Transcription Start Site Usage in Cancer Heterogeneity

Paige-Louise White

A thesis submitted to the University of Birmingham for the degree of **DOCTOR OF PHILOSOPHY**

UNIVERSITY OF BIRMINGHAM

Institute of Cancer and Genomic Sciences
College of Medical and Dental Sciences
University of Birmingham

April 2024

Abstract

Resolving intra tumour heterogeneity (ITH) remains an extensive therapeutic challenge in cancer research. Despite an emphasis on the importance of the transcriptional landscape in understanding cancer cell behaviours, the role of the core promoter remains widely ignored. Recent studies have shown the presence of alternative transcription start sites (TSS) and found a distinctive YC TSS enrichment in cancers that correlated with global phenotypes. However, tumour dynamics are underpinned by a complex network of sub clonal diversity therefore, the true biological significance of TSS usage remains elusive. Here we further characterise the phenotypic landscape of YC enriched cancers on both an inter and intra level. Through successful undertaking of single cell CAGE sequencing, we identify ITH of TSS usage even at a single cell level. We find differential YC enrichment associated with cellular physiological states that facilitate a growth phenotype. Next, we identify cell cycle phase associated TSS shifting and find YC enrichment associated with G0 and G1 phases. Scrutinization of genes utilising YC enrichment showed subsets critically involved in directing the metabolic state of the cell. Overall, our findings highlight an important TSS association to cancer cell growth dynamics that act in a cell cycle dependent manner, even at a single cell resolution. Finally, attempts to characterise the spatiotemporal mapping of cycling phases through imaging emphasised the dynamic landscape of cancer behaviours. The work presented here further highlights the importance of TSS usage investigations that are masked in conventional gene expression analysis.

Acknowledgements

First and foremost, I would like to express sincere gratitude and appreciation for the support and guidance from Ferenc. From continual understanding to extensive encouragement of my ideas no matter how absurd they may seem. I hope to instil the same confidence in future students as he did in me. I am extremely thankful to Andy and Dov for never hesitating to provide their invaluable insight, enthusiasm, and encouragement in both the project and me.

To all the past, present and visiting members of the Mueller lab, I am forever grateful for the endless support you have given me. From finding the perfect balance between kindness and tough love, you have made the whole process undoubtedly easier. Neal and Nelson, your inability to say no to long runs, snack trips and cat café outings was the fuel that kept me going. Ada your no contact, high patience, and compassionate approach to everything you do has taught me how to be both a better scientist and person. To Yavor, the backbone of the Mueller lab, without whom much of this work would not be possible, I am extremely grateful for the patience you held in teaching me the computational ways and never complaining about my repetitive questions. Finally, to Joe who initiated and supported so many of the ideas that fuelled my PhD.

To my extended family, my friends many of whom I had the privilege of meeting because of this PhD. You were the support system. You stepped up as family and made the distance from my own family much more bearable. Sophie, your ability to listen to my daily rants in ridiculously long voice notes without ever complaining has made sure I never went lonely. Phoebe, my fellow P, without you, sunset walks and countless cups of tea and therapy the success of my final year would not have been possible. Kass, my soul mate, for always reminding me I could retire and just open a bookshop but continuously ensuring I didn't. To the many others who provided shoulders to cry on, or lean on in fits of laughter (Connor, I am so grateful your humour is as silly as mine). I owe so much to the Pattersons you all made me who I am I cannot express enough how much I appreciate you all.

Finally, but most importantly to my family those with me now and those who I have lost along the way. Thank you for never letting me doubt myself and always ensuring I had the strength to keep trying. To my parents, who always encouraged me to be undoubtedly myself even if it meant constant waves of chaos and unpredictability. You constantly adapted and made sure I did to. Thank you for providing me with the will to be unapologetically independent and for giving me the 4 best friends I could ever ask for. My brothers, for always treating me like a brother and making sure I always knew what I was truly capable of.

I dedicate this thesis to my Grandparents who didn't get to see me finish but who fuelled my desire to make it happen.

Contents

List of Figures

List of Tables

List of Abbreviations

AP	Alternative Promoter
ATG	Alternative Translation Genes
ATM	Alternative Translation Machinery
ATP	Adenosine Triphosphate
BME	Basement Membrane Extract
CAGE	Cap Analysis of Gene Expression
CC	Consensus Cluster
CCA	Canonical Correlation Analysis
CCID	Consensus Cluster Identification
CDK	Cyclin Dependent Kinases
CMS	Consensus Molecular Subtypes
CRC	Colorectal Cancer
CRE	Cis-regulatory Elements
CSC	Cancer Stem Cell
CTSS	Consensus Transcription Start Site
DDR	DNA Damage Repair
DEG	Differentially Expressed Genes
DGE	Differential Gene Expression
eIF4B	Eukaryotic Initiating Factor 4B
FUCCI	Fluorescence Ubiquitination based Cell Cycle Indicator
GO	Gene Ontology
GSEA	Gene Set Enrichment Analysis
IRES	Internal Ribosome Entry Site
ISC	Intestinal Stem Cell
ITH	Intra Tumour Heterogeneity
LARP	La-Related Protein
LMP	Low Melting Point
MET	Multitask Evolution Theory
MSI	Microsatellite Instability
mTOR	Mammalian Target of Rapamycin
OHS	Organoid Harvesting Solution
PABPC1	Poly(A)-Binding Protein C1
PB	Processing Bodies
PCA	Principal Component Analysis
PFA	Paraformaldehyde
PS	Post Seeding
PDO	Patient Derived Organoids
PIC	Pre Initiation Complex
RBP	Ribosomal Binding Proteins
ROS	Reactive Oxygen Species
RT	Reverse Transcriptase
SC	Sub Clone

SCAFE	Single-Cell Analysis of Five-prime Ends
SD	Standard Deviation
SG	Stress Granules
SLIC-CAGE	Super-Low Input Carrier CAGE
snoRNA	Small Nucleolar RNA
SOM	Self-organising Maps
TA	Transit Amplifying
TAF	TBP-Associated Factors
TBP	TATA Binding Protein
TF	Transcription Factors
TFBS	Transcription Factor Binding Sites
TIM	Translation Initiation Machinery
TME	Tumour Microenvironment
TNM	Tumour Node Metastasis
TOP	Terminal Oligopyrimidine
TPM	Tags Per Million
TRF2	TATA Box Binding Protein Related Factor
TSC1/2	Tuberous Sclerosis Complex ½
TSO	Template Switching Oligonucleotide
TSS	Transcription Start Site
UMAP	Uniform Manifold Approximation and Projection
UMI	Unique Molecular Identifier

Chapter 1: Introduction

1.1 Cancer Heterogeneity

1.1.1 Inter vs Intra tumour heterogeneity

The sheer variability that exists in cancer is simply highlighted in its definition as an umbrella term for over 200 different diseases (Iacona & Lutz, 2019; Miller et al., 2019). Despite these morphologically characterised subtypes, often little correlation in treatment response exists between tumours of the same subtype (Burrell et al., 2013; Dagogo-Jack & Shaw, 2018). Thus, highlighting that the complexity of a tumour's aetiology is not just bound by factors such as the location in which it is found. Tumour heterogeneity is one of the biggest hurdles for diagnostic and treatment progression (Dagogo-Jack & Shaw, 2018; Fisher et al., 2013; Stanta & Bonin, 2018a). This heterogeneity occurs in two ways. Firstly, inter heterogeneity whereby tumour phenotype differs between subtypes and patients. Secondly, intra heterogeneity (ITH) refers to the differences seen within a single tumour (Hausser & Alon, 2020; Prasetyanti & Medema, 2017). Akin to developing embryos, tumour architecture is comprised of distinct sub cellular populations defined by differing fate potentials and phenotypic presentations (Y. Ma et al., 2010). These subclones (SC) can vary considerably in their genetic, epigenetic, and even transcriptomic backgrounds and impact overall tumour behaviour (Fisher et al., 2013; McGranahan & Swanton, 2017). By understanding the molecular contributions for both inter and intra heterogeneity and their subsequent interaction, we can better resolve the genotype-phenotype map encompassing tumour formation, development, and response (Lenz et al., 2021; Cho and Przytycka, 2013b. Hence deepening tumour stratification criteria facilitating the goals of precision medicine (Al-Hamaly et al., 2023).

1.1.2 Gastrointestinal morphology and Colorectal Cancer

The impact of tumour heterogeneity is widely seen in colorectal cancer (CRC) which has the third highest cancer mortality rate worldwide (Bray et al., 2018; Morgan et al., 2023). CRC is a complex disease that develops through accumulation of erroneous genetic and non-

genetic alterations in the gastrointestinal cells that line the crypt-villus axis (De Rosa et al., 2015; Punt et al., 2017). The intricate morphology of the gastrointestinal system outlined in **Figure 1.1**, is composed of diverse cell types to maintain the core barrier function essential for molecular uptake and protection of other tissues (Shen, 2009). This function is held up by the compartmentalisation of specialised cells that exist and maintain the villi that line mucosal layers (Shen, 2009; Sumigray et al., 2018). The crypt, found at the very base harbours not only Paneth and goblet cells, but also LGR5+ intestinal epithelial stem cells (ISC) and undifferentiated secretory cells in a region known as the transit amplifying zone (Foerster et al., 2022; Shen, 2009). Here cells are terminally differentiating and migrating vertically towards the villi tip into absorption specialised cells such as enterocytes and tuft, before being expelled from the epithelium (Foerster et al., 2022; Shen, 2009; Sumigray et al., 2018). This repair and replenishment axis is maintained by vital Wnt signalling acting on ISC controlling their division and progression up the villi in a concentrated gradient (Krausova & Korinek, 2014; Mah et al., 2016; Walter et al., 2022). This highly active and dynamic architecture is often mimicked in CRC aetiology, resulting in tumour cellular composition and disease outcome being vastly heterogeneous (Fedi et al., 2021; Punt et al., 2017).

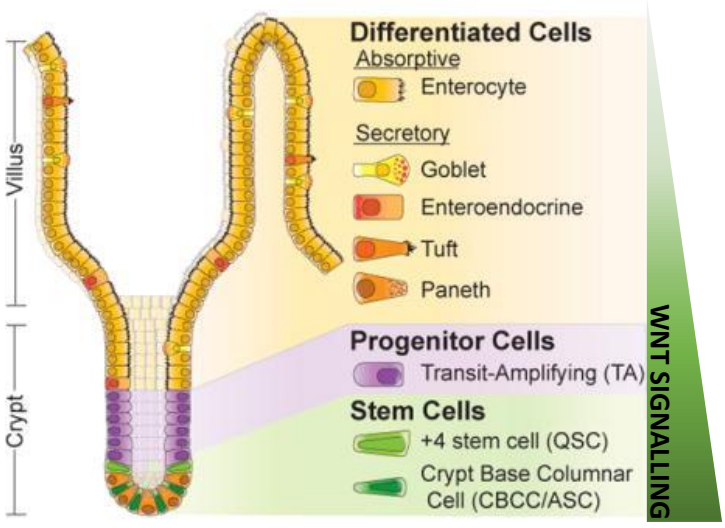


Figure 1.1: Morphological characterisation of intestinal crypt structure. Cellular composition and ordering of the intestinal crypt begin at the base where Paneth cells are intertwined with naïve cells such as crypt base columnar cells and stem cells. The path of differentiation extends vertically up the crypt to the villus through a gradient of tightly controlled Wnt signalling. Just above the crypt base is a highly proliferative zone harbouring the transit amplifying cells with resultant trajectories into functional secretory and absorptive differentiated cells lines the villus tip. *Adapted from* (Carulli et al., 2014).

1.1.3 Drivers of CRC and disease stratification

For years CRC stratification was mainly built on the conventional tumour, node, metastasis (TNM) staging system with microsatellite instability (MSI) status later also providing prognostic value (Punt et al., 2017; M. P. Singh et al., 2021; Zaborowski et al., 2022). However, efforts to better characterise CRC resulted in the identification of consensus molecular subtypes (CMS) with translational prognostic capabilities (Guinney et al., 2015; Khaliq et al., 2022). Utilising a wealth of sequencing data, the CMS classification system provides an insight and foundation of understanding into the heterogeneity of the biological underpinnings of CRC (Guinney et al., 2015; M. P. Singh et al., 2021). The first layer of classification comes from the genomic-focused aberrations with CMS1 defined by the hypermutable microsatellite instable genotype (Guinney et al., 2015). Contrastingly the remaining 3 are characterised by significant somatic copy number alterations underpinned by high chromosomal instability particularly in oncogenes (Guinney et al., 2015; Rejali et al., 2023). The second layer utilises transcriptomic and epigenomic analysis and highlighted stark differences and establishes the now accepted taxonomy. CMS1, the MSI immune subtype, shows distinct upregulation of pathways vital for immune response (Guinney et al., 2015). CMS2 the canonical subtype making up the largest portion of CRC cases, shows significant oncogenic amplification coupled with both WNT and MYC upregulation (Guinney et al., 2015). CMS3 has been identified as the metabolic subtype with tumours characterised as this displaying both KRAS mutations and upregulation of metabolic signatures (Guinney et al., 2015). Finally, CMS4, identified as the poorest prognoses subtype, is named aptly as mesenchymal due to the upregulation of genes known to be important for epithelial to mesenchymal transition and other stromal infiltrating genes (Guinney et al., 2015; Hoorn et al., 2022).

The necessity to combine both genomic and transcriptomic features to define the 4 diverse subtypes highlights the importance of the transcriptome in cancer heterogeneity, further solidifying the belief that cancer cannot be resolved by the genome alone (R. J. Chen et al., 2022). A finding further supported by Wanigasooriya et al, who used transcriptomic analysis and identified the dysregulation of the mTOR pathway in CRC correlated with irradiation responsiveness (Wanigasooriya et al., 2022). Through

combination therapy of blocking the mTOR/PI3K pathway they were able to sensitise non-responsive CRC to irradiation, highlighting an important role of these pathways in CRC (Wanigasooriya et al., 2022). This research resulted in a novel approach to treating resistant CRC that would not have been achieved by investigating genomic alterations alone.

1.1.4 Modelling CRC using organoids

A challenge in resolving heterogeneity is the inability to understand and observe its true extent in tumour dynamics due to time point bias (Cyll et al., 2017). However, the development of CRC patient derived organoids (PDO), which sufficiently recapitulate the tumour cellular organisation missing from 2D cultures, allows for this to be further studied (Kaushik et al., 2018; Wanigasooriya et al., 2022). The methodology often requires the use and isolation of either healthy or cancerous tissue specific stem cells (CSC), which are then embedded in a 3D dome of nutrient containing Matrigel (Yin et al., 2016). An advanced pooling approach which combines CSC and other cancer, and normal donor cells shows better retention of heterogeneity and cellular composition seen in tumours (Lancaster & Huch, 2019; Xiaolei et al., 2017). Co-culturing with mature cells such as fibroblasts also resulted in organoids that represent the CMS (Atanasova et al., 2023; Sayed et al., 2021). Furthermore, PDOs have been shown to retain the molecular features and morphology of their organ of origin, thus creating an exciting modelling system to investigate the full extent of the phenotypic landscape of tumour growth and treatment response (Betge et al., 2022; Jeong et al., 2022; Sayed et al., 2021). In combination with the advent of single cell technologies, CRC organoids represent an opportunity to better understand ITH (Bowes et al., 2022; F. Wu et al., 2021).

1.2 Intra-tumour heterogeneity and the role of cellular states

1.2.1 Theories of Intra-tumour heterogeneity

Although there is merit in investigating the differences between tumours (inter), whole tumour analysis may overlook important findings happening within a single tumour

(intra) (Ramón y Cajal et al., 2020). The degree of heterogeneity within a tumour is known to indicate a poorer prognosis (Ramón y Cajal et al., 2020; T. Yu et al., 2021). Further to this, research has shown populations of cells within a single tumour can respond differently to treatment and may provide explanations to why some patients relapse with more aggressive tumours (Ling et al., 2015; Marusyk & Polyak, 2010; Ohshima & Morii, 2021; Punt et al., 2017; Roerink et al., 2018). Many have likened ITH dynamics to that seen in evolution. Clonal evolution theory states that SC trajectories occur through adaptations obtained from driver events to maintain a clonal advantage, thus coinciding with Darwinian natural selection (Hausser & Alon, 2020; Prasetyanti & Medema, 2017).

Conversely, the identification of CSC in many subtypes, fuelled the theory that ITH and thus production, growth and maintenance of tumours relies on a subset of cells, that abide to similar principles of regeneration (Prasetyanti & Medema, 2017). These CSC are capable of both robust self-renewal, thus maintaining plasticity, and differentiation into further defined progeny, therefore producing heterogenous populations (Clevers, 2011; Sasaki & Clevers, 2018). Despite previously attributed as opposing theories, the coalition of both clonal evolution theory and CSC theory highlights the plasticity of tumour cell behaviour and the significant role the environment plays in defining tumour dynamics (Marusyk et al., 2012). By better understanding the adaptive and resultant dysregulated behaviours that uphold tumour phenotypes we can greaten our understanding of cancer aetiology (Marusyk et al., 2020; Marusyk & Polyak, 2010).

1.2.2 Cellular states within tumours

Resolving tumour evolution requires understanding SC molecular diversity and the intrinsic and adaptive functional capabilities that drive tumour function and response (X. X. Sun & Yu, 2015). Characterising these subpopulations has identified many layers of ITH with differentially defined phenotypes (Maleki et al., 2024a; Stanta & Bonin, 2018a). For example, tumour composition is shown to be vastly heterogeneous in defined cell types, often mimicking host organ cellular architecture (Jögi et al., 2012). The building of this ITH is believed to be mainly due to CSCs (Naz et al., 2021; Wahl & Spike, 2017). CSCs harbour distinct phenotypic plasticity with resultant trajectories to cell types resulting through accumulation of genetic and non-genetic alterations driven by tumour microenvironment

(TME) interaction (Nairuz et al., 2023). For example, CDX1, a transcription factor vital for intestinal differentiation, is believed to play a central role in CSC differentiation in CRC (Ashley et al., 2013; Jones et al., 2015). CDX1 levels are greatly affected by extrinsic factors such as nutrient availability and more specifically hypoxia (Ashley et al., 2013; Jones et al., 2015). It is hypothesised that the promoter region harbours HIF1a target sites therefore, in hypoxic environments differentiation is prevented and cells are kept in a stem like quiescent state driving resistance (Ashley et al., 2013; Najafi et al., 2020). This highlights the necessity of adaptability and hence plasticity of cellular states to maintain or regain homeostasis to aid tumour growth and survival (Wahl & Spike, 2017).

1.2.3 Cancer stem cells and Cell cycle regulation

To maintain continual plasticity, CSC undergo asymmetric division resulting in more differentiated and stemlike progeny thus self-renewal (Chao et al., 2023; Z. Li et al., 2022). The asymmetrical dissemination of fate-determining factors including Notch and Wnt during mitosis direct the progeny fates (Z. Li et al., 2022). Therefore, the cell cycle is believed to be one of the key driving forces in CSC fates and consequently in cancer formation and progression (Lindell et al., 2023). The cell cycle is tightly regulated and highly dysregulated in cancer making it one of the most prominent targets in anticancer therapies (Jingwen et al., 2017; Patra et al., 2023). In many cancers the inhibitors shown in **Figure 1.2**, known to regulate the formation of the cyclin and cyclin dependent kinases (CDK) complexes are often silenced through mutation or enhanced protein degradation (Yin Liu et al. 2017; Otto and Sicinski 2017). The cyclin-CDK complexes are then able to phosphorylate protein targets allowing for many transcription factors (TF) to activate genes associated with progression through the cell cycle (Yin Liu et al. 2017). Many of these genes show cyclic expression in addition to a distinctive delay between active transcription and steady state RNA expression, highlighting how fine-tuned transcriptional regulation in cellular division is (Boström et al., 2017; Y. Liu et al., 2017). Further resolution of the role of the transcriptome in the dysregulation of the cell cycle in diseases such as cancer may provide more therapeutic targets and markers for the disease.

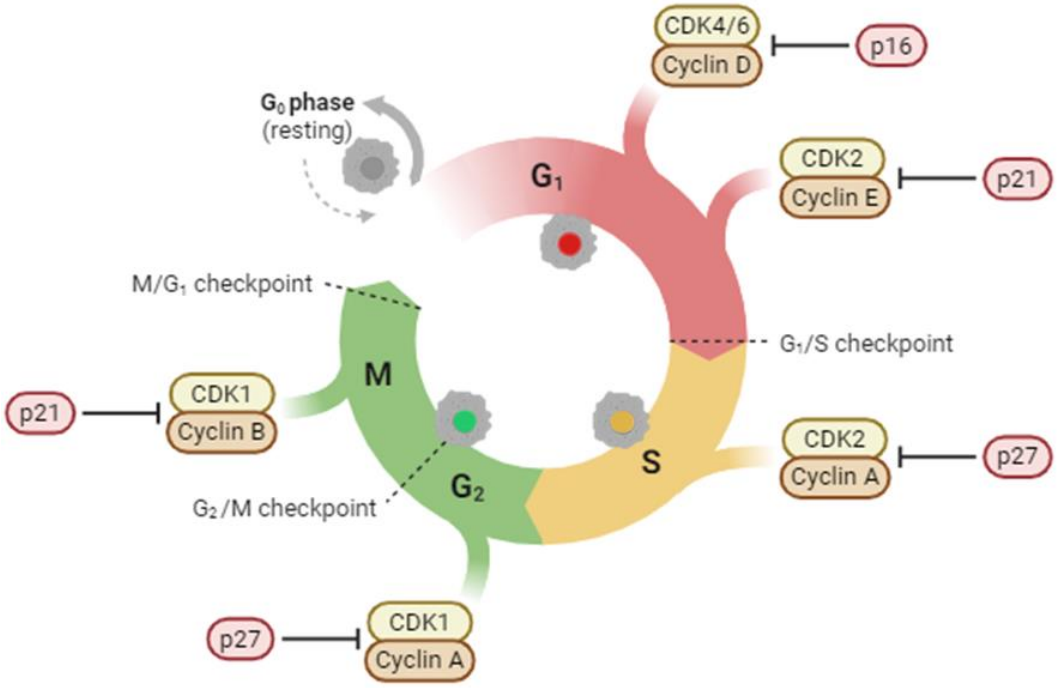


Figure 1.2: Cell Cycle regulation and characterisation of phases using FUCCI. Cell cycle progression is mediated by the formation of phase dependent CDK complexes and are essential for transitions through checkpoints. Progression can be halted through the upregulation of inhibitors p21 and p16. Cycle dynamics have been characterised through labelling of key phase associated genes with fluorescent reporters. FUCCI labels the G1 specific highly expressing CDT1 with mKO2 resulting in a red or orange nuclei, as cells transition into S phase CDT1 degradation coupled with Geminin upregulation that has been tagged with the Monomeric Azami-Green fluorescent protein, results in a yellow emitting nucleus, following complete degradation of CDT1 and hence G2M phase nuclei will be green before losing all fluorescence upon exiting the cell cycle into G0. *Adapted from* (Jingwen et al., 2017; Yano et al., 2020). *Made in Biorender.*

1.2.4 Cell cycle dysregulation and resistance

The advent of visualisation techniques such as FUCCI (fluorescence ubiquitination-based cell cycle indicator), which fluorescently label the different cycle phases: G1,S,G2M **Figure 1.2**, has also shown ITH populations have different cycle physiology (Newman & Zhang, 2008; A. M. Singh et al., 2016; Zielke & Edgar, 2015). In addition to differentiation status slowing cycle iterations, cellular spatial location also shows altered dynamics with more proliferative cells located at the outer tumour regions (Yano et al., 2014, 2020). Through efficient upregulation of DNA damage repair (DDR) mechanisms during G2M, cycle and phase lengths can also be tuned (Maleki et al., 2024b). Utilisation of FUCCI labelling also revealed proliferative cells were more responsive to cytotoxic agents whereas quiescent cells show a distinctive resistance (Yano et al., 2020). The ability of

cells to enter this quiescent state in response to stressors is vital to cellular survival highlighting an important role of the cell cycle in cell adaptability. Maintaining the balance between driving to and away from a proliferative state is vital for cancer cell survival (Kumari & Jat, 2021; Loftus et al., 2022). Fuelling two core principles established in the notable hallmarks of cancer: sustaining proliferative signals and suppressing antigrowth signals and highlighting the significance of cycle regulation in cancer (Hanahan & Weinberg, 2000, 2011).

1.2.5 Metabolic adaptability in cancer (The proliferation equilibrium)

A key mediator of maintaining proliferative capabilities is the metabolic state of the cell (Leal-Esteban & Fajas, 2020; J. Liu et al., 2022; Nong et al., 2023). To survive the high energy demand required to continue cycling, cancer cells undergo metabolic reprogramming (Leal-Esteban & Fajas, 2020; J. Liu et al., 2022). The Warburg effect, an aspect of reprogramming established in the 1920s, highlights cancer cell preference for adenosine triphosphate (ATP) and lactate production through glycolysis (Liberti & Locasale, 2016; Navarro et al., 2022; Nong et al., 2023). The absence of complete aerobic respiration represents a more efficient and thus beneficial approach to energy and metabolic intermediate production in stressed states (X. Li et al., 2022; Zheng, 2012). The complex multifaceted shift also produces metabolic precursors for amino acids, nucleotide, lipid, and carbohydrate biosynthesis all essential for growth and survival (Nenkov et al., 2021; Nong et al., 2023; Schiliro & Firestein, 2021). This abundancy of metabolic intermediates enables cells to drive their proliferation (DeBerardinis et al., 2008). The ability of cells to perform this is believed to be core in defining cellular behaviour and may even be a vital mechanism underpinning retention of stemness (Papadaki & Magklara, 2022; G. M. Wen et al., 2022; X. Zhu et al., 2020). This dynamic exists also in reverse, the ability of a cell particularly in stressed states to restrain from proliferation and focus on metabolic production and adaptation to reactive oxygen species (ROS) determines survival (Endo et al., 2020; Pranzini et al., 2021; Varghese et al., 2020). Therefore, understanding the regulatory mechanisms underlying this equilibrium needs to be explored.

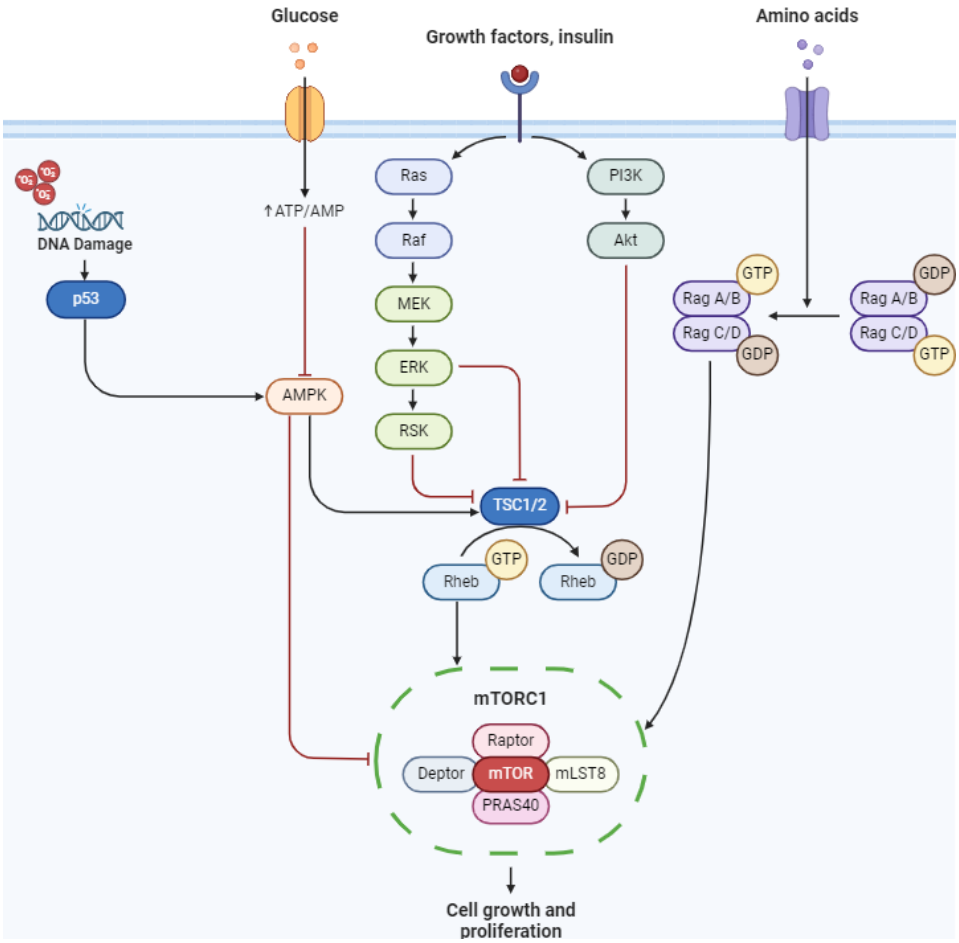


Figure 1.3: mTORC1 nutrient sensing cascade. Nutrient sensing pathways and their interconnection with mTORC1 resulting in inhibition or promotion of its growth and proliferation function. High levels of glucose results in an increase in ATP subsequently blocking AMP kinase pathway which is activated in nutrient dense environments that often results in ROS accumulation and DNA damage facilitating a p53 activation, AMPK phosphorylates mTOR directly inactivating it and indirectly through interaction with the heterodimer tuberous sclerosis complex (TSC1/2) resulting in its activation and hence inhibition of mTOR promoting Rheb. PI3k and RAS pathways and their downstream interactions block TSC1/2 thus preventing Rhen inactivation and enhancing mTOR activation. *Adapted from* (J. Huang & Manning, 2008; Jeon, 2016; Mendoza et al., 2011)*, Created in Biorender.*

1.2.6 The role of lipogenic metabolic reprogramming in cancer

Glycolysis is not the only pathway significantly reprogrammed to provide survival advantages in cancer. Pathways involved in regulating lipid biogenesis are notably altered in cancers, unsurprising given the essential role fatty acids play in cellular function and survival (Lee et al., 2022; Y. Ma et al., 2018). Not only do lipids provide a further energy source to cells through their catabolism to ATP, but they are the backbone of functional processes including signalling pathways and cellular membrane formation (Batchuluun et al., 2022; H. R. Jin et al., 2023; Lee et al., 2022; Y. Ma et al., 2018). The increased uptake and de novo production of fatty acids enables both the proliferative phenotype in cancers but also allows

for retention of homeostasis in stressed environments (Koundouros & Poulogiannis, 2020). The lipogenic role even extends to the TME through interaction with immune and stromal cells and can even result in immunosuppression thus supporting tumour survival (H. R. Jin et al., 2023). A significant upregulation in the expression of the transporters responsible for the maintenance of exchange has been well characterised in cancers and high expression of CD36, an essential fatty acid transporter, is a notable marker of poor prognosis in many cancers. Furthermore, studies have also shown deleting CD36 was successful in slowing tumour growth and resistance, highlighting the sheer dependency of lipids as a resource in some cancers (Beloribi-Djefaflia et al., 2016; Koundouros & Poulogiannis, 2020). Lipogenesis has since been identified as a universal state found within cancers, highlighting its critical role in sustaining tumour function, influencing tumour phenotype and even supporting cancer cell survival (Hausser et al., 2019).

1.2.7 Role of mTORC1 in metabolic reprogramming

Growth signalling pathways are key mediators in determining cell state and fate and are often found hyperactivated in cancer (Sever & Brugge, 2015). These signals converge on the key nutrient sensing pathway governed by the two mTOR complexes, mammalian target of rapamycin complex 1 and 2 (mTORC1 and mTORC2) **Figure 1.3** (Zou et al., 2020). The pathway has extensive downstream interactions that are widely implicated in cellular response specifically metabolic reprogramming (Navarro et al., 2022; Nong et al., 2023; Schiliro & Firestein, 2021). Unsurprisingly, mTOR is central to maintaining a functional level of homeostasis in cancer cells predominantly through mediating macromolecule catabolism and anabolism (Laplante & Sabatini, 2013). Firstly, on a translational level through direct activation of translation associated machinery, including the phosphorylation of S6K1 central to regulating MYC translation (Csibi et al., 2014; Mori et al., 2014). This well characterised oncogene functions as a global transcriptional re modeler through its TF role and results in the upregulation of many genes predominantly associated with enhancing cancer growth (Dang, 2012; Das et al., 2023). mTOR has also shown to upregulate other TF associated with controlling metabolic shifting including the well-known glycolysis regulator HIF1α and the lipogenesis regulating TF SREBPs (Laplante & Sabatini, 2013). More recently mTOR has also been shown to interact with ATF4 a vital

TF in regulating the stress response commonly associated with amino acid depletion (Torrence et al., 2021). Unsurprisingly mTOR inhibition has become a key anti-cancer therapeutic target, and as highlighted in **1.1.3** a successful one in CRC and sensitising resistance (Wanigasooriya et al., 2022). However, further understanding of how mTOR fully facilitates cell dynamics is needed.

1.2.8 Multi-task evolution theory

The appreciation for the dynamic equilibrium that exists in cancer cells has been further strengthened by a growing theory in cancer called multitask evolution theory (MET). MET highlights how cellular plasticity is a combination of extrinsic pressures and intrinsic adaptations specifically on the transcriptional landscape (Hausser & Alon, 2020; Wölfl et al., 2022). MET states that certain biological pressures, result in specific adaptations within the tumour transcriptome and cause selective trade-off in cellular function and thus cause ITH (Hausser & Alon, 2020; Plutynski, 2021). The selection pressure and resultant performance of an essential function or task owes to the expression of certain genes that allow the cells to respond to the pressure **Figure1.4A** (Hausser et al., 2019; Hausser & Alon, 2020). Selective trade-offs define cellular diversity as certain populations are under different pressures resulting in distinct phenotypes, thus abiding to the premise that no cell can complete all functions essential for tumour survival (Hart et al., 2015; Hausser et al., 2019; Hausser & Alon, 2020).

Using Pareto task inference algorithm which resolves high dimensionality data, the bounds of the gene expression and hence cell function can be plotted in gene expression space (Hart et al., 2015; Plutynski, 2021). The resultant plot will exist as a polyhedral with vertices representing a single archetype with a distinct gene expression profile **Figure 1.4B** (Hart et al., 2015; Hausser & Alon, 2020). Determination of a cell's archetype, positioning in the polyhedral, and hence function is through distinct spatiotemporal factors highlighting the fluidity of transcriptional regulation as described in **Figure 1.4**. Archetypes represent an exciting new way to define cellular states in cancer appreciating the continuum to which cells transcriptional landscape exists and collectively contribute to tumour function.

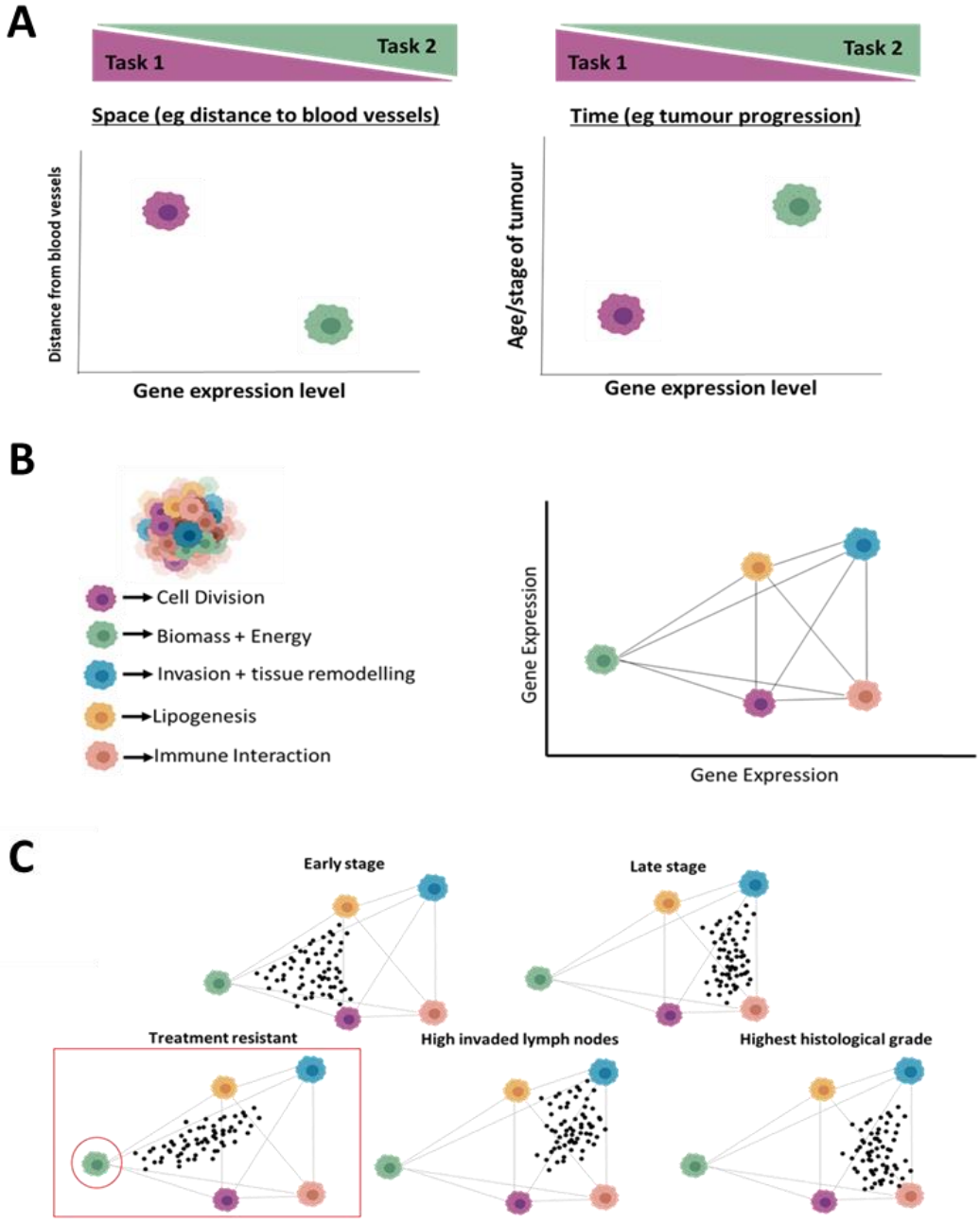


Figure 1.4. Selective constraints cause evolutionary trade-offs in task decision. (A) Trade off in gene expression occur depending on environmental factors that promote specific cellular function to enable survival. Spatially and temporally determined cells closer to blood vessels with more access to nutrients will be more proliferative compared to cells further away and thus in a nutrient scarce state will need to shift away from proliferation to metabolism remodelling to survive. (B) Polyhedral depiction of cancer cell task determination and associated gene expression derived from multi-task theory, which states under selective pressures and resultant tasks, optimal gene expression will result in a trade-off depending on task need and quantity of task. Abiding to the premise that no cell can complete all functions essential for tumour survival. Using Pareto task inference algorithm and a wealth of tumour sequencing data 5 core archetypes have been universally found in tumours and will sit on a polyhedral plotted in gene expression space. Whereby each vertices representing a single archetype with a distinct gene expression profile. (C) The overall known clinical phenotype of a tumour was consistent with the distribution of gene expression on the polyhedral. *Figure adapted from* (Hausser et al., 2019; Hausser and Alon, 2020; Nath et al., 2021).

1.2.9 Role of archetypes in cancer

Using a wealth of bulk and single cell RNA sequencing data from a broad selection of cancers, 5 core, universal cancer archetypes have been identified and highlighted in **Figure 1.4B** (Hausser et al., 2019; Hausser & Alon, 2020). Archetype analysis provides insight into not only the state of a single cell, but also the overall clinical phenotype of the tumour through the distribution of cells on the polyhedral **Figure 1.4C** (Hart et al., 2015; Hausser et al., 2019). For example, tumours known to have invaded many lymph nodes sat closer to the invasion and tissue remodelling archetype (Hausser et al. 2019). Other clinical features impacting distribution on the polyhedral included stage, where early-stage tumours were found closer to cell division, biomass and energy and the lipogenesis archetype (Groves et al., 2021; Hausser et al., 2019). Later stages were found in the regions of the immune interaction and invasion and tissue remodelling archetypes highlighting a more differentiated tumour (Hausser et al. 2019; Combes et al. 2021). The tumours that sat closer to the biomass and energy archetype profile and had considerable cancer metabolic reprogramming, a later defined sub archetype, were least responsive to treatments such as irradiation (Nath et al., 2021). MET and archetypes provide a novel approach to exploring the continuum of cellular states in cancer and the dynamics in which they exist in cancer growth, progression, and response.

Despite its relative novelty, MET and the approach of using archetypes to define universal states in cancer is becoming more popular. Application of MET to small cell lung cancer enabled the identification of subtypes in what was previously believed to be relatively homogenous tumours (Groves et al., 2022). Previous approaches had relied on determining subtypes using expression of specific TF however its utilisation was not successful across model systems, the use of archetypes however was (Groves et al., 2022). A recent paper by Combes et al, also extended the use of archetypes to define the immune state of tumours through analysis of their TME and resultant expression changes (Combes et al., 2023). They identified dominant immune archetypes that not only provide predictive capabilities, but outlined potential deficiencies that could be exploited in immunotherapies (Combes et al., 2022, 2023). These immune archetypes have since also been used by other groups to achieve similar profiling to identify optimal treatment paths (Anderson et al., 2023). Furthermore, a recent longitudinal study tracked the changes in archetype presentation in ovarian cancer

over treatment courses in patients (Nath et al., 2021). Tumour heterogeneity and hence archetype dominance changed over treatments, they found consistent metabolic and proliferative states that were highly present in resistant patients (Nath et al., 2021). This paper and the concept of MET highlights the significance of the transcriptional landscape and its plasticity in determining cellular state and a need to further our understanding of the transcriptional regulators involved.

1.3 Gene regulatory networks

1.3.1 Transcriptional regulation in Cancer

As described by archetypes, expression changes can give essential insight into the molecular functions underpinning a specific condition (Rodriguez-Esteban & Jiang, 2017). Therefore, the study of gene regulation pathways in cancer cellular state dynamics will contribute to both efficient marking as well as offer opportunity to identify targets of interference (Assi et al., 2019). Transcription regulation is an extremely complicated process involving many stages and vast molecular recruitment (Couvillion et al., 2022; Sperling, 2007). The basis of the regulation mechanism includes recruitment and construction of the transcription machinery, followed by initiation, elongation, release and termination, and finally RNA modification events (Casamassimi & Ciccodicola, 2019; Villard, 2004). Many regulatory elements interconnect and uphold this mechanism firstly chromatin accessibility, driven by remodelling proteins, histone modifications and other epigenetic modifiers (Casamassimi & Ciccodicola, 2019). Furthermore, TF are highly involved and widely studied due to their enhancer influence through recruitment of transcription initiation machinery to promoters (Bushweller, 2019; Lambert et al., 2018). Not only are mutations in TF commonly associated with cancer but their levels and usage have shown correlations to maintenance of stemness in CSC (Islam et al., 2021; Lambert et al., 2018).

1.3.2 Cis-regulatory elements and gene expression

The key premise of transcriptional regulation analysis is to understand how transcription factors and other components target the cis-regulatory elements (CRE) of DNA surrounding genes and how such interactions mediate signals received by the nucleus of the cell (K. P. Singh et al., 2018; R. Zhao et al., 2009). CRE are defined into two categories, the

promoter, the sequence in which transcription initiation starts and enhancers, proximal and distal TF binding sites (TFBS) (Andersson & Sandelin, 2020; Panigrahi & O'Malley, 2021; Schoenfelder & Fraser, 2019). The latter has been the predominate research focus in understanding differential gene expression dynamics in disease states (Z. Sun et al., 2023). The acquisition of super enhancers at oncogenes is believed to be a vital driving force of uncontrolled proliferation in cancer and cellular phenotype determination and have shown therapeutic target relevance (Z. Sun et al., 2023). The promoter, however, is often overlooked despite being the recipient of signals emanating from the complex TF and CRE interactions that consequently initiate and drive RNA polymerase recruitment (Casamassimi & Ciccodicola, 2019). The promoter not only interacts with transcription initiation machinery, including the large multiprotein TFIID composed of TBP and TAFs essential for core promoter recognition but also directly and indirectly with TF, epigenetic modifiers and many other regulatory signals (Andersson & Sandelin, 2020; X. Chen & Xu, 2022; Patel et al., 2020). Thus, highlighting the promoter as the signal integration hub and the complexity of its vital involvement in defining the transcriptional regulatory response.

1.3.3 Cap dependent translation

Differential regulation post transcription also exists and adds a further layer of complexity in understanding the significance of gene expression (Corbett, 2018). Alternative mechanisms of translation highlight a potential differential post transcriptional regulatory mechanism that utilises transcript diversity and may facilitate differential gene expression. The canonical pathway of protein synthesis is through cap dependent translation and the initiation process is composed of the formation and hence binding of the translation initiation machinery (TIM) (Hinnebusch & Lorsch, 2012). The addition of the 7-methylguanosine cap (m7G) to 5' of mRNA transcripts not only aids in preserving transcript stability but is also the target for the key translation initiation machinery binding also termed eIF4F complex (**Figure 1.5**) (Cowling, 2010; Furuichi, 2015). The eIF4F containing complex consists of EIF-4e proteins responsible for m7G cap binding, eIF4G a core scaffolding protein aiding in recruitment of the remainder of the preinitiation complex (PIC) eIF4A and B (Blagosklonny, 2013; Borden & Volpon, 2020; Marques-Ramos et al., 2017; Merrick, 2004). Cap-dependent translation is mainly regulated by the mTOR pathway. When mTOR is active,

mTORC1 phosphorylates the translational repressor protein 4E-BP1, thus preventing it from binding to EIF-4e protein (Böhm et al., 2021; X. Qin et al., 2016). EIF-4e binding is the key rate limiting step of translation ultimately resulting in the recruitment of 40S ribosomal subunit to the 5'-terminal cap of mRNA and driving initiation (Nandagopal & Roux, 2015). mTOR also activates ribosomal protein S6 kinase, that regulates many cellular processes including translation through the activation of eukaryotic translation initiation factor 4B (eIF4B) also important for formation of the PIC (M. Yang et al., 2022).

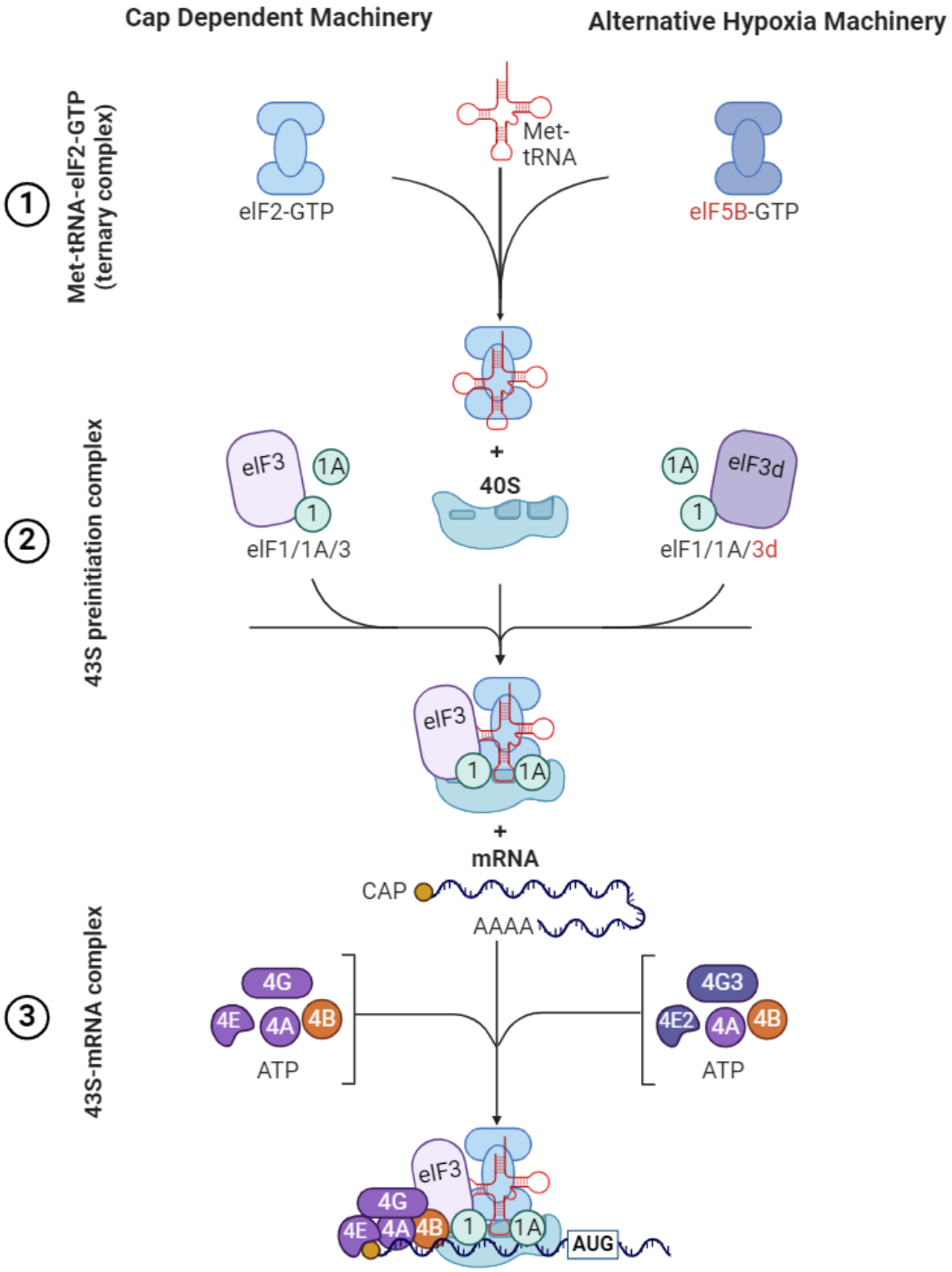


Figure 1.5: Alternative routes of translation machinery formation. Formation of cap-dependent translation machinery involves a core 3 step process centred around recruiting multiple elongation factors that form the preinitiation complex that subsequently binds to mRNA with the facilitation of the key Cap binding EIF4F highlighted in step 3. Many of the key factors are regulated by the mTOR pathways and when in a stressed state a core element of each step is inactivated predominantly through phosphorylation thus preventing the formation of the machinery. The factors which are shown to be affected by this are replaced in stressed states by alternative elongation factors and form the alternative translation machinery also terms EIF4H. The right panel shows the relative composition of the ATM and the altered components replaced during mTOR inactivation. *Adapted from* (Ho et al., 2021; Vaysse et al., 2015) *and created in Biorender.*

1.3.4 Alternative translation in cellular stress

Translation is one of the most energy taxing cellular processes, therefore the mTOR pathway is highly receptive to nutrient sensing and thus plays a vital role in mediating cellular stress response (Roux & Topisirovic, 2012). In stressed states mTORC1 is inactivated and unable to phosphorylate several downstream targets resulting in translation repression often aiding in cell survival (Roux & Topisirovic, 2012; X. Wang & Proud, 2006). However, translation of certain transcripts is also vital for survival and an intricate alternative translation machinery (ATM) network exists to bypass the use of mTOR dependent translation (Uniacke et al., 2012; Vaysse et al., 2015). The ATM is mainly driven by metabolic stress demands originating from hypoxia environments and hence this machinery is often termed oxygen-sensing and results in complex translatome remodelling to maintain translation efficiency (de la Parra et al., 2018; Ho et al., 2016, 2021; Uniacke et al., 2012). The key proteins involved and regulated by mTOR, eIF4E1 and eIF4G1, are replaced by eIF4E2 and eIF4G3 respectively, resulting in the newly defined hypoxia cap-binding complex or eIF4FH (Ho et al., 2021; Ho & Lee, 2016; Uniacke et al., 2012). This complex is further aided by a basally dormant protein eIF5B which is responsible for transporting met-tRNA onto ribosomes **Figure1.5** (Ho et al., 2018; Ho & Lee, 2016). The transcripts targeted via this mechanism are those associated with genes known to be vital in metabolic shifting to the more energy conservative anaerobic glycolysis (Ho et al., 2020). Furthermore, it was found the remodelling even extends to ribosomal binding proteins (RBP) with a distinct subset activating in these hypoxic environments (Ho et al., 2020). This adaptive process represents a complex survival mechanism potentially through exploiting transcript diversity. Such mechanism maybe utilised in cancers to enable resistance is stressed states.

