs. SOCS3-C-Ft	0.2005, $p < 0.0001$	-0.06478, $p = 0.3991$
Control-Ft vs. SOCS3-N-Ft	0.1006, $p = 0.0791$	0.01994, $p = 0.9604$
SOCS3-C-Ft vs. SOCS3-N-Ft	-0.09986, $p = 0.0704$	0.08472, $p = 0.1562$
Tukey's multiple comparisons test computed using $\alpha = 0.05$ and adj. $p < 0.05$.		

7 References

1. M'Koma, A.E., *Inflammatory Bowel Disease: An Expanding Global Health Problem.* Clinical Medicine Insights. Gastroenterology, 2013. **6**: p. 33-47.

2. Burisch, J., et al., *The burden of inflammatory bowel disease in Europe.* Journal of Crohn's and Colitis, 2013. **7**(4): p. 322-337.

3. Ostanin, D.V., et al., *T cell transfer model of chronic colitis: concepts, considerations, and tricks of the trade.* Gastrointestinal and Liver Physiology, 2009. **296**(2): p. 135-146.

4. Gadaleta, R.M., et al., *Farnesoid X receptor activation inhibits inflammation and preserves the intestinal barrier in inflammatory bowel disease.* Gut 2011. **60**(4): p. 463-472.

5. Williams, J., et al., *Interaction of suppressor of cytokine signalling 3 with cavin-1 links SOCS3 function and cavin-1 stability*. Vol. 9. 2018.

6. Duricova, D., et al., *Age-related differences in presentation and course of inflammatory bowel disease: an update on the population-based literature ☆.* Journal of Crohn's and Colitis, 2014. **8**(11): p. 1351-1361.

7. Abel Alexandra, G., et al., *Patchy colitis, and young age at diagnosis and at the time of surgery predict subsequent development of Crohn's disease after ileal pouch-anal anastomosis surgery for ulcerative colitis.* JGH Open, 2018. **2**(1): p. 8-14.

8. Biancheri, P., et al., *The challenges of stratifying patients for trials in inflammatory bowel disease.* Trends in Immunology, 2013. **34**(11): p. 564-571.

9. Lees, C.W., et al., *New IBD genetics: common pathways with other diseases.* Gut, 2011. **60**(12): p. 1739.

10. Tontini, G.E., et al., *Differential diagnosis in inflammatory bowel disease colitis: State of the art and future perspectives.* World Journal of Gastroenterology : WJG, 2015. **21**(1): p. 21-46.

11. Rigoli, L. and R.A. Caruso, *Inflammatory bowel disease in pediatric and adolescent patients: A biomolecular and histopathological review.* World Journal of Gastroenterology, 2014. **20**(30): p. 10262-10278.

12. Magro, F., et al., *European consensus on the histopathology of inflammatory bowel disease.* Journal of Crohn's and Colitis, 2013. **7**(10): p. 827–851.

13. Khor, B., A. Gardet, and R.J. Xavier, *Genetics and pathogenesis of inflammatory bowel disease.* Nature, 2011. **474**: p. 307–317.

14. Jin, J., *Inflammatory bowel disease.* JAMA, 2014. **311**(19): p. 2034-2034.

15. Li, J., et al., *Functional Impacts of the Intestinal Microbiome in the Pathogenesis of Inflammatory Bowel Disease.* Inflammatory Bowel Diseases, 2015. **21**(1): p. 139-153.

16. Zhang, Y.-Z. and Y.-Y. Li, *Inflammatory bowel disease: Pathogenesis.* World Journal of Gastroenterology : WJG, 2014. **20**(1): p. 91-99.

17. Matricon, J., N. Barnich, and D. Ardid, *Immunopathogenesis of inflammatory bowel disease.* Self Nonself, 2010. **1**(4): p. 299-309.

18. Guslandi, M., *A natural approach to treatment of inflammatory bowel disease.* British Journal of Clinical Pharmacology, 2008. **65**(4): p. 468–469.

19. Rabišková, M., et al., *Coated chitosan pellets containing rutin intended for the treatment of inflammatory bowel disease: in vitro characteristics and in vivo evaluation.* International Journal of Pharmaceutics, 2012. **422**(1-2): p. 151-159.

20. Raghu Subramanian, C. and G. Triadafilopoulos, *Care of inflammatory bowel disease patients in remission.* Gastroenterol Rep (Oxf), 2016. **4**(4): p. 261-71.

21. Vasudevan, A., P.R. Gibson, and D.R. van Langenberg, *Time to clinical response and remission for therapeutics in inflammatory bowel diseases: What should the clinician expect, what should patients be told?* World Journal of Gastroenterology, 2017. **23**(35): p. 6385-6402.

22. Oikonomou, K.A., et al., *Drug-Induced Nephrotoxicity in Inflammatory Bowel Disease.* Nephron Clinical Practice, 2011. **119**(2): p. c89-c96.

23. Triantafillidis, J.K., E. Merikas, and F. Georgopoulos, *Current and emerging drugs for the treatment of inflammatory bowel disease.* Drug Design, Development and Therapy, 2011. **5**: p. 185-210.

24. Cohen, B.L. and D.B. Sachar, *Update on anti-tumor necrosis factor agents and other new drugs for inflammatory bowel disease.* BMJ, 2017. **357**.

25. Dalal, S.R. and R.D. Cohen, *What to Do When Biologic Agents Are Not Working in Inflammatory Bowel Disease Patients.* Gastroenterology & Hepatology, 2015. **11**(10): p. 657-665.

26.   Amiot, A. and L. Peyrin-Biroulet, *Current, new and future biological agents on the horizon for the treatment of inflammatory bowel diseases.* Therapeutic Advances in Gastroenterology, 2015. **8**(2): p. 66-82.

27.   Lopez, J. and A. Grinspan, *Fecal Microbiota Transplantation for Inflammatory Bowel Disease.* Gastroenterology & Hepatology, 2016. **12**(6): p. 374-379.

28.   Jostins, L., et al., *Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease.* Nature, 2012. **491**: p. 119.

29.   Ek, W.E., M. D'Amato, and J. Halfvarson, *The history of genetics in inflammatory bowel disease.* Annals of Gastroenterology : Quarterly Publication of the Hellenic Society of Gastroenterology, 2014. **27**(4): p. 294-303.

30.   Flamant, M., et al., *Advances in the Development of Janus Kinase Inhibitors in Inflammatory Bowel Disease: Future Prospects.* Drugs, 2017. **77**(10): p. 1057-1068.

31.   Zundler, S. and M.F. Neurath, *Integrating Immunologic Signaling Networks: The JAK/STAT Pathway in Colitis and Colitis-Associated Cancer.* Vaccines, 2016. **4**(1): p. 5.

32.   Liu, Q.-F., et al., *Association of STAT4 rs7574865 polymorphism with susceptibility to inflammatory bowel disease: A systematic review and meta-analysis.* Clinics and Research in Hepatology and Gastroenterology, 2015. **39**(5): p. 627-636.

33.   Uk, C.R. *Statistics by cancer type*. [Online] Accessed 15/04/2018; Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer.

34.   Mokarram, P., et al., *New frontiers in the treatment of colorectal cancer: Autophagy and the unfolded protein response as promising targets.* Autophagy, 2017. **13**(5): p. 781-819.

35.   Vasen, H.F.A., et al., *Guidelines for the clinical management of familial adenomatous polyposis (FAP).* Gut, 2008. **57**(5): p. 704-713.

36.   Rustgi, A.K., *The genetics of hereditary colon cancer.* Genes & Development, 2007. **21**(20): p. 2525-2538.

37.   Alnabulsi, A. and G.I. Murray, *Proteomics for early detection of colorectal cancer: recent updates.* Expert Review of Proteomics, 2018. **15**(1): p. 55-63.

38.   Sandouk, F., F. Al Jerf, and M. H D Bassel Al-Halabi, *Precancerous Lesions in Colorectal Cancer.* Vol. 2013. 2013. 457901.

39.   Tanaka, T., *Colorectal carcinogenesis: Review of human and experimental animal studies.* Journal of Carcinogenesis, 2009. **8**: p. 5.

40. Subramaniam, R., A. Mizoguchi, and E. Mizoguchi, *Mechanistic roles of epithelial and immune cell signaling during the development of colitis-associated cancer.* Cancer research frontiers, 2016. **2**(1): p. 1-21.

41. Pancione, M., A. Remo, and V. Colantuoni, *Genetic and Epigenetic Events Generate Multiple Pathways in Colorectal Cancer Progression.* Pathology Research International, 2012. **2012**: p. 11.

42. Irrazábal, T., et al., *The Multifaceted Role of the Intestinal Microbiota in Colon Cancer.* Molecular Cell, 2014. **54**(2): p. 309-320.

43. Li, Y., et al., *IL-6-induced DNMT1 activity mediates SOCS3 promoter hypermethylation in ulcerative colitis-related colorectal cancer.* Carcinogenesis, 2012. **33**(10): p. 1889-1896.

44. Pino, M.S. and D.C. Chung, *The Chromosomal Instability Pathway in Colon Cancer.* Gastroenterology, 2010. **138**(6): p. 2059-2072.

45. Yamagishi, H., et al., *Molecular pathogenesis of sporadic colorectal cancers.* Chin J Cancer, 2016. **35**.

46. Aziz, M.A., et al., *Towards personalized medicine of colorectal cancer.* Critical Reviews in Oncology/Hematology, 2017. **118**: p. 70-78.

47. Mattar, M.C., et al., *Current Management of Inflammatory Bowel Disease and Colorectal Cancer.* Gastrointestinal Cancer Research, 2011. **4**(2): p. 53–61.

48. Bernstein, C.N., et al., *World Gastroenterology Organization Practice Guidelines for the Diagnosis and Management of IBD in 2010.* Inflammatory Bowel Diseases, 2010. **16**(1): p. 112–124.

49. Triantafillidis, J.K., G. Nasioulas, and P. Kosmidis, *Colorectal cancer and inflammatory bowel disease: epidemiology, risk factors, mechanisms of carcinogenesis and prevention strategies.* Anticancer Research, 2009. **29**(7): p. 2727-2737.

50. Hedrick, T.L. and C. Friel, *Colonic Crohn Disease.* Clinics in Colon and Rectal Surgery, 2013. **26**(2): p. 84–89.

51. Neufert, C., et al., *Tumor fibroblast–derived epiregulin promotes growth of colitis-associated neoplasms through ERK.* The Journal of Clinical Investigation, 2013. **123**(1428-1443).

52. Dupaul-Chicoine, J., et al., *Control of Intestinal Homeostasis, Colitis, and Colitis-Associated Colorectal Cancer by the Inflammatory Caspases.* Immunity, 2010. **32**(3): p. 367-378.

53. Lutgens, M., et al., *High frequency of early colorectal cancer in inflammatory bowel disease.* Gut, 2008. **57**(9): p. 1246-1251.

