Study of Drug-Resistant Colon Cancer Cells

Jack Morton

School of Medicine

Bio-discovery Institute 3

Supervisors: Dr Paloma Ordóñaz Morán (1st) and Prof Kevin Gaston (2nd)

1 Disclaimer

At the start of 2023, part way through my MRes degree, my father took ill with an aggressive Sino-Nasal Undifferentiated Carcinoma (SNUC). This resulted in an incredibly difficult time for myself and my family, a time of uncertainty that would only worsen with my mother's diagnosis of Multiple Sclerosis (MS). The treatment pathway devised for my father entailed 4 rounds of, as the doctors described it, an extreme chemotherapy cocktail of TPF, followed by a 7-week program of high dose radiotherapy for which he would need to be present for every weekday. With this in mind, I decided to return home for the duration of my studies to be there to support my family during this incredibly tyring time. In order to accommodate this change of circumstance, we had to dramatically change the layout of my MRes to what was originally planned. Because I wasn't able to physically be in the lab, the data I had at my disposal was all that I had generated up until Christmas. This would mean a lack of replicated results to work with as well as some missing data points, as will be explained later in this paper. Additional lab work that was planned, such as the use of 3D models and primary cell types, would no longer be possible and an alternative route had to be devised. The following paper was completed in a time of incredible emotional and physical hardship to myself, with many sections being completed in the cancer centre reception when I had taken my father to radiotherapy appointments to allow my mother the break she needed to deal with her own diagnosis. I would like to take this opportunity to give my greatest gratitude to my supervisor, Dr Paloma Ordóñez Morán, who's support throughout all of this made it possible. Her commitment to adjusting my project to enable me to continue my studies from home, as well as her constant supervision and assistance, has allowed me to complete this project under very difficult circumstances. The kindness and care displayed by her was invaluable and of great comfort to both myself and my family. I would also like to thank both my parents whose belief in me and support, even throughout their own hardships, kept me confident in my abilities to complete this project. And lastly, but by no means least, I would like to thank all my lab colleagues and the wider team for the support and help offered, which enabled me to complete any last lab work required. Thank you all.

2 Abstract

With CRC maintaining to be the 4th most common cancer worldwide, with large portions of mortality being contributed to treatment resistant tumours and patient relapse, more knowledge into the biology of this disease is required. With the intestine being a rapidly renewing tissue, stem cells are located at the crypt base while progenitor and differentiated cells are localized at the upper part of the crypt. Stem cells are responsible for damage repair and maintain cells mass, a reasonable assumption is that cancers would be able to hijack this for their own accord. Previous studies have highlighted the ability for a subsect of cancer cells, known as cancer stem cells, to harbour a drug resistance phenotype as well as being able to persist and regrow after treatment end. To better understand the role that these cancer stem cells may play in treatment response, this research focused on the gene expression differences in response to treatment administration between differentiated and undifferentiated cell types. Caco-2 cells lines were expanded in lower confluence to generate a stable and reliable cell line with fewer differential characteristics. qPCR was perfumed on these cells cDNA as well as on the ones grown as stranded to analyse the effect of chemotherapy on cell lines with alternative levels of differentiation. This study was complimented with Bioinformatic and RNA sequencing analyses of these cell lines when treated with drugs to better understand the full scope of changing gene expressions. Finally, data generated in this project was compared with previous data sets of colon cancer cells treated with chemotherapeutic agents to study how other research differs or enforces my own. Overall, we aimed to gain insights into the genetic pathways harboured by the cancer stem cells that present drug resistance in order to better understand what is seen in clinic and provide a target in increasing chemotherapy efficacy.

3 Table of Contents

Table of Contents

4 Introduction

4.1 Anatomy and functions of the large intestine

The large intestine is a member of the digestive tract along with the mouth, oesophagus, stomach, small intestine and rectum. It is on average 5 feet long making it around one fifth of the total gastrointestinal tract length[1]. The small intestine is much longer in length but has a shorter lumen. It is further different due to the presence of omental appendices, haustra and teniae coli[2]. The large intestine is comprised of 5 distinct parts: the cecum, ascending colon, transverse colon, descending colon and sigmoid colon[1]. Most of the important nutrients are absorbed by the small intestine leaving the large intestine responsible for water and some nutrient absorption, vitamin absorption, faeces compaction and moving waste material towards the rectum[2]. By the time the indigestible material reaches the colon, almost all the nutrients and around 90% of water has been absorbed. The ascending colon will then absorb the rest of the water and compact and solidify the indigestible material into stool. The large intestine has many mechanisms to aid it in these responsibilities[1].

Motility

The intestinal wall is multi layered containing the mucosa, submucosa, muscular layer and the serosa. The muscular layer is constructed of 2 layers, an inner circular layer and an outer longitudinal layer. The layers of smooth muscle aid in the motility of the large intestine. There is seen to be 2 types of motilities in the large intestine, haustral contraction and mass movement. Haustal contraction, which is responsible for the large intestines segmented appearance, is activated by the presence of chyme (indigestible food material) and serves to move the food slowly along the large intestine and mix the chyme to aid in water absorption. Mass movements are much faster and enable quick movement of the chyme to the rectum[1].

Water and Electrolyte Absorption

Water absorption occurs via osmosis. This happens when water diffuses across an osmotic gradient generated by the absorption of electrolytes. Sodium is actively absorbed via sodium channels and potassium can be absorbed or secreted depending on lumen concentration[1].

Production and Absorption of Vitamins

The colon contains trillions of bacteria which are essential for protection of our gut and vitamin production. They produce these vitamins via fermentation. Vitamin K and B are produced by these colonic bacteria. These vitamins are then absorbed into the blood[1].

4.2 Intestinal Stem cells

Due to the nature of the intestinal environment, this organ represents one of the most highly renewing tissue in an adult mammal. It is in constant flux between mesenchymal to progenitor proliferation, lineage commitment and ultimately cell death and renewal. Recently it has been identified that a certain cell population noted as 'Stem Cells' fuel this process of cell renewal in most organs. The anatomy of the epithelium of the intestinal crypts makes it one of the most accessible models to study this stem cell biology[3]. As stated previously the intestines play an important role in vitamin and mineral absorption which are vital to life[2]. The epithelium of the intestine consists of 2 primary components: the proliferative crypts of Lieberkühn and long projects known as villi. The intestinal stems cells are seen to reside at the crypts in the intestine and constantly divide to maintain the high cell turnover rate brought on by the intestinal environment. These cells can give rise to either more stem cells, to continue the cycle of self-renewal, or progenitor cells that enters the trans amplifying (TA) compartment to rapidly divide before a terminal differentiation into a specific cell[4]. These progenitor cells themselves fall into two broad categories, they can be either absorptive or secretory. Absorptive progenitors are known to differentiate into the most abundant cell type of the intestinal epithelium, enterocytes. Alternatively, secretory progenitors can differentiate into Paneth cells, Enteroendocrine cells and Goblet cells. These differentiated cells arise in the TA zone and then begin to move up the villi. Once they reach the tip of the villus, they die, are shed into the intestinal environment and

replaced. However, Paneth cells are they only cell type that will move back down towards the stem cell zone (**Figure 1**)[5].

Whilst it is generally agreed that in each intestinal crypt there are four to six independent stem cells, there are still 2 disputed schools of thought on their exact identity. The first, known as the '+4 position' states that these stem cells reside in the position of +4 relative to the crypt bottom[6]. Experimental data has shown these +4 cells to be radiation sensitive which is thought to be a characteristic of stem cells aimed at protecting from genetic damage. In this model it is thought that damaged stem cells are replaced by TA cells which contain a more effective repair capacity. These then fall back into the +4 position and regain their stem cell properties[7]. The second school of thought, known as the 'Stem Cell Zone' is based on the identification of crypt base column cells (CBC). These CBC cells are small, undifferentiated, cycling cells that reside between then Paneth cells. In this model it is thought that the cells within the +4 position are offspring of these CBC cells[8]. Due to the lack of effective stem cell markers, it has been difficult to prove which of these arguments is the most accurate representation of the stem cell landscape within the crypt[3].

In this next session I will discuss the distinct functional and cellular properties of intestinal stem cells and explain the DNA markers associated with them for identification and finally, the signalling pathways that control the fate of these cells.

4.2.1 Functional Properties

Adult stem cells can be defined as having 2 distinct functional features. Firstly, the stem cell population needs to be maintained over long periods of time (longevity). Secondly, these cells populations need to have the capacity to produce additional populations of differentiated cells and tissues. Most stem cells are said to be 'multipotent', meaning they can generate multiple different cell types, yet some are only able to generate one specific type of differentiated cell. In addition to these factors, stem cells are often seen to have 2 other functional properties.

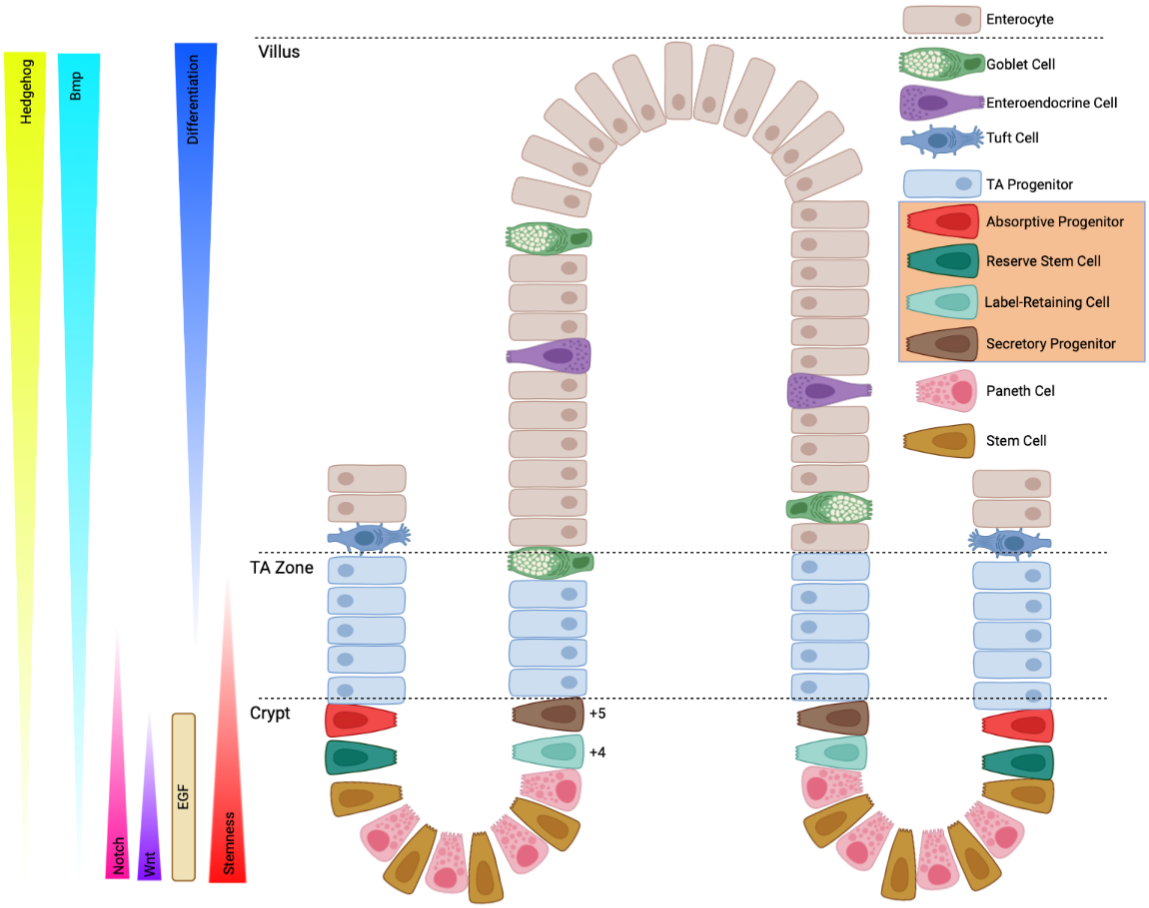


Figure 1: Cycling stem cells are located at the bottom of the intestinal crypts. Differentiated cells are seen to move towards the tip of the villus. Progenitor cells located above the stem cell crypt are highly plastic and can be called upon to fulfil many roles. Even though they are highly differentiated, Paneth cells are in the intestinal crypt and contribute to the intestinal stem cell niche. Several signalling pathways such as Notch also contribute to the fate of ISC [5]. (Adapted from citation)

Stem cells are often found to divide infrequently (quiescence). When they do divide, they produce one rapidly cycling daughter cell and another daughter cell that replaces the parental stem cell to thus not reduce the effective stem cell population (asymmetric division). These rapidly cycling daughter cells are the TA cells described above and have the responsibility for the building and maintaining of the tissue mass. These TA cells undergo a set number of division cycles before terminally differentiating[3].

4.2.2 Cellular Properties

Although the signalling pathways have been extensively studied for ISC, their metabolic state has not. It is believed that the metabolism of ISC plays an important role in the maintenance and homeostasis of these cells[5]. One study showed that calorie restriction inhibited mTORC1 signalling in the neighbouring Paneth cells of the intestinal crypts but not

in the ISC. This promoted stem cells self-renewal and proliferation[9]. Short term fasting was also shown to enhance the function of the ISC by inducing the nuclear receptor peroxisome proliferator activated receptor (Ppar) mediated fatty acid oxidation program[10]. However, high fat diet induced obesity was shown to activate this signalling program in specifically ISC to promote LGR5+ regenerative capacity and increase the cellular numbers. This enhances the ability to form tumours with APC loss[11]. Ppar programs have been seen to activate the Wnt/β-catenin signalling in bone tissue so may be amplifying the Wnt signalling in ISC[12]. Additionally, metabolomic analysis of adult ISC and Paneth cells has shown a diversity in their metabolic landscape. Paneth cells were shown to have an increase in glycolysis, whereas ISC showed an increase of mitochondrial oxidation. This study also showed that Paneth cells were adapted to produce lactate which is in turn transformed into pyruvate in ISC fueling the mitochondrial oxidative phosphorylation (OXPHOS) which is responsible for the promotion of ISC proliferation[5].

In addition to the altered metabolic landscape of ISC, another important cellular characteristic is their cellular plasticity and ability to regenerate when injury occurs[5,13]. Epithelial plasticity has been reported in many studies of injury induced regeneration. It involved the intestinal epithelium dedifferentiating into a more immature state upon injury[13]. A common observation amongst these studies is that the plastic event occurs at the +4 position were the ISC are located[5]. Many findings all agree that stemness in the intestine is not hardwired, and the +4 progenitor cells are highly plastic and regain ISC identity in response of an injury. Additionally, studies have shown that Paneth cells gain plasticity upon irradiation induced Notch activation[14]. This demonstrates that the plasticity of the intestinal epithelium is not restricted to the early progeny[5].

4.2.3 Stem Cell Signalling Pathways

Multiple signalling pathways are thought to orchestrate the stem cell to daughter cell transition in the intestinal epithelium.

Wnt Signalling

The Wnt signalling gradient from the bottom of the crypt is pivotal for the proliferation and maintaining ISC[15]. Wnt ligands are secreted by both epithelial and stromal cells[16] (**Figure 2**). Epithelial Paneth cells secrete Wnt3 over short distances for ISC maintenance[17]. FOX11+ and CD34+ stromal mesenchymal cells also secrete stromal derived Wnt ligands[5,18]. Blocking Wnt ligands causes a rapid cessation of ISC proliferation followed by loss of the crypt cell bulk illustrating the pivotal role that Wnt plays in the survival of ISC[5]. Stromal cells are also seen to secrete R-Spondins which is a known Wnt agonist[19]. This enhances Wnt activity by binding to Lgr family receptors[5].

EGF Signalling

Epidermal growth factor (EGF) is an essential factor required for ISC proliferation. It is shown to be secreted by the Paneth cells and is a driver of Leucine-rich repeat-containing G-protein coupled receptor 5+ (LGR5+) stem cells proliferation in mouse intestinal organoid models[20,21]. Inhibition of EGF signalling was discovered to induced quiescence's in LGR5+ ISC whilst maintaining Wnt signalling. Restoration is shown to revert the ISC back to its normal cycling state[20]. This demonstrates that whilst Wnt defines the identity of ISC, both Wnt and EGF are required for stem cells proliferation[5].

Notch Signalling

Notch signalling plays a valuable role in the maintenance of the ISC pool. Notch inhibition is shown to decrease the number of LGR5+ cell numbers and the proliferation of these cells. Many studies have shown that Paneth cells are responsible for releasing DII/4 ligands (**Figure 2**) that activates Notch signalling via histone deacetylase (HDAC) in ISCs[22]. Transgenic activation of HDAC 1 increases the proliferation of ISC whereas deletion of HDAC 1 and 2 induces stem cells loss and impairs tissue regeneration after radiation wounding. This is also seen in the removal of Notch ligands[5].

Bmp signalling

Bone morphogenic protein (Bmp) plays an important role in intestinal epithelium differentiation in the post mitotic villi by forming an opposing gradient to Wnt signalling. This is formed by ligand secretion from mesenchymal cells[23]. It has been reported that Bmp signalling represses ISC stemness by its downstream effector Smad4 that in turn repressed

the expression of LGR5 and other core stem cell signature genes[24]. The opposing gradients of Bmp and Wnt are important to the maintenance of ISC homeostasis[5].

Hedgehog Signalling

There are 2 hedgehog ligands, Sonic (Shh) and Indian (Ihh) which are expressed in the intestinal epithelium. The receptors (Patch1/2) are expressed on mesenchymal cells[25]. In the adult intestine the epithelial will release hedgehog ligands which in turn regulate the secretion of Bmp[26]. Reduced hedgehog signalling is shown to induce crypt hyperproliferation and reduced differentiation[5].

4.2.4 Epigenetic regulation on ISC

As explained above, transcriptional regulation of ISC is important in controlling their fate. However, recent evidence has suggested that epigenetic regulation is coordinated with this transcriptional regulation in order to fulfil the cellular state[5,27]. Histone acetylation and methylation, both important epigenic events in other cellular processes, are shown to regulate the chromatin conformational change to aid or negate transcriptional activity. This in turn effects cellular processes such as proliferation and differentiation[27]. An example of this is the loss of Hdac1 and Hdac2 which in turn impairs the proliferation and stem cell

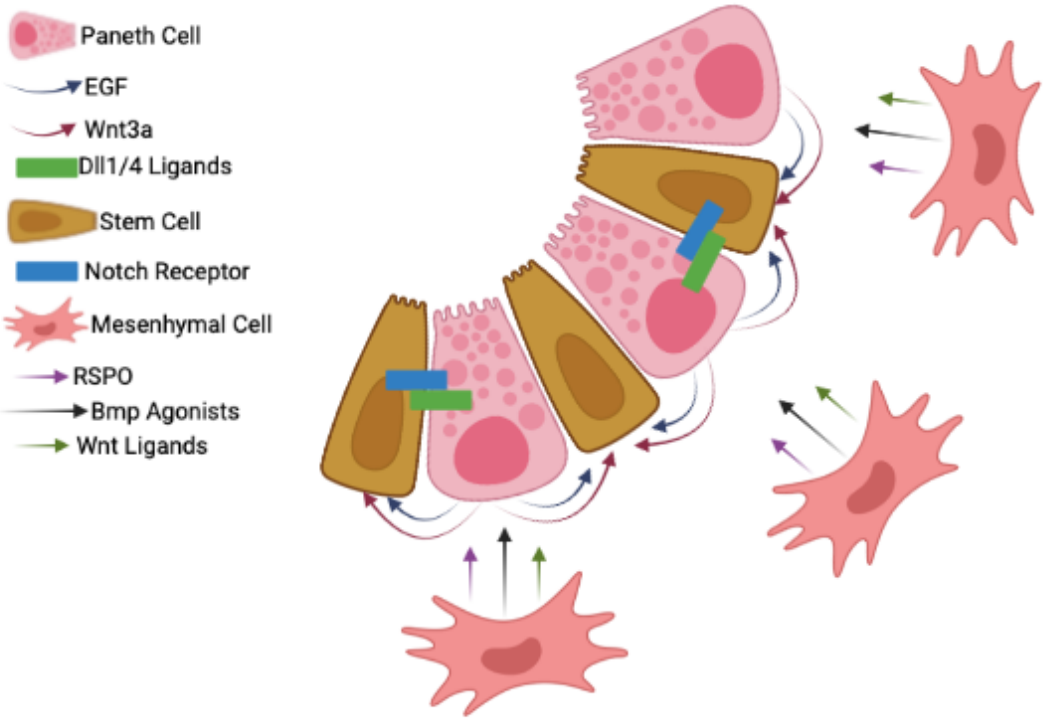


Figure 2: Intestinal stem cell niche. Paneth and stromal mesenchymal cells secrete certain ligands and agonists to control the fate of stem cells within the intestinal crypt[5]. (Adapted from citation)

gene expression in the intestinal crypt[28]. Another example is the loss of DNA methyltransferase 1 will result in stem cell expansion and reduced differentiation in the intestinal crypt[29]. To summarise, the research shows that epigenetic modification, such as DNA methylation, plays a crucial role in the definition of ICS fate[5].

4.3 Colorectal Cancer

Colorectal cancer (CRC) is the world's 4th most common cancer, and its mortality is expected to rise every year[30]. Whilst it was a rare form of cancer in the 1950s, there is a concerning increase in incidence among the western and developing world of CRC[31]. It is expected that by 2035 the number of new cases per year may reach that of 2.5 million with nearly 10% of all cancer related deaths being attributed to CRC[32]. With early detection, CRC can be highly treatable with surgery and adjuvant chemotherapy. Unfortunately, CRC has a high recurrence rate with survival being much lower in drug resistant patients[30]. The 5-year survival rate of CRC is around 64% but drops to 12% in metastatic and recurrent CRC. This shows the need for knowledge into the resistant cells responsible for this recurrence[31]. CRC occurs exclusively in the colon or rectum and is caused primarily by the aberrant proliferation of the epithelial lining forming tumours. CRC can be characterised into 3 distinct types: Sporadic, Hereditary and Colitis-associated. With harmful risk factors, such as diet and lifestyle becoming more apparent in western culture, the strain of CRC on health services is an ever-increasing burden[30].

CRC development is seen to take around 10-15 years for transformation from polyp to malignant growth. Regular screening for polyps and abnormalities has been pivotal in early detection of CRC development allowing for treatment at a more effective time. Unfortunately, only 40% of cases are detected at early stages and with CRC recurrence being a problem, there is still much need for development of strategies to target this disease.[30] In this section I will discuss the development and epidemiology of CRC, as well as the pathophysiology and risk factors associated with this tumour type.

4.3.1 Development.

CRC can be highly genetically diverse but will develop via a set of unique mechanisms as seen in the fact that many CRC cells contain dozens of soma clonal mutations which inevitably result in a distinct level of gene expression. This causes CRC to have a large mutational load of all the malignancies. Additionally, based on the number of soma clonal mutations, CRC can be categorised as either hyper or non-hyper mutated[30]. Point mutations are non-inherited mutations which individuals can acquired throughout their life. Cancers that arise from this are known as sporadic cancers and they account for 70% of CRC[33]. Usually, these cancers are heterogenous and the point mutations can be on different target genes. Conversely, many CRC cases are seen to follow a specific mutational landscape which drives the health tissue to an adenoma and then eventually a carcinoma. The first mutation is seen on the well-established tumour suppressor gene APC which begins its development from the benign adenomas, termed 'polyps'. Around 15% of these benign growths will become malignant. APC mutation is followed by KRAS, DCC and TP53 in that order[34].

CRC is seen to develop from epithelial cells when they acquire genetic or epigenetic changes which gear them towards hyperproliferation. These cells proliferate to form a benign adenoma which can eventually form a cancerous mass and metastasis around the body. This

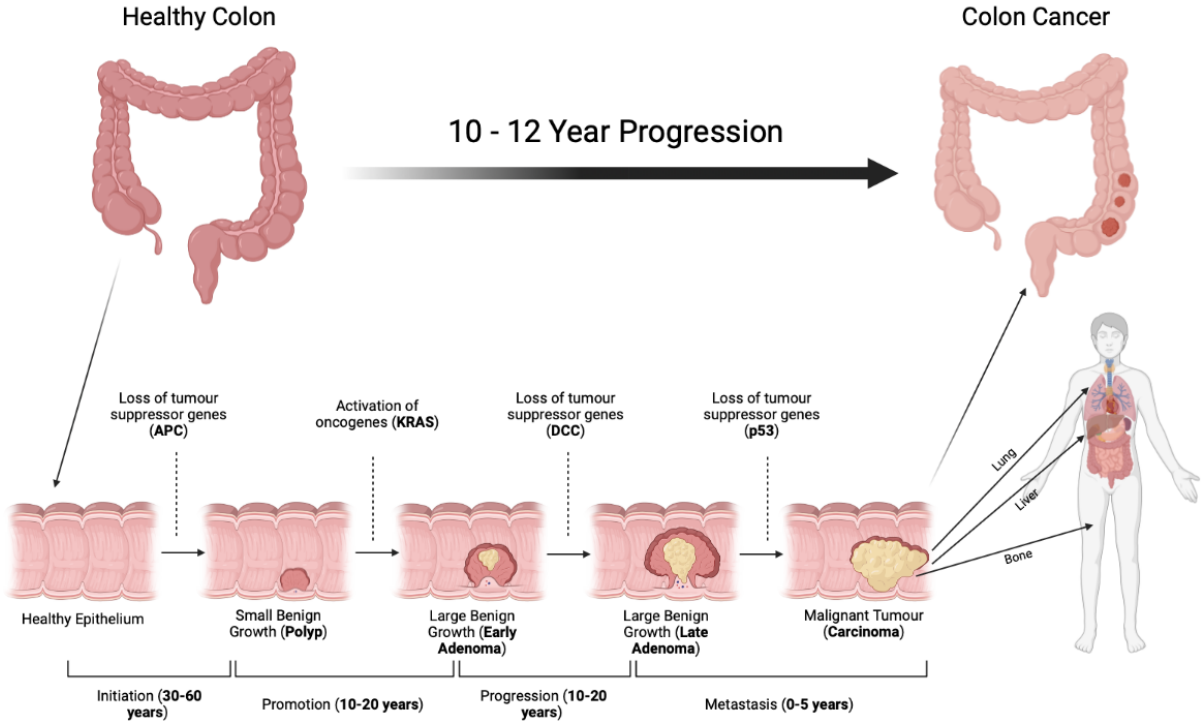


Figure 3: The 4 stages of CRC development: Initiation, promotion, progression and metastasis. There are many metastatic sites associated with CRC but the liver, lung and bone are the most commonly scene in that order[30]. (adapted from citation).

metastasis can be facilitated by chromosomal instability or microsatellite instability. It is shown develop from a tiny adenoma into a large adenoma and then carcinoma[30]. When an adenocarcinoma becomes invasive, the cells can metastasise through the body via the lymphatic or blood system. These invasive adenocarcinomas account for around 96% of all CRC but a period of up to 18 years can occur between the development of a polyp and the formation of an invasive cancer[35,36]. It has been shown that it takes an average of 9 years to metastasis and can be classified from stage 0 to IV. Only 10% of benign polyps (stage 0) will result in a malignancy and invade further into the tissue of origin (stage I). The tumour will grow in volume (stage II) and then invade into the peritoneum (stage III). This will then develop into a metastasis through the lymphatic and blood system (stage IV) (**Figure 3**)[37]. The stage determines the severity of the disease and helps guide clinicians on treatment options[30]. Environmental and genetic factors are the leading cause of the hallmark behaviour of cancer accumulation and action in CRC. As stated earlier, the progressive accumulation of genetic changes activates oncogenes and represses tumour suppressor genes. This loss of genomic stability is commonly associated with the development of neoplastic lesions within the colon, which accelerates the development of mutations on key sites to trigger malignant changes. It is believed that these changes begin in the stem cells located in the crypt base and results in the development of cancer stem cells which are essential for the development and expansion of the tumour mass[31].

4.3.2 Risk Factors

Developing CRC can be divided into modifiable and non-modifiable risk factors. Someone's personal medical history (which includes their age, sex and family history for example) cannot be controlled by the individual and are thus deemed non modifiable. Modifiable factors can be controlled by the individual and include but are not exclusive to personal habits and lifestyle. By altering these factors, an individual can reduce the chances of them developing CRC[30]. With the worldwide probability of developing CRC sitting at around 4%-5% (or 15%-20% for an individual with a family history of CRC[31]) there is great importance in understanding what factors contribute to its development and how this effects the development landscape and treatment options for individual patients[34].

Personal and Family History

Most CRC will develop in individuals over the age of 50 with the mean age being 68 in men and 72 in women. Whilst both men and women can develop CRC evidence shows males to be at a greater risk of developing the disease[30]. As previously mentioned, the risk of developing CRC is increased when a first degree relative has also been diagnosed with CRC. This risk is further increased when the relative was diagnosed before the age of 60[38].

It is believed that around 2%-8% instances arise from 2 hereditary symptoms, Lynch syndrome and Familial Adenomatous Polyposis (FAP). Lynch syndrome is autosomal dominant and arises from mutations in genes via mismatch repair errors. It is often associated with mutations in MLH1 and MSH2. This condition confers a 20% risk of CRC by 50 years of age and an 80% risk by 85 years of age. FAP is also autosomal dominant but is caused by Adenomatous Polyposis Coli (APC) gene defects. Individuals will develop thousands of colon polyps by their mid teen years with a high probability that they will go on to develop into a malignant tumour[39,40]. It is assumed that patients with FAP will have a 100% chance of developing CRC by the age of 40[38].