1.4 Promoter level transcriptional regulation

1.4.1 Promoter classification

A potential source of transcript diversity is believed to originate from transcriptional levels of regulation such as the promoter (Strausberg & Levy, 2007). To better understand the role of promoters the last two decades has seen their classification underpinned by distinct

regulatory interactions leading to defined biological roles (Müller and Tora, 2014a; Nepal et al., 2013). Characterisation of classes is dependent on a diverse sequence of events encompassing the following components: composition of the initiation machinery, organisation of enhancer interaction, chromatin accessibility and epigenetics features, nucleosomal ordering at transcription start sites (TSS) and positioning of general TF protein binding sites (Müller & Tora, 2014; Nepal et al., 2013). On the single gene level further complexity emerges from the use of multiple, alternative promoters (AP) by the same gene and by differential reading of multiple DNA sequence codes within the same promoter (Haberle et al., 2014). AP usage can further diversify transcriptional output through differing transcript and protein production (Huin et al., 2017; Singer et al., 2008). Promoter architecture is therefore much more complex than previously thought and understanding the subsequent interactions with these diverse sets of protein complexes represents a potentially unexplored level of gene regulation.

1.4.2 Alternative promoters and disease

Research into the role of promoters in cancer predominantly focused on epigenetic regulation such as DNA methylation resulting in gene silencing and the role of enhancers in initiation complex recruitment (Belinsky, 2005; Gonzalgo & Jones, 1997; Herman & Baylin, 2003; Kazanets et al., 2016). Although insightful, many of these analyses miss crucial gene regulatory roles of the promoter, beyond a passive response to silencing. Following the identification of tissue specific AP usage, the role of AP in cancer and other disease states were investigated (Demircioğlu et al., 2019). Isoform and thus transcriptional diversity have not only been found originating from the pre transcriptional regulation via AP usage, but such usage shows interesting correlations with patient survival (Demircioğlu et al., 2019; Valcárcel et al., 2021). Promoter usage may therefore play a role in disease phenotypes such as resistance thus, opening a new realm of potential biomarkers and even therapeutic targets. Further to this, epigenetic and other regulatory landscapes surrounding AP appear to be vastly different, including methylation sites (Nepal & Andersen, 2023). The heterogeneity in epigenetic and other regulatory landscapes that surround AP maybe missed if the promoter is not explored.

1.4.3 Transcription start sites and promoter phenotype

To further investigate promoter structure, usage and promoter-associated transcript diversity, a high throughput approach called cap analysis of gene expression, or CAGE was developed (Kodzius et al., 2006; Shiraki et al., 2003). CAGE uses a cap trapper and biotinylated linker on the 5'end of cDNA strands synthesised from total RNA to produce CAGE tags (Takahashi, Kato, et al., 2012; Takahashi, Lassmann, et al., 2012). CAGE tags not only measure RNA expression, but also maps them to their associated TSS down to a single-nucleotide resolution, thus showing overall RNA expression but also expression levels associated to specific TSS (Nepal et al., 2020; Shiraki et al., 2003). CAGE analysis broadens our understanding of distinct promoter characteristics and allows for better classifications that have interesting biological and functional relevance (Danks et al., 2018; H. Li et al., 2015). For example, TSS distribution informs promoter architecture: genes can have a single dominant peak, these sharp promoters are associated with tissue specific structural genes and ribosome biogenesis genes (H. Li et al., 2015). The absence of a single dominant peak or broad TSS distribution are found in housekeeping genes. Genes can also harbour multiple dominant peaks, this multimodal classification was attributed to developmental genes, highlighting a potential role of utilising alternative TSS in different tissues and regulatory contexts (H. Li et al., 2015). This highlights how even TSS distribution and usage have a potential interesting role in transcriptional regulation.

1.4.4 Alternative promoter architecture in cell cycle dynamics

Utilisation of CAGE analysis has highlighted the role of AP usage and promoter architecture differences in cellular states specifically through its link to cell cycle transitions into fate specific states during development (Wragg et al., 2020). Stem cells are distinguishable via their fast proliferative state and hence short cell cycle characterised by a shorter G1 phase driven by distinct regulations of vital cycle genes (Shyh-Chang et al., 2013; Zaveri & Dhawan, 2018). In developing zebrafish embryos, cells are rapidly cycling before slowing when transitioning away from pluripotency to defined cell fates (Wragg et al., 2020; Zaveri & Dhawan, 2018). Both fast and slow cycling cells showed utilisation of TATA-like motifs in their upregulated genes, however the fast cycling also utilised CCAAT-boxes and W-box motifs distinguished by a broad promoter usage landscape (Wragg et al., 2020). Conversely in slow

cycling cells, TSS usage narrowed to a distinct region of the promoter creating a sharper peak specifically on tissue specific genes (Wragg et al., 2020). These changes were specifically resulting from shifts in cell cycle dynamics rather than tissue or lineage specific associations (Wragg et al., 2020). This work highlighted an interesting link between the role of alternative promoter usage in the regulation of cell cycle dynamic transitioning during differentiation stages in development. Further analysis is needed to identify a potential further pervasive role of promoters in cell cycle dynamics in both development and disease.

1.4.5 Advancements in CAGE methodology

With the increase interest in promoter level sources of transcriptional regulation and the insight CAGE analysis provided, a push has been seen for advancements in the technology. As described above conventional CAGE attains the 5' end of transcripts using cap trapping and attachment of a biotinylated linker which contains an endonuclease site allowing for production of CAGE tags (Carninci et al., 1996; Takahashi, Kato, et al., 2012). This technology has aided many significant research efforts including large scale atlas production like the FANTOM5 consortium (Cvetesic et al., 2018; Forrest et al., 2014). Furthermore, to counteract the high yield of RNA required and potential PCR amplification and enzymatic digestion biases that exist in the approach, nAnT-iCAGE was developed (Murata et al., 2014). This approach has been even further advanced to accommodate extremely low input through the utilisation of selectively degradable carrier RNA, thus allowing for even nanogram volumes of input called Super-Low Input Carrier-CAGE (SLIC-CAGE) (Cvetesic et al., 2018). Alterations in the conventional methodology also allowed for accommodation of low input, nanoCAGE uses template switching oligonucleotide (TSO) technology which utilises the reverse transcriptase (RT) innate methodology of incorporating additional deoxycytosines to the cDNA strand occurring in a cap dependent manner (Poulain et al., 2017; Salimullah et al., 2011). The TSO contains relevant complementary riboguanosine repeats meaning the RT will switch from the newly synthesised cDNA to the TSO (Poulain et al., 2017; Salimullah et al., 2011). This process outlined in **Figure 1.6** allows for specific alterations including PCR primer sites and relevant unique molecular identifier (UMI) into the oligonucleotides that allow for amplification (Poulain et al., 2017).

Through combining the TSO methodology of nanocage with that of 5' single cell RNA sequencing established by 10x genomics single cell CAGE has since been made possible

(Kouno et al., 2019). The barcoding attributable to the single cell RNA sequencing approach allows for associating transcript to individual cellular identity and the subsequent nanoCAGE TSO approach results in capturing TSS (Kouno et al., 2019; Moody et al., 2022). Therefore, this means we can identify cellular specific gene expression and their relative TSS usage. To accommodate the G bias outlined in **Figure 1.6** that hinders TSO methodology, a newly devised pipeline has been developed named Single-Cell Analysis of Five-prime Ends (SCAFE). The TSO linker contains complementary riboguanosine repeats that facilitate the switch to 5' linker after the RT reaches the cap, these riboguanosine repeats can also be found downstream of the cap meaning incorrect early integration of the 5' can occur. This G bias can therefore produce artefacts (Tang et al., 2013). The SCAFE software enables the identification and removal of G associated mismatches and hence TSS artefacts that occur as a result of the TSO G-bias (Moody et al., 2022). The use of single cell CAGE is so far limited to this paper and their sole scrutinization of CRE, it is yet to be explored its applicability to other cellular inputs and TSS capture and alternative usage.

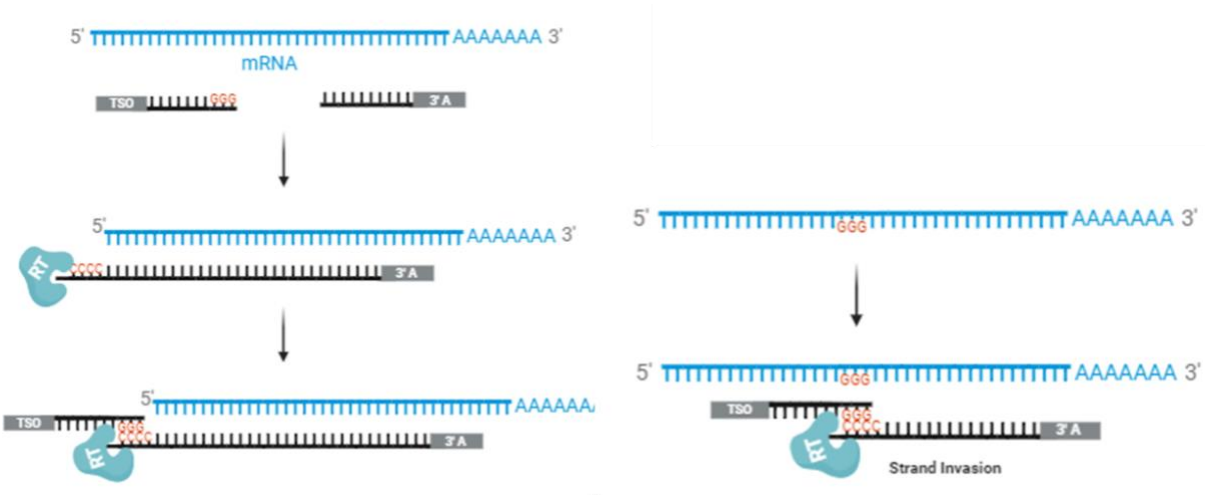


Figure 1.6: Characterisation of mRNA 5' ends using template switching oligonucleotides. Capture of 5' ends can utilise the innate function of reverse transcriptase and the addition of relevant linkers. Notably the 5' linker which includes the complementary riboguanosine repeats meaning following reaching the CAP the RT will switch from the mRNA to the TSO. Strand invasion can occur and is highlighted in the right panel, when complementary structures are found downstream of the 5'end and result in early integration of the TSO and thus creates and artefact transcript that will map and highlight and incorrect TSS. Adapted from (Tang et al., 2013) and created in Biorendor.

1.4.6 Canonical and non-canonical start sites

CAGE analysis also allows for promoter classification through TSS usage and represents an opportunity for exploration into TSS usage in promoter dynamics. The majority (~80%) of

TSS initiations occur at pyrimidine (+1), purine (-1 / TSS) dinucleotides (CA, TA, CG, TG) and these are classified as canonical or YR, **Figure 1.7** (Carninci et al., 2006; Kodzius et al., 2006). Non-canonical or YC initiate with a pyrimidine (+1), cytosine (-1 / TSS) dinucleotide (Haberle & Stark, 2018). In *Drosophila*, the YC motif, or TCT as it was previously defined, showed selective transcription via TATA box binding protein related factor (TRF2), highlighting differential regulation of TSS (Y. L. Wang et al., 2014). An interesting and notable subset of YC, characterised by a +5 polypyrimidine stretch following the cytosine initiation are known as terminal oligopyrimidine (TOP) and were initially predominantly associated with genes coding translation associated machinery such as ribosomal genes (Nepal et al., 2013; Parry et al., 2010). However recent investigation that began in zebrafish, highlighted that YC initiation is far more widespread and even constituting a minor, but notable component of transcription initiation in thousands of genes with an evolutionary conserved dual initiator (Nepal et al., 2020). Thus, meaning initiation of transcription in these genes harbouring dual initiation capabilities, or classified as dual genes, could start at either its YR or YC TSS. Further scrutinization of dual genes has shown the diversity of the intermingling of these initiators, including the distance from each other and even the preferential usage (Nepal et al., 2020; Wragg et al., 2023). Therefore, alternative TSS usage represents a previously unexplained and unexplored contributor to promoter level transcription regulation unrestricted to just AP usage.

1.4.7 Unclassified promoters

Current research has focused on these two promoter classifications predominantly resulting from a lack of clear understanding of the role of the other potential TSS sequences not defined as YC or YR. Furthermore, dinucleotide frequency analysis achieved through CAGE shows the alternative initiator sequences appear in a significantly low proportion especially in comparison to the YR and YC (Nepal et al., 2020; Wragg et al., 2023). Despite their sparsity, initiation occurring from the purine A has shown some functional relevance on a translational regulatory level. Meyer et al, showed that transcripts that begin with an A at their 5'UTR are targeted by methyltransferases (Meyer et al., 2015). The resultant methyl modification is believed to enable these transcripts to bypass cap-dependent translation and is able to be targeted by alternative translation machinery including RBP (Meyer et al., 2015). Meyer et al also showed the levels of these transcripts can alter firstly through depletion of

relevant methyltransferases and even in stress environments (Meyer et al., 2015). These findings highlight that other promoter classifications are likely needed, and future work should consider the relevance of the currently unclassified.

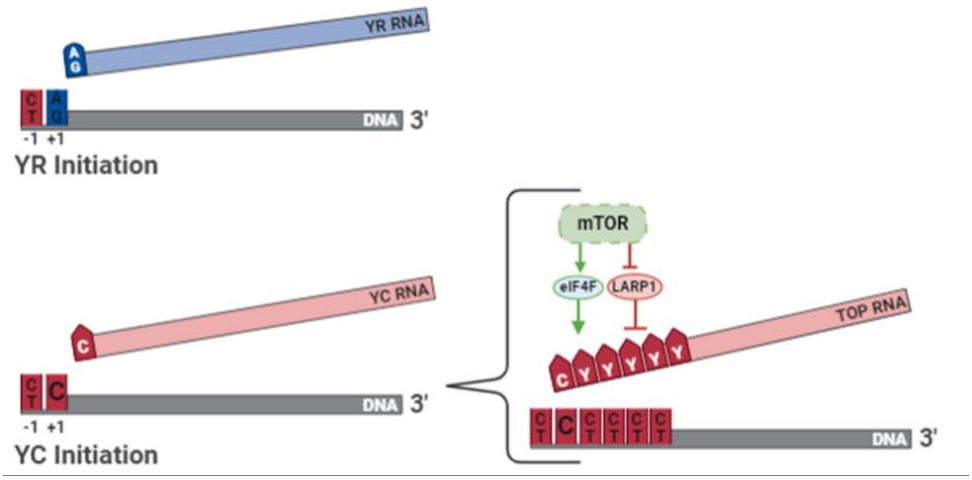


Figure 1.7: Transcription initiation classifications. Characterisation of promoter classes using TSS initiator sequence. YR transcripts are produced from TSS that initiate from a Purine with a preceding pyrimidine. YC transcripts are produced from TSS initiating from a cytosine with a preceding pyrimidine. A subset of YC also known as TOP also initiate from a cytosine but contain a 5 bp sequence of polypyrimidines positively targeted by translation initiation factors. LARP1 also targets this polypyrimidine stretch through its DM15 domain and is believed to inhibit the binding of the TIM through its higher affinity binding. Both mechanisms are regulated by the mTOR pathway. *Figure adapted from* (Wragg et al., 2023) *and created in Biorender.*

1.4.8 TOP initiation in cancer

The functional relevance of TOP initiation has also been investigated in disease, in particular cancer. As TOP transcripts often code translation machinery elements including both RBP and even translation factors, it is highly believed to play a role in driving translation levels (Hochstoeger & Chao, 2024; Weber et al., 2023). This mechanism is hypothesised to be exploited in cancers to maintain the protein production demand for uncontrolled growth (Weber et al., 2023). Furthermore, many TOP genes are believed to be host genes for small nucleolar RNAs (snoRNA), meaning their introns are excised to produce both coding and noncoding snoRNAs with distinct functions in ribosomal RNA control (De Turris et al., 2004; Mourksi et al., 2020; M. S. Scott & Ono, 2011). snoRNAs have been highly implicated in cancers and evidence also suggests oncogenic potential through its interconnection with p53 (Liang et al., 2019; Su et al., 2014; Williams & Farzaneh, 2012). Many snoRNA host genes harbour dual initiation, here it was found that the canonical initiation population was driving

protein coding transcription from the promoter, whilst the TOP component encoded the snoRNA, with regulation of each independent of one another (Nepal et al., 2020). This finding suggests that alternative transcription initiation complexes are utilised to drive canonical and TOP initiation from these promoters (Nepal et al., 2020; Yamashita et al., 2008).

1.4.9 Regulation of TOP transcripts

Further understanding of the functional relevance of differential TSS usage comes from the evidence that resultant transcripts are differentially regulated. Upon identifying that levels of TOP mRNA were sensitive to inhibitors of mTOR, it was hypothesised the pathway played a role in TOP translation through cap-dependent translation (Hsieh et al., 2012; Nepal et al., 2020; Thoreen et al., 2012; M. Yang et al., 2022). The polypyrimidine stretch that defines TOP transcripts is believed to be the target and potential binding site of translation initiation machinery that is specifically regulated by mTOR (Cockman et al., 2020; Hochstoeger et al., 2024). In activated states mTOR phosphorylates 4E-BP1, preventing it from binding to the EIF4F TIM components allowing their formation (X. Qin et al., 2016; Thoreen et al., 2012). Although this form of cap dependent translation targets most mRNAs research has shown an increased translation efficiency of TOP transcripts (Hochstoeger et al., 2024; Meyuhas & Kahan, 2015). TOP transcripts often code ribosome biogenesis proteins and hence aid in further driving translation, this represents a potential feedback loop that cells utilise to aid in satisfying protein production demands (Cockman et al., 2020). As mTOR is often highly upregulated in most cancers, the increase in translation required to maintain and drive proliferative states underpinning tumour growth, may utilise TOP transcript production and this cap dependent translation mechanism (Bouyahya et al., 2022; T. Tian et al., 2019). Studies have shown an altered affinity of TIM on TOP transcripts that may result in their heightened sensitivity to mTOR inhibition (Hochstoeger et al., 2024; Tamarkin-Ben-Harush et al., 2017). Therefore, this may highlight a biological understanding of the success attributed to blocking mTOR to sensitise cancers to treatment.

1.4.10 Role of LARP1 and TOP translation regulation

Interestingly deletion of La-related protein 1 (LARP1) using CRISPR CAS9 showed mTOR inhibition no longer affected TOP mRNA levels, suggesting its predominant role in the

regulation (Jia et al., 2021). LARP1 belongs to the family of LARP RBPs and like many other RBPs, has shown significant involvement in cancer aetiology and research has even identified it as oncogenic (Schwenzer et al., 2021). Unlike many of the other LARP proteins LARP1 is believed to be able to bind mRNA in two places indicating a potential dual role (Schwenzer et al., 2021; Stavraka & Blagden, 2015). Firstly, the La domain composed of an accompanying RNA recognition motif but also the presence of a DM15 domain in its C terminus (Mura et al., 2015; Stavraka & Blagden, 2015). Despite the presence of extensive LARP1 isoforms both the aforementioned domains appear in most highlighting their significance in function (Schwenzer et al., 2021; Stavraka & Blagden, 2015). mTORC1 has been demonstrated to phosphorylate and hence inhibit LARP1. When in its active unphosphorylated state LARP1, through its DM15 domain, simultaneously binds to the polypyrimidine stretch of TOP transcripts and the 5'cap preventing the binding and hence formation of the PIC (Mura et al., 2015; Ogami et al., 2022). In addition to repressing TOP transcript translation, LARP1 is also believed to be involved in regulating their stabilisation and decay through its binding to poly(A)-binding protein C1 (PABPC1) (Mura et al., 2015; Ogami et al., 2022). This binding facilitates preservation and even extension of poly(A) tail length through polyadenylation and hence aiding in prevention of transcript degradation (Ogami et al., 2022; Slomovic et al., 2006). A key target for LARP1 appears to be mTOR transcripts, resulting in their stabilisation and a complex feedback network regulating post transcriptional events (Hopkins et al., 2016; Mura et al., 2015).

LARP1 has shown to anchor TOP transcripts to both stress granules (SG) and processing bodies (PB) (Farooq et al., 2022; Wilbertz et al., 2019). The resultant fate of these transcripts has been shown to differ, including re-entry into the translation cycle following diminishing of stressors, potentially representing a store to allow the driving of translation when energy levels permit (Hsieh et al., 2012; Philippe et al., 2020; Schneider et al., 2022; Thoreen et al., 2012). However, recent work in starvation whereby chronic mTOR inactivation is present and hence global translation is lower, baseline levels of TOP transcripts are still translated through shuttling of transcripts to monosomes (Schneider et al., 2022). Therefore, our true understanding of the dynamics and the regulation of these transcripts in different states remains fully understood.

1.4.11 Alternative initiation and cancer

Despite research highlighting differential regulation and hence potential functional significance of alternative TSS usage, the understanding of TOP usage in disease was predominantly limited to its role in ribosomal and translation machinery production (Alfonso-Gonzalez & Hilgers, 2024). However, work in the Mueller lab wanted to further resolve alternative TSS usage to explore the previously unappreciated YC and TOP component in core promoters of dual genes, and its potential implication in cancers beyond just ribosomal genes and AP usage (Wragg et al., 2023). Utilising CAGE data obtained through FANTOM5 datasets it was established that not only is TOP enriched in cancer when compared to healthy tissues, the enrichment was seen in the dual genes compared to its YR counterpart **Figure 1.8** (Wragg et al., 2023). This means that in cancer TSS usage shifts away from its canonical YR start site, resulting in the production of TOP transcripts with potentially different post transcriptional regulation. Interestingly, such findings would be masked using standard RNA sequencing approaches as overall transcript levels remain relatively unchanged but their TSS usage and hence transcript production differ (Wragg et al., 2023). Further to this it was also established that this usage dynamic was not just limited to TOP but was in fact seen in other YC start sites not harbouring the conventional TOP motif sequence (Wragg et al., 2023). Not only does this indicate a potentially similar regulation and functional usage to TOP but also expands the number of genes of interest undergoing TSS switching in disease states (Wragg et al., 2023). The dynamic shifting between TSS in these functional contexts was almost exclusively occurring between the YC and YR classifications and such phenomena was not significant in the other initiator sequences. Therefore, the resultant dual gene classification was predominantly assigned to genes containing both YC and YR.

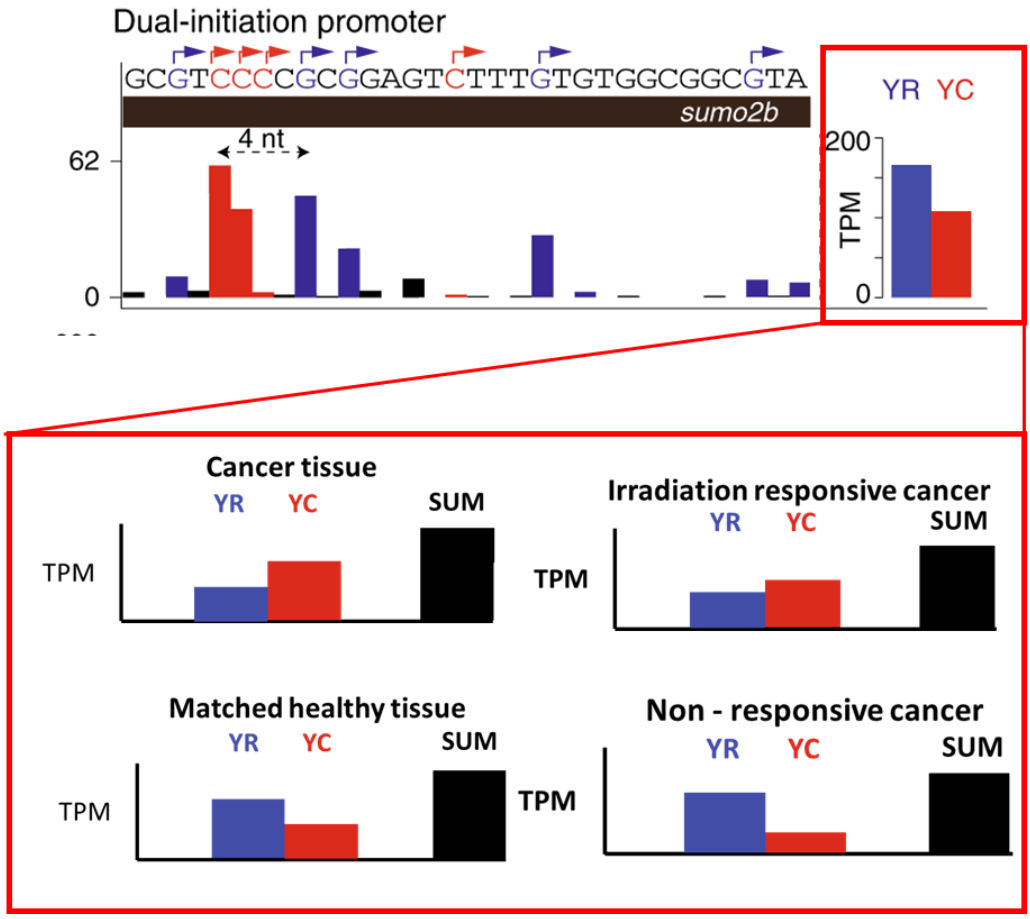


Figure 1.8: YC usage Is enriched in Cancer relative to irradiation responsiveness. Representative view of individual dual initiator gene *sumo2b* genome browser track highlighting YC (in red) and YR (in blue) initiation start site usage and relative levels in form of CAGE tags and the relative sum of Tags per million per initiator class. Relative YC:YR usage in these genes provides a biological readout that has since been associated with features. Bottom panel highlights schematic representing such associations. Left shows YC usage in dual initiator genes is enriched in comparison to matched healthy tissue. Right shows the relative YC enrichment correlates with irradiation responsiveness. In both cases the overall SUM remains relatively similar meaning would not highlight differential expression and the closeness of initiators means these findings would be masked in conventional RNA sequencing. Adapted from (Wragg et al., 2023).

1.4.12 YC initiation marks irradiation responsiveness

To further explore the TSS shifting dynamic in cancers, CAGE was performed on 5 CRC PDO with different clinical phenotypes. It was found that YC enrichment, including both TOP and non-TOP YC TSS, correlated to irradiation responsiveness (Wragg et al., 2023). PDOs that responded well to irradiation treatment had a higher YC enrichment in dual genes whereas PDOs which did not respond well to irradiation treatment had a preferential usage of the YR TSS (Wragg et al., 2023). The overall expression of many of these dual genes showing this dynamic switching remained relatively unchanged between the different lines, hence

expanding our understanding of genes previously unexplored in tumour evolution. Furthermore, analysis of promoter behaviour upon irradiation of CRC organoids revealed a prominent depletion of TOP initiation both globally and specifically in dual initiators, relative to YR initiation, with the level of depletion also correlating with therapeutic response (Wragg et al., 2023). This analysis not only highlighted potential novel therapeutic targets to explore through the identification of newly relevant functional gene sets but also aids in stratifying tumours to establish appropriate treatment choices through TSS enrichment. The inter heterogeneity of TSS usage indicates a potential functional relevance of TSS and hence function of core promoter dynamics in cancer aetiology and may represent a fundamental transcriptional regulatory element involved in determining tumour behaviour.

The advent of 5' sequencing approaches such as CAGE has broadened our appreciation for the sheer complexity of promoter architecture and provided newfound intrigue into the fundamental role the promoter plays in cell behaviour. Work in the Mueller lab has highlighted an appreciation for a previously unexplored promoter level regulation in the form of TSS and their dynamic usage in development and disease. Although we have established an interesting shift in usage that represents an exciting biomarker for treatment response, the functional relevance of this core promoter TSS switching is still not fully understood. Furthermore, as previously highlighted individual tumours are comprised of a complex network of phenotypically diverse subcellular populations, and it is not understood the pervasive existence of TSS usage differences in ITH and therefore may represent a previously unexplored level of transcriptional regulation in cellular state determination.

1.5 Aim of this study

In this thesis we aim to deepen our understanding of TSS usage and its role in cellular biology, particularly in cancer heterogeneity, with the future goal of identifying potential therapeutic vulnerabilities. By investigating TSS dynamics, we seek to uncover their functional relevance in cancer, specifically focusing on how TSS usage contributes to ITH and influences treatment responsiveness and other cancer phenotypes. While current literature has begun to elucidate the role of TSS usage in transcriptional regulation, highlighting differential downstream regulatory mechanisms and dynamic shifts across cancer classifications, our knowledge remains limited. Most studies have linked TSS usage to treatment response, but without comprehensive insight into the link to the molecular drivers of these phenotypes. We hypothesise that YC usage may provide a growth advantage and hence aid in driving highly proliferative cancers. Thus, to test this hypothesis and fill the gaps in our understanding this thesis aims to explore the biological significance of TSS usage through transcriptomic and imaging analyses.

This will be achieved by the direct aims:

Firstly, to further phenotypically characterise organoids with global TSS classifications to understand the molecular features including pathway enrichments of high YC cancers.

Secondly to explore the penetrance of alternative TSS usage by successfully carrying out single cell CAGE sequencing. This data will be used to identify cellular states linked to TSS usage, allowing us to better define the genes and molecular features affected, as well as the significance of the initiator motifs on dynamics.

Finally, to develop methodology to visualise cellular dynamics, including cycle phasing, to aid in linking genomic insights with spatiotemporal patterning of organoid architecture.

Chapter 2: Materials and Methods

2.1 Molecular Biology

Table 2.1. Molecular biology material compositions

Name	Materials and methods
LB broth for Agar	10g of NaCl and Tryptone, 5g of yeast extract and NaOH were dissolved in 900ml of distilled H_20, pH adjusted and made up to 1litre. This was split and 5g of Agar was added to 500mL, autoclaved and Ampicillin (Sigma-Aldrich, Dorset, UK) was added and poured into plates.
E3 Buffer	5 mM NaCl, 0.17 mM KCl, 0.33 mM CaCl2, 0.33 mM MgSO4 dissolved in H_20.
DNA extraction buffer	10 mM Tris pH 8, 2 mM EDTA. 0.2% Triton X-100, 200 µg/ml Proteinase K
HEK Media	DMEM 12 (ThermoFisher Scientific) 5ml of non-essential amino acids, 5ml penicillin streptomyces, 5ml L-Glutamine, 55ml 10% FBS.
Transfection master mix 1	For T75 flask: 5.1µg of Lentivirus envelope vector PMD2.G, 7.8µg of Lentiviral packaging plasmid PSPAX and 45µl of P300 enhancer added to 1875µl of Optimem.
Transfection master mix 2	For T75 flask: 1875µl Optimem + 52.5µl LIPO300
Organoid media	Stem cell - 50ml IntestiCult OGM Human Basal medium was added to 50ml of Stem cell organoid supplement. 100ul of Prim antibiotic was added.
Low melting point agarose	5% agarose solution was created by dissolving 5g in 100ml of PBS and stored at RT. Each consecutive use was done by reheating the solution to 100c and aliquoted out.

2.1.1 Bacterial Transformation

Glycerol stocks of transformed of FUCCI plasmid (PLL3.7m-clover-geminin (1-110)-IRES-mKO2 -Cdt(30-120)) (Addgene 83841), Plasmid map outlined in **Appendix A** and Puromycin donor plasmid (pLJM1-EGFP) were produced by transformation of plasmids into Competent E.coli cells (New England Biolabs) and grown on pre-warmed LB-agar plates (**Table 2.1**) at 37°C overnight. FAST-FUCCI pBOB-EF1-FastFUCCI-Puro (Addgene 86849) (**Appendix A**) containing competent E.coli stab pipette tip was inserted into stab and spread onto pre-warmed LB-Agar plates. Selection using ampicillin (Sigma Aldrich, A53554) (100 µg/mL) and colonies picked and cultured in 6mL LB broth and placed on a shaker overnight. 500µl of bacteria culture was added to 500µl of glycerol and frozen at -80°c.

2.1.2 Plasmid purification

Plasmid purification of remaining culture was conducted using Nucleospin-Plasmid mini kit (Qiagen, 27104) and quantified using Nanodrop G908 (ThermoFisher Scientific).

2.1.3 Plasmid Construction

Primers for Puromycin resistance gene extraction of the (pLJM1-EGFP) plasmid, for a later selection marker, were designed using in-fusion cloning primer design tool by Takarabio and were added to PCR master mix containing CloneAmp HiFi PCR premix (TakaraBio). 5µl of the amplified puromycin inserts were loaded into 1.5% agarose gel containing ethidium bromide alongside 100bp ladder (ThermoFIsher) and run at 120v for 20 minutes. PCR products were cleaned up using Genejet kit (ThermoFIsher) and eluted in 20µl of elution buffer. The FUCCI plasmid was linearised using restriction enzyme MIVI incubated at 37°c, run on a 1.5% agarose gel and purified using Genejet kit (ThermoFIsher). The plasmids were ligated using In-Fusion HD cloning kit (Takarabio) and transformed into bacteria, picked, and purified as per previous protocol. The ligated plasmid is here on out referred to as FUCCI-PLJM1.

2.1.4 RNA extraction

All RNA (unless otherwise stated) was extracted by pelleting cells submerged in 1% PBS-BSA centrifuged at 400g for 4 minutes at 4°c. Supernatant discarded and 350µl of β-mercaptoethanol containing Buffer RLT and homogenised. The remainder of the steps was carried out using the RNeasy Micro kit (Qiagen, 74004) including on DNA digestion using DNase1 and resultant RNA was eluted in 50µl of nuclease free water.

2.1.5 RNA and DNA quantification and quality assessment

RNA was quantified using RNA-50 and DNA using DNA-50 on the Nanodrop 2000 using water as a blank. Further Qubit™ RNA quantification was used for our low yield RNA using the high sensitivity assay kit following kit specifications to generate relative standards and 1µl of sample to 199µl of working solution and quantified using Qubit™ 3.0 fluorometer. Pre sequencing RNA quality was checked using Tapestation High sensitivity RNA (Agilent

Cat: 5067-5579). Performed by University of Birmingham Genomics Facility.

2.1.6 Gel Visualisation

All DNA sample analyses requiring size determination or relevant extraction were loaded into 1.5% agarose gel (Sigma Aldrich; A9539) containing ethidium bromide (BioRad Cat:1610433) alongside 100bp ladder (ThermoFisher, 15628050) and run at 120v for 20 minutes.

2.2 Cell Biology

2.2.1 Organoid Culturing

Organoids were split once a week at a ratio of 1:12 to maintain a low confluency. Cells were disassociated using pre warmed TrypLE Express (1x) (ThermoFisher, 257943) and suspended to a 1 cell suspension using the big tip little tip method before being washed in PBS without Mg^{2+} Ca^{2+} (Sigma Aldrich, RNBL2265). Cells were counted using Tryphan Blue (ThermoFisher, 15250061) and BioRad TC20 automated cell counter replated at a seeding concentration of 50-100,000 cells in 50ul of 8mg Matrigel (Corning, 354234) in initial experiments and later 11mg BME-2 select (Bio Techne BME001-05) allowing for gel polymerisation for 15 minutes in 37°c before media is added. Organoids were maintained in 500ul of Intesticult media as prepared in **Table 2.1**, which is replenished twice a week, and kept in 37°c 5% $C0_2$. All work was carried out under cat 2 fume cupboard. Experiments that required growth past day 7, media was changed every 3 days.

5 Patient derived Colorectal Cancer organoid lines were used and provided by Professor Andrew Beggs:

- S345653 (p29) – Radio-resistant
- S366557 (p15) – Radio-resistant
- S309884 (p33) – Radio-sensitive
- S302389 (p12) - Radio-sensitive
- S292064 (p29) - Radio-sensitive

Full patient source, treatment and genotype of the lines is outlined in **Appendix B**.

2.2.2 Organoid Rejuvenation

Frozen organoid lines were kept in liquid nitrogen for rejuvenation samples were kept on dry ice before thawing in water bath and added to 5ml of DMEM F12 centrifuged at 400 x g for 4 minutes, removed the supernatant and resuspend the pellet in 50µl of BME. Intesticult media as prepared in **Table 2.1** is used supplemented with RHoK inhibitor (Stem Cell Technologies 72307) at a final concentration of 10uM. After 3 days normal intesticult media is used.

2.2.3 Doubling time calculation

To establish proliferation and hence doubling time, all lines are disassociated to single cell suspension using protocol in 2.2.1, cells are counted using Tryphan blue at equal volumes and using Bio Rad TC20 automated cell counter. 50,000 cells are seeded and grown for 4, 9 or 14 days. At the final time point each well is disassociated, BME dissolved using TRPLE and cell count repeated. Doubling time is calculated using the below equation.

DT = [D x (LN2)] / [LN (Ce / 50,000)]

D = Days grown for (4,9 or 14)

Ce = Cell count at end of experiment

2.2.4 Lenti-viral particle production

HEK293 P5, fast growing and transducible (FT) were rejuvenated in 5ml HEK media and seeded into T5 flasks. Cells were maintained and split 1:6 into T75 flasks and allowed to grow for 3 days. The media was washed and disassociated using TryplE and cells were counted using Tryphan blue and cell count. 2 million cells were seeded in a T25 flask for the control and 6 million cells for the viral particles, cells were then grown to 95% confluency. 10.2µg of FUCCI-PLJM1 or FAST-FUCCI plasmid was added to transfection master mix 1 (**Table 2.1**). Master mix 2 was inverted and added to Master mix 1 and 3750µl was added to T75 containing HEK293 and incubated at 37°c for 5 hours, before replacing with fresh clean HEK media and allowed to grow for 48 hours. Viral particles were collected into cryovials and stored at -80°c.

2.2.5 Genetically modified organoid line production

2.2.5.1 FUCCI-PLJM1 line

3 colorectal cancer organoid lines were used:

- S345653 (p134) – Radio-resistant
- S366557 (p81) – Radio-resistant
- S309884 (p129) – Radio-sensitive

2.2.5.2 FAST-FUCCI

2 colorectal cancer organoid lines were used:

- S302389 (p16) – Radio-sensitive
- S366557 (p19) – Radio-resistant

Each line was transduced using the FUCCI-PLJM1 or FAST-FUCCI plasmid containing lentiviral particles. Organoids were disassociated to single cell plated with propylene containing media and 100ul of viral particles per well before being span for 2 hours and then incubated at 37c for 3 hours. The viral particles were then washed away, the cells pelleted and replated in Matrigel domes into individual wells and Intesticult media was added. Control lines containing no viral particles were plated for all 3 lines to maintain controls at the same passage.

2.2.5.3 Puromycin Selection

To check the transduction was successful puromycin containing media was added to the organoids including one of the control wells. To identify the optimal selection dosage varying concentrations of puromycin (Gibco, 2260194) were added: 2ug, 4ug, 6ug, 8ug, 10ug and 12ug per well was added after splitting, at one cell stage and allowed to incubate at 37c for 4 days. Optimal 4ug was identified for all lines. Successful selection was identified by cell survival and presence of fluorescent cells compared to control cells treated with puromycin which did not survive.

2.2.6 Irradiation Protocol

Organoids undergoing irradiation for the single cell sequencing experiment were grown for 4 days with no treatment. From days 5 to 9 lines were irradiated using the CellRad Irradiator once a day at 5Gy, a previously determined optimal dose. The lines were then left to recover or establish irradiation associated affects for 5 days before cell recovery and hence sequencing on day 14.

2.2.7 Standard imaging

To identify live cells, each well of organoids was stained with 500ul of Hoechst (ThermoFisher, 62249), the plate covered with foil and incubated at 37c 5% CO_2 for 30 minutes. Organoids were imaged short-term using EVOS AMEFC4300 Thermo Fisher microscope.

2.2.8 Fixation and high-resolution imaging

To visualise organoid lines in high resolution Zeiss Z1 lightsheet, organoids were recovered by removing media, washing with ice cold PBS followed by 500µl of Organoid Harvesting Solution (OHS) (Corning, 1252001) using a Pasteur pipette were transferred into 15ml falcon tube and kept on ice and shaker at 60rpm for 30 minutes to dissolve BME. Samples are then span at 100g for 3 minutes at 4°c supernatant discarded, and pellet of organoids washed in 1% PBS-BSA span again and supernatant discarded. Pellet is then resuspended in 1ml of 4% PFA (Sigma Aldrich, MKCH0868) kept at 4°c for 45 minutes before adding of PBST to inactivate formalin. To image, samples were spun at 100g for 3 minutes, supernatant discarded and resuspended in 1.5% low melting point agarose (Sigma Aldrich, A4018) and pulled into appropriate lightsheet capillary before solidification of the agarose. Samples were imaged in the Zeis Z1 lightsheet using and lasers 488-30 (488nm) and laser 561-20 (561nm) at 1.8 and 1.2 intensity respectively for 287ms exposure and following filters LBF 405/488/561/640. Lightsheet chamber is filled with PBS without Mg^{2+} and Ca^{2+}.

2.2.9 Image analysis

All Image analysis was carried out using ZEN blue 3.1 or ZEN black (software 2.3)

2.3 Longitudinal high-resolution imaging protocol

2.3.1 Long-term high-resolution imaging protocol overview

The experimental protocol for sample preparation to sufficiently carry out our designed long-term high-resolution imaging protocol to facilitate the spatiotemporal patterning of fluorescent reporters in 3D cultures is outlined in **Figure2.1**. To set up the vertical culturing, the organoids are seeded into BME or Matrigel as previously outlined with the exception of setting into a 24 well plate, instead the organoid suspension is set in the Z1 lightsheet black capillary (GmbH, size 2, 701910). Upon gel polymerisation the suspension is the

inserted into an Eppendorf containing relevant holding medium. The hybrid gel is then pulled into a larger green Z1 lightsheet capillary (GmbH, size 3, 701910) resulting in BME containing organoids being encased in the holding medium gel as outlined in **Figure 2.1c**. The capillary can then be loaded into the lightsheet as per standard protocol. The sample is suspended into the chamber containing PBS using the plunger ensuring only ¼ of the gel is exposed to prevent breaking. Imaging is done using conditions specified in 2.2.8. In between imaging the capillary is inserted into a falcon tube containing Intesticult media (**Figure 2.1b**), through a hole equal diameter of the capillary, inserted into a modified T25 tissue culture lid. The vertical culturing falcon tube is kept in a test tube rack in a standard tissue culture incubator at 37°c 5% C02.

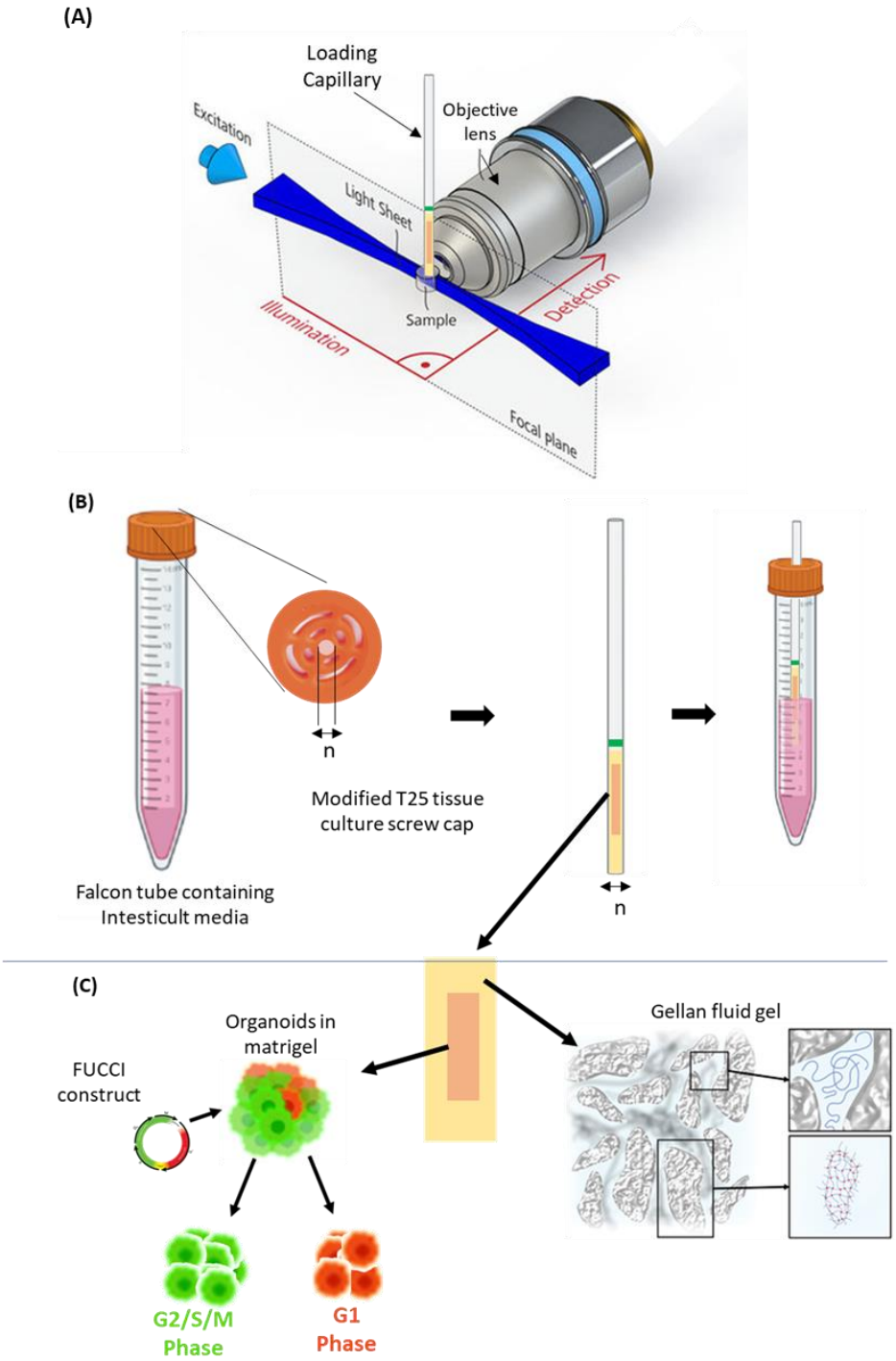


Figure 2.1. Schematic representing sample preparation for long term imaging protocol. (A) Zeiss Z1 lightsheet sample set up showing suspension of holding capillary into imaging chamber. Loaded gel embedded sample is pushed into focal plane and hence hanging from capillary. Adapted from (Eberle et al., 2015) (B) Schematic showing modified vertical culturing design. Organoids are embedded into the Z1 lightsheet capillary and suspended into organoid media through a modified T75 tissue culture lid retained on a 15ml falcon containing the media. The capillary is removed from the falcon and placed into the lightsheet as per standard protocol (C) Schematic of gel embedding set up, with organoid suspended in Matrigel or BME encased in Gellan fluid gel. Microstructure of Gellan sheering, and force properties adapted from (Cooke et al., 2018).

2.3.2 Agarose formation and embedding

1.5% low melting point agarose solution was created using PBS and heated until dissolved and allowed to cool before complete polymerisation the solution is pulled into glass Z1 lightsheet loading capillary. Using smaller glass capillary containing organoids embedded in BME is inserted into the capillary containing LMP agarose and dispensed into the gel.

2.3.3 Gellan preparation

2g of Gellan gum powder (Apollo Scientific, 71010-52-1), 2% weight/volume ratio, was added to 90ml de-ionised water, 10ml PBS and 5ml 0.2M NaCl and stirred on a hot plate at 175°C and 450rpm until dissolved. The solution was then autoclaved. Following this, the Gellan gum was transferred to a hot plate and span at room temperature until solidified and cooled. The Gellan gum was stored in a 4°c refrigerator until experimental use.

2.3.3.1 Gellan microstructure optimisation

Using the above protocol the preparation of Gellan meets many of the required criteria for the holding medium needed to suit the imaging protocol. Gellan is optically clear and retains the same refractive index as water meaning imaging visualisation is not distorted. The microstructure enables sufficient permeability allowing for delivery of nutrients from the media to reach the embedded organoids. However, the structure of the conventional preparation approach was not strong enough to enable the embedding of the organoids or to withstand the vertical culturing and imaging. Therefore, alterations to the Gellan reagent compositions, notably an increase in the NaCl concentration was advised and used to enable the formation of stronger ionic bonds. The optimisation process resulted in the variations of Gellan compositions outlined in **Table 2.2.** Initial testing showed that Gel ID 3 and 6 lacked optical clarity, furthermore viability testing showed cell growth was significantly impacted in these gels. Our testing showed Gel ID 4, was able to withstand the structural demands of the culturing and imaging set up and did not affect the growth of the organoids, therefore all proceeding experiments were carried out using Gel ID 4.

Table 2.2. Gellan reagent composition modifications

Gel ID	Gellan %	NaCl (mM)	PBS
1	1.5	10	50ul
2	1.5	25	50ul
3	1.5	50	50ul
4	2	10	50ul
5	2	25	50ul
6	2	50	50ul

2.3.3.2 Gellan viability testing

Viability testing assessing continued organoid growth was performed on Gels: 1,2,4 and 5. Gellan was dispensed into a well of 24 well Cornig tissue culture plate, at around 2mm thickness. Utilising the conventional plating outlined in 2.2.1, the BME containing organoids were added on top and in the centre of the Gellan base. After the BME had polymerised a further 2mm of Gellan was added on encasing the BME in Gellan, 700µl of media was added. Lines were grown for 4 days, and their viability assessed by imaging.