54. Antoniou, E., et al., *Cytokine Networks in Animal Models of Colitis-associated Cancer.* Anticancer Research, 2015. **35**(1): p. 19-24.

55. Grivennikov, S.I., *Inflammation and colorectal cancer: colitis-associated neoplasia.* Seminars in immunopathology, 2013. **35**(2): p. 229-244.

56. Luo, C. and H. Zhang, *The Role of Proinflammatory Pathways in the Pathogenesis of Colitis-Associated Colorectal Cancer.* Mediators of Inflammation, 2017. **2017**: p. 5126048.

57. Samadi, A.K., et al., *A multi-targeted approach to suppress tumor-promoting inflammation.* Seminars in Cancer Biology, 2015. **35**: p. S151-S184.

58. Werner, S.L., D. Barken, and A. Hoffmann, *Stimulus Specificity of Gene Expression Programs Determined by Temporal Control of IKK Activity.* Science, 2005. **309**(5742): p. 1857.

59. Ashall, L., *Pulsatile Stimulation Determines Timing and Specificity of NF-κB-Dependent Transcription.* Science 2009. **324**(5924): p. 242-246.

60. Nan, Y., C. Wu, and Y.-J. Zhang, *Interplay between Janus Kinase/Signal Transducer and Activator of Transcription Signaling Activated by Type I Interferons and Viral Antagonism.* Frontiers in Immunology, 2017. **8**: p. 1758.

61. Yoshiura, S., *Ultradian oscillations of Stat, Smad, and Hes1 expression in response to serum.* Proceedings of the National Academy of Sciences, 2007. **104**(27): p. 11292–11297.

62. Kroemer, G., G. Mariño, and B. Levine, *Autophagy and the integrated stress response.* Molecular cell, 2010. **40**(2): p. 280-293.

63. Hwang, W.C., et al., *Inhibition of phospholipase D2 induces autophagy in colorectal cancer cells.* Experimental &Amp; Molecular Medicine, 2014. **46**: p. e124.

64. Gardet, A. and R.J. Xavier, *Common alleles that influence autophagy and the risk for inflammatory bowel disease.* Current Opinion in Immunology, 2012. **24**(5): p. 522-529.

65. Koay, L.C., R.J. Rigby, and K.L. Wright, *Cannabinoid-induced autophagy regulates suppressor of cytokine signaling-3 in intestinal epithelium.* American Journal of Physiology - Gastrointestinal and Liver Physiology, 2014. **307**(2): p. G140-G148.

66. Waldner, M.J. and M.F. Neurath, *Master regulator of intestinal disease: IL-6 in chronic inflammation and cancer development.* Seminars in Immunology, 2014. **26**(1): p. 75-79.

67. O'Shea, J.J. and R. Plenge, *JAK and STAT signaling molecules in immunoregulation and immune-mediated disease.* Immunity, 2012. **36**(4): p. 542–550.

68. Zhang, W.-N., et al., *CUEDC2 (CUE Domain-containing 2) and SOCS3 (Suppressors of Cytokine Signaling 3) Cooperate to Negatively Regulate Janus*

Kinase 1/Signal Transducers and Activators of Transcription 3 Signaling. The Journal of Biological Chemistry, 2012. **287**(1): p. 382-392.

69. Baker, S.J., S.G. Rane, and E.P. Reddy, *Hematopoietic cytokine receptor signaling.* Oncogene, 2007. **26**(47): p. 6724-6737.

70. Coskun, M., et al., *Involvement of JAK/STAT signaling in the pathogenesis of inflammatory bowel disease.* Pharmacological Research, 2013. **76**: p. 1-8.

71. Speil, J., et al., *Activated STAT1 Transcription Factors Conduct Distinct Saltatory Movements in the Cell Nucleus.* Biophysical Journal, 2011. **101**(11): p. 2592-2600.

72. Nardozzi, J.D., K. Lott, and G. Cingolani, *Phosphorylation meets nuclear import: a review.* Cell Communication and Signaling : CCS, 2010. **8**: p. 32-32.

73. Wirthmueller, L., et al., *Hop-on hop-off: importin-α-guided tours to the nucleus in innate immune signaling.* Frontiers in Plant Science, 2013. **4**(149).

74. Levy, D.E. and J.E. Darnell Jr, *STATs: transcriptional control and biological impact.* Nature Reviews Molecular Cell Biology, 2002. **3**: p. 651.

75. Palmer, D.C. and N.P. Restifo, *Suppressors of Cytokine Signaling (SOCS) in T cell differentiation, maturation, and function.* Trends in immunology, 2009. **30**(12): p. 592-602.

76. Sasi, W., A.K. Sharma, and K. Mokbel, *The Role of Suppressors of Cytokine Signalling in Human Neoplasms.* Molecular Biology International, 2014. **2014**: p. 24.

77. Yoshimura, A., T. Naka, and M. Kubo, *SOCS proteins, cytokine signalling and immune regulation.* Nature Reviews Immunology, 2007. **7**: p. 454.

78. Zhang, T., et al., *The Coiled-Coil Domain of Stat3 Is Essential for Its SH2 Domain-Mediated Receptor Binding and Subsequent Activation Induced by Epidermal Growth Factor and Interleukin-6.* Molecular and Cellular Biology, 2000. **20**(19): p. 7132-7139.

79. Bowman, T., et al., *STATs in oncogenesis.* Oncogene, 2000. **19**: p. 2474.

80. Lesina, M., et al., *Stat3/Socs3 Activation by IL-6 Transsignaling Promotes Progression of Pancreatic Intraepithelial Neoplasia and Development of Pancreatic Cancer.* Cancer Cell, 2011. **19**(4): p. 456-469.

81. Bromberg, J., *Stat proteins and oncogenesis.* The Journal of Clinical Investigation, 2002. **109**(9): p. 1139-1142.

82. Atreya, R. and M.F. Neurath, *Involvement of IL-6 in the pathogenesis of inflammatory bowel disease and colon cancer.* Clinical Reviews in Allergy & Immunology, 2005. **28**(3): p. 187-195.

83. Lee, H., et al., *Stat3-induced S1PR1 expression is critical for persistent Stat3 activation in tumors.* Nature medicine, 2010. **16**(12): p. 1421-1428.

84. Yu, H. and R. Jove, *The STATs of cancer — new molecular targets come of age.* Nature Reviews Cancer, 2004. **4**: p. 97.

85. O'Shea, J.J., M. Gadina, and R.D. Schreiber, *Cytokine Signaling in 2002.* Cell, 2002. **109**(2): p. S121-S131.

86. Gibson, D.L., et al., *Interleukin-11 Reduces TLR4-Induced Colitis in TLR2-Deficient Mice and Restores Intestinal STAT3 Signaling.* Gastroenterology, 2010. **139**(4): p. 1277-1288.

87. Neufert, C., et al., *Activation of epithelial STAT3 regulates intestinal homeostasis.* Cell Cycle, 2010. **9**(4): p. 652-655.

88. Goldsmith, J.R., J.M. Uronis, and C. Jobin, *Mu Opioid Signaling Protects Against Acute Murine Intestinal Injury in a Manner Involving Stat3 Signaling.* The American Journal of Pathology, 2011. **179**(2): p. 673–683.

89. Grivennikov, S., et al., *IL-6 and STAT3 are required for survival of intestinal epithelial cells and development of colitis associated cancer.* Cancer cell, 2009. **15**(2): p. 103-113.

90. Becker, C., et al., *TGF-β Suppresses Tumor Progression in Colon Cancer by Inhibition of IL-6 trans-Signaling.* Immunity, 2004. **21**(4): p. 491-501.

91. Bollrath, J., et al., *gp130-Mediated Stat3 Activation in Enterocytes Regulates Cell Survival and Cell-Cycle Progression during Colitis-Associated Tumorigenesis.* Cancer Cell, 2009. **15**(2): p. 91-102.

92. Kershaw, N.J., *SOCS3 binds specific receptor–JAK complexes to control cytokine signaling by direct kinase inhibition.* Nature Structural & Molecular Biology, 2013. **20**(4): p. 469-478.

93. Yoshimura, A., et al., *SOCS, inflammation, and autoimmunity.* Frontier of Immunology, 2012. **3**: p. 1-9.

94. Tamiya, T., et al., *Suppressors of cytokine signaling (SOCS) proteins and JAK/STAT pathways: Regulation of T-cell inflammation by SOCS1 and SOCS3.* Arteriosclerosis, Thrombosis, and Vascular Biology, 2011. **31**: p. 980-985.

95. Piessevaux, J., et al., *The many faces of the SOCS box.* Cytokine & Growth Factor Reviews, 2008. **19**(5): p. 371-381.

96. Babon, J.J., L.N. Varghese, and N.A. Nicola, *Inhibition of IL-6 family cytokines by SOCS3.* Seminars in Immunology, 2014. **26**(1): p. 13-19.

97. Kershaw, N.J., et al., *SOCS3 binds specific receptor–JAK complexes to control cytokine signalling by direct kinase inhibition.* Nature Structural & Molecular Biology, 2013. **20**(4): p. 469-476.

98. Yasukawa, H., et al., *The JAK-binding protein JAB inhibits Janus tyrosine kinase activity through binding in the activation loop.* EMBO Journal, 1999. **18**(5): p. 1309-1320.

99. Sasaki, A., et al., *Cytokine-inducible SH2 protein-3 (CIS3/SOCS3) inhibits Janus tyrosine kinase by binding through the N-terminal kinase inhibitory region as well as SH2 domain.* Genes to Cells, 1999. **4**(6): p. 339-351.

100. Kamura, T., et al., *VHL-box and SOCS-box domains determine binding specificity for Cul2-Rbx1 and Cul5-Rbx2 modules of ubiquitin ligases.* Genes & Development, 2004. **18**(24): p. 3055-3065.

101. Gupta, I., et al., *Delineating Crosstalk Mechanisms of the Ubiquitin Proteasome System That Regulate Apoptosis.* Frontiers in Cell and Developmental Biology, 2018. **6**: p. 11.

102. Sasaki, A., et al., *The N-terminal Truncated Isoform of SOCS3 Translated from an Alternative Initiation AUG Codon under Stress Conditions Is Stable Due to the Lack of a Major Ubiquitination Site, Lys-6.* Journal of Biological Chemistry, 2003. **278**(4): p. 2432-2436.

103. Babon, J.J., et al., *The Structure of SOCS3 Reveals the Basis of the Extended SH2 Domain Function and Identifies an Unstructured Insertion That Regulates Stability.* Molecular Cell, 2006. **22**(2): p. 205-216.

104. Kato, H., et al., *Association of single-nucleotide polymorphisms in the suppressor of cytokine signaling 2 (SOCS2) gene with type 2 diabetes in the Japanese.* Genomics, 2006. **87**(4): p. 446-458.