Inflammatory bowel disease (IBD), such as Crohn's, is seen to increase the risk of CRC development due to the chronic inflammation of mucosal tissue. This causes increased cell turnover which is turn makes them more susceptible to sporadic mutation in tumour suppressors and oncogenes[30]. It is the third most likely condition to develop into CRC after the two mentioned previously and is a chronic, incurable disease. The cause is unknown but is thought to be contributed to by genetic, environmental and immunological factors[41]. The elevated risk can be anywhere between 2-6 times more likely depending on the severity of the disease[38].

Type 2 diabetes mellitus is another condition that has been evidenced to increase the risk of CRC in individuals, it is shown to be closely related to proximal CRC over distal[30]. Increase to the insulin concentration, hyperinsulinemia, is thought to aid the development of CRC by stimulating cell division because of the indirect increase of insulin derived growth factor 1 (IGF-1). IGF-1 increases cell division and represses cell death thus increasing the chance for cell lineages to acquire harmful mutations[42]. Additionally, chronic inflammation is

associated with a favourable carcinogenic environment and facilitates tumour initiation and growth[38].

Lifestyle and Diet

Due to its location and biology, diet plays and important role in assessing an individual's risk in developing CRC. Red and processed meat are 2 known factors that increase your risk of developing CRC. Both have been classified as carcinogenic and it has been estimated that your risk of developing CRC will increase by 17% for every 100g of daily intake[43]. It is believed that these substances influence the development of CRC by 3 main products: heterocyclic amines (HACs), polycyclic aromatic hydrocarbons (PAHs) and N-nitroso compounds (NOCs). HACs and PAHs are genotoxic, meaning they have the potential to cause point mutations and initiate carcinogenesis. NOCs are carcinogenic agents and are known to interact with DNA. Another substance found in red meat that is thought to contribute to CRC is heme. This iron containing compound causes oxidative stress in intestinal cells and leads to DNA damage and genetic mutations causing hyper proliferation within the cells which can result in the formation of polyps and then a malignant mass[30,38]. However, consuming a high fibre diet that includes many fruits and vegetables is associated with low CRC prevalence[30].

Obesity is also associated with an elevated risk of CRC. A meta-analysis summary showed that CRC risk increased by 10% every increasing BMI of 8 kg/m^2 [30]. It is believed that due to adipose tissue being an endocrine organ it controls energy consumption and inflammatory response. Overweight individuals are seen to secrete more anti apoptotic factors, such as IL-1, which are known to exhibit mitogenic effect on epithelial cells and inhibit apoptosis[38].

Additional lifestyle factors also have an important effect on CRC development. Regular physical activity is shown to decrease inflammation and stress and aid in immune response and hormonal regulation[38]. Alcohol is also shown to increase the chance of CRC development by 20-40% depending on the daily intake[44]. Ethanol found in the alcohol is metabolised and influences the gut microbiome via the development of DNA adducts, triggering of oxidative stress and generation of harmful immunomodulatory effects[41]. Cigarette smoking is associated with a 2-3-fold increase in CRC development chance. It is

also considered that cigarette smoke is responsible for 12% of CRC deaths[41]. Being a mixture of thousands of chemicals with over 60 of those being carcinogenic, cigarette smoke causes DNA damage and activation of carcinogenic pathways[30,38].

Gut Microbe

The gut microbiome is a collection of thousands of bacteria, viruses and other microorganisms that play a valuable role in nutrient and drug metabolism and absorption. It also plays an important role in maintaining the intestinal barrier and protects against pathogens. Research in CRC patients has shown that alterations in the gut microbiome may be a deciding factor in CRC initiation and progression. Toxic products from bacteria have shown to cause DNA damage and immune stimulation which disturbs the intestinal barrier function. All of this causes a microenvironment that is favourable for carcinogenesis and facilitates tumour growth and expansion[38].

4.3.3 Treatment Strategies

As with many malignancies, chemotherapy and surgery remain the first line offence for patients presenting with CRC. Nearly a quarter of all CRC patients will present with metastatic sites-thus rendering surgery an ineffective control method alone[45]. Chemotherapy is often used in combination as an adjuvant treatment to aid with surgery. It can be administered as a single target agent compromising of 5-fluorouracil (5FU) or a cocktail containing many drugs that often includes Oxaliplatin. The chemotherapy can either be delivered prior to or post-surgery. The aim when delivering the chemotherapy prior to surgery is shrink the tumour so that it becomes surgically viable. The aim when delivering the chemotherapy post-surgery is as an aid to clearing up any malign cells left around the surgical site and or to target micro lesions that may have spread around the body. Patients with poor performance are recommended the lone drug therapy, as additive agents seem to have a similar effect with varying side effects. However, this treatment option has several drawbacks including, but not exclusive to, systemic toxicity, poor response rates, acquired tumour resistance and low tumour specificity. As a result, large amounts of research has been undertaken in an attempt to develop novel approaches to targeting CRC, some of which will be discussed in the next section[30,32].

Immunotherapy's (Check Point Inhibitors)

Malignancies with several specific mutations can be identified and destroyed by the host immune system via the recognition of T cells and their binding via the MHC molecules on antigen presenting cells (APC). This must be supported by secondary signals which play and important role in immune cell tolerance. This double check mechanism is required to avoid excessive immune activation and auto immune conditions. Tumours can escape and avoid the immune system in a variety of ways, such as the secretion of immunosuppressive chemokines or the recruitment of immunosuppressive cells. Another explanation is the activation of co-inhibitory receptors or checkpoint receptors. These receptors are located on the surface of T cells and include program death ligand 1 (PD1) and cytotoxic T lymphocyte antigen 4 (CTLA-4). CTLA-4 downregulates IL-2 secretion by competitively binding to B7-1/2 to reduce CD28 immune cell affect. PD-1 activation leads to the diminishment of downstream pathways such as PI3K which is responsible for the suppression of immune cell proliferation[46]. CRC are shown to have a higher abundance of PDL-1[32]. A phase 2 clinical trial is currently underway utilising a dual mixture of Nivolumab and Ipilimumab, agents which block these ligands to stop their activation and thus reduce immune avoidance[30].

Targeted Therapies

Targeted therapies are a novel approach that has been taken to tackling many different malignancies including CRC. In CRC it has shown increasing promise and has increased considerably the overall survival in patients. This therapy works by blocking certain genetic pathways and receptors that tumours can use in order to facilitate tumour growth[30]. EGFR is one such receptor, being a unique target among the tyrosine kinase receptors. It effects the RAS/RAF/MEK/ERK, PI3K/AKT and JAK/STAT pathways utilised by cancers for their ability to regulate cell growth, survival and migration. EGFR is shown to be upregulated in many different malignancies including CRC and is a potential target[32]. Drugs such as Cetuximab and Panitumumab are approved by the FDA for targeting this pathway[30,32].

VEGF is another such pathway showing promise as a target for treatment. VEGF is known to promote the formations of new blood vessels in the tumour mass (angiogenesis). This is a hallmark of cancer and is vitally important to facilitate the ability to supply oxygen and nutrients required for tumour growth and metastasis[32].

4.4 Cancer Stem Cells

Proposed almost four decades ago, the cancer stem cell (CSC) theory states that tumour growth is fuelled by a small subpopulation of dedicated malignant cells. This small subpopulation is also believed to be responsible for the inevitable tumour relapse after treatment and the facilitation of metastasis[47]. With over 70% of CRC deaths being attributed to metastatic sites in the liver, CRC CSC provide a highly attractive target for therapies and understanding[48]. With relapse after treatment being a significant problem in cancer treatment, understanding the tumour mechanisms of such relapses presents a gap in our knowledge that requires attention. Multiple resistance mechanisms can be present within a patient or even in a single tumour. Random genetic events confer resistant to drugs and push for genetic diversity. When administering drugs, a small population of cells can survive by entering a drug tolerant persister state. Here, there is little population growth during those treatments, such as chemotherapy, that target fast cycling cells and are thereby ineffective. After long term treatment, these resistant cells regain the ability to grow and divide and thus create a treatment resistant tumour. It is not known whether the survival and expansion through this drug tolerant state mediates the acquisition of genetically driven resistant mechanisms. Additionally, whether these resistant mechanisms occur through a persister bottleneck is unclear[49]. There has been proposed three distinct strategies for targeting these cell pools; target the tumour cell plasticity, prevent quiescent and find drug to target the persister cells or reverse the quiescence change[50].

4.4.1 Cancer Stem Cells Model

CSC share many properties with embryonic stem cells (ESC). However, unlike ESC, CSC have lost control of their replicative ability which can result in tumorigenesis. They are often defined by the ability to self-renew and produce differentiated cells[48]. Recently, the CSC models have relied on four main principles. First, some of the cellular hierarchy of tumours results from the hierarchical organisation reminiscent of the tissue of origin. Second, these hierarchies are fuelled by self-renewing, quiescent CSC with the bulk of the tumour mass being non-CSC capable of transient proliferation and therefore are unable to contribute to long term growth. Third, CSC identity is hardwired which is illustrated by the fact that non-CSC rarely initiate tumours in xenograft models and thus shows limited plasticity. Finally,

CSC are resistant to chemotherapy and radiotherapy which target non-CSC which provides insight into treatment relapse[47].

4.4.2 Stem Cell Plasticity

Lineage trancing experiments have highlighted the ability for committed differentiated cells to move up and down the hierarchy of differentiation. This is termed plasticity and whilst has been acknowledged for a while, it is now believed to be more prevalent than recently thought. This can be seen in the epithelium of the mouse intestine. Removing the ISC that possess LGR5 does not viably influence the integrity of the epithelium as would be expected. Instead, dedicated progenitor cells will revert in the multipotent LGR5+ stem cells and replace the removed ones. This can be understood when we observe the signals that are present in the surrounding environment. As stated earlier, ISC revives Wnt to promote self-renewing and Notch to block differentiation. They can also receive powerful mitotic stimuli that triggers EGFR. When progenitor cells are provided with these signals, and blocked BMP/TGF-β, they rapidly regain stemness[47,51]. This demonstrates that committed cells next to the crypt base can regain stemness and revert to a multipotent fate[47]. It has been shown that both CSC and non-CSC are plastic and can undergo several unique phenotypic changes. A study took stem cell, basal cell and luminal cell types from breast cancer. It was evidenced that in vitro these three populations were able to generate cells of the other two phenotypes demonstrating their ability to change their phenotype to different stimuli. Additionally, when environmental stimuli were controlled, all three cell types were equally tumorigenic and able to produce xenograft[48,52]. What this shows us is that the CSC state is not hardwired, and cell types are able to change their fate depending on the environment they find themselves in[47]. Additionally, work in CRC has shown the ability of the microenvironment to control cell plasticity. Many CRC are initiated by genetic changes to activate Wnt signalling that transform the crypt progenitor cells into a CRC phenotype. Within single tumours they can display distinct Wnt signal levels despite sharing downstream mutations[53]. These Wnt levels is shown to correlate to tumour initiating capacity in xenotransplantation assays[47].

4.4.3 CSC and EMT

Epithelial to mesenchymal transition (EMT) is the process whereby epithelial cells can acquire a mesenchymal gene pool that triggers a phenotypic change facilitating migration and invasion, two factors important for cancer metastasis. Recently, the connection between CSC and EMT has been gaining attention. It is known that the overexpression of EMT transcription factors not only supports the more migratory phenotype but also enhances tumour initiating capacity, a factor important to CSC[47]. This can be seen in tumours with enhanced Snail Family Transcription Receptor 1 (SNAI1) (a master EMT transcription factor) and displays greatly enhanced tumour initiating capacity in both mouse and human models[54].

However, something that is still poorly understood is that metastasis in many carcinoma types remain an epithelial organisation and lacks mesenchymal traits[55]. This suggests that EMT is not needed for metastasis in certain situations and is displayed in the fact that cells frozen in a constant state of EMT are poorly metastatic and only begin to move when the EMT genes are silenced[56]. Additionally, it is known that an EMT transcription factor known as TWIST1 regulates CSC properties and tumour progression in skin cancer. These findings suggest that EMT is uncoupled with CSC properties[47].

There are two viewpoints that may be able to explain these observations. Firstly, is that EMT in cancer may be transient, meaning that epithelial tumour cells can adopt reversable mesenchymal states facilitated by environmental cues which would then result in a plastic CSC phenotype. This can be witnessed in breast cancer were cancer cells swap between CSC and non-CSC states depending on the expression of the EMT transcription factor ZEB1[57]. Secondly, it has been shown that expression of TWIST1 primes mammary cells towards the CSC state that continues to persist after TWIST1 is ablated[58]. All of this demonstrates that CSC hierarchy is not fixed and that conversations between CSC and non-CSC cell types may be a common phenomenon driven by environmental cues[47].

5.4.4 CSC Metabolism

Stems cells are shown to utilise glycolysis to generate energy. This led to it being hypothesised that CSC utilised the same methods as stranded stem cells to generate their energy. It was demonstrated in a series of studies on multiple different cancer types,

including glioblastoma and lung cancer, that the glycolytic rate of CSC was significantly higher than that of non-CSC[59]. Supporting this was the fact that glucose uptake and lactate production, along with other factors associated with glycolysis, were significantly increased in CSC[60]. Additionally, genes important to glycolysis, such as PDK1 and PKM2, were seen to be upregulated in brain CSC[61]. The glycolytic switch in CSC seems to play more of a role in stemness then being a direct consequence of it. Studies demonstrated that the transition from oxidative phosphorylation (OXPHOS) to glycolysis can increase the stem cell population. This switch seems to be facilitated by HIF-1, whose role is related to stem cells characteristics, such as self-renewal and therapy resistance. In tumours, one of the ways HIF-1 affects the glucose uptake though the upregulation of GLUT1 expression during glycolysis. Additionally, it reduces the amount of reactive oxygen species (ROS) by increasing glycolysis[62]. This homeostasis of ROS is necessary to facilitate stemness in cancers, such as breast cancer, in response to hypoxia of chemotherapy toxicity[63].

OXPHOS is a crucial metabolism used in most complex tissues. However, this produces reactive oxygen species which are known to cause stem cell dysfunction. As observations of haemopoietic stem cells demonstrate that they reside within a hypoxic environment, it is assumed that stem cells avoid OXPHOS by employing glycolysis. Long term OXPHOS induction in stem cells is shown to reduce the stem cell pool and loss of quiescence. In the intestine, it is seen that highly proliferative LGR5+ ISC have elevated OXPHOS and that Paneth cells perform glycolysis, providing lactate for ISC for later metabolism[64]. It is seen that, unlike other stem cells, reactive oxygen species do not cause damage but instead promote differentiation via the MAPK pathway[65].

Altered metabolism is one of the main hallmarks of cancer and has inspired many targeted treatment strategies. Additionally, just as cancers have a distinct metabolism compared to normal cells, CSC also have a unique metabolism compared to cancer cells. Metastatic cancer stem cells display distinct metabolic pattens depending on their organ of spread. Metastatic breast cancer that spreads to the liver (but not lungs or bone) shows reduced glutamine and OXPHOS metabolism. However, it also transforms glucose derived pyruvate into lactate which is a phenomenon known as the Warburg effect[47].

Not only is the metabolism of a CSC controlled by the cell itself, but it is also recognised that the surrounding microenvironment plays an important role in CSC metabolism. Cancer Associated Fibroblasts (CAFS) are reported to promote the formation of tumour spheres and induce epigenetic metabolism changes in CSC in pancreatic cancer[66]. It is also shown that the CAFs have metabolic adaptations that support glycolysis and nutrient use in CSC. CAFs are seen to be influenced to begin glycolysis by metabolic changes such as HIF-1a[67]. They then produce metabolites that are taken up by the CSC to promote the production of oxidative energy associated to tumour progression and treatment resistance. This is known as the reverse Warburg effect[63].

Inflammation is another hallmark of a cancer environment. Macrophages are an abundant cell type located in the tumour microenvironment. Some are characterised by their attraction to pro inflammatory cytokines produced by CSC and are known as Tumour Associated Macrophages (TAMs). It is here that they must alter their metabolism to compete with the CSC for nutrients in the harsh environment. Under hypoxia, a condition often associated with solid tumours, TAMs upregulate an angiogenic response and form abnormal blood vessels[63].

The evidence here demonstrates that CSC have a unique metabolic landscape with specific energy requirements. This may present an opportunity for targeted treatments in both early and late-stage disease.

4.4.5 CSC and Therapy Resistance

Chemotherapy, along with radiotherapy and surgery, is still one of the leading front-line treatments for many cancer types. However, studies have shown that the use of chemotherapy enhances tumour heterogeneity which in turn lessens the effectiveness of treatment and causes treatment failure and disease progression[68]. These populations of chemoresistance stem cells are thought to be enriched by CSC[47]. One of the main drivers of this chemoresistance in CSC have been demonstrated to originate from their stemness-related pathways. This is witnessed in the downregulation of Hedgehog (Hh), Wnt/βcatenin and Hippo pathways. Hh is shown to regulate the properties of CSC by upregulating stemness genes such as SRY Box Transcription Factor 2 (Sox2)[69]. In CRC specifically, Hh-GL1 pathways supports the survival and metastasis of CSC to chemotherapy[70]. Wnt/catenin

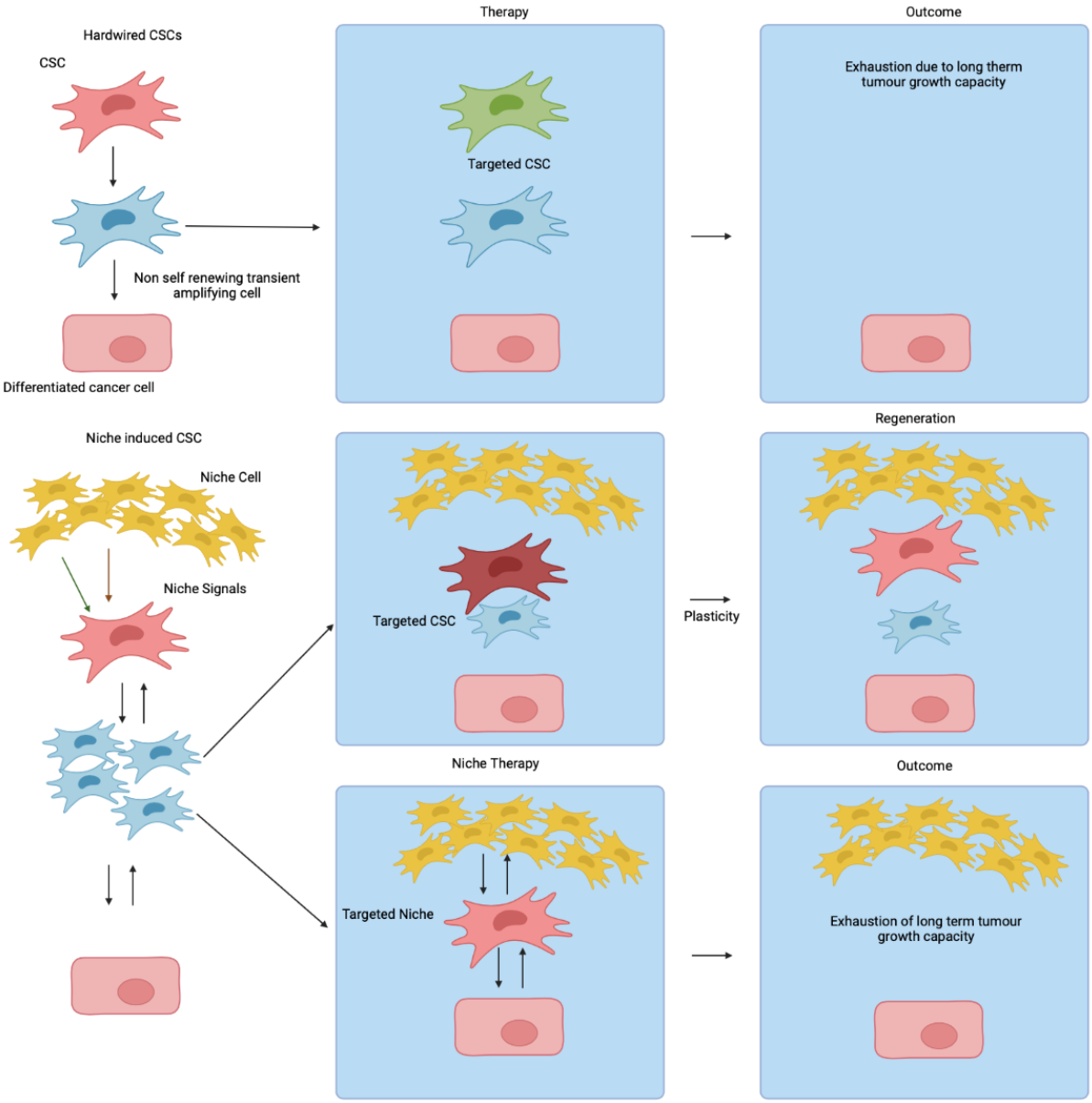


Figure 4: Difficulty with targeting CSC to overcome therapy resistance. In tumours with unidirectional CSC communication, destruction of the therapy resistance CSC pool is usually capable of complete tumour ablation. However, many CSC show a high degree of cell plasticity. This allows niche signals from the surrounding environment to re program the stem cell population to fulfil either the progenitor or differentiated role after loss of CSC. This in turn will result in tumour regeneration and ultimately, therapy failure. Blocking the niche signals that provide this may provide a viable therapeutic target for further investigation.[47]. (Adapted from citation)

activation is shown to foster the development of CRC. Knock down of this pathway causes a decrease of stemness related genes which in turn causes a reduction of the CSC pool in breast cancer cells. Wnt/β catenin activation has also been linked to chemoresistance in several studies. This is evident in LGR5, a Wnt target gene, promoting resistance to the chemotherapy 5FU in CSCs[71]. In this next section I will summarise some of the other ways in which CSC are thought to contribute to therapy resistance.

CSC and EMT

EMT and stemness are 2 factors often observed together. EMT and stemness markers are expressed in circulating tumour cells and EMT markers confer a stem like phenotype to cancer cells[72]. EMT is also related to drug resistance. For example, the invasive phenotype gained by cancer cells undergoing EMT can be used to escape anti EGFR therapies, this is due to mesenchymal cells having an increased expression of genes involved in metastasis and invasion (**Figure 4**)[73]. This shows EMT maintaining stem like characteristics that grant therapy resistance[68].

Multi-Drug Resistance and Detoxification Proteins

Side population cells show a CSC like phenotype and are often found in a variety of solid malignancies including gastrointestinal cancers. These cells exhibit a high expression of drug transporter proteins that act to expel chemotherapeutic agents from the environment which, in turn, confer a better survival and relapse chance to the cancer cells[68,74]. These are known as ABC transporter proteins. One such ABC transporter protein, ABCG2, is known to have the ability to effect drugs, such as doxorubicin, and is expressed on more than half of chemo resistant tumours[75]. ALDH superfamily is a detoxification enzyme and is responsible for the oxidation of aldehydes and retinol to carboxylic acid and retinoic acid respectively. This detoxifies the cells from the drugs and the ROS. ALDH+ cells are used as a diagnostic marker for chemotherapy resistance[71].

CSC Dormancy

It has been discovered that, in addition to the intra tumour heterogeneity, there are also functionally distinct clones within the tumour mass. These clones promote cells with greater survivability in times of stress from treatments such as chemotherapy. This is evidenced when chemotherapy is administered to the tumour which in turn causes growth to slow and ultimately for the tumour to revert to a dormant state[68,76]. In this dormant state the cells are confined to the G0 phase but are still able to respond to mitotic signals[75]. These cells still maintain a potent tumour propagation potential which leads to both chemotherapy resistance and tumour progression[68]. This is due, in part, to chemotherapy and radiotherapy

being effective at destroying reproducing cells, whereas these dormant cell population, comprising of both CSC and non-CSC, are avoided and ultimately survive[75].

CSC and DNA Damage Resistance

Conventional therapies, such as chemo and radiotherapies, induce DNA damage to destroy cells. However, CSC can be resistant to this form of damage. This is possible as they can protect themselves from oxidative DNA damage, promoting DNA repair pathways such as ATM and anti-Apoptotic signalling[68]. The relationship between apoptotic and anti-apoptotic signals controls the sensitivity of cells to death[77]. CSC are seen to weaken the expression of death receptors as seen in leukaemia, where CSC like populations display a significantly lower expression of FAS ligands then the differentiated cells, which in turn causes chemoresistance. It was discovered that this could be reversed by introducing synthetic FAS ligands to glioma stem cells[78]. This is also supported by the studies undertaken with soluble TRAIL (an apoptosis inducing ligand) which impairs the tumour growth in mouse models of Glioblastoma[79]. However, the short half-life of this reduced its efficacy but could be overcome with TRAIL-engineered mesenchymal stromal cells to induce apoptosis in lung and breast stem like cancer cells[80]. Additionally, CSC usually show TRAIL resistance due to the overexpression of cFLIP which supports the self-renewal capacity[81]. CSC also showed higher expressions of BCL-2, an anti-apoptotic factor, compared to the differentiated cancer cell pool[82]. Furthermore, the CSC pool can activate the Src/Slug signalling pathway to support the self-renewal ability of the CSC. Pharmacological inhibition of Src with Dasatinib reverses this change, increasing the cells sensitivity to treatment and reduces apoptosis avoidance[71,83].

Due to chemo and radiotherapies being specialised to destroy rapidly dividing cells, the quiescent CSC are often missed. This is further enhanced by the efficient DNA damage response found within this cell population. This has been demonstrated when treatment of ovarian and lung cancers with Cisplatin caused an enrichment of the CSC subpopulation[71].

Tumour Environment

It is well established that tumours are supported by surrounding stromal cells and other products known as the tumour microenvironment. It has also been shown that there is an additional microenvironment, known as the CSC microenvironment[84], that supports CSC

maintenance and contains native stomal[85] and vasculature tissue[86]. This environment accelerates the CSC divisional process and thus allows them to differentiate into daughter cells whilst continuing to maintain their capacity for self-renewal and their primitive developmental state. The cells possessed by this specific microenvironment stimulate certain signalling pathways[72], such as notch[86] and Wnt[87], that can alter the state of the CSC, support metastasis, cell division and repopulation[68]. CSC are also seen to facilitate interactions via regulation of autocrine and paracrine signals between the tumour and TME. This is particularly evident in the presence of the extracellular matrix, cancer associated fibroblasts (CAF) and immune cells[71].

Cancer Associated Fibroblasts

Cancer Associated Fibroblasts (CAF) are a cell type found within the tumour microenvironment that secrete many molecules including pro-inflammatory cytokines and growth factors. CAF secrete factors that are known to have a close relationship with CSC and play a major component in their self-renewal capacity, plasticity and chemoresistance[88]. It has been evidenced that in colorectal cancers, CAF presence guarantees a CSC population and supports their expansion[89]. Additionally, cells grown in CAF conditioned media are seen to be more aggressive than those not. It does this via the modulation of stemness features of the cancer cells by the secretion of chemokines, cytokines and growth factors[90]. Amongst the many of these produced by CAFs, research has highlighted that two, IL-6 and IL-8, play a major role in maintaining the stem like state of cancer cells and facilitate metastasis and chemoresistance[91]. Furthermore, the production of TGF-β induced the acquisition of the stem like phenotype and promotes EMT and chemoresistance in cancers[92].

Many studies have shown the importance of CAFS in chemoresistance of solid tumours. It was discovered that CAFs secrete high levels of TGF-β2 and IL-6 which promotes the transcription of GLI-2 and allows resistance to 5FU and oxaliplatin treatment[93]. Additionally, in lung and breast cancer, CAF production of IL-6 and IL-8 enriched the CSC pool and provided resistance to Cisplatin treatment[71].

4.5 Colorectal Cancer Stem Cells

Whilst conventual therapies including chemotherapy and surgical resection are a powerful and effective tool with treating CRC, treatment resistance and relapses remain a major

problem in CRC care. It is believed that this resistance is driven by colorectal cancer stem cells (CCSC), in addition to their role in tumorigenesis and metastasis. With multidrug resistance and minimal residual disease being too often seen within CRC care, attention has turned to the subpopulation of CCSC. It is believed that, due to its tumour initiating capacity and plasticity, as well as a complex relationship to the TME, CCSC is the main culprit to these issues. In this section, I will discuss the origin of CCSC as well as the identification and the role of this population in CRC[94].

4.5.1 Origins

The discovery of CSC within myeloid leukaemia opened the doorway to the discovery of CSC within other tumour types[95]. Whilst the specific origin of CCSC remains controversial, there are several hypotheses that have emerged. CCSC contains many tumour-like characteristics, such as treatment resistance and uncontrolled growth, that may be acquired from aberrant genetic and epigenetic changes. These changes may be specific to progenitor cells or by the dedifferentiation of somatic cells. This dedifferentiation may be facilitated by genetic and environmental factors[87]. It is well established that the CCSC play an important role in the initiation and invasion of CRC. CCSC have been shown to give rise to heterogenous tumours with the ability to be transplanted into immunodeficient mice and maintain a tumour initiation capacity. Additionally, the tumour that forms will share characteristics of the original tumour[96]. These tumours then can go on to metastasis and produce secondary tumour sites. Whilst the only make up a small population of the tumour, CCSC plays an invaluable role in forming and sustaining the tumour[94].