2.3.4 Imaging and culturing set up

Organoids embedded in BME were encased in the Gellan Gel ID 4 as outlined in **Figure 2.1**. The sample was kept in the vertical culturing condition and then imaged at 7am and 7pm in the Z1 lightsheet. The chamber was filled with PBS containing both NaCl and Mg2+ to ensure we did not alter osmotic balance; temperatures were also kept at 37°c. Zeiss Z1 lightsheet microscope was used for this the first track set for the FUCCI the selected lasers 488 and 561nm at 1.8 and 1.3 intensity and 219.71 millisecond (ms) exposure, this was kept consistent across all imaging. The Z1 lightsheet enables full 3D reconstruction of a sample ad thus allows for altering 3 dimensions, X Y and Z axis to ensure full free movement. The relative coordinates, the axis values, of each organoid or organoid cluster were noted and upon each imaging set up we reinserted them into the Zen software to ensure we were imaging the same cells.

2.4 Sequencing

2.4.1 Single Cell experiment

General outline of the single cell protocol is outlined in **Figure 2.2** organoid lines and

plating passage point used:

- S302**389** (p24) – Radio Sensitive (High YC):
    - Control
    - Irradiated – prepared as per protocol in 2.2.6
- S366**557** (p28) – Radio-resistant (High YR)
    - Control

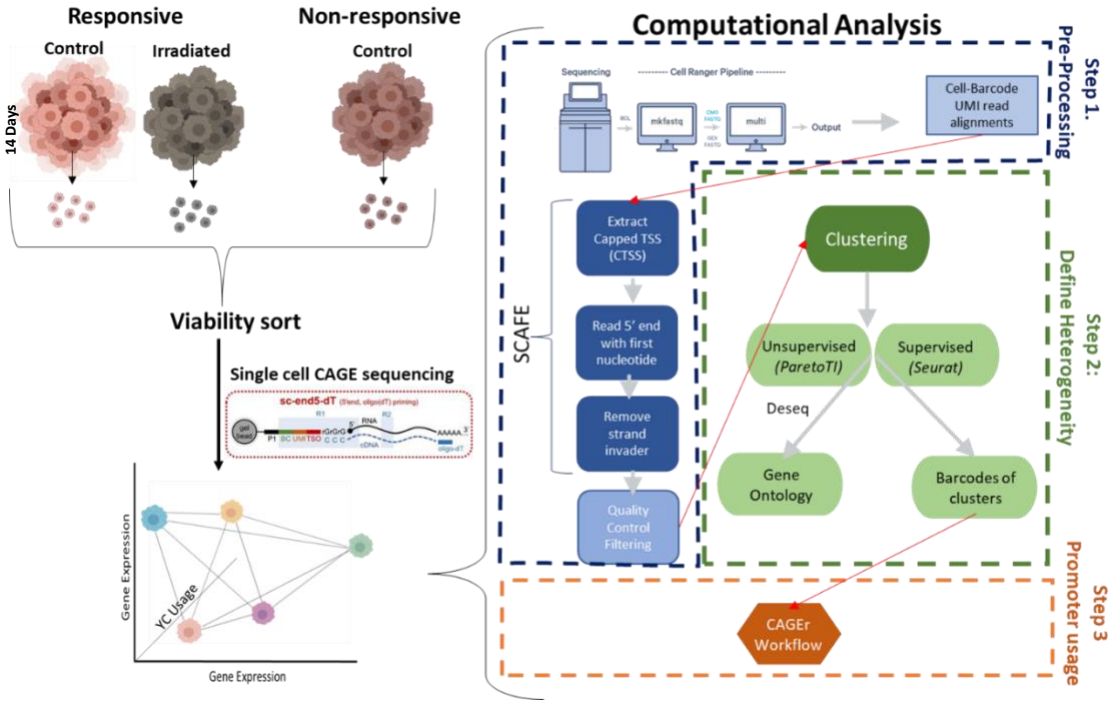


Figure 2.2. Schematic representation of Single cell CAGE protocol. Experimental and sample preparation of single cell suspension from 14-day old organoids, viability sort using cell death staining and negative fluorescent gating in FACS sort. Computational outline to establish YC usage attributable to cellular states.

Single cell Sample preparation:

5 wells for each control were grown for 14 days as per protocol in 2.2.1 and 10 wells of S302389 as per irradiation protocol outlined in 2.2.6. On the 14th day cells BME domes were washed with ice cold PBS, removed from 24 well plate using Pasteur pipette and transferred to 15ml Falcon tube, span at 400 x g for 4 minutes at 4°c, PBS discarded and 1ml of OHS used to gently resuspend the pellet, kept on ice and shook at 60rpm for 30 minutes to dissolve BME span at 400 x g for 4 minutes and pellet resuspended in 1ml of Tryple to disassociate cells for 5 minutes and neutralised with 2ml of ice cold PBS followed by light mechanical disassociation using a 1% PBS-BSA coated P10 pipette tip attached to a

P1000 pipette tip. Cells were span at 400 x g for 4 minutes and resuspended in 1% PBS BSA and filtered using FlowMI Cell Strainer (40uM) (Bel-Art J332333). Post filtered cells are counted to assess viability and span again then resuspended in 100µl of 1:100 dilution of Zombie Nir cell death stain (Biolegend, 423105), only permeable and, hence staining the dead cells. These are covered with foil and incubated for 20 minutes on ice before being washed with ice cold PBS-BSA and resuspended in 1ml of 1% PBS-BSA. In all occurrences, unless stated otherwise, samples are kept at 4°c.

Flow cytometry of live cells:

Post filtered cells were then sorted using Flow cytometer using Cyan flow cytometer machine, to ensure retention of high cell viability and thus limit cellular stress we ensured sorting was carried out on the lowest speed of using 100um nozzle and frequency of 28.3KHz using the gating strategy using laser at 561nm to detect stained cells and hence non-fluorescing gating was kept the same across all 3 samples and scatters outlined in **Figure 2.3**. Flow sorting was carried out by Dr Mary Clarke.

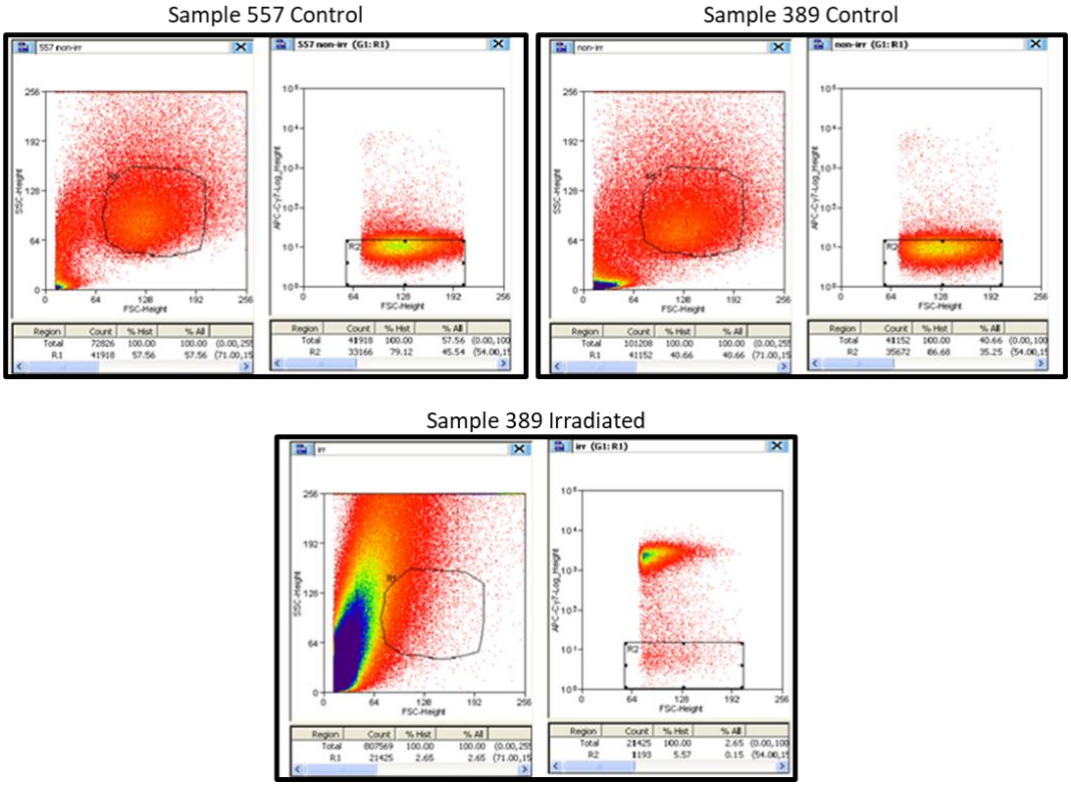


Figure 2.3. Cell Viability Flow cytometry gating strategy. Left panel of each indicates forward scatter and right indicates cell death staining, negative cells highlighted in black box.

Single Cell Library Preparation:

Post sorted cells were kept on ice and viability checked before proceeding with library preparation was assessed using Tryphan Blue and Bio Rad TC20 counting as previously outlined and only continued if viability as above 80%. Both Control samples had 98% viability and irradiated had 79%. The single cell protocol was the standard single cell 5' mRNA V2 protocol with the key alteration of the read sequencing cycles outlined in the protocol by (Moody et al., 2022). Conventional single cell uses 90 cycles from read 2 and 26 cycles from read 1, however, to sufficiently capture the 5' end the read cycles are switched with read 1 at 90 cycles.

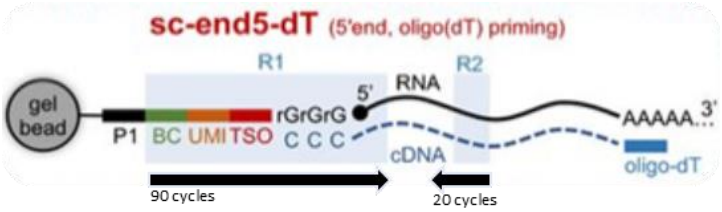


Figure 2.4. Single cell adaptor structure. Chromium 10x linkers, 5' linker containing Barcode (BC) Unique molecular Identifier (UMI) Template switching oligonucleotide (TSO). Switched read PCR preference to 90 cycles from Read 1 (R1) and 20 cycles from Read 2 (R2). Adapted from (Moody et al., 2022).

Table 2.3 Cell Input for single cell library preparation:

Sample no	**Av cell count**	**Av live %**	cells	water
Sample 1 -389 irradiated	1.53x 10^5	79	33	5.7
Sample 2 -389 Control	2.87 x 10^5	98	16..5	22.2
Sample 3- 557 control	2.57x 10^5	98	16.5	22.2

Table 2.4 cDNA profile for single cell library

Sample	**cDNA PCR cycle no**	**cDNA stock pg/ul**	**Total yield in 40 ul (ng)**	**50 ng input for library prep**	**PCR cycles for library prep**
1	12	1,630	65,200	30.67484663	15
2	12	29,300	1,172,000	1.706484642	15
3	12	25,500	1,020,000	1.960784314	15

Individual cells are barcoded using the bead and barcode system outlined in **Figure 2.4** using the Chromium Next GEM single cell 5' Kit V2 (10xGenomics, PN-1000263) with differential sample specific indexing before pooling and loading onto the Chromium™

Single Cell K Chip to add gel beads before being sequenced on the Illumina Nova Seq 6000 using the S1 v1.5 Flowcell (Illumina, 20028318) . Two control samples were sequenced at average 289,401 reads per cell for sample 3 557 and 266,347 reads per cell for sample 2 389. Irradiated sample 1 (389) was sequenced at average 71,080 reads per cell. cDNA profiles were all checked for sufficient RNA quantity and quality. Library preparation and sequencing was carried out by Birmingham university Genomics department.

2.4.2 FUCCI SLIC CAGE

Sample preparation:

5 wells seeded at 50,000 density of Sample 557 FUCCI-transduced lines generated in **2.2.5.2** were grown to day 14 as per previously outlined protocol. Single cell suspension was prepared as described in 2.4.1. Cells were then sorted into relative cycling populations based on fluorescent reporter expression presence using the 530/30 filter and 488nm laser to detect the Geminin and hence GFP^{+} G2M population. The 582/15 filter and 561nm laser to detect the mKO2 and hence RFP+ G1 cells, those not fluorescing were also captured and determined as G0 outlined in **Figure 2.5**.

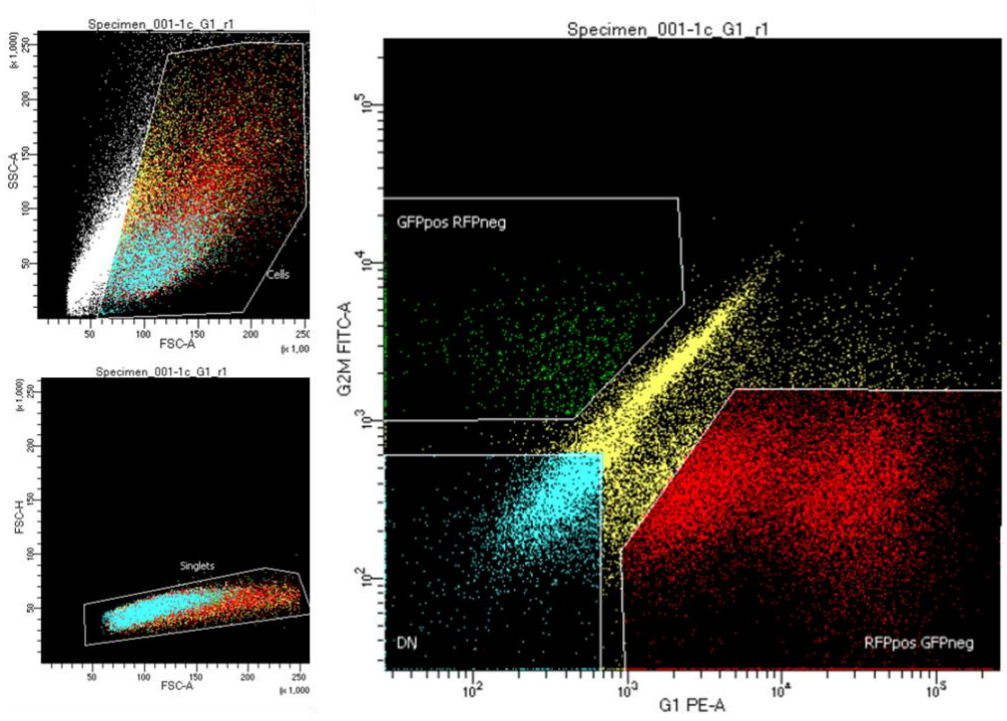


Figure 2.5. FACS gating strategy for FUCCI cycling phase separation. Gating of 557 FUCCI line. Left top indicates forward vs side scatter to determine cell vs debris and bottom individual cells. Right panel highlights mKO2 (or G1) fluorescence on X axis vs Geminin (or G2M) on Y axis. Gating of relevant G1 populations: RFP positive, GFP negative in red. G2M populations: G2M cells, RFP negative in green). G0 cells: RFP negative and GFP negative in blue. Uncollected cells depicted in yellow.

RNA was immediately extracted and quantified from each population using protocol outlined in 2.1.4 and 2.1.5. RNA Tapestation quality assessment.

SLIC-CAGE library preparation:

The following protocol was carried out with Dr Yavor Hadzhiev

Table 2.5 SLIC-CAGE Reagent compositions

Name	Reagents	Catalogue /Lot
Reverse Transcription Enzyme Mix (vol per 1 reaction)	Nuclease free water (4.6µl)	Thermofisher 1039-498
	5x First strand buffer (7.6µl)	Thermofisher 18080-044
	0.1M DTT (1.9µl)	Thermofisher 18080-044
	10mM dNTPs (1µl)	Thermofisher 18427-088
	Trehalose/Sorbitol (7.6µl)	Sigma Aldrich 618-234
	Superscript III RT (3.8µl)	Thermofisher 18080-044
250nM $NaIO_4$	1mg $NaIO_4$	Sigma Aldrich MKBW2436V
	Nuclease free water 18.7µl	Thermofisher 1039-498
10mM Biotin	50mg biotin	Sigma Aldrich B4501
	DMSO (13.5ml)	Sigma Aldrich 102664507
1M Tris-HCL (vol per 1 reaction)	1M TRIS-HCL (14µl)	Sigma Aldrich SLBV0814
	Glycerol (2µl)	Sigma Aldrich G5516-1L
RNAseOne mix (vol per 1 reaction)	10x RNAseONE buffer (4.5µl)	Promega M4265
	RNAseONE 10U/µl (0.5µl)	Promega M4265
Wash Buffer A	4.5M NaCl2	Sigma Aldrich BCBZ6730
	50mM EDTA pH 8	Sigma Aldrich BCCD1632
	0.1% Tween	ThermoFisher BP337-500
Wash Buffer B	10mM Tris-HCL	Sigma Aldrich SLBV0814
	1mM EDTA	Sigma Aldrich BCCD1632
	0.5M NaOAc	Alfa Aesar Z06E505
	0.1% Tween20	ThermoFisher BP337-500
Wash Buffer C	0.3M NaCl2	Sigma Aldrich BCBZ6730
	1mM EDTA	Sigma Aldrich BCCD1632
	0.1% Tween	ThermoFisher BP337-500
Rnase mix	Nuclease free water (2.4µl)	Thermofisher 1039-498
	10x RNAseONE buffer (0.5µl)	ThermoFisher EN0531
	Rnase H (0.1µl)	New England Biolabs M0297
	RNAseONE (2µl)	ThermoFisher EN0531
2nd Strand Synthesis Mix (vol per 1 reaction)	10x ThermoPol reaction buffer (5µl)	New England Biolabs B9004
	Nuclease free water (2µl)	Thermofisher 1039-498
	10mM dNTPs (1µl)	Thermofisher 18427-088
	DeepVen(exo-) DNA Pol (2U/ul) (1µl)	New England Biolabs M0259
Enzyme Degradation mix	I-SceI (RI)	Biolabs 10128048
	ICeuI (RI)	Biolabs 10033476

Reverse Transcription to produce cDNA

The SLIC-CAGE library was prepared following the nAnT-iCAGE protocol outlined in (Murata et al., 2014) with the appropriate modifications for SLIC-CAGE specified in Cvetesic et al., 2018 to accommodate the low-level RNA input obtained from our FUCCI populations. 800ng of each sample was added to 4200ng of previously prepared carrier mix as per specification, to this 1µl of 2.5mM N6 primer mix was added per solution and incubated at 65°c for 5 minutes and immediately placed on ice to denature the RNA. 28µl of reverse transcription (RT) enzyme mix was added to each sample mixed and incubated at 25°c fir 30 seconds followed by 50°c for 1 hour. The samples are then purified using RNAClean XP beads at a ratio of 1.8 to sample volume, incubated at room temperature (RT), separated using a magnetic stand, washed twice using freshly made 70% ethanol before eluting with 42µl of water ensuring pipetting up and down 40 times before incubating at 37°c for 5 minutes returning to magnetic stand and transferring supernatant to new PCR tube.

5' cap oxidation and biotinylation

The next step involved oxidation of the 5' cap by combining the supernatant with 1M NaOAc at pH 4.5 mixing and adding 2µl of NaIO4 as prepared in **Table 2.5**, covered with foil and kept on ice for 45 minutes. To stop the oxidation reaction 16ul per reaction of 1M Tris-HCL (pH8.5) solution is added. RNAClean XP purification is repeated as previously detailed. Biotinylation of the samples is carried out by adding 4µl of NaOAc (pH 6) mixed and 4µl of Biotin solution outlined in **Table 2.5** mixed and incubated at 23°c in dark. 12µl of 2-propanol (Fisher Scientific, 12608655) is added to the biotinylated sample followed by another round of RNAClean XP purification.

RNase digestion and Cap-trapping

Purified biotinylated samples were digested with 5µl of RNAseOne mix outlined in **Table 2.5** incubated at 30 minutes. For cap -trapping M-270 beads were prepared by adding 30µl of beads per sample to 0.75µl tRNA (20µg/µl) (double standard amount as we used yeast tRNA) and incubated on ice for 30 minutes and ensuring mixing by flicking the tube every 5 minutes. The beads are then washed twice in Wash buffer A, outlined in **Table 2.5**, followed by resuspension in Wash Buffer A containing tRNA (20µg/µl). 105µl of the

resuspended beads are added to the RNAseOne treated sample and incubated at 37°c for 30 minutes followed by sequential washing with Wash Buffer A, Wash Buffer B and then Wash buffer C on the magnetic stand. cDNA is then released from the beads by adding 10x RNaseOne buffer in equal volume incubated for 5 minutes at 95°c and immediately on ice for 2 minutes before returning to magnetic stand and transferring supernatant to new tubes. 30µl of RNaseOne buffer was added to beads mixed and the supernatant transferred into the previous aliquot resulting in 65µl of sample. Cap-trapped samples incubated with RNase Mix for 15 minutes at 37°c and purified now using AMPure XP purification beads using same protocol as the RNAClean XP beads then mixed with 5µl of RNAseOne mix incubated at 30 minutes for 37°c before repeating AMPure XP purification. Samples are then Speed Vacuumed at 37°c for 40 minutes and the pellet resuspended in 7µl of water.

Linker Ligation

Single Strand linker ligation (SSLL) mix was created by incubating cap-trapped sample at 95°c for 5 minutes and 5’ single strand linker (2.5µM) at 60°c for 5 minutes before placing both on ice, both are then combined with 16µl of Mighty ligation Mix and incubated at 16°c for 16 hours. cDNA purification of the 5’SSLL mix and hence removal of additional linker was performed by 2 rounds of AMPure XP bead purification, repeating speed vac and resuspending pellet in 7µl of water. Second SSL was carried out this time for the 3’ linkers following same protocol as the 5’linker addition.

Linker Degradation and second strand synthesis

Following AMPure XP bead purification cDNA was dephosphorylated and hence treated with 10µl of SAP mix and incubated at 37°c for 30 minutes followed by 65°c for 15 minutes. 3’ linker degradation was achieved by adding 2µl of Uracil specific excision enzyme and incubated at 37°c for 30 minutes, 95°c for 5 minutes and immediately on ice for 2 minutes to prevent reannealing. The samples were then purified using AMPure XP beads. To carry out second strand synthesis we added 10µl of second strand synthesis mix to the sample incubate at 95°c for 5 minutes, 55°c for 5 minutes, 72°c for 30 minutes and held at 4°c. Degradation of the second strand primer was carried out through adding 1µl of Exonuclease I and incubated at 37°c for 30 minutes and purified using SPRI magnetic beads as previously described.

Carrier degradation

To remove the carrier, we mixed the cDNA samples with the degradation mix outlined in **Table 2.5** and incubated at 37°c for 3 hours and deactivation of the enzyme by incubating at 65°c for 20 minutes. The sample was then purified using SPRI beads in a 1:8 ratio as previously described. To amplify the prepared libraries the samples were mixed with PCR mix and amplified using the following cycle: 95°c for 3 minutes, 98°c for 20 seconds, 60°c for 15 seconds, 72°c for 2 minutes, steps 2 to 4 were repeated for 8 cycles followed by 72°c for 2 minutes and held at 4°c. A second carrier digestion was performed using a 0.7x bead ratio to remove the shorter strand linkers that may have retained. RNA quality and quantity was characterised using protocol in 2.1.5 and 6.

2.5 Computational analysis

2.5.1 Preprocessing and mapping of sequencing libraries

2.5.1.1 Single cell data

The single cell generated data was demultiplexed, trimmed and mapped using the Cell ranger 7.0 pipeline, the latter steps incorporate the STAR alignment algorithm using the Hg18 genome reference followed by recalibration using MAPQ adjustment to identify potential misalignments (Cline et al., 2020; Dobin et al., 2013). Reads are then filtered to correct for barcode identity and UMI alterations and UMI count is generated and called to relevant barcodes and hence real cell identity producing a filtered UMI count file. To account for G biases that exists in the TSO methodology, the reads were further filtered for TSO causing artefacts using the SCAFE algorithm (Moody et al., 2022). Firstly, identifying the extra G-mismatches, this information is then fed into the multiple logistic regression model to identify actual TSS clusters, removes stand invaders or 'artefacts' and validates TSS clusters through parametric clustering of the TSS 5'ends using *Paraclu*.

2.5.1.2 SLIC-CAGE

Demultiplexing, trimming and mapping were carried out using Bowtie (v.1.3.1) with mapping criteria set to include only 2 mismatches and a MAPQ score greater than 20 (Langmead et al., 2009).

Both above analyses were performed by Dr Yavor Hadzhiev.

2.5.2 Quality control assessment

Post filtered reads per cell were analysed for the presence of multiplets using Scrublet algorithm running default parameters (Wolock et al., 2019). Resultant reads were imported into a Seurat object and cells with mitochondrial mapping above 10% were removed (Butler et al., 2018).

2.5.3 Cellular state analysis

To investigate ITH originating from actual biological and not technical heterogeneity, reads per cell were analysed using Seurat V5, data was normalised using scTransform V2 utilising Pearson residuals generated from negative binomial regression to correct gene weighting scaling (Choudhary & Satija, 2022). To identify clusters dimensionality reduction analysis was performed by principle component analysis (PCA) using default settings computing 50 PCs. Clustering was performed using K-nearest neighbour graph generation using previously generated PCA space followed by application of the FindClusters function set to 0.5 resolution parameter. Nonlinear dimensional reduction uniform manifold approximation and projection (UMAP) embedding using the umap-learn package (McInnes et al., 2018).

2.5.3.1 DGE analysis

Markers defining clusters were calculated using the Seurat FindMarkers function using either Wilcoxon Rank sum testing default or for more in depth comparative analysis where specified using DESeq2 (Love et al., 2014). Analysis of upregulated gene sets was performed using Seurats ModuleScore function, identifying average score of whole gene set. DGE testing of data without repeats was performed using SOM clustering analysis using the Kohnen R package, utilising normalised CAGE tags from all signal not separated by initiator. Gene ontology analysis was calculated using the ShinyGO 0.8 online platform http://bioinformatics.sdstate.edu/go/ (Ge et al., 2020).

2.5.3.2 Pathway Enrichment analysis

To identify the upregulation of gene subsets and pathways, Seurats Module score function was used (varying bins of reference genes were used, the standard 100 was

used unless states otherwise). Gene sets were obtained from GSEA database, utilising the identifiers outlined in **Table 2.6**.

Table 2.6. Gene set enrichment source list

Gene Set	Source
Proliferation	GSEA – Acquisition: GO:0008283
Wnt signalling	GSEA – Acquisition: PAHS-043A
Myc Signalling	GSEA – Acquisition: (Schaefer et al., 2009)
mTOR signalling	GSEA – Acquisition: hsa04150
Differentiation	(X. H. Zhao et al., 2023)
Hypoxia	GSEA – Acquisition: M5891 (Liberzon et al., 2015)
Hypoxia ATM	(Ho et al., 2020)
Hypoxia ATG	(Ho et al., 2020)

2.5.3.3 Integration

Comparative analysis between datasets was achieved using Harmony (Korsunsky et al., 2019) embedded in the Integrate layers function of Seurat and referencing mapping of cell type and universal archetype comparison using Anchor-based CCA integration (Stuart et al., 2019).

2.5.4 Archetype analysis

Polytope analysis of single cell data sets to determine archetype vertices was carried out using the ParetoTI algorithm (Hart et al., 2015)utilising the ParetoTI package in R. The polytope fitting was achieved by determining the lowest vertices achieves (hence archetypes) that retained statistical significance. Genes enriched at each archetype were extracted and analysed in ShinyGo to determine general Gene ontology. To determine likeness to universal archetypes (for both single cell and bulk comparisons) enriched genes were compared to the universal markers outlines in (Hausser et al., 2019). 3D projections were created using Plotly (Sievert, 2020).

2.5.5 Promoter analysis

To identify TSS the generated BAM files containing filtered reads were imported and analysed using the CAGEr V4.3.1 (Haberle et al., 2015). Data was normalised using power-law distribution with appropriate α values calculated through reverse cumulative plotting and set to x 10^6. Consensus clusters were calculated as CTSS within 100bp and those falling

within the promoter region defined by mapping to the hg38 Ensembl reference genome (Martin et al., 2023). TSS visualisation was achieved using rtracklayer (Lawrence et al., 2009)to generate bedgraph files subsequently uploaded to UCSC genome browser (Kent et al., 2002).

2.5.5.1 TSS identification and Ratio calculations

Classification of TSS class identity was achieved by summing CTSS based on initiator sequence YR = 'CG' 'CA' 'TG' 'TA' and YC = 'CC' 'TC'. TOP transcripts were characterised by presence of conventional TOP motif 'CYYYY' in the first 5bp. To calculate dinucleotide frequency, initiation site (Inr) were extracted and CTSS per Inr were summed and divided by the total signal to calculate % of Inr. Finally, we summed the YC signal, TSS dinucleotides represented as 'TC' or 'CC', and YR dinucleotides defined as 'CA', 'CG' , 'TA' or 'TG' to define % tags per million (TPM) to initiator class. To calculate dual gene ratios TPM per initiator class were summed per CC and divided by total signal at that CC. Full list of consensus Dual genes identified are outlines in **Appendix C**.

2.5.5.2 Ratio comparisons between states and subpopulations and assessment of ribosomal Bias in trends

To compare YC:YR between subpopulations, frequency distribution analysis as performed on YC:YR values for all Dual genes present in the samples compared. This was achieved by calculating YC signal to YR per gene for each sample divided by the average ratio across all samples. The date in then normally distributed through log transformation. High YC samples or YC enriched are identified as those who's median (peak of the normal distribution plot) is greater than 0.5, contrastingly those whose median is less than -0.5 are classified as YC low (YR enriched). Secondly to assess the significance of the shift in ratio between usages removal of known high YC ribosomal genes (**Appendix D**) was performed, and frequency distribution analysis repeated.

2.5.5.3 Motif analysis

TOP associated motif analysis was achieved by determining defining TOP decreasing motifs using the first 5bp from initiation. YC4 = 'CYYYR', YC3 = 'CYYRR' and pure YC or YC2 = 'CY'.

2.5.6 Bulk Correlation and PCA analysis

Comparison between bulk CAGE from previously generated data was further analysed using normalised TPM count matrix provided by Dr Joseph Wragg (Wragg et al., 2023). Comparison between single cell CAGE and matched bulk was achieved by pseudo bulking and summing reads per sample and normalising reads using power law distribution. PCA analysis was performed using corrplot package (Wei et al., 2017).

2.5.7 Statistical analysis

Doubling time dot plots were performed using GraphPad Prism V8.0. Venn diagram analysis using VennDiagram package on r (H. Chen & Boutros, 2011). All other statistical analysis and graphical visualisations were carried out using ggplot2 (Wilkinson, 2011).

Chapter 3: Characterisation of high YC organoids

3.1 Introduction

Core promoter TSS switching represents a previously unappreciated level of transcriptional regulation in cancer. As discussed in Chapter 1, thousands of genes can be differentially transcribed through alternative TSS usage, resulting in the production of either YR or YC transcripts. Research has begun indicating that these transcripts may undergo different pre and post transcriptional regulation mechanisms (Nepal et al., 2020). Alternative TSS usage is not detectable in RNA sequencing, therefore its functional relevance is yet to be truly understood. A shift to YC usage at dual initiator promoters has been found pervasive in cancer. Furthermore, the YC usage increase provides a novel biomarker for irradiation responsive CRC cancers and the depletion upon treatment highlights a potential important promoter level regulation role in tumour dynamics (Wragg et al., 2023).

Differing global TSS phenotype between cancers of the same subtype highlights an unexplored level of inter tumour heterogeneity. However, our understanding of the functional relevance of YC enrichment in cancers is currently limited to the evidence that it correlates with the tumour's irradiation responsiveness. Such clinical phenotypes are underpinned by a complex network of molecular mechanisms and gene expression profiles. Therefore, this link doesn't provide biological insight into why TSS switching in subsets of genes may provide an evolutionary advantage in tumour development. In this chapter I aim to better establish the functional importance and significance of TSS usage preferences in organoid physiological states. I hypothesised that YC enriched organoids would be phenotypically distinct from low YC organoids characterised by gene expression profiles that would provide more functional insight beyond just an association with irradiation responsiveness. To test this firstly I will further phenotypically characterise CRC organoids with known global TSS profiles using imaging and transcriptomic analysis.

To truly understand the molecular phenotypes of a tumour, literature suggests it is

imperative to explore ITH (Ramón y Cajal et al., 2020). Organoids are shown to be an excellent model system for CRC in that they successfully recapitulate both inter and intra heterogeneity seen in the disease (Kashfi et al., 2018). Therefore, we hypothesised that the YC enriched and YC low organoids would have different SC populations with distinctive physiological states that we could characterise by their expression profiles. To test this, I aimed to optimise methodology to successfully carry out single cell CAGE.

3.2 Results

3.2.1 Characterisation of high YC organoids through imaging and proliferation analysis.

To better understand the molecular underpinnings and hence functional relevance of TSS usage differences, we set out to further identify the phenotypes associated with the high YC enrichment beyond just irradiation responsiveness. We selected the 5 CRC PDOs that Wragg et al., 2023 had already established their global TSS profile of using CAGE, 2 YC enriched (064 and 389), one moderate (884) and 2 YC low (653 and 557), YC enrichment was calculated as stated in **Chapter 2.2.5.2**. As discussed in Chapter 1 regulation of YC transcripts is heavily dependent on growth pathways such as mTOR, therefore we hypothesised that YC enrichment may contribute to growth dynamics. To test this, I explored the proliferative capabilities of these lines with differential YC enrichment. To determine proliferation rates all lines were seeded at 50,000 cells at day 0, grown, harvested, and counted again at days 4, 9 and 14 post seeding (PS). Interestingly culturing doubling time analysis revealed that YC enrichment also correlated with fast proliferation phenotype (**Figure 3.1**). Initially the experimental standard 4-day PS counts were performed however significant inter variability was seen within lines but still showed the aforementioned trend. To rule out that differences were less likely to be organoid initiating ability rather than proliferation rate I also grew the lines to day 9 and day 14, the latter being the timepoint at which all sequencing was carried out. Doubling time shortened for all lines in the later counts however all the time points showed proliferation rates correlate with YC enrichment.

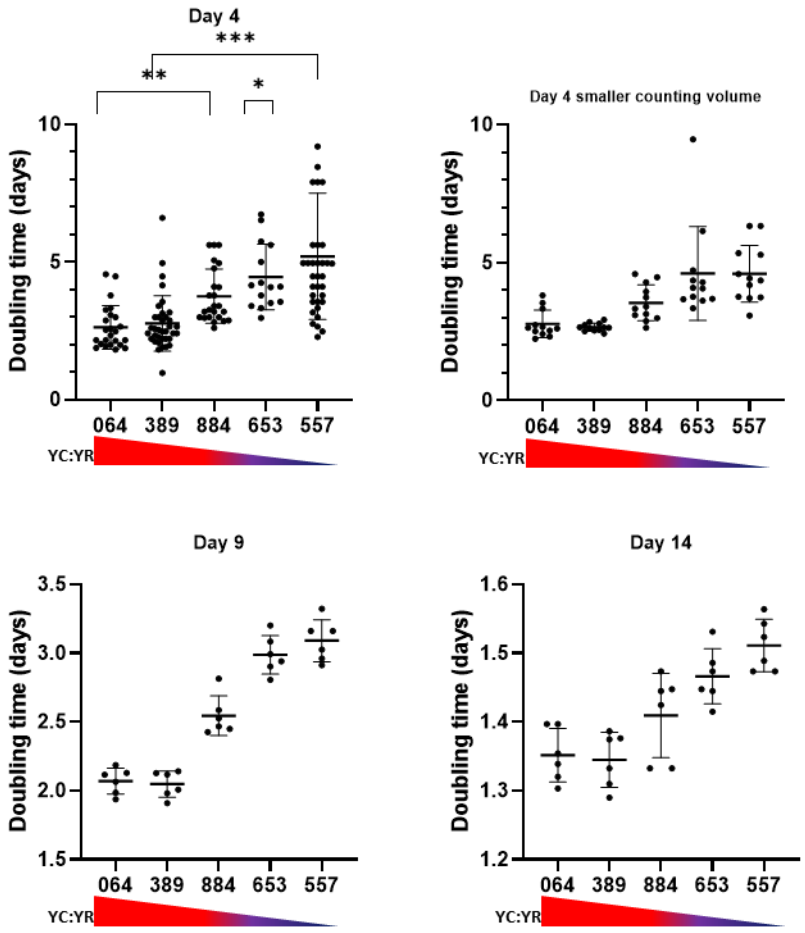


Figure 3.1. Proliferation rates decrease following global YC depletion. Doubling time analysis of CRC organoids. Dot plot representing calculated doubling time for each line ordered by YC enrichment with highest enriched on the left and YC depleted on the right. Each individual point represents 1 well counted at day post seeding (PS) (A) 4 days PS, (B) 9 days PS and (C) 14 days PS. All experiments were a minimum of n=3. * Represents n=5, ** is n=8, and *** n=10.

Uncontrolled growth and the ability to evade antigrowth signals are well characterised hallmarks of cancer formation and progression, however the dependency on this feature differs between cancers and can be a result of other tumour characteristics (Hanahan & Weinberg, 2011). As seen in in embryo development proliferation dynamics change upon lineage differentiation of cells, with growth slowing as cells move away from pluripotency (Wragg et al., 2020). Such feature is believed to be mimicked in tumour development, with fast proliferative tumours often being less differentiated (Jögi et al., 2012). I aimed to explore whether our different CRC PDO lines had different levels of cellular differentiation and whether it correlated with proliferative state and hence YC enrichment. As organoids sufficiently recapitulate host organ and hence tumour architecture including cellular composition, previous studies have characterised distinctive morphological features

associated with differentiation status through structural formations present in the organoids visible using imaging (Fujii et al., 2016; Kashfi et al., 2018).

To explore the morphological characteristics and hence differentiation status of our organoids, I grew each line under standard culturing conditions to day 14, harvested, fixed use PFA and immobilised in low melting point (LMP) agarose to image in the high resolution Z1 lightsheet microscope. Utilising our doubling time analysis information that 14 days would provide sufficient time for organoid structures to form and to allow for comparison to sequencing data. The morphology of all lines can be seen in the brightfield images in **Figure 3.2a.** The two fast growing organoids (064 and 389) showed a distinct cystic morphology, harbouring the attributable 'hollow' core a tell-tale feature of fast proliferating, undifferentiated cells lacking both segregation and polarisation capabilities (Fujii et al., 2016; Kashfi et al., 2018). Contrastingly slower growing lines 653 and 557, appeared to mimic host and hence healthy intestinal structures in the form of crypt and villus formations, highly indicative of more differentiated cellular lineages. Combining all current summary characteristics of each organoid in **Figure 3.2b** allows for identification of TSS associated trends appearing. For example, YC enriched organoids have a characteristic fast proliferating undifferentiated phenotype that responds well to treatment. Strikingly opposite correlations are seen in the slow growing well differentiated unresponsive lines harbouring a YR dominant (YC low) global TSS profile. Finally, the middle organoid further supported this intermediate observation sitting in between in all features. Interestingly, using genomic mutation data undertaken by Professor Andrew Beggs does not show a causative link to YC enrichment further highlighting the need to explore transcriptomics.

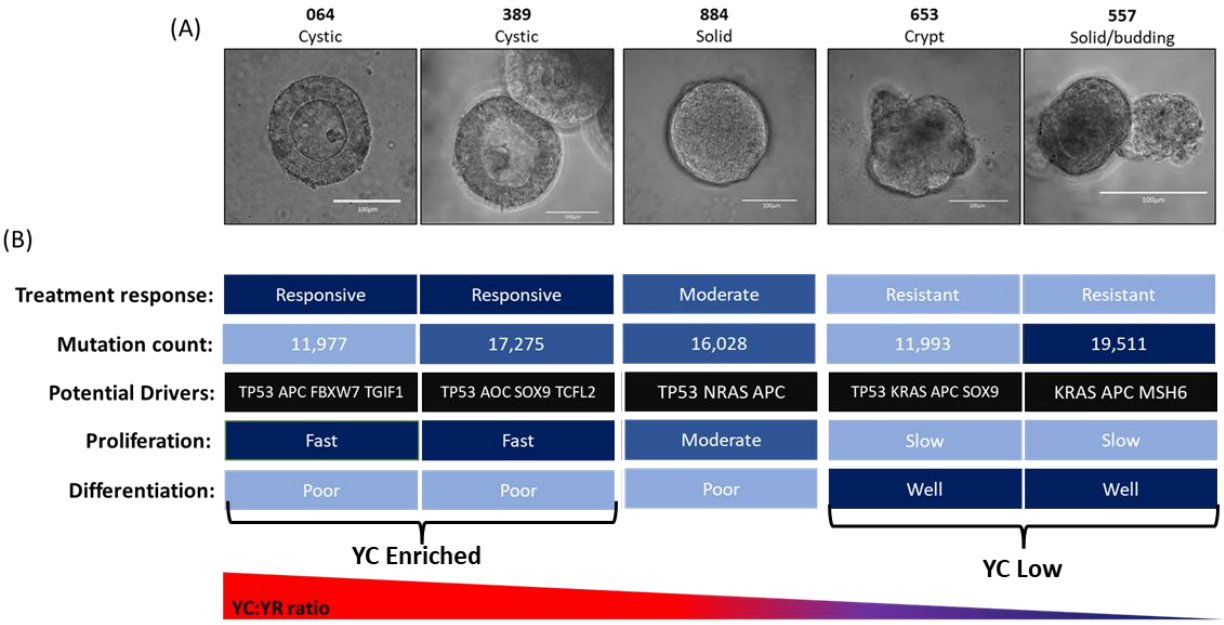


Figure 3.2. Distinct organoid characteristics correlate with YC enrichment. Summary characteristics of CRC organoids and brightfield morphology imaging. (A) Brightfield 14-day fixed images, white lined scale bar represents 100µm. (B) Previously characterised features including mutation burden and potential driver mutations (performed by Professor Andrew Beggs), irradiation treatment responsiveness and YC enrichment of YC:YR dual genes (performed by Dr Joseph Wragg) and newly characterised features from doubling time analysis and imaging morphological association.

These findings have since been published and are my own contributions *to Intra-promoter switch of transcription initiation sites in proliferation signalling dependent RNA metabolism* (Wragg et al., 2023).

3.2.2 Molecular Phenotyping of high YC organoids using bulk analysis

An advantage of CAGE sequencing is that not only does it provide transcript counts, but also captures the 5'end of these tags allowing for base resolution of TSS, the resultant sum of these tags (TPM) and their distribution allows for profiling of overall TSS usage (Takahashi, Kato, et al., 2012). Previous work performed in our laboratory characterised the global TSS profile of these organoids and hence defined them based on their YC enrichment of dual genes, depicted in **Figure 3.2**. As with RNA sequencing, CAGE also provides insight into the gene expression profiles and hence molecular state of these organoids. Dr Wragg was able to obtain insight into molecular functioning through analysis of dual genes whereby shifting of TSS usage was present and exploration of the gene ontology (GO) of these gene sets, this revealed an extensive broad range of functionalities including, proliferation, metabolism, translation, and others (Wragg et al., 2023).

To better understand the overall molecular phenotype of the organoid and see if expression data could validate and add to our current imaging and culturing analysis, I aimed to further scrutinise the expression sequencing data. We were limited in the analysis we could statistically and successfully carry out due to the absence of repeats and hence differential gene expression (DGE) was not possible. I used a generated count matrix from Dr Wraggs CAGEr analysis that encompassing the normalised TPM per gene for each organoid. Principle component analysis (PCA) of the underlying gene expression profiles highlights the variance that exists between our samples. As hypothesised the lines that display similar culturing and morphological phenotypes cluster closer to each other (**Figure 3.3**). Therefore, I identified that global TSS phenotype appears to have a distinctive gene expression profile pervasive across lines of different origin.

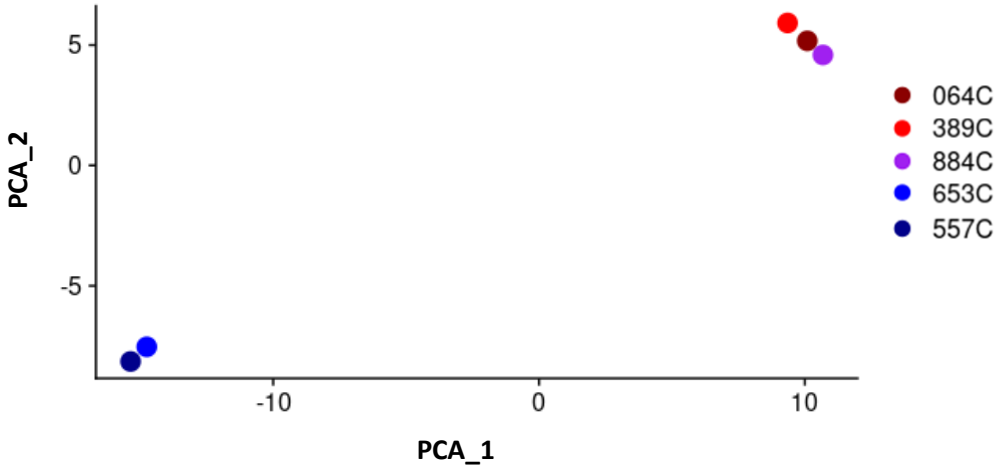


Figure 3.3. PCA plot of bulk sequencing of CRC organoids. Total TPM counts per gene normalised to account for sequencing depth bias differences (put in values of normalisation). Coloured by TSS phenotype, 064 389 coloured Red shades to represent high YC, 884 purple to indicate moderate to no dramatic TSS shifting and 653 and 557 coloured blue to represent high YR TSS phenotype

I next set out to understand the gene ontology (GO) underpinning the expression profiles through comparing the gene expression of key pathways and functional features against the average expression of randomly selected bins of genes. Gene sets were obtained from the GSEA molecular signatures database (Subramanian et al., 2005). Module scoring generates a DGE like analysis to determine whether the expression of a predefined set of genes of interest are more up or downregulated when compared to random gene selection. Bins of 200 random genes were used with the aggregated expression of these genes subtracted from the average expression of the interest set. A positive score indicates the genes are

more upregulated, negative delimits downregulated and close to 0 represents the gene set is similar to the average expression.

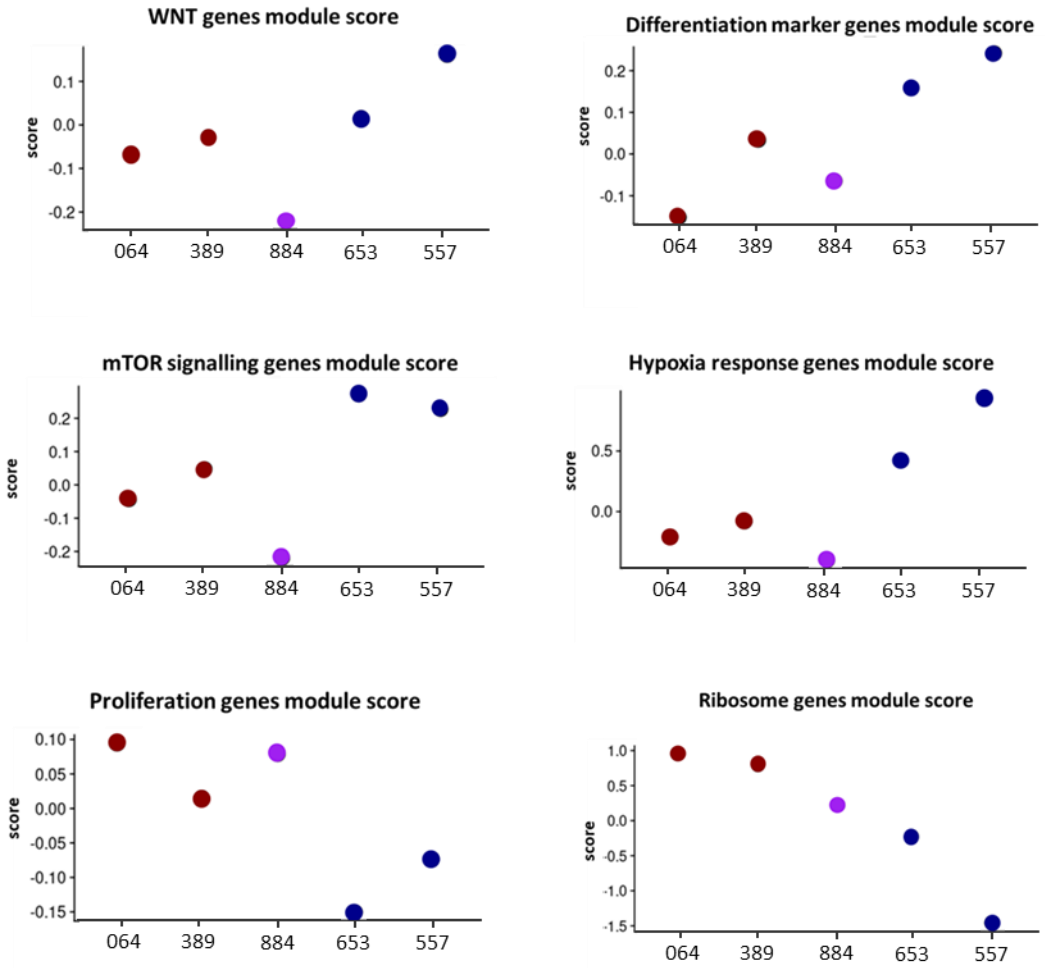


Figure 3.4. YC enrichment correlates with upregulation of specific functional pathways. Gene set enrichment analysis in Bulk sequenced organoids. Module score of gene sets compared against average binned gene expression. Top panel gene sets associated with differentiation, middle translation and metabolism and bottom growth. Dots coloured TSS global phenotype, red = high YC, purple = moderate YC and blue = low YC.

Comparing the scores of gene sets previously identified to be functionally relevant in driving cancer phenotypes between samples, shows distinct correlations and hence similarities between lines with the same TSS phenotype. Previous studies have highlighted the fundamental role WNT and its downstream signalling network play in establishing and maintaining homeostasis of intestinal morphology (Krausova & Korinek, 2014; Mah et al., 2016; Vermeulen et al., 2010). Unsurprisingly, the two more differentiated lines which imaging analysis in **Figure 3.4** showed crypt-like structure formation displayed higher WNT gene scores when compared to the more undifferentiated lines. Further supported by the same

observed correlations seen in the differentiation marker gene module score comparisons. The same trend was also seen whereby high YR scored more highly when analysing mTOR signalling genes and hypoxia response. The latter gene set encompasses many genes previously identified to be essential in reprogramming metabolic state and hence advanced tumour formation (Eales et al., 2016; C. Yang et al., 2014). High YC organoids scored more highly when ribosomal genes were assessed, having previously established that ribosomal and translation associated genes are YC dominant and their subsequent expression is often highly correlated with growth, our hypothesis that the high YC organoids would score more highly with these gene sets was confirmed. Furthermore, the gene expression of high YC organoids is also more primed to proliferation associated genes.