105. Starr, R., et al., *A family of cytokine-inducible inhibitors of signalling.* Nature, 1997. **387**: p. 917.

106. Roberts, A.W., et al., *Placental defects and embryonic lethality in mice lacking suppressor of cytokine signaling 3.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(16): p. 9324-9329.

107. Evans, M.K., et al., *Expression of SOCS1 and SOCS3 genes is differentially regulated in breast cancer cells in response to proinflammatory cytokine and growth factor signals.* Oncogene, 2006. **26**: p. 1941.

108. Liu, X., et al., *Suppressors of Cytokine-Signaling Proteins Induce Insulin Resistance in the Retina and Promote Survival of Retinal Cells.* Diabetes, 2008. **57**(6): p. 1651.

109. Suzuki, A., et al., *Cis3/Socs3/Ssi3 Plays a Negative Regulatory Role in Stat3 Activation and Intestinal Inflammation.* The Journal of Experimental Medicine, 2001. **193**(4): p. 471-482

110. White, G.E., et al., *Suppressor of cytokine signalling protein SOCS3 expression is increased at sites of acute and chronic inflammation.* Journal of Molecular Histology, 2011. **42**(2): p. 137-151.

111. Niemand, C., et al., *Activation of STAT3 by IL-6 and IL-10 in Primary Human Macrophages Is Differentially Modulated by Suppressor of Cytokine Signaling 3.* The Journal of Immunology, 2003. **170**(6): p. 3263.

112. Hontecillas, R., et al., *Immunoregulatory mechanisms of macrophage PPAR-γ in mice with experimental inflammatory bowel disease.* Mucosal Immunology, 2010. **4**: p. 304.

113. Gil, M.P., et al., *Regulating type 1 IFN effects in CD8 T cells during viral infections: changing STAT4 and STAT1 expression for function.* Blood, 2012. **120**(18): p. 3718-3728.

114. Miyagi, T., et al., *High basal STAT4 balanced by STAT1 induction to control type 1 interferon effects in natural killer cells.* The Journal of Experimental Medicine, 2007. **204**(10): p. 2383-2396.

115. Isomoto, H., et al., *Sustained IL-6/STAT-3 Signaling in Cholangiocarcinoma Cells due to SOCS-3 Epigenetic Silencing.* Gastroenterology, 2007. **132**(1): p. 384-396.

116. Li, Y., et al., *Disease-related expression of the IL6/STAT3/SOCS3 signalling pathway in ulcerative colitis and ulcerative colitis-related carcinogenesis.* Gut, 2010. **59**(2): p. 227.

117. Pickert, G., et al., *STAT3 links IL-22 signaling in intestinal epithelial cells to mucosal wound healing.* The Journal of Experimental Medicine, 2009. **206**(7): p. 1465-1472.

118. Rigby, R.J., et al., *Suppressor of cytokine signaling 3 (SOCS3) limits damage-induced crypt hyper-proliferation and inflammation-associated tumorigenesis in the colon.* Oncogene, 2007. **26**: p. 4833.

119. Thagia, I., et al., *Intestinal epithelial suppressor of cytokine signaling 3 enhances microbial-induced inflammatory tumor necrosis factor-α, contributing to epithelial barrier dysfunction.* American Journal of Physiology - Gastrointestinal and Liver Physiology, 2015. **308**(1): p. 25-31.

120. Cheng, X., et al., *miR-19b downregulates intestinal SOCS3 to reduce intestinal inflammation in Crohn's disease.* Scientific Reports, 2015. **5**(10397): p. 1-15.

121. Carow, B. and M.E. Rottenberg, *SOCS3, a Major Regulator of Infection and Inflammation.* Frontiers in Immunology, 2014. **5**(58).

122. Shaw, A. and E. Filbert, *Scaffold proteins and immune-cell signalling.* Nature Reviews Immunology 2009. **9**(1): p. 47-56.

123. Mengel, B., et al., *Modeling oscillatory control in NF-κB, p53 and Wnt signaling.* Current Opinion in Genetics & Development, 2010. **20**(6): p. 656–664.

124. Pigolotti, S., S. Krishna, and M.H. Jensen, *Oscillation patterns in negative feedback loops.* Proceedings of the National Academy of Sciences, 2007. **104**(16): p. 6533–6537.

125. Ankers, J.M., et al., *Spatio-temporal protein dynamics in single living cells.* Current Opinion in Biotechnology, 2008. **19**(4): p. 375–380.

126. Shimojo, H., T. Ohtsuka, and R. Kageyama, *Oscillations in notch signaling regulate maintenance of neural progenitors.* Neuron, 2008. **58**(1): p. 52–64.

127. Nelson, D.E., et al., *Oscillations in NF-.B Signaling Control the Dynamics of Gene Expression.* Science, 2004. **306**(5696): p. 704-708.

128. Moore, R., et al., *MiR-192-Mediated Positive Feedback Loop Controls the Robustness of Stress-Induced p53 Oscillations in Breast Cancer Cells.* PLOS Computational Biology, 2015. **11**(12): p. e1004653.

129. Isomura, A., *Ultradian oscillations and pulses: coordinating cellular responses and cell fate decisions.* Development, 2014. **141**(19): p. 3627-3636.

130. Alon, U., *Network motifs: theory and experimental approaches.* Nature Reviews Genetics, 2007. **8**: p. 450.

131. Kolch, W., et al., *The dynamic control of signal transduction networks in cancer cells.* Nature Reviews Cancer, 2015. **15**: p. 515.

132. Moss Bendtsen, K., et al., *The role of mRNA and protein stability in the function of coupled positive and negative feedback systems in eukaryotic cells.* Scientific Reports, 2015. **5**: p. 13910.

133. Novák, B.l. and J.J. Tyson, *Design principles of biochemical oscillators.* Nature Reviews Molecular Cell Biology, 2008. **9**: p. 981-991.

134. Shi, D. and W. Gu, *Dual Roles of MDM2 in the Regulation of p53.* Genes & Cancer, 2012. **3**(3-4): p. 240–248.

135. Cheong, R., S. Paliwal, and A. Levchenko, *Models at the single cell level.* Wiley Interdisciplinary Reviews: Systems Biology and Medicine, 2009. **2**(1): p. 34-48.

136. Bar-Or, R.L., et al., *Generation of oscillations by the p53-Mdm2 feedback loop: A theoretical and experimental study.* PNAS, 2000. **97**(21): p. 11250–11255.

137. Batchelor, E., et al., *Stimulus-dependent dynamics of p53 in single cells.* Molecular Systems Biology, 2011. **7**(1): p. 488.

138. Purvis, J.E., et al., *p53 Dynamics Control Cell Fate.* Science, 2012. **336**(6087): p. 1440-1444.

139. Batchelor, E., A. Loewer, and G. Lahav, *The ups and downs of p53: understanding protein dynamics in single cells.* Nature Reviews Cancer, 2009. **9**: p. 371-377.

140. Batchelor, E., et al., *Recurrent Initiation: A Mechanism for Triggering p53 Pulses in Response to DNA Damage.* Molecular Cell, 2008. **30**(3): p. 277-289.

141. Vera, J., et al., *Systems biology of JAK-STAT signalling in human malignancies.* Progress in Biophysics and Molecular Biology, 2011. **106**(2): p. 426-434.

142. Raia, V., et al., *Dynamic Mathematical Modeling of IL13-Induced Signaling in Hodgkin and Primary Mediastinal B-Cell Lymphoma Allows Prediction of Therapeutic Targets.* Cancer Research, 2011. **71**(3): p. 693.

143. Rateitschak, K., et al., *Mathematical modelling of interferon-γ signalling in pancreatic stellate cells reflects and predicts the dynamics of STAT1 pathway activity.* Cellular Signalling, 2010. **22**(1): p. 97-105.

144. Nguyen, K.B., et al., *Critical Role for STAT4 Activation by Type 1 Interferons in the Interferon-γ Response to Viral Infection.* Science, 2002. **297**(5589): p. 2063.

145. Der, S.D., et al., *Identification of genes differentially regulated by interferon α, β, or γ using oligonucleotide arrays.* Proceedings of the National Academy of Sciences, 1998. **95**(26): p. 15623.

146. Pertsovskaya, I., et al., *Transient oscillatory dynamics of interferon beta signaling in macrophages.* BMC Systems Biology, 2013. **7**(1): p. 59.

147. Khalili, M.B., R. Yazdanparast, and A. Nowrouzi, *Induction of transient cell cycle arrest by H2O2 via modulation of ultradian oscillations of Hes1, Socs3, and p-Stat3 in fibroblast cells.* Journal of Cellular Biochemistry, 2018. **119**(2): p. 1453-1462.

148. Li, S. and L. Zhang, *Circadian Control of Global Transcription.* BioMed Research International, 2015. **2015**: p. 8.

149. Bass, J., *Circadian topology of metabolism.* Nature, 2012. **491**: p. 348.

150. Leone, V., et al., *Effects of diurnal variation of gut microbes and high fat feeding on host circadian clock function and metabolism.* Cell host & microbe, 2015. **17**(5): p. 681-689.

151. Dowse, H.B. and J.M. Ringo, *Further evidence that the circadian clock in Drosophila is a population of coupled ultradian oscillators.* Jouranl of Biological Rhythms, 1987. **2**(1): p. 65-76.

152. Barrio, R.A., L. Zhang, and P.K. Maini, *Hierarchically coupled ultradian oscillators generating robust circadian rhythms.* Bulletin of Mathematical Biology, 1997. **59**(3): p. 517-532.

153. Ono, D., K.-i. Honma, and S. Honma, *Circadian and ultradian rhythms of clock gene expression in the suprachiasmatic nucleus of freely moving mice.* Scientific Reports, 2015. **5**(12310).

154. Hussain, M.M., *Regulation of Intestinal Lipid Absorption by Clock Genes.* Annual Review of Nutrition, 2014. **34**(1): p. 357-375.

155. Reinke, H. and G. Asher, *Circadian Clock Control of Liver Metabolic Functions.* Gastroenterology, 2016. **150**(3): p. 574-580.

156. Saha, S. and P. Sassone-Corsi, *Circadian Clock and Breast Cancer: A Molecular Link.* Cell Cycle, 2007. **6**(11): p. 1329-1331.

157. Shi, M. and X. Zheng, *Interactions between the circadian clock and metabolism: there are good times and bad times.* Acta Biochimica et Biophysica Sinica, 2013. **45**(1): p. 61–69.

158. Sahar, S. and P. Sassone-Corsi, *Metabolism and cancer: the circadian clock connection.* Nature Reviews Cancer, 2009. **9**: p. 886.

159. Green, C.B., J.S. Takahashi, and J. Bass, *The Meter of Metabolism.* Cell, 2008. **134**(5): p. 728-742.

160. Fu, L. and C.C. Lee, *The circadian clock: pacemaker and tumour suppressor.* Nature Reviews Cancer, 2003. **3**: p. 350.