4.5.2 Identification

One of the first markers used to identify CCSC is the cell surface marker CD133[97]. It is known as the prominin-1 glycoprotein and is responsible for the organisation of cell membrane topology[98]. Isolation of CD133+ cells were shown to be capable of tumour formation in mouse models and remained undifferentiation when cultured in serum free media. They were also shown to have an increasingly more aggressive phenotype along the generations[99]. Additionally, most tumours compromising CD133- cells were unable to initiate tumour formation, demonstrating its promising potential as a stem cell marker[100]. Another study revealed that CD133+ cells produced IL-4 which is utilised to avoid apoptosis

and thus grants the tumour chemotherapy resistance[101]. What this demonstrated was that CD133 provides both tumours initiating capacity and treatment resistance, two important CSC factors[48]. However, the prognostic value of CD133 has given inconsistent results and was not shown to affect tumour course[102]. Additionally, mutations within the RAS-RAF program were shown to affect CD133 expression without effecting the tumours' ability to initiate in immunodeficient mice[97]. Whilst CD133 was the first to be proposed as a CCRC marker, many more have emerged since. LGR5 is expressed only by crypt base cells and is proven to have the ability to differentiate into different epithelium linages[103]. These LGR5+ cells are found near to Paneth cells in the crypt base and are said to play essential roles in the maintenance of stem cells[104]. Additionally, LGR5 knock down has been shown to cause tumour regression and recovery of LGR5+ cells are shown to cause tumour growth and recurrence[105]. SOX2 is known to prevent differentiation and is shown to be expressed in late-stage tumours in CRC. SOX2 is located at the crypt base and high levels of SOX2 correlates to a more aggressive and metastatic disease[106,107]. STAT3 activation leads to tumour invasion and proliferation[108]. It is located within the crypt base epithelium and is

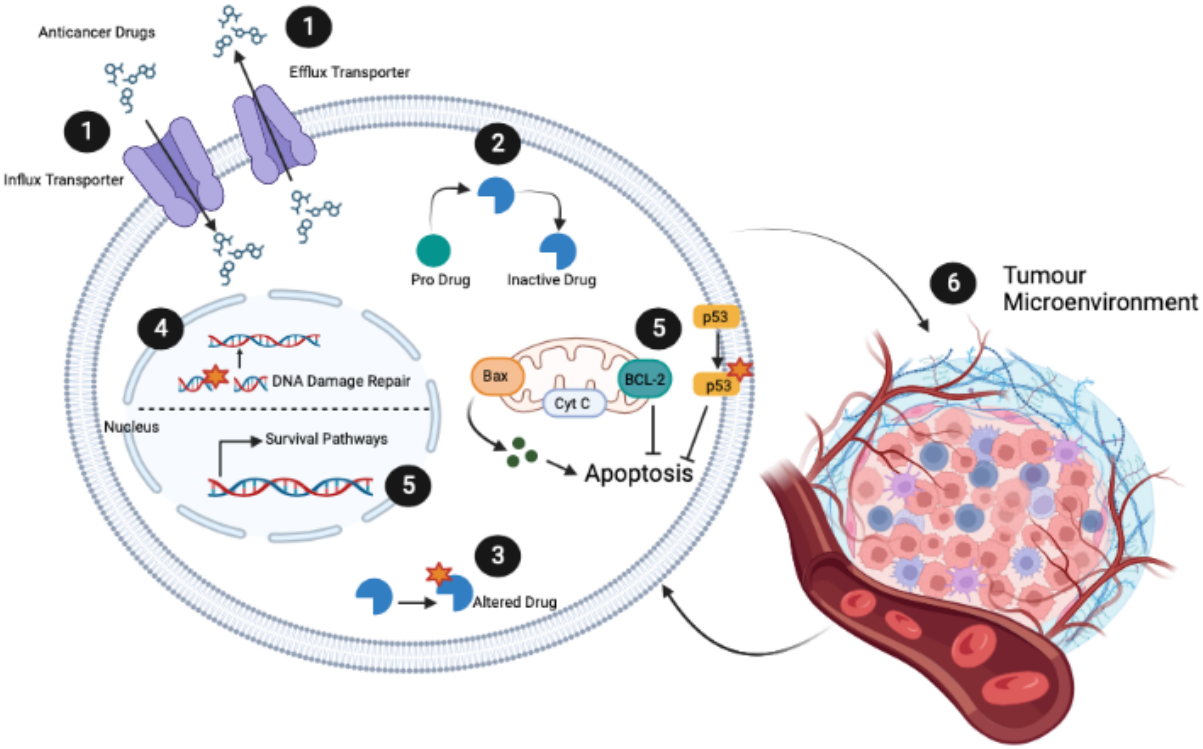


Figure 5: A collection of the major mechanisms that CSC can employ to enhance a tumours drug resistance capability. These include: changes to the transportation of chemotherapy drugs in and out of the cell (1), changes and impairment to the metabolism of chemotherapy drugs (2), altering the targets of the drugs (3), enhancing the DNA damage repair to reverse the damage caused by chemotherapy (4), dysregulation of the balance between pro and anti-apoptosis pathways (5) and the ability of the tumour microenvironment to influence these factors (6).[94] (Adapted from citation)

related to chemotherapy resistance with STAT3 inhabitation leading to 5FU sensitivity[109]. Additionally, STAT3 is related to other proposed CSC markers as they are shown to be present in higher concentrations within CD133+ cells[48].

4.5.3 Role of CSC in colorectal cancer

With 20% of new CRC patients presenting with metastatic disease, effective treatment is vital in combating this disease. Resistance to anti-cancer therapies presents a difficult challenge one which is desperately needed to be overcome to deliver effective treatment. The properties of CSC make them forerunners in allowing the tumour to reduce the efficacy of chemotherapies[94]. A limitation to the influx and efflux of anti-cancer drugs could be a contender in combating the resistance shown in CRC. ABC transporter proteins play a pivotal role in the efflux of anti-cancer drugs from the cells (**Figure 5**)[110]. In colon cancer specifically, ABCB1 is seen to be overexpressed thus reducing the build-up of chemotherapy drug within the cell and ultimately leading to treatment failure and disease progression[111]. SLC is another important transporter protein and has shown to be affected in cancer, reducing the ability for chemotherapies to move into the cell and thus diminish the influx of drug into the cell[112]. Higher expression of these transporters of CCSC may explain their innate drug resistance[94]. Overexpression of drug metabolising enzymes, such as cytochrome P450, may also play a role in treatment resistance[110]. These enzymes play a key role in the ability to metabolise CRC drugs such as SN-38[113]. The higher concentrations of these drugs within CCSC may play a central role in their ability to resist drugs with mutations in certain genes resulting in changes to the targets of these drugs[94]. This is witnessed by the alterations to the KRAS gene in anti EGFR therapies[114]. Drug resistance can also be explained by CCSC ability to effectively repair damaged DNA. DNA damage, by the addition of bulky adducts, is a mechanism of how platinum-based chemotherapies, such as Oxaliplatin, effectively destroy cancer cells[110]. Upregulation of excision repair pathways has been demonstrated to result in resistance to this drug in CRC as the balance between pro and anti-apoptotic signals controls cell death[115]. However, this balance is often seen to be dysregulated in CCSC. TP53 is a powerful gene responsible for the induction of apoptosis when cells are introduced to chemotherapy and is seen to be mutated in around 85% of CRC cases. This leads to a resistance to chemotherapies, such as 5FU, compared to unmutated

types[116]. With this often being seen in CCSC it is yet another factor that may explain their drug resistance[94].

5 Materials and Methods

5.1 Caco2 Cell Line

For the experimentation in this project an undifferentiated Caco2 cell line was generated. This is used as an ease-of-use terminology as the cell type is not strictly undifferentiated. Previous work on Caco2 cell lines described the ability for Caco2 cells to dedifferentiate when maintained at a lower confluence of 50% rather the usual operating confluence of 80%. When this was done, the cells maintained a more homogonous and polarized mono-layer. The cells that I used were seeded at 50% confluence and maintained this way for a minimum of 10 passages before being subject to experimentation[117].

5.2 Cell Culture

5.2.1 Cell Passage

Cells were maintained in DMEM high glucose medium and 10% Fetal Boovine Serum (FBS) and Pen-Strep (P/S) in T75 flasks. When passage was required, media was removed and then cell plate was washed with 5 ml PBS to maximise media removal. PBS was then removed and Trypsin EDTA solution (Sigma T4049) added in the amount of 2.5 ml and incubated at 37℃ at 5% CO_2 for 5 mins until all cells were detached from the flask. This was then washed down using 5 ml of media and aspirated along the bottom of the flask to ensure highest cell yield. This was transferred to a 15ml flacon tube and centrifuged at 600 g for 5 minutes at room temperature. During this time a new flask was prepared with 10 ml of media and placed in an incubator to reach 37℃. Once centrifugation was completed, the old media was removed and 5ml of new media added. The pellet was then re suspended via aspiration. Then around 500 µl (depending on initial confluence of cells and cell type) was added to the prepared flask. Cells were maintained in this way at 37℃ and 5% CO_2. They were checked and passaged every 2 days until they had reached around 50% confluence for the undifferentiated type and 70% confluence for the differentiated. Aseptic conditions were used throughout.

5.2.2 Cryopreservation of cells

A freezing solution of 10ml media with 10% dimethyl sulfoxide (DMSO) was prepared. Media was removed from cell flask and washed with PBS to remove excess media. Trypsin was added as mentioned above and incubated for 5 minutes. Trypsin was then deactivated with 5 ml media containing FBS and mixed via aspiration. Cells were then transferred to a 15 ml falcon tube and centrifuged as mentioned above. Superintendent was removed and cells resuspended in the freezing solutions prepared earlier. 1 ml of the new solution was transferred to a cryovial and stored at -80℃.

5.2.3 Thawing cells

A new flask was prepared on 10 ml media and brought to 37℃ in the incubator. Cryovial was retrieved from -80℃ freezer and defrosted at room temperature. Once defrosted, the contents within the cryovial were transferred to the prepared flask.

5.2.4 Cell Imaging

Microscopy was used to examine cells during chemotherapy treatment and general cell maintenance. They were viewed under an EVOS M7000 Microscope (bright field at x50 magnification)

5.3 Chemotherapy Drug Treatments

Caco2 cell lines were incubated in Gibco DMEM high glucose medium with same additives as mentioned above on a 6 well plate. Cells were transferred to the plate 24 hours prior to drug administration to allow cell attachment and reach desired confluence. Undifferentiated cell group were maintained at 40-50% confluence and differentiated cell group maintained at 60-70% confluence. All three drugs were dissolved into master mixes using DMSO in drug dependent amounts. The final drug concentrations required were found using previously published work of using the specified drugs in colon cancer samples[118]. This was further optimised with testing toxicities in varying concentrations to achieve desired cell death. Drugs were administered for 48 hours in consternation of 5, 10, 15, 20, 25 µM along with new media 2 days after passage when desired confluence was reached. Control samples were solely treated with DMSO to ensure the results obtained were from the drug administration and not the dissolving agent as it is known to be toxic to cells. Aseptic conditions were used throughout.

5.4 Gene Expression analysis using RT-qPCR

RT-qPCR was used in order to compare multiple different subpopulations. These populations were differentiated and undifferentiated samples as well as samples treated with chemotherapy (5FU, Oxaliplatin and SN-38) in ranging concentrations on both differentiated and undifferentiated populations.

5.4.1 RNA Extraction

RNA was prepared using the mRNeasy Mini Kit (Qiagen) from the Caco2 cells used in the experiment. The media was removed and washed using PBS from cell plate and then transferred to a fume hood on ice. In the fume hood, the Qiagen set lysis buffer and β-mercaptoethanol were added in amounts according to the manufacturer's instructions. Cells were then homogenised using a cell scraper and left to stand for 5 minutes then an equal amount of 70% ethanol was added to the cell plate and mixed via pipette before being extracted to a sterile Eppendorf tube. 700 µl of sample was then transferred to one of the RNeasy spin columns with 2 ml collection tube and centrifuged at 12000 g for 15 seconds. After this 700 µl of the wash buffer (RW1) was added and centrifuged for the same time and at the same speed. Next, 500 µl of RPE buffer was added to the spin column and spun for the same time and speed. This was followed by another 500 µl of RPE buffer but centrifuged for 2 minutes at the same speed. Lastly the spin column was placed in a new 1.5 ml collection tube and 50 µl of RNAse free water added and spun for 1 minute at the same speed. The spin column was then discarded, and the flow analysed using the Nanodrop-2000c (thermoscientific) and stored at -80℃. Additional information can be found in the manufacture's instructions[119].

5.4.2 cDNA Synthesis

GoScript Reverse Transcription System was used for cDNA synthesis. To anneal the primers, RNA sample was mixed with olido dT primers (d18T and d24T) and dNTP nucleotides in a sterile RNase free Eppendorf tube. This was then made up to 12 µl with RNase-free water (see **Table 1** bellow for volume mixture). This was heated at 65℃ for 5 minuets to melt the secondary structure and provide optimal temperature for the oligodT to attach. The tubes were then placed on ice immediately after in order to prevent the secondary structure from

reforming. This was then centrifuged briefly to collect all the sample in the bottom of the tube. In order to complete the reverse transcription reaction, DTT 0.1M, 5X reaction buffer, Ribonuclease inhibitor (RNasin) and RTenzyme superscript were added (see **Table 2** for volume mixture). Tubes were mixed gently by flicking the bottom and then centrifuged briefly to collect all sample at the bottom of the tube. This was then heated at 42℃ for 50 minutes, followed by 70℃ for 15 minutes to inactivate the enzyme and then placed on ice to cool. These were then diluted with RNase free water and stored at -20℃.

Table 1: *Primer annealing mix for cDNA synthesis*

	Samples (RT)	Negative Control (-RT
RNA (1µ)	Depends on concentration	Depends on concentration
Primers (oligo DT)	1 µl	1 µl
dNTPs	1 µl	1 µl
Nuclease free water	Depends on concentration	Depends on concentration
Total Volume	12 µl	12 µl

Table 2: *Reverse transcription reaction master mix*

Mix	Sample
Annealed primer	12 µl
DTT, 0.1M	2 µl
5X Reaction Buffer	4 µl
Ribonuclease Inhibitor (RNasin)	1 µl
RT enzyme Superscript	1 µl
Total Volume	20 µl

5.4.3 Performing qPCR

For each sample set, a separate SYBR-Green Master mix (Applied Biosystems) was prepared by multiplying the required number of wells by the amount of each reagent volume. This was necessary to help reduce pipetting error. To account for pipetting errors, an additional 10% of master mix was prepared (see **Table 3** for mixture amount). Samples were added to a MicroAmp optical 96 well plate (Applied Biosystems) and ran in triplicate. Samples were also done for GAPDH housekeeping gene to act as a control. Amplification was analysed

using the Applied Biosystems Viia7 Real-time PCR system. When the program was finished the ΔΔCt values, as well as the melting curves, were generated with the built-in software and utilising GAPDH as the housekeeper control gene and exported to excel for further analysis. Each sample plated had their own GAPDH delta curve to be compared to. Primers can be found at **Table 4**.

Table 3: *qPCR master mix volume and calculations*

Master Mix	**Volume in µl**
Gene Primers Mix	1.5
SYBR Green	7.5
H_2O	3
Total Volume	12
Sample amount	3
Total Volume needed	3 x 18 wells = 54
10% extra	5.4 = 6
Final Volume to make	60

Table 4: *Gene primer sequences*

Gene	**Forward Primer (5'-3')**	**Reverse Primer (3'-5')**
GAPDH	GAGTCAACGGATTTGGTCGT	TTGATTTTGGAGGGATCTCG
ALDOB	AGCTGGCTTGTCATAATTC	CATAGAAAAGTCCACCCAACTCC
MUC2	ACCCGCACTATGTCACCTTC	GGACAGGACACCTTGTCGTT
MKI67	ATGCAGAATCAGAAAGGGAAAGG	TTGTCTTTCTTGATCTCAGGCAC
SOX9	CCCGCAACAGATCTCCTACA	GAAGGTGGAGTAGAGCCCTG
ANPEP	CATTATGACACACCCTACCCACT	CTCATGAGCAATCACAGTGACC

5.5 Analysis of RNA Sequencing

RNA sequencing was sent to and completed by Novogene UK. Control Caco2 RNA samples along with samples treated with 5FU and SN-38 were collected as described above.

Table 5: *Concentrations for each sample sent for RNA sequencing*

Sample Name	Concentration in ng/µl
Undifferentiated treated with 5FU and SN-38	286.3
Undifferentiated treated with 5FU	40.8
Undifferentiated treated with SN-38	708.6
Differentiated treated with 5FU and SN-38	185.4
Differentiated treated with 5FU	1280.3
Differentiated treated with SN-38	136.9

5.5.1 Galaxy Analysis

Raw data files from RNA sequencing were uploaded to Galaxy (https://usegalaxy.org)[120]. The homo sapiens hg19 was used as the reference genome. FastQC was used to analyse read quality. Once the reads had been analysed for quality, Cutadapt was used to remove contaminated reads to purify the read quality. RNA star was then used on these Cutadapt files in order to be able to visualise the data sets in software such as IGV. Finally, FeatureCounts was used to generate read count numbers for each individual gene expression for each sample type and MiltiQC was used to group each data file together for each part and visualise.

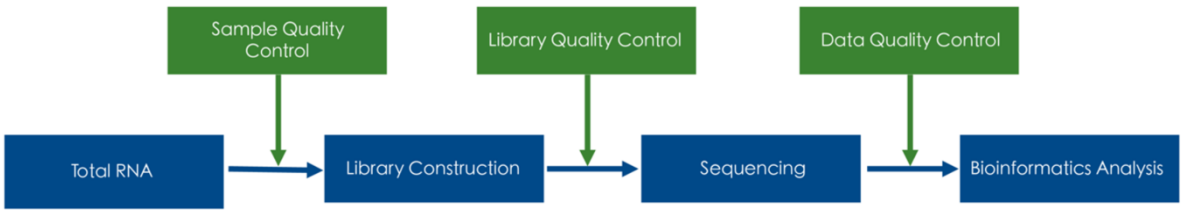


Figure 6: RNA sequencing flow diagram prepared by Novogene.

5.5.2 IDEP.96

These count table files were inputted into an excel sheet where they were organised and exported as a CSV file to IDEP.96 (integrated differential expression and pathway analysis) (http://bioinformatics.sdstate.edu/idep96/)[121]. From here I was able to analyse and create visual tables of the count files for individual genes of interest. In order to achieve this, I used the pre-process section to identify the relative read counts. I was also able to select target genes to map and compare between the different sample types. Prior to this I researched what genes would be appropriate to identify the effect on differentiation that the drugs had

on the samples[122]. I was also able to generate heat maps to compare sample types and use K mean analysis to identify clusters within the sample.

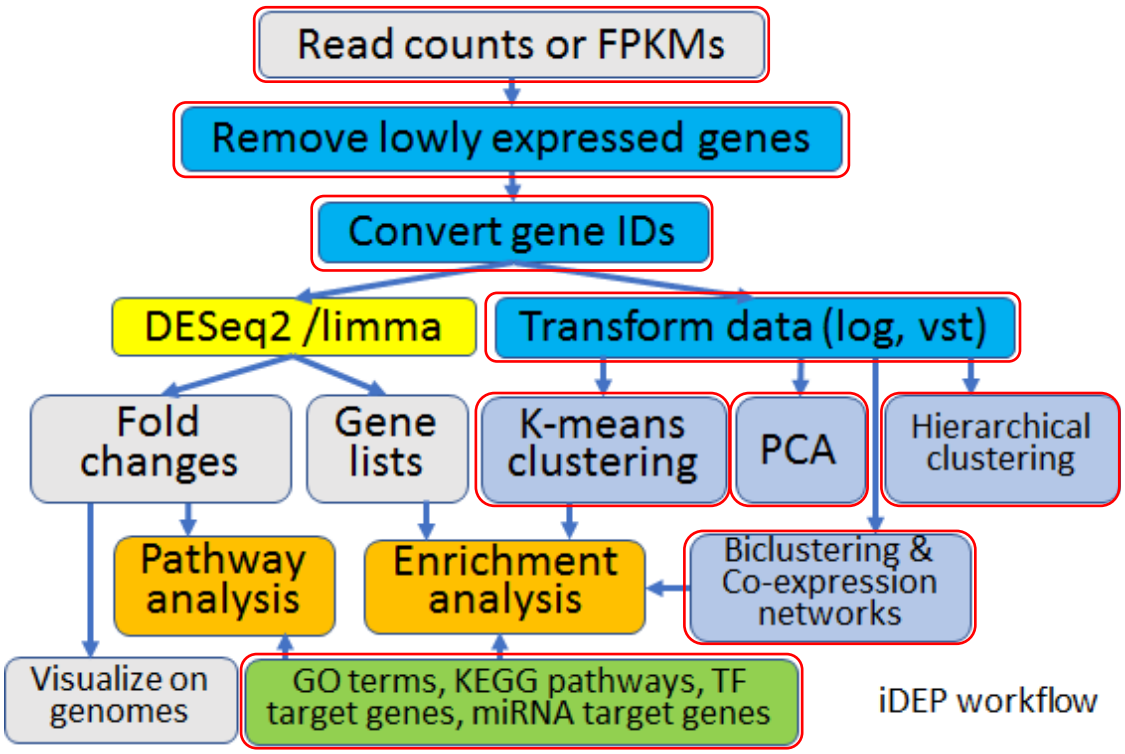


Figure 7: iDEP workflow. Its shows the multiple pathways that can be taken in analysis after the upload of read counts. Red outlines highlight the pathway I took when analysis my data in this software.

5.5.3 Additional Analysis

In additional to the use of galaxy, Brinda Balasubramanian generated read counts in a different way for comparison. She used similar tools in the use of FASTQC software (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), Trim Galore! Software (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore) [123] , HISAT 2 alignment tool (http://daehwankimlab.github.io/hisat2/)[124] and Feature counts function for the subread package (https://subread.sourceforge.net/) [125].

The data obtained from this pathway was compared to the one that I created and data generated here was used to analyse single target genes used later.

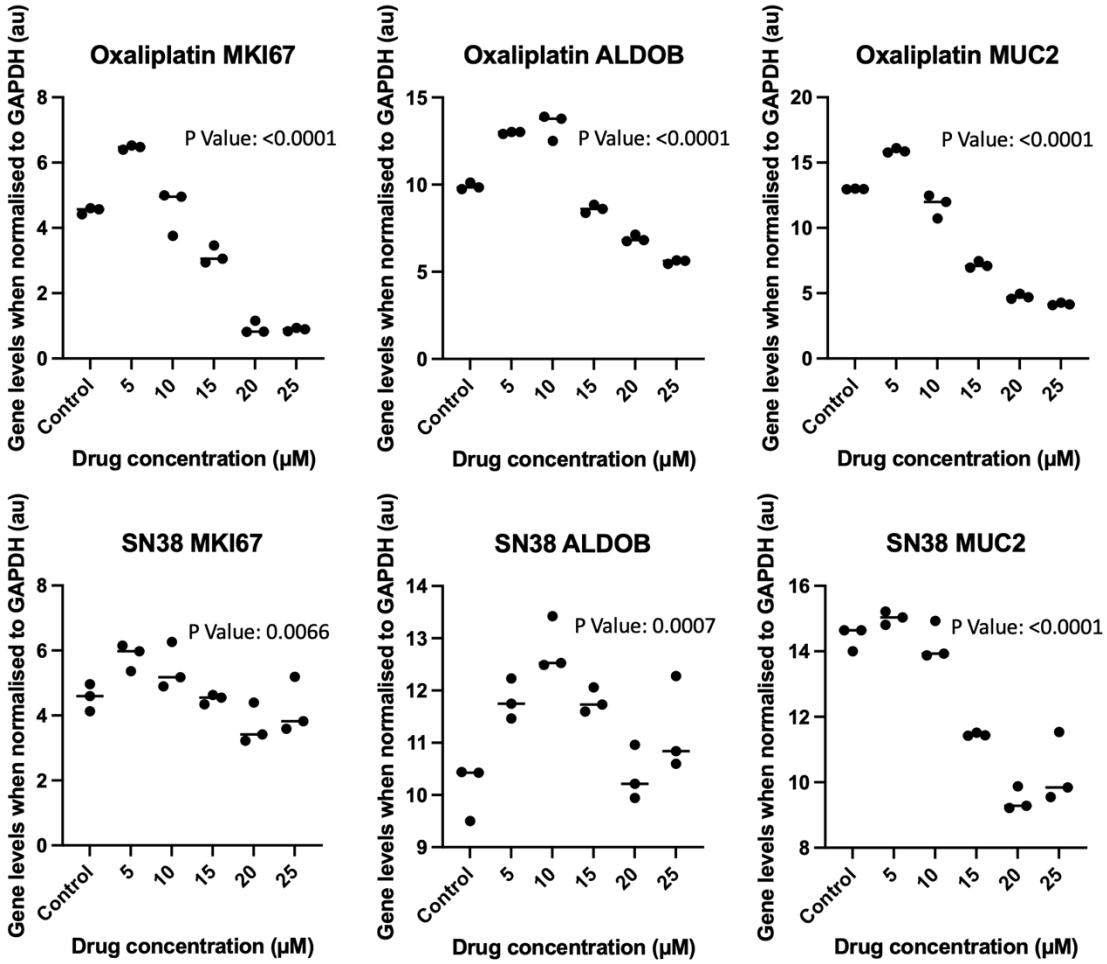


Figure 8: Graphs generated in Graph Pad Prism. Shows the distribution of gene analysis in a One-way ANOVA from pPCR compared to GAPDH housekeeper gene.

5.6 Statistics

qPCR data was analysed using Graph Pad Prism. All data sets were shown to be significant when p value was above 0.05 and not significant when the P value was <0.05. All data sets had normal distribution and a one-way ANOVA test was used to compare the data sets.

Table 6: *Reagent List with location of where to find them*

Reagent	Use	Source	Stock Solution	Concentration used
DMEM high glucose medium	Growth medium for cells	Gibco (15480564)	X1	X1

Pen-Strep	Antibiotic for medium	Gibco (10378016)	X100	X1
Fetal Bovine Serum (FBS)	Growth Supplement for medium	Sigma-Aldrich	10%	10%
DMSO	Freezing solution	Sigma-Aldrich	X1	10%
B-mercaptoethanol	Aids in RNA extraction via denaturing disulphide bonds in RNase	Sigma-Aldrich	X1	0.143M
Ethanol	Precipitates nucleaic acid out of solution in RNA extraction		99%	70%
Oligo dT 18 and 24 mix	Primers	Invitrogen	X1	0.02
dNTP mix	Nucelotides for cDNA structure	Invitrogen	X1	100nM
DTT	Prevents secondary structure reformation	Invitrogen	X1	0.1M
5X Reaction Buffer	Aids in first strand formation	Promega	X1	1.5mM
RNasin ribonuclease inhibitor	RNase inhibitor	Fisher BioReagents™	X1	20-40U/µL

RT Enzyme Superscript	Reverse transcriptase	Invitrogen (18064022)	X1	200U/µL
SYBR™ Green Master Mix	Amplification marker	ThermoFisher (A25742)	X2	
5-Flurouracil (5FU)	Chemotherapy Drug	Merck (F6627-5G)	Pure Powder	Varies
Oxaliplatin	Chemotherapy Drug	Merck (09512-5MG)	Pure Powder	Varies
7-Ethyl-10-Hydroxycamptothecin (SN-38)	Chemotherapy Drug	Merck (H0165-10MG)	Pure Powder	Varies

6 Objectives

Disease relapse and drug resistance to standard of care therapies continues to be an evolving problem in CRC patient care. Therefore, due to their drug resistive and tumour initiating nature, understanding the biology and function of CSC is vital in improving survival and care for patients burdened with this disease.

In this study the goal set were:

1. Test whether the maintenance of low confluence Caco-2 cells could produce a cell line with reduces differentiation.
2. Understand the effect of chemotherapeutic agents on the differentiation of Caco-2 cell lines.
3. Identify the effect of chemotherapeutic agents on specific pathways and compare these between differentiated and less differentiated cell lines.
4. Compare gathered data to previous research to identify similarities and differences.

7 Results

7.1 Generation of a low differentiated ("undifferentiated") colorectal cancer cell line

The Caco2 cell line is known to have the ability to undifferentiate when kept in low confluence for several passages[117]. In order to achieve this, the cells were seeded at a low confluence (30-50%) and maintained in this state with regular passages for 2 weeks. This was done to have a sample selection that would have a lower number of differentiated cells and presumably a higher amount of cancer stem and progenitor cells. RNA extraction and qPCR was used to quantify whether the lower differentiated nature had taken place. Previous studies demonstrated that ALDOB and ANPEP were two target genes that are expressed mainly in the enterocyte cell type[122]. Enterocytes are an epithelial cell type that lines the small and large intestine. They play a key role in digestion with specialisations such as villi that facilitate ion and nutrient uptake as well as protection from harmful bacterium[126]. I decided that this would serve as a viable benchmark to highlight the differentiated nature of the cell type. If the expression level of these genes is lower than the parental cell line, then that would correlate to a lower yield of differentiated cell type thus demonstrating an undifferentiated cell population. The qPCR results showed that in the cell pool kept in lower confluence for 10 passages, the expression levels of both ALDOB and ANPEP were lower than those kept at the higher confluence (**Figure 9**). From this it was concluded that the cells kept in a lower confluence expressed a lower number of differentiated enterocyte cells than those that had not.