3.2.2.1 Characterising global archetypes

To identify more universal characteristics beyond just analysing a selected set of specific pathways, we next scored the lines for marker genes associated with the universal archetypes previously established by Hausser et al (Hausser et al., 2019). Previous studies highlighted tumour behaviours and molecular phenotypes repeatedly showed distinct archetype profiles and thus provide a formative way of linking the tumour genotype-phenotype map (Hausser et al., 2019; Hausser & Alon, 2020). Stratification of our organoids using the archetype module scoring revealed similar trends seen in the literature where the same analysis was performed on end state tumour samples (**Figure 3.5a)** (Hausser et al., 2019; Heiser et al., 2012). The fast-proliferating undifferentiated organoids show a less dominant archetype profile with the highest expressed genes emanating from the proliferative archetype (**Figure 3.5b**). A dominant invasion and tissue remodelling archetype profile was seen in the two slower growing differentiated lines (**Figure 3.5b**). This feature commonly associated with a more progressive resistant disease was found to have lower expression in the responsive lines. The most resistant line 557, showed an increase in genes associated with the biomass and energy archetype, many of the genes underpinning this molecular feature are known to be vital in metabolic reprogramming and hence aid tumour survival in stressed conditions.

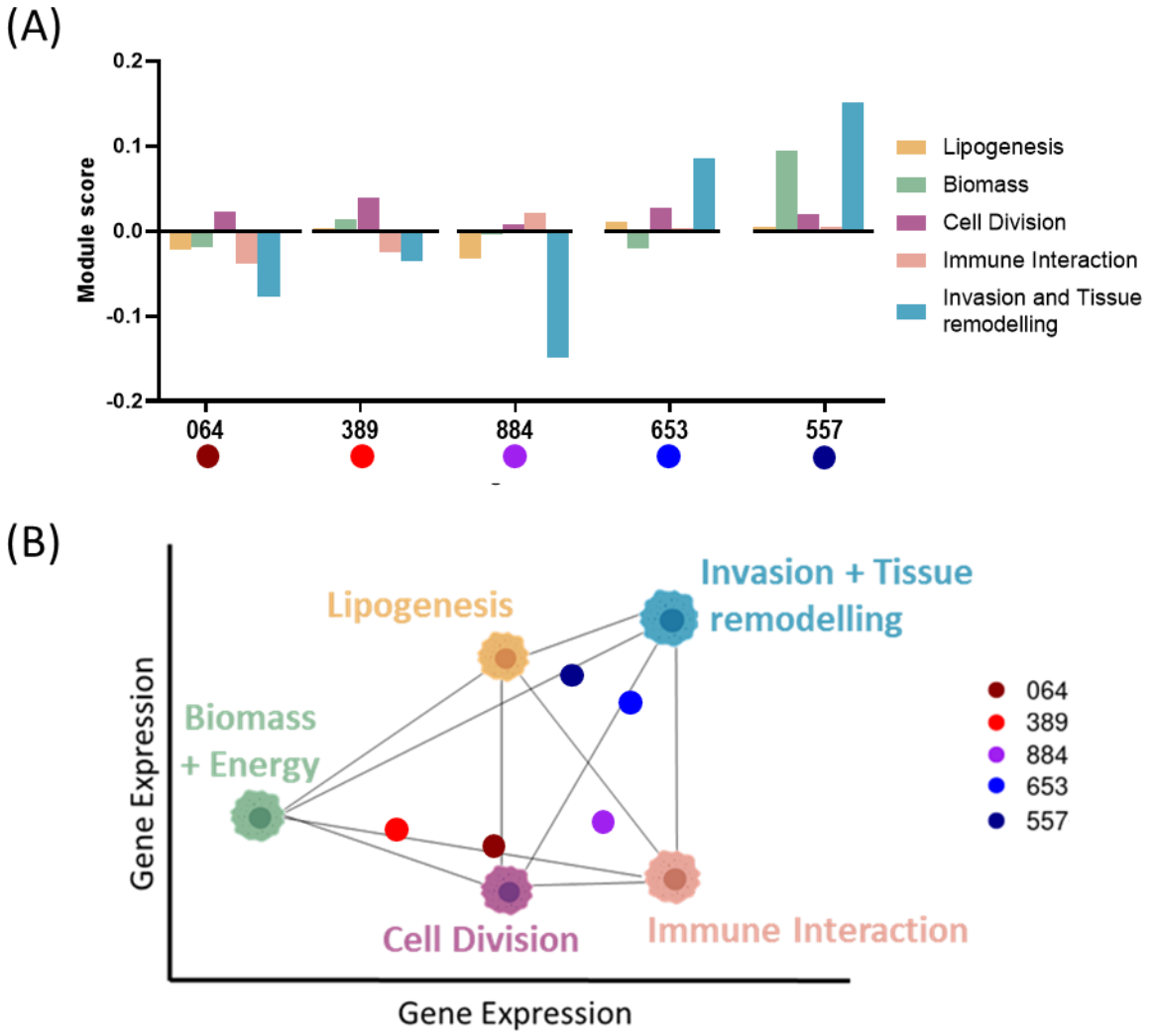


Figure 3.5. Stratification of organoids using universal archetype referencing. (A) Bar chart of module scoring for the 5 universal archetypes represented in right legend panel in each organoid line. Positive value indicates upregulated average expression compared to randomly assigned gene pines, negative shows downregulation. (B) Projection of each organoid positioning onto universal archetype polyhedral coloured by YC enrichment. The YC enriched organoids plot within the cell division and metabolic associated region. The two YC low plot closer to the defined late-stage archetype invasion and tissue remodelling.

Utilising the reference states and the accompanying growing literature of the clinical phenotypes associated with archetype profiling, I have further confirmed organoid success in recapitulating tumour behaviours. Furthermore, this characterisation provided insight into global phenotype associated with each organoid and helped inform which organoids to choose for following analysis.

3.2.3 Methodology for single cell CAGE experiment

Our current understanding of molecular features and hence biological relevance of TSS

usage in cancer states is vastly limited due to the bias accompanying bulk analysis that greatly masks the intra variability that exists within a tumour. Many previous studies have highlighted the sheer variability that exists within a tumour and through the advent of single cell technologies has highlighted the complex SC populations that behave differently but collectively contribute to tumour function (Caiado et al., 2016; Marusyk et al., 2012). Therefore, to truly understand the molecular phenotypes that underpin high YC organoids we aimed to perform single cell CAGE analysis. As depicted in **Figure 3.6** our experimental workflow would allow us to elucidate firstly the cellular physiological states that make up our organoids, providing us with a better understanding of the biological relevance of observed TSS profiles. Is it simply that high YC organoids harbour a very distinct population of cells hence responsible for the YC characteristic or is this a global characteristic not associated with state and hence gene expression? Secondly, how heterogeneous are the organoids and is the global TSS profile pervasive through all populations?

In order to allow for sufficient YC comparisons between organoids, we picked two opposing organoids. Firstly, the YC enriched 389 organoids with fast proliferating phenotype (**Figure 3.2b)** showing cell division archetype with little differentiation (**Figure 3.5a**). Secondly, the YC low 557 organoids, showing extensive differentiation in both imaging (**Figure 3.2a**) and late stage archetype presentation through enrichment of invasion cellular states (**Figure 3.5a**). Although single cell 5' CAGE sequencing had recently been developed and performed by Moody et al., 2022 the experimental aims of this paper focused on other CRE rather than TSS calling meaning no study had yet explored TSS variation at a single cell resolution. To explore and characterise ITH of both physiological state and TSS usage I aimed to adapt the methodology outlined in Moody et al., 2022 to accommodate single cell preparation of organoids and successful capture of 5'ends. The experimental design outlined in **Figure 3.6** highlights each step required to successfully perform this experiment resulting from multiple rounds of pre sequencing optimisation to ensure each step did not induce cellular stress. The full methodology is outlined in **Chapter 2.4.1**.

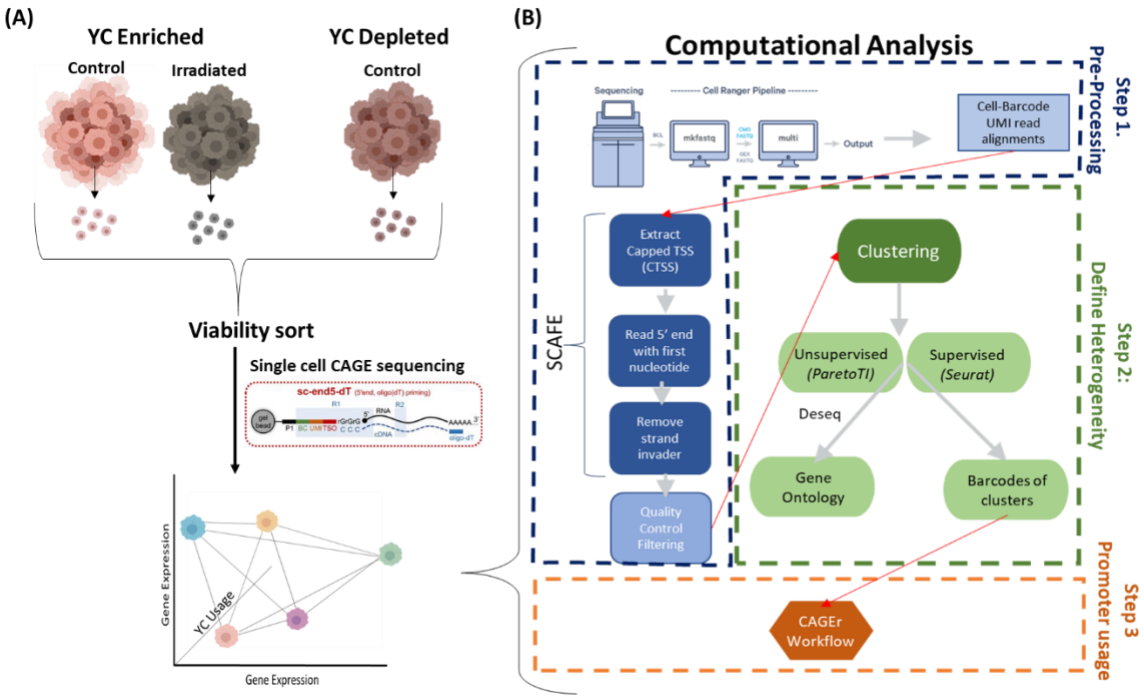


Figure 3.6. Experimental workflow of Single Cell CAGE of CRC. Schematic representation of experimental design of single cell experiment from sample and library preparation to computations analysis for ITH and TSS investigations.

Firstly, the single cell sequencing approach relies on live cells upon barcoding, therefore high cellular viability was essential. As dead or dying cells will contain large amounts of degraded RNA and degradation predominantly begins at the 5' end of transcripts, captured transcripts may result in incorrect TSS mapping reducing robustness and representation of correct 5'ends (Houseley & Tollervey, 2009). To ensure high viability following disassociation to single cell, organoids were stained with a Zombie NIR (Biolegend) viability dye that only penetrated the cellular walls of dead or dying cells. This allowed us to gate and hence sort out the population of viable cells using flow cytometry based on cells not fluorescing outlined in **Chapter 2 Figure 2.3**. The flow sort was done at the lowest speed, collected, washed and sorted in PBS BSA to prevent cellular adhesions and the collection tube was held directly below the expelling nozzle. All conditions including the sorting were kept at 4°c. Each of these steps were essential to prevent post sorting stress as previous attempts not using these steps showed decrease in viability assessed using a Tryphan blue viability count post sorting. The experiment only progressed to library preparation when viability was above 80%. Each of the lines were cultured in the same conditions as the previous bulk sequencing

to allow for comparisons and hence sequenced at 14 days PS. We determined that 14 days provided sufficient growth through presence of secondary structures indicative of differentiation. Furthermore, BME has shown to begin degrading post 14 days and thus may affect organoid growth (Kozlowski et al., 2021; Z. Zhao et al., 2022).

Following cell ranger filtering analysis outlined in **Chapter 2.5.1.1**, that removes background sources of sequencing reads not resulting from an actual cell, we successfully achieved single cell sequencing of both control samples. 6283 cells were captured at a mean read depth of 289,401 with median gene coverage of 4,859 per cell for YC depleted 557 line as depicted in **Table 3.1**. The control sample of the YC enriched 389 line had a higher number of cells at 7,099 and lower number of reads per cell (266,347) however a larger number of median genes captured.

Table 3.1: Single cell sequencing library statistics

Sample	Number of Cells	Mean Reads (per cell)	Median Genes (per cell)	Number of cells (post filtering)
557 control (YC low)	6283	289,401	4859	5506
389 control (YC enriched)	7099	266,347	5634	5182
389 Irradiated (YC low)	647	71,080	817	605

Previous work by Dr Wragg in the lab showed a drastic loss in YC enrichment post irradiation of the YC enriched lines such as 389 (Wragg et al., 2023). I aimed to explore whether this was simply the result of the loss of a specific population of cells or whether shifting in TSS usage occurred to mediate cellular survival. However, performing single cell CAGE on post irradiated samples was exceptionally challenging as desired viability was initially not met even post sorting. Only 5% of cells from organoid 389 survive post irradiation and the initial workflow appeared to inflict further stress on the surviving cells meaning the sequencing could not proceed. However, final optimisation that incorporated the above steps allowed us to collect viable cells and thus include these in the sequencing at a much lower depth as depicted in **Table 3.1**, although this may not allow for direct TSS calling on a single cell level it would still provide insight into the SC gene expression changes that occur post irradiation.

3.2.4 Evaluation and validation of single cell experiment success

I aimed to assess the success of the single cell sequencing and hence quality of data produced initially by analysing the distribution of established quality control metrics in both samples. Firstly, molecule count, defined as number of molecules or UMI detected within a cell which indicates sequencing reads. Secondly feature count, defined as number of genes detected per cell were plotted to explore the distribution between and within the samples (**Figure 3.7a**). Filtering for cells containing less than 200 genes and 10 UMI is the pre-defined standard as these often indicate background or empty droplets (Ilicic et al., 2016). These feature plots also indicated a significant bimodal split of data in both samples. To rule out that this split in data was resulting from the existence of multiplets, the presence of more than one cell per droplet, we ran the data through the Scrublet pipeline (Wolock et al., 2019). Using this pipeline outlined in **Chapter 2.5.2** we established a low multiplet rate of 4.6% for sample 389 and 6.8% for sample 557, cells identified as likely multiplet were discarded, however it did not disrupt the bimodal split. It has recently been highlighted the impact ambient RNA can have on single cell data analysis, with both quality control metrics and even cellular state determination being affected thus resulting in inaccurate results (Floriddia, 2022). To remove the contamination of these freely floating and hence irrelevant transcripts we used the DecontX workflow and removed the cells however we found this was also not impacting the split (S. Yang et al., 2020). Further analysis into this bimodal split of populations is needed as such finding was not seen in literature. However, given that conventional RNA sequencing is performed at a much lower depth, this could be a result of sheer disparity between a cell state known to significantly decrease transcription or could result from a technical bias not previously found in the data.

A significant positive correlation is seen in both samples when number of genes against number of UMI detected per cell is plotted thus indicating reads are sufficiently distributed throughout the cells per gene **(Figure 3.7b)** (Ilicic et al., 2016). Although our library preparation cDNA profiles indicated high quality RNA, we further validated the absence of stressed, dead or dying cells through filtering of out those whereby % of reads mapping to mitochondrial DNA was above 10% (**Figure 3.7c**), a threshold recently established most appropriate for single cell human analysis (Osorio & Cai, 2021).

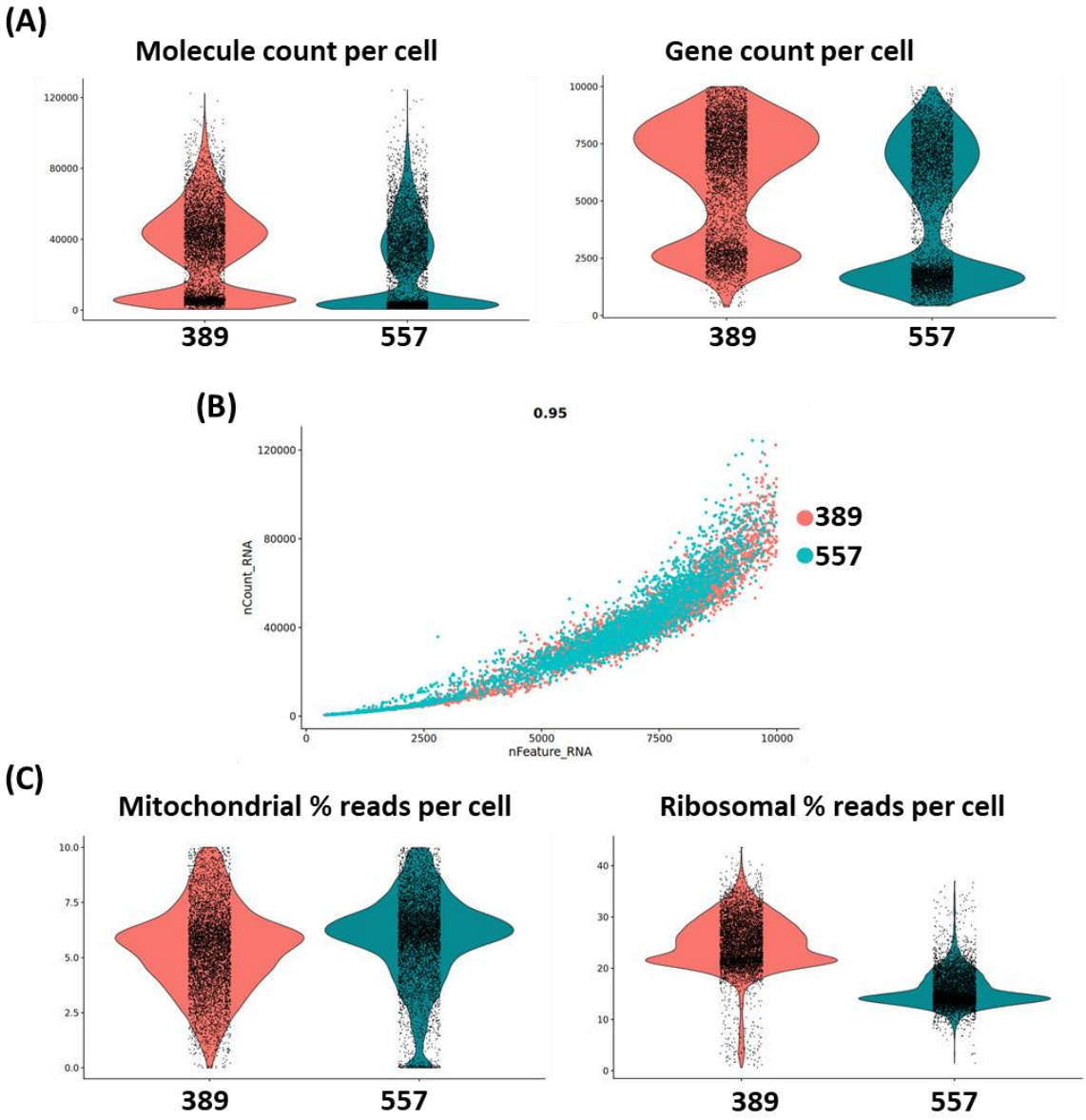


Figure 3.7. Feature plots evaluating quality of sequencing per single cell. Coloured by sample of origin, red indicates sample 389 YC enriched line control, blue is 557 YC depleted control line **(A)** Data distribution per cell. Total number of molecules in each individual cell (left) Number of genes detected in each cell (right). **(B)** Correlation analysis of per cell molecule count vs Gene count, Pearsons coefficient =0.95 **(C)** Read mapping per cell. Percentage of reads that map to the mitochondrial genome post 10% filtering (left). Percentage of reads that map to ribosomal genes (right).

The Feature plots outlined in Figure 3.7 show similarity between the two samples however a shift in % of ribosomal genes in the YC enriched 389 sample was seen (**Figure3.7c**). Although this indicated ribosomal genes are more expressed in this line compared to the YC depleted 557, also supported from bulk analysis highlighted in **Figure 3.4**, we cannot rule out that this is resulting from a technical bias. Furthermore, research shows ribosomal genes are often YC enriched, therefore downstream analysis must take into consideration this bias

when analysing YC associated trends (Cockman et al., 2020; Hochstoeger & Chao, 2024).

3.2.4.1 Validation to pre-existing data

To explore complexity and representation of our single cell data, I compared expression profiles to the previously generated bulk data. To achieve this, single cell data was pseudo bulked per sample and the counts were normalised across single cell and bulk samples using a power law distribution supported in the CAGEr workflow that allows for disparities between library sizes (Balwierz et al., 2009; Haberle et al., 2015). Reverse cumulative plots were generated depicted in **Figure 3.8a**, and optimal parameters for power law normalisation determined as $\alpha = 1.05$ fit in range of tags 10 to 15,000. Gene by Gene TPM count matrices were then generated and principal component analysis revealed similar variance firstly existing between samples in both single cell and matched bulk (**Figure 3.8b**). Interestingly, a similar variance between experimental conditions, the variation between single cell 389 and single cell 557 was the same degree of variance seen between the matched bulk, indicating likely PC2 variation highlights sample differences whereas PC1 variance accounts for sequencing and experimental differences.

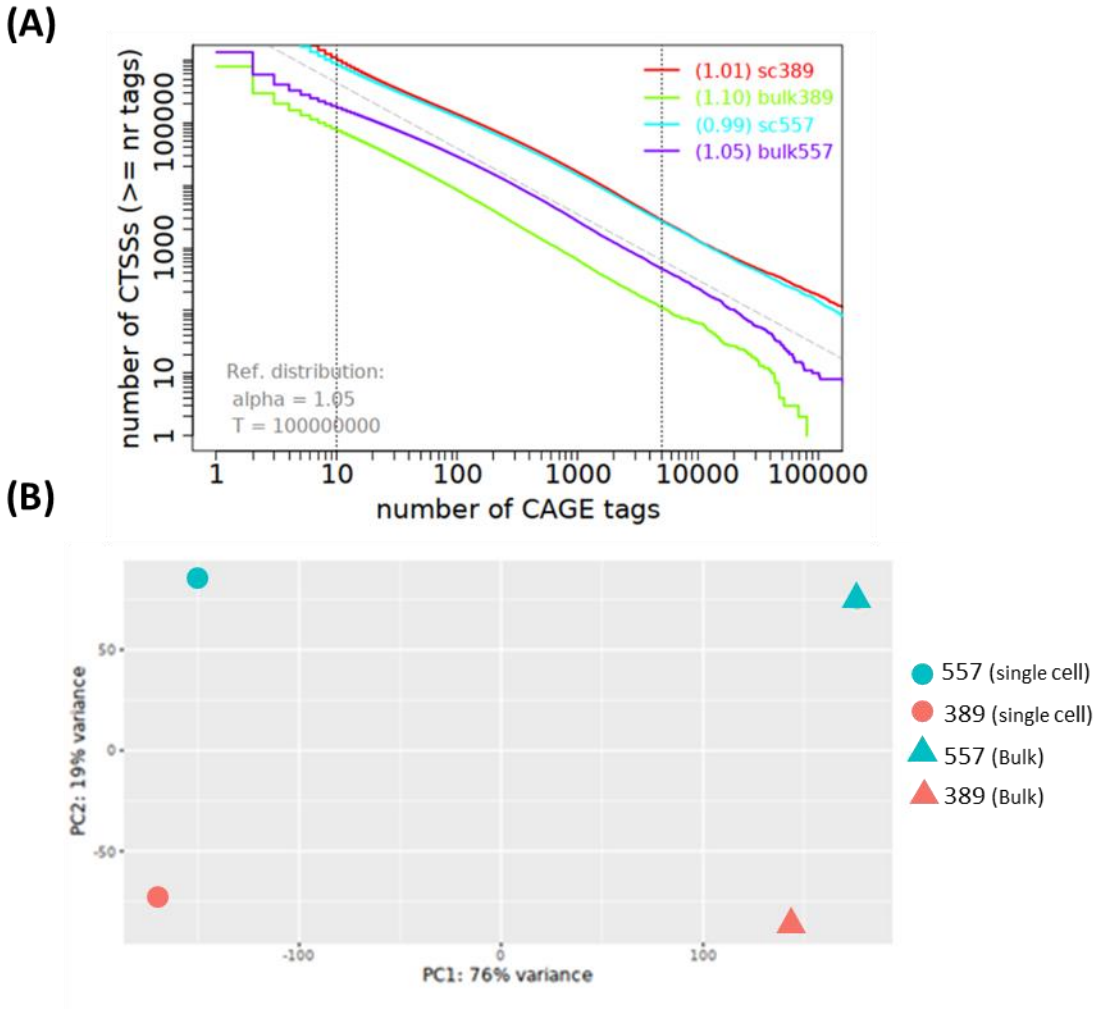


Figure 3.8. Validation of expression profile retention between single cell and bulk CAGE sequencing. (A) Reverse cumulative of raw TPM counts per single cell sample and matched bulk. (B) PCA comparison of bulk CAGE sequenced samples and single cell pseudo bulked. Coloured by sample of origin, shaped by experimental design.

To further compare, log transformed TPM counts per gene from single cell against matched bulk were compared (**Figure 3.9**). Linear regression analysis revealed in both samples a significant positive correlation R= 0.77 and 0.72 for non-responsive and responsive respectively. Thus, indicating similar gene expression retention between experimental conditions. Furthermore, in both single cell the top 50 highly expressed genes were predominantly ribosomal genes, removal of these improved Pearsons's coefficient value. In both samples we see a population of genes captured only in the single cell. Both plots also show genes captured in single cell that are not captured in bulk, and vice versa. Although this can be a result of increased depth in the single cell increasing chance of capturing low levelled genes, the different sequencing approaches, outlined in **Chapter 1.4.5**, can also introduce technical bias favouring certain RNA capture.

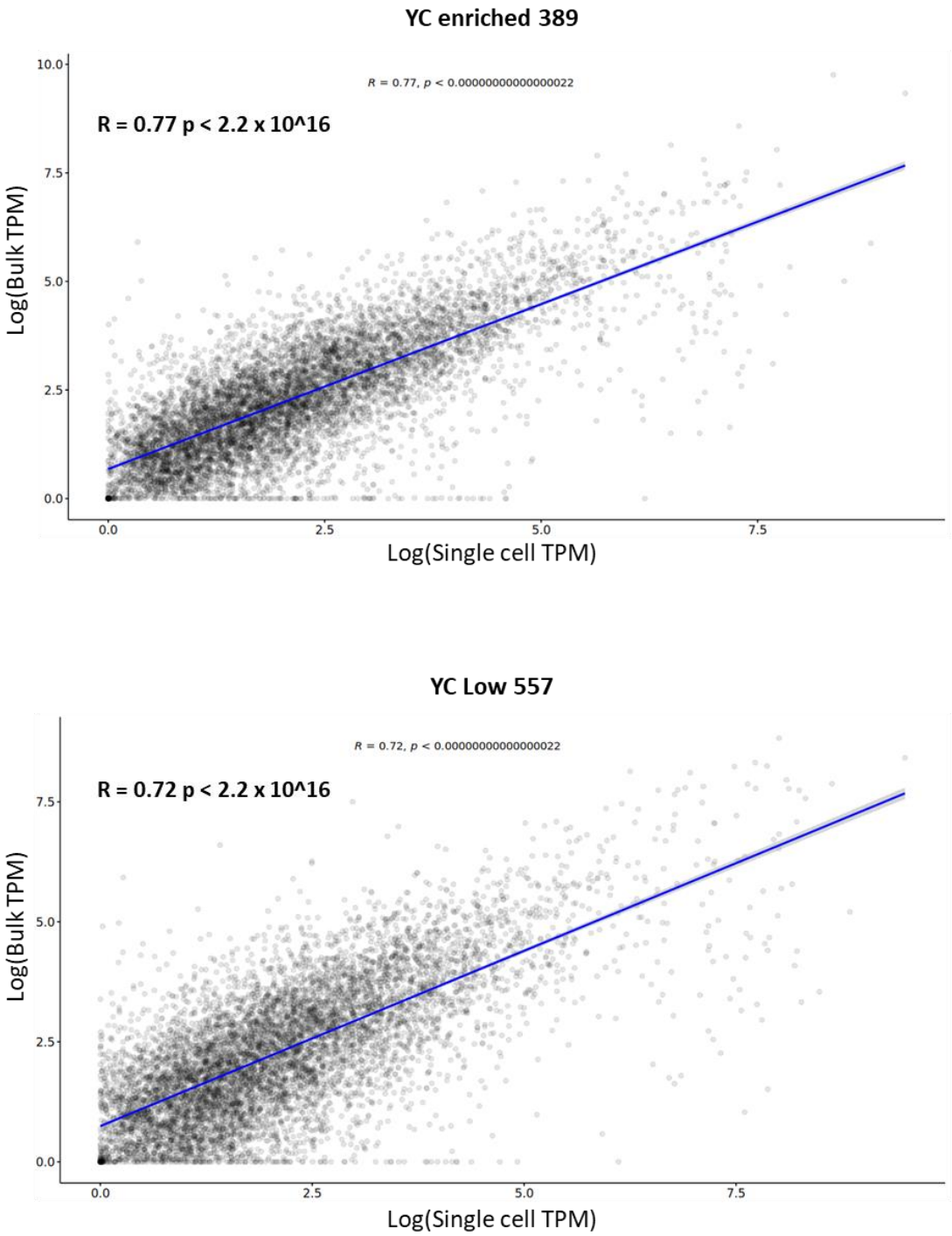


Figure 3.9. Gene expression profiles are retained between bulk and single cell sequencing. Gene by gene TPM count Scatter plot of single cell vs matched bulk. Log transformed TPM counts per gene by sample of origin. X axis represents single cell whereas Y is bulk. Top shows YC enriched 389 single cell vs matched bulk. Bottom shows YC low 557 single cell vs matched bulk.

3.2.5 Molecular Phenotyping of high YC organoids using single cell analysis

3.2.5.1 Intra heterogeneity analysis of organoids

As the bulk data indicated interesting phenotypic differences such as increased

differentiation and metabolic pathway genes in YC low lines, I aimed to investigate whether similar could also be observed in the single cell. Initial gene set enrichment analysis was performed and showed some different observations seen in the bulk data. Violin plots in **Figure 3.10** show that sample 389 now on average scores higher than 557 in the mTOR pathway genes however the opposite trend is observed in the bulk. 389 is characteristically much more proliferative and hence growth was driven meaning higher mTOR is expected, the lack of complexity associated to the bulk data may not capture this. The same trend is seen in the other growth associated pathway MYC. Changes to the increase in hypoxia genes is likely a result of the responsiveness of sample 389 and the single cell workflow could be inducing expression changes to accommodate stress generated. However, the WNT signalling analysis showed that 557 single cell was still higher than 389.

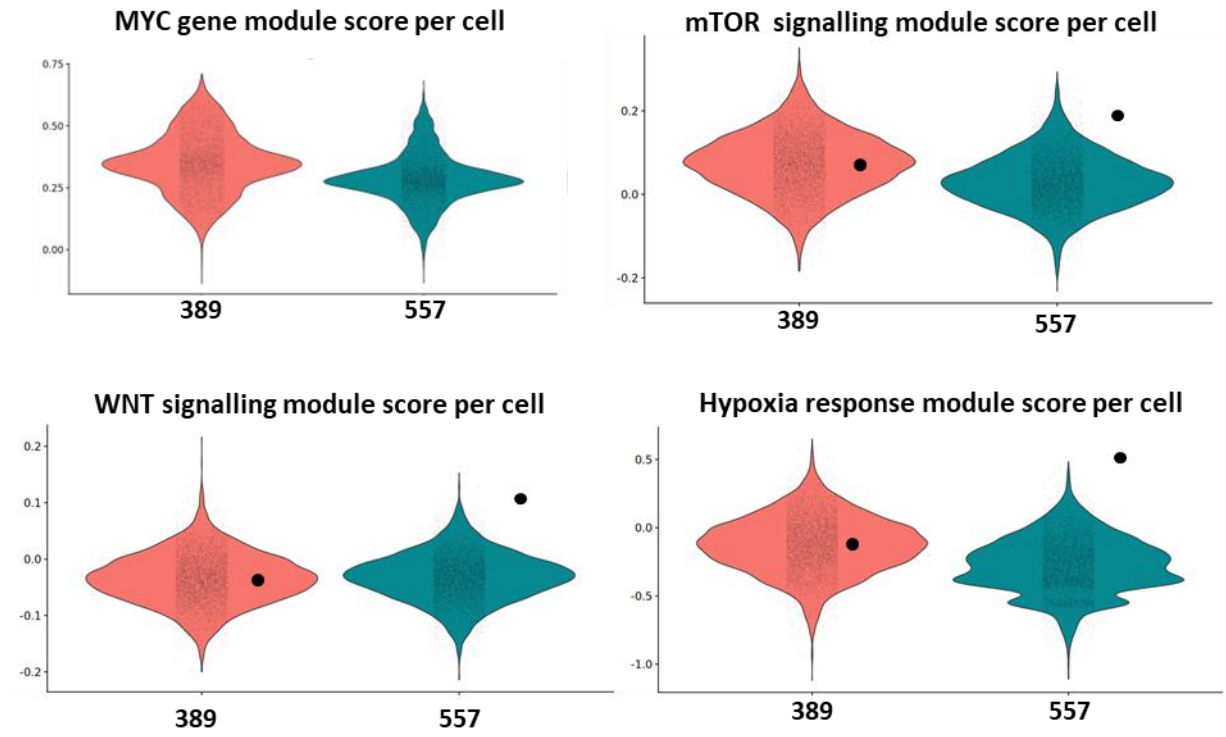


Figure 3.10. Gene set enrichment analysis in single sequenced organoids per cell. Coloured by sample of origin, red indicates YC enriched 389 sample control, blue is YC low 557 control sample. Black dots represent the previous matched sample bulk CAGE module scoring for that pathway **(A)** MYC pathway genes **(B)** mTOR pathway genes **(C)** Wnt signalling genes **(D)** Hypoxia associated genes.

To identify gene expression associated with high YC usage, DGE between the YC enriched and YC low lines was performed. The two single cell data sets were integrated and normalised using SCtransform V2 as outlined in **Chapter 2.5.3**, this method utilises the Pearsons residuals

approach meaning better retention of biologically variable and meaningful genes are kept even if not highly expressed when compared to conventional log normalisation (Lause et al., 2021). Counts were aggregated per line and differentially expressed genes (DEG) were identified as those with log fold change above or below +/- 0.6 with adjusted p value >0.05, the distribution of DEG of sample 389 vs 557 represented in **Figure 3.11**. This analysis highlighted how transcriptionally different the samples were and with more genes upregulated in 389 an indicative feature of fast proliferation (Bowry et al., 2021). Upregulated genes in the YC low organoid show a dominance for functions involved in regulating and guiding differentiation. For example, both *HOXC18* and *SOX8* have recently shown to be upregulated in more differentiated CRC and may potentially act as a prognostic marker for more advanced poorer outcome cancers (Bhatlekar et al., 2018; Y. Wang et al., 2019; J. Yu et al., 2022). The most significantly upregulated in YC low was the *CD69* gene shown to play a significant involvement in metabolic reprogramming and is direct target of HIF1α (Cibrián & Sánchez-Madrid, 2017). Conversely upregulated genes in the YC enriched organoid appear to have distinctive roles in promoting proliferation (Jiang & Liu, 2015; Mao et al., 2021).

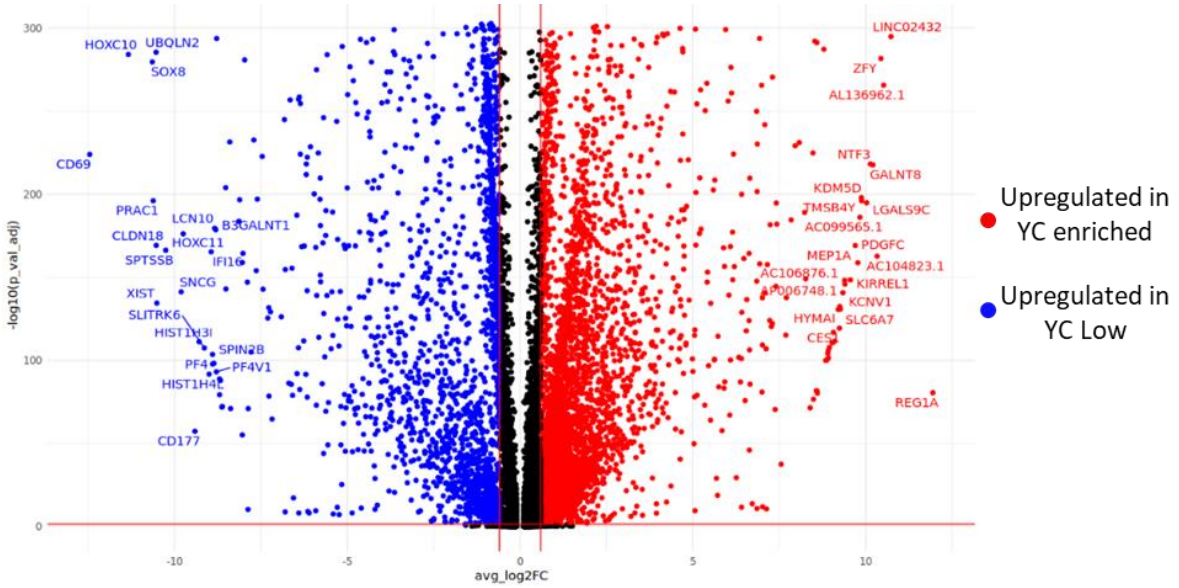


Figure 3.11. Differential Gene Expression Analysis of YC enriched organoid vs YC low organoid. Volcano plot of average fold change against log transformed p adjusted value of differentially expressed genes of YC enriched organoids (389) vs YC low (557). Up regulated genes shown in red, down regulated in blue, those insignificantly unchanged in black.

The violin plots in **Figure 3.10** also showed ITH of pathway enrichment within the samples highlighting that the global phenotypes may not represent all cells within the organoid. I aimed to characterise ITH and hence identify the identified sub populations with potentially

different physiological states that exist within our organoids. Clustering analysis was performed using Seurats SCTransform V2, utilising Pearsons's residuals negative binomial distribution analysis to identify gene markers and set cluster identity. Dimensionality reduction projections of the clusters was performed and represented in **Figure 3.12** and **3.13**. To identify physiological state or phenotype DGE analysis was performed on each cluster and GO was analysed using the identified DEGs, using all genes expressed in the line as background in ShinyGo8.0. The dominant GO was selected per cluster and annotated as cluster identity.

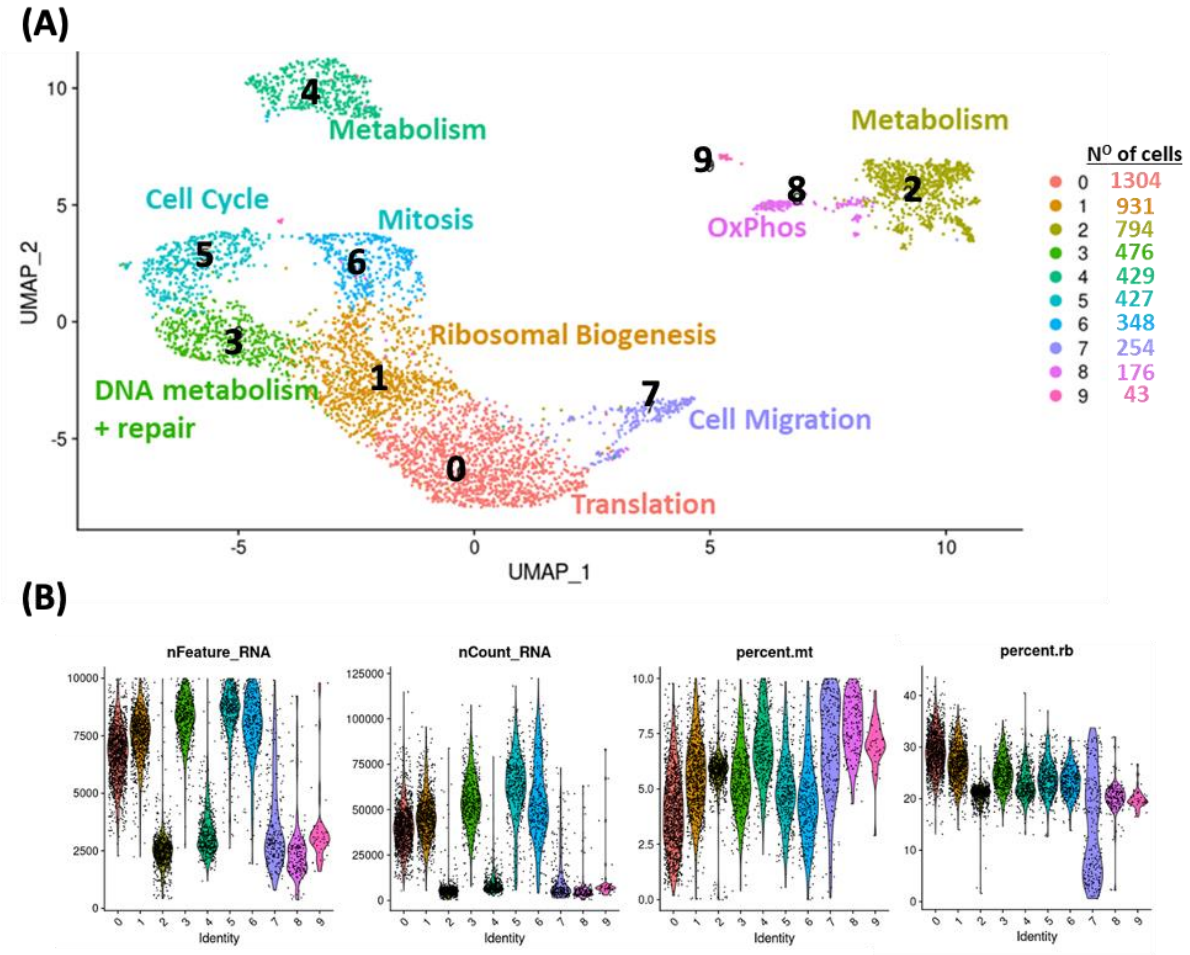


Figure 3.12: Identification of subpopulations in YC enriched organoid (389). (A) UMAP projection of clusters with annotated identity achieved through Wilcoxon Rank sum testing and pathway enrichment and gene network fold enrichment analysis of cluster associated marker genes. Number of cells per cluster highlighted in far-right panel. (B) Violin feature plots grouped by cluster showing per single cell feature count, UMI count, % of reads mapped to mitochondria and ribosomal genes.

Analysis showed the dominant cluster was translation in 389 followed by ribosomal biogenesis **Figure 3.12**, many genes associated with these regulatory networks are involved

in growth and are characteristically associated with YC or TOP TSS usage. The clusters also varied in feature count (genes captured), a likely explanation for some of the smaller cluster separations. Two metabolic populations, cluster 2 and 4, were identified with both showing less gene and read counts compared to the other clusters, a similar cluster was identified in sample 557 **Figure 3.13** (cluster 0), showing a marked decrease in gene count. Comparison of the enriched genes in these 3 metabolic associated clusters show similarity and association to pathways responsible for metabolic reprogramming. Under stress conditions, cells reprogramme many biological processes to aid in survival, nascent RNA sequencing has shown global down regulation of transcription also occurs (Vihervaara et al., 2017). However, stress response genes are upregulated many of which are associated with metabolic reprogramming. As we filtered out the highly stressed cells using mitochondrial DNA cut offs, we hypothesise these metabolic cluster could be functionally stressed cells potentially located in nutrient deprived regions of the organoid.

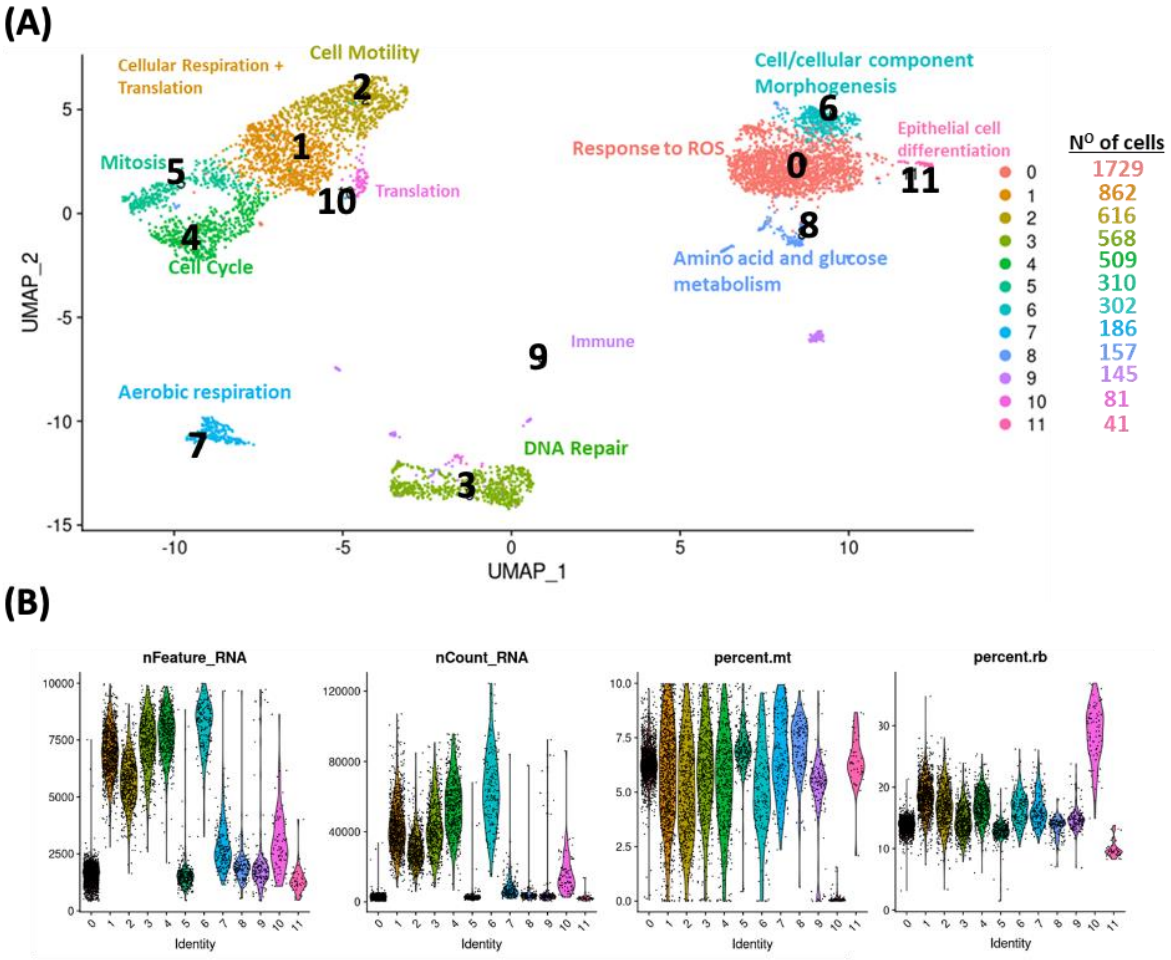


Figure 3.13: Identification of subpopulations in YC low organoid (557). (A) UMAP projection of clusters with annotated identity achieved through Wilcoxon Rank sum testing and pathway enrichment and gene network fold enrichment analysis of cluster associated marker genes. Number of cells per cluster highlighted in far-right panel. (B) Violin feature plots grouped by cluster showing per single cell feature count, UMI count, % of reads mapped to mitochondria and ribosomal genes**.**

The high YR 557 samples largest cluster was response to reactive oxygen species (ROS) with genes involved in pathways associated with metabolic reprogramming. Furthermore, this sample is more heterogeneous with more clusters and separation in clusters. Both these features are linked in cancer research with resistance to treatment hence their presence appears to align with the clinical phenotype of the organoid line. Although both samples show distinct phenotypic differences, similar clusters cellular were identified in both samples. Both had subpopulations defined by enriched cell cycle genes. Furthermore, these cell cycle clusters showed a distinctive circular pattern on the UMAP, potentially highlighting dynamic cell cycle progression.

3.2.5.2 Identification of cellular states represented by archetypes.

Recent research has indicated the limitations of UMAP projections to determine subpopulations in single cell data resulting from the variation that can be generated through slight alterations in the choice of parameters (Zhai et al., 2022). An alternative to UMAPs discussed in **Chapter 1.2.7** is archetypal analysis and was also pursued to explore ITH. Using MET to analyse the expression data would identify key cellular states with clearly defined expression profiles that exist within a sample, these are known are archetypes, all cells will sit within these archetypes dependent on how similar they are. Archetype analysis has the advantage of utilising a 3D approach preventing the artificial reduction of dimensions that result in us drawing arbitrary boundaries which may bias clustering approaches. Therefore, we hypothesised that archetypal analysis would provide useful insight into the continuum of cellular states that exist within our organoids.

Archetypal analysis may also aid in understanding the role of TSS usage in universally characterised states (archetypes) whereby cells have shown to switch or transition into different states in response to the environment. I identified that many of the marker genes underpinning the universal archetypes, identified by Hausser et al., 2019, harbour a large proportion of dual initiator genes as seen in **Figure 3.14**. Furthermore, many marker genes are shared between the universal archetypes with less than 20% of their dual initiator genes being unique to a specific archetype, highlighted in **Figure 3.14b**. As noted in Chapter 1, an important balance exists between cellular growth driving and function. A key regulation of this balance are the networks that underpin proliferative and metabolic phenotypes and the cellular transition from one to the other state to survive (Keibler et al., 2016; Schiliro & Firestein, 2021; J. Zhu & Thompson, 2019). We found that the largest shared dual initiator genes were between the cell division and biomass and energy archetypes. Our new hypothesis is that TSS shifting occurs in these dual genes to aid in transitioning archetype states from one to another aiding cellular function.