161. Abdullah, S., et al., *Circadian Computing: Sensing, Modeling, and Maintaining Biological Rhythms.* 2017. 35-58.

162. Chen, S.-T., et al., *Deregulated expression of the PER1 , PER2 and PER3 genes in breast cancers.* Carcinogenesis, 2005. **26**(7): p. 1241-1246.

163. Flynn-Evans, E.E., et al., *Shiftwork and Prostate-Specific Antigen in the National Health and Nutrition Examination Survey.* JNCI: Journal of the National Cancer Institute, 2013. **105**(17): p. 1292-1297.

164. Shih, M.-C., et al., *Promoter methylation in circadian genes of endometrial cancers detected by methylation-specific PCR.* Molecular Carcinogenesis, 2006. **45**(10): p. 732-740.

165. Mostafaie, N., et al., *Correlated downregulation of estrogen receptor beta and the circadian clock gene Per1 in human colorectal cancer.* Molecular Carcinogenesis, 2009. **48**(7): p. 642-647.

166. Hu, S., et al., *Melatonin and Tryptophan Circadian Profiles in Patients with Advanced Non-Small Cell Lung Cancer.* Vol. 26. 2009. 886-92.

167. Tokunaga, H., et al., *Clinicopathological significance of circadian rhythm-related gene expression levels in patients with epithelial ovarian cancer.* Acta Obstetricia et Gynecologica Scandinavica, 2010. **87**(10): p. 1060-1070.

168. Lin, Y.-M., et al., *Disturbance of circadian gene expression in hepatocellular carcinoma.* Molecular Carcinogenesis, 2008. **47**(12): p. 925-933.

169. Dallmann, R., A. Okyar, and F. Lévi, *Dosing-Time Makes the Poison: Circadian Regulation and Pharmacotherapy.* Trends in Molecular Medicine, 2016. **22**(5): p. 430-445.

170. Jung, C.-H., et al., *Bmal1 suppresses cancer cell invasion by blocking the phosphoinositide 3-kinase-Akt-MMP-2 signaling pathway.* Oncology Reports, 2013. **29**(6): p. 2109-2113.

171. Luo, W. and A. Sehgal, *microRNA-279 acts through the JAK/STAT pathway to regulate circadian behavioral output in Drosophila.* Cell, 2012. **148**(4): p. 765-779.

172. Kaeffer, B., *Mammalian intestinal epithelial cells in primary culture: A mini-review.* In Vitro Cellular & Developmental Biology - Animal, 2002. **38**(3): p. 123-134.

173. Ferraretto, A., et al., *New Methodological Approach to Induce a Differentiation Phenotype in Caco-2 Cells Prior to Post-confluence Stage.* Anticancer Research, 2007. **27**(6B): p. 3919-3925.

174. Sun, H., et al., *The Caco-2 cell monolayer: usefulness and limitations.* Expert Opinion on Drug Metabolism & Toxicology, 2008. **4**(4): p. 395-411.

175. Sambuy, Y., et al., *The Caco-2 cell line as a model of the intestinal barrier: influence of cell and culture-related factors on Caco-2 cell functional characteristics.* Cell Biology and Toxicology, 2005. **21**(1): p. 1-26.

176. Liévin-Le Moal, V. and A.L. Servin, *Pathogenesis of Human Enterovirulent Bacteria: Lessons from Cultured, Fully Differentiated Human Colon Cancer Cell Lines.* Microbiology and Molecular Biology Reviews : MMBR, 2013. **77**(3): p. 380-439.

177. Ferruzza, S., et al., *A protocol for differentiation of human intestinal Caco-2 cells in asymmetric serum-containing medium.* Toxicology in Vitro, 2012. **26**(8): p. 1252-1255.

178. Matsumoto, H., et al., *Biosynthesis of Alkaline Phosphatase During Differentiation of the Human Colon Cancer Cell Line Caco-2.* Gastroenterology, 1990. **98**(5, Part 1): p. 1199-1207.

179. Moore, S.R., et al., *Robust circadian rhythms in organoid cultures from PERIOD2::LUCIFERASE mouse small intestine.* Disease Models &amp; Mechanisms, 2014. **7**(9): p. 1123.

180. Bieler, J., et al., *Robust synchronization of coupled circadian and cell cycle oscillators in single mammalian cells.* Molecular Systems Biology, 2014. **10**(7).

181. Balsalobre, A., F. Damiola, and U. Schibler, *A Serum Shock Induces Circadian Gene Expression in Mammalian Tissue Culture Cells.* Cell, 1998. **93**(6): p. 929-937.

182. Buhr, E.D., S.-H. Yoo, and J.S. Takahashi, *Temperature as a Universal Resetting Cue for Mammalian Circadian Oscillators.* Science, 2010. **330**(6002): p. 379.

183. Tamaru, T., et al., *Synchronization of Circadian Per2 Rhythms and HSF1-BMAL1:CLOCK Interaction in Mouse Fibroblasts after Short-Term Heat Shock Pulse.* PLOS ONE, 2011. **6**(9): p. e24521.

184. Colella, A.D., et al., *Comparison of Stain-Free gels with traditional immunoblot loading control methodology.* Analytical Biochemistry, 2012. **430**(2): p. 108-110.

185. Edwards, R.A., G. Jickling, and R.J. Turner, *The Light-induced Reactions of Tryptophan with Halocompounds¶.* Photochemistry and Photobiology, 2002. **75**(4): p. 362-368.

186. Barclay Johanna, L., et al., *SOCS3 as a tumor suppressor in breast cancer cells, and its regulation by PRL.* International Journal of Cancer, 2009. **124**(8): p. 1756-1766.

187. Cooper, S., *Reappraisal of serum starvation, the restriction point, G0, and G1 phase arrest points.* The FASEB Journal, 2003. **17**(3): p. 333-340.

188. Saksena, S., et al., *Mechanisms underlying modulation of monocarboxylate transporter 1 (MCT1) by somatostatin in human intestinal epithelial cells.* Gastrointestinal and Liver Physiology, 2009. **297**(5): p. G878–G885.

189. Darmoul, D., et al., *Aberrant Expression and Activation of the Thrombin Receptor Protease-Activated Receptor-1 Induces Cell Proliferation and Motility in Human Colon Cancer Cells.* The American Journal of pathology 2003. **162**(5): p. 1503-1513.

190. Stierum, R., et al., *Proteome analysis reveals novel proteins associated with proliferation and differentiation of the colorectal cancer cell line Caco-2.* Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 2003. **1650**(1): p. 73-91.

191. Buhrke, T., I. Lengler, and A. Lampen, *Analysis of proteomic changes induced upon cellular differentiation of the human intestinal cell line Caco-2.* Development, Growth & Differentiation, 2011. **53**(3): p. 411-426.

192. Bailey, C.A., P. Bryla, and A.W. Malick, *The use of the intestinal epithelial cell culture model, Caco-2, in pharmaceutical development.* Advanced Drug Delivery Reviews, 1996. **22**(1): p. 85-103.

193. Lea, M.A., et al., *Induction of Differentiation of Colon Cancer Cells by Combined Inhibition of Kinases and Histone Deacetylase.* Anticancer Research, 2007. **27**(2): p. 741-748.

194. Yang, B., et al., *The Transition from Proliferation to Differentiation in Colorectal Cancer Is Regulated by the Calcium Activated Chloride Channel A1.* PLoS ONE, 2013. **8**(4): p. e60861.

195.   Kitano, H., *Cancer as a robust system: implications for anticancer therapy.* Nature Reviews Cancer, 2004. **4**: p. 227-235.

196.   Rossetti, S., et al., *Entrainment of breast (cancer) epithelial cells detects distinct circadian oscillation patterns for clock and hormone receptor genes.* Cell Cycle, 2012. **11**(2): p. 350-360.

197.   van der Valk, J., et al., *Optimization of chemically defined cell culture media – Replacing fetal bovine serum in mammalian in vitro methods.* Toxicology in Vitro, 2010. **24**(4): p. 1053-1063.

198.   Mbeunkui, F., O. Fodstad, and L.K. Pannell, *Secretory Protein Enrichment and Analysis: An Optimized Approach Applied on Cancer Cell Lines Using 2D LC−MS/MS.* Journal of Proteome Research, 2006. **5**(4): p. 899-906.

199.   Codeluppi, S., et al., *Influence of rat substrain and growth conditions on the characteristics of primary cultures of adult rat spinal cord astrocytes.* Journal of Neuroscience Methods, 2011. **197**(1): p. 118-127.

200.   Perreault, N. and J.-F. Beaulieu, *Use of the Dissociating Enzyme Thermolysin to Generate Viable Human Normal Intestinal Epithelial Cell Cultures.* Experimental Cell Research, 1996. **224**(2): p. 354-364.

201.   Forger, D.B. and C.S. Peskin, *Stochastic simulation of the mammalian circadian clock.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(2): p. 321-324.

202.   Tay, S., *Single-cell NF-κB dynamics reveal digital activation and analogue information processing.* Nature Letters, 2010. **466**: p. 267–271.

203.   Hirata, H., et al., *Oscillatory expression of the bHLH factor Hes1 regulated by a negative feedback loop.* Science, 2002. **298**(5594): p. 840-843.

204.   Longo, D. and J. Hasty, *Dynamics of single-cell gene expression.* Molecular Systems biology, 2006. **2**(64).

205.   Hoffmann, A., et al., *The IκB-NF-κB signaling module: Temporal control and selective gene activation.* Science, 2002. **298**(5596): p. 1241-1245

206.   Lahav, G., et al., *Dynamics of the p53-Mdm2 feedback loop in individual cells.* Nature Genetics, 2004. **36**(2): p. 147-150.

207.   Fletcher, T.C., A. DiGiandomenico, and J. Hawiger, *Extended anti-inflammatory action of a degradation-resistant mutant of cell-penetrating suppressor of cytokine signaling 3.* The Journal of biological chemistry, 2010. **285**(24): p. 18727-18736.

208.   Boyle, K., et al., *Deletion of the SOCS box of suppressor of cytokine signaling 3 (SOCS3) in embryonic stem cells reveals SOCS box-dependent regulation of JAK but not STAT phosphorylation.* Cellular signalling, 2009. **21**(3): p. 394-404.

209. Dai, C.-L., et al., *Inhibition of protein synthesis alters protein degradation through activation of protein kinase B (AKT).* The Journal of biological chemistry, 2013. **288**(33): p. 23875-23883.

210. Laudisi, F., et al., *STAT3 Interactors as Potential Therapeutic Targets for Cancer Treatment.* International journal of molecular sciences, 2018. **19**(6): p. 1787.

211. Arora, M., *Cell Culture Media: A Review.* Materials and Methods, 2013. **3**.

212. Balsalobre, A., L. Marcacci, and U. Schibler, *Multiple signaling pathways elicit circadian gene expression in cultured Rat-1 fibroblasts.* Current biology : CB, 2000. **10**: p. 1291-4.