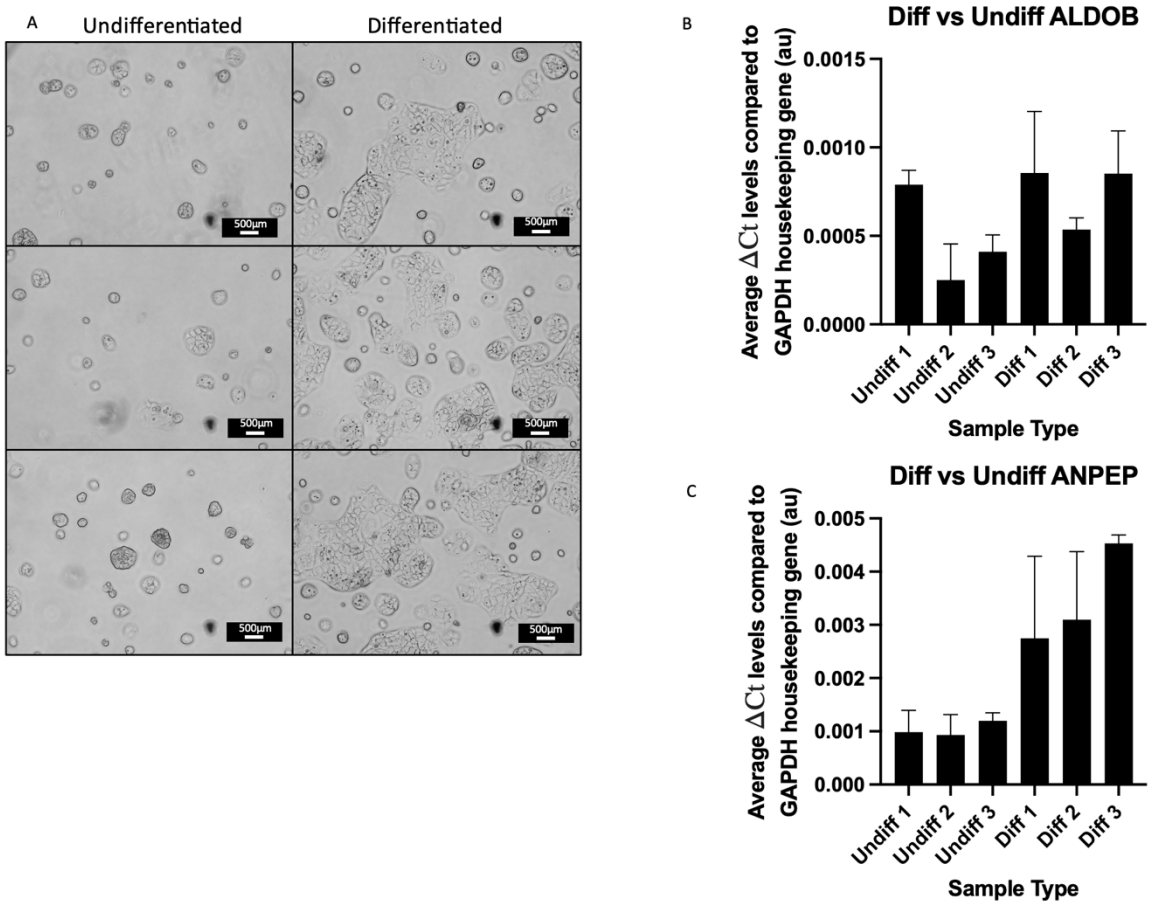


Figure 9: A; Cell images from Caco2 cell lines. Cells that were desired to be less differentiated were kept at a lower confluence in order to maintain an undifferentiated phenotype. **B,C;** qPCR results on the differentiated and undifferentiated cell lines. Genes of interest are ALDOB and ANPEP respectively with GAPDH used as a housekeeping gene to normalise results. Graphs show the changes in the normalised gene expression levels within the standard grown Caco2 cell lines (Diff 1, 2 and 3) and the custom generated undifferentiated Caco2 cell lines (Undiff 1, 2 and 3). Error bars show gene expression difference between each replicate. For each set n=3. 500µm scale bar.

7.2 Effects of chemotherapeutic agents on the undifferentiated phenotype of Caco2 cells

To further assess the viability of the statement that stem cells are the population responsible for allowing tumours to acquire chemotherapy resistance, undifferentiated Caco2 cell lines were treated with several drugs used in colorectal cancer treatment in the clinics. These drugs were oxaliplatin and 7-ethyl-10-hydroxycamptothecin (SN-38). Drug concentrations and timings were obtained from previous work as mentioned earlier. After administration of drugs, the cells were harvested and analysed via qPCR. Photos were taken at the time of drug administration and 24 and 48 hours after. In the oxaliplatin drug group, cell photos that showed the higher concentrations (20-25 µmol/ml) of drugs caused mass cell death and breaking apart of cell clusters (**Figure 10a**). Whilst amongst the SN-38 treated

group there was fewer cell clusters to begin with, the higher drug concentrations were still shown to stop them from forming later as seen in the T48h control group. Additionally higher concentrations (20-25 µmol/ml) of SN-38 were shown to cause mass cell death as with the oxaliplatin (**Figure 10b**).

The first gene of interest examined by qPCR was ALDOB. This is due to ALDOB being an enterocyte marker gene as mentioned above[127]. This was undertaken to assess the effect of the differentiation of the cell populations when introduced to chemotherapeutic stress. In the oxaliplatin group, ALDOB gene expression level was seen to increase with the higher levels of drug concentrations. This is showing that cell death is observed and cells present have a more differentiated state (**Figure 11a**). In the SN-38 gene population, there was seen to be a steep drop on ALDOB expression with the introduction of the drug but then towards the higher concentrations (20-25 µmol/ml) it increases again though not to the levels of the control group (**Figure 11e**). This would suggest that low levels of drug cause the populations to shift to an undifferentiated phenotype, whereas the higher levels seem to force a differentiation of the cell population.

The second gene investigated was MUC2. Previous studies identified MUC2 to be a Goblet cell marker gene[122]. Goblet cells are another type of differentiated cell found within the intestines and are responsible for the secretion of mucin to create a protective mucus layer in order to protect the intestine from harm[128]. This gene was chosen for investigation to act as a secondary identifier of cellular differentiation. In the oxaliplatin drug group, a similar pattern was seen as within the ALDOB group, suggesting again that drug treatment causes cells to differentiate (**Figure 11b**). In the SN-38 group, there was seen to be a reduced amount of MUC2 in the control group but still followed a similar trend towards the higher drug concentrations (20-25 µmol/ml) (**Figure 11f**). This also suggests that drug administration pushes the cells towards a more differentiated cell type.

The next target gene searched for was MKI67. This gene is a proliferation marker so was used to judge how effective the drugs were by measuring cell proliferation[129]. The oxaliplatin group showed a gradual increase in expression and then a much greater increase when the drug concentration reached 20 µmol/ml (**Figure 11c**). In the SN-38 group, MKI67

showed an initial decrease from the control group before a great increase at the same concentration mentioned above (**Figure 11g**).

Finally, SOX9 was tested aa it is a progenitor and Wnt pathway marker gene. This was only done in the oxaliplatin drug group. SOX9 is a progenitor marker so can be used to infer a less differential nature of the treated cells[130]. Conversely, compared to the ALDOB and MUC2 gene types, SOX9 showed a large increase in the higher drug concentrations (25 µmol/ml) which suggests a greater undifferentiated phenotype and an upregulation of Wnt pathway in these cells (**Figure 11d**).

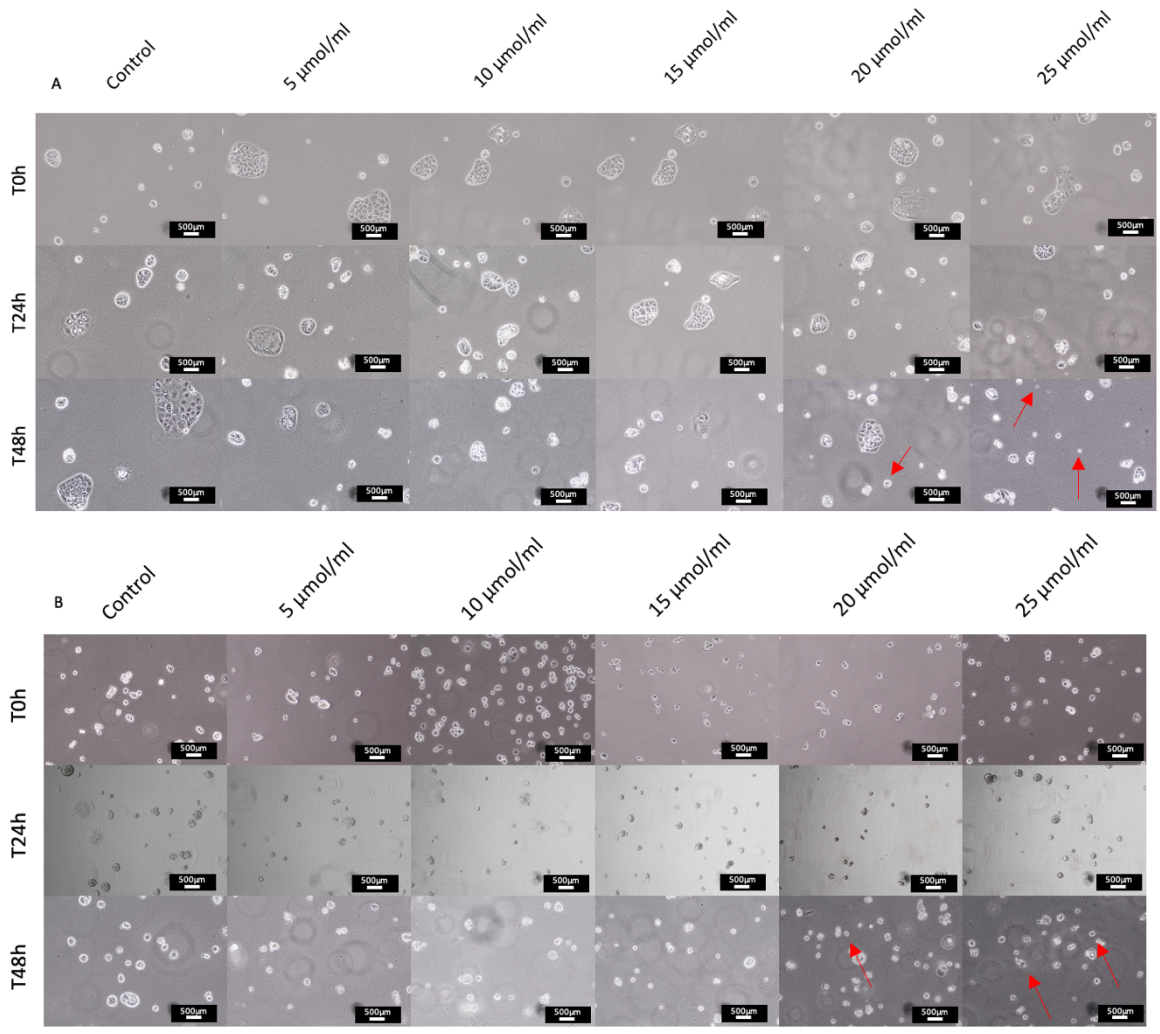


Figure 10: A; Cell images from Caco2 cell lines treated with oxaliplatin. Cell samples were generated to be the undifferentiated type as described earlier. They were treated and left for 48 hours but photos were taken at 24 hours. Red arrows highlight dying cells at higher concentrations. **B;** Cell images from Caco2 cell lines treated with SN-38. Cell samples were generated to be the undifferentiated type as described earlier. They were treated and left for 48 hours but photos were taken at 24 hours. Red arrows highlight dying cells at higher concentrations. Scale bar at 500µm

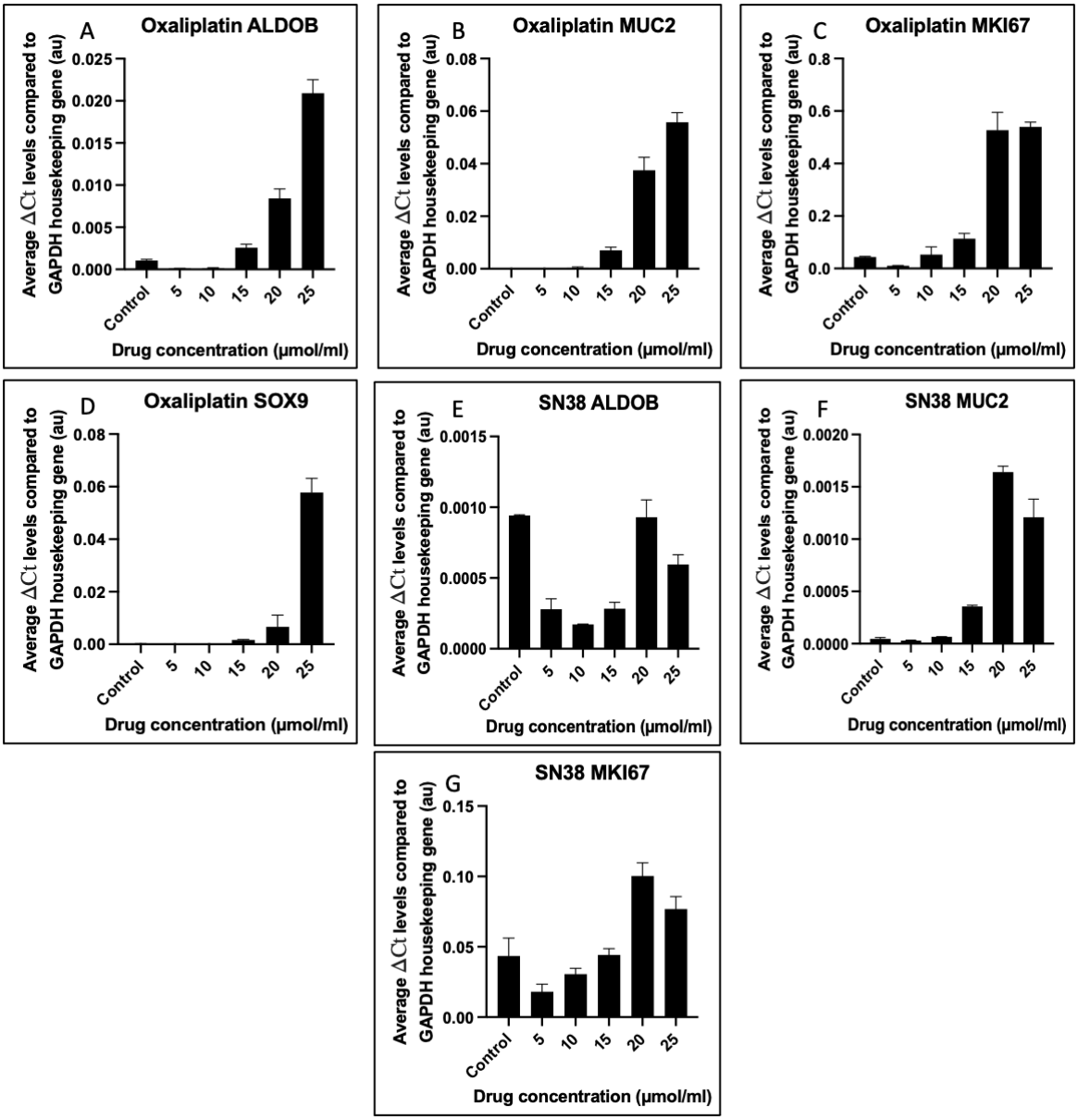


Figure 11: A; qPCR results from oxaliplatin experiments on undifferentiated cells line. Gene of interest is ALDOB. **B;** qPCR results from oxaliplatin experiments on undifferentiated cell line. Gene of interest is MUC2. **C;** qPCR results from oxaliplatin experiments on undifferentiated cell line. Gene of interest is MKI67. **D;** qPCR results from oxaliplatin experiments on undifferentiated cell line. Gene of interest is SOX9. **E;** qPCR results from SN-38 experiments on undifferentiated cells line. Gene of interest is ALDOB. **F;** qPCR results from SN-38 experiments on undifferentiated cell line. Gene of interest is MUC2. **G;** qPCR results from SN-38 experiments on undifferentiated cell line. Gene of interest is KI67. All graphs show the ΔCT value for each gene and each concentration. In all instances the house keeping gene it is compared to was an individual plot of GAPDH. Error bars show the variation between the replicates on each data point. In all instances N=3 500μm scale bar.

7.3 Gene expression profile of undifferentiated vs differentiated subtypes: analysing the raw data obtained from RNA sequencing.

This study then analysed the gene expression profile of these two subtypes of Caco2 cells treated with 5FU, SN-38 and a combination of both. RNA was extracted, checked for quality (RIN) and sent to Novogene to perform RNA sequencing. As previously mentioned, the raw

data obtained from the RNA sequencing was analysed using the web-based platform Galaxy. This was chosen due to its user-friendly interface and there being existing knowledge of using this software amongst my colleagues. 6 samples were analysed:

A1- Undifferentiated cell lines treated with both 5FU and SN-38 (UNDIFF 5FU and SN-38)

B1- Undifferentiated cell line treated with just 5FU (UNDIFF 5FU)

C1- Undifferentiated cell line treated with just SN-38 (UNDIFF SN-38)

D1- Differentiated cell lines treated with both 5FU and SN-38 (DIFF 5FU and SN-38)

E1- Differentiated cell line treated with just 5FU (DIFF 5FU)

F1- Differentiated cell line treated with just SN-38 (DIFF SN-38)

Whilst there were replicates taken from the cell line samples and additional controls, unfortunately due to time constraints (as mentioned in the disclaimer at the beginning), only one set of replicates were investigated from A1 to F1 (N=1).

7.3.1 *Using Galaxy to check data quality*

Steps taken during this process can be found in the methods section. **Figure 12 and Figure 13** contain *FastQC* and *MultiQC* analysis of data imported into Galaxy. All data sets were within the green section when the programs were ran showing that the data is of good quality.

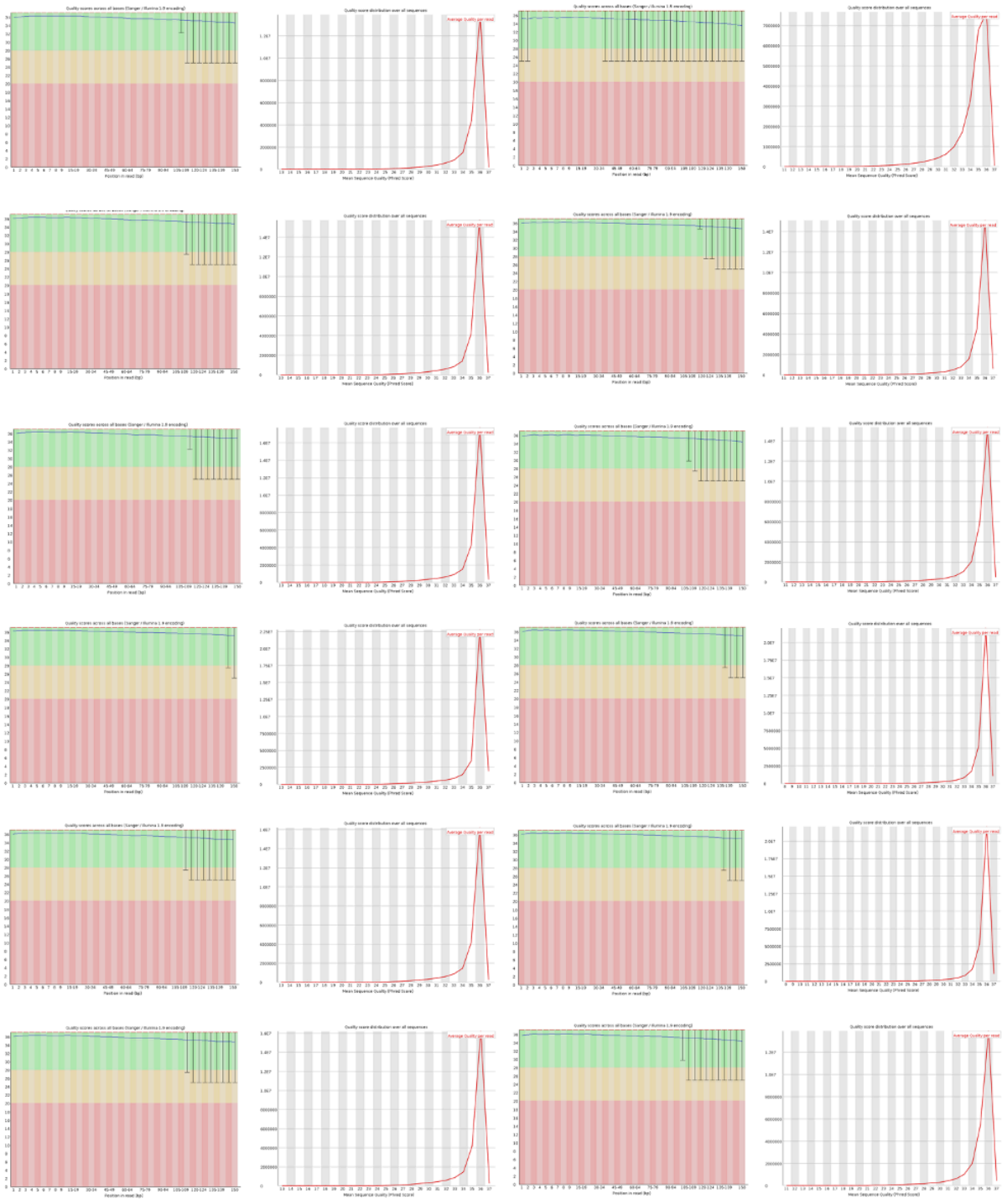


Figure 12: *FastQC* file analysis on samples included in the experiment. On the left is the per base quality reach for each sample and on the right is per sequence quality score graph for each sample. From top to bottom is sample A1-F1 with sample 1 on the left and sample 2 on the right.

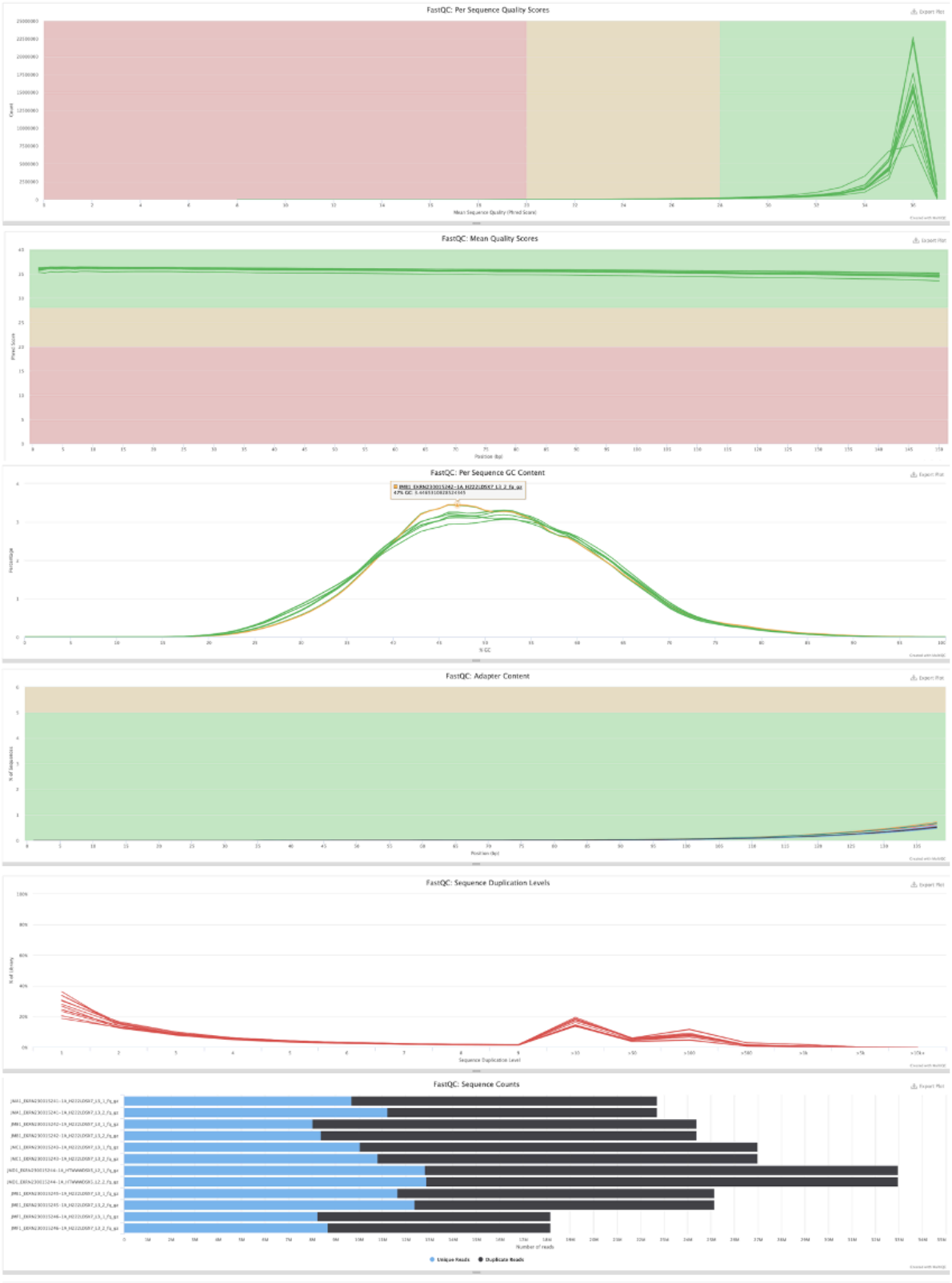


Figure 13: *MultiQC* analysis of the *FastQC* files. It shows from top to bottom: the adapter content, sequence GC content, per sequence quality score, sequence counts, mean quality scores and the sequence duplication levels.

7.4 Pre-processing the raw data obtained from Galaxy

Using the iDEP software the read counts file was uploaded as previously described. Firstly, the number of reads per sample was checked along with how this data was distributed. Then a scatter plot was generated for the three data sets that were linked by the drug which were used to treat the cells (A1-D1, B1-E1, C1-F1) (**Figure 14a, 14e. 14f, 14g**). Additionally, a principal component analysis (PCA) plot was generated which showed that the samples treated with the same drugs as being close together. The single treatment groups were seen to be close together between the undifferentiated and differentiated group suggesting a closer correlation. However, the group treated with both drugs showed a greater variation between the 2 cell types (**Figure 14d**).