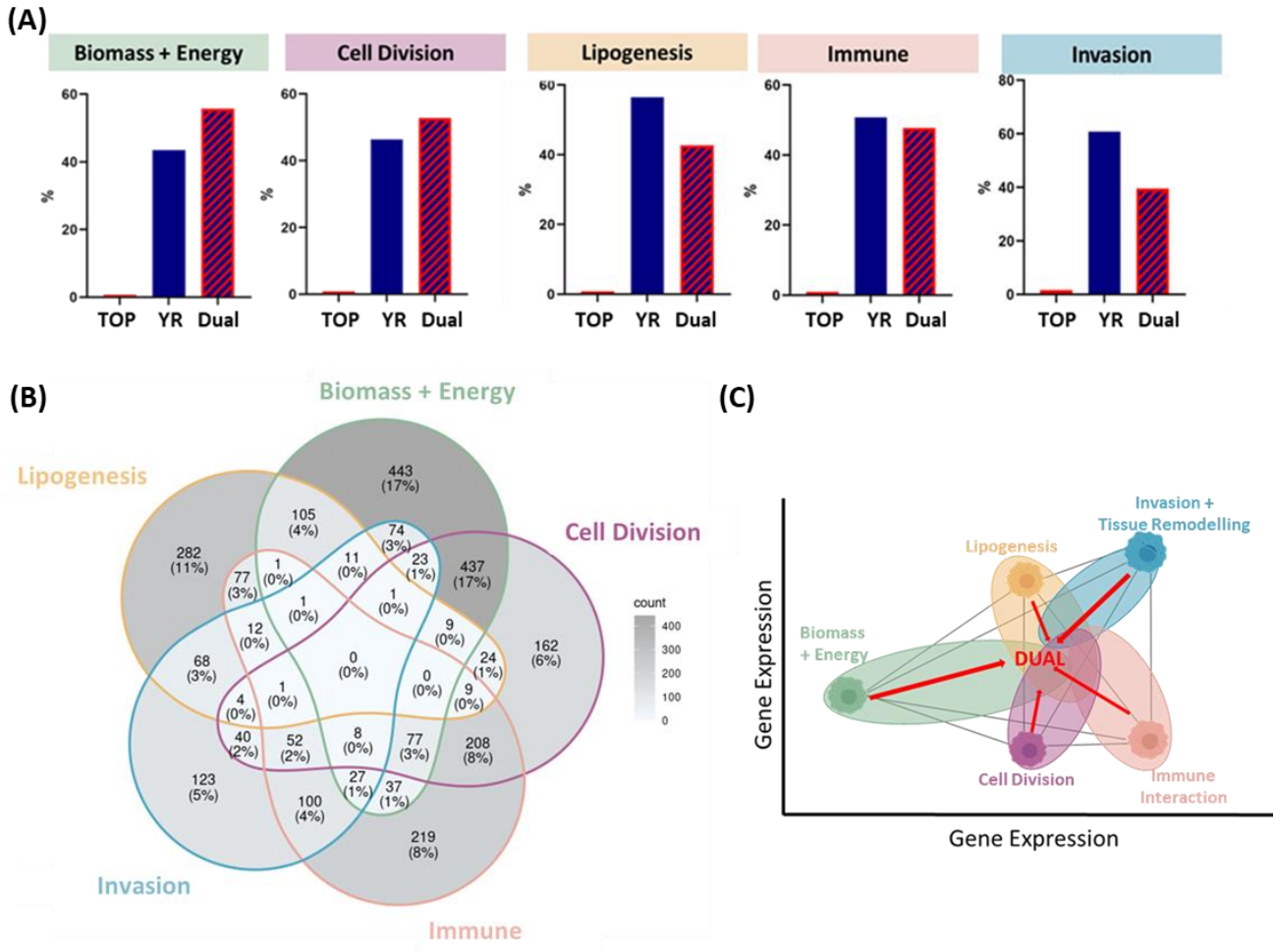


Figure 3.14: **Marker genes of universal archetypes harbour dual initiation.** (A) Bar chart showing percentage of the genes that make up the markers for each archetype that are Top only, Canonical (YR) only and are Dual. (B) Venn diagram of the dual initiator genes only for each universal archetype markers and their percentage sharedness across other archetypes. (C) Schematic representation of hypothesis overlaying shared Dual gene usage between archetypes.

To test our hypothesis and hence better understand the cellular exist in our organoids we ran the ParetoTI algorithm on our single cell sequencing data (Hart et al., 2015). To determine polyhedral distribution, I calculated the minimum simplex polytope that best encompasses our data, with the most distinctive and defined cells at the vertices. I then plotted our individual cells within the calculated polytope based on their gene expression likeness to the vertices, or archetypes. Both samples were required 5 archetypes (vertices of defined states) to encompass all data. The archetype distribution polytopes are shown in **Figure 3.15**, further validating that YC depleted 557, is more heterogeneous with distinct specialised populations compared to the high YC sample 389 validating our previous clustering analysis. We took the

marker genes for each of the vertices and performed GO analysis and characterised their likeness to the universal archetypes defined in literature by assigning each vertex to the universal archetype which it shares the most marker genes expression with **Figure 3.15**.

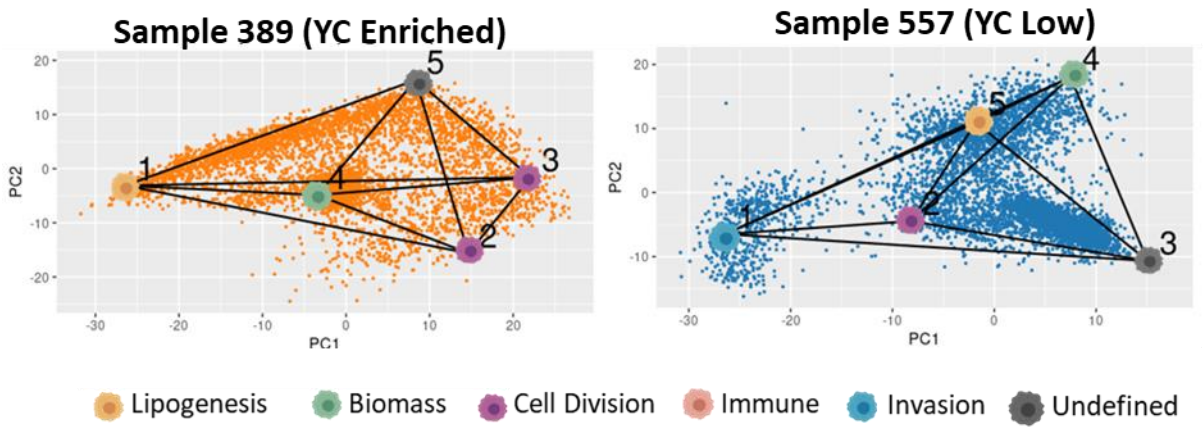


Figure 3.15: Single Cell Archetype distribution plots. Unsupervised characterisation of single cell using ParetoTI analysis. 2D scatterplot of polytope and annotated vertices (Archetypes) using dominant gene ontology for YC enriched sample 389 and for YC low sample 557.

As seen in the UMAP projections from **Figure 3.12** and **3.13** similar cellular clusters and hence some cell states exist between the samples this was further validated in the archetype analysis firstly in the individual sample comparisons showing shared defined universal states. To explore this further we integrated the datasets to find similarities in cellular states between the sample using the Harmony algorithm and normalisation using SC transform V2 (Korsunsky et al., 2019; Lause et al., 2021). The integrated dataset was first analysed using UMAP clustering (**Figure 3.16a**) and when the plot is split by sample of origin (**Figure 3.16b**) few clusters are shared between both samples indicating distinct differences between the two samples.

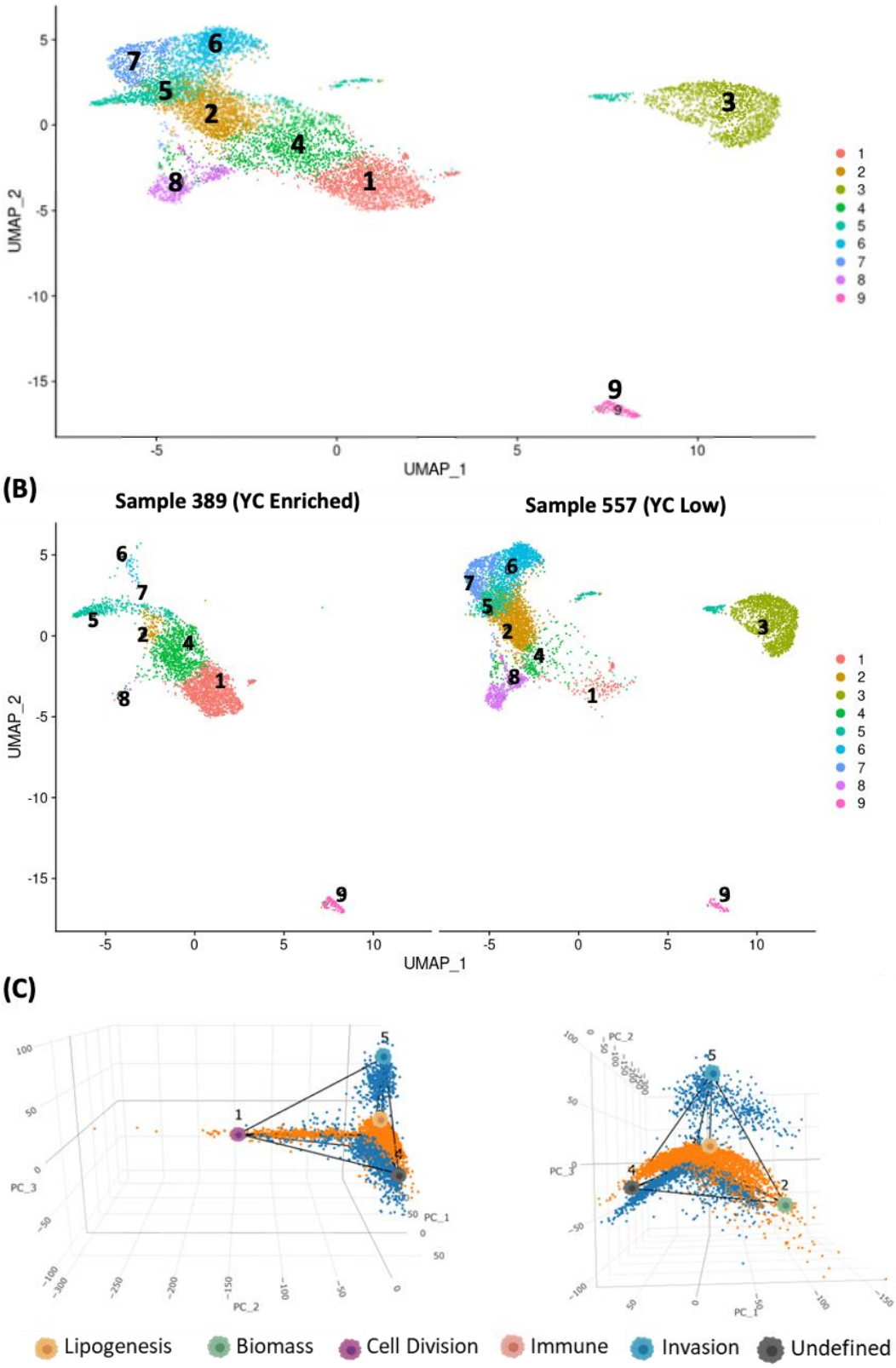


Figure 3.16: Archetype analysis shows similarities between samples missed in UMAP clustering. Integration of both single cell sequencing data sets. (A) UMAP clustering of integrated YC enriched and YC low samples. (B) Integrated UMAP split by sample of origin showing few shared clusters populated by both samples. (C) 3D projection of archetype polytopes for integrated dataset containing both 389 (Orange) and 557 (blue) control samples shown from 2 different perspectives. Associated universal archetypes annotated.

However, archetype analysis of the integrated data (**Figure 3.16c**) shows both have cells located near the same vertices or defined cell state. Their dispersion to the specialised cell differed, indicating a shared similar universal state with some disparities in the underlying expression profile. Such low-resolution analysis is missed in UMAP projections. Archetype 5 in **Figure 3.16c** shows a vertex and hence cell state only occupied by cells from the YC low 557. GO analysis of the archetype enriched genes revealed a high degree of similarity to the invasion and tissue remodelling universal archetype. As highlighted in literature the invasion archetype is often associated with aggressive and differentiated phenotypes, thus further supporting our previous findings in this chapter (**Figure 3.2 and Figure 3.10**).

3.2.5.3 ITH analysis of intestinal cell types.

So far, our analysis has been focused on cellular states that underpin the high YC phenotype to identify a potential cellular feature associated with YC usage. Previous research in zebrafish showed promoter architecture changes as cells differentiate away from pluripotency (Wragg et al., 2020). The imaging analysis in **Figure 3.2** indicated distinct morphological differences between our lines with literature highlighting correlations to different cellular functions from the different host organ cell types (Fujii et al., 2016; Kashfi et al., 2018). To investigate whether cell type and hence differentiation could also be associated with YC usage I generated an integrative cell type reference UMAP using the raw preexisting adult gut cell atlas generated by the Teichmann lab (Zilbauer et al., 2023). Firstly, using the full single cell RNA sequencing of 428K intestinal cells to establish the broad cell types, this revealed both samples only harboured cells of epithelial lineage. To characterise the cell types, I generated a cell type annotation map using the epithelium lineage data using Seurat following the same processing outlined in (Zilbauer et al., 2023). Following this I integrated the annotation map with our datasets using anchor based CCA integration to calculate anchors (defined cells with distinctive gene expression markers) outlined in **Chapter 2.5.3.3**. The data is then transferred onto our query datasets to establish prediction scores for cell types and hence identity established through highest prediction score.

Our analysis showed that the YC enriched organoid line is composed of mainly naïve stem like cell populations, the predominant cell type annotation is intestinal stem cells followed by transit amplifying cells (**Figure 3.17**). The latter have been widely documented in literature as

being the daughter cells of stem cells and function predominantly to continuously proliferate to eventually produce more differentiated progeny under altered signalling including WNT (Mah et al., 2016; Vermeulen et al., 2010; Walter et al., 2022). A small population of differentiated cells do exist, mainly the secretory goblet cells, likely explaining the cystic morphology these organoids display. Contrastingly the YC low sample 557 is composed of more differentiated cells (**Figure 3.17**) This supported our previous imaging analysis in **Figure 3.2** that showed intricate intestinal lumen morphology and high differentiation markers (**Figure 3.4**). Although the sample does not contain identifiable stem cell population it does have a large population of the naïve transit amplifying cells likely responsible for maintaining the organoid growth. This analysis suggests a TSS choice association with cellular differentiation state.

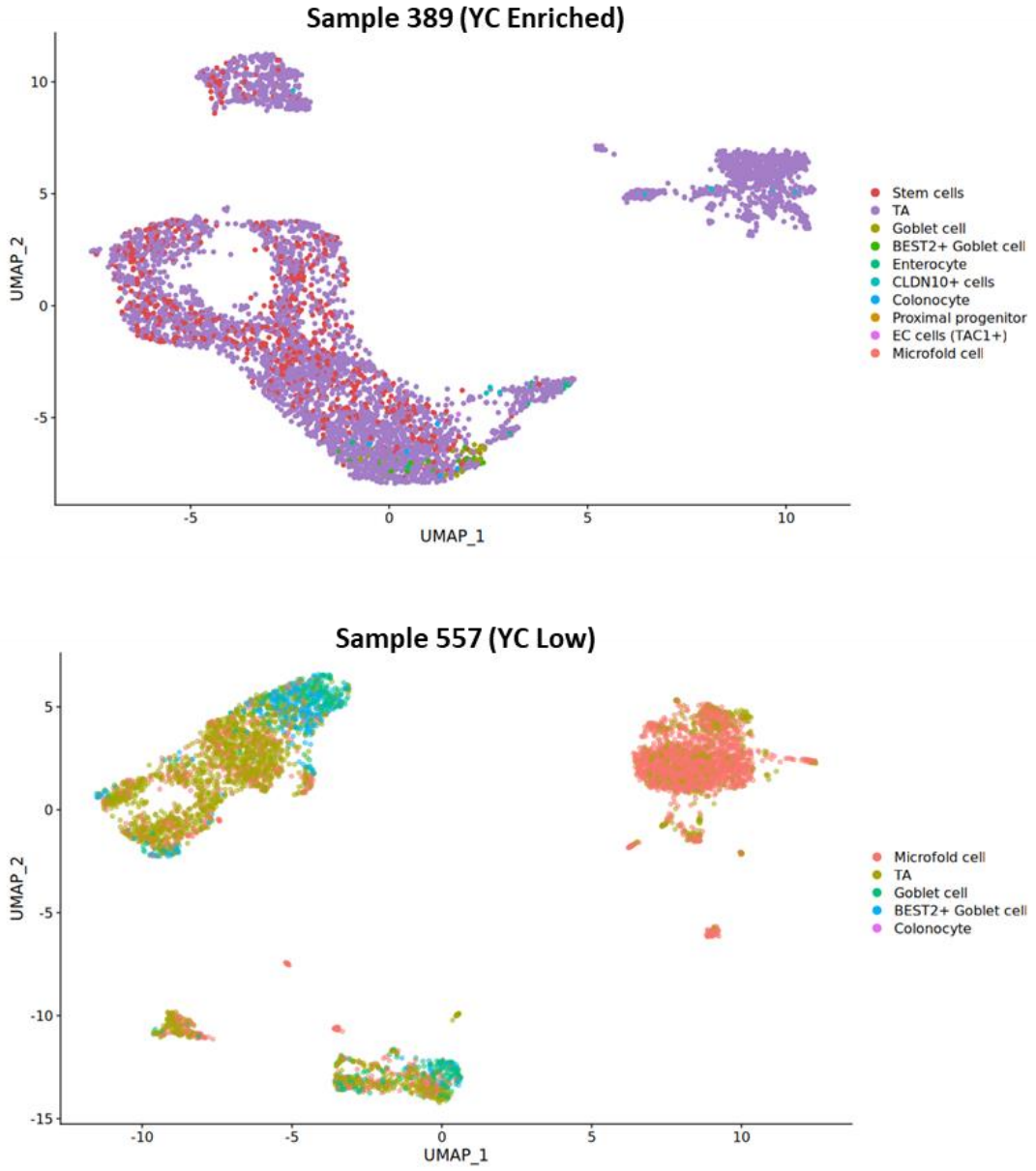


Figure 3.17: **Heterogeneity in cell type and differentiation between organoids.** UMAP projection with cell type annotation. Single cell predicted annotation of intestinal epithelial cell type classification using Single cell gut atlas reference projection onto clustered UMAPs using integrative anchors.

3.2.6 Molecular phenotyping of organoids post irradiation

Previous research in our lab has shown that YC usage is a distinctive transcriptional biomarker for organoids that show high cell death in response to irradiation. Furthermore, YC enrichment is significantly lost post irradiation whereas the depletion in YC usage in irradiated non-responsive does not change post irradiation (Wragg et al., 2023). YC enriched organoid 389 is shown to respond well to irradiation treatment however around 5% of the

population do survive and post treatment bulk CAGE reveals a shift in YC:YR dynamics to a dominant YR usage (Wragg et al., 2023). What is not known however is whether this surviving population represent a distinct population of cells present in the control sample and whether their TSS profile is altered upon irradiation. This would indicate a potential survival regulatory mechanism in TSS usage shifting. Although it was our goal to explore the latter, as previously explained the single cell CAGE proved challenging on our irradiated sample due to inability of obtaining sufficient high-quality cells. However, we were able to sequence a small population of post irradiated cells at a lower depth highlighted in **Table 3.1**. Although we will not be able to characterise promoter usage at this depth, we hypothesised that we would gain essential biological insight into the molecular features and hence cellular state required for cellular survival.

To investigate this our first approach was to identify whether a specific cluster of cells from the control group survived post treatment. This was achieved by using the initial control data and hence UMAP projection from **Figure 3.12** as a reference map to annotate the irradiated sample. Utilising the same approach used in **Figure 3.17** where I used a reference atlas to define cell type. The irradiated sample was initially normalised and clustered as per previous experimental design and dimensionality reduction to produce the representative UMAP projection in **Figure 3.18**. The samples were integrated using anchor based canonical correlation analysis (CCA) integration using the control sample cluster markers, the DEGs, as anchors and transferred onto the irradiated data. To visualise the control cluster identity of the irradiated cells, these were projected onto the original control UMAP outlined in **Figure 3.12**. As shown in **Figure 3.18b** cluster retention post irradiated appears not to be associated with just one cluster or cell state, that in fact a diverse number of clusters and hence cellular states from the control sample exist post irradiation.

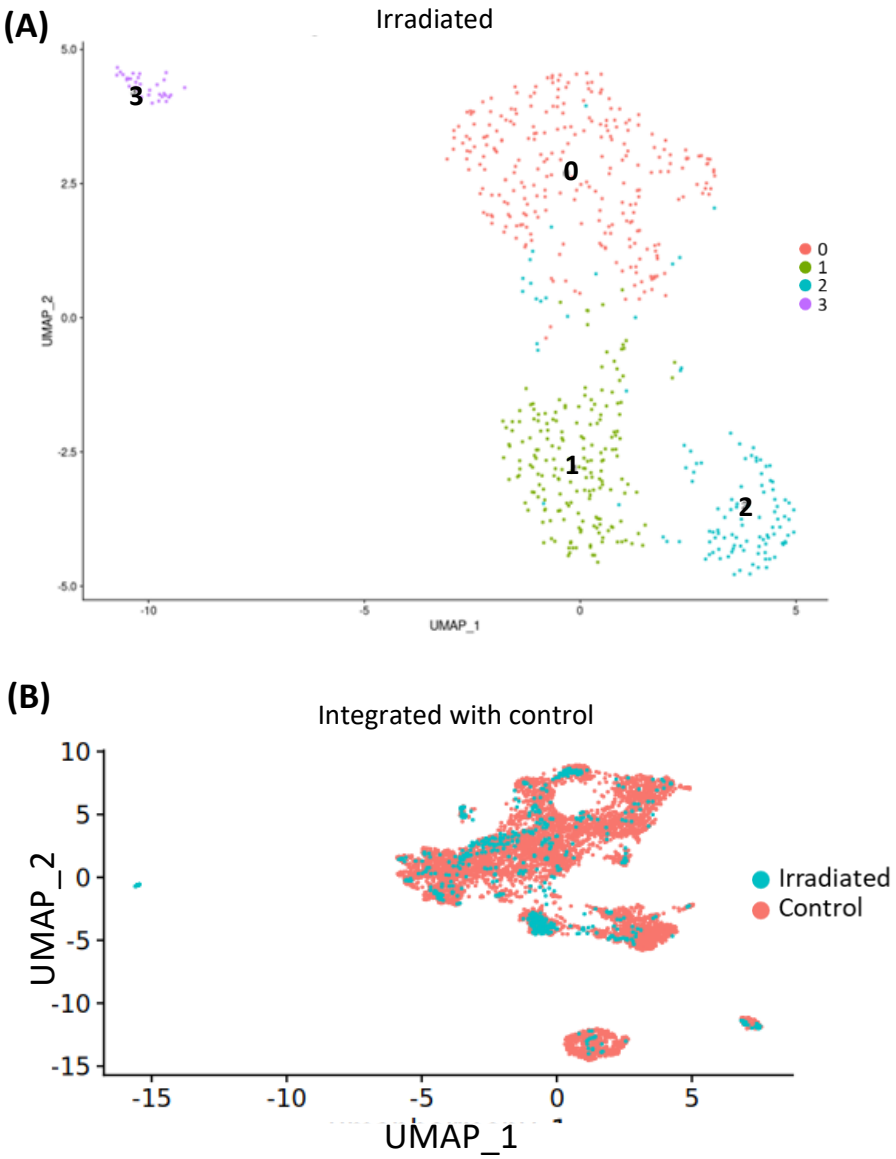


Figure 3.18: **Identification of cellular states post irradiation.** (A) UMAP projection of YC enriched line post irradiation. Intra heterogeneity sub population identification of sample 389 irradiated achieved using Seurat Sctransform version 2 through statistical dimensionality reduction testing and clustering through Pearson residuals normalisation approach. (B) Single cell UMAP projection of sample 389 irradiated mapped onto control. Reference mapping of sample389irr onto control.

To make further relevant comparisons between pre and post treatment we had to account for and attempt to correct the sheer sequencing depth bias between our data sets. Although literature suggests down sampling is unnecessary given many normalisation techniques are sufficient to account for these biases, most analysis done in the statistical testing of these approaches do not deal with data as depth diverse as ours (Lötsch et al., 2021). To achieve a representative down sampling, we firstly calculated the dispersion of our read count per cell in the irradiated sample through interquartile range analysis. We then randomly down

sampled our control using Seurat's SampleUMI function to a max UMI per cell as 10,602, one IQW above the median UMI count per cell in the irradiated. Harmony integration and normalisation was performed on the samples and the result clustering is shown in **Figure 3.19**, further emphasising the variability of cell clusters retained post irradiation.

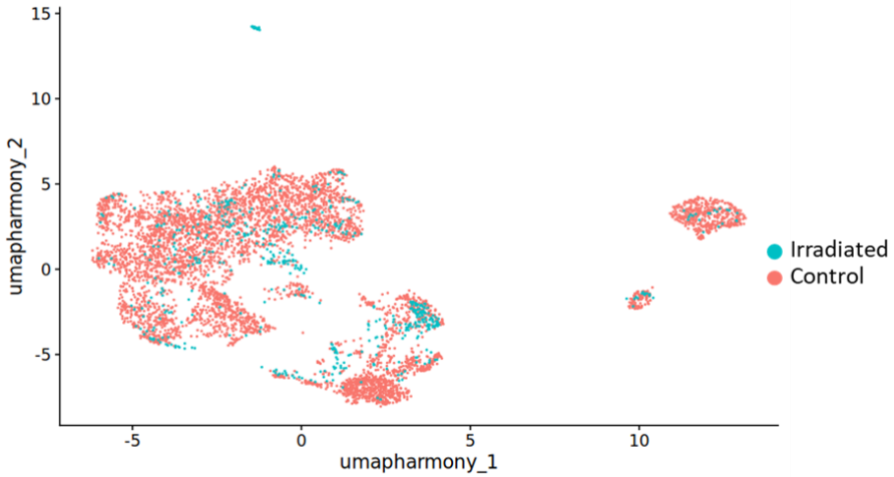


Figure 3.19: Integration of control and irradiated organoids through down sampling (A) Single cell UMAP integration of sample 389 irradiated and control read count down sampled. (B) Post control down sampling UMI count (left) and feature count per cell.

To better understand the features of the cells that survive we performed gene set enrichment analysis between the integrated samples (**Figure 3.20**). We found that most pathways were downregulated post irradiated, including the key growth associated features that underpin the control organoid phenotype, proliferation and mTOR. However, two interesting associated gene sets were on average enriched more highly post irradiation, these were the hypoxia ATM and their gene targets highlighted in previous literature, the hypoxia alternative target genes (ATG). Previous studies have highlighted the role of these genes in the stress response via retention and maintenance of metabolic reprogramming that is essential for survival (Ho et al., 2020, 2021).

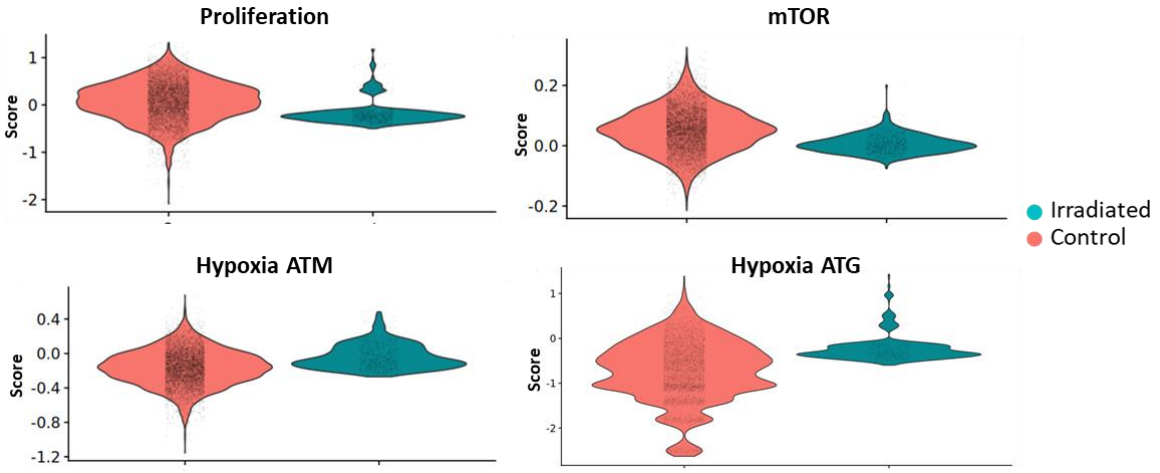


Figure 3.20: Hypoxia alternative machinery target genes are upregulated post irradiation. Gene set enrichment analysis score of pathway associated genes in single sequenced organoids per cell irradiated (blue) vs control (pink). Key growth associated pathway genes, proliferation and mTOR signalling are downregulated post irradiation (top). Hypoxia alternative translation machinery (ATM) genes and their target genes (ATG) are upregulated post irradiation.

Given our previous analysis in **Figure 3.16** showed UMAP clustering to insufficiently highlight similarities between datasets I performed archetype analysis on the control and irradiated samples. Interestingly, this revealed our surviving population did appear to group together, unlike the UMAP clustering suggested. **Figure 3.21** shows post irradiation population plotted relatively central in the polytope at almost equidistance away from the vertices and hence archetypes found in the control sample. This highlights an almost naïve population of cells given that it does not disperse to a vertex thus archetype. Studies highlight that these undefined states are often stem like meaning they do not have a defined function, furthermore such feature is attributable to senescence or stress response (Groves et al., 2021; Hausser et al., 2019). Senescence has been described in literature as a potential survival mechanism to cellular stress induced by some anti-cancer therapies (Schmitt et al., 2022). This supports the downregulation of sensing pathways such as mTOR and upregulation of stress response pathways (**Figure 3.20**).

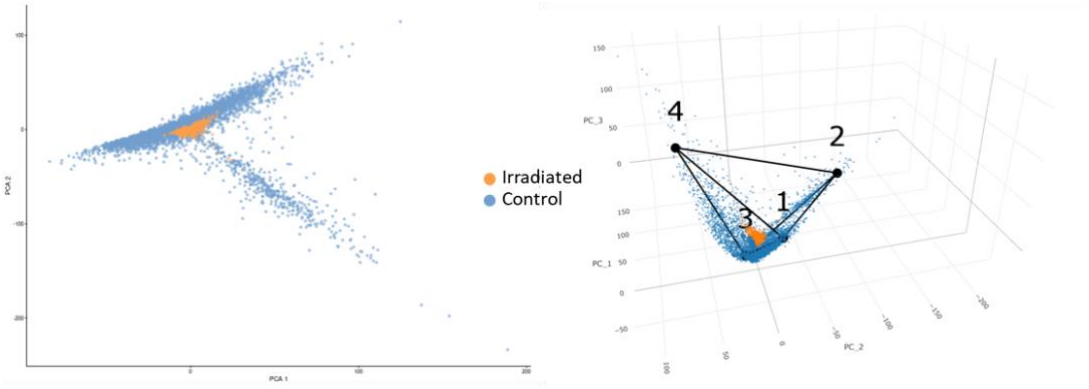


Figure 3.21: **Single Cell Archetype distribution plots integrated control vs irradiated.** Unsupervised characterisation of single cell using ParetoTI analysis. 2D scatterplot of polytope integrating high YC (B) 3D projection of archetype polytopes for integrated dataset containing both 389 Control (Blue) and 389 irradiated (orange) samples.

3.3 Discussion

Cancer heterogeneity is one of the biggest challenges being faced in cancer research, with extensive efforts attempting to resolve the molecular features that contribute to different tumour behaviours (Caiado et al., 2016; Marusyk et al., 2012). Work in our lab has identified a novel TSS choice feature implicated in cancers. The relative TSS usage showed varying YC enrichment correlating with treatment irradiation responsiveness in culture (Wragg et al., 2023). The underlying mechanism of this stratification involves alternative initiation sites present in the core promoter of thousands of genes, characterised into two distinct TSS classes, YR and YC (Nepal et al., 2020; Wragg et al., 2023). The resultant transcript produced from either class are hypothesised to have different pre and post transcriptional regulation. The subsequent transcript ratios both globally and at a gene level represent an interesting biological readout. However, we still do not understand why we see a YC enrichment in cancers and the potential role it plays in tumour dynamics beyond its correlations with treatment responsiveness.

YC enrichment correlates with growth associated organoid phenotypes.

In this chapter, I demonstrate through various experimental approaches that global TSS usage in CRC organoids correlates with molecular phenotypes beyond just treatment response. This finding indicates an interesting link between a new transcriptional initiation regulatory level and tumour state determination. Given the highly proliferative nature of

cancers, extensive research efforts have begun resolving the regulatory mechanisms that facilitate this cancer feature. I show that the high YC TSS profile is underpinned by a highly proliferative phenotype. Firstly, through the culturing doubling time analysis, which revealed doubling time decreases in correlation with YC usage (**Figure 3.1**). It must be noted however that such approach is limited as a higher doubling time could be due to more cells in the high YC organoids simply having organoid formation capabilities. Literature suggests that more stem-like features are required to facilitate organoid formation and hence division into daughter cells (Xiaolei et al., 2017; Yin et al., 2016; Z. Zhao et al., 2022). The imaging analysis in **Figure 3.2** revealed the YC low organoids, 653 and 557, display more differentiated morphology, therefore it is likely that upon splitting many cells do not contain regenerative stem like capabilities in these lines. Although I later confirm these features through sequencing, our analysis is then limited to time point bias. Therefore, to truly understand the growth patterns of our different organoid lines, including the proliferative and differentiation rates, long term 4D imaging of the lines would provide invaluable spatiotemporal insight into their dynamics allowing for better characterisation into tumour evolution.

Further evidence of the proliferative phenotype association to YC usage was seen in our molecular characterisation of bulk CAGE sequencing data (**Figure 3.4**). Although the absence of repeats makes it challenging to make statistically driven comparisons, the module scoring analysis showed that the high YC lines have a higher enrichment for proliferation associated genes. Furthermore, work by Dr Joseph Wragg in our lab identified that the YC enrichment seen in these bulk CAGE analyses was not just a result of an increase in TOP genes, previously identified to be involved in cellular growth dynamics through driving translation (Wragg et al., 2023). Therefore, these findings indicate that YC usage plays an important role in facilitating cell growth beyond just the upregulation of ribosomal associated genes. This analysis also suggests that the non-TOP YC genes may share the same regulatory dynamics as their TOP counterpart.

Single cell CAGE sequencing showed similar transcriptional profile retention to bulk sequencing approaches.

Although bulk analysis proved insightful, we are limited to understanding the biological significance of these findings due to the sheer ITH that exists (Marusyk et al., 2020; Meacham & Morrison, 2013). Therefore, in this chapter a key aim was to successfully undertake and validate single cell CAGE sequencing to characterise the molecular phenotypes within our organoids and later establish TSS choice usage. Utilising the backbone of the sequencing approach outlined in (Moody et al., 2022), we identified key criteria that must be achieved and hence optimised to carry out such an experiment. Upon experimental designing we acknowledged that we must prioritise cellular viability, given that we require high quality RNA to successfully map correct TSS. As a result, we opted to ensure high cell viability through cell sorting based on negative presence of a cell death stain. This step could result in the loss of relevant cell populations therefore true ITH could not be achieved in this analysis. To overcome the potential loss of populations a spatial transcriptomic experiment would be insightful and allow for exploration of true ITH (Brady et al., 2021; Waclaw et al., 2015). This experiment would also allow us to see how representative our identified SC populations are across multiple organoids, or if certain organoids within our lines only contain some of the clusters we identified. Furthermore, to achieve high depth, we had to limit the number of cells we sequenced and therefore we may not be capturing all populations. Both these issues could be resolved using a spatial transcriptomic analysis whilst utilising the cluster marker genes we identified in our sequencing to explore the validity of our analysis.

Given that single cell CAGE had not previously been performed to capture TSS, we were limited on validation and thus relied on characteristics previously identified in bulk sequencing. Our first step of validation and the premise of such highlighted in this chapter was the retention of transcriptional profiles between the single and bulk sequencing. After normalisation to account for library depth differences we identified a strong correlation of TPM counts between the pseudo bulked single cell sequenced and matched bulk. Although PCA results still showed variation between them (**Figure 3.8**), the same variation was retained in both experiments between samples likely owing to the significant difference in sample and library preparation. For example, recent literature has highlighted transcriptional signatures associated with disassociation techniques, a necessary step in single cell preparation (Van Den Brink et al., 2017). Furthermore, the bulk CAGE protocol

relied on a capping technique to identify 5' ends, however the single cell protocol utilised a TSO approach, which has showed some preferential bias for transcript capture and therefore could alter transcriptional profiles (Tang et al., 2013).

Single cell sequencing showed ITH differences in cell types and physiological states between YC low and YC enriched organoids.

Intra tumour heterogeneity analysis can help provide essential insight into the tumour function and even tumour evolution through the scrutinization of the SC populations (Caiado et al., 2016; Marusyk et al., 2012; McGranahan & Swanton, 2017). Previous studies have already begun highlighting the clinical relevance of ITH in determining treatment responsiveness, with more heterogeneity showing more aggressive phenotypes (Caiado et al., 2016; Stanta & Bonin, 2018a; X. X. Sun & Yu, 2015). Furthermore, many of the ITH SC populations share genotypes meaning their behaviours appear to be more dependent on other regulatory mechanisms such as epigenetic and transcription (Caiado et al., 2016). We show that our organoids follow the same trends seen in literature and extend findings to correlations with TSS usage. Firstly, the YC enriched irradiation responsive line shows less heterogeneity compared to the YC low non-responsive line (**Figure 3.12 and 3.13**). YC enriched organoids also show SC populations associated with translation and ribosomal biogenesis, both of which are highly associated with maintenance of growth dynamics and provide likely reasoning for the global growth phenotype we see in these lines. Furthermore, our referencing mapping and cell type annotation revealed that it was composed of stem and stem-like transit amplifying cells. These naïve, undifferentiated cells have been vastly characterised as being growth driven specifically in CRC (Y. S. Ma et al., 2020; Magee et al., 2012). These features provide an explanation into the cystic morphology we see in our high-resolution imaging (**Figure 3.2**). The imaging analysis also highlighted the presence of secondary structures in our YC low lines attributable to differentiated morphology (Kashfi et al., 2018). I later confirmed this in our bulk CAGE through an increase in enrichment in differentiation associated pathways such as WNT and with our single cell reference mapping annotations (**Figure 3.17**). The latter showed the absence of ISC in this population, contrasting research which has highlighted the link between CSC and resistance (Y. Li et al., 2021; Makena et al., 2020; Prieto-Vila et al., 2017). This is likely due to the single cell atlas used to reference map, originating from healthy intestinal tissues, thus missing the

characteristics of CSC. Many studies have shown that as cells differentiate their growth patterns slow and some even become quiescent (Jögi et al., 2012; Lindell et al., 2023). We hypothesise that the shift towards YR usage in this phenotype represents a potential TSS regulatory mechanism. YR usage utilises cap independent translation making it less sensitive to stressors that block cap-dependent translation.

The ability of cells to transition to this quiescent state have shown remarkable associations to cell survival (Kumari & Jat, 2021; Lindell et al., 2023; Schmitt et al., 2022). Such phenotype maybe present in the post irradiated samples, whereby we see a drastic decrease in the number of genes being captured. Interestingly however our analysis shows an upregulation of genes associated with metabolic reprogramming (**Figure 3.20**). A newly identified characteristic of this gene set is their continued translation during stress. Translation continues through the formation of an oxygen dependent translation machinery capable of forming and bypassing the inactive mTOR pathway (Ho et al., 2020; Uniacke et al., 2012). Although we were unable to perform high depth CAGE on our post irradiated sample, previous global bulk analysis showed a significant decrease in YC post treatment (Wragg et al., 2023). Further research is needed to understand whether the change from YC to YR post irradiation is a regulatory mechanism or more simply a transcriptional response, such as a depletion of YC transcript abundance. To begin to explore this, we must further understand the role of these gene sets in the control sub populations and their relative TSS usage. Do we see heterogeneity in TSS usage in these genes between clusters? Are these genes always utilising the YR TSS despite being dual or does TSS usage differ in these genes dependent on the physiological state of the cell?

In this chapter, I characterised the SC populations present in our organoids and showed both distinct differences and similarities between lines. A pervasive challenge was sufficiently comparing cellular physiological states given the many possibilities of GO definitions and our approach in utilising the dominant GO masks many underlying phenotypes. A solution that also provided interesting insight into the continuum of states and hence appreciation for cellular plasticity, was utilising archetypal theory (Combes et al., 2021; Groves et al., 2021; Hart et al., 2015; Hausser et al., 2019; Hausser & Alon, 2020). By using the previously characterised universal archetypes by Hausser et al., 2019 as reference GOs we were more easily able to

compare our SC populations between samples (**Figure 3.16b**). Furthermore, distinct clinical phenotypes have been associated to the different archetypes which provided both a level of validation in our own phenotypes but also greater understanding in the molecular mechanisms underpinning them (Combes et al., 2021; Groves et al., 2021; Hausser et al., 2019; Nath et al., 2021). The identification of shared archetype states between the lines represents an exciting opportunity to explore their TSS usage between samples with different global TSS profiles. Is TSS usage attributable to a specific SC state irrespective of its environment? Does TSS shift between differently characterised clinical phenotypes? Such questions are the drivers for later chapters.

Chapter 4: Investigating intra-promoter switch of TSS in cellular physiological states.

4.1 Introduction

In the previous chapter we identified that TSS usage correlates with a distinct molecular phenotype predominantly associated with enhancing growth dynamics. So far however the analysis has focused on the gene expression profiles associated with a global TSS phenotype and our previous results highlight variable cell states within both organoids. The use of single cell CAGE allows us to map the transcripts to TSS, meaning we can characterise ITH of TSS usage beyond its association to global bulk phenotype. Previous studies have already shown the clinical significance of ITH and its association with treatment response (Stanta & Bonin, 2018a; X. X. Sun & Yu, 2015). The ability of cells to behave differently in the tumour and potentially transition into different states following environmental pressures represents an intrinsic mechanism aiding in cellular survival (Clairambault & Shen, 2020; Pérez-González et al., 2023; Torborg et al., 2022). Transcriptional landscapes are a vital component of this intrinsic mechanism and understanding the regulatory components will help in identifying these tumour evolution dynamics and hence cellular plasticity. In this chapter, I aim to understand the penetrance of TSS usage differences and its contribution to ITH. Firstly, by characterising the TSS phenotype of cellular physiological states to see whether YC enrichment is unique to a state. Understanding whether TSS usage changes between molecular states will provide invaluable insight into the biological relevance of TSS base choice variation. Secondly, to explore this phenomenon on a single cell level, a resolution not previously achieved.

Chapter 3 also identified shared states within these differing organoids including both metabolic and cell division associated subpopulations. This poses the question of whether the TSS usage and hence regulatory mechanisms of these states are also shared? For example, do the metabolic subpopulations between the samples have the same TSS phenotype or do they differ? Therefore, the aim of this chapter is to understand the TSS

usage of both different and shared cellular physiological states. As cellular physiological states are underpinned by specific gene expression profiles, I aim to identify the gene networks responsible for YC enriched and YC low states. By identifying functionally relevant gene sets we can further explore the dependence on the initiator motif. The YC TSS consists of different subsets mainly dependent on their initiator sequence and similarity to the TOP motif. Previous research has highlighted the functional relevance of the TOP motif through differential post transcriptional regulation driven by LARP1 binding and mTOR regulated translations machinery (Hochstoeger et al., 2024; Jia et al., 2021; Mura et al., 2015; Ogami et al., 2022; Philippe et al., 2020). Furthermore, work in the lab has identified that YC TSS without the TOP feature behave transcriptionally similar to TOP on a global scale. Therefore, I aim to see if the initiator motif of YC subsets is attributable to certain genes linked to a cell state.

4.2 Results

4.2.1 Validation of single cell CAGE TSS capture

As previously highlighted in chapter 3, single cell CAGE has not previously been performed to study TSS variation dynamics. Having established a level of validation in the previous chapter through investigating the general gene expression profile in our lines that has been sufficiently retained across sequencing approaches, we now aimed to validate the TSS capturing success. The development of this single cell sequencing approach was designed, performed, and outlined by (Moody et al., 2022), although their aim was to characterise cis regulatory elements and not TSS usage, they developed a pipeline (SCAFE) to remove G bias artefacts that can dominate the sequencing data as discussed in Chapter 1. The reads and hence bam files generated from the *cellranger* pipeline were analysed through the SCAFE pipeline outlines in **Chapter 2.5.1**. This analysis was carried out by Dr Yavor Hadzhiev. To show that the single cell sequencing approach is reflective of TSS usage and hence no bias of base capture I aimed to compare TSS profiles against the previously sequenced bulk data. To do this I pseudo bulked both single cell samples by summing filtered, corrected reads obtained from the SCAFE pipeline. The bam files for single cell pseudo bulk and matched bulk were then analysed in the CAGEr package outlines in **Chapter 2.5.5**. To compare between samples, normalisation of CAGE tags was performed as described in **Chapter 3.2.4**

using a power law distribution supported in the CAGEr workflow (Balwierz et al., 2009; Haberle et al., 2015). This step was essential to attempt to account for extensive disparities between library sizes of the single cell and bulk sequencing highlighted in **Table 4.1**.

Table 4.1: Library size differences between sequencing approaches

Sample	Library Size
Single cell 389	225,644,802
Bulk 389	7,755,829
Single cell 557	181,357,951
Bulk 557	26,650,370

Reverse cumulative plots were generated depicted in **Figure 3.8**, and optimal parameters for power law normalisation determined as $\alpha = 1.05$ fit in range of tags 10 to 15,000 and converted to TPM metric. Tag clusters representing likewise promoter regions defined as genomic regions within 100bp, harbouring multiple spatially separated Tags are then calculated using a distance-based clustering approach using the *distclu* argument in the function, keeping singletons (clusters that have only one Tag) with a normalised signal over 3. The repeated presence across samples means it is found frequently enough to be valid. A genomic ranges object is produced from the normalised consensus TSS (CTSS) with the total TPM signal. Consensus clusters (CC) were then calculated between the samples to identify and define shared promoter regions by aggregating Tag clusters within the quantile boundaries (set to standard specified in CAGEr workflow) therefore producing CAGE signal sum in all samples mapped to genomic coordinates for each CC.

Previous findings by Dr Wragg showed that global dinucleotide frequency changes characterise the organoids (Wragg et al., 2023). Therefore, I aimed to validate the TSS classes are retained in our single cell pseudo bulked data. I achieved this by summing CTSS signal per initiator and dividing by total signal to calculate % of initiator class as outlined in **Chapter 2.5.5**. Previous bulk CAGE sequencing analysis showed %YC increase in sample 389 compared to 557, our pseudo bulk single cell sequencing also matched this trend as seen in **Figure 4.1**.

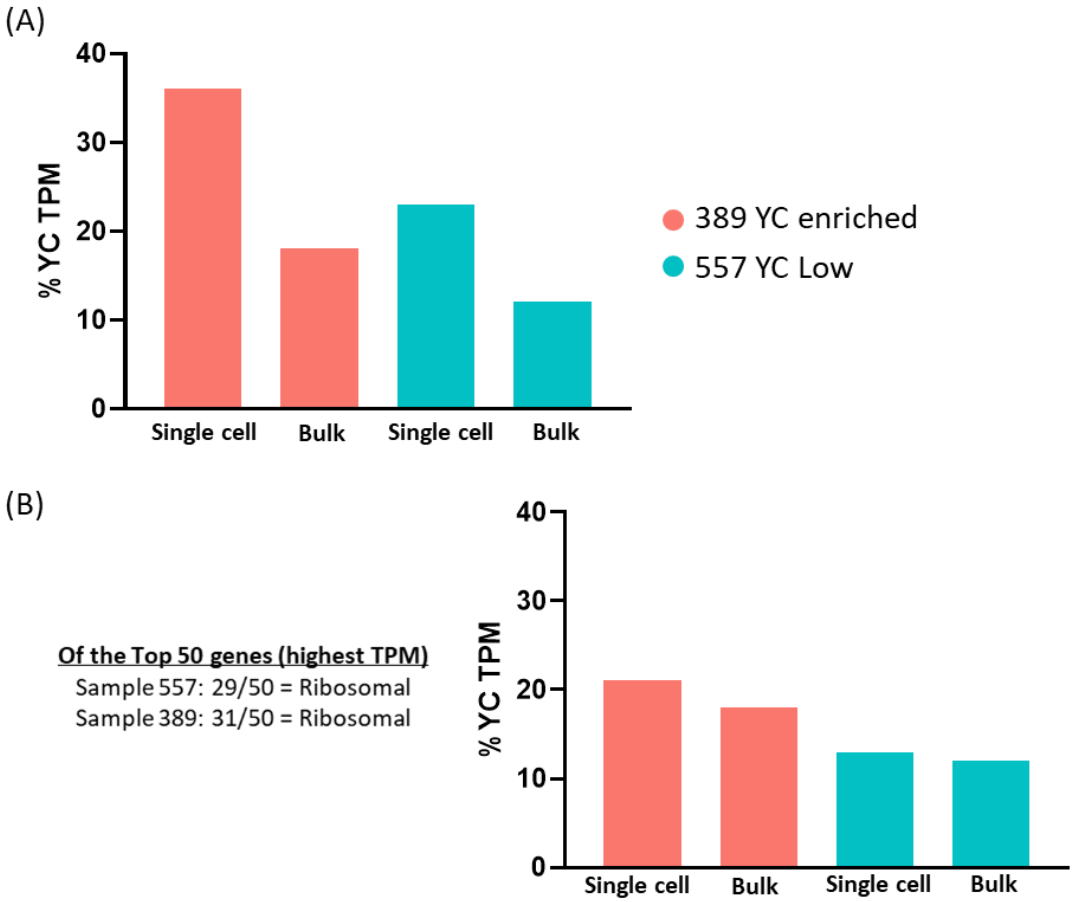


Figure 4.1 Global YC enrichment is retained in matched samples across sequencing approaches. Dinucleotide frequency analysis between pseudo bulked single cell sequencing and matched bulk. (A) Percentage of total TSS is YC in single cell pseudo bulked compared to its matched bulked. (B) Top 50 expressed genes in the single cell data show high ribosomal bias. Right panel shows YC % in single cell post removal of the top 50 genes.

Our analysis revealed that despite using different sample preparation, sequencing technique, and library depth differences, we still see the same YC usage trends in our organoid lines. However, we did see a dramatic increase in YC enrichment in both pseudo bulked samples compared to the matched control bulk. Our previous analysis in Chapter 3 revealed that on average 10-25% of the reads map to ribosomal genes (**Figure 3.7**) additionally our TPM per gene analysis in **Figure 3.8** showed the highly expressed genes that differ to bulk are ribosomal genes. Previous studies have already identified ribosomal genes as being TOP sequence containing genes, we hypothesised that our high YC values in our single cell data was a result of a more efficient capture of abundant ribosomal protein genes (Cockman et al., 2020; Hochstoeger & Chao, 2024; Meyuhas & Kahan, 2015) . We explored this bias by removing the signal of the top 50 genes in our single cell datasets and repeated the dinucleotide frequency comparisons. We found, global YC and YR levels more closely

represented the bulk data but still retained the difference between organoids (**Figure 4.1b**). These results suggest that our single cell CAGE was successfully represents global TSS base choice.

4.2.2 Characterisation of TSS base choice phenotype on a single cell level

Our single cell CAGE data provides the opportunity to explore TSS base choice phenotype on a single cell level which has not yet been described in the literature. Although the organoids show a bulk YC:YR TSS ratio variation, Chapter 3 showed these lines harbour degrees of ITH in cellular physiological state and cell type. Therefore, we hypothesised that there would be ITH variation in TSS usage between the cells within the same sample resulting from these features. A challenge however would be in visualisation of the usage on a single cell resolution, given that we had around 5,000 cells per sample. Firstly, we generated a single cell matrix profiling the individual cells with information generated from previous pipeline analysis including read and feature counts. To this matrix the relative TPM signals per initiator (YC or YR) were added. Additionally using the generated CC ID we were able to calculate the average YC:YR usage ratio per cell, including all promoters with both signals, therefore including only dual genes. Collation of this data on a single cell level allows us to scrutinise the variation of YC:YR in and between samples and compare against other features.