213. Wiejak, J., et al., *Flavanoids induce expression of the suppressor of cytokine signalling 3 (SOCS3) gene and suppress IL-6-activated signal transducer and activator of transcription 3 (STAT3) activation in vascular endothelial cells.* The Biochemical journal, 2013. **454**(2): p. 283-293.

214. Taylor, S.C., et al., *A Defined Methodology for Reliable Quantification of Western Blot Data.* Molecular Biotechnology, 2013. **55**(3): p. 217-226.

215. Farmer, S.R., et al., *Regulation of actin mRNA levels and translation responds to changes in cell configuration.* Molecular and Cellular Biology, 1983. **3**(2): p. 182-189.

216. Castaño, Z. and R.M. Kypta, *Housekeeping Proteins: Limitations as References During Neuronal Differentiation.* The Open Neuroscience Journal, 2008. **2**: p. 36-40.

217. Khan, S.A., et al., *Cell-type specificity of β-actin expression and its clinicopathological correlation in gastric adenocarcinoma.* World Journal of Gastroenterology, 2014. **20**(34): p. 12202-12211.

218. Böhmer, F.-D. and K. Friedrich, *Protein tyrosine phosphatases as wardens of STAT signaling.* JAK-STAT, 2014. **3**(1): p. e28087.

219. Lu, D., et al., *The phosphatase DUSP2 controls the activity of the transcription activator STAT3 and regulates TH17 differentiation.* Vol. 16. 2015.

220. Liu, Y., et al., *Protein inhibitor of activated STAT 1 (PIAS1) is identified as the SUMO E3 ligase of CCAAT/enhancer-binding protein β (C/EBPβ) during adipogenesis.* Molecular and cellular biology, 2013. **33**(22): p. 4606-4617.

221. Croker, B.A., et al., *SOCS3 negatively regulates IL-6 signaling in vivo.* Nature Immunology, 2003. **4**: p. 540.

222. Lang, R., et al., *SOCS3 regulates the plasticity of gp130 signaling.* Nature Immunology, 2003. **4**: p. 546.

223. Wang, L., et al., *IL-6 Induces NF- B Activation in the Intestinal Epithelia1.* The Journal of Immunology, 2003. **171**(6): p. 3194-3201.

224. Kallal, L.E. and C.A. Biron, *Changing partners at the dance: Variations in STAT concentrations for shaping cytokine function and immune responses to viral infections.* JAK-STAT, 2013. **2**(1): p. e23504.

225. Altschuler, S.J. and L.F. Wu, *Cellular heterogeneity: when do differences make a difference?* Cell, 2010. **141**(4): p. 559-563.

226. Cristinelli, S. and A. Ciuffi, *The use of single-cell RNA-Seq to understand virus–host interactions.* Current Opinion in Virology, 2018. **29**: p. 39-50.

227. Ingram, P.J., M.P.H. Stumpf, and J. Stark, *Nonidentifiability of the Source of Intrinsic Noise in Gene Expression from Single-Burst Data.* PLOS Computational Biology, 2008. **4**(10): p. e1000192.

228. Sherman, M.S., et al., *Cell-to-cell variability in the propensity to transcribe explains correlated fluctuations in gene expression.* Cell systems, 2015. **1**(5): p. 315-325.

229. Lei, X., et al., *Biological Sources of Intrinsic and Extrinsic Noise in cI Expression of Lysogenic Phage Lambda.* Scientific Reports, 2015. **5**: p. 13597.

230. Vlad, E., et al., *Connecting protein and mRNA burst distributions for stochastic models of gene expression.* Physical Biology, 2011. **8**(4): p. 046001.

231. Elowitz, M.B., et al., *Stochastic Gene Expression in a Single Cell.* Science, 2002. **297**(5584): p. 1183.

232. Iadevaia, S., et al., *Identification of Optimal Drug Combinations Targeting Cellular Networks: Integrating Phospho-Proteomics and Computational Network Analysis.* Cancer Research, 2010. **70**(17): p. 6704.

233. Hanahan, D. and R.A. Weinberg, *Hallmarks of Cancer: The Next Generation.* Cell, 2011. **144**(5): p. 646–674.

234. Stelling, J., et al., *Robustness of Cellular Functions.* Cell, 2004. **118**(6): p. 675–685.

235. Yi, T.-M., et al., *Robust perfect adaptation in bacterial chemotaxis through integral feedback control.* PNAS, 2000. **97**(9): p. 4649-53.

236. David, R., *Calibrating the clock.* Nature Reviews Molecular Cell Biology, 2009. **10**: p. 816.

237. Meir, E., et al., *Robustness, Flexibility, and the Role of Lateral Inhibition in the Neurogenic Network.* Current Biology, 2002. **12**(10): p. 778-786.

238. Rao, C.V., D.M. Wolf, and A.P. Arkin, *Control, exploitation and tolerance of intracellular noise.* Nature, 2002. **420**: p. 231.

239. Albert, R., H. Jeong, and A.-L. Barabási, *Error and attack tolerance of complex networks.* Nature, 2000. **406**: p. 378.

240. Turgeon, J.L. and D.W. Waring, *Chapter 20 - Endocrinology of the Single Cell: Tools and Insights A2 - Ulloa-Aguirre, Alfredo*, in *Cellular Endocrinology in Health and Disease*, P.M. Conn, Editor. 2014, Academic Press: Boston. p. 325-340.

241. Iwasaki, S. and N.T. Ingolia, *The Growing Toolbox for Protein Synthesis Studies.* Trends in Biochemical Sciences, 2017. **42**(8): p. 612-624.

242. Ratz, M., et al., *CRISPR/Cas9-mediated endogenous protein tagging for RESOLFT super-resolution microscopy of living human cells.* Scientific Reports, 2015. **5**: p. 9592.

243. Styers, M.L., J. Lowery, and E. Sztul, *Transient Expression of Epitope-Tagged Proteins in Mammalian Cells*, in *Immunoelectron Microscopy: Methods and Protocols*, S.D. Schwartzbach and T. Osafune, Editors. 2010, Humana Press: Totowa, NJ. p. 43-61.

244. Cui, Y., et al., *Using Fluorescent Protein Fusions to Study Protein Subcellular Localization and Dynamics in Plant Cells*, in *High-Resolution Imaging of Cellular Proteins: Methods and Protocols*, S.D. Schwartzbach, O. Skalli, and T. Schikorski, Editors. 2016, Springer New York: New York, NY. p. 113-123.

245. Collings, D.A., *Subcellular Localization of Transiently Expressed Fluorescent Fusion Proteins*, in *Legume Genomics: Methods and Protocols*, R.J. Rose, Editor. 2013, Humana Press: Totowa, NJ. p. 227-258.

246. Rizzo, M.A., M.W. Davidson, and D.W. Piston, *Fluorescent protein tracking and detection: applications using fluorescent proteins in living cells.* Cold Spring Harbor Protocols, 2009. **12**.

247. Gibson, T.J., M. Seiler, and R.A. Veitia, *The transience of transient overexpression.* Nature Methods, 2013. **10**: p. 715.

248. Harder, Z., R. Zunino, and H. McBride, *Sumo1 Conjugates Mitochondrial Substrates and Participates in Mitochondrial Fission.* Current Biology, 2004. **14**(4): p. 340-345.

249. Voeltz, G.K., et al., *A Class of Membrane Proteins Shaping the Tubular Endoplasmic Reticulum.* Cell, 2006. **124**(3): p. 573-586.

250. Birchler, J.A. and R.A. Veitia, *Gene balance hypothesis: Connecting issues of dosage sensitivity across biological disciplines.* Proceedings of the National Academy of Sciences, 2012. **109**(37): p. 14746.

251. Zhang, X., et al., *Epitope tagging of endogenous proteins for genome-wide ChIP-chip studies.* Nature Methods, 2008. **5**: p. 163.

252. Dean, K.M. and A.E. Palmer, *Advances in fluorescence labeling strategies for dynamic cellular imaging.* Nature Chemical Biology, 2014. **10**: p. 512.

253. Uemura, T., et al., *Fluorescent protein tagging of endogenous protein in brain neurons using CRISPR/Cas9-mediated knock-in and in utero electroporation techniques.* Scientific Reports, 2016. **6**: p. 35861.

254. Hsu, P.D., E.S. Lander, and F. Zhang, *Development and Applications of CRISPR-Cas9 for Genome Engineering.* Cell, 2014. **157**(6): p. 1262-1278.

255. Mei, Y., et al., *Recent Progress in CRISPR/Cas9 Technology.* Journal of Genetics and Genomics, 2016. **43**(2): p. 63-75.

256. Sander, J.D. and J.K. Joung, *CRISPR-Cas systems for editing, regulating and targeting genomes.* Nature Biotechnology, 2014. **32**(4): p. 347-355.

257. Barrangou, R. and L.A. Marraffini, *CRISPR-Cas Systems: Prokaryotes Upgrade to Adaptive Immunity.* Molecular Cell Review, 2014. **54**: p. 234-244.

258. Shmakov, S., et al., *Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems.* Molecular Cell Article, 2015. **60**: p. 1-13.

259. Pearson, B.M., et al., *Differential Distribution of Type II CRISPR-Cas Systems in Agricultural and Nonagricultural Campylobacter coli and Campylobacter jejuni Isolates Correlates with Lack of Shared Environments.* Genome Biology and Evolution, 2015. **7**(9): p. 2663-2679.

260. Qiu, Y., et al., *An Active Type I-E CRISPR-Cas System Identified in Streptomyces avermitilis.* PLoS One, 2016. **11**(2).

261. Rath, D., et al., *The CRISPR-Cas immune system: Biology, mechanisms and applications.* Biochimie, 2015. **117**: p. 119-128.

262. Sorek, R., C.M. Lawrence, and B. Wiedenheft, *CRISPR-Mediated Adaptive Immune Systems in Bacteria and Archaea.* Annual Review of Biochemistry, 2013. **82**: p. 237-266.

263. Reeks, J., J.H. Naismith, and M.F. White, *CRISPR interference: a structural perspective.* Biochemical Journal, 2013. **453**(2): p. 155-166.

264. Barrangou, R., *Diversity of CRISPR-Cas immune systems and molecular machines.* Genome Biology, 2015. **16**(247): p. 1-11.

265. Salmon, G.P.C. and P.C. Fineran, *A century of the phage: past, present and future.* Nature Reviews Microbiology, 2015. **13**: p. 777-787.

266. Yang, H., et al., *One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering.* Cell, 2013. **154**(6): p. 1370–1379.

267. Ran, F.A., et al., *Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity.* Cell, 2013. **154**(6): p. 1380-1389.

268. Cong, L., et al., *Multiplex Genome Engineering Using CRISPR/Cas Systems.* Science, 2013. **339**(6121): p. 819-823.

269. Bikard, D., et al., *Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system.* Nucleic Acids Research, 2013. **41**(15): p. 7429-7437.

270. Gilbert, L.A., et al., *CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes.* Cell, 2014. **154**(2): p. 442-451.