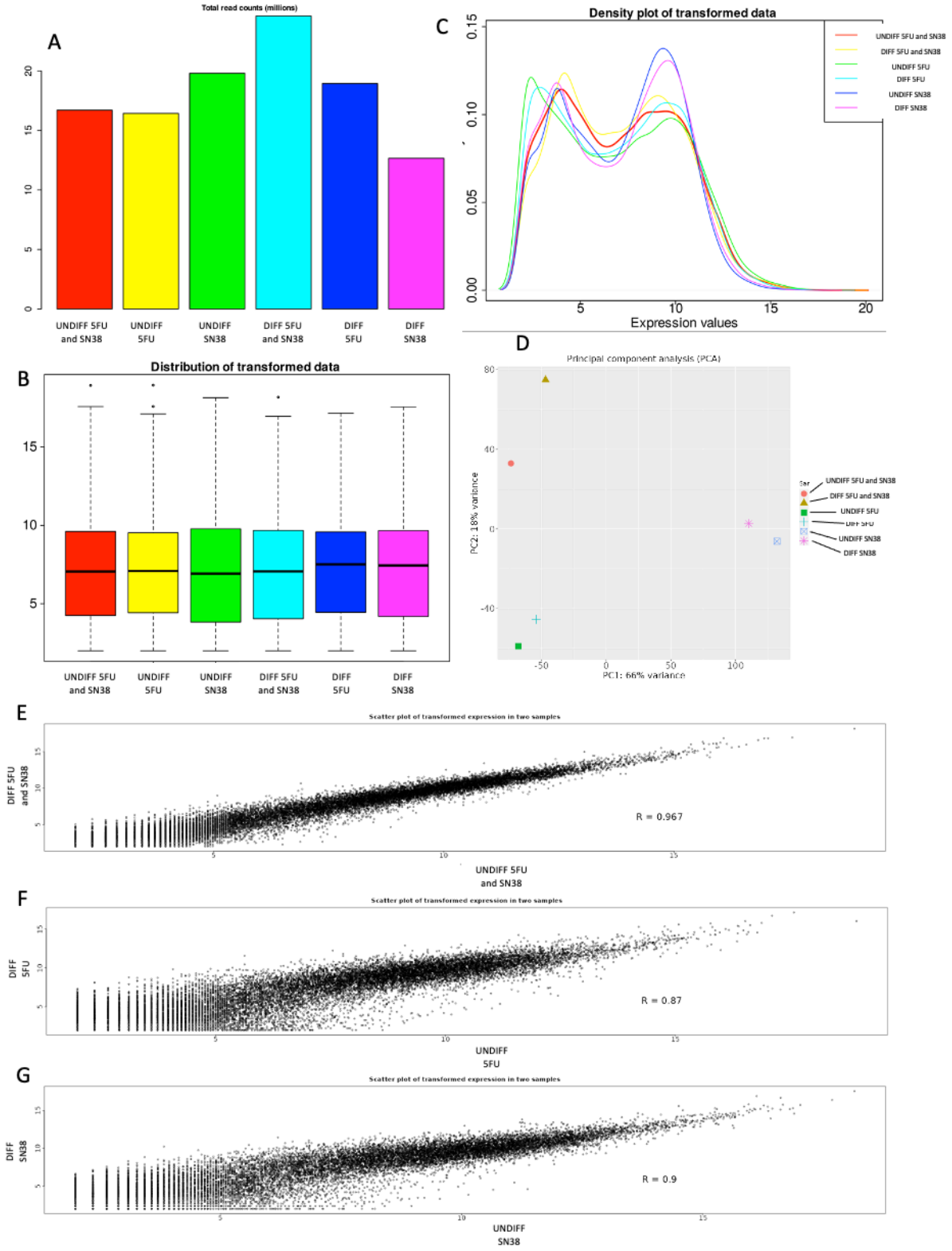
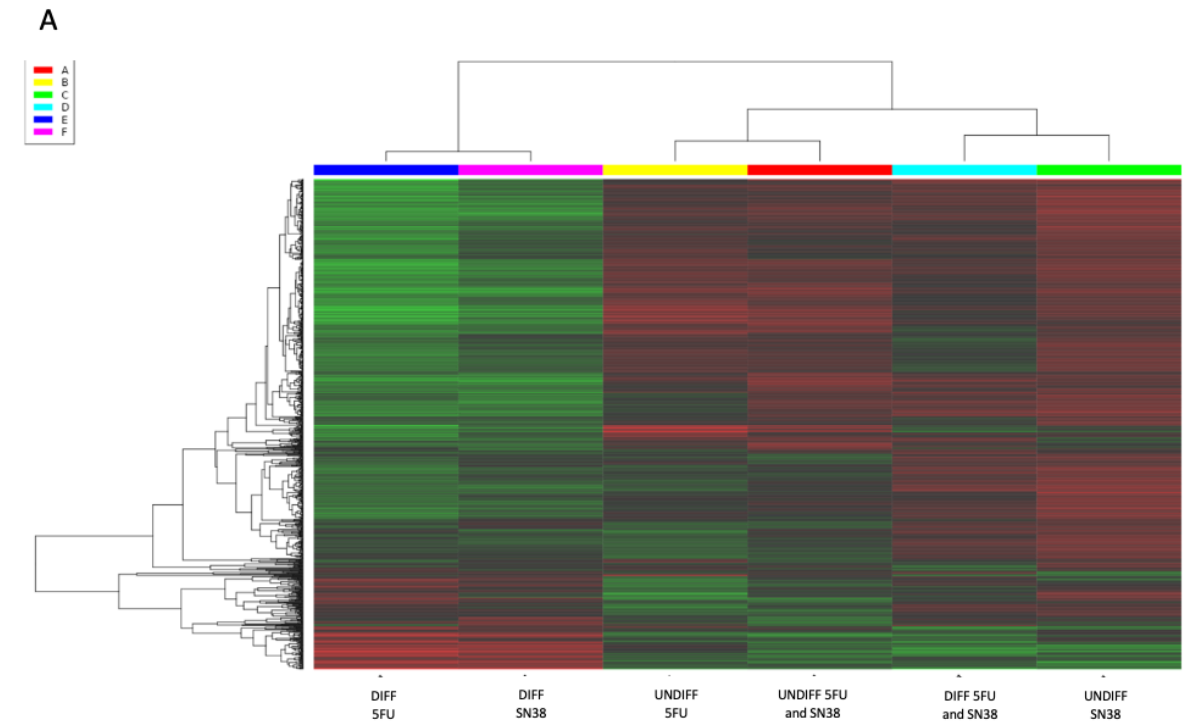
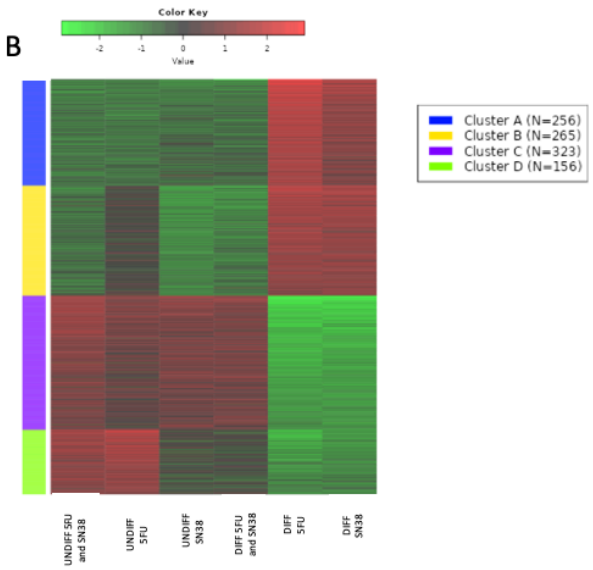


Figure 14: Pre-processed raw data using the IDEP.96 software. **A, B, C;** Graphs to show the total reach counts of each sample and the distribution of the data points. **D;** PCA plot with principle X axis as 1 and y axis as 2. **E, F, G;** Scatter graphs that show R correlation comparisons for each data set. It compares A to D, B to E and C to F as they were treated with the same drugs. Scatter plots showed undifferentiated groups as the differentiated showed similar results.

7.5 Gene expression analysis

To study the effect of the chemotherapy drug on the gene expression levels between differentiated and undifferentiated cells a heatmap was generated with each sample for the reads per gene. Interestingly, the samples clustered together but there was also another cluster between the groups (**Figure 15a**). The k-means global showed the generation of 4





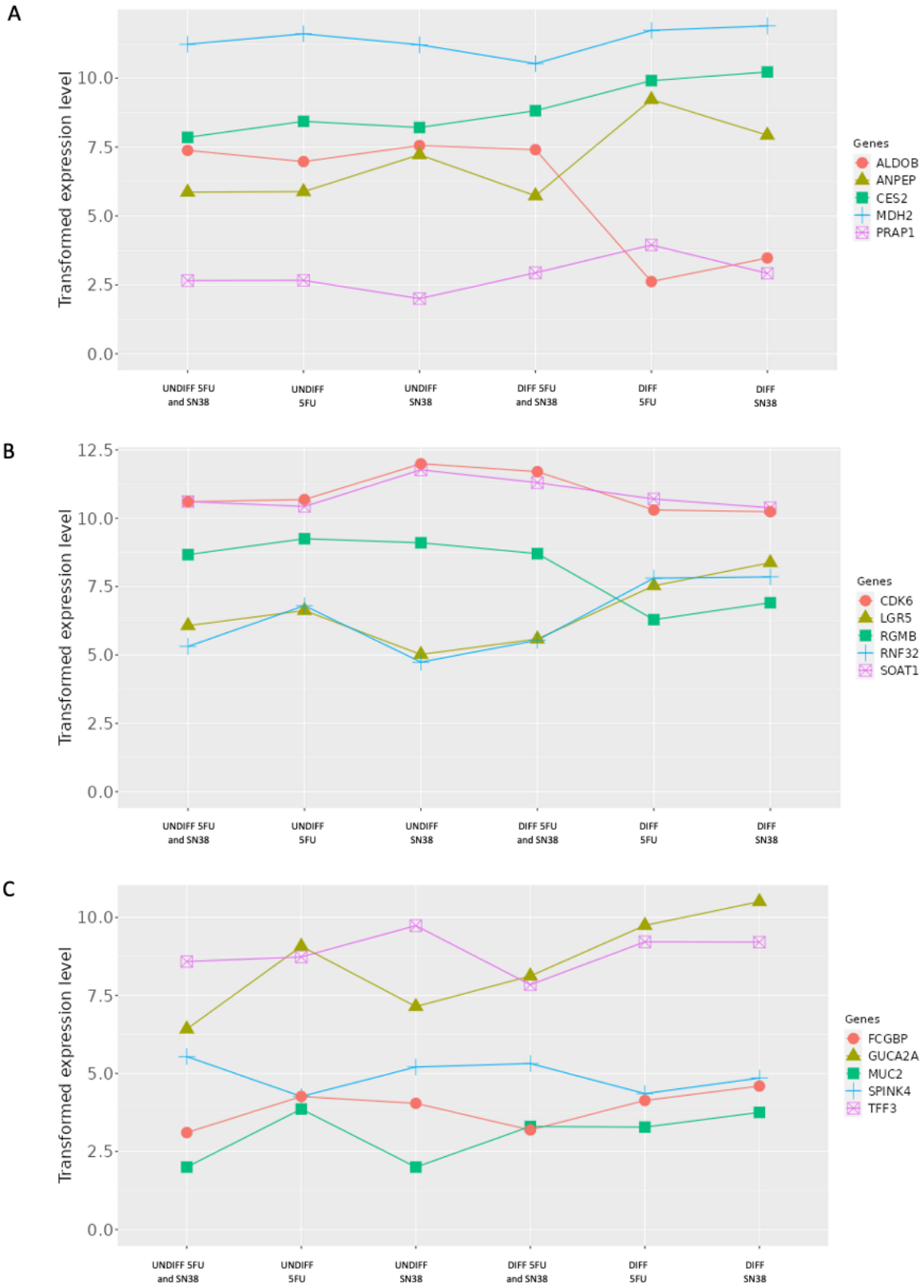
Enriched pathways for each cluster

Cluster	adj.Pval	nGenes	Pathways
A	8.7e-04	18	Small molecule catabolic process
	1.8e-03	13	Carboxylic acid catabolic process
	1.8e-03	19	Organic hydroxy compound metabolic process
	1.9e-03	13	Organic acid catabolic process
	5.5e-03	39	Small molecule metabolic process
D	4.8e-06	6	DNA replication-dependent nucleosome assembly
	1.6e-05	6	RDNA heterochromatin assembly
	1.6e-05	5	Negative regulation of megakaryocyte differentiation
	1.6e-05	6	Nucleolar chromatin organization
	3.7e-05	6	Nucleolus organization
	3.9e-05	13	Response to steroid hormone
	4.0e-05	8	Nucleosome assembly
	4.0e-05	20	Response to organic cyclic compound
	4.0e-05	5	Depurination
	4.0e-05	7	Heterochromatin organization
	4.0e-05	5	Regulation of histone exchange
	4.0e-05	5	Positive regulation of histone exchange
	6.0e-05	8	Regulation of gene silencing by miRNA
	6.4e-05	6	Regulation of androgen receptor signaling pathway

Figure 15: A; Heatmap to show the clustering of the 6 samples. **B;** k-mean analysis of samples showing which are the 4 clusters when the top 1,000 genes are selected as a variable. The left shows a table containing the most abundant pathways within the A and D cluster. Data shows normalised count values when compared to housekeeper gene.

clusters with a highlight in the A and D cluster. The A cluster contained many pathways included in metabolic process and the D cluster included a host of different pathways such as chromatin and nucleus organisation as well as histone exchange regulation (**Figure 15b**).

It was then decided that this project was to investigate the effect on specific target genes to see how the chemotherapy drugs effect these pathways. Using the reference[122] mentioned before genes related to specific pathways in intestinal cells were investigated and selected a few of them that demonstrated expression in these samples[122]. The project then investigated enterocyte cell markers (ALDOB, ANPEP, CES2, MDH2, PRAP1) (**Figure 16a**), intestinal stem cell markers (CDK6, LGR5, RGMB, RNF32, SOAT1) (**Figure 16b**), Paneth cell markers (GUCA2B, HABP2, ITLN1, PLA2G2A, Wnt3) (**Figure 16c**) and Goblet cell markers (FCGBP, GUCA2A, MUC2, SPINK4, TFF3) (**Figure 16d**). Then finally Wnt pathway gene markers (AXIN2, BTRC, CTNNB1, EPHB3, RARA) (**Figure 16e**) was investigated. Goblet cells were investigated to see whether the cells are differentiating into specific cell types or remaining undifferentiated.


A
Transformed expression level
10.0
7.5
5.0
2.5
0.0
UNDIFF 5FU and SN38
UNDIFF 5FU
UNDIFF SN38
DIFF 5FU and SN38
DIFF 5FU
DIFF SN38
Genes
ALDOB
ANPEP
CES2
MDH2
PRAP1
B
Transformed expression level
12.5
10.0
7.5
5.0
2.5
0.0
UNDIFF 5FU and SN38
UNDIFF 5FU
UNDIFF SN38
DIFF 5FU and SN38
DIFF 5FU
DIFF SN38
Genes
CDK6
LGR5
RGMB
RNF32
SOAT1
C
Transformed expression level
10.0
7.5
5.0
2.5
0.0
UNDIFF 5FU and SN38
UNDIFF 5FU
UNDIFF SN38
DIFF 5FU and SN38
DIFF 5FU
DIFF SN38
Genes
FCGBP
GUCA2A
MUC2
SPINK4
TFF3

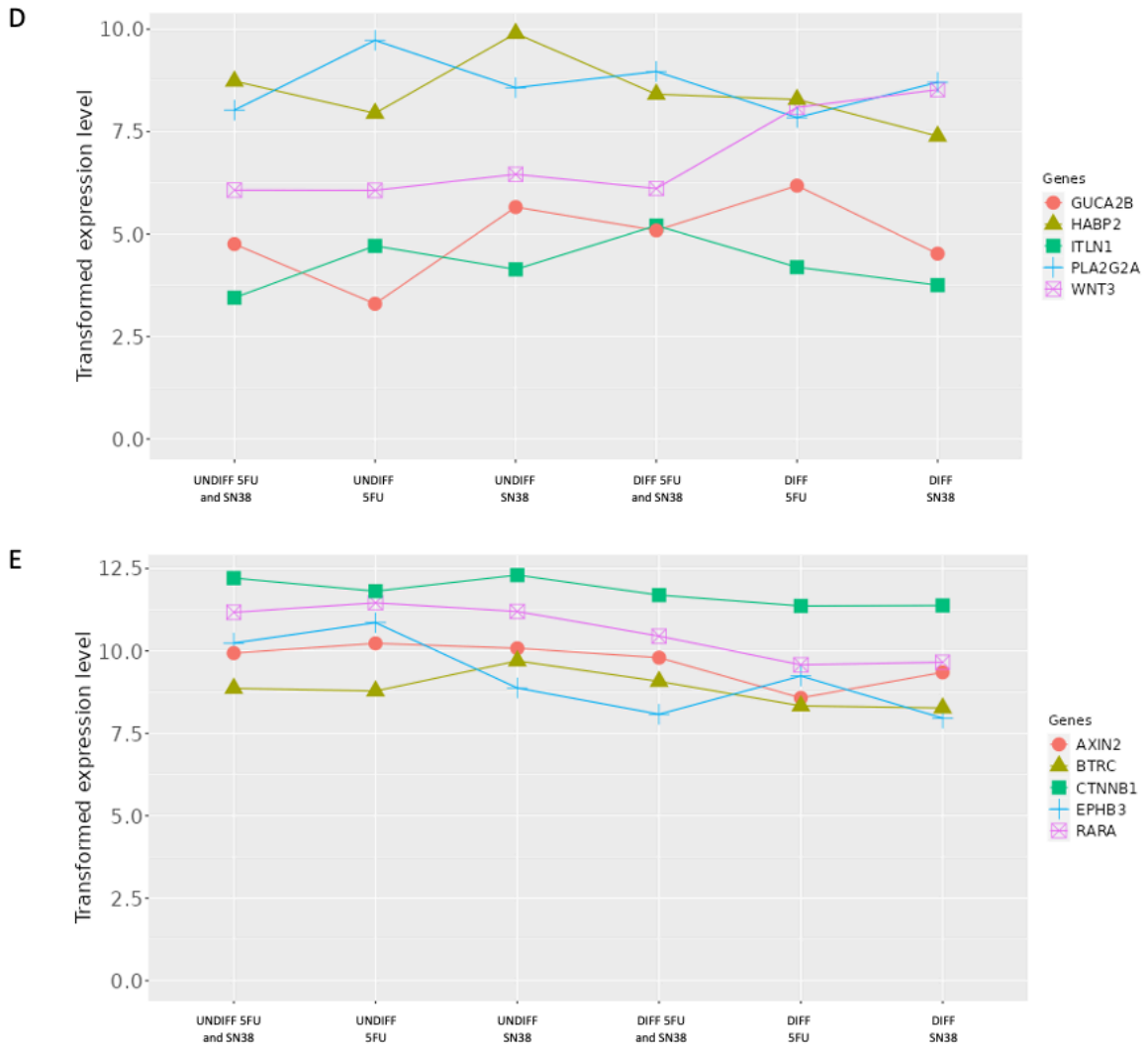


Figure 16: Individual gene of interest expression levels for the 6 markers. **A;** Enterocyte markers. **B;** Intestinal Stem Cell markers. **C;** Goblet cell markers. **D;** Paneth cell markers **E;** Wnt pathway markers. Transformed expression levels is the log expression levels of normalised data.

7.6 Analysis of independent drug treated groups

After, samples were separated into 3 groups dictated by which drug the cells were treated with. The aim of this was to have a closer look into the differential effect of the drugs in the specific cell markers in each group. I chose a gene target from each of the graphs above (**Figure 16**) and plotted them for each of the pairs.

7.6.1 Investigating specific genes of interest

In the dual drug group, the graphs showed that the differentiated cell type showed a higher expression level of specific cell markers (such as the Goblet and Enterocyte gene markers) but a lower level of LGR5, an intestinal stem cell marker (**Figure 17a**). This suggests that the

differentiated group remained more differentiated than the undifferentiated cell type with lower number of intestinal stem cells. Additionally, the differentiated cell type had a lower level of AXIN2, a Wnt pathway marker. This would also suggest a lower level of intestinal stem cells within this population. In the samples treated with only 5FU and only SN-38, there were some differences to the other samples. The first similarity is higher levels of CES2 and GUCA2A2. These are the enterocyte and Paneth cell markers which suggests a higher proportion of differentiated cell types within the differentiated sample group. Converse to the first sample type, both samples had high levels of LGR5 in the differentiated cell type. This implies a higher proportion of intestinal stem cells and/or Wnt pathway activation in these samples. This is weakened however, by the fact that there is lower level of Wnt pathway markers in these samples. The final difference is the lower levels of ITLN1 gene. This demonstrates a lower level of Goblet cell markers which is converse to what we observed in the first sample type (**Figure 17b, 17c**).

To investigate further, graphs were generated of multiple genes encoding for enterocytes (ALDOB, ANPEP, CES2, MDH2, GSTA4) and Wnt pathway markers (AXIN2, BTRC, CTNNB1, EPHB3, RARA) which were the groups most interesting to study. In addition, on the Wnt pathway graph, LGR5 was also searched for as it is a well-established intestinal stem cell gene in vivo (**Figure 18**). Regarding the enterocyte gene expression, the results varied between the three groups. In the dual drug group, the variation was minimal with many genes, including ALDOB, ANPEP and GSTA4, having little difference between the two groups. CES2 showed a more pronounced difference with an upregulation in the differentiated group, whereas MDH2 showed a slight increase in the undifferentiated group (**Figure 18a**). Both the individual drug treated groups showed a trend in ANPEP, CES2 and MDH2 with an upregulation in the differentiated group. ALDOB, however, was seen to be upregulated dramatically in the undifferentiated cell group. Regarding the GSTA4 gene, there was seen to be little overall change compared to the other target genes but was upregulated in the 5FU group and downregulated in the SN-38 group (**Figure 18c, 18e**). Wnt pathway genes showed similarities across all three groups with an upregulation in the undifferentiated cell groups (**Figure 18b, 18d. 18f**). Conversely, LGR5 showed variations across the three groups. In the dual drug treated group, LGR5 was upregulated in the undifferentiated cell group.

However, in the single drug treated groups it was shown to be greatly downregulated in the undifferentiated group (**Figure 18b, 18d. 18f**).

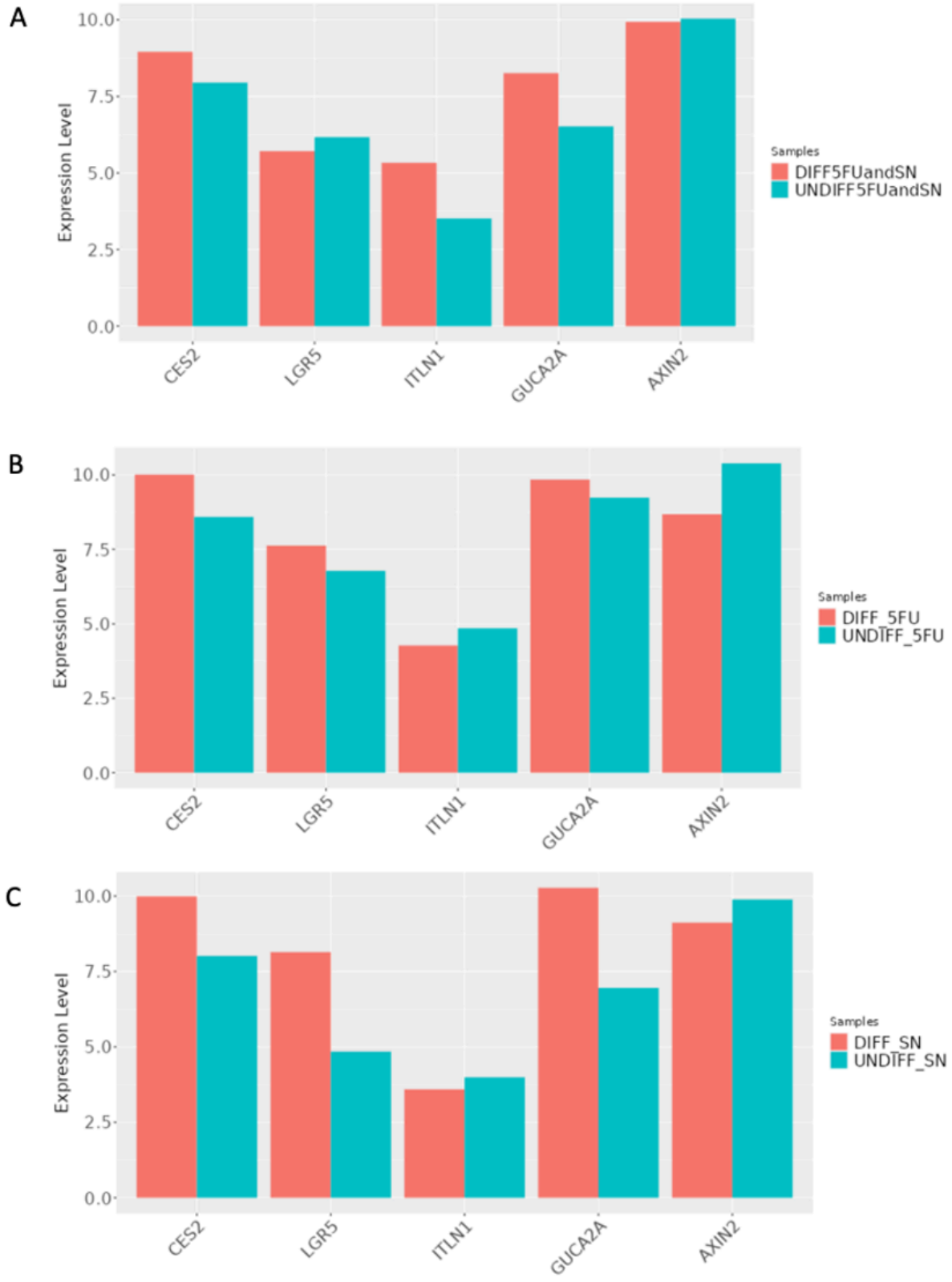
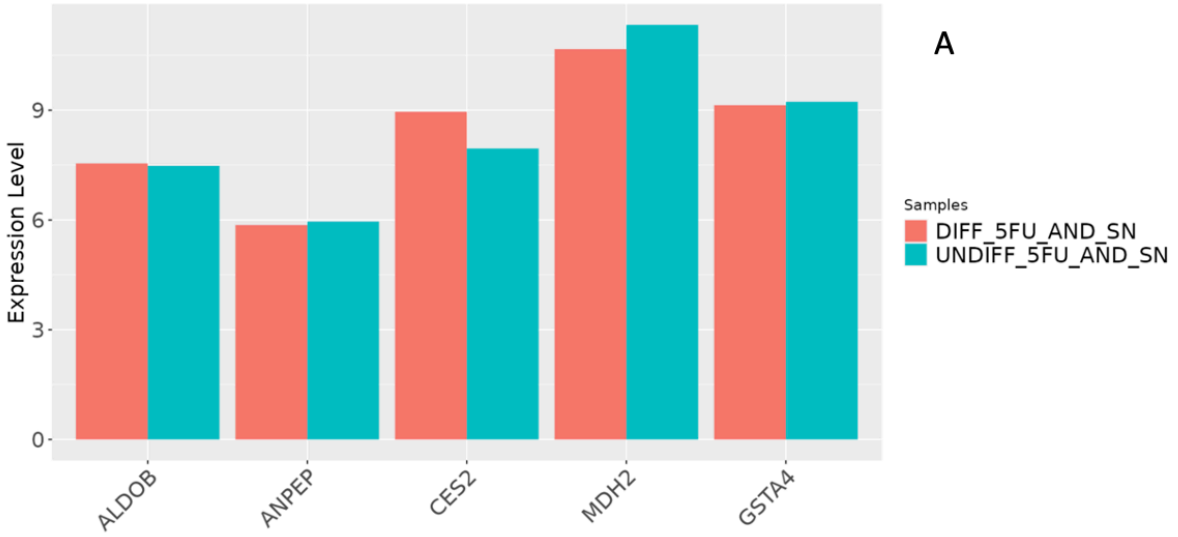
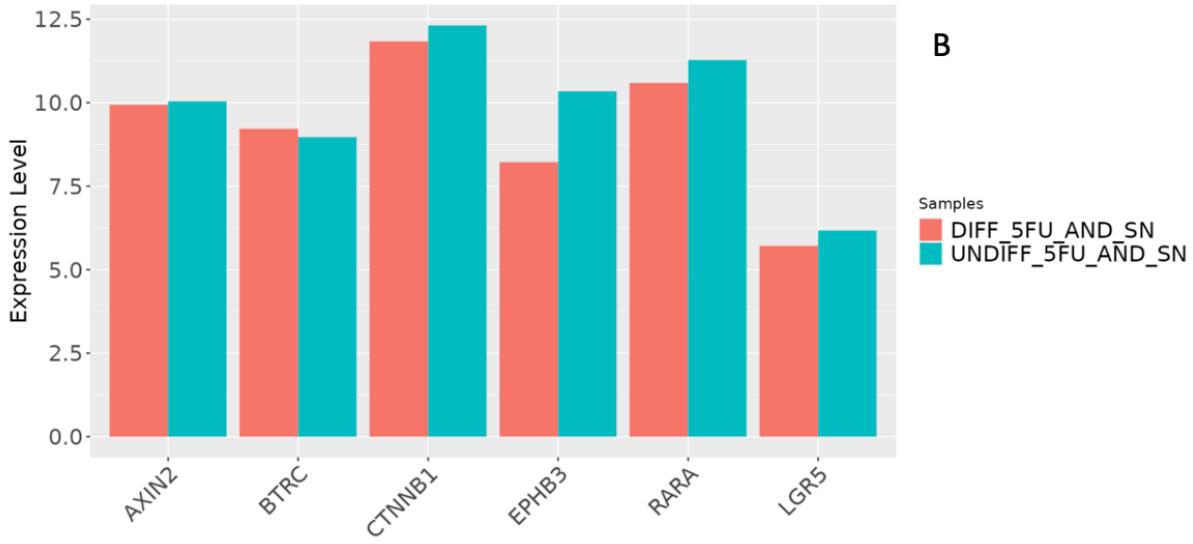
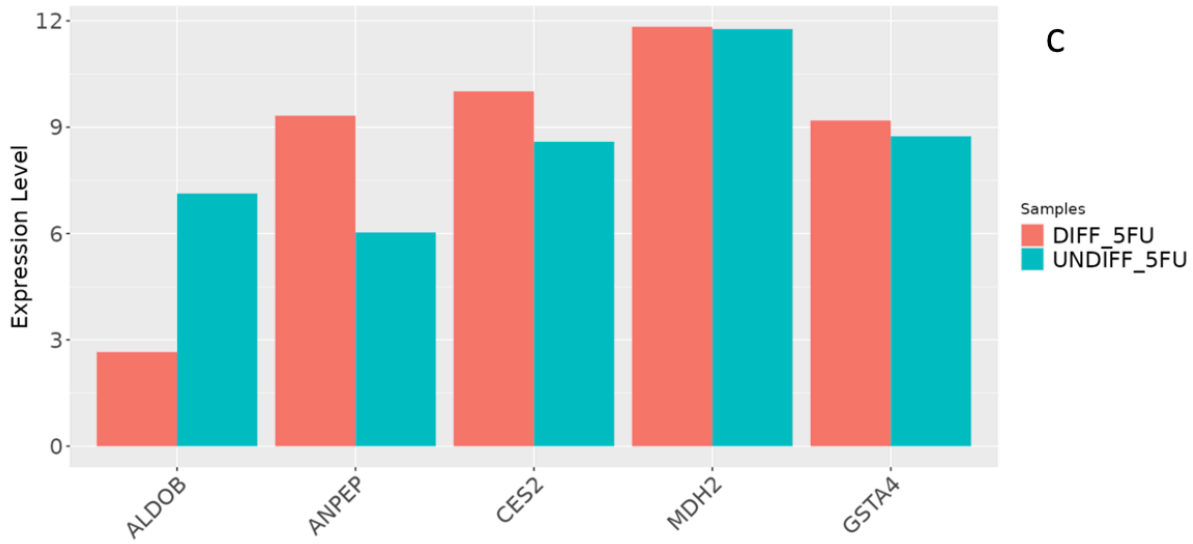


Figure 17: Graphs generated in IDEP.96 comparing sample types treated with the same drugs. **A;** 5FU and SN-38. **B;** 5FU. **C;** SN-38. Each graphs contains one gene relating to each of the graphs previously generated. Expression level showing the degree of upregulation of normalised data set compared to base data set.