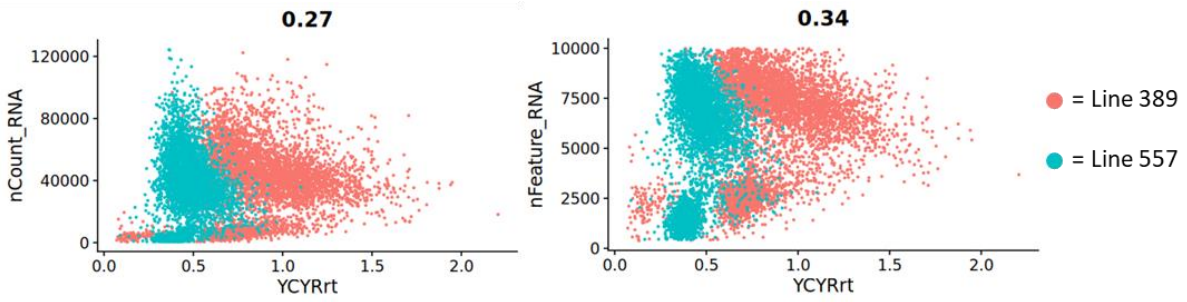


Figure 4.2. Scatter YC:YR ratio shows marked variation within samples plot comparison of YC:YR vs read depth per cell. Scatter plot comparisons of YC:YR per cell vs cell features. Average YC:YR ratio per cell against number of molecules, hence read depth (left) and against Feature or gene count captured per cell, coloured by sample of origin.

As seen when plotting the YC:YR vs genes captured per cell, not only do we see validation in the previously seen bulk trend that the 389 responsive line has a higher YC usage in dual initiator genes, we also now see variation in this usage within the sample including cells with

similar ratio values between samples (**Figure 4.2**). This raised the question of whether these cells in different samples were in the same cellular physiological state. To initially explore this, we plotted the calculated YC:YR ratio per cell values onto our UMAP projections developed in Chapter 3, to see if TSS phenotype clustered together based on gene expression determined clusters. Interestingly in both samples the high YC cells predominantly clustered close together indicating many harboured a similar gene expression profile (**Figure 4.3**). In sample 389 another cluster of high YC cells appears in cluster 4, located distinctively away from the large high YC cluster revealing a different gene expression profile.

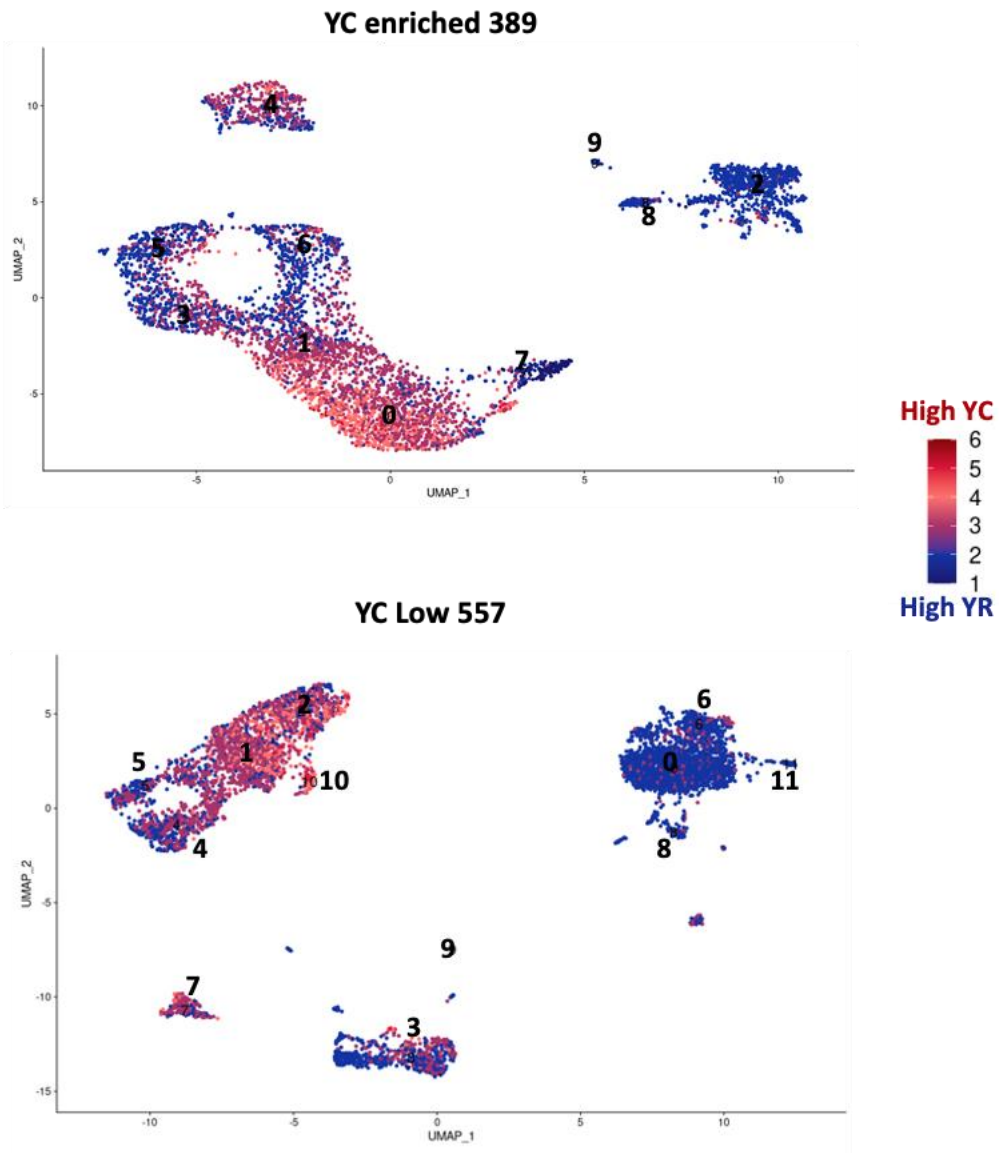


Figure 4.3: **YC enriched cells cluster close together.** Single cell UMAP projection of single cell data coloured by average YC:YR of all DUAL genes calculated per cell. Sample 389 Top, sample 557 bottom. YC shift in DUAL TSS dynamic depicted by colour scale on right of panel.

To identify the cellular state of these high YC cells and see if they are shared across samples, we performed gene set enrichment module scoring for states of interest, as had been previously performed on whole sample level in **Chapter 3.10**. To make visual comparisons against the YC:YR, the expression heat map colour scales were kept the same across all UMAPs. Additionally, to statistically validate the trends, we performed linear correlation analysis between each gene set feature score per cell against its average YC:YR ratio to produce Pearson correlation coefficient (r) for each feature to measure both strength and direction of the relationship.

Firstly, gene set analysis showed heterogeneity in scores within both samples, further validating previous ITH analyses. We also saw expected expression correlations between gene sets. For example, as can be seen in **Figure 4.4** the highest expression of proliferation markers correlated with the *Myc* gene and its target gene expression. Studies have shown that Myc proteins are master TF that upregulate and induce the expression of many genes, many of which are associated with proliferation (Ahmadi et al., 2021). A widely characterised feature of Myc proteins is their role in cell cycle progression through its multifunctional interactions including transcriptional activation of pro-cycle genes and the repression of the inhibitors (Ahmadi et al., 2021; García-Gutiérrez et al., 2019). Correlation analysis revealed that both these proliferative features negatively correlated to YC:YR in both samples, highlighting a further link between proliferation and TSS usage, however opposite to the overall phenotype analysis which shows fast growing and hence highly proliferative lines correlate higher with YC usage. We aim to explore the reasoning for this in a later chapter.

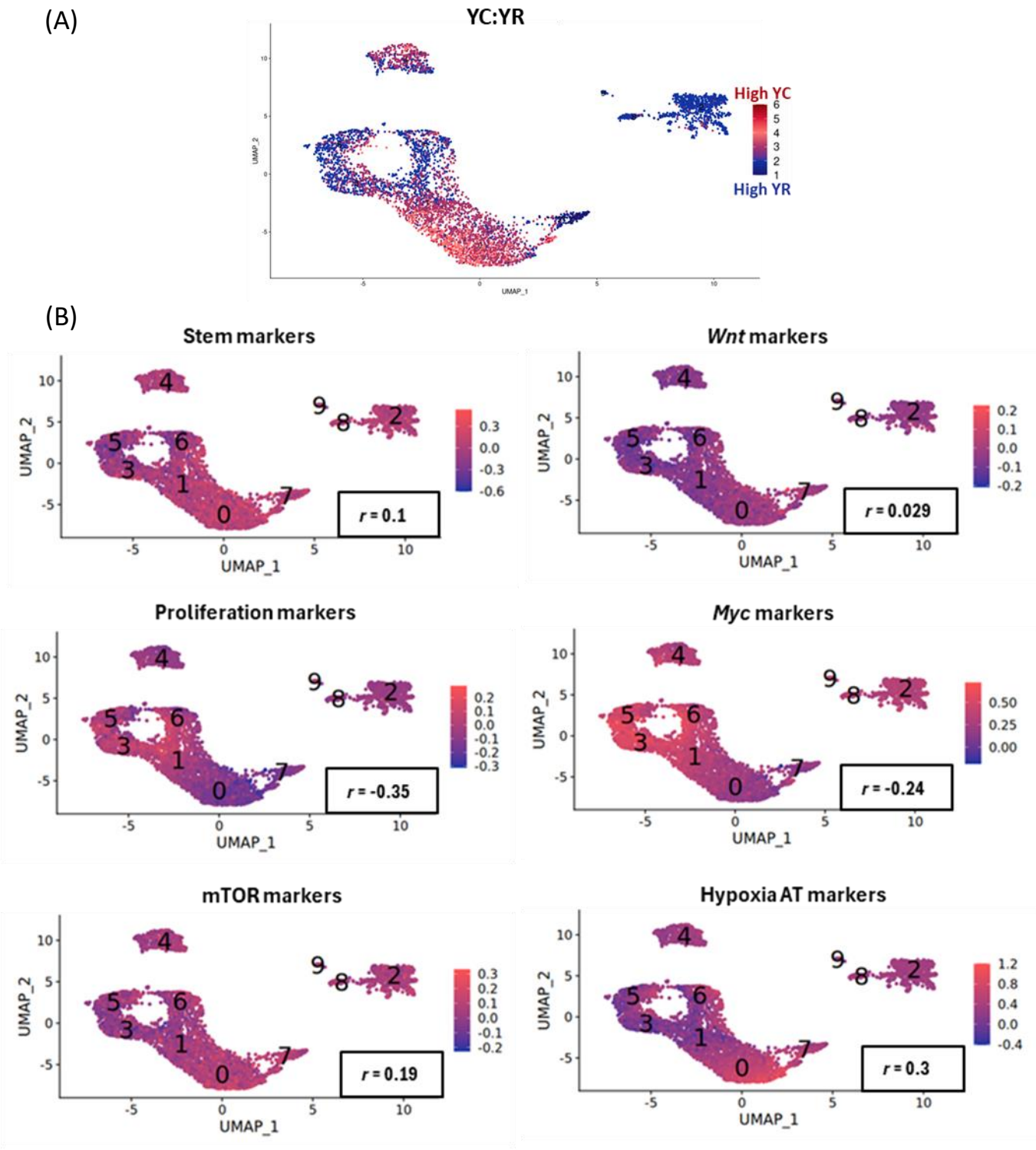


Figure 4.4: High YC cells are enriched for metabolic reprogramming genes in fast growing organoid. (A) YC:YR average per cell. (B) Single cell UMAP projection of gene set enrichment per cell sample 389. Calculated module score for gene sets shown in scale bar. Top panel includes differentiation associated gene sets, Middle proliferation and Bottom cap dependent translation and cap independent translation. R in each panel shows Pearsons correlation coefficient (*r*) of module score per cell vs average YC:YR.

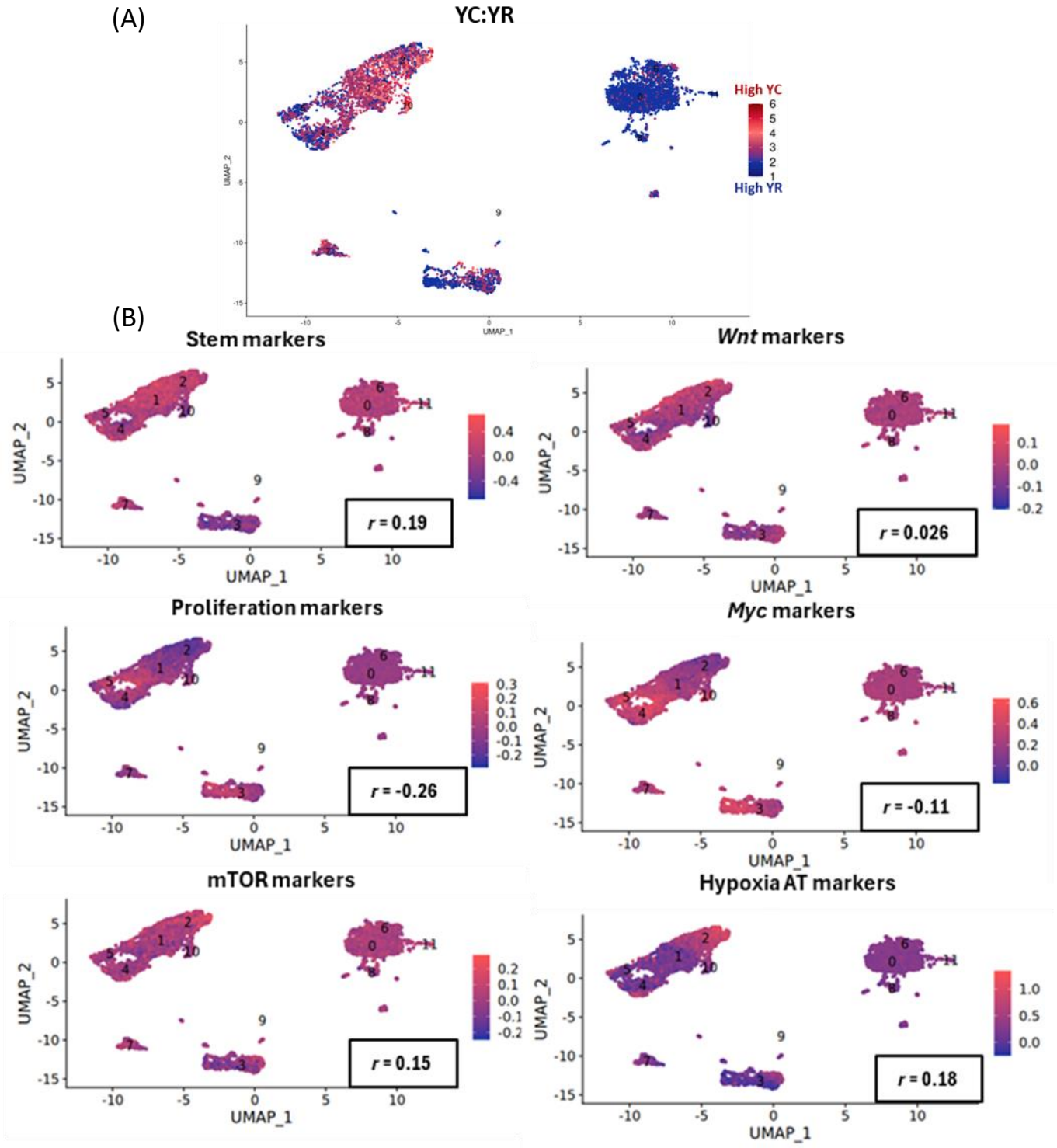


Figure 4.5: High YR cells are enriched for proliferation genes associated with cell division processes. (A) YC:YR average per cell. (B) Single cell UMAP projection of gene set enrichment per cell sample 557. Calculated module score for gene sets shown in scale bar. Top panel includes differentiation associated gene sets, Middle proliferation and Bottom cap dependent translation and cap independent translation. R in each panel shows Pearsons correlation coefficient (*r*) of module score per cell vs average YC:YR.

Although none of our correlations showed strong associations to YC:YR ratio, clear trends were seen within the samples. For example, in sample 389 the more proliferative line the largest positive correlation was with the hypoxia ATG. ATG is a gene set widely associated with metabolic reprogramming essential for maintaining the redox balance and thus enabling cellular survival, and are targeted by ATM in low oxygen levels (Ho et al., 2020; Uniacke et al., 2012). The anticorrelation between the proliferative phenotype vs YC:YR and the metabolic remodelling phenotype vs YC:YR better highlighted in Figure 4.6 and could indicate a TSS level regulation of growth dynamics through the upregulation of genes responsible for production of key genes and proteins that provide the building blocks essential for cellular division. TSS usage may play a role in the proliferation vs metabolism equilibrium discussed in Chapter 1, a hypothesis explored later in this chapter. Although a similar trend was seen in the slower growing 557-line (**Figure 4.5**), the correlations were weaker. The strongest correlation was with stem cell markers, many of which are associated with lineage differentiation. This analysis showed that the YC:YR was highest in the cells that displayed the predominant phenotype of the whole line: sample 389 is highly proliferative and YC:YR was highest in the genes responsible for aiding in growth through metabolic maintenance whereas sample 557 was defined as highly differentiated and the highest YC:YR was in the cells showing the highest expression for differentiation markers.

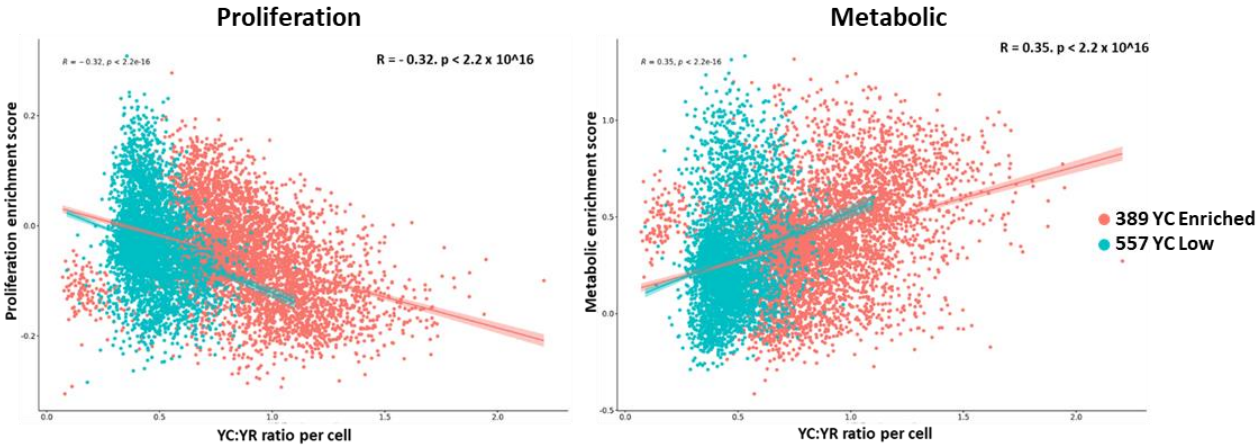


Figure 4.6: **TSS usage vs proliferation and metabolic features.** Scatter plot correlation analysis of module score for proliferation associated gene sets and hypoxia associated metabolic reprogramming gene vs average YC:YR usage per cell coloured by sample of origin. Linear regression analysis performed using pearsons coefficient.

To show that the high YC and hence larger YC:YR values were not simply due to higher expression of dual initiator genes, we performed correlation analysis between average dual gene expression

and YC:YR per cell (**Figure 4.7**). A low positive correlation was seen for both however it can be argued that the greater dispersion in data is seen between the YC:YR axis than the dual gene expression score, meaning that not only is there a level of TSS usage in these genes not fully explored but that conventional RNA-sequencing would not be able to detect such changes.

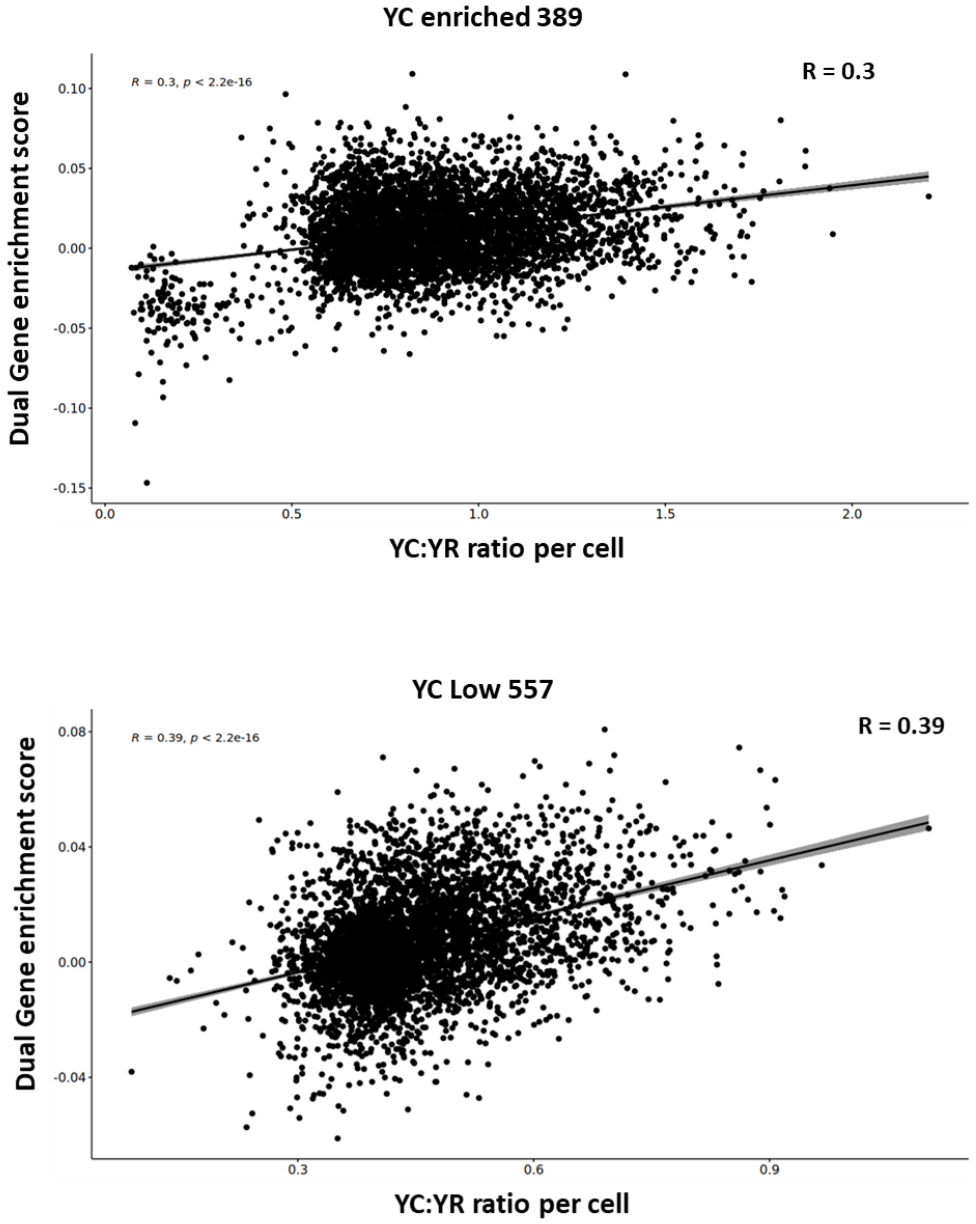


Figure 4.7. YC:YR ratio does not strongly correlate with dual initiator gene differential expression. Scatter plot comparison of average YC:YR vs dual initiator gene expression module score per cell. Regression analysis linear model displayed with calculated Pearson Coefficient R score.

Furthermore to show this was not a TOP motif specific feature but also attributable to YC TSS not harbouring the conventional initiator TOP motif, we calculated the non-TOP YC signal per cell through extracting the first 5 bp of the TSS and summing TPM for those without the presence of the polypyrimidine stretch and calculated YC:YR per CC and thus average non-TOP-YC : YR per cell.

UMAP projection shows ITH variability of non-TOP-YC dynamics in likewise dynamics as all YC:YR (**Figure 4.8**). Correlation analysis vs all-YC:YR shows a positive correlation indicating the YC dynamics are not we see are not solely resulting from TOP associated YC. This suggests a similar regulatory and hence usage dynamic between all YC initiation sites.

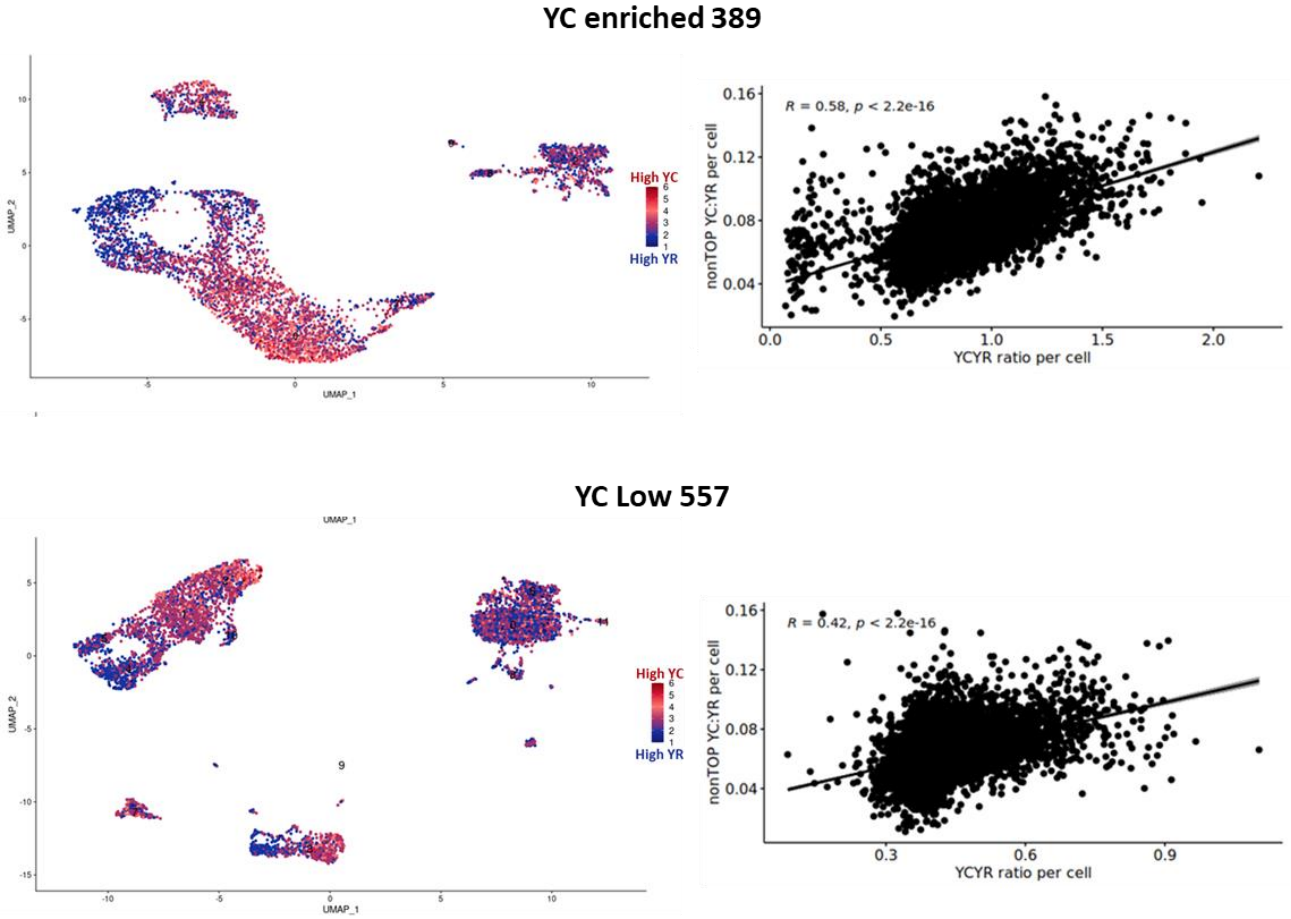


Figure 4.8: **Non-TOP YC initiators show similar ITH enrichment.** Single cell UMAP projection of TSS profiles per cell. Average non-TOP-YC:YR of all DUAL genes calculated per cell. YC shift in DUAL TSS dynamic depicted by colour scale on right of panel. Right panel shows non-TOP YC: YR vs all YC:YR correlation analysis with linear regression Pearson coefficient values.

4.2.3 TSS usage in cancer cell physiological states

Now that I have confirmed there is ITH of TSS usage phenotypes I aimed to characterise the cells that show a higher YC usage and consequently the genes responsible for the dynamic shifting as we hypothesised this would give us insight into biological relevance. A challenge with continuing our analysis on a single cell level was although the depth per cell provided some insight into the global gene usage, resolution to a single gene wasn't efficient due to the absence of sufficient molecules per CCID identified making quantitative comparison gene by gene between single cells

difficult. Therefore we strived to identify the relevant genes using a pseudo bulking approach which we could later explore on a single cell, single gene level. We hypothesized that as our UMAP projection in **Figure 4.3** showed cells with similar YC:YR average clustered together, we could pseudo bulk the cells based on previously characterised features determined in Chapter 3. Our first approach was to see if the high YC dynamic was attributable to cell differentiation types. As previous work in zebrafish embryos show cells alter TSS architecture at genes responsible for tissue function as they move away from pluripotency (Wragg et al., 2020), I aimed to see if the more stem-like pluripotent cells in our organoids showed a distinct YC:YR dynamic. To explore this, we extracted all barcodes attributed to each cell type identity in the Seurat object. The SCAFE-filtered and processed CTSS for each barcode were collated based on cell type identity thus treating each cell type as an individual sample (**Figure 4.9a**).

To identify YC:YR shifting I identified all dual initiator CCID and calculated YC signal to YR (YC:YR) per gene for each cell type and then the average YC:YR per gene across the sample (all cell types). Next I normally distributed the data by log transforming the YC:YR per gene per sample / the average YC:YR per gene, equation shown in **Figure 4.9b**. Cell types where there is a YC shift in the YC:YR ratio will plot with their relative frequency plot shifted above 0, the opposite is true for the samples with a higher YR shift than the other cell types. Finally if there is no changes in the YC:YR dynamic between the samples then the density median points will stay close to 0. Only the stem and TA cells were included in the analysis for sample 389 (**Figure 4.9b**) as the other populations contained minimal cells and comparisons between wouldn't be possible.

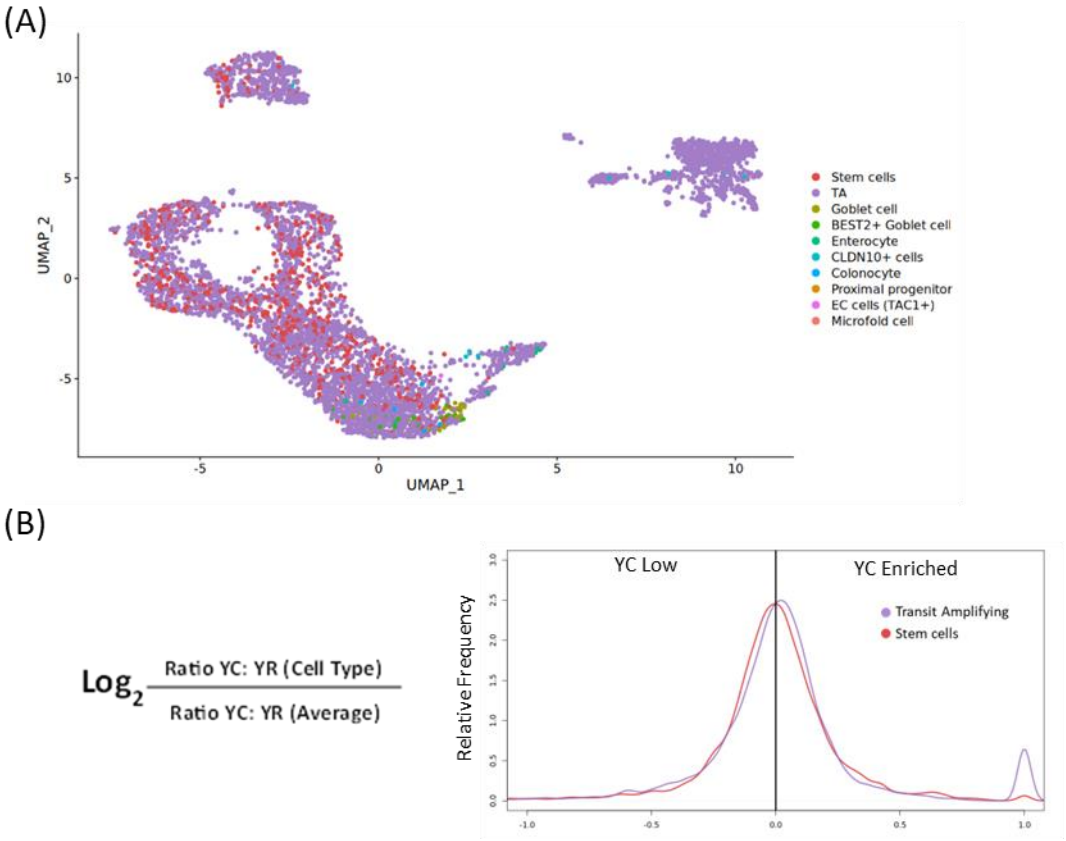


Figure 4.9: **YC:YR dynamics do not differ between naïve intestinal cell types.** (A) Single cell predicted annotation of intestinal epithelial cell type classification using Single cell gut atlas reference projection (Zilbauer et al., 2023) onto clustered UMAPs using integrative anchors. (B) Frequency distribution of YC:YR ratio per cell type pseudo bulk compared to total pseudo bulked average.

Although some separation between YC:YR is seen in the more heterogeneous organoid 557 (**Figure 4.10a**). Organoid 557 shows more differentiated lineages, however given the plasticity of cancer cells they are unlikely to be terminally differentiating like the cells present in the developing embryo. This retention of some stem like capabilities may indicate disparities between the cancer and embryo development, thus all TSS trends may not be conserved. Frequency distribution analysis did show the more distinctive proliferative transit amplifying cells appear to be YR enriched (**Figure 4.10b**). Future comparisons will seek to compare likewise cell types between the organoids. For example, integrating both organoids and comparing the YC:YR of the transit amplifying cells seen in both.

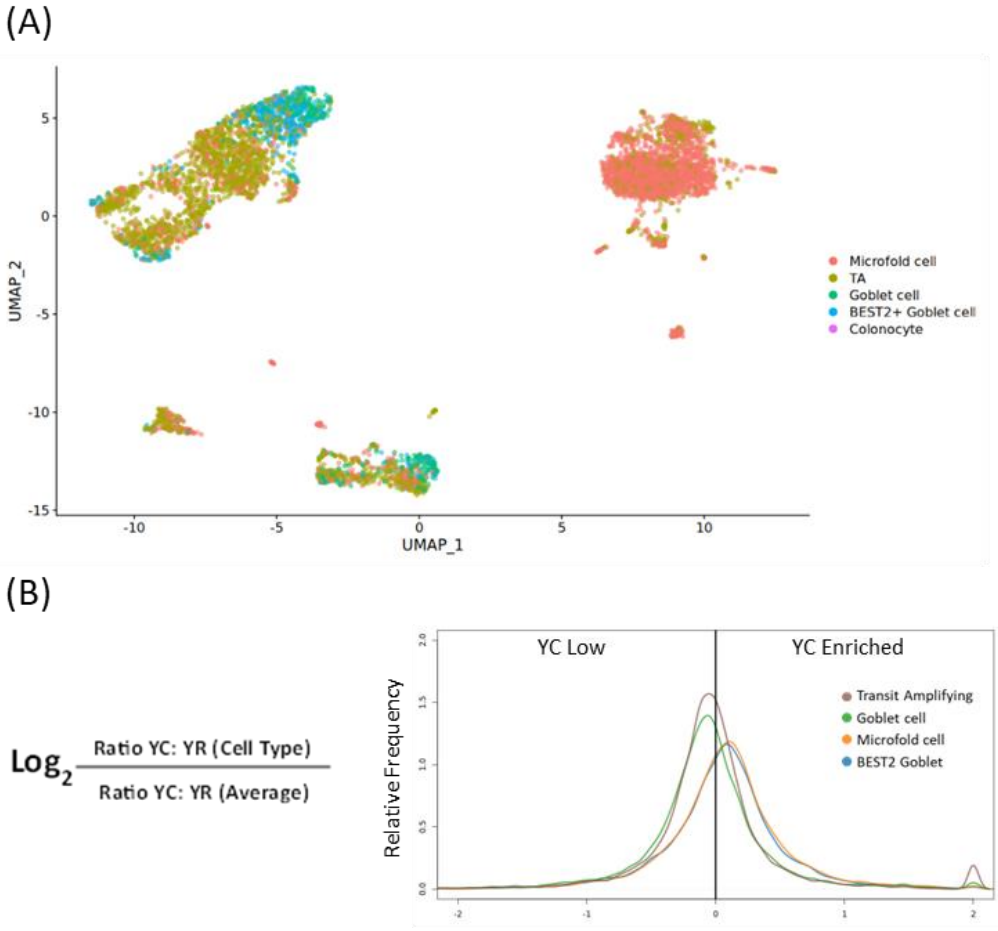


Figure 4.10: **YC is lower in progenitor cells in more differentiated organoid.** (A) Single cell predicted annotation of intestinal epithelial cell type classification onto YC low sample 557 using Single cell gut atlas reference projection (Zilbauer et al., 2023) onto clustered UMAPs using integrative anchors. (B) Frequency distribution of YC:YR ratio per cell type pseudo bulk compared to total pseudo bulked average.

As our single cell TSS analysis in **Figure 4.3** showed the high YC cells cluster together on the UMAP I decided to repeat the above pseudo bulking approach using UMAP generated clusters we defined in Chapter 3 (**Figure 3.12** and **3.13**). Our frequency distribution of YC:YR analysis showed a greater separation between clusters in the YC enriched sample 389 (**Figure 4.11a**). Furthermore, the high YC phenotype did not appear to be a result of just one population unique to this sample, with 3 clusters (translation, metabolism and cell migration) all showing a marked YC shift. Despite the global whole organoid TSS phenotype being significantly different, this analysis showed that the same GO defined clusters, or cell physiological states, also show a YC shift in the 557 sample (**Figure 4.11b**). This highlights a potential TSS usage dependence of these physiological cell states many of which appear to be characteristically energy demanding or focused. For example, literature highlights that

translation is one of the most energy demanding processes of a cell (Roux & Topisirovic, 2012). In both samples the YC low or YR enriched populations are consistently cellular physiological states associated with cell division including the DNA repair. This highlights a potential cell cycle associated dynamic which will be investigated in a later chapter.

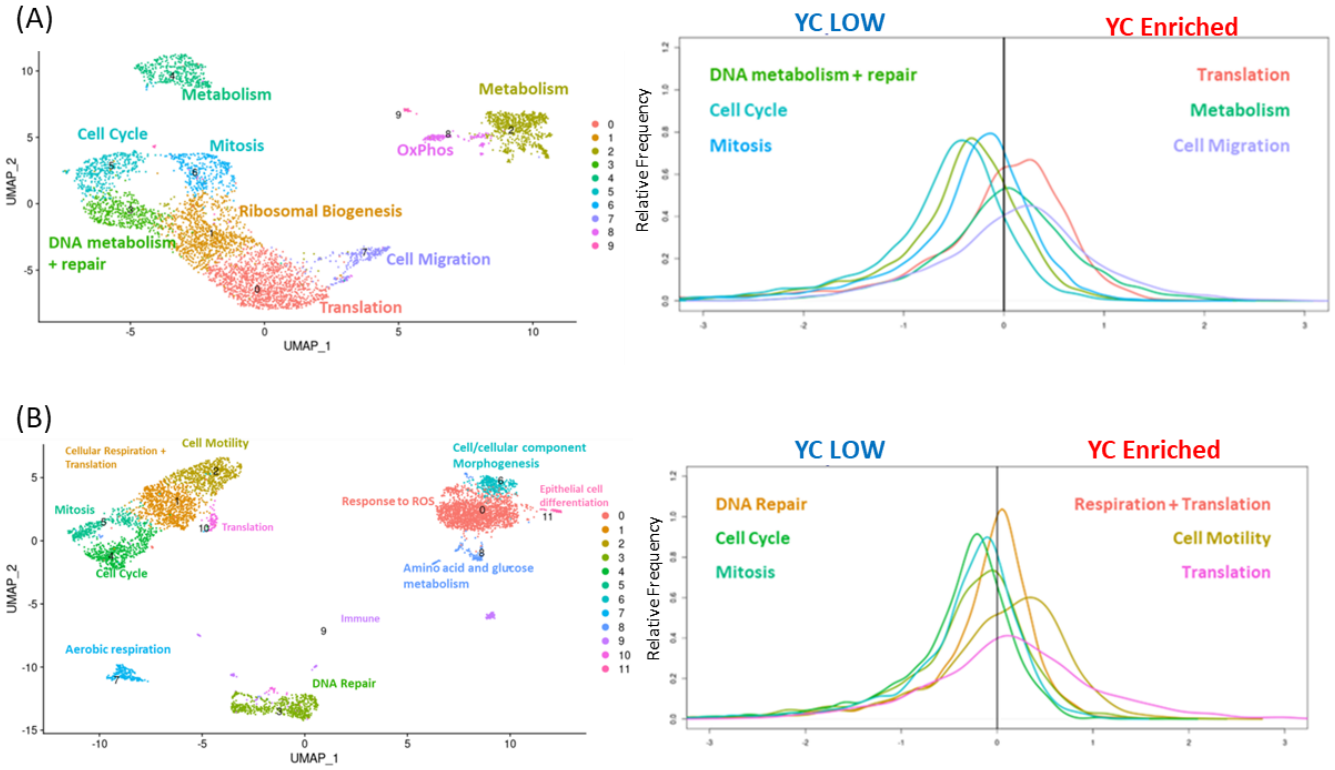


Figure 4.11. YC enrichment is seen in likewise clusters across different lines. Comparison of YC:YR dynamics in subpopulations within samples. UMAP projection of subpopulations clustered and annotated with dominant gene ontology. Frequency distribution of YC:YR ratio per sample compared to pseudo bulked average including only clusters where YC:YR shifted. YC:YR expression ratio calculated per DUAL initiator per cluster divided by average YC:YR ratio for that promoter. Panel (A) represents sample 389 (n=6265) (B) represents sample 557 (n = 3,360).

Previous work by Dr Joseph Wragg in our lab, identified a signature marker set of dual genes whose TSS usage and hence YC shift is sufficient stratification criteria for irradiation (Wragg et al., 2023). Repeating this analysis between the YC lower 389 clusters shows we see greater dynamic shifting in our cluster YC:YR seen in the frequency distribution plot in **Figure 4.12.** Further validating the significance of these genes beyond just association to irradiation phenotype and thus biological relevance they may provide into understanding YC usage.

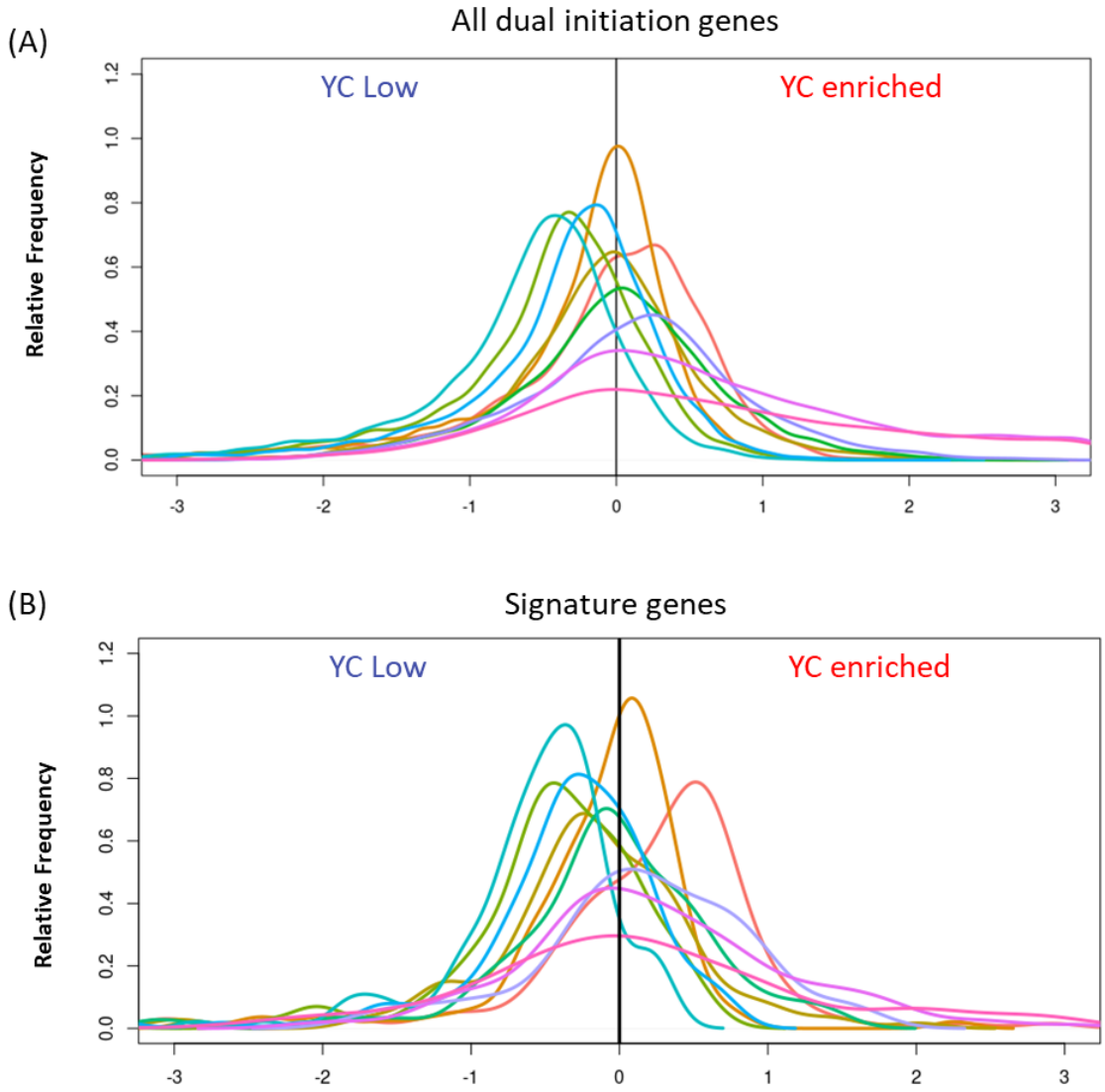


Figure 4.12. YC shifting signature genes validate ITH TSS cluster separation. (A) Frequency distribution analysis of organoid 389 clusters for all dual initiator genes. (B) Organoid 389 cluster YC:YR dynamics of signature genes. Frequency distribution of YC:YR ratio per sample compared to pseudo bulked average for signature Dual genes. YC:YR expression ratio calculated per DUAL initiator per cluster divided by average YC:YR ratio for that promoter.

The presence of translation associated GOs in the YC enriched clusters was expected as many studies have shown that TOP initiator motif is highly associated to translation associated genes (Hochstoeger & Chao, 2024; Meyuhas & Kahan, 2015). To investigate whether this shift was the result of TOP specific transcription enrichment we calculated the nonTOP-YC : YR ratio and repeated the normal distribution. As can be seen in **Figure 4.13** the YC enriched shifts between the same clusters are still present, indicating a potential similar regulatory mechanism for all YC transcripts. It must be noted however that these genes could still contain a TOP motif downstream of their TSS and hence still be targeted by TOP regulation

mechanisms including LARP1 (Philippe et al., 2018, 2020). In this analysis we define non-TOP-YC as the absence of a TOP signal detected in the CCID, However, to fully explore non-TOP-YC genes we must select genes where no TOP motif is found downstream of the TSS or identify if distance of the TOP motif from the TSS affects the dynamic by repeating frequency distribution analysis.

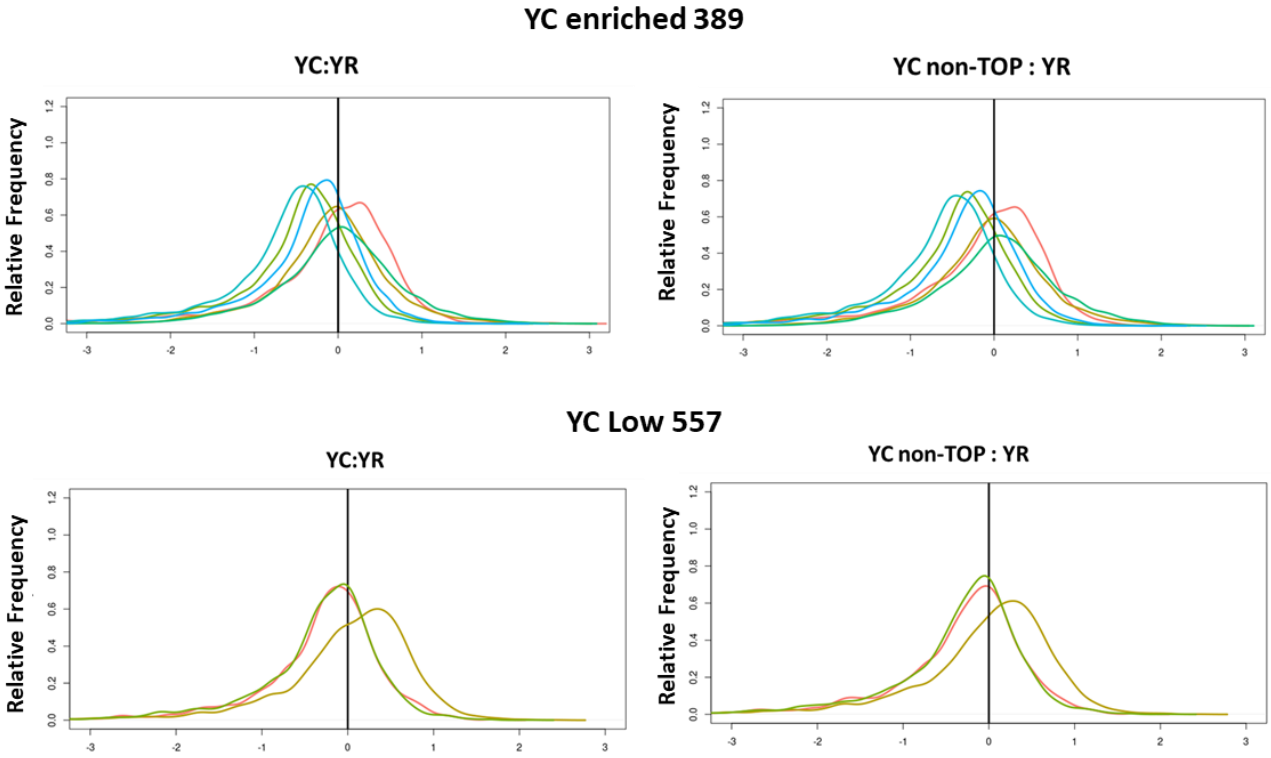


Figure 4.13. YC enrichment is retained upon removing TOP signal Comparison of YC:YR dynamics in YC only and YC-TOP initiators. Frequency distribution of YC:YR ratio per sample compared to pseudo bulked average. YC:YR expression ratio calculated per DUAL initiator per cluster divided by average YC:YR ratio for that promoter. Right indicates sub setting of genes to only include DUAL genes where the YC does not have the conventional TOP initiator.