271. Xu, T., et al., *Cas9-Based Tools for Targeted Genome Editing and Transcriptional Control.* Applied and Environmental Microbiology, 2014. **80**(5): p. 1544-1552.

272. Zhuchi, T., et al., *CRISPR/Cas9: a powerful genetic engineering tool for establishing large animal models of neurodegenerative diseases.* Vol. 10. 2015. 35.

273. Mali, P., et al., *CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering.* Nature Biotechnology, 2013. **31**: p. 833-838.

274. Maruyama, T., et al., *Increasing the efficiency of precise genome editing with CrIsPr-Cas9 by inhibition of nonhomologous end joining.* Nature Biotechnology, 2015. **33**(5): p. 538-542.

275. Svitashev, S., et al., *Genome editing in maize directed by CRISPR–Cas9 ribonucleoprotein complexes.* Nature Communications, 2016. **7**(13274).

276. Maeder, M.L. and C.A. Gersbach, *Genome-editing Technologies for Gene and Cell Therapy.* Molecular Therapy 2016. **24**(3): p. 430–446.

277. Orthwein, A., et al., *Mitosis Inhibits DNA Double-Strand Break Repair to Guard Against Telomere Fusions.* Science, 2014. **344**(6180): p. 189-193.

278. Dueva, R. and G. Iliakis, *Alternative pathways of non-homologous end joining (NHEJ) in genomic instability and cancer.* Translational Cancer Research, 2013. **2**(3).

279. Jasin, M. and R. Rothstein, *Repair of Strand Breaks by Homologous Recombination.* Cold Spring Harbor Perspectives in Biology, 2013. **5**(11): p. 1-40.

280. Walczak, A., et al., *Evaluation of DNA Double Strand Breaks Repair Efficiency in Head and Neck Cancer.* DNA and Cell Biology, 2012. **31**(3): p. 297-304.

281. Ran, F.A., et al., *Genome engineering using the CRISPR-Cas9 system.* Nature Protocols, 2013. **8**(11).

282. Ciccia, A. and S.J. Elledge, *The DNA Damage Response: Making it safe to play with knives.* Molecular cell, 2010. **40**(2): p. 179-204.

283. Liu, T. and J. Huang, *DNA End Resection: Facts and Mechanisms.* Genomics Proteomics Bioinformatics, 2016. **14**(3): p. 126-130.

284. Cho, S.W., et al., *Heritable Gene Knockout in Caenorhabditis elegans by Direct Injection of Cas9–sgRNA Ribonucleoproteins.* Genetics, 2013. **195**(3): p. 1177-1180.

285. Zhang, C., et al., *Efficient Editing of Malaria Parasite Genome Using the CRISPR/Cas9 System.* mBio - American Society for Microbiology, 2014. **5**(4): p. e01414-14.

286. Ma, Y., et al., *Generation of eGFP and Cre knockin rats by CRISPR/Cas9.* The FEBS Journal, 2014. **281**(17): p. 3779-3790.

287. Kimura, Y., et al., *Efficient generation of knock-in transgenic zebrafish carrying reporter/driver genes by CRISPR/Cas9-mediated genome engineering.* Scientific Reports, 2014. **4**: p. 6545.

288. Zhang, C., et al., *Systematic CRISPR-Cas9-Mediated Modifications of Plasmodium yoelii ApiAP2 Genes Reveal Functional Insights into Parasite Development.* mBio, 2017. **8**(6).

289. Yu, Z., et al., *Various applications of TALEN- and CRISPR/Cas9-mediated homologous recombination to modify the Drosophila genome.* Biology Open, 2014. **3**(4): p. 271.

290. Yu, Z., et al., *Highly Efficient Genome Modifications Mediated by CRISPR/Cas9 in Drosophila.* Genetics, 2013. **195**(1): p. 289-291.

291. Park, A., et al., *CRISPR/Cas9 Allows Efficient and Complete Knock-In of a Destabilization Domain-Tagged Essential Protein in a Human Cell Line, Allowing Rapid Knockdown of Protein Function.* PLOS ONE, 2014. **9**(4): p. e95101.

292. Banaszynski, L.A., et al., *A Rapid, Reversible, and Tunable Method to Regulate Protein Function in Living Cells Using Synthetic Small Molecules.* Cell, 2006. **126**(5): p. 995-1004.

293. Heigwer, F., G. Kerr, and M. Boutros, *E-CRISP: fast CRISPR target site identification.* Nature Methods, 2014. **11**: p. 122.

294. Montague, T.G., et al., *CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing.* Nucleic Acids Research, 2014. **42**(W1): p. W401-W407.

295. Ma, M., et al., *A Guide RNA Sequence Design Platform for the CRISPR/Cas9 System for Model Organism Genomes.* BioMed Research International, 2013. **2013**: p. 4.

296. Wu, X., A.J. Kriz, and P.A. Sharp, *Target specificity of the CRISPR-Cas9 system.* Quantitative biology, 2014. **2**(2): p. 59-70.

297. Hsu, P.D., et al., *DNA targeting specificity of RNA-guided Cas9 nucleases.* Nature Biotechnology, 2013. **31**: p. 827.

298. Fonfara, I., et al., *Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems.* Nucleic Acids Research, 2014. **42**(4): p. 2577-2590.

299. Fu, Y., et al., *Improving CRISPR-Cas nuclease specificity using truncated guide RNAs.* Nature Biotechnology, 2014. **32**: p. 279.

300. Thyme, S.B., et al., *Internal guide RNA interactions interfere with Cas9-mediated cleavage.* Nature Communications, 2016. **7**: p. 11750.

301. Wang, T., et al., *Genetic Screens in Human Cells Using the CRISPR-Cas9 System.* Science, 2014. **343**(6166): p. 80.

302. Tsai, S.Q., et al., *GUIDE-Seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases.* Nature biotechnology, 2015. **33**(2): p. 187-197.

303. Romanienko, P.J., et al., *A Vector with a Single Promoter for In Vitro Transcription and Mammalian Cell Expression of CRISPR gRNAs.* PLoS ONE, 2016. **11**(2): p. e0148362.

304. Yanik, M., et al., *In vivo genome editing as a potential treatment strategy for inherited retinal dystrophies.* Progress in Retinal and Eye Research, 2017. **56**: p. 1-18.

305. Liang, X., et al., *Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA.* Journal of Biotechnology, 2017. **241**: p. 136-146.

306. Thomas, K.R., K.R. Folger, and M.R. Capecchi, *High frequency targeting of genes to specific sites in the mammalian genome.* Cell, 1986. **44**(33): p. 419-28.

307. Shy, B.R., et al., *Co-incident insertion enables high efficiency genome engineering in mouse embryonic stem cells.* Nucleic Acids Research, 2016. **44**(16): p. 7997-8010.

308. Yao, X., et al., *Homology-mediated end joining-based targeted integration using CRISPR/Cas9.* Cell Research, 2017. **27**: p. 801.

309. Zhang, J.-P., et al., *Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage.* Genome Biology, 2017. **18**: p. 35.

310. Klepsch, O., et al., *Intragenic regulation of SOCS3 isoforms.* Cell Communication and Signaling, 2019. **17**(1): p. 70.

311. Ehlting, C., et al., *Regulation of Suppressor of Cytokine Signaling 3 (SOCS3) mRNA Stability by TNF-α Involves Activation of the MKK6/p38<sup>MAPK</sup>/MK2 Cascade.* The Journal of Immunology, 2007. **178**(5): p. 2813.

312. Hölter, K., et al., *Analysis of sequence variations in the suppressor of cytokine signaling (SOCS)-3 gene in extremely obese children and adolescents.* BMC Medical Genetics, 2007. **8**: p. 21-21.

313. Zhang, M.Y., et al., *Methylation profiling of SOCS1, SOCS2, SOCS3, CISH and SHP1 in Philadelphia-negative myeloproliferative neoplasm.* Journal of Cellular and Molecular Medicine, 2013. **17**(10): p. 1282-1290.

314. Gibson, D.G., et al., *Enzymatic assembly of DNA molecules up to several hundred kilobases.* Nature Methods, 2009. **6**: p. 343.

315. Yu, C., et al., *Small Molecules Enhance CRISPR Genome Editing in Pluripotent Stem Cells.* Cell Stem Cell, 2015. **16**(2): p. 142-147.

316. Savic, D., et al., *CETCh-seq: CRISPR epitope tagging ChIP-seq of DNA-binding proteins.* Genome Research, 2015. **25**(10): p. 1581-1589.

317. Mali, P., et al., *RNA-Guided Human Genome Engineering via Cas9.* Science, 2013. **339**(6121): p. 823-826.

318. Heyer, W.-D., K.T. Ehmsen, and J. Liu, *Regulation of homologous recombination in eukaryotes.* Annual review of genetics, 2010. **44**: p. 113-139.

319. Lin, S., et al., *Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery.* eLife, 2014. **3**: p. e04766.

320. Li, G., et al., *Small molecules enhance CRISPR/Cas9-mediated homology-directed genome editing in primary cells.* Scientific reports, 2017. **7**(1): p. 8943-8943.

321. Pinder, J., J. Salsman, and G. Dellaire, *Nuclear domain 'knock-in' screen for the evaluation and identification of small molecule enhancers of CRISPR-based genome editing.* Nucleic Acids Research, 2015. **43**(19): p. 9379-9392.

322. Brinkman, E.K., et al., *Easy quantitative assessment of genome editing by sequence trace decomposition.* Nucleic acids research, 2014. **42**(22): p. e168-e168.

323. Guha, T.K., A. Wai, and G. Hausner, *Programmable Genome Editing Tools and their Regulation for Efficient Genome Engineering.* Computational and Structural Biotechnology Journal, 2017. **15**: p. 146-160.

324. Hafez, M. and G. Hausner, *Homing endonucleases: DNA scissors on a mission.* Genome, 2012. **55**(8): p. 553-569.

325. Carroll, D., *Genome Engineering With Zinc-Finger Nucleases.* Genetics, 2011. **188**(4): p. 773.

326. Joung, J.K. and J.D. Sander, *TALENs: a widely applicable technology for targeted genome editing.* Nature Reviews Molecular Cell Biology, 2012. **14**: p. 49.

327. Mohr, G., et al., *A Targetron System for Gene Targeting in Thermophiles and Its Application in Clostridium thermocellum.* PLOS ONE, 2013. **8**(7): p. e69032.

328. Harrison, M.M., et al., *A CRISPR view of development.* Genes & Development, 2014. **28**(17): p. 1859-1872.

329. Urnov, F.D., et al., *Genome editing with engineered zinc finger nucleases.* Nature Reviews Genetics, 2010. **11**: p. 636.

330. Malzahn, A., L. Lowder, and Y. Qi, *Plant genome editing with TALEN and CRISPR.* Cell & Bioscience, 2017. **7**: p. 21.