A
Expression Level
0
3
6
9
ALDOB
ANPEP
CES2
MDH2
GSTA4
Samples
DIFF_5FU_AND_SN
UNDIFF_5FU_AND_SN


B
Expression Level
0.0
2.5
5.0
7.5
10.0
12.5
AXIN2
BTRC
CTNNB1
EPHB3
RARA
LGR5
Samples
DIFF_5FU_AND_SN
UNDIFF_5FU_AND_SN


C
Expression Level
0
3
6
9
12
ALDOB
ANPEP
CES2
MDH2
GSTA4
Samples
DIFF_5FU
UNDIFF_5FU

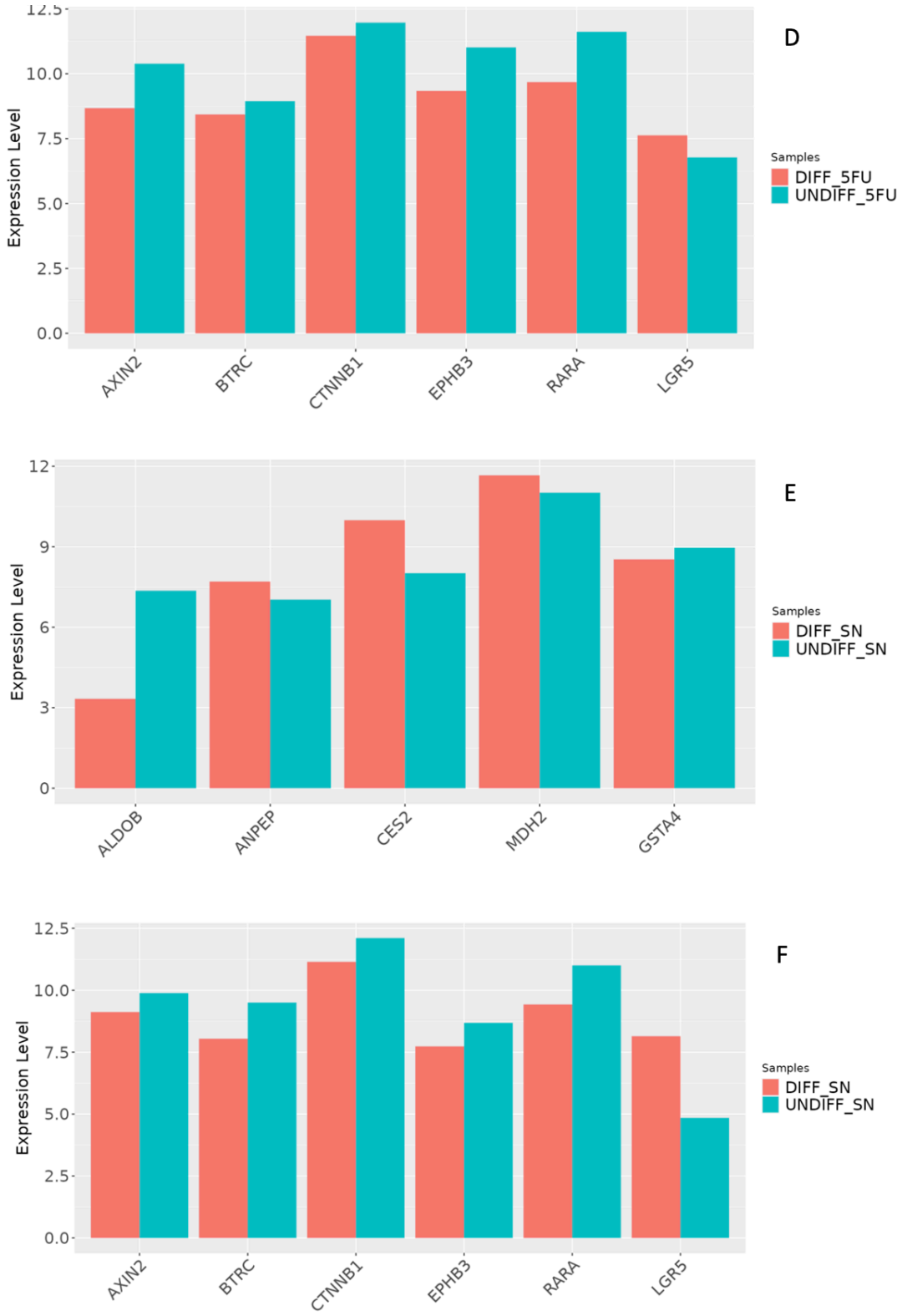


Figure 18: Graphs generated in IDEP.96 expanding on sample types treated with the same drugs. **A, C and E;** Highlights genes encoding for enterocytes in CRC. **B, D and F;** Highlights genes encoding for Wnt pathway as well as LGR5, a stem cell gene marker. Expression level showing the degree of upregulation of normalised data set compared to base data set.

7.6.2 Generation of heatmaps and k means analysis

The final step taken with the data generated was to explore what effect the cell differentiation had on different pathways when treated with drugs. Continuing to use the separated files, dictated by the type of drug administered, additional heatmaps were generated and K means data was re-ran. Unfortunately, as explained in the disclaimer, the lack of replicates meant that it was not possible to analyse using DeSEQ2 but this software was used later. In order to get a better idea of which genes were upregulated or downregulated, the number of shown genes on the heatmap was reduced to 50. This resulted in only the top 50 genes that were most significantly different between the two groups would be shown. Then, using the K means tool, it reduced the number of enrichments to around 1,000-1,200 to only show the most significantly enriched pathways (**Figure 19**).

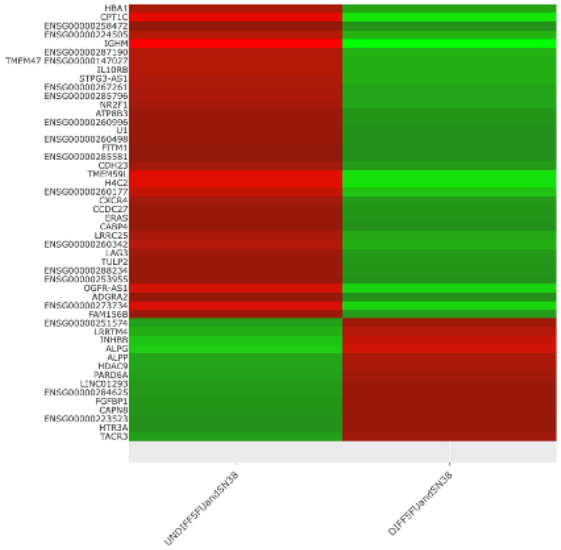
In the 5FU and SN-38 combined group, the cluster with the greatest enrichment was cluster C (**Figure 19c**). This cluster showed pathways controlling the response to chemokines and G protein coupled responses. These pathways are most noted for their ability to control and stimulate the migration of cells and an EMT shift[131]. Upregulation of these can be attributed to the fact that this data set is obtained from the undifferentiated cell lines, demonstrating that in chemotherapeutic stress, the cells begin to reverse into a cell type with a greater stemness phenotype.

With the administration of just 5FU, three clusters appeared to be enriched (**Figure 16f**). The first, cluster A, contained few pathways but they were attributed to metabolic process and again the G protein coupled receptor pathway. This was shown to be downregulated in the undifferentiated group demonstrating a metabolic shift in this cell set. Cluster B, with the largest number of pathways, showed responsibility for DNA packing and cell division pathways. These pathways were seen to be enriched in the undifferentiated cell type suggesting a greater quantity of cell division.

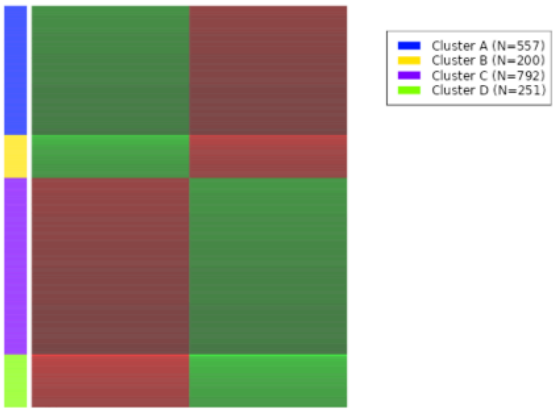
Finally, within the SN-38 alone drug group, the cluster with the greatest enrichment was cluster A (**Figure 16a**). This showed pathways reasonable in the inductions of chemotaxis.

However, the most interesting pathway found was the upregulation of the pathway responsible for Catecholamine transport and secretion. These are stress hormones usually secreted by the adrenal gland but have been shown in studies to stimulate the migration and metastasis of colon cancer cells[132]. The enrichment of these pathways could be demonstrating the greater migratory potential of the undifferentiated cell type.

A



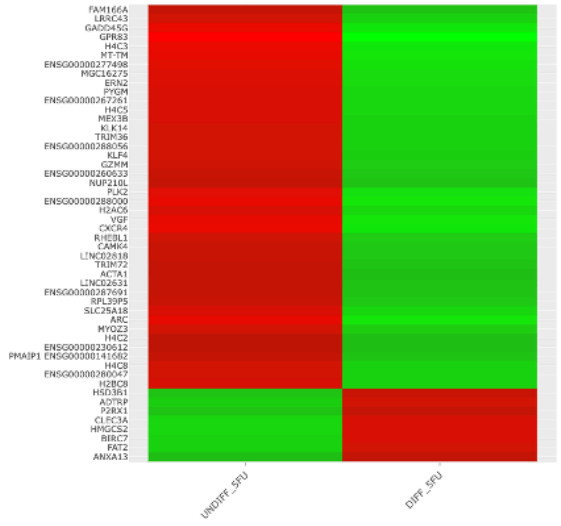
B



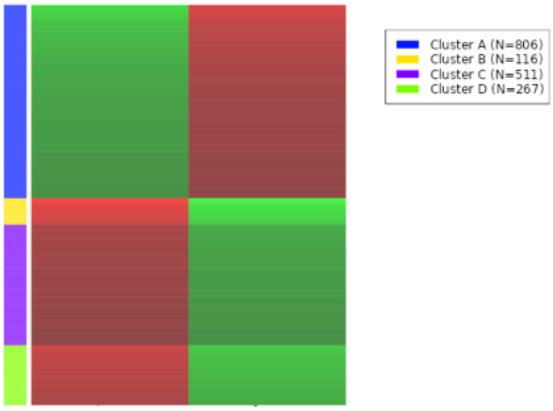
C

Cluster	adj.Pval	nGenes	Pathways
C	7.7e-04	20	Sodium ion transport
	7.7e-04	31	Sensory perception
	1.2e-03	75	System process
	4.1e-03	9	Response to chemokine
	4.1e-03	9	Cellular response to chemokine
	5.4e-03	36	G protein-coupled receptor signaling pathway
	5.4e-03	45	Nervous system process

D

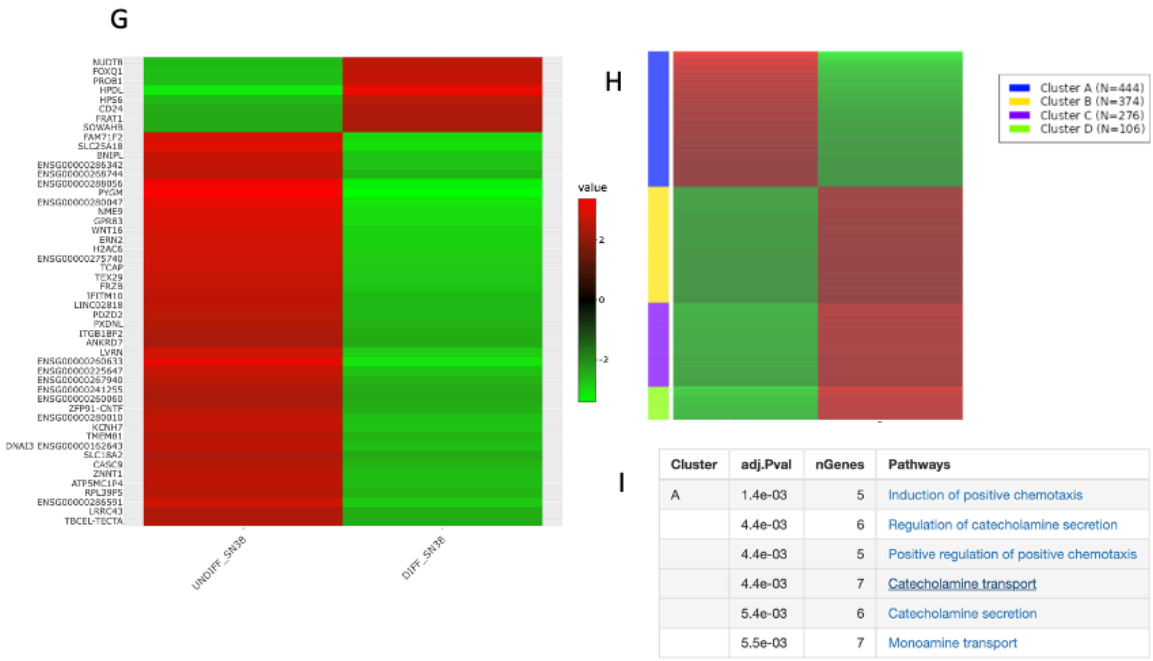


E



F

Cluster	adj.Pval	nGenes	Pathways
A	5.6e-04	41	Organic hydroxy compound metabolic process
	2.5e-03	43	G protein-coupled receptor signaling pathway
	6.6e-03	29	Alcohol metabolic process
B	6.2e-07	9	Nucleosome assembly
	6.2e-07	6	DNA replication-dependent nucleosome assembly
	1.5e-06	6	RDNA heterochromatin assembly
	1.5e-06	10	DNA packaging
	1.5e-06	9	Chromatin assembly
	1.5e-06	5	Negative regulation of megakaryocyte differentiation
	1.5e-06	6	Nucleolar chromatin organization
	1.8e-06	19	Response to organic cyclic compound
	3.1e-06	6	Nucleolus organization
	3.3e-06	9	Nucleosome organization
	4.9e-06	9	Chromatin assembly or disassembly
	5.3e-06	5	Depurination
	6.0e-06	5	Regulation of histone exchange
	6.0e-06	5	Positive regulation of histone exchange
	9.6e-06	6	Regulation of androgen receptor signaling pathway
D	1.9e-03	9	Nucleosome assembly
	5.6e-03	9	Chromatin assembly



Cluster	adj.Pval	nGenes	Pathways
A	1.4e-03	5	Induction of positive chemotaxis
	4.4e-03	6	Regulation of catecholamine secretion
	4.4e-03	5	Positive regulation of positive chemotaxis
	4.4e-03	7	Catecholamine transport
	5.4e-03	6	Catecholamine secretion
	5.5e-03	7	Monoamine transport

Figure 19: Heatmap and K means generated for drug separate data sets in IDEP.96. It shows a heatmap of the top 50 enriched genes, K means heatmap with clusters and a table containing the most enriched pathways. **A, B and C;** relate to the group treated with both 5FU and SN-38. **D, E and F;** relate to the group treated with 5FU. **G, H and I;** relate to the group treated with SN-38.

7.7 Analysis of previously published data of intestinal cancer treated with chemotherapeutic drugs.

As an additional step to compare to the data generated, a paper that used the same drug combination of 5FU and SN-38 on CRC in mouse organoids was investigated[118]. Organoids are in vitro models that are engineered to grow in a 3D environment. They replicate many responses and functions of there in vivo tissue counterpart[133]. The data was downloaded from the Gene Expression Omnibus which has already been subjected to quality control, trimmed and normalised for subsequent analysis. The data set in question is titled GSE207974[118] and the gene counts were used for analysis.

7.7.1 Pre-process of GSE207974

Once downloaded, the data was prepared for IDEP.96 and uploaded it in the same manner as mentioned before for the original data generated for this project. In IDEP.96 the data was handled in the same way as before. Heatmaps and PCA plots were generated in order to see the effect of the drug treatment on the intestinal CSC.

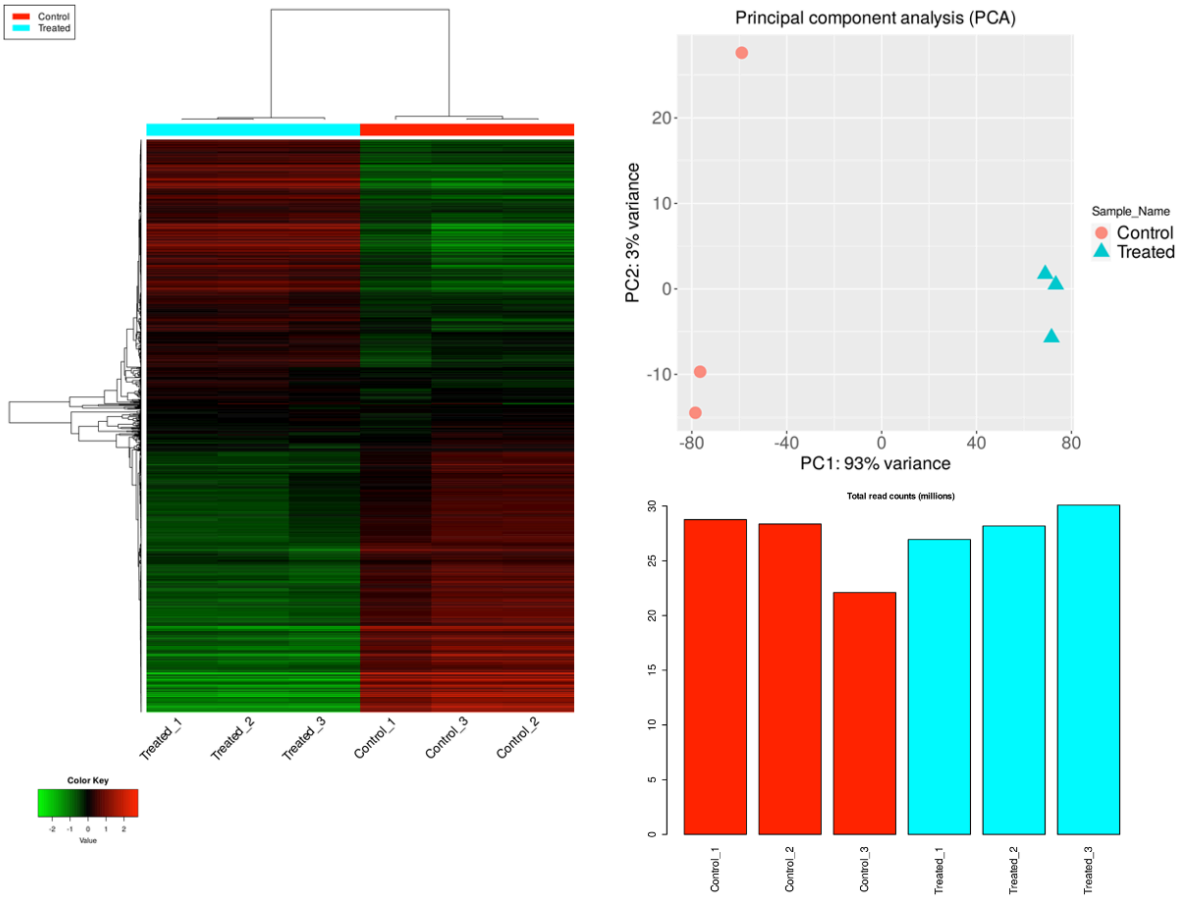


Figure 20: Pre-process analysis for the previously studied data sets using iDEP software. Right, Heatmap shows the relation between the treated and control groups for a broad array of gene expressions. Top left, PCA plots showing the similarities of the samples within the group. Bottom left, the graph shows the total read counts for each sample.

7.7.2 K means analysis and cluster enrichment of GSE207974

K means was used to establish the related clusters that were affected (**Table 8**). Cluster A showed an enrichment of many genes relating to hypoxia response and metabolism. The treated cell types showed a reduced enrichment of genes in response to hypoxia than those that were untreated. The fact that this is reduced with the drug treatment could suggest a metabolic shift in order to avoid the chemotherapy stress. Cluster B showed the enrichment of migratory genes and chemokine response pathways. These genes were reduced in the control sample and upregulated in the treated group. Increased response to chemokines and higher motility in the treated group could be attributed to a mesenchymal shift in response to chemotherapeutic stress. The last cluster to show significant enrichment was to do with immunity and response to other organisms. Like cluster B, cluster C was seen to be reduced in the control group with the treated group showing an upregulation in these genes. There was no significant enrichment in cluster D (**Figure 20**).

7.7.3 Genes of interest investigation of GSE207974

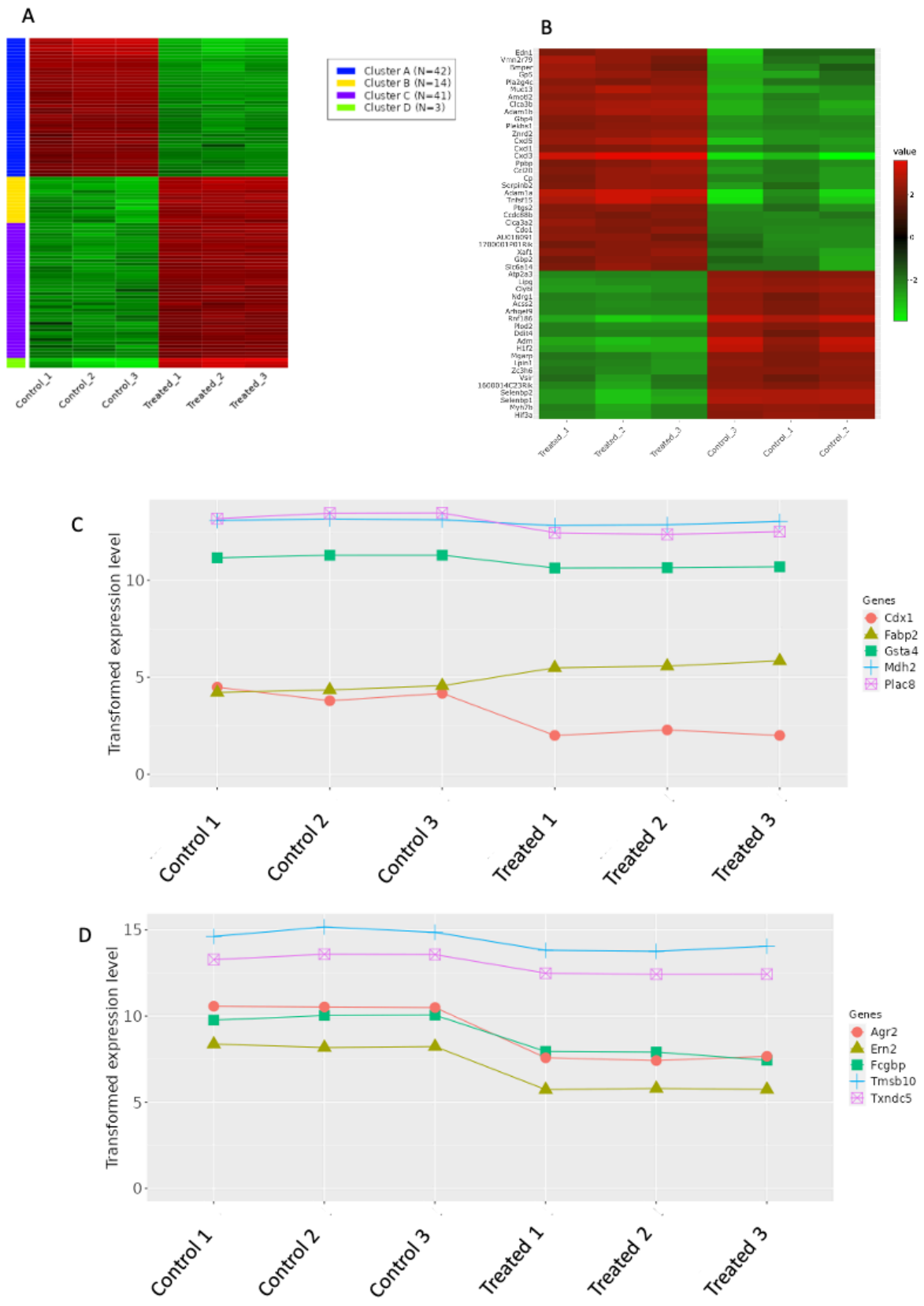
The next step in this process was to analysis the relative expression of specific genes between the treated and untreated groups. This project chose to investigate genes that were related to similar groups as with the analysis of my data. It began with the enterocyte genes. The data showed a slight decrease in these gene types within the treated groups suggesting a lower degree of differentiation. This was also reflected in both the gene types relating to Goblet and Paneth cells. The result of this demonstrates that the treatment is more effective on the differentiated cell and is not as effective on the cells expressing lower levels of differentiation genes. The genes relating to stem cells is harder to conclude on. Some of the gene's signatures were shown to increase in the treated group whereas other were shown to decrease. LGR5 (one of the most studied CSC genes in CRC) was shown to decrease. This may suggest that the cells are becoming quiescent with the chemotherapeutic stress. Finally, the target genes of the Wnt pathway showed little to no change between the groups suggesting that drug treatment does not affect it.

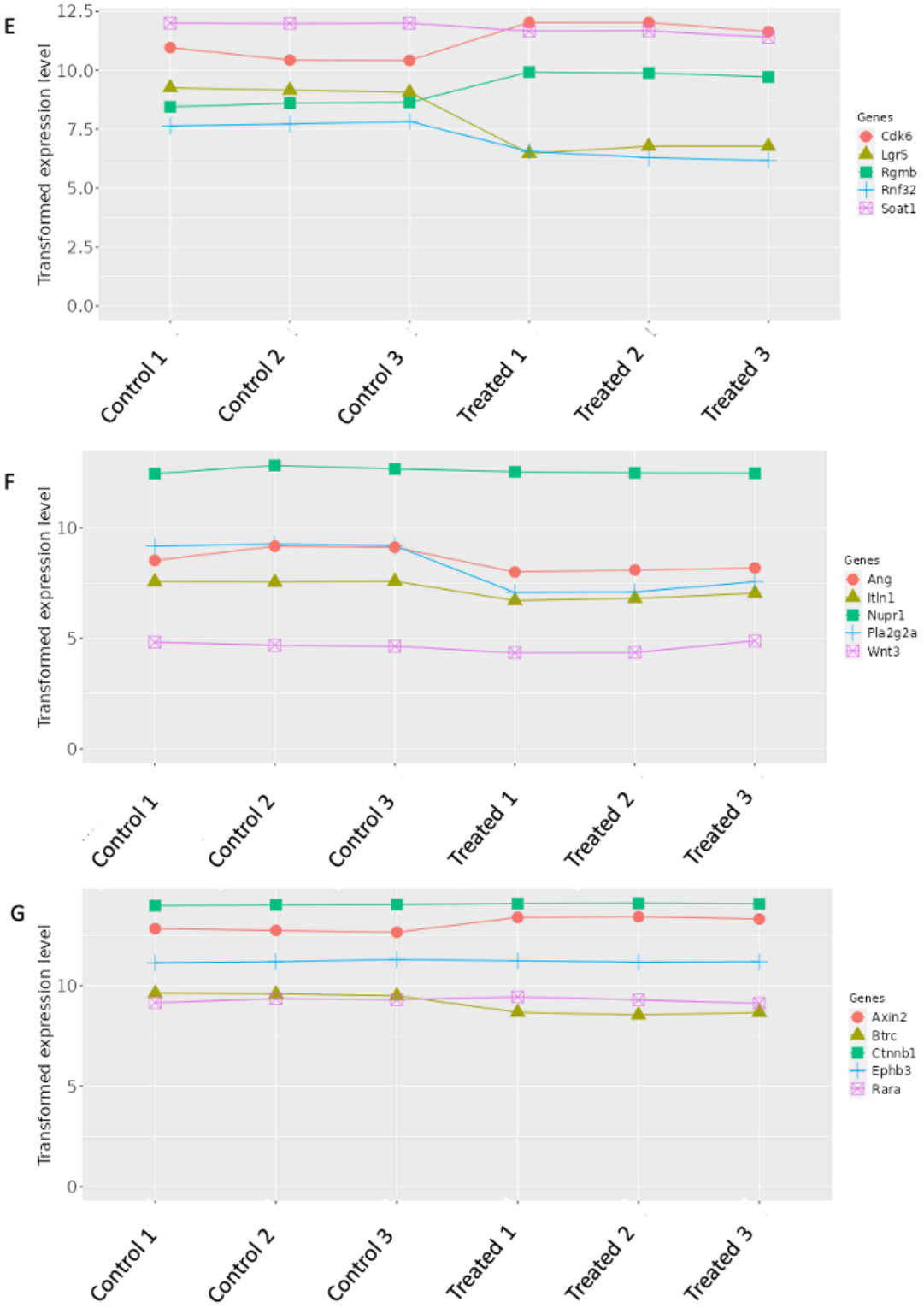
7.7.4 DeSEq2 and volcano plot visualisation of gene regulation in GSE207974

Lastly using the DEG2/DeSEq2 program on IDEP.96, a volcano plot was generated to display the genes with the highest change between the treated and untreated cell group showing significance with the greatest fold change (**Figure 21h**). Next, Enrichr was used (https://maayanlab.cloud/Enrichr/)[134] to study the specific genes and what pathways they are involved in. I went into detail of the three different genes from both the upregulated and downregulated genes. Starting with the downregulated genes, the most prominent is AMOTL2. This gene is involved in the Hippo signalling pathway. This pathway is highly conserved and plays roles in organ development and the control of stem cell renewal[135]. The next gene that was seen to have reduced expression was KRT4, a gene involved in the Oncostatin M pathway. This pathway is often upregulated in bone cancer and promotes cells renewal of the osteoblasts and osteoclasts. However, research has shown that in CRC Oncostatin M promotes cell plasticity and the acquisition of CSC like phenotype[136]. Finally, the third lowly expressed gene is PTGS2, involved in a host of different pathways. One of which is Oncostatin M, but also several lipid metabolism pathways. This could be

demonstrating CSC quiescence and reduced metabolic activity when induced to chemotherapeutic stress.