To explore the penetrance of promoter usage heterogeneity and how the cluster bulk is built from single cell variation, we generated UCSC genome browser tracks of the top 50 cells with the highest TPM per each cluster. As many genes did not contain sufficient CAGE tags found in multiple individual cells in different clusters, I selected for genes which had sufficient capture and were known dual initiator genes. I compared two clusters with opposing dynamics from organoid 389; the translation GO cluster 0 (YC enriched) and Cluster 3 the DNA metabolism and repair cluster (YC low). I show that TSS usage differences are seen on both a single gene level and on a single cell level. **Figure 4.14** is a genome browser view of the pseudo bulked cluster (black) and 5 single cells within that cluster and their CAGE tags.

Firstly, this analysis shows the dominant single cell TSS usage is representative of the bulk cluster summed phenotype. Secondly, both base choices can occur in the same cell. In cell 2 of cluster 0 and in cell 3 of cluster 3 both YC and YR TSS are used highlighting for the first time that alternative TSS usage is not mutually exclusive.

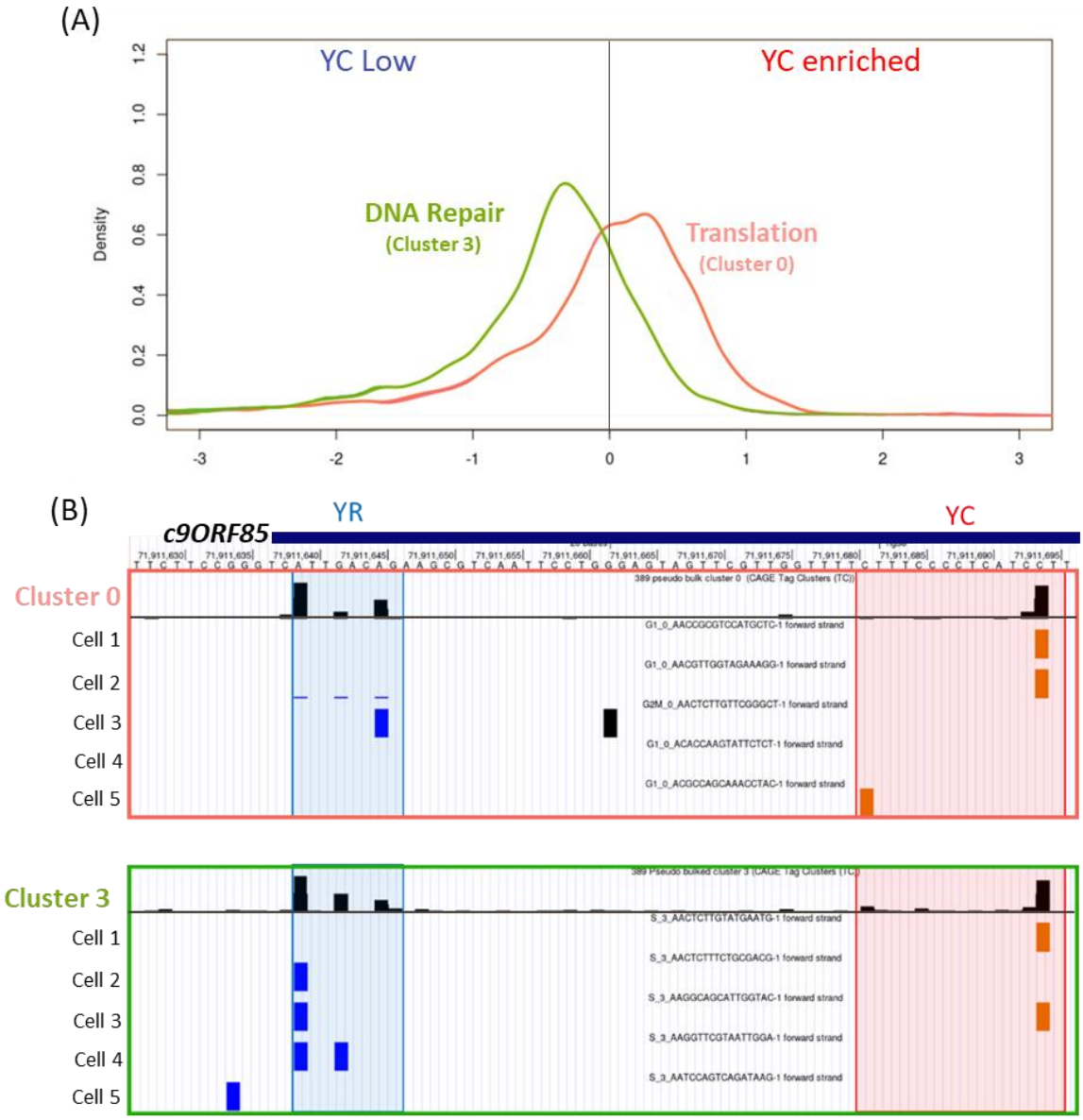


Figure 4.14. Genome browser view of single cell YC:YR usage of signature DUAL gene. (A) Frequency distribution of cluster 0 and cluster 3 with distinct YC enrichment and depletion. (B) UCSC snapshot of CAGE sequencing tracks of the signature dual gene c9ORF85. Cluster summed TSS TPM signals shown at top of each panel in black scale for both summed clusters is equal. Followed by 5 tracks of 5 representative cells in that cluster with YR initiation sites highlighted in blue and YC initiation in red. Single cell scales are all equal of 4.

Our analysis showed that in both lines the YC enriched clusters appear to share GO, **Figure 4.11** shows both samples have Translation, migration and a metabolic like cluster in the YC

enriched. Furthermore, the same is seen for the YC depleted clusters all associating with cell division in both samples. This raised the question as to whether TSS usage, specifically YC usage, correlates to a specific gene expression profile and hence physiological state independent of the global whole tumour phenotype. To begin to explore this we must first understand how similar the TSS defined clusters are between the samples. To achieve this, we grouped the clusters in both samples as either YC enriched, showing a marked deviation of their median point, or peak of plot, + 1 SD from the 0 in the frequency distribution plots (**Figure 4.11**) thus indicating a higher dependency and usage of YC in dual initiator genes. YC low clusters were defined as having shifted their median point -1 SD from 0, those not in this range were defined as no shift. The newly defined groupings appear to be grouped together, further supporting a similar gene expression profile underpinning TSS usage (**Figure 4.15**).

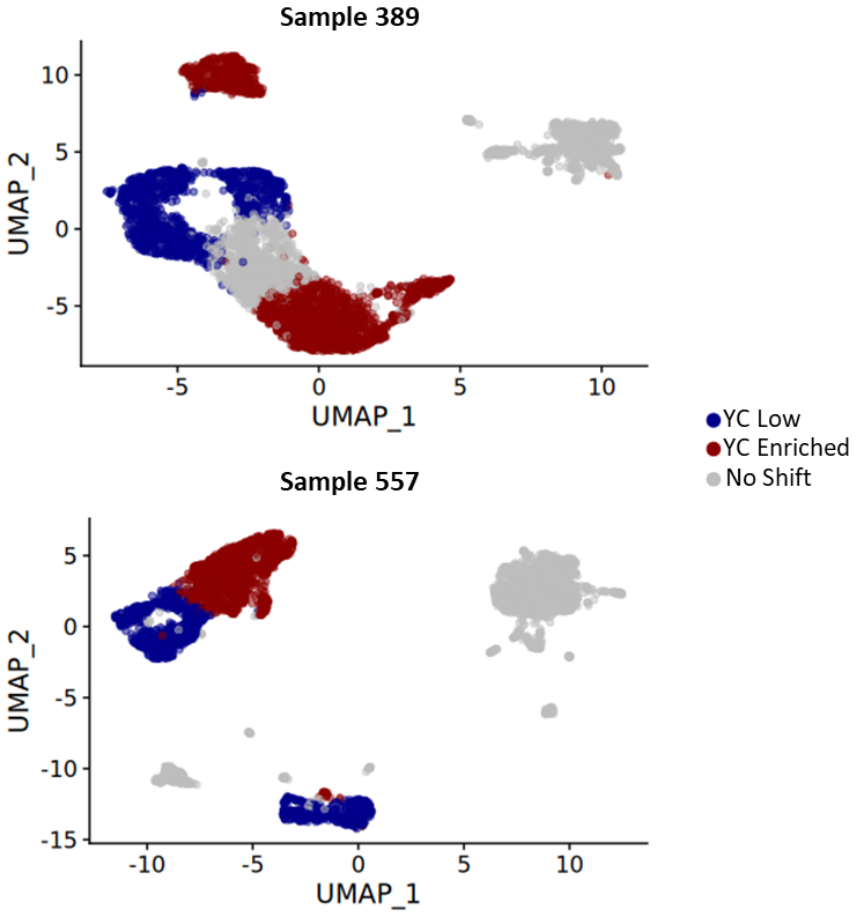


Figure 4.15: **Clustering visualisation of high YC and high YR clusters in both samples** UMAP projection of sample 389, high global YC TSS phenotype top, and sample 557, high global YR phenotype bottom. Coloured by cluster determined TSS profile characterised using YC:YR dynamic shifting of dual genes in frequency distribution analysis. Clusters with high YC usage in YC:YR coloured in dark red and clusters with high YR in dark blue. Those that did not show significant change are coloured in grey.

Next, I pseudo bulked the clusters based on these groupings and performed DGE analysis between the YC enriched population and YC depleted population within each sample using non-parametric Wilcoxon rank sum test through Seurats FindMarkers function. Differentially expressed genes were defined as those with whose log fold-change was greater than 0.6 for upregulated genes and less than -0.6 for downregulated and both where adjusted p-value, calculated using Bonferroni correction, was less than 0.05. The volcano plots of this analysis in **Figures 4.16** and **4.17** show of the top 10 most differentially expressed genes neither are shared between the samples. To investigate this further I performed GO analysis to identify the biological processes and Kegg pathways enriched in populations using ShinyGo0.80 with an FDR cut off of 0.05. I found that the low YC clusters in both samples were enriched in cell cycle associated genes and pathways. The YC enriched clusters share less similarity between the samples, sample 389 was predominantly associated with metabolic pathways whereas 557 shows enrichment mainly in cell motility with some metabolic pathways including glycolysis.

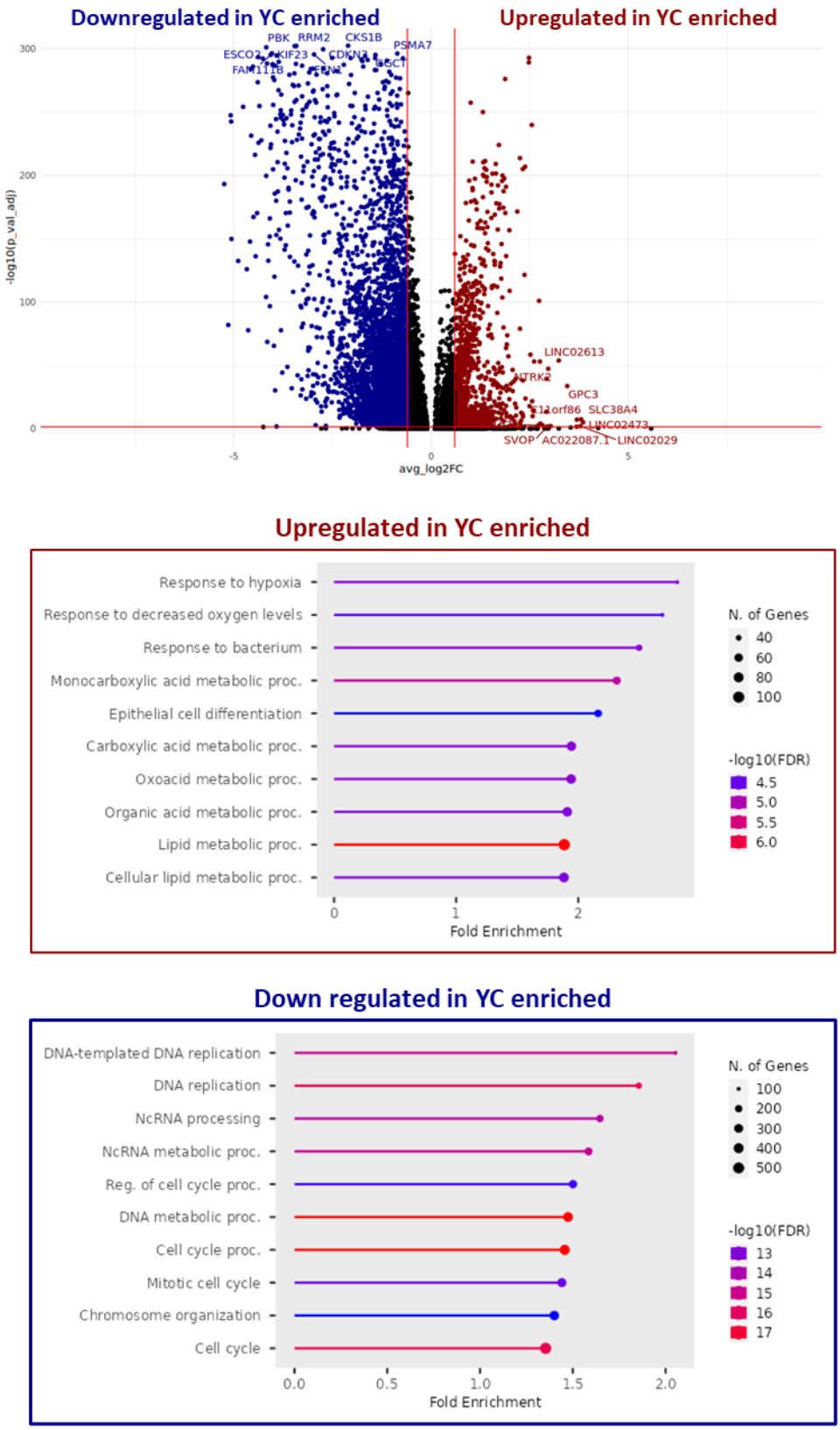


Figure 4.16: **Differential gene expression analysis of high YC cluster vs high YR cluster sample 389** (A) Volcano plot of DEG with top 10 genes both up and down regulated highlighted. Differential regulation determined as having log2FC greater than 0.6 (for upregulation) and less than -0.06 (for downregulation and adjusted p value less than 0.05. (B) Gene ontology analysis of down regulated genes in high YC clusters (Bottom and up regulated in YC clusters (Top).

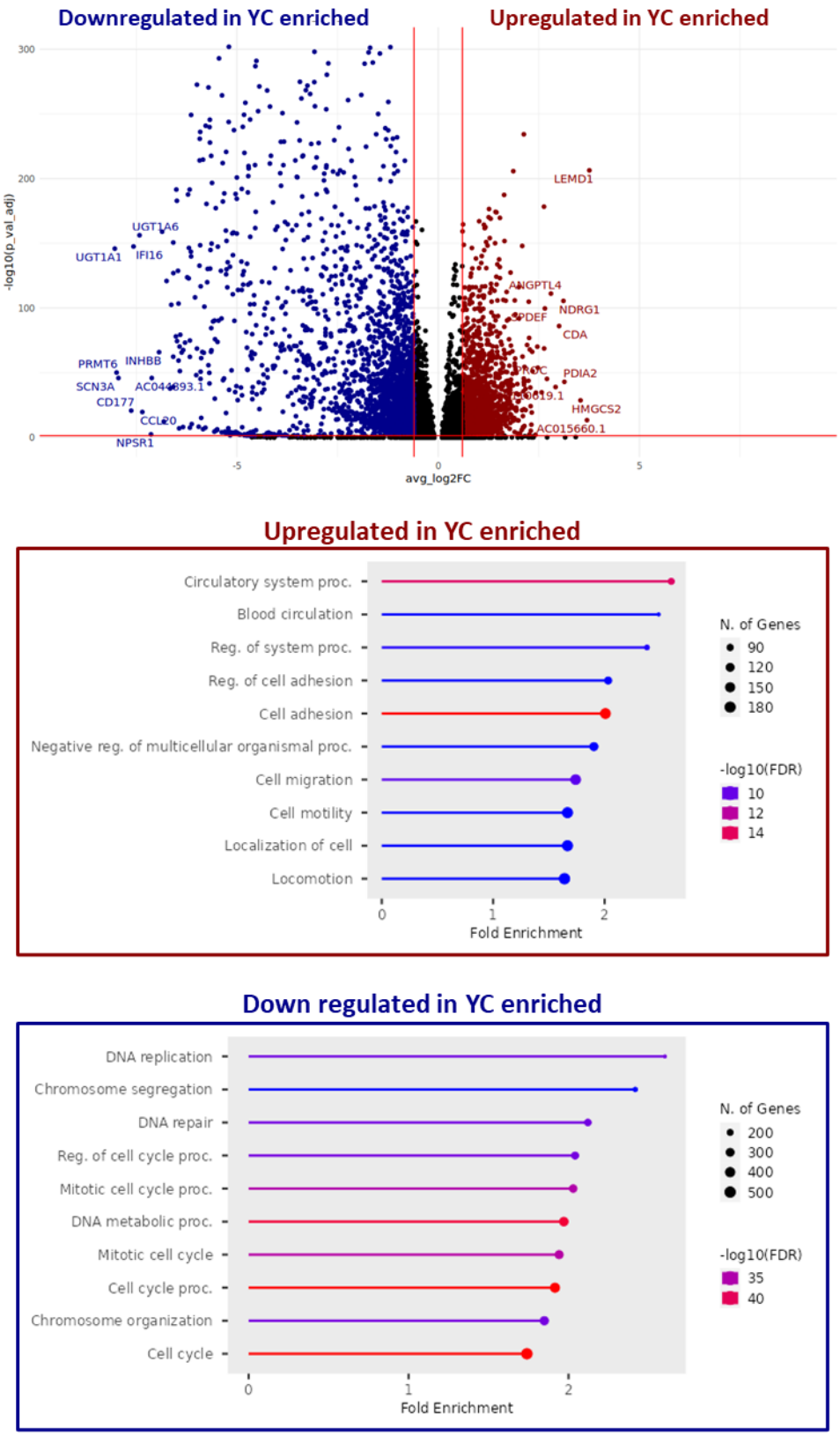


Figure 4.17: **Differential gene expression analysis of high YC cluster vs high YR cluster sample 557** (A) Volcano plot of DEG with top 10 genes both up and down regulated highlighted. Differential regulation determined as having log2FC greater than 0.6 (for upregulation) and less than -0.06 (for downregulation and adjusted p value less than 0.05. (B) Gene ontology analysis of down regulated genes in high YC clusters (Bottom and up regulated in YC clusters (Top).

The YC clusters appear to be enriched in genes associated with the predominant function and hence characteristic phenotype of that sample. Comparative analysis of the genes shows 18% are shared between the high YC populations in both samples, GO analysis of these shared genes shows upregulation in metabolic associated pathways including the reprogramming driving hypoxia response pathway **Figure 4.18**. The shared genes upregulated in the high YR cells in both sample shows further enrichment and validation of the cell cycle associated pathways. The continuous finding of YC depletion associated with cell cycle gene enrichment and hence cell cycle function, further highlights a TSS usage role in cell cycle dynamics that warrants further validation and investigation.

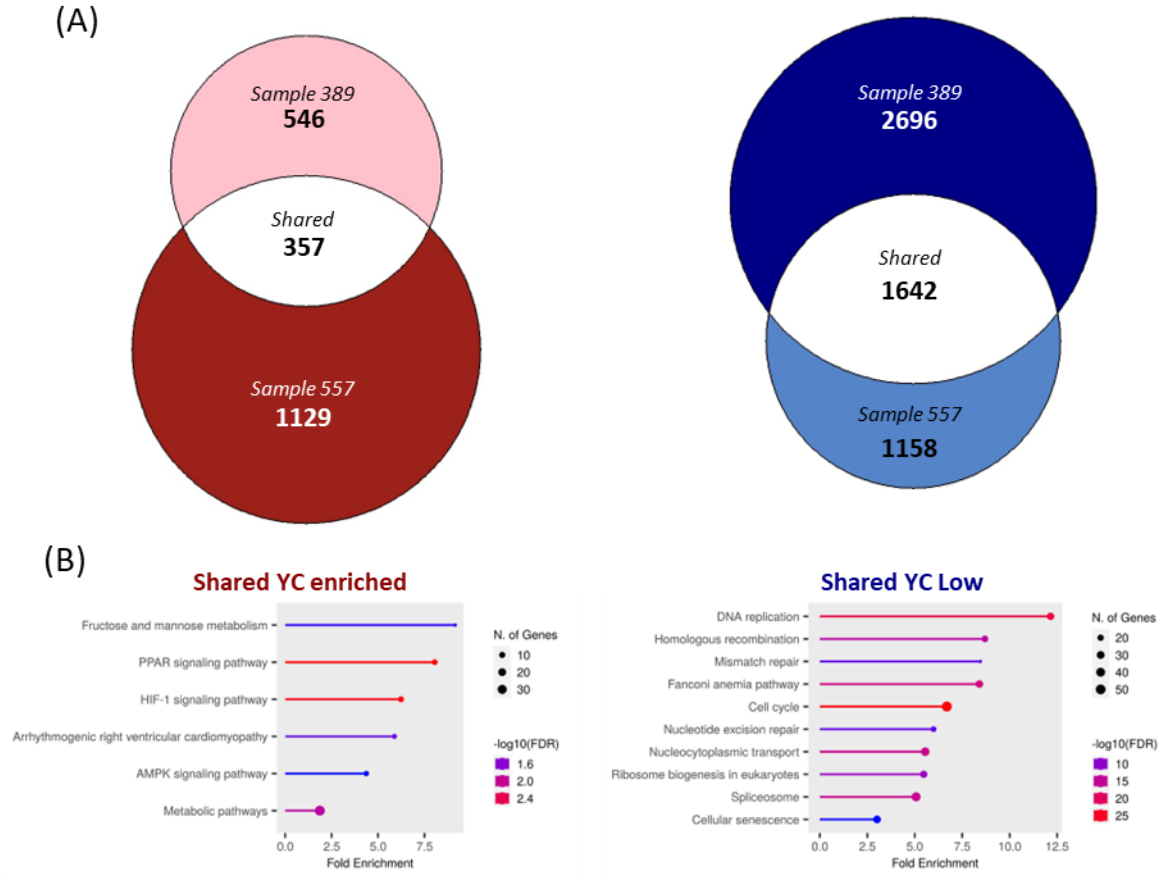


Figure 4.18: High YC and High YR clusters show similar GO between samples. Comparative analysis of shared enriched genes. (A) size proportional Venn analysis of the DEG in YC enriched clusters (Red) and YC low clusters (blue) separated by genes enriched in just sample 389 (top) and just sample 557 (bottom) with shared genes in the middle. (B) GO analysis of shared DEGs between samples.

4.2.4 YC shifting in cancer cell archetype states

A clear challenge appears when defining cellular states as the expression profile of cells and clusters of cells can be explained by a complex network of GO and therefore our previous conclusions of cluster definition by dominant GO maybe masking underlying phenotypes. Furthermore, the vast depth of GO possibilities makes it particularly challenging to compare between samples. The use of archetype analysis represents a growing computational approach to exploring ITH and the cellular states that exist (Hausser et al., 2019). The pan cancer analysis identified 5 universal archetypes allowing for convenient and informative comparisons between samples (**Figure 4.19a**). I aimed to use this analysis to gain better appreciation of the continuum of cellular states that exist allowing for better visualisation of the TSS usage in ITH at a single cell level. Using our polytope determination in Chapter 3, I projected the previously calculated cell YC:YR ratio average onto the 2D PCAs. **Figure 4.19** show that the high YC state shows a gradual transition to a defined state defined by marker genes associated with the lipogenesis universal archetype in both samples. As can be seen in **Figure 4.19c**, the high YC:YR cells are more dispersed showing that there are dissimilarities in their gene expression. When plotting proliferation score (**Figure 4.19d**) calculated using known marker genes of cell division from the GSEA (Subramanian et al., 2005), the opposite correlation to the YC:YR is seen. With the highest proliferating scoring cells existing at the furthest distance away from the highest YC indicating distinct gene expression differences.

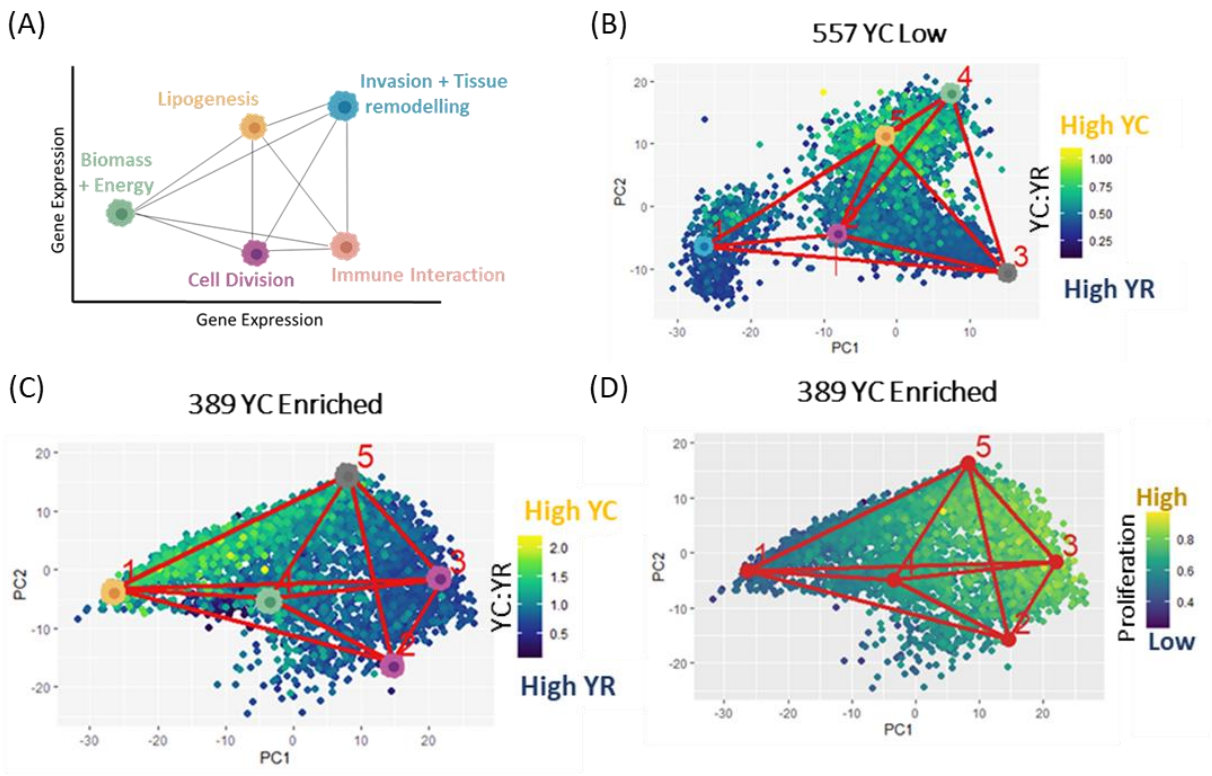


Figure 4.19: YC enrichment shows dominant archetype cell state preference. (A) Universal archetype polytope schematic. (B) Archetype polytope coloured by average YC:YR ratio per cell with vertices and hence dominant archetypes present coloured by universal archetype reference in schematic. sample 557 (C) and sample 389 (D) Archetype polytope coloured by proliferation score for sample 389.

As archetype analysis shows transitioning cells, meaning cells that are not a completely defined archetype and hence sit in between two archetypes, I aimed to see if YC enrichment also shows a gradual shifting as cells become a more defined state. Therefore, to further investigate the relationship of YC enrichment and the lipogenesis archetype depicted in **Figure 4.19c**, I extracted the PCA coordinate of this state. Using a distance based metric analysis using the archetype vertex coordinates and each individual cell coordinates, I ranked the cells based on distance from the vertex coordinate, explained in **Figure 4.20a**. I only included cells that did not exceed the midpoint of the polyhedral and thus not based near another archetype vertex. These centre cells represent the more undefined stem-like states (Hart et al., 2015; Hausser et al., 2019). Plotting the cell rank distance from the vertex to against its average YC:YR ratio shows a strong negative correlation representative in **Figure 5.20c** with Pearsons coefficient value of -0.55. This indicates that as cells become more

defined into this lipogenesis state, they have a higher dependency on YC usage in dual gene.

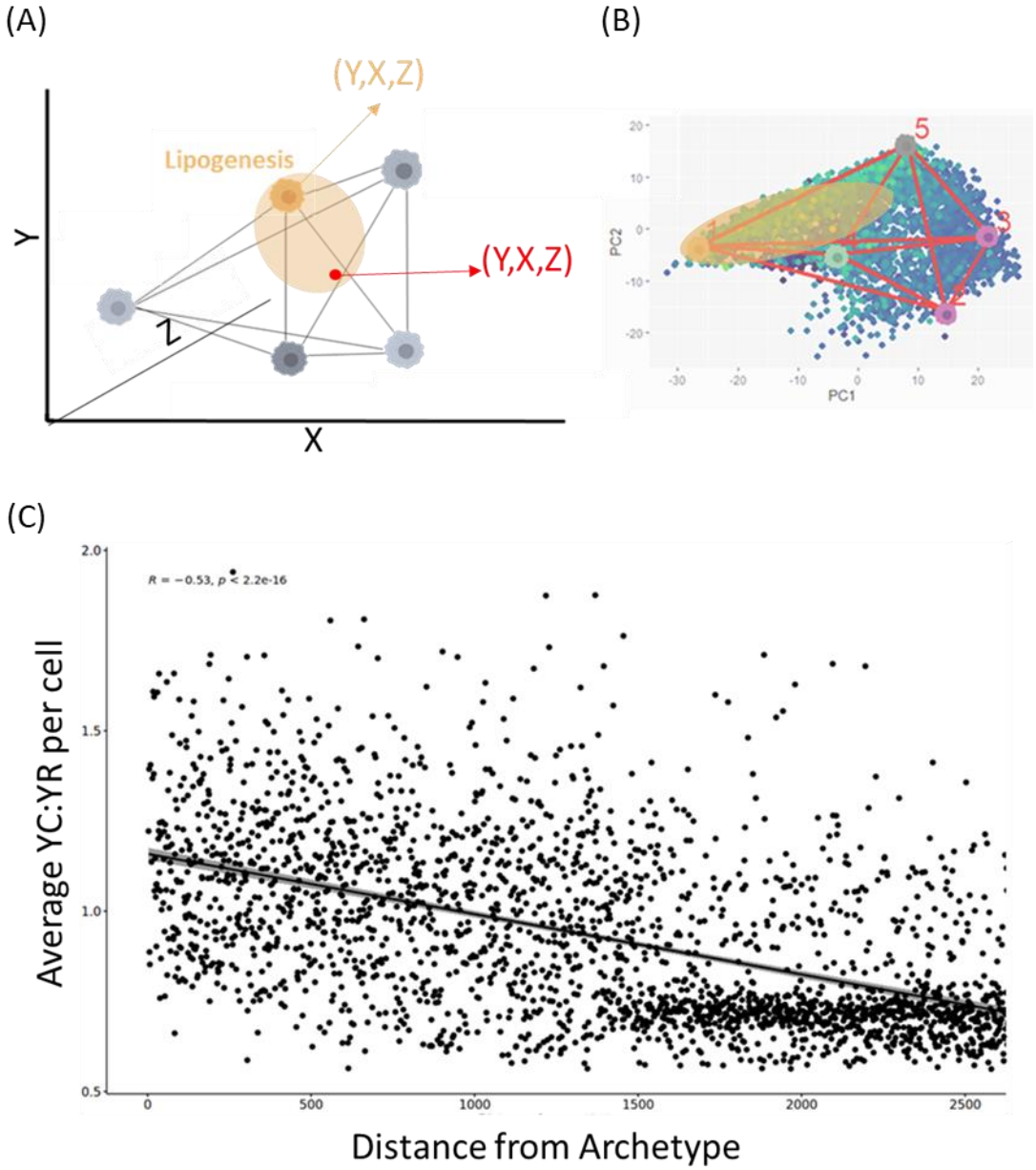


Figure 4.20: YC enrichment shows gradual shifting as distance from lipogenesis archetype increases. (A) Schematic showing how distance base analysis of archetype transitioning cells. Coordinate of defined lipogenesis state identified and the coordinate of the middle point of the polyhedral (point equidistant from all archetypes indicating an undefined cell). Ranking of cells based on their PCA derived coordinates and distance from the archetype coordinate. (B) overlay of distance schematic onto the 2D archetype polytope used in the analysis. (C) Scatter plot of cell average YC:YR ratio and distance from high YC lipogenesis archetype vertex.

Hausser et al., 2019 state that the universal archetypes share a large proportion of marker genes, given the appreciation that many genes form a complex network of context dependent interactions and thus their expression may not be attributable to just one function or hence GO. The analysis in **Figure 3.14** also showed that many of these shared genes were dual initiator genes meaning they harbour base choice-dependent transcript abundance. As there are shared genes between these archetypes it raises the question of whether differential TSS usage maybe a distinguishing factor between archetypes. Do we see a shift in TSS usage in these shared dual initiator genes dependent on the cell's archetype state? To investigate this, I aimed to identify cells attributable to each of the universal archetypes. Using the defined marker genes in Hausser et al., 2019, I scored each cell for each archetype and assigned the cell identity on the highest score obtained, I projected these identities onto previously generated cluster UMAPs to see how they compared (**Figure 4.21**). Interestingly, we see more presence of different archetype throughout the previously assigned clusters. It must be noted that this approach does not allow for the 'undefined' cells that would exist in normal archetype analysis and would be seen central in the polytope distinctively distant from defined vertices (Hart et al., 2015; Hausser et al., 2019). This approach assumes identity to its closest archetype. To identify whether YC enrichment was attributable to a universal archetype state **Figure 4.21** shows the YC:YR average per cell grouped by archetype identity. This analysis found across both samples the lipogenesis archetype has the highest median YC enrichment further supporting our previous ParetoTI driven archetype analysis in **Figure 4.20**.

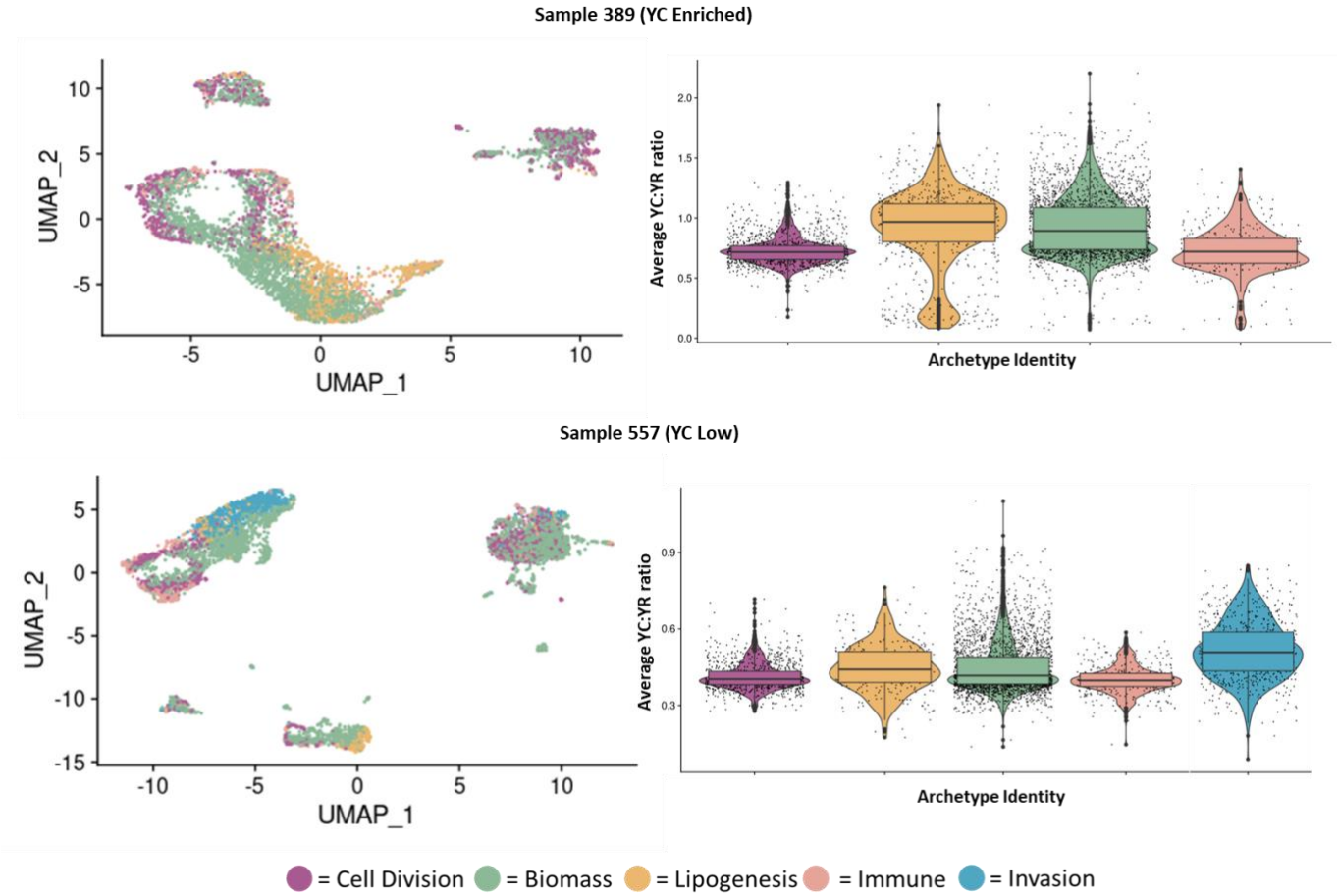


Figure 4.21: Universal archetype states show differing TSS usage. UMAP cluster coloured by cellular universal archetype identity. Average YC:YR per cell grouped by universal archetype identity. Violin plots of grouped data overlayed showing median and upper and lower quantile of data distribution.

To see if the YC:YR shift between the archetypes is a result of shared dual initiator genes, I pseudo bulked the samples based on archetype and extracted cell barcodes, collated CTSS to perform CAGEr analysis workflow on each archetype. As previously highlighted in **Figure 3.14** the Biomass and Cell Division archetypes share the largest number of dual initiator genes. Furthermore, the biomass archetype state shows YC enrichment compared to the cell division archetype which shows global TSS usage is more YR dominant **Figure 4.21.** I repeated the previous frequency distribution analysis of the YC:YR per CCID per archetype / the average YC:YR per CCID however this time only including the genes shared between the two archetypes. The GO of these shared genes is outlined in **Figure 4.22b** and shows two distinct groups of biological processes. Interestingly, when comparing the promoter TSS usage in shared marker genes we see a YC usage shift dependent on the cell archetype or state (**Figure 4.22a**).

As many of the YC:YR analyses so far have focused on cellular states or clusters that are

defined by gene expression profiles and hence comparisons will be affected by DEGs. This makes YC:YR dynamics difficult to interpret as it is unclear if they exist between subpopulations because of DEGs. However, by comparing genes known to be upregulated in both archetypes it allows us to potentially rule out this bias and explore a potential role for alternative TSS shifting in multifunctional genes. **Figure 4.22a** shows that as cells enter a biomass cell state, they shift to a higher YC usage in these shared genes. Contrastingly when cells enter a cell division state YR TSS are more dominantly used. The GO of these shared genes (**Figure 4.22b**) shows two groups. Firstly, translation and associated pathways including many translation initiation and elongation factors and secondly metabolic pathways. This highlights a potential base choice regulatory role in cellular state determination in these genes. Further investigation is needed into the differential regulatory mechanisms of these TSS and potential surrounding CRE It also highlights a transcriptional level regulatory mechanism of translation associated genes beyond just TOP usage.

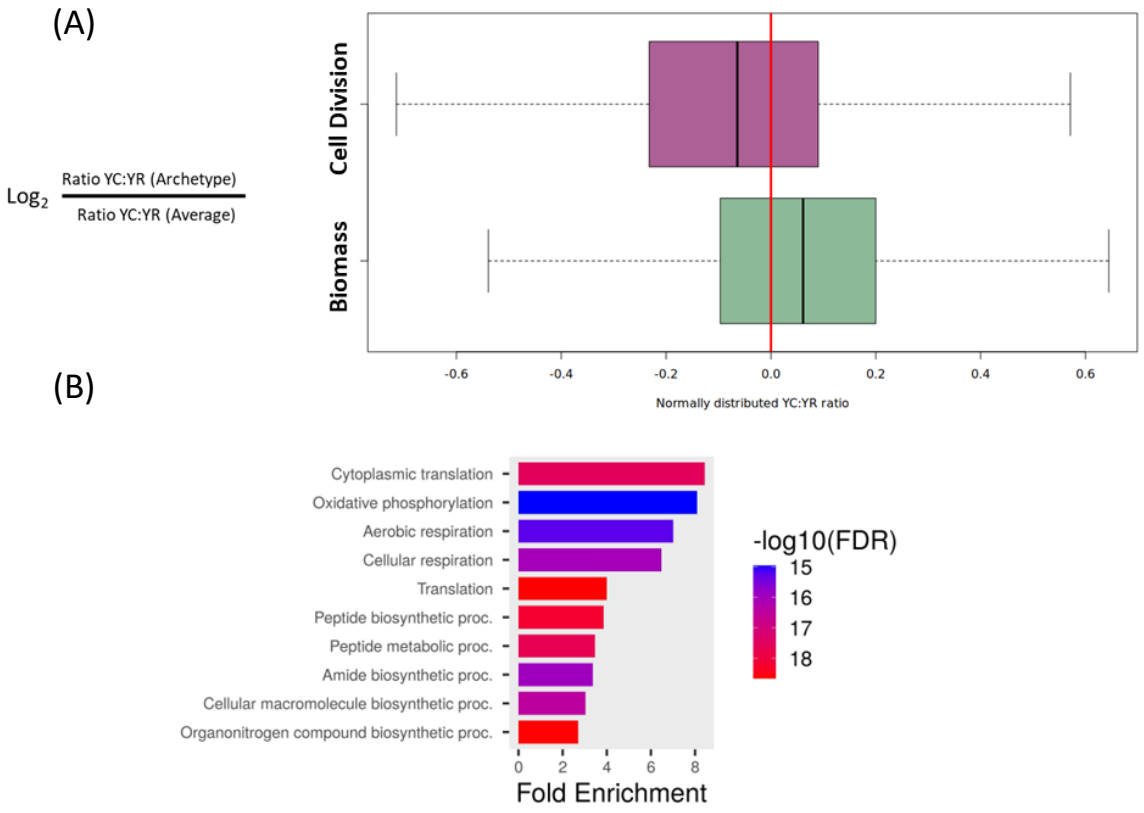


Figure 4.22: **YC:YR projection onto Archetype distribution plots.** A) Frequency distribution analysis of YC:YR ratio per archetype pseudo bulk compared to total pseudo bulked average including only shared enriched gene. (B) Gene ontology of the genes that show shared enrichment between the biomass and cell division archetype.

Shifting into a more conservative or energy depleted state is vital in metabolic reprogramming and is mediated by nutrient sensing pathways that interact with TF that upregulate genes that enable this process (Cerezo & Rocchi, 2020; Faubert et al., 2020; Phan et al., 2014). A core TF in this process shown to be activated by both the mTOR pathway in growth associated driving and separately by stress in nutrient depleted environments is ATF4 (Linares et al., 2017; Torrence et al., 2021). We found that ATF4 was a shared marker gene between the cell division and biomass archetypes. Furthermore, our analysis highlighted it as a dual gene that showed distinct dynamic shifting from YR to YC in cancers compared to healthy and its YC usage correlated to irradiation responsiveness (Wragg et al., 2023). Like other TF, the relevant YC and YR TSS of ATF4 are within a few base pairs of each other **Figure 4.23**. When we compare the relative YC:YR usage between the two archetypes at this gene we see a slight enrichment of YR in the cell division states however a significant depletion in YC compared to the biomass state. We hypothesise that this shifting could be a result of biomass states requiring an upregulation of ATF4 to aid in metabolic intermediate production. The higher usage of YC may represent a higher dependence of the mTOR driven cap-dependent translation. Contrastingly given ATF4 role as a sensor, it could also be hypothesised the shift from YC may simply be a result of nucleotide bioavailability being higher in biomass states.

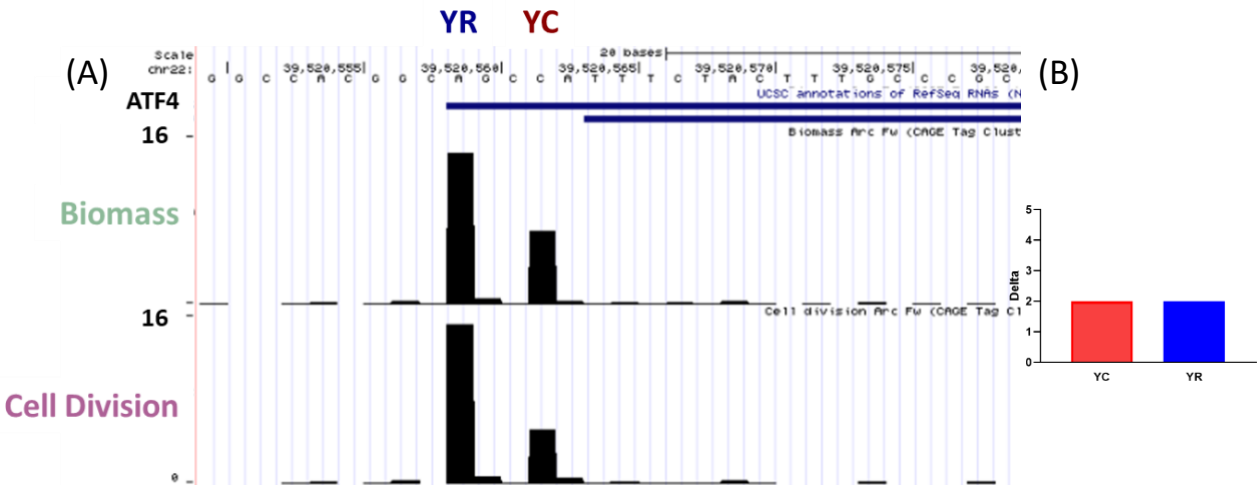


Figure 4.23: ATF4 shows state dependent TSS usage. UCSC genome browser view of ATF4 gene and TSS specific TC for each archetype state. YR initiator distinguished by a 'CA' TSS and YC by a 'CC'. (B) The delta change of TSS usage between the archetype shows equal change. Total signal between the samples are equal but YC has increased in the biomass archetype to the same proportion that its YR has decreased when compared to cell division.

4.2.4.1 Hidden initiator motifs

An additional critical TF mediated nutrient sensor is SREBF1 and is essential for controlling lipid and amino acid metabolism. Recent studies have also shown the importance of lipogenesis in cancer progression and directly linked *SREBF1* expression levels with proliferation rate (Ricoult et al., 2016; Y. A. Wen et al., 2018). SREBF1 is also a shared marker gene for both cell division and the other metabolic archetype lipogenesis. Although this is a dual gene, we see TSS usage predominantly associating with YR, however the two dominant TSS are separated by a distinctive TOP motif meaning only initiation occurring from the first YR, depicted on the far right of the browser view in **Figure 4.24** as this is a reverse strand gene, will contain this motif. It is speculated that although initiation does not occur from YC, the presence of an internal top, especially one so close to the initiator, could still be targeted by TOP specific regulation. Interestingly we see a decrease in the first YR usage between the states however YR initiator proceeding the TOP motif remains the same. As the predominant TSS are YR our YC:YR ratio metric would miss this interesting dynamic and thus warrants further investigation into the internal TOP containing motifs and whether their regulation is targeted by LARP1. The decrease in the TOP containing motif of the cell division state could be a result of the upregulation of mTOR driving growth and division and hence these TOP containing transcripts being translated in this state. Contrastingly the enrichment of this transcript in the metabolic state could be resulting from LARP1 binding and hence stabilisation. Furthermore, the TOP containing motif has an A initiator, it has been shown in literature that methylation can occur at these sites and enable the bypassing of cap-dependent translation thus allowing for their continued translation in stress blocked mTOR states (Meyer et al., 2015). The preferential use of this initiator in a metabolic state associated with reprogramming could also result from this.

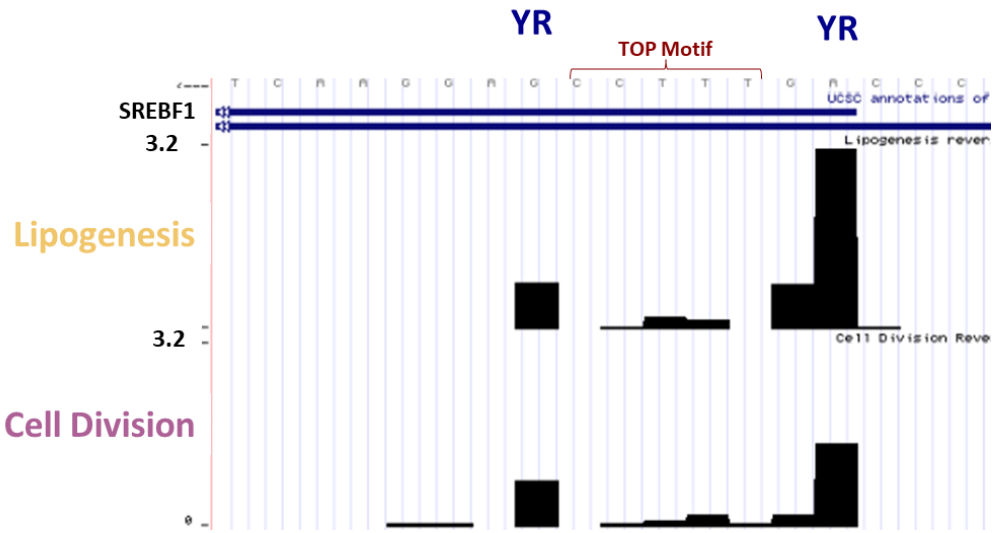


Figure 4.24: SREBF1 decreases TOP motif transcript production in lipogenesis state. UCSC genome browser view of SREBF1 gene and TSS specific TC for each archetype state. Reverse strand direction shows two YR initiator distinguished by a 'CA' TSS and a 'CG' separated by a TOP polypyrimidine motif 'TTTCC'.