331. Bortesi, L. and R. Fischer, *The CRISPR/Cas9 system for plant genome editing and beyond.* Biotechnology Advances, 2015. **33**(1): p. 41-52.

332. Engler, C., R. Kandzia, and S. Marillonnet, *A One Pot, One Step, Precision Cloning Method with High Throughput Capability.* PLoS ONE, 2008. **3**(11): p. e3647.

333. Pattanayak, V., et al., *High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity.* Nature Biotechnology, 2013. **31**: p. 839.

334. Kim, Y., et al., *A library of TAL effector nucleases spanning the human genome.* Nature Biotechnology, 2013. **31**: p. 251.

335. Lamb, B.M., A.C. Mercer, and I.I.I.C.F. Barbas, *Directed evolution of the TALE N-terminal domain for recognition of all 5' bases.* Nucleic Acids Research, 2013. **41**(21): p. 9779-9785.

336. Liang, X., et al., *Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection.* Journal of Biotechnology, 2015. **208**: p. 44-53.

337. Kouranova, E., et al., *CRISPRs for Optimal Targeting: Delivery of CRISPR Components as DNA, RNA, and Protein into Cultured Cells and Single-Cell Embryos.* Human gene therapy, 2016. **27**(6): p. 464-475.

338. Chen, Y., et al., *Using local chromatin structure to improve CRISPR/Cas9 efficiency in zebrafish.* PLOS ONE, 2017. **12**(8): p. e0182528.

339. Wu, X., et al., *Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells.* Nature Biotechnology, 2014. **32**: p. 670.

340. Xu, H., et al., *Sequence determinants of improved CRISPR sgRNA design.* Genome research, 2015. **25**(8): p. 1147-1157.

341. Doench, J.G., et al., *Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9.* Nature Biotechnology, 2016. **34**: p. 184.

342. Isaac, R.S., et al., *Nucleosome breathing and remodeling constrain CRISPR-Cas9 function.* eLife, 2016. **5**: p. e13450.

343. Doench, J.G., et al., *Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation.* Nature biotechnology, 2014. **32**(12): p. 1262-1267.

344. Uusi-Mäkelä, M.I.E., et al., *Chromatin accessibility is associated with CRISPR-Cas9 efficiency in the zebrafish (Danio rerio).* PLOS ONE, 2018. **13**(4): p. e0196238.

345. Chen, X., et al., *Probing the impact of chromatin conformation on genome editing tools.* Nucleic acids research, 2016. **44**(13): p. 6482-6492.

346. Chari, R., et al., *Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach.* Nature Methods, 2015. **12**: p. 823.

347. Wang, H., et al., *One-step generation of mice carrying mutations in multiple genes by CRISPR/cas-mediated genome engineering.* Cell, 2013. **153**(4): p. 910-918.

348. Rojas-Fernandez, A., et al., *Rapid generation of endogenously driven transcriptional reporters in cells through CRISPR/Cas9.* Scientific Reports, 2015. **5**: p. 9811.

349. Dow, L.E., et al., *Inducible in vivo genome editing with CRISPR-Cas9.* Nature Biotechnology, 2015. **33**(4): p. 390-394.

350. Chu, V.T., et al., *Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells.* Nature Biotechnology, 2015. **33**: p. 543.

351. Song, F. and K. Stieger, *Optimizing the DNA Donor Template for Homology-Directed Repair of Double-Strand Breaks.* Molecular therapy. Nucleic acids, 2017. **7**: p. 53-60.

352. Ceasar, S.A., et al., *Insert, remove or replace: A highly advanced genome editing system using CRISPR/Cas9.* Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 2016. **1863**(9): p. 2333-2344.

353. Rico, E., et al., *Inducible high-efficiency CRISPR-Cas9-targeted gene editing and precision base editing in African trypanosomes.* Scientific Reports, 2018. **8**(1): p. 7960.

354. Nguyen, N., M.M.F. Quail, and A.D. Hernday, *An Efficient, Rapid, and Recyclable System for CRISPR-Mediated Genome Editing in Candida albicans.* mSphere, 2017. **2**(2): p. e00149-17.

355. Lam, F.H., D.J. Steger, and E.K. O'Shea, *Chromatin decouples promoter threshold from dynamic range.* Nature, 2008. **453**: p. 246.

356. Pelet, S., et al., *Transient Activation of the HOG MAPK Pathway Regulates Bimodal Gene Expression.* Science, 2011. **332**(6030): p. 732.

357. Colman-Lerner, A., et al., *Regulated cell-to-cell variation in a cell-fate decision system.* Nature, 2005. **437**: p. 699.

358. Zimmer, M., *Green Fluorescent Protein (GFP): Applications, Structure, and Related Photophysical Behavior.* Chemical Reviews, 2002. **102**(3): p. 759-782.

359. Aymoz, D., et al., *Real-time quantification of protein expression at the single-cell level via dynamic protein synthesis translocation reporters.* Nature Communications, 2016. **7**: p. 11304.

360. Skotheim, J.M., et al., *Positive feedback of G1 cyclins ensures coherent cell cycle entry.* Nature, 2008. **454**: p. 291.

361. Hu, J., et al., *Direct activation of human and mouse Oct4 genes using engineered TALE and Cas9 transcription factors.* Nucleic Acids Research, 2014. **42**(7): p. 4375-4390.

362. Zheng, Q., et al., *Precise gene deletion and replacement using the CRISPR/Cas9 system in human cells.* BioTechniques, 2014. **57**(3): p. 115-124.

363. Böttcher, R., et al., *Efficient chromosomal gene modification with CRISPR/cas9 and PCR-based homologous recombination donors in cultured Drosophila cells.* Nucleic Acids Research, 2014. **42**(11): p. e89-e89.

364. Friedland, A.E., et al., *Heritable genome editing in C. elegans via a CRISPR-Cas9 system.* Nature Methods, 2013. **10**: p. 741.

365. Katic, I. and H. Großhans, *Targeted Heritable Mutation and Gene Conversion by Cas9-CRISPR in <em>Caenorhabditis elegans</em&gt.* Genetics, 2013. **195**(3): p. 1173.

366. Brandman, O. and T. Meyer, *Feedback loops shape cellular signals in space and time.* Science (New York, N.Y.), 2008. **322**(5900): p. 390-395.

367. Ferrell, J.E., *Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability.* Current Opinion in Cell Biology, 2002. **14**(2): p. 140-148.

368. Dougherty, E.R. and I. Shmulevich, *On the limitations of biological knowledge.* Current genomics, 2012. **13**(7): p. 574-587.

369. Nguyen, L.K., *Regulation of oscillation dynamics in biochemical systems with dual negative feedback loops.* Journal of the Royal Society, Interface, 2012. **9**(73): p. 1998-2010.

370. Hoffmann, A. and D. Baltimore, *Circuitry of nuclear factor κB signaling.* Immunological Reviews, 2006. **210**(1): p. 171-186.

371. Jiang, M., et al., *Dysregulation of SOCS-Mediated Negative Feedback of Cytokine Signaling in Carcinogenesis and Its Significance in Cancer Treatment.* Frontiers in immunology, 2017. **8**: p. 70-70.

372. Qin, H., et al., *Expression and functional significance of SOCS-1 and SOCS-3 in astrocytes.* Journal of immunology (Baltimore, Md. : 1950), 2008. **181**(5): p. 3167-3176.

373. Monk, N.A.M., *Oscillatory Expression of Hes1, p53, and NF-κB Driven by Transcriptional Time Delays.* Current Biology, 2003. **13**(16): p. 1409-1413.

374. Hirata, H., et al., *Instability of Hes7 protein is crucial for the somite segmentation clock.* Nature Genetics, 2004. **36**: p. 750.

375. Claydon, A.J. and R. Beynon, *Proteome dynamics: revisiting turnover with a global perspective.* Molecular & cellular proteomics : MCP, 2012. **11**(12): p. 1551-1565.

376. Pratt, J.M., et al., *Dynamics of Protein Turnover, a Missing Dimension in Proteomics.* Molecular &amp; Cellular Proteomics, 2002. **1**(8): p. 579.

377. Aunan, J.R., W.C. Cho, and K. Søreide, *The Biology of Aging and Cancer: A Brief Overview of Shared and Divergent Molecular Hallmarks.* Aging and disease, 2017. **8**(5): p. 628-642.

378. Martin-Perez, M. and J. Villén, *Determinants and Regulation of Protein Turnover in Yeast.* Cell systems, 2017. **5**(3): p. 283-294.e5.

379. Shaner, N.C., P.A. Steinbach, and R.Y. Tsien, *A guide to choosing fluorescent proteins.* Nature Methods, 2005. **2**: p. 905.

380. Miyawaki, A., T. Nagai, and H. Mizuno, *Mechanisms of protein fluorophore formation and engineering.* Current Opinion in Chemical Biology, 2003. **7**(5): p. 557-562.

381. Lippincott-Schwartz, J. and G.H. Patterson, *Photoactivatable fluorescent proteins for diffraction-limited and super-resolution imaging.* Trends in cell biology, 2009. **19**(11): p. 555-565.

382. Subach, F.V., et al., *Monomeric fluorescent timers that change color from blue to red report on cellular trafficking.* Nature Chemical Biology, 2009. **5**: p. 118.

383. Terskikh, A., et al., *"Fluorescent Timer": Protein That Changes Color with Time.* Science, 2000. **290**(5496): p. 1585.

384. Tsuboi, T., et al., *Age-dependent preferential dense-core vesicle exocytosis in neuroendocrine cells revealed by newly developed monomeric fluorescent timer protein.* Molecular biology of the cell, 2010. **21**(1): p. 87-94.

385. Khmelinskii, A., et al., *Tandem fluorescent protein timers for in vivo analysis of protein dynamics.* Nature Biotechnology, 2012. **30**: p. 708.

386. Khmelinskii, A. and M. Knop, *Analysis of Protein Dynamics with Tandem Fluorescent Protein Timers*, in *Exocytosis and Endocytosis*, A.I. Ivanov, Editor. 2014, Springer New York: New York, NY. p. 195-210.

387. Barry, J.D., et al., *TimerQuant: a modelling approach to tandem fluorescent timer design and data interpretation for measuring protein turnover in embryos.* Development (Cambridge, England), 2016. **143**(1): p. 174-179.

388. Pédelacq, J.-D., et al., *Engineering and characterization of a superfolder green fluorescent protein.* Nature Biotechnology, 2005. **24**: p. 79.

389. Shaner, N.C., et al., *Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein.* Nature Biotechnology, 2004. **22**: p. 1567.

390. Khmelinskii, A., et al., *Incomplete proteasomal degradation of green fluorescent proteins in the context of tandem fluorescent protein timers.* Molecular biology of the cell, 2016. **27**(2): p. 360-370.

391. Duncan, R.R., et al., *Functional and spatial segregation of secretory vesicle pools according to vesicle age.* Nature, 2003. **422**: p. 176.