The most common upregulated gene in the treated group is PDK1. This is involved in pathways responsible for DNA damage repair and pyruvate metabolism. Upregulation of this gene could explain the resistance to treatment found in stem cells and their metabolic shift. Next is EGLN3, a gene involved in hypoxia response. Upregulation of this could be due to hypoxia brought on by the drug treatment. The next gene to be upregulated was ESPN, a gene that is involved in sensory conduction in ear hair follicles, also shown to play roles in chemosensory reception in cells[137]. Upregulation of this gene is interesting but may not play a role in the functions investigated for this study.


A
Cluster A (N=42)
Cluster B (N=14)
Cluster C (N=41)
Cluster D (N=3)
Control_1
Control_2
Control_3
Treated_1
Treated_2
Treated_3
B
value
C
Transformed expression level
Genes
Cdx1
Fabp2
Gsta4
Mdh2
Plac8
Control 1
Control 2
Control 3
Treated 1
Treated 2
Treated 3
D
Transformed expression level
Genes
Agr2
Ern2
Fcgbp
Tmsb10
Txndc5


E
Transformed expression level
12.5
10.0
7.5
5.0
2.5
0.0
Control 1
Control 2
Control 3
Treated 1
Treated 2
Treated 3
Genes
Cdk6
Lgr5
Rgmb
Rnf32
Soat1
F
Transformed expression level
10
5
0
Control 1
Control 2
Control 3
Treated 1
Treated 2
Treated 3
Genes
Ang
Itln1
Nupr1
Pla2g2a
Wnt3
G
Transformed expression level
10
5
0
Control 1
Control 2
Control 3
Treated 1
Treated 2
Treated 3
Genes
Axin2
Btrc
Ctnnb1
Ephb3
Rara

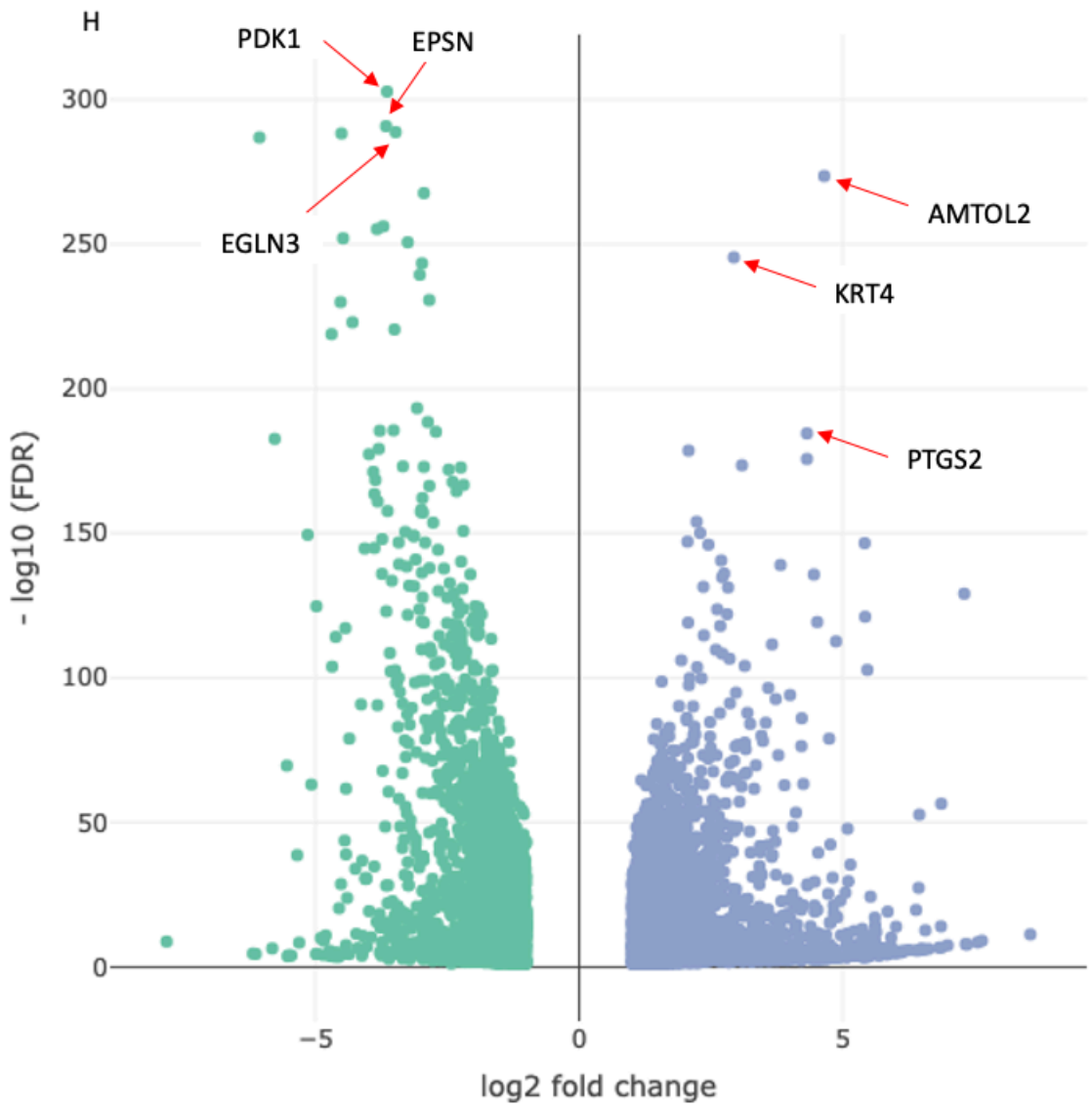


Figure 21: Analysis using IDEP.96 of previously published data GSE20774. **A;** K means heatmap analysis of enriched clusters. **B;** heatmap generated to show the top to genes that are most significantly changed between the data sets. **C, D, E, F and G;** Graphs to show individual genes expressions relating to enterocyte cells, Goblet cells, intestinal stem cells, Paneth cells and Wnt pathway respectively. Transformed data is showing the normalised data expression levels compared to that of the base data set **H;** Volcano plot generated in IDEP.96 using the DEG2 system. Arrows added specify target genes that were discussed above.

Table 8: K means enrichment table for the previously studied data set.

Cluster	adj.Pval	nGenes	Pathways	Genes
A	0.000466801562395659	3	Histone H3-K27 trimethylation	H1f2 H1f4 H1f3
A	0.000466801562395659	8	Response to hypoxia	Hif3a Ndrg1 Ddit4 Ero1a Vegfa Adm Plod2 Mgarp
A	0.000466801562395659	8	Response to decreased oxygen levels	Hif3a Ndrg1 Ddit4 Ero1a Vegfa Adm Plod2 Mgarp
A	0.000479843416588905	3	Nucleosome positioning	H1f2 H1f4 H1f3

A	0.002833751248828	4	Intrinsic apoptotic signaling pathway in response to endoplasmic reticulum stress	Atp2a3 Ero1a Creb3l1 Rnf186
A	0.002833751248828	3	Histone H3-K27 methylation	H1f2 H1f4 H1f3
A	0.002833751248828	5	Cellular response to hypoxia	Hif3a Ndrg1 Ero1a Vegfa Mgarp
A	0.002833751248828	3	Histone H3-K4 trimethylation	H1f2 H1f4 H1f3
A	0.00363063294654988	5	Cellular response to decreased oxygen levels	Hif3a Ndrg1 Ero1a Vegfa Mgarp
A	0.00714251633161085	9	Carboxylic acid metabolic process	Aldoc Ddit4 Aldh1l2 Lpin1 Acss2 Ido1 Plod2 Insig1 Lipg
A	0.00725619740983012	9	Oxoacid metabolic process	Aldoc Ddit4 Aldh1l2 Lpin1 Acss2 Ido1 Plod2 Insig1 Lipg
A	0.00756692700976322	3	Chromosome condensation	H1f2 H1f4 H1f3
B	0.000699521843375409	3	Response to chemokine	Edn1 Ccl20 Cxcl5
B	0.000699521843375409	3	Cellular response to chemokine	Edn1 Ccl20 Cxcl5
B	0.000699521843375409	3	Chemokine-mediated signaling pathway	Edn1 Ccl20 Cxcl5
B	0.00130544719684616	3	Neutrophil chemotaxis	Edn1 Ccl20 Cxcl5
B	0.0017428247038449	3	Neutrophil migration	Edn1 Ccl20 Cxcl5
B	0.0017428247038449	3	Granulocyte chemotaxis	Edn1 Ccl20 Cxcl5
B	0.00220696709157988	3	Granulocyte migration	Edn1 Ccl20 Cxcl5
B	0.00220696709157988	3	Cellular response to interferon-gamma	Edn1 Ccl20 Gbp4
B	0.00254479137176114	3	Response to interferon-gamma	Edn1 Ccl20 Gbp4
B	0.00254479137176114	2	Nitric-oxide synthase biosynthetic process	Edn1 Ccl20
B	0.00254479137176114	2	Regulation of nitric-oxide synthase biosynthetic process	Edn1 Ccl20
B	0.00436078647841214	3	Myeloid leukocyte migration	Edn1 Ccl20 Cxcl5
B	0.00436078647841214	3	Leukocyte chemotaxis	Edn1 Ccl20 Cxcl5
B	0.00580897392795435	2	Killing of cells of other organism	Ccl20 Cxcl5
C	2.54427941444854E-05	14	Response to biotic stimulus	Csf1 Il18rap Lcn2 Gbp2 Gpatch3 Ppbp Cxcl1 Plscr2 Ptgs2 Vnn1 Akap12 Ccdc88b Cxcl2 Gbp6
C	2.54427941444854E-05	14	Response to other organism	Csf1 Il18rap Lcn2 Gbp2 Gpatch3 Ppbp Cxcl1 Plscr2 Ptgs2 Vnn1 Akap12 Ccdc88b Cxcl2 Gbp6
C	2.56549620501236E-05	14	Defense response	Dnase1 Csf1 C1qtnf12 Il18rap Lcn2 Gbp2 Gpatch3 Ppbp Cxcl1 Ptgs2 Vnn1 Ccdc88b Cxcl2 Gbp6

C	4.12233205369118E-05	8	Response to lipopolysaccharide	Gbp2 Ppbp Cxcl1 Plscr2 Ptgs2 Akap12 Cxcl2 Gbp6
C	4.12233205369118E-05	14	Biological process involved in interspecies interaction between organisms	Csf1 Il18rap Lcn2 Gbp2 Gpatch3 Ppbp Cxcl1 Plscr2 Ptgs2 Vnn1 Akap12 Ccdc88b Cxcl2 Gbp6
C	5.29126328316452E-05	8	Response to molecule of bacterial origin	Gbp2 Ppbp Cxcl1 Plscr2 Ptgs2 Akap12 Cxcl2 Gbp6
C	0.000161140533820456	13	Immune response	Dnase1 Csf1 Lcp1 Il18rap Lcn2 Gbp2 Gpatch3 Ppbp Cxcl1 Vnn1 Ccdc88b Cxcl2 Gbp6
C	0.000172756743626572	9	Inflammatory response	Dnase1 Csf1 C1qtnf12 Il18rap Ppbp Cxcl1 Ptgs2 Vnn1 Cxcl2
C	0.000172756743626572	9	Response to bacterium	Lcn2 Gbp2 Ppbp Cxcl1 Plscr2 Ptgs2 Akap12 Cxcl2 Gbp6
C	0.000172756743626572	6	Cellular response to lipopolysaccharide	Gbp2 Ppbp Cxcl1 Plscr2 Cxcl2 Gbp6
C	0.000178104696285257	17	Response to external stimulus	Efnb3 Dnase1 Csf1 C1qtnf12 Il18rap Lcn2 Gbp2 Gpatch3 Ppbp Cxcl1 Plscr2 Ptgs2 Vnn1 Akap12 Ccdc88b Cxcl2 Gbp6
C	0.000181349540831208	6	Cellular response to molecule of bacterial origin	Gbp2 Ppbp Cxcl1 Plscr2 Cxcl2 Gbp6
C	0.000351738603877	6	Cellular response to biotic stimulus	Gbp2 Ppbp Cxcl1 Plscr2 Cxcl2 Gbp6
C	0.000363901934201574	16	Immune system process	Efnb3 Dnase1 Csf1 Lcp1 Il18rap Lcn2 Gpatch4 Gbp2 Gpatch3 Ppbp Cxcl1 Vnn1 Ccdc88b Cxcl2 Cebpd Gbp6
C	0.00053329504946166	10	Response to cytokine	Csf1 Il18rap Gbp2 Ppbp Cxcl1 Ptgs2 Akap12 Xaf1 Cxcl2 Gbp6

8 Discussion

With CRC continuing to be one of the most common cancers, its burden on health services forever increases. Combined with its tendency for relapse and post treatment metastatic disease, a greater depth of knowledge is still ultimately required for effective treatment with fewer cases of reoccurrence. A host of previous data published has related cancer cell stemness to treatment resistance and metastatic potential. However, with stemness lacking effective designation markers in colon cancer disease, it remains hard to describe. For this reason, we aimed in this research project to identify whether differentiation in cancer cells can interfere with chemotherapeutic response.

8.1 Caco2 differentiation as an in vitro model

The first experiment conducted was using Caco2 cell lines in order to generate a less differentiated cell line by using a cancer cell line that can differentiate in vitro[117]. The aim of this was to compare cells grown in alternative conditions to the ones grown standard and compare the effect of drug treatment of them. As mentioned earlier, the cells grown in low confluence are not truly undifferentiated but do have less differentiated cells and a probably a greater quality of stemness than those grown standard. Additionally, this would provide an easy way of comparing the effect of differentiation on the same cell line. The qPCR data in the results does show a higher level of differentiated genes in the differentiated cell group (via the expression of certain enterocyte markers). With this data we concluded that the lower confluence group has lost some of its differentiated capacity and reverted to a more stem like state. With this result, we were confident to move on to the next stage of treatments.

8.2 Drug treatment and qPCR

Once it was established that the Caco2 cells were showing reduced differentiation to the standard set, the project moved onto the treatment of these cells with a variety of drugs. Oxaliplatin and SN-38 was used as it is an established treatment regime for CRC. From the results collected after the treatment, it demonstrated that the chemotherapy was selectively destroying the undifferentiated cell types. Alternatively, it is possible that the

stress brought on by the chemotherapy was causing the undifferentiated cell types to shift back into a differentiated type.

A previous study into the role of LGR5 on CRC metastasis and treatment resistance may contain some answers to the questions my observations posed. LGR5 is an established stem cell marker in CRC and can be used to identify cell populations with high stemness. In order to control this, CRC cell population that was LGR5+ with a diphtheria toxin receptor was used. This allowed the removal of LGR5+ cells via the administration of diphtheria toxin (DT), allowing for control of the stem cell population. Transcriptional analysis exhibited an increase in intestinal stem cell markers as well as Wnt target genes supporting the fact that these were CSC. Additionally, tumour engraftment analysis showed that these LGR5+ tumours had enhanced tumour initiating capacity, a trait established to belong to CSC. With the administration of DT, tumour stasis was observed. This persisted as long as the treatment was administered but resumed with the discontinuation of the treatment. This illustrates the plasticity of the CSC model and how differentiated cells can become more stem-like with treatment removal. They concluded that, in the absence of the LGR5+ CSC pool, LGR5- cells maintained the tumour via proliferation whilst attempting to continually replenish the CSC pool139.

Whilst the experimentation explained above differs from this project, it may contain some answers to what was observed. Due to the reduced differentiation of the Caco2 cells without true stemness, the chemotherapy treatments could still be effective in the targeting and removal of these cell types. This in turn causes the cell population to become more differentiated and endothelial to survive the chemotherapy stress.

8.3 Comparison between undifferentiated and differentiated cells groups in response to drug treatment

The next step took in this project was to investigate the effect of differentiation on cellular response to chemotherapy drug treatment. 5FU and SN-38 was chosen as the chemotherapy drugs to use as they are already established treatment regimens for patients with CRC and previous data sets analysed later also utilised them[118,140]. Additionally, to

simulate a more realistic chemotherapy response, one drug treatment contained both 5FU and SN-38 in combination. This was done to see if cell populations would react differently when exposed to multiple drug types compared to just one. Previous studies have revealed that cell populations with 5FU resistance have stem cell marker genes indicating that it is a good drug to model differences between undifferentiated and differentiated cell types[141].

Single gene marker analysis

When looking at the expression of single gene markers obtained from the individual treated groups, one of the things observed was the upregulation of AXIN2, a Wnt pathway gene, in all three of the undifferentiated groups. Wnt pathway, as explained in the introduction, fosters CRC development and ablation of this pathways causes a reduction on the CSC pool[142]. An upregulation of genes relating to this pathway, with the administration of chemotherapy, could be showing a cell population shift towards a greater stem like phenotype in order to survive the treatment stress.

In the group treated with both drugs, there was seen to be an increase in the expression of LGR5. This is a Wnt target gene and is often described as a marker for CSC. LGR5 is known to trigger therapy resistance in 5FU treated patients and demonstrated the greater stem like phenotype in the undifferentiated cell type[143].

When a broader analysis of Wnt target genes and enterocyte genes were investigated, some interesting points arose. In the dual drug group, for the enterocyte targets, there seemed to be little change between the two groups. This suggests that the cell populations have a similar differential pattern. However, before treatment, the qPCR data obtained showed a distinct change in their differential pattern. Combination of both drugs may be forcing the undifferentiated group into a more differentiated role. Wnt target genes showed a greater degree of upregulation in the undifferentiated group which is in line to what would be expected[144].

In both the single drug groups, the most interesting change was the upregulation of ALDOB in the undifferentiated cell group. This is a stark contrast to what was expected and what was observed in the other genes associated with enterocyte cells. Previous data has shown an interesting role for ALDOB in CRC. ALDOB enhancement allows for a greater fructose

metabolism and is associated with liver metastasis from CRC[145]. The k means data analysis showed an upregulation of chemotaxis and migratory genes as well as alternative metabolism pathways. This suggests that the individual drug treatments may not be strong enough to target the drug resistant persister cells. This allowed for metabolic reprogramming in order to facilitate alternative environments for growth via metastasis. Wnt target genes were also seen to be upregulated in the undifferentiated type which can be associated with the accusation of a more stem like phenotype to support migration. However, LGR5, a known stem cell marker in colon cells, was seen to be downregulated. This is differing to the upregulation of the Wnt target genes and requires further investigation.

Evaluation of the enrich clustered identified by k means analysis

K means data generated showed that in the group treated with both drugs, there was an upregulation of genes relating to EMT. It is well established that an increase in migratory potential via the facilitation of EMT increases the drug resistive capacity of tumours, with many previously published articles and research confirming this[146–148]. One such research data showed that an upregulation of extra cellular matrix protein 1 (ECM1) fostered CRC chemoresistance via the PI3K pathway. Their research showed that cell groups resistant to 5FU displayed higher levels of ECM1 and acquired a more mesenchymal phenotype and poorer cell prognosis. Additionally, there was seen to be a morphological change to spindle like formations displaying the cells have undergone a EMT shift[149]. Upregulation of EMT genes, within the undifferentiated group, with drug treatment could be demonstrating the treatment resistance in the persister cell populations. In the group treated with just 5FU, it was witnessed that the undifferentiated cell group displayed an enrichment of genes associated with cell division, suggesting that this group maintained a proliferative capability in the drug environment. Finally, the SN-38 group demonstrated an increase in the regulation of chemotaxis and catecholamines in the undifferentiated group. Catecholamines have been shown to regulate cytoskeletal production and stimulate a more invasive phenotype[150]. This, in addition to the data showing induction of chemotaxis, suggests that the drug treatment is causing the undifferentiated group to become more mobile and invasive. This fits with the theory that the stem-like phenotype confers drug resistance as mesenchymal stem cells are seen to have anti-cancer drug properties[151].

8.4 Comparison of acquired data to previously published data sets on CRC organoids treated with chemotherapeutic agents

Finally, after all the data generated in the lab was acquired, it was decided the next step would be to compare this to other data sets found relating to my topic. The data sets analysed were acquired from a paper in which authors investigated drug resistance in CRC stem cells. They had generated CRC organoids of CSC in mouse models and treated them with 5FU and SN-38. One of the benefits of this study was the fact that they had a drug-free control group to investigate as well as a group treated with both drugs. However, the limitations were that they did not contain any single treatment samples or different time points.

In this group of samples, the data point were subjected to the same investigations taken with this projects data. In the single gene analysis, it was observed that between the treated and control group there was no change in the Wnt target genes. This suggests that Wnt is not induced in response to treatment and is more likely to be a relevant identifier of cellular differentiation. Whilst genes like Cdk6, that encode for intestinal stem cells, were shown to increase with drug treatment suggesting a push towards a greater stem like phenotype, LGR5 was shown to be decreased. This demonstrates that LGR5 is downregulated with treatment and is particularly sensitive to chemotherapy. This may provide insight into why a decrease of LGR5 was observed in the undifferentiated cell population.

K mean data analysis showed an upregulation of pathways involving migration and movement in the treated type. This may also be mirrored in some of my data showing that the addition of drug treatment and the induced stress from the cytotoxic compounds on stem cells, causes a genetic and/or phenotypic shift in the cells to prime them for migration to escape the affected area. It is established that mobile, mesenchymal cells are more resistant to drug treatment and can escape the drug saturated environment. It also showed an upregulation of immune cross talk in the treated cell type. Increased communication between cancer and immune cells may be facilitating immune dependent cell death and thus establishing treatment resistance.

8.5 Limitations and future directions

Limitations

As mentioned above, the nature of the undifferentiated Caco2 cells may cause issues with this experimental model. Additionally, the monolayer nature of the cell culture dose not replicate the true tumour environment found within the intestine. Whilst the chemotherapy drugs are utilised in CRC, a wider range and use of drug combinations will create a more accurate representation of what is used in clinic.

One of the most obvious limitations to this model is the lack of true stemness. Whilst the undifferentiated Caco2 cell lines provided an easy and relatively stable way to study the effect of differentiation, a true stem cell pool would provide a greater model to test on. Additionally, Caco2 cell lines are limited by their relatively homogenous monolayer. Due to the lack of stem cells within the cell population, Caco2 cells are only able to differentiate into enterocytes. This is not representative of the complex environment of the colon and may affect the drug sensitivity of cells. However, this tool is less time consuming than using organoids or mouse models and provides information about the cancer cell line responds depending on enterocyte content which are up to 80% of the total intestinal cells[138].

One of the greatest limitations during the experimentations is the lack of a drug-free control. This would allow for a comparison between the individual sample types to see whether the differences experienced are due to the undifferentiated nature or the drug treatment. The reason for the lack of this control group can be found in the disclaimer section. Furthermore, whilst repeats were generated in the lab, they were not analysed using RNA sequencing. Additional analysed repeats would have allowed for a more in-depth study as several pieces of software on IDEP.96, such as DeSEQ2, were not available without the addition of repeated experiments and data sets. The use of 3D models would have provided a situation with greater in clinic relevance which may have an effect of the overall gene expression of treated cells.

Whilst the public data set the same dual drugs (5FU and SN-38) as my experiments did, the lack of individual drug treatment groups do not allow for comparison of these data sets.

Additionally, there were no concentration ranges so comparison at different concentration points is not possible. Furthermore, the model of experimentation was 3D organoids in mouse models. Both the 3D nature and mouse environment would differ to the parameters within my experiment and therefor may be influencing the gene expression levels observed.

Future directions

A future direction to improve the method followed in this project would be the utilisation of intestinal stem cells. A comparison between these populations and the undifferentiated Caco2 cell lines would give a good indication to the degree of differentiation possessed by them. Additionally, the use of 3D models would provide a better landscape of the colon and allow for a truer response to the drug treatments. In the second semester of my studies, it was planned to differentiate 3D-patient-derived organoids by using valproic acid that are currently being grown by several team members. However, for reasons outlined in the disclaimer section, these experiments were substituted by an expanded bioinformatic analysis.

Additionally in the future the next logical step would be to identify the effect of the treatment of different genes. Whilst the genes investigated are a good measure of the differentiation of the cell types, additional properties may unveil interesting results. LGR5 would be one candidate as it is well established as a stem cell marker, so rather than looking at the differentiation, investigation into the proportion of stemness with drug treatment may be interesting. Likewise, investigation into survival pathways may also uncover the effect of treatment and cell response. This would pair nicely with using the same drug treatments of standard Caco2 cell lines and comparing both, as this would give a greater understanding how the differentiation affects depending on the cell type used. A further step would be to compare it to intestinal stem cells treated with these same compounds to see how comparable the undifferentiated Caco2 cell line is. If this direction was taken, performing scRNAseq would give greater insight in how tumour response would depend on differentiation and stemness content.

A step that can be taken to better this experiment would be to analyse more replicates and drug-free control groups. This would allow for a greater understanding of what is observed

and what it can be attributed to. Multiple replicates of the same experimentation would allow the use of DeSEQ2 software as well as the generation of volcano plots to specify individual genes which show a greater difference in their regulation. Additionally, the use of intestinal stem cell models would provide a more accurate representation of drug effect on cancer persister cells. Another experiment that could be conducted would be survival analysis to compare survival rates between the two groups. One experiment that was started but unfortunately was not completed is a recovery experiment. Here I removed the drug saturated media after treatment and allowed cells to recover in drug free media. The idea was to analyse the cells to see which group could recover from drug treatment more effectively, simulating relapse, as has been done in 2D models in recent publications[152]. RNA-seq analysis of this cell group could provide a greater understanding of how the stem cells recover from cytotoxic drug treatment and re-initiate tumour formation.

As previously stated, the use of 3D models and organoids provides a situation with greater in clinic relevance. When grown in 3D, tumours cells will have areas that are exposed to differing amounts of oxygen and therefor have, usually towards the centre, hypoxic regions occuring[153]. We can see in the K means analysis that there is an enrichment of hypoxia response pathways in the untreated group. This is expected as with drug treatment and cell death, there would be a shrinkage in the tumour mass and therefore less hypoxic areas. Interestingly, previous research has highlighted the effect of hypoxia on 5FU resistance. 5FU along with oxaliplatin was administered to CRC cell lines under normoxic and hypoxic conditions. Whist in normoxia, the chemotherapeutic effect was evidenced and inhibited the proliferative effect of HT29 cell lines whereas under hypoxia, there was seen to be a significant reduction in cell death. This was said to be observed due to the cell cycle arrest in the G_0 phase that is brought on by hypoxia. However, SN-38 administration was not seen to be impaired by hypoxia[154]. Hypoxia is known to be able to induce and maintain a stem cell phenotype whilst preventing differentiation in CRC cell lines[155] . Testing the chemotherapeutic effects of SN-38 on designated stem cells alone in hypoxic conditions may give greater insights into their resistance mechanisms and ways to overcome them.

9 Conclusion

The drive to greater our understanding and widen our breadth of knowledge into CSC will be one that continues for many years to come. Treatment resistance and disease relapse continues to be the greatest challenge in clinic resulting in many CRC deaths. Through the bioinformatic analysis and identification of enriched pathways and upregulated genes in undifferentiated CRC cells, this project has provided the first steps into identifying how CSC avoid treatment response and how treatment can be designed to overcome this. Furthermore, the enforcement of the undifferentiated Caco2 cell line generation compared to stem cell organoids can provide future researchers with an easy to generate and handle stem cell model. This can pave the way for many other research groups to investigate and test new methods and ideas without the need for complex systems. Ultimately, the findings that this project brought forth may aid other researchers in identifying resistance pathways and mechanisms and developing targeted treatment options, therefore reducing clinical relapses with improved outcomes and disease-free survival, addressing an unmet need for CRC patients.

10 Acknowledgments

I would like to thank my project supervisor Paloma Ordóñez Morán, without whose expertise and guidance, none of this would be possible. Her continued support both professionally and personally was a great driver to complete this project. I would like to thank my lab colleagues Brinda Balasubramanian for all her support and help in completing the bioinformatics section of this project as well as Magda Martinez and Shivam Patel for their aid and expertise in maintaining cell culture and other laboratory practices. **Figures 1-5** were created using Biorender (https://www.biorender.com/).

11 References

1. Physiology, Large Intestine - StatPearls - NCBI Bookshelf.
2. Anatomy, Abdomen and Pelvis, Large Intestine - StatPearls - NCBI Bookshelf.
3. Barker, N., De Wetering, M. Van & Clevers, H. The intestinal stem cell. *Genes and Development* vol. 22 1856–1864 Preprint at https://doi.org/10.1101/gad.1674008 (2008).
4. Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* **449**, 1003–1007 (2007).
5. Baulies, A., Angelis, N. & Li, V. S. W. Hallmarks of intestinal stem cells. *Development (Cambridge)* vol. 147 Preprint at https://doi.org/10.1242/dev.182675 (2020).
6. *Searching for adult stem cells in the intestine*. www.embomolmed.org.
7. POTTEN, C. S. Extreme sensitivity of some intestinal crypt cells to X and γ irradiation. *Nature* **269**, 518–521 (1977).
8. 4440635.
9. Yilmaz, Ö. H. *et al.* MTORC1 in the Paneth cell niche couples intestinal stem-cell function to calorie intake. *Nature* **486**, 490–495 (2012).
10. Mihaylova, M. M. *et al.* Fasting Activates Fatty Acid Oxidation to Enhance Intestinal Stem Cell Function during Homeostasis and Aging. *Cell Stem Cell* **22**, 769-778.e4 (2018).
11. Beyaz, S. *et al.* High-fat diet enhances stemness and tumorigenicity of intestinal progenitors. *Nature* **531**, 53–58 (2016).
12. Scholtysek, C. *et al.* PPARβ/δ governs Wnt signaling and bone turnover. *Nat Med* **19**, 608–613 (2013).

13. de Sousa e Melo, F. & de Sauvage, F. J. Cellular Plasticity in Intestinal Homeostasis and Disease. *Cell Stem Cell* vol. 24 54–64 Preprint at https://doi.org/10.1016/j.stem.2018.11.019 (2019).
14. Yu, S. *et al.* Paneth Cell Multipotency Induced by Notch Activation following Injury. *Cell Stem Cell* **23**, 46-59.e5 (2018).
15. Clevers, H. XThe intestinal crypt, a prototype stem cell compartment. *Cell* vol. 154 274 Preprint at https://doi.org/10.1016/j.cell.2013.07.004 (2013).
16. Farin, H. F., Van Es, J. H. & Clevers, H. Redundant sources of Wnt regulate intestinal stem cells and promote formation of paneth cells. *Gastroenterology* **143**, (2012).
17. Farin, H. F. *et al.* Visualization of a short-range Wnt gradient in the intestinal stem-cell niche. *Nature* **530**, 340–343 (2016).
18. Degirmenci, B., Valenta, T., Dimitrieva, S., Hausmann, G. & Basler, K. GLI1-expressing mesenchymal cells form the essential Wnt-secreting niche for colon stem cells. *Nature* **558**, 449–453 (2018).
19. de Lau, W., Peng, W. C., Gros, P. & Clevers, H. The R-spondin/Lgr5/Rnf43 module: Regulator of Wnt signal strength. *Genes and Development* vol. 28 305–316 Preprint at https://doi.org/10.1101/gad.235473.113 (2014).
20. Basak, O. *et al.* Induced Quiescence of Lgr5+ Stem Cells in Intestinal Organoids Enables Differentiation of Hormone-Producing Enteroendocrine Cells. *Cell Stem Cell* **20**, 177-190.e4 (2017).
21. Sato, T. *et al.* Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology* **141**, 1762–1772 (2011).
22. Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J. S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res* **24**, 1012–1019 (2014).
23. He, X. C. *et al.* BMP signaling inhibits intestinal stem cell self-renewal through suppression of Wnt-β-catenin signaling. *Nat Genet* **36**, 1117–1121 (2004).
24. Qi, Z. *et al.* BMP restricts stemness of intestinal Lgr5 + stem cells by directly suppressing their signature genes. *Nat Commun* **8**, (2017).
25. Madison, B. B. *et al.* Epithelial hedgehog signals pattern the intestinal crypt-villus axis. *Development* **132**, 279–289 (2005).
26. Kosinski, C. *et al.* Indian hedgehog regulates intestinal stem cell fate through epithelial-mesenchymal interactions during development. *Gastroenterology* **139**, 893–903 (2010).
27. Martinez-Redondo, P. & Izpisua Belmonte, J. C. Tailored chromatin modulation to promote tissue regeneration. *Semin Cell Dev Biol* **97**, 3–15 (2020).
28. Zimberlin, C. D. *et al.* HDAC1 and HDAC2 collectively regulate intestinal stem cell homeostasis. *The FASEB Journal* **29**, 2070–2080 (2015).
29. Sheaffer, K. L. *et al.* DNA methylation is required for the control of stem cell differentiation in the small intestine. *Genes Dev* **28**, 652–664 (2014).
30. Hossain, M. S. *et al.* Colorectal Cancer: A Review of Carcinogenesis, Global Epidemiology, Current Challenges, Risk Factors, Preventive and Treatment Strategies. *Cancers* vol. 14 Preprint at https://doi.org/10.3390/cancers14071732 (2022).
31. Kuipers, E. J. *et al.* Colorectal cancer. *Nat Rev Dis Primers* **1**, (2015).

32. Xie, Y. H., Chen, Y. X. & Fang, J. Y. Comprehensive review of targeted therapy for colorectal cancer. *Signal Transduction and Targeted Therapy* vol. 5 Preprint at https://doi.org/10.1038/s41392-020-0116-z (2020).
33. Fearon, E. F. & Vogelstein, B. *A Genetic Model for Colorectal Tumorigenesis*. *Cell* vol. 61 (1990).
34. Mármol, I., Sánchez-de-Diego, C., Dieste, A. P., Cerrada, E. & Yoldi, M. J. R. Colorectal carcinoma: A general overview and future perspectives in colorectal cancer. *International Journal of Molecular Sciences* vol. 18 Preprint at https://doi.org/10.3390/ijms18010197 (2017).
35. Pickhardt, P. J. *et al.* Assessment of volumetric growth rates of small colorectal polyps with CT colonography: A longitudinal study of natural history. *Lancet Oncol* **14**, 711–720 (2013).
36. Stryker, S. J. *et al. Natural History of Untreated Colonic Polyps*. *GASTROENTEROLOGY* vol. 93 (1987).
37. https://www.nih.gov/research- training/advances-colorectal-cancer-research.
38. Sawicki, T. *et al.* A review of colorectal cancer in terms of epidemiology, risk factors, development, symptoms and diagnosis. *Cancers* vol. 13 Preprint at https://doi.org/10.3390/cancers13092025 (2021).
39. Sehgal, R. *et al.* Lynch Syndrome: An updated review. *Genes (Basel)* **5**, 497–507 (2014).
40. Win, A. K., MacInnis, R. J., Hopper, J. L. & Jenkins, M. A. Risk prediction models for colorectal cancer: A review. *Cancer Epidemiology Biomarkers and Prevention* vol. 21 398–410 Preprint at https://doi.org/10.1158/1055-9965.EPI-11-0771 (2012).
41. Rawla, P., Sunkara, T. & Barsouk, A. Epidemiology of colorectal cancer: Incidence, mortality, survival, and risk factors. *Przeglad Gastroenterologiczny* vol. 14 89–103 Preprint at https://doi.org/10.5114/pg.2018.81072 (2019).
42. Ma, Y. *et al.* Type 2 diabetes and risk of colorectal cancer in two large U.S. prospective cohorts. *Br J Cancer* **119**, 1436–1442 (2018).
43. Chan, D. S. M. *et al.* Red and processed meat and colorectal cancer incidence: Meta-analysis of prospective studies. *PLoS One* **6**, (2011).
44. *Colorectal Cancer Causes, Risk Factors, and Prevention Colorectal Cancer Risk Factors*.
45. 31455888 (1).
46. Seidel, J. A., Otsuka, A. & Kabashima, K. Anti-PD-1 and anti-CTLA-4 therapies in cancer: Mechanisms of action, efficacy, and limitations. *Frontiers in Oncology* vol. 8 Preprint at https://doi.org/10.3389/fonc.2018.00086 (2018).
47. Batlle, E. & Clevers, H. Cancer stem cells revisited. *Nature Medicine* vol. 23 1124–1134 Preprint at https://doi.org/10.1038/nm.4409 (2017).
48. Munro, M. J., Wickremesekera, S. K., Peng, L., Tan, S. T. & Itinteang, T. Cancer stem cells in colorectal cancer: A review. *Journal of Clinical Pathology* vol. 71 110–116 Preprint at https://doi.org/10.1136/jclinpath-2017-204739 (2018).
49. Ramirez, M. *et al.* Diverse drug-resistance mechanisms can emerge from drug-tolerant cancer persister cells. *Nat Commun* **7**, (2016).
50. Mullard, A. Stemming the tide of drug resistance in cancer. *Nature reviews. Drug discovery* vol. 19 221–223 Preprint at https://doi.org/10.1038/d41573-020-00050-y (2020).
51. Sato, T. *et al.* Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature* **469**, 415–418 (2011).

52. Gupta, P. B. *et al.* Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* **146**, 633–644 (2011).
53. Van De Wetering, M. *et al. The-Catenin/TCF-4 Complex Imposes a Crypt Progenitor Phenotype on Colorectal Cancer Cells renewal occurs in these crypts through a coordinated series of events involving proliferation, differentiation, and migration toward the intestinal lumen. Pluripotent cation compartment divide approximately every 12 hr*. *Cell* vol. 111 http://www.cell.com/cgi/content/full/111/2/241/ (2002).
54. Ye, X. *et al.* Distinct EMT programs control normal mammary stem cells and tumour-initiating cells. *Nature* **525**, 256–260 (2015).
55. Nieto, M. A., Huang, R. Y. Y. J., Jackson, R. A. A. & Thiery, J. P. P. EMT: 2016. *Cell* vol. 166 21–45 Preprint at https://doi.org/10.1016/j.cell.2016.06.028 (2016).
56. Celià-Terrassa, T. *et al.* Epithelial-mesenchymal transition can suppress major attributes of human epithelial tumor-initiating cells. *Journal of Clinical Investigation* **122**, 1849–1868 (2012).
57. Chaffer, C. L. *et al.* XPoised chromatin at the ZEB1 promoter enables breast cancer cell plasticity and enhances tumorigenicity. *Cell* **154**, 61 (2013).
58. Schmidt, J. M. *et al.* Stem-cell-like properties and epithelial plasticity arise as stable traits after transient twist1 activation. *Cell Rep* **10**, 131–139 (2015).
59. 18519656.
60. Liu, P. P. *et al.* Metabolic regulation of cancer cell side population by glucose through activation of the Akt pathway. *Cell Death Differ* **21**, 124–135 (2014).
61. Goidts, V. *et al.* RNAi screening in glioma stem-like cells identifies PFKFB4 as a key molecule important for cancer cell survival. *Oncogene* **31**, 3235–3243 (2012).
62. Yuen, C. A., Asuthkar, S., Guda, M. R., Tsung, A. J. & Velpula, K. K. Cancer stem cell molecular reprogramming of the Warburg effect in glioblastomas: a new target gleaned from an old concept. *CNS oncology* vol. 5 101–108 Preprint at https://doi.org/10.2217/cns-2015-0006 (2016).
63. Zhu, X. *et al.* Energy metabolism in cancer stem cells. *World J Stem Cells* **12**, 448–461 (2020).
64. Suda, T., Takubo, K. & Semenza, G. L. Metabolic regulation of hematopoietic stem cells in the hypoxic niche. *Cell Stem Cell* vol. 9 298–310 Preprint at https://doi.org/10.1016/j.stem.2011.09.010 (2011).
65. Rodríguez-Colman, M. J. *et al.* Interplay between metabolic identities in the intestinal crypt supports stem cell function. *Nature* **543**, 424–427 (2017).
66. Lonardo, E., Frias-Aldeguer, J., Hermann, P. C. & Heeschen, C. Pancreatic stellate cells form a niche for cancer stem cells and promote their self-renewal and invasiveness. *Cell Cycle* **11**, 1282–1290 (2012).
67. Lisanti, M. P., Martinez-Outschoorn, U. E. & Sotgia, F. Oncogenes induce the cancer-associated fibroblast phenotype: Metabolic symbiosis and 'fibroblast addiction' are new therapeutic targets for drug discovery. *Cell Cycle* vol. 12 2723–2732 Preprint at https://doi.org/10.4161/cc.25695 (2013).
68. Thi, L. *et al.* Cancer Stem Cells (CSCs) in Drug Resistance and their Therapeutic Implications in Cancer Treatment. (2018) doi:10.1155/2018/5416923.
69. Merchant, A. A. & Matsui, W. Targeting Hedgehog - A cancer stem cell pathway. *Clinical Cancer Research* vol. 16 3130–3140 Preprint at https://doi.org/10.1158/1078-0432.CCR-09-2846 (2010).

70. Varnat, F. *et al.* Human colon cancer epithelial cells harbour active HEDGEHOG-GLI signalling that is essential for tumour growth, recurrence, metastasis and stem cell survival and expansion. *EMBO Mol Med* **1**, 338–351 (2009).
71. Gaggianesi, M. *et al.* Messing Up the Cancer Stem Cell Chemoresistance Mechanisms Supported by Tumor Microenvironment. doi:10.3389/fonc.2021.702642.
72. Oskarsson, T., Batlle, E. & Massagué, J. Metastatic stem cells: Sources, niches, and vital pathways. *Cell Stem Cell* vol. 14 306–321 Preprint at https://doi.org/10.1016/j.stem.2014.02.002 (2014).
73. Kitai, H. *et al.* Epithelial-to-mesenchymal transition defines feedback activation of receptor tyrosine kinase signaling induced by MEK inhibition in KRAS-mutant lung cancer. *Cancer Discov* **6**, 754–769 (2016).
74. v11a86-seigel.
75. Li, Y., Wang, Z., Ajani, J. A. & Song, S. Drug resistance and Cancer stem cells. *Cell Communication and Signaling* vol. 19 Preprint at https://doi.org/10.1186/s12964-020-00627-5 (2021).
76. Kreso, A. *et al.* Variable clonal repopulation dynamics influence chemotherapy response in colorectal cancer. *Science (1979)* **339**, 543–548 (2013).
77. Catalano, V. *et al.* Colorectal cancer stem cells and cell death. *Cancers* vol. 3 1929–1946 Preprint at https://doi.org/10.3390/cancers3021929 (2011).
78. Costello, R. T. *et al. The immunophenotype of minimally differentiated acute myeloid leukemia (AML-M0): reduced immunogenicity and high frequency of CD34 + /CD38 − leukemic progenitors*. *Leukemia* vol. 13 http://www.stockton-press.co.uk/leu (1999).
79. Unterkircher, T. *et al.* Bortezomib primes glioblastoma, including glioblastoma stem cells, for TRAIL by increasing tBid stability and mitochondrial apoptosis. *Clinical Cancer Research* **17**, 4019–4030 (2011).
80. Loebinger, M. R., Eddaoudi, A., Davies, D. & Janes, S. M. Mesenchymal stem cell delivery of TRAIL can eliminate metastatic cancer. *Cancer Res* **69**, 4134–4142 (2009).
81. Zobalova, R. *et al.* CD133-positive cells are resistant to TRAIL due to up-regulation of FLIP. *Biochem Biophys Res Commun* **373**, 567–571 (2008).
82. Liu, G. *et al.* Analysis of gene expression and chemoresistance of CD133+ cancer stem cells in glioblastoma. *Mol Cancer* **5**, (2006).
83. Sun, Q. *et al.* Proapoptotic PUMA targets stem-like breast cancer cells to suppress metastasis. *Journal of Clinical Investigation* **128**, 531–544 (2018).
84. Kise, K., Kinugasa-Katayama, Y. & Takakura, N. Tumor microenvironment for cancer stem cells. *Adv Drug Deliv Rev* **99**, 197–205 (2016).
85. Adisetiyo, H. *et al.* Dependence of castration-resistant prostate cancer (CRPC) stem cells on CRPC-associated fibroblasts. *J Cell Physiol* **229**, 1170–1176 (2014).
86. Hovinga, K. E. *et al.* Inhibition of notch signaling in glioblastoma targets cancer stem cells via an endothelial cell intermediate. *Stem Cells* **28**, 1019–1029 (2010).
87. Schwitalla, S. *et al.* Intestinal tumorigenesis initiated by dedifferentiation and acquisition of stem-cell-like properties. *Cell* **152**, 25–38 (2013).
88. LeBleu, V. S. & Kalluri, R. A peek into cancer-associated fibroblasts: Origins, functions and translational impact. *DMM Disease Models and Mechanisms* **11**, (2018).
89. Liao, C. P. *et al.* Androgen receptor in cancerassociated fibroblasts influences stemness in cancer cells. *Endocr Relat Cancer* **24**, 157–170 (2017).
90. Tsuyada, A. *et al.* CCL2 mediates cross-talk between cancer cells and stromal fibroblasts that regulates breast cancer stem cells. *Cancer Res* **72**, 2768–2779 (2012).

91. Li, Y. *et al.* Cancer-associated fibroblasts promote the stemness of CD24 + liver cells via paracrine signaling. *J Mol Med* **97**, 243–255 (2019).
92. Hasegawa, T. *et al.* Cancer-associated fibroblasts might sustain the stemness of scirrhous gastric cancer cells via transforming growth factor-β signaling. *Int J Cancer* **134**, 1785–1795 (2014).
93. Tang, Y. A. *et al.* Hypoxic tumor microenvironment activates GLI2 via HIF-1α and TGF-β2 to promote chemoresistance in colorectal cancer. *Proc Natl Acad Sci U S A* **115**, E5990–E5999 (2018).
94. Hervieu, C., Christou, N., Battu, S. & Mathonnet, M. The role of cancer stem cells in colorectal cancer: From the basics to novel clinical trials. *Cancers (Basel)* **13**, 1–26 (2021).
95. Bonnet, D. & Dick, E. *Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell*. http://www.nature.com/naturemedicine (1997).
96. 22236375 (1).
97. De Angelis, M. L., Francescangeli, F., Zeuner, A. & Baiocchi, M. Colorectal cancer stem cells: An overview of evolving methods and concepts. *Cancers* vol. 13 Preprint at https://doi.org/10.3390/cancers13235910 (2021).
98. Dalerba, P. *et al. Phenotypic characterization of human colorectal cancer stem cells*. www.pnas.org/cgi/content/full/ (2007).
99. Ricci-Vitiani, L. *et al.* Identification and expansion of human colon-cancer-initiating cells. *Nature* **445**, 111–115 (2007).
100. O'Brien, C. A., Pollett, A., Gallinger, S. & Dick, J. E. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature* **445**, 106–110 (2007).
101. Todaro, M. *et al.* Colon Cancer Stem Cells Dictate Tumor Growth and Resist Cell Death by Production of Interleukin-4. *Cell Stem Cell* **1**, 389–402 (2007).
102. Horst, D. *et al.* The cancer stem cell marker CD133 has high prognostic impact but unknown functional relevance for the metastasis of human colon cancer. *Journal of Pathology* **219**, 427–434 (2009).
103. Ricci-Vitiani, L., Fabrizi, E., Palio, E. & De Maria, R. Colon cancer stem cells. *Journal of Molecular Medicine* vol. 87 1097–1104 Preprint at https://doi.org/10.1007/s00109-009-0518-4 (2009).
104. Schepers, A. G. *et al.* Lineage Tracing Reveals Lgr5+ Stem Cell Activity in Mouse Intestinal Adenomas. *Science (1979)* **337**, 730–735 (2012).
105. Shimokawa, M. *et al.* Visualization and targeting of LGR5 + human colon cancer stem cells. *Nature* **545**, 187–192 (2017).
106. Beiraghdar, M., Talebi, A. & Kianersi, K. Comparison of gene expression of SOX2 and OCT4 in normal tissue, polyps, and colon adenocarcinoma using immunohistochemical staining. (2015) doi:10.4103/2277-9175.167958.
107. Neumann, J. *et al.* SOX2 expression correlates with lymph-node metastases and distant spread in right-sided colon cancer. *BMC Cancer* **11**, (2011).
108. Yu, H., Pardoll, D. & Jove, R. STATs in cancer inflammation and immunity: A leading role for STAT3. *Nature Reviews Cancer* vol. 9 798–809 Preprint at https://doi.org/10.1038/nrc2734 (2009).
109. Spitzner, M. *et al.* STAT3 inhibition sensitizes colorectal cancer to chemoradiotherapy in vitro and in vivo. *Int J Cancer* **134**, 997–1007 (2014).

110. Marin, J. J. G. *et al.* Chemoprevention, chemotherapy, and chemoresistance in colorectal cancer. *Drug Metab Rev* **44**, 148–172 (2012).
111. Holohan, C., Van Schaeybroeck, S., Longley, D. B. & Johnston, P. G. Cancer drug resistance: An evolving paradigm. *Nature Reviews Cancer* vol. 13 714–726 Preprint at https://doi.org/10.1038/nrc3599 (2013).
112. Li, Q. & Shu, Y. *Role of solute carriers in response to anticancer drugs*. http://www.molcelltherapies.com/content/2/1/15 (2014).
113. Buck, E. *et al.* Tumor response to irinotecan is associated with CYP3A5 expression in colorectal cancer. *Oncol Lett* **17**, 3890–3898 (2019).
114. Misale, S. *et al.* Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. *Nature* **486**, 532–536 (2012).
115. 11731512.
116. Hu, T., Li, Z., Gao, C. Y. & Cho, C. H. Mechanisms of drug resistance in colon cancer and its therapeutic strategies. *World Journal of Gastroenterology* vol. 22 6876–6889 Preprint at https://doi.org/10.3748/wjg.v22.i30.6876 (2016).
117. Natoli, M., Leoni, B. D., D'Agnano, I., Zucco, F. & Felsani, A. Good Caco-2 Cell Culture Practices. *Toxicology in Vitro* **26**, 1243–1246 (2012).
118. Cañellas-Socias, A. *et al.* Metastatic recurrence in colorectal cancer arises from residual EMP1+ cells. *Nature* **611**, 603–613 (2022).
119. *Sample to Insight__ RNeasy ® Plus Mini Handbook*. (2020).
120. Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Res* **50**, W345–W351 (2022).
121. Ge, S. X., Son, E. W. & Yao, R. iDEP: An integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinformatics* **19**, (2018).
122. Norkin, M., Ordóñez-Morán, P. & Huelsken, J. High-content, targeted RNA-seq screening in organoids for drug discovery in colorectal cancer. *Cell Rep* **35**, (2021).
123. Krueger F. Trim Galor. (2021).
124. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907–915 (2019).
125. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
126. Moor, A. E. *et al.* Spatial Reconstruction of Single Enterocytes Uncovers Broad Zonation along the Intestinal Villus Axis. *Cell* **175**, 1156-1167.e15 (2018).
127. Gassler, N. *et al.* Molecular characterisation of non-absorptive and absorptive enterocytes in human small intestine. *Gut* **55**, 1084–1089 (2006).
128. Diem-Phuong D. Dao 1, P. H. L. 2. Histology, Goblet Cells. *In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan. 2023 Mar 15.*
129. 16151588.
130. Lizarraga, E. *et al.* Clinical and functional analysis of SOX9 in colorectal cancer. *Journal of Clinical Oncology* **37**, 519 (2019).
131. Hughes, C. E. & Nibbs, R. J. B. A guide to chemokines and their receptors. *FEBS Journal* vol. 285 2944–2971 Preprint at https://doi.org/10.1111/febs.14466 (2018).
132. Yang, E. V. Role for catecholamines in tumor progression: Possible use for β-blockers in the treatment of cancer. *Cancer Biology and Therapy* vol. 10 30–32 Preprint at https://doi.org/10.4161/cbt.10.1.12260 (2010).
133. Zhao, Z. *et al.* Organoids. *Nature Reviews Methods Primers* **2**, (2022).

134. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90–W97 (2016).
135. Fu, M. *et al.* The Hippo signalling pathway and its implications in human health and diseases. *Signal Transduction and Targeted Therapy* vol. 7 Preprint at https://doi.org/10.1038/s41392-022-01191-9 (2022).
136. Junk, D. J. *et al.* Oncostatin M promotes cancer cell plasticity through cooperative STAT3-SMAD3 signaling. *Oncogene* **36**, 4001–4013 (2017).
137. Donaudy, F. *et al.* Espin gene (ESPN) mutations associated with autosomal dominant hearing loss cause defects in microvillar elongation or organisation. *J Med Genet* **43**, 157–161 (2006).
138. De Sousa E Melo, F. *et al.* A distinct role for Lgr5 + stem cells in primary and metastatic colon cancer. *Nature* **543**, 676–680 (2017).
139. Barone, C. *et al.* Schedule-dependent activity of 5-fluorouracil and irinotecan combination in the treatment of human colorectal cancer: In vitro evidence and a phase I dose-escalating clinical trial. *Br J Cancer* **96**, 21–28 (2007).
140. Telang, N. Drug-Resistant Stem Cells: Novel Approach for Colon Cancer Therapy. *International Journal of Molecular Sciences* vol. 23 Preprint at https://doi.org/10.3390/ijms23052519 (2022).
141. Ordóñez-Morán, P., Dafflon, C., Imajo, M., Nishida, E. & Huelsken, J. HOXA5 Counteracts Stem Cell Traits by Inhibiting Wnt Signaling in Colorectal Cancer. *Cancer Cell* **28**, 815–829 (2015).
142. Hsu, H. C. *et al.* Overexpression of Lgr5 correlates with resistance to 5-FU-based chemotherapy in colorectal cancer. *Int J Colorectal Dis* **28**, 1535–1546 (2013).
143. Yuan, S. *et al.* Role of Wnt/ β-Catenin Signaling in the Chemoresistance Modulation of Colorectal Cancer. *Biomed Res Int* **2020**, (2020).
144. Bu, P. *et al.* Aldolase B-Mediated Fructose Metabolism Drives Metabolic Reprogramming of Colon Cancer Liver Metastasis. *Cell Metab* **27**, 1249-1262.e4 (2018).
145. De Las Rivas, J. *et al.* Cancer drug resistance induced by EMT: novel therapeutic strategies. *Archives of Toxicology* vol. 95 2279–2297 Preprint at https://doi.org/10.1007/s00204-021-03063-7 (2021).
146. Ni, Q., Li, M. & Yu, S. Research Progress of Epithelial-mesenchymal Transition Treatment and Drug Resistance in Colorectal Cancer. *Technology in Cancer Research and Treatment* vol. 21 Preprint at https://doi.org/10.1177/15330338221081219 (2022).
147. Sabouni, E. *et al.* Unraveling the function of epithelial-mesenchymal transition (EMT) in colorectal cancer: Metastasis, therapy response, and revisiting molecular pathways. *Biomedicine and Pharmacotherapy* vol. 160 Preprint at https://doi.org/10.1016/j.biopha.2023.114395 (2023).
148. Long, S. *et al.* ECM1 regulates the resistance of colorectal cancer to 5-FU treatment by modulating apoptotic cell death and epithelial-mesenchymal transition induction. *Front Pharmacol* **13**, (2022).
149. Hanns, P., Paczulla, A. M., Medinger, M., Konantz, M. & Lengerke, C. Stress and catecholamines modulate the bone marrow microenvironment to promote tumorigenesis. *Cell Stress* vol. 3 221–235 Preprint at https://doi.org/10.15698/cst2019.07.192 (2019).

150. Houthuijzen, J. M., Daenen, L. G. M., Roodhart, J. M. L. & Voest, E. E. The role of mesenchymal stem cells in anti-cancer drug resistance and tumour progression. *British Journal of Cancer* vol. 106 1901–1906 Preprint at https://doi.org/10.1038/bjc.2012.201 (2012).
151. pmc_2435618.
152. Oren, Y. *et al.* Cycling cancer persister cells arise from lineages with distinct programs. *Nature* **596**, 576–582 (2021).
153. Bhattacharya, S., Calar, K. & De La Puente, P. Mimicking tumor hypoxia and tumor-immune interactions employing three-dimensional in vitro models. *Journal of Experimental and Clinical Cancer Research* vol. 39 Preprint at https://doi.org/10.1186/s13046-020-01583-1 (2020).
154. KOJI MURONO, NELSON H. TSUNO & KAZUSHIGE KAWAI. *SN-38 Overcomes Chemoresistance of Colorectal Cancer Cells Induced by Hypoxia, through HIF1alpha*. (2012).
155. Yeung, T. M., Gandhi, S. C. & Bodmer, W. F. Hypoxia and lineage specification of cell line-derived colorectal cancer stem cells. *Proc Natl Acad Sci U S A* **108**, 4382–4387 (2011).