392. Wiegand, U., et al., *Red, yellow, green go!–A novel tool for microscopic segregation of secretory vesicle pools according to their age.* Vol. 31. 2003. 851-6.

393. Donà, E., et al., *Directional tissue migration through a self-generated chemokine gradient.* Nature, 2013. **503**: p. 285.

394. Takashima, Y., et al., *Intronic delay is essential for oscillatory expression in the segmentation clock.* Proceedings of the National Academy of Sciences, 2011. **108**(8): p. 3300.

395. Lu, C., W.E. Bentley, and G. Rao, *A High-Throughput Approach to Promoter Study Using Green Fluorescent Protein.* Biotechnology Progress, 2004. **20**(6): p. 1634-1640.

396. Bielefeld, P., et al., *Transcription factor oscillations in neural stem cells: Implications for accurate control of gene expression AU* Neurogenesis, 2017. **4**(1): p. e1262934.

397. Gerlach, R.G., et al., *Rapid engineering of bacterial reporter gene fusions by using Red recombination.* Applied and environmental microbiology, 2007. **73**(13): p. 4234-4242.

398. Leclerc, G.M., et al., *Development of a Destabilized Firefly Luciferase Enzyme for Measurement of Gene Expression.* BioTechniques, 2000. **29**(3): p. 590-601.

399. Kitsera, N., A. Khobta, and B. Epe, *Destabilized green fluorescent protein detects rapid removal of transcription blocks after genotoxic exposure.* BioTechniques, 2007. **43**(2): p. 222-227.

400. Krebs, D.L. and D.J. Hilton, *SOCS Proteins: Negative Regulators of Cytokine Signaling.* STEM CELLS, 2001. **19**(5): p. 378-387.

401. Auernhammer, C.J., C. Bousquet, and S. Melmed, *Autoregulation of pituitary corticotroph SOCS-3 expression: characterization of the murine SOCS-3 promoter.* Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(12): p. 6964-6969.

402. Matthews, J., et al., *Estrogen receptor-α regulates SOCS-3 expression in human breast cancer cells.* Biochemical and Biophysical Research Communications, 2005. **335**(1): p. 168-174.

403. Barclay, J.L., et al., *Characterization of the SOCS3 Promoter Response to Prostaglandin E2 in T47D Cells.* Molecular Endocrinology, 2007. **21**(10): p. 2516-2528.

404. Invitrogen. *Flp-In System: For Generating Constitutive Expression Cell Lines.* 2010; Available from: https://www.thermofisher.com/uk/en/home/references/protocols/proteins-expression-isolation-and-analysis/protein-expression-protocol/flp-in-system-for-generating-constitutive-expression-cell-lines.html#prot5.

405. Jun-ichi, M., et al., *Expression vector system based on the chicken β-actin promoter directs efficient production of interleukin-5.* Gene, 1989. **79**(2): p. 269-277.

406. Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis.* Nature methods, 2012. **9**(7): p. 676-682.

407. Doerr, A., *(Protein) timing is everything.* Nature Methods, 2012. **9**: p. 783.

408. Inagaki-Ohara, K., et al., *SOCS, inflammation, and cancer.* JAK-STAT, 2013. **2**(3): p. e24053-e24053.

409. Guo, N. and Z. Peng, *MG132, a proteasome inhibitor, induces apoptosis in tumor cells.* Asia-Pacific Journal of Clinical Oncology, 2013. **9**(1): p. 6-11.

410. Römisch, K., *Endoplasmic REeticulum–Associated Degradation.* Annual Review of Cell and Developmental Biology, 2005. **21**(1): p. 435-456.

411. Al-Bari, M.A.A., *Targeting endosomal acidification by chloroquine analogs as a promising strategy for the treatment of emerging viral diseases.* Pharmacology research & perspectives, 2017. **5**(1): p. e00293-e00293.

412. Levine, B. and G. Kroemer, *Autophagy in the pathogenesis of disease.* Cell, 2008. **132**(1): p. 27-42.

413. Sarkar, S., et al., *Rapamycin and mTOR-independent autophagy inducers ameliorate toxicity of polyglutamine-expanded huntingtin and related proteinopathies.* Cell Death And Differentiation, 2008. **16**: p. 46.

414. Tasaki, T., et al., *The N-end rule pathway.* Annual review of biochemistry, 2012. **81**: p. 261-289.

415. Manna, P.R. and D.M. Stocco, *The Role of Specific Mitogen-Activated Protein Kinase Signaling Cascades in the Regulation of Steroidogenesis.* Journal of Signal Transduction, 2011. **2011**: p. 13.

416. Nakamura, O., et al., *The combination of rapamycin and MAPK inhibitors enhances the growth inhibitory effect on Nara-H cells.* International journal of molecular medicine, 2014. **33**(6): p. 1491-1497.

417. Wang, J., et al., *A non-canonical MEK/ERK signaling pathway regulates autophagy via regulating Beclin 1.* The Journal of biological chemistry, 2009. **284**(32): p. 21412-21424.

418. Favata, M., et al., *Identification of a Novel Inhibitor of Mitogen-activated Protein Kinase Kinase.* Vol. 273. 1998. 18623-32.

419. Yokota, S.-i., et al., *Induction of suppressor of cytokine signaling-3 by herpes simplex virus type 1 contributes to inhibition of the interferon signaling pathway.* Journal of virology, 2004. **78**(12): p. 6282-6286.

420. He, B., et al., *Cloning and characterization of a functional promoter of the human SOCS-3 gene.* Biochemical and Biophysical Research Communications, 2003. **301**(2): p. 386-391.

421. Kholodenko, B.N., *Cell-signalling dynamics in time and space.* Nature reviews. Molecular cell biology, 2006. **7**(3): p. 165-176.

422. Feng, S., et al., *Improved split fluorescent proteins for endogenous protein labeling.* Nature communications, 2017. **8**(1): p. 370-370.

423. Cherney, M.M. and B.E. Bowler, *Protein dynamics and function: Making new strides with an old warhorse, the alkaline conformational transition of cytochrome c.* Coordination Chemistry Reviews, 2011. **255**(7): p. 664-677.

424. Burry, R.W., *Controls for Immunocytochemistry: An Update.* Journal of Histochemistry and Cytochemistry, 2011. **59**(1): p. 6-12.

425. Baetz, A., et al., *Identification of a nuclear localization signal in suppressor of cytokine signaling 1.* The FASEB Journal, 2008. **22**(12): p. 4296-4305.

426. Lee, K.-H., et al., *Increased cytoplasmic levels of CIS, SOCS1, SOCS2, or SOCS3 are required for nuclear translocation.* FEBS Letters, 2008. **582**(15): p. 2319-2324.

427. Mizushima, N., T. Yoshimori, and B. Levine, *Methods in mammalian autophagy research.* Cell, 2010. **140**(3): p. 313-326.

428. Haller, M., et al., *Ubiquitination and proteasomal degradation of ATG12 regulates its proapoptotic activity.* Autophagy, 2015. **10**(12): p. 2269-2278.

429. White, C.A. and N.A. Nicola, *SOCS3: An essential physiological inhibitor of signaling by interleukin-6 and G-CSF family cytokines.* JAK-STAT, 2013. **2**(4): p. e25045-e25045.

430. Sacchetti, A., et al., *Efficient GFP mutations profoundly affect mRNA transcription and translation rates.* FEBS Letters, 2001. **492**(1): p. 151-155.

431. Nordholt, N., et al., *Effects of growth rate and promoter activity on single-cell protein expression.* Scientific Reports, 2017. **7**(1): p. 6299.

432. Yan, B., et al., *STAT3 association with microtubules and its activation are independent of HDAC6 activity.* DNA and cell biology, 2015. **34**(4): p. 290-295.

433. Sumit, M., S. Takayama, and J.J. Linderman, *New insights into mammalian signaling pathways using microfluidic pulsatile inputs and mathematical modeling.* Integrative biology : quantitative biosciences from nano to macro, 2017. **9**(1): p. 6-21.

434. Grivennikov, S.I., F.R. Greten, and M. Karin, *Immunity, inflammation, and cancer.* Cell, 2010. **140**(6): p. 883-899.

435. Chen, X., et al., *Immune Cell Types and Secreted Factors Contributing to Inflammation-to-Cancer Transition and Immune Therapy Response.* Cell Reports, 2019. **26**(7): p. 1965-1977.e4.

436. Shalapour, S. and M. Karin, *Immunity, inflammation, and cancer: an eternal fight between good and evil.* The Journal of Clinical Investigation, 2015. **125**(9): p. 3347-3355.

437. Coleman, O.I. and D. Haller, *Bacterial Signaling at the Intestinal Epithelial Interface in Inflammation and Cancer.* Frontiers in Immunology, 2017. **8**: p. 1927.

438. Pencik, J., et al., *JAK-STAT signaling in cancer: From cytokines to non-coding genome.* Cytokine, 2016. **87**: p. 26-36.

439. Inagaki-Ohara, K., et al., *Enhancement of leptin receptor signaling by SOCS3 deficiency induces development of gastric tumors in mice.* Oncogene, 2012. **33**: p. 74.

440. Yoshida, T., et al., *SOCS1 is a suppressor of liver fibrosis and hepatitis-induced carcinogenesis.* The Journal of experimental medicine, 2004. **199**(12): p. 1701-1707.

441. Imayoshi, I., et al., *Oscillatory Control of Factors Determining Multipotency and Fate in Mouse Neural Progenitors.* Science, 2013. **342**(6163): p. 1203.

442. Kobayashi, T., et al., *The cyclic gene Hes1 contributes to diverse differentiation responses of embryonic stem cells.* Vol. 23. 2009. 1870-5.

443. Kageyama, R., et al., *Oscillatory gene expression and somitogenesis.* Wiley Interdisciplinary Reviews: Developmental Biology, 2012. **1**(5): p. 629-641.

444. Tan, S.-L., et al., *MicroRNA9 regulates neural stem cell differentiation by controlling Hes1 expression dynamics in the developing brain.* Genes to Cells, 2012. **17**(12): p. 952-961.

445. Bonev, B., P. Stanley, and N. Papalopulu, *MicroRNA-9 Modulates Hes1 Ultradian Oscillations by Forming a Double-Negative Feedback Loop.* Cell Reports, 2012. **2**(1): p. 10-18.

446. Raj, A. and A. van Oudenaarden, *Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences.* Cell, 2008. **135**(2): p. 216-226.

447. Cathomen, T. and J. Keith Joung, *Zinc-finger Nucleases: The Next Generation Emerges.* Molecular Therapy, 2008. **16**(7): p. 1200-1207.

448. Wright, Addison V., James K. Nuñez, and Jennifer A. Doudna, *Biology and Applications of CRISPR Systems: Harnessing Nature’s Toolbox for Genome Engineering.* Cell, 2016. **164**(1): p. 29-44.

449. Cheol-Min, G., et al., *The art of reporter proteins in science: past, present and future applications.* BMB Rep., 2010. **43**(7): p. 451-460.