To ensure that this finding wasn't a result of a bias that can occur with specific CAGE sequencing specifically the G bias that occurs in the template switching approach we used in our experiment, I compared the peaks against our bulk and the Fantom 5 datasets. We validated that this is a universal captured TSS and appears to show a potential colon specific usage.

4.3 Discussion

The work in this chapter begins to highlight the functional relevance of TSS usage even to single cell resolution through its association to cellular states. Although previous work in our lab and in the previous chapter had identified a global phenotype associated with high YC enrichment we still didn't understand the biological significance beyond its potential growth role (Wragg et al., 2023). This is because bulk whole tumour analysis masks the variability that exists within a tumour (Lawson et al., 2018). Therefore, identifying functionally relevant genes is challenging given the variation of gene expression that occurs within.

Alternative TSS usage represents a new level of ITH.

In this chapter I validate our single cell TSS capture success and show that despite global TSS usage phenotype, ITH of TSS usage exists. Initial variability observed between matched bulk sequencing was resolved upon removal of ribosomal genes (**Figure 4.1**). Although in principle combining the individual single cell data should give representative analysis of the whole tumour, and thus be comparable to the actual matched bulk. Sequencing often has a bias towards ribosomal genes because of their high expression in comparison to other genes (Subramanian et al., 2022). Therefore, combining the individual cell data, we are increasing the ribosomal bias and inadvertently increasing YC bias.

Our initial attempt to explore the TSS usage at a single cell level was to use our previously defined metric of YC:YR ratio. However, analysing this ratio at a single cell, single gene level proved challenging as molecule count per CTSS was very low. This meant few CC were identified across genes and across sufficient cells to provide comparisons. However, mining of genes with sufficient data is possible and is a future analysis that will be attempted. To overcome the low level of single cell gene capture we used the average ratio per all dual genes in each cell. Visualisation of single cell TSS phenotype on our UMAP showed that the high YC cells predominantly cluster together in both samples (**Figure 4.3**). This highlights an underlying gene expression profile underpinning this TSS phenotype and a potential functional cellular advantage to YC usage.

Cellular physiological states show differential YC enrichment.

Cellular physiological state provides invaluable insight into relevant function of a cell and is characterised by a defined gene expression profile underpinned by a complex regulatory network spanning multiple layers (Kotliar et al., 2019). Vast efforts have been underway to better our understanding of the genetic, epigenetic and even transcriptomic profiles that are involved in determining cell function and hence state and how these are exploited in cancer (H. Jin et al., 2023; Kotliar et al., 2019; Trapnell, 2015). Understanding these mechanisms paves the way for both biomarker discovery and therapeutic targets as interesting correlations of ITH cell states and behaviours including treatment response are emerging (Caiado et al., 2016; Stanta & Bonin, 2018b). Here I aimed to begin exploring if TSS base choice variation is one of the transcription level regulatory mechanisms.

Pseudo bulking by the previously generated clusters in Chapter 3 (**Figure 3.13 and Figure 3.14**) allowed us to identify phenotypes associated with high YC shared across both

samples. Despite the sheer phenotypic differences I identified between our two samples, I initially saw homogeneity between the clusters that showed high YC (**Figure 4.11**). As expected, the GO of the high YC clusters represented translation in both samples. As previously stated, the predominant gene expression in this experiment was attributable to ribosomal genes (**Figure 3.7c**). The high bias will significantly influence data and using our approach of defining state based on the dominant GO, could be masking the underlying expression profiles and other relevant biological processes attributable to that cell. This was validated when we compared the gene expression profiles of the high YC clusters between the samples and found they only shared 18% of upregulated genes (**Figure 4.18**), meaning other biological processes beyond translation were underpinning these clusters. Contrastingly however, we identified YC appears to be significantly lower in cells enriched in cell cycle associated genes in both samples representing a potentially previously unexplored transcriptional level regulation of the cell cycle.

TSS usage differs between universal archetypes.

The identification of universal archetypes in cancer has highlighted shared states consistently found in cancer that show clinically relevant phenotypic presentation associations (Hausser et al., 2019; Heiser et al., 2012). Therefore, archetypes represent an opportunity to use them as reference states in our samples to make more intuitive comparisons. Firstly, I showed that the results support current literature with the more late-stage aggressive archetype populations, the invasion and immune, being present mainly in the non-responsive differentiated line (**Figure 4.19**) (Groves et al., 2021; Hausser et al., 2019). Previous archetype analysis in cancer had relied on tumour sequencing, however our findings show that organoids also sufficiently recapitulate evolutionary and phenotypic dynamics of tumours.

Characterisation of the TSS usage attributable to archetype states found that in both samples the lipogenesis archetype showed a distinctive YC enrichment (**Figure 4.21**). This functional state has been characterised in literature as being involved in production of fatty acids and is an important metabolic regulating process (Hausser et al., 2019). Further interrogation of the genes underpinning the lipogenesis state show significant involvement in protein anabolism and catabolism. Lipogenesis is a multifunctional process widely exploited in cancer to ensure a proliferative state is retained through sufficient production of metabolites and even ATP (Koundouros & Poulogiannis, 2020; Mounier et al., 2014; Rysman et al.,

2010). The regulatory mechanisms of these genes have been widely studied to identify potential therapeutic targets that will result in inhibiting tumour growth (Guo et al., 2015; Vettore et al., 2020). Many studies have identified the role of *SREBF* the highly activated TF that plays an essential mediator role interconnecting the oncogenic signalling including PI3k and metabolic reprogramming (Guo et al., 2015). Interestingly I have identified many of the nutrient sensors including *SREBF* are dual genes. Furthermore, I have identified that we see differential TSS usage, represented by higher YC usage, in these TF in metabolic states (**Figure 4.23** and **Figure 4.24**)

The challenge when comparing cellular states is they are often characterised by distinctive gene expression profiles. Therefore, it is unclear if the YC enrichment is simply a result of an increase in transcription and hence expression. However, archetype analysis appreciates cellular state is more of a continuum. Furthermore, universal archetypes share marker gene expression as many genes have multiple functional roles that are often environmentally context dependent (Hausser et al., 2019). I previously showed that the biomass and the cell division archetype share the largest amount of marker genes (**Figure 3.14**). We hypothesise that this is due to the cellular survival equilibrium that exists and is dynamic in cancer cells. Cellular growth and hence division is intertwined with and highly dependent on the metabolic state of the cell (Feitelson et al., 2015; Nong et al., 2023; Phan et al., 2014). In this chapter I showed TSS changes in genes with roles in two contrasting states: metabolism and cell division (**Figure 4.22**), with the YC usage associating with the metabolic states. The GO of the shared genes reveals hypoxia response genes. Many of these genes play a vital role in the metabolic reprogramming that occurs in cancer, similar to those identified in the lipogenesis state (Ho et al., 2020; Van Den Beucken et al., 2011). However, these genes are also essential for maintaining redox balance in stress environments often widely present in cancers (Ho et al., 2020). A key initiator of upregulating these genes is the nutrient sensing TF ATF4. ATF4 has been shown to be involved in both the driving of growth process through its interaction with mTORC1 and mediating cellular stress gene response through nutrient sensing (Ebert et al., 2022; Kreß et al., 2023; Torrence et al., 2021). Interestingly we see like the case with SREBP1, ATF4 is also a dual gene, and its expression remains high in two contrasting states, cell division and biomass and energy (**Figure 4.23**). TSS usage however and hence YC enrichment differs between them. This highlights a potential association of TSS usage and maintenance of the proliferation to metabolism

equilibrium. Recent research has highlighted an ATM that occurs when the mTOR driven cap-dependent translation is inactivated (Ho & Lee, 2016; Tamarkin-Ben-Harush et al., 2017). This stress induced pathway allows for translation of genes vital to the proliferation to metabolism equilibrium during stress environments (Tamarkin-Ben-Harush et al., 2017). The presence of dual TSS at these genes highlights a potential transcription level regulation whose translation could be maintained even in stress environments dependent on their 5' end motif. Further work is needed to see if shifting to a different TSS in these genes in stress. Direct interference experiments to test the hypothesis that YR transcripts that lack TOP or YC motif are continued to be translated in these stress environments unlike the YC transcripts, are needed.

Many of the nutrient sensing genes vital for both maintaining the proliferation phenotype and metabolic reprogramming and sustenance during stress harbour dual TSS. Interestingly, the distance between the relevant YC or YR motif are significantly shorter in these TF than many other dual genes identified previous analysis by Wragg et al (**Figure 4.23 and 4.24**) (Wragg et al., 2023). We hypothesise that differential usage could result from nucleotide bioavailability rather than other regulatory differences. Given the close proximity of the TSS it is unlikely they will be targeted by alternative CRE or conformational changes. Experiments that directly alter purine and pyrimidine levels in cells and then analysis of relative TSS usage would allow this hypothesis to be further explored.

In summary, I show that YC enrichment is correlative with metabolic reprogramming genes that are widely upregulated in cancers. Although our findings are currently through association, it has fuelled an ongoing hypothesis that YC usage may enable the production of sufficient metabolic intermediates required to maintain the hallmark proliferative state in tumours. We postulate this as the translation of YC transcripts is shown to be regulated by growth signal pathways specifically the mTOR pathway and the resultant translation has shown to have a higher efficiency (Borden & Volpon, 2020; Schneider et al., 2022). It therefore raises the possibility that YC translation may represent a potentially more efficient way of protein production. When cells meet the threshold required for proliferation and cell division begins, these transcripts are translated resulting in lower YC. Lower YC in cell division may also represent a shift is to a more controlled low translation state to sufficiently retain energy demands required for cell division. Although we cannot currently answer this

definitively, we have begun to identify associative trends. Firstly, differential YC usage in the sub populations as previously mentioned but also across cancers whereby the more proliferative organoid has a higher YC enrichment. Given our ITH findings we can now hypothesise that YC usage plays a role in retaining the proliferative state of these lines.

A further association can be made when considering the YC transcripts are heavily reliant on mTOR regulated translation, such pathways are extremely receptive to external stress stimuli and become inactivated (Csibi et al., 2014; Thoreen et al., 2012; T. Tian et al., 2019; Zou et al., 2020). Therefore, if a tumour is heavily dependent on YC usage and enter a stressed state they are likely to be unable to translate the cap-dependent YC transcripts. Contrastingly the slower growing organoids showing a YR dependency appear to respond better to stress. Future work is needed to understand the potential link between TSS usage and the ATM that has recently been elucidated in hypoxia and stress induced states. Furthermore, as we show that shifting to YR usage correlates with the enrichment of cell division genes, this highlights a potential cell cycle level regulatory process likely enabling the proliferative capacity of the cell, such hypothesis is further explored in the following chapter.

Chapter 5: Mapping TSS usage in cell cycle dynamics

5.1 Introduction

The previous chapters demonstrated YC enrichment associated with increased growth and proliferation, but low YC was associated with enrichment of cell cycle gene expression. We hypothesised these contrasting results may indicate a potential TSS regulatory role in cell cycle phases. The cell cycle and its regulation are central in cancer initiation, progression, and response (Hanahan & Weinberg, 2011). Unsurprisingly it is one of the most investigated phenomena in cancer research and is one of the most frequently targeted cellular processes in anti-cancer therapy (Castelli et al., 2021; Feitelson et al., 2015; Sazonova et al., 2024). As discussed in Chapter 1, maintaining the proliferative phenotype and transition into quiescence in response to environmental pressures is a vital equilibrium for cell survival and is also a marker for cellular differentiation (Basu et al., 2022; Castelli et al., 2021; Dalton, 2015).

Core promoter architecture differences have already been identified in differentially cycling genes in our lab. Previous work from Dr Joseph Wragg identified that promoter usage became sharper as cells became more differentiated and slower cycling (Wragg et al., 2020). Although this sharpening showed TSS usage reduced to a small number of bases, they did not establish specific TSS initiator class usage. Therefore, TSS usage dynamics have yet to be explored in the context of cell cycle phases. In this chapter, I aim to see if the ITH TSS usage we see in our organoids can be resolved by cell cycle phase differences.

So far, we have identified a relative proliferative state association to TSS usage, however we are currently limited to single time points that prohibits temporal analysis. We are yet to appreciate the evolutionary dynamics of cellular states such as cycling phasing in organoid development and how these cycling dynamics differ between two differentially growing organoids. Spatiotemporal resolution of cellular physiological states will allow us to potentially understand better the role of TSS usage in tumour growth. Organoids sufficiently recapitulate tumour heterogeneity and spatial organisation (Rios & Clevers, 2018; Sayed et al., 2021; Z. Zhao et al., 2022). Furthermore, previous archetype analysis in Chapter 4 shows

retention of evolutionary dependent ITH in our organoids. Therefore, they represent an exciting model to explore cellular physiological states and TSS usage in tumour formation and to interpret cellular genomics data with imageable phenotypes.

In this chapter I aim to develop a longitudinal high-resolution imaging protocol for 3D cultures that can be utilised to resolve cell cycles. A recognised ongoing challenge of using organoids is the inability to truly capture their whole structures using live imaging due to the complexity of their culturing conditions (Alladin et al., 2020; Hof et al., 2021; Serra et al., 2019). Often benchtop light microscopes do not provide sufficient resolution to capture the full tumour and requires a more powerful high-resolution scope. The lightsheet microscope enables Z stack scanning and has minimal illumination plane, which limits bleaching and enables timed lapse imaging (Oksdath Mansilla et al., 2021). The conventional Z1 lightsheet apparatus was deemed inappropriate for live organoid imaging and later resulted in the development of a microscope specifically designed for organoid imaging (de Medeiros et al., 2022; Oksdath Mansilla et al., 2021; Rios & Clevers, 2018). However, the costing of such equipment and its limitations to organoids makes it an often-unjustified purchase. Therefore, I aim to explore whether the Z1 lightsheet imaging protocol previously designed for Zebrafish imaging, can be adapted to organoids, to enable spatiotemporal mapping of physiological states in organoid growth.

5.2 Results

5.2.1 YC initiation usage correlates with cell cycle phase

5.2.1.1 Assigning cell cycle phase identity

Our initial findings showed YC enrichment correlated with fast proliferative global phenotype. However, Intra-cluster analysis in Chapter 4 (**Figure 4.11**) consistently found YC was lower in cells enriched in proliferative genes associated with cell division. This contrasting finding raised our hypothesis that there was a cell cycle phasing role underpinning these results. Therefore, to see if the ITH in the organoids could be better characterised by cell cycle phases, I labelled individual cells using the well characterised cell cycle scoring function in the Seurat analysis, utilising the previously identified cycle phase specific marker genes (Kowalczyk et al., 2015). This analysis allows us to identify cells in the G2M and S phase, cells scoring negatively are assigned as G1 phase. Since there are no

universal markers for G1, given the variation in expression profiles dependent on the cell state, we cannot distinguish them from G0. Single cell resolution of cycle phase identity was labelled on the UMAPs for both organoids (**Figure 5.1a**). Despite the organoids showing differences in growth dynamics (**Figure 3.1**), we don't see a dominant cycle phase difference between them with the largest population in both being G1/G0 cells (**Figure 5.1b**). A recent paper highlighted the significance of replication dynamics on gene expression profiles. These dynamics feature creates a distinctive cycling pattern on UMAPs and interlinks the intricacy of transcription replication interaction (Pountain et al., 2024). Interestingly we see a similar cyclic pattern in our UMAPs, represented by black arrows in **Figure5.1a** highlighting that these patterns coincide with cell cycle phasing further emphasising transcriptional cycling effects.

Studies have shown that a significant decrease in transcription occurs following the G1 to S transition to retain energy for the remainder of the cycle and division process and to limit replication stress (Riba et al., 2022; Ryl et al., 2017; J. Wang et al., 2021). Our initial quality control analysis of our sequencing data in Chapter 3 (**Figure 3.7a**) highlighted a bimodal split in our cells with two distinct populations showing different read and feature count per cell. We hypothesised that the potential cause of this was cell cycle dependent. Feature plots in **Figure 5.1** show that this is unlikely. Although we still see a high expression count indicating the likely well documented G1 transcription wave, other phases show a large number of cells in this high feature count population. It could be argued that this is resulting from the scoring system is binary and cycling is dynamic so cells could be just entering or transitioning away from their assigned phase therefore is not truly representative. Furthermore, distinct differences in the bimodal split are seen between the two organoids, could highlight a potential technical capture bias.

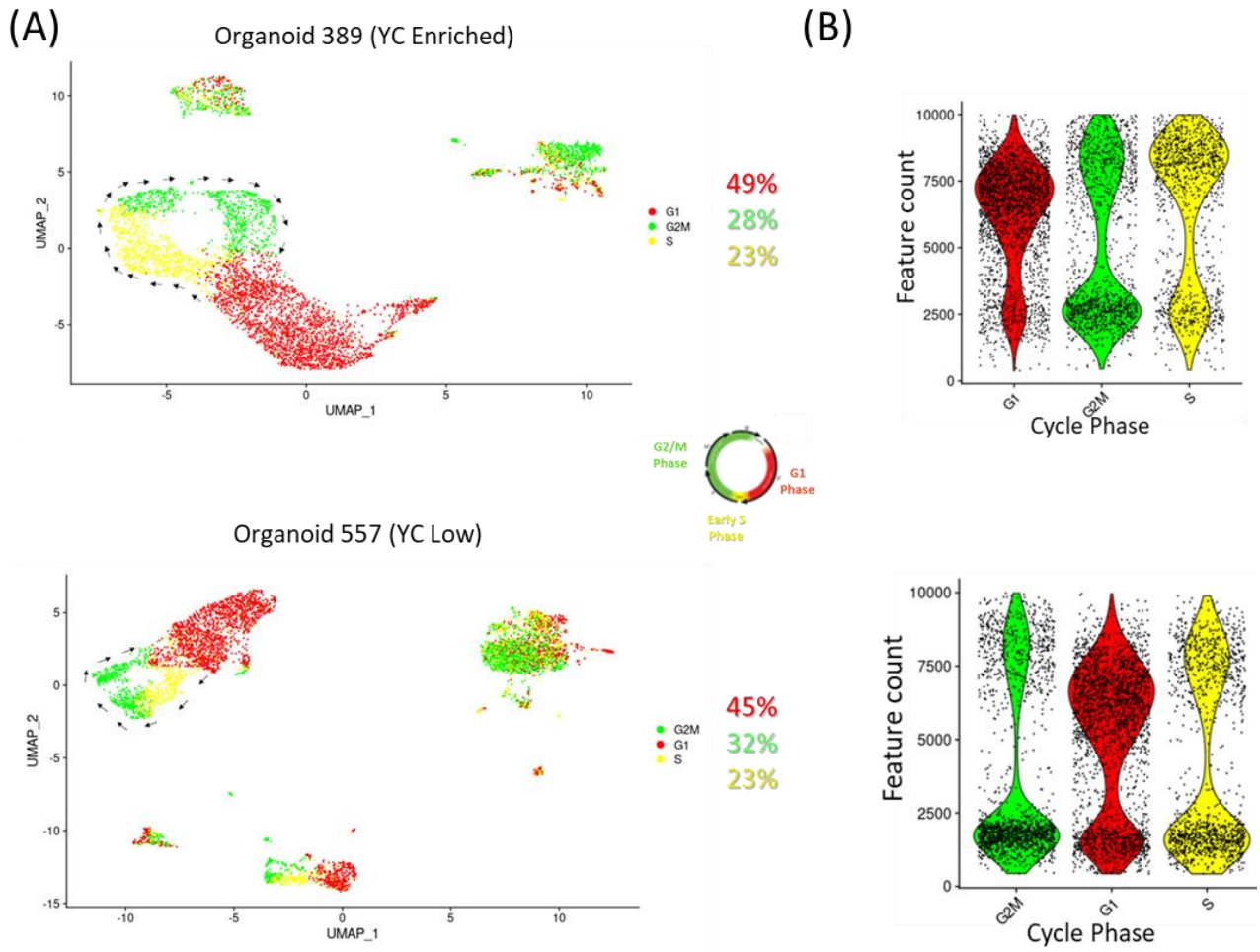


Figure 5.1: **ITH of cell cycle phase identity in organoids.** (A) UMAP clustered organoids coloured by cell cycle phase identification through cycle gene scoring. Percentage of each phase present in the total population coloured by relevant phase (red = G1, yellow = S and green = G2M) throughout. Arrows indicate cycling pattern of expression data recently outlined in (Pountain et al., 2024) (B) Feature count (genes captured) per cell grouped by cycle phase.

5.2.1.2 Identifying TSS usage dynamics between cycle phases

Next, I repeated our pseudo bulking analysis, generating CTSS files per cell cycle phase for both samples independently, ran CAGEr workflow normalising between phase populations with α = 1.05 fit in range of tags 10 to 15,000 and converted to TPM. Identified promoter regions and TSS type and performed dual gene YC:YR frequency distribution analysis using the equation in **Figure 5.2a**. This analysis shows that in both samples we see a distinctive YC shift in TSS usage in cells in G1/G0. Contrastingly in both samples YC is depleted in S and G2M phased cells, supporting finding from Chapter 4. Further to this, after integrating the datasets to compare likewise phases between samples we see still see the same trend but

YC enrichment in G1/G0 phase is greater in the more proliferative 389 sample, Figure 5.2. Literature highlights distinct gene expression profiles underpin each cycle phase. Therefore, I aimed to rule out that the cycle dependent YC shift was just resulting from an upregulation in dual initiator genes in G1/G0 phase. Correlation analysis of YC:YR against dual gene expression score per cell, coloured by phase shows no distinct separation in the phases in dual initiator expression but clear separation is seen in YC:YR (**Figure 5.2c**). Further supporting that conventional RNA sequencing approaches that only focus on expression differences may miss underlying TSS associated dynamics.

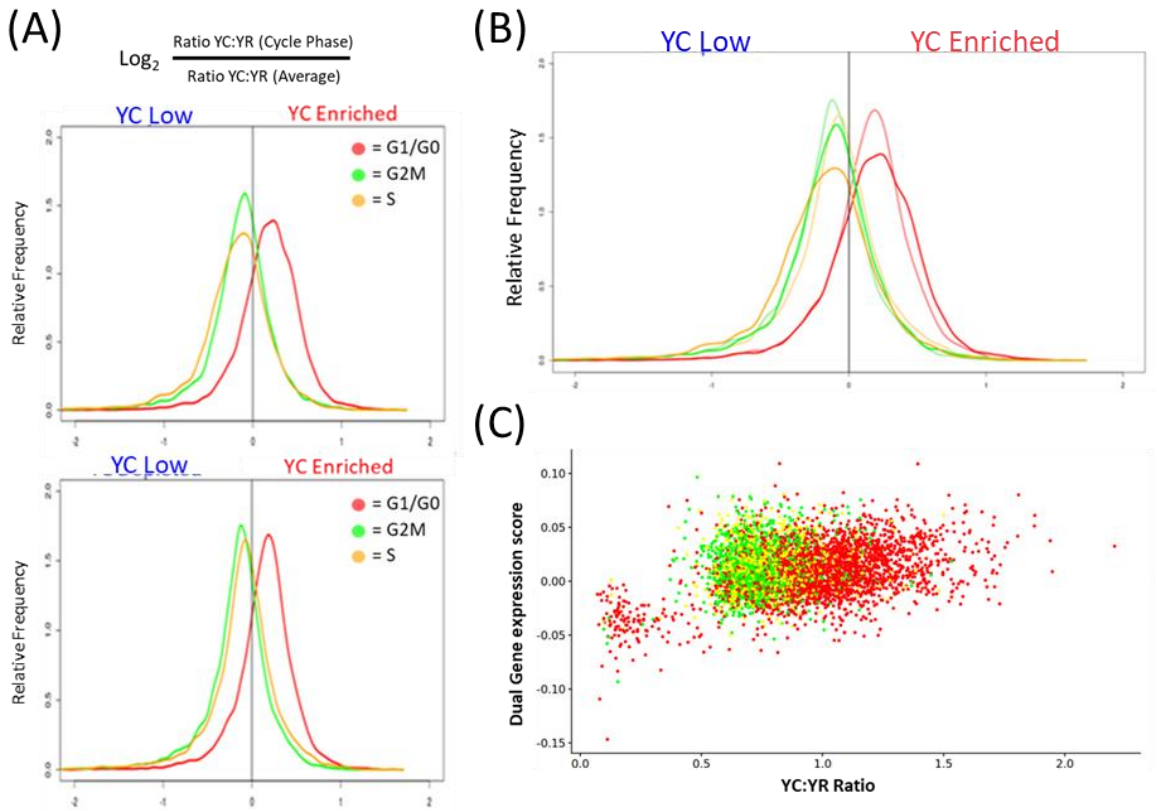


Figure 5.2: YC enrichment differs between cycling phases. Cell cycle scored TSS usage analysis. (A) Frequency distribution of YC:YR ratio per cycle phase compared to whole average for signature Dual genes. YC:YR expression ratio calculated per DUAL initiator per cycle phase divided by average YC:YR ratio for that promoter. Top = 389, bottom = 557. (B) Integrated frequency distribution analysis of cycle phases for both samples, darker = 389, lighter =557. (C) Average YC:YR ratio vs Dual gene module score per cell, coloured by cycle phase.

To visualise TSS usage on a single gene level and compare between population, we generated UCSC genome tracks from the normalised CTSS from the genomic ranges object supported through CAGEr workflow for each cycle phase population. The bigwig files were uploaded to UCSC genome browser and custom sessions were generated. **Figure 5.3** shows

an example of a dual gene *ATF4* with differential TSS usage between phases where overall expression is relatively unchanged indicated in the bar charts proceeding each track. We see in *ATF4* an enrichment for YC usage shift compared to a YR enrichment shift in G2M. Many studies have explored the roles of *ATF4* particularly in cancer as it plays a significant role in stress mediated response through its TF function and plays a vital role in controlling cellular proliferation and in high levels can induce cycle arrest (Torrence et al., 2021; Wortel et al., 2017). Expression levels of *ATF4* are significant in driving the balance between maintaining proliferative capabilities and surviving cellular stress through mediating metabolite production and initiating pro-apoptotic signalling (X. Tian et al., 2021; Wortel et al., 2017). We also previously identified ATF4 as a candidate dual gene. Differential TSS usage of *ATF4* highlights a potentially novel regulatory mechanism of cellular stress regulation, furthermore high YC enrichment in G1 could indicate a cancer cell pro-growth mechanism.

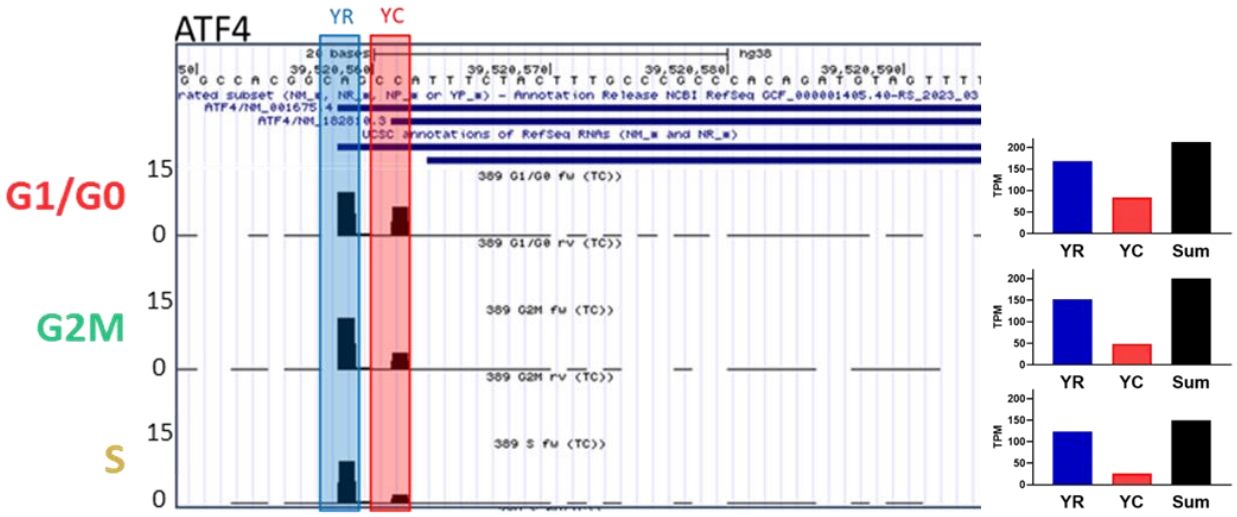


Figure 5.3. ATF4 shows cell cycle dependent TSS usage. Genome browser view of *ATF4* gene TSS. TPM per cycle phase per TSS. UCSC genome browser view of ATF4 gene and TSS specific TC for each Cell cycle phase. YR initiator distinguished by a 'CA' TSS and YC by a 'CC'. Bar chart showing total TPM per TSS and total sum TPM.

Previous studies have shown ribosomal gene expression is significantly enriched in G1 phase to facilitate the protein production demand required to enable cycle progression (Nosrati et al., 2014). Feature plots in our samples further validate this showing in G1 cells we see a higher % of reads mapping to ribosomal genes (**Figure 5.4a**). I removed 80 ribosomal genes and repeated YC:YR frequency distribution analysis to determine if our ratio shifts between phases were resulting from these genes. Overlaying the frequency plots shows little

alteration in our plots (**Figure 5.4b**). This analysis highlights that the YC enrichment we previously identified in G1 is not just a result of an increase in the TOP containing ribosomal genes. Furthermore, the absence of significant expression differences, highlights genes with potential differential regulation through alternative TSS usage potentially involved in cell cycle dynamics. These findings would be missed in conventional RNA sequency due to the absence of differentially expressed genes.

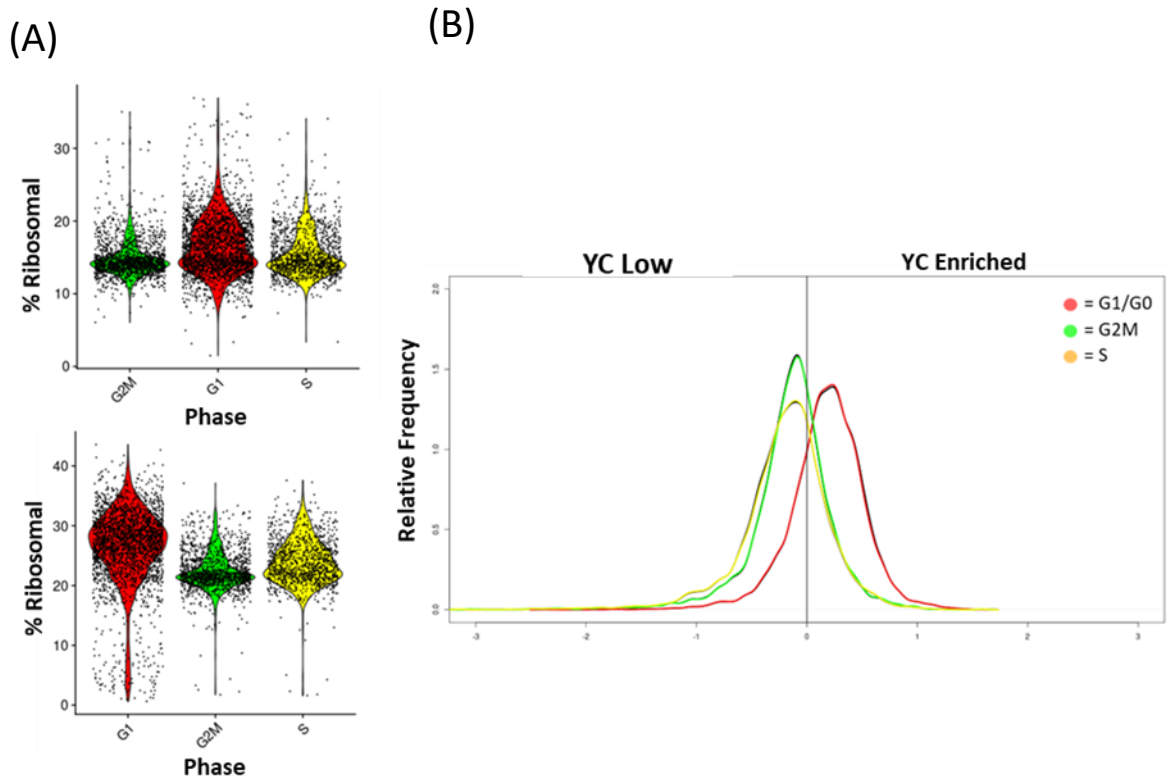


Figure 5.4. Removal of high YC ribosomal genes does not change YC:YR shift in cycle phases. Removal of ribosomal gene mapped counts (A) Percentage of reads mapped to ribosomal genes split by cycle phase, sample 389 Top and sample 557 Bottom. (B) Cell cycle Dual gene YC:YR frequency distribution represented by the black lines and post removal of ribosomal genes in coloured.

I aimed to test if the cycle phase dependent shift observed was specific to the YC TSS or a global shift of transcript abundance. I repeated the ratio frequency analysis calculating TPM signal for all dinucleotides including conventional YC and YR, the recently identified non-biological dinucleotide combinations and the uncharacterised others outlined in **Figure 5.5**. The greatest shift was attributed to YC:YR with lower degrees of shifting occurring when comparing YC to the other groups and almost no cycle phase separation when these were compared to YR dinucleotides (**Figure 5.5**).

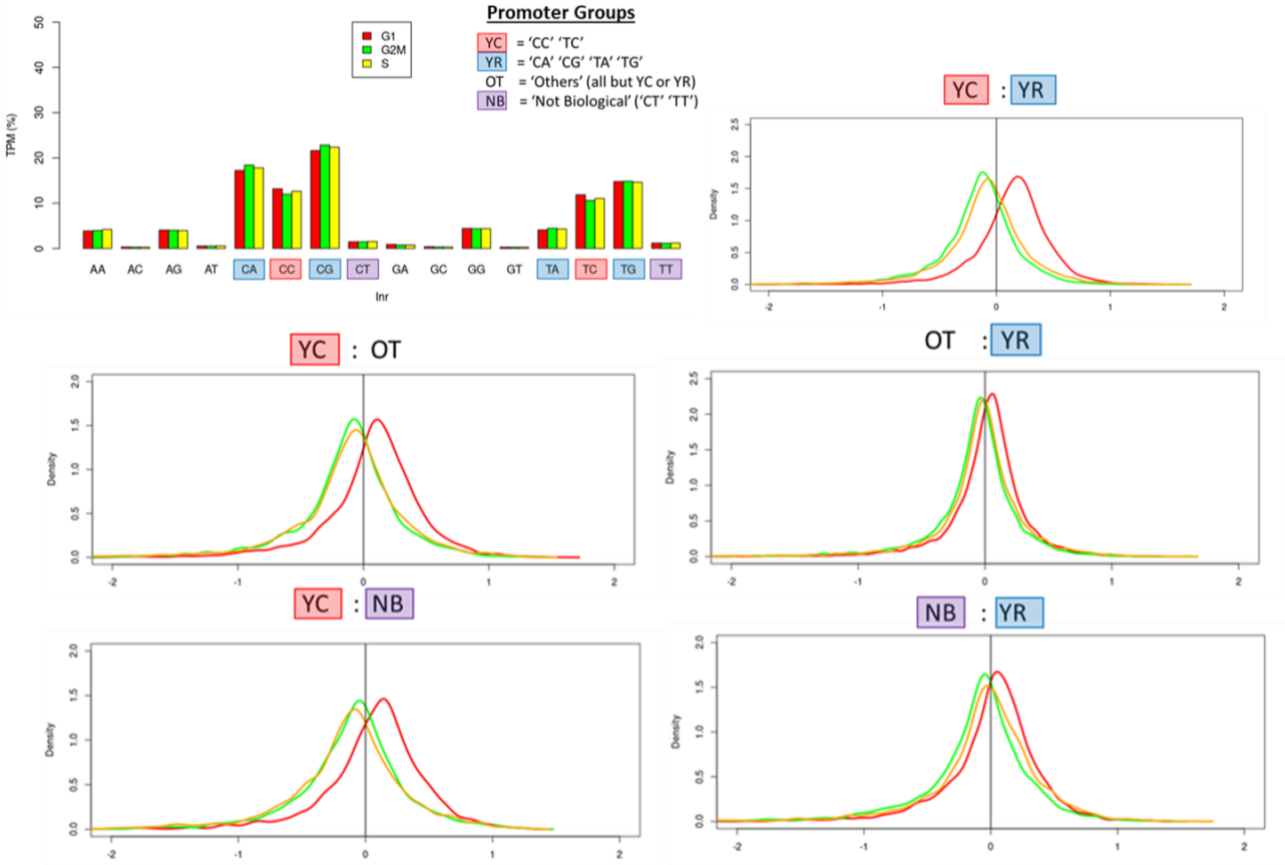


Figure 5.5. Dinucleotide shifting is unique to YC initiators. Comparison of dinucleotide usage shifting. (A) Bar showing % TPM per dinucleotide pairs defined by promoter group in upper legend. Frequency distribution analysis for all dual genes with both signals (B) YC:YR (C) YC and YR : all other dinucleotides (OT) (D) YC and YR : Non-biological dinucleotides (NB).

5.2.2 Validation of TSS shifting in cell cycle phase sorted populations

So far, our analysis has been on sequencing generated from one experiment, as we are currently unable to generate a repeat for this single cell experiment, we wanted to validate our findings by identifying and performing CAGE sequencing on relevant subpopulations. Given that we see a significant TSS shifting in cell cycle phases across both samples and that a well-developed and characterised cell cycle fluorescent tracer, FUCCI, has been developed we opted to choose cycle phase subpopulations. We generated FUCCI organoids by transducing our organoid lines with Lenti-viral particles containing the FAST-FUCCI plasmid, outlined in Appendix 1, that we generated from HEK293 FT cells. Initial attempts relied on the use of a different FUCCI plasmid: FUCCI-PLJM1 plasmid map outlined in **Appendix 1**. However, this plasmid contained truncated forms of the phase specific reporters. I found essential degradation of the G2M protein marker Geminin was not sufficient and led to

cytoplasmic accumulation making resolution of cycle phases not possible (**Figure 5.6a**). A new FUCCI construct was obtained (FAST-FUCCI) the previous line generation steps were repeated, and imaging showed only nucleic reporter expression. Comparative image analysis of the two FUCCI generated lines are shown in **Figure 5.6b**, proceeding experiments relied solely on the FAST-FUCCI construct.

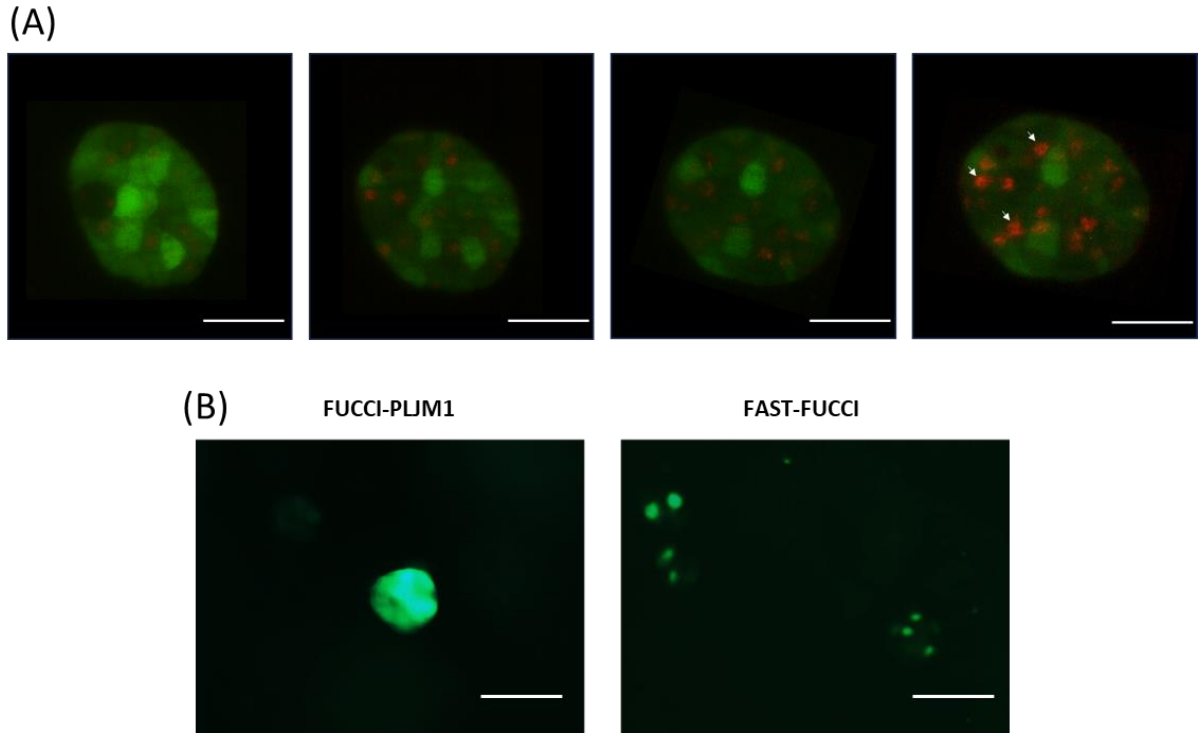


Figure 5.6. Comparison of different FUCCI construct generated organoid lines. (A) FUCCI-PLJM1 transduced 557 organoid showing inefficient Geminin (green) degradation overlapping with red nuclear CDT1 expression (white arrows) preventing G1 resolution. (B) Left FUCCI-PLJM1 transduced 557 line and right FAST-FUCCI transduced 557 organoid showing nuclear expression of both fluorescent reporter proteins. Scale bar 100µm.

To compare the relative cycling phase populations, present in our organoids and therefore validate our single cell scoring we recovered and fixed our FUCCI transduced organoids at day 14 and imaged in the high resolution Z1 lightsheet. As shown in **Figure 5.7** fluorescent signals demonstrates colour variation and spatial differences which correspond with their characterised morphology. The more proliferative cystic line (389) shows more disperse and more cycling cells representative of higher quantity of GFP, such feature is indicative of a more uncontrolled growth with many cells with proliferation association capacity. Contrastingly the slower growing more differentiated 557 sample has a more compact, structured lumen like morphology with a more synchronous phasing of cycling cells.

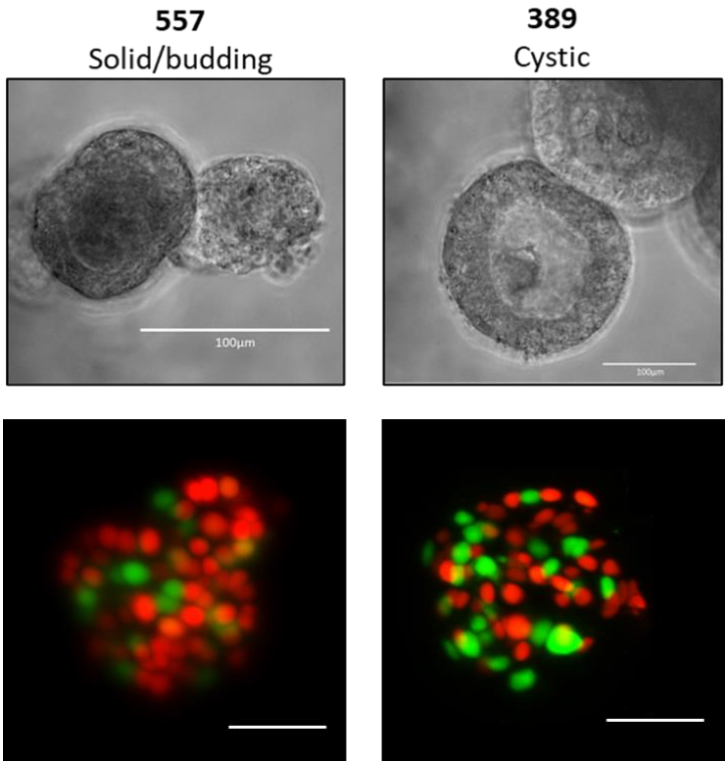


Figure 5.7. Organoid lines show differential cycling populations. Morphological characterisation of organoid lines using brightfield imaging. (B) Comparison of FUCCI generated organoid lines with cells expressing nuclear mKO2-CDT1 red G1 identity or mAG-GEM green G2M cells. Overlapping of expressing and hence yellow nuclei represent early S phase. Scale bar 100µm.

As our single cell analysis was carried out on non-transduced lines we aimed to see if proliferative dynamics are potentially altered post transduction and antibiotic selection. Therefore, we repeated our previously outlined doubling time analysis on the FUCCI positive organoid lines following both the day 4 and day 14 counting. As seen in **Figure 5.8** we see little alterations compared to the growth dynamics characterised in **Figure 3.1**. The previous general trend was retained, whereby Sample 389 is still proliferating faster represented by a shorter doubling time. Despite this organoid 389 being indicatively sensitive, the antibiotic selection required to generate the FUCCI lines does not appear to have affected the global phenotype.

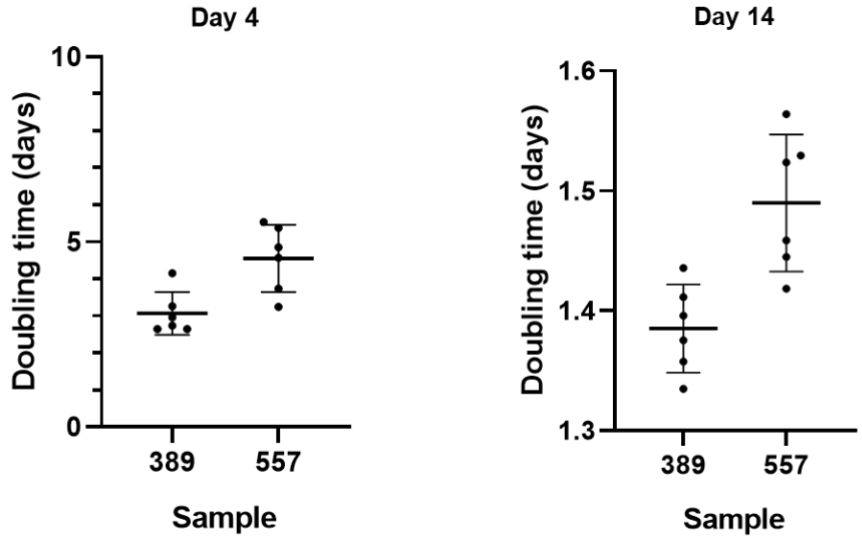


Figure 5.8. Doubling time analysis on FUCCI Organoid lines. Dot plot representing calculated doubling time for each line. Each individual point represents 1 well counted at day post seeding (PS) (A) 4 days PS, (B) 14 days PS. All experiments were a minimum of n=3.

I cultured our FUCCI organoids to day 14 following the same culturing, harvesting and single cell suspension protocol performed in the single cell experiment. To bulk according to cell cycle phase, I FACS-sorted according to the FUCCI fluorescent reporters using the highlighted FACs gates defined in Chapter 2.4.2, ensuring consistent and similar environmental conditions as that used in the optimising for the single cell experiment. The use of the FUCCI reporter also allowed us to distinguish between G1 cells and those in G0 as the latter will be the non-fluorescing population denoted in blue in **Figure 2.5**. Therefore, I obtained 3 populations from sample 557: G1, G2M and G0 cells, since we yielded different quantities of cells to make comparable sequencing libraries, I down sampled both G1 and G0 to match the 50,000 cells obtained in the G2M population. Lower concentrations of RNA were obtained than were necessary for standard CAGE workflow, therefore we opted to perform the small cell number optimised SLIC-CAGE protocol outlined in **Chapter 2.4.2**. Initial QC metrics obtained from library preparation showed little carrier RNA retention used in the SLIC-CAGE protocol and good quality RNA was identified.

BAM files were generated per population and analysed using CAGEr workflow. Promoter specific CTSS identified and TSS defined into promoter group (YC or YR) as before and relative TPM signals were used to calculate YC:YR ratio. Frequency distribution analysis of YC:YR

demonstrated the similar trend of YC enrichment in G1 compared to G2M (**Figure 5.9**) as seen in the scored pseudo bulked single cell experiment validating our findings (**Figure 5.2**). Furthermore, G0 followed a similar YC enrichment as G1, with little separation occurring between the two populations (**Figure 5.9**).

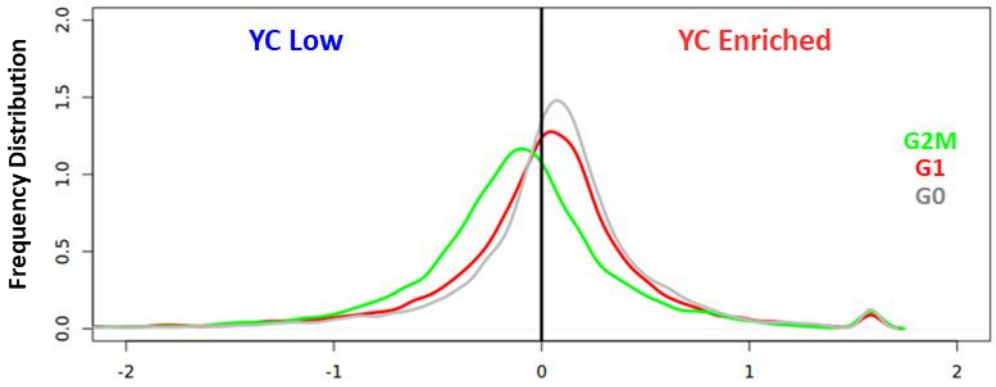


Figure 5.9. YC enrichment in dual genes is seen in G1 and G0 cells. Frequency distribution of YC:YR ratio per cycle phase compared to whole average for signature Dual genes. YC:YR expression ratio calculated per DUAL initiator per cycle phase divided by average YC:YR ratio for that promoter.

To identify whether any genes differ in their TSS usage between G1 and G0 I plotted a gene by gene comparison of normalised YC:YR for G1 vs G0. Genes with high YC preference in G1 but high YR usage in G0 would plot in the bottom right quadrant of the scatter (+0.5 value for G1 > x >-0.5 value for G0) where x = a dual gene. Genes with the reverse trend, High YC in G0 but high YR in G1 would plot in the upper left quadrant (+0.5 in G0 > x > - 0.5 in G1) (**Figure 5.10**). I performed GO analysis on these populations to identify enriched pathways using KEGG enrichment and identified metabolic genes are enriched in YC genes in G0, as cells enter the cell cycle and hence G1, YC enrichment shifts to the regulator genes with known translation associated functions. The enrichment in YC usage in the metabolic genes is likely needed to enable sufficient cellular metabolite composition to aid in cellular division before cells enter the cell cycle. Upon entry to G1 TSS usage shifts to the translation associated genes potentially to drive protein production from the previously generated G0 transcripts, highlighting a potentially transcription level regulation of translation in cell cycle regulation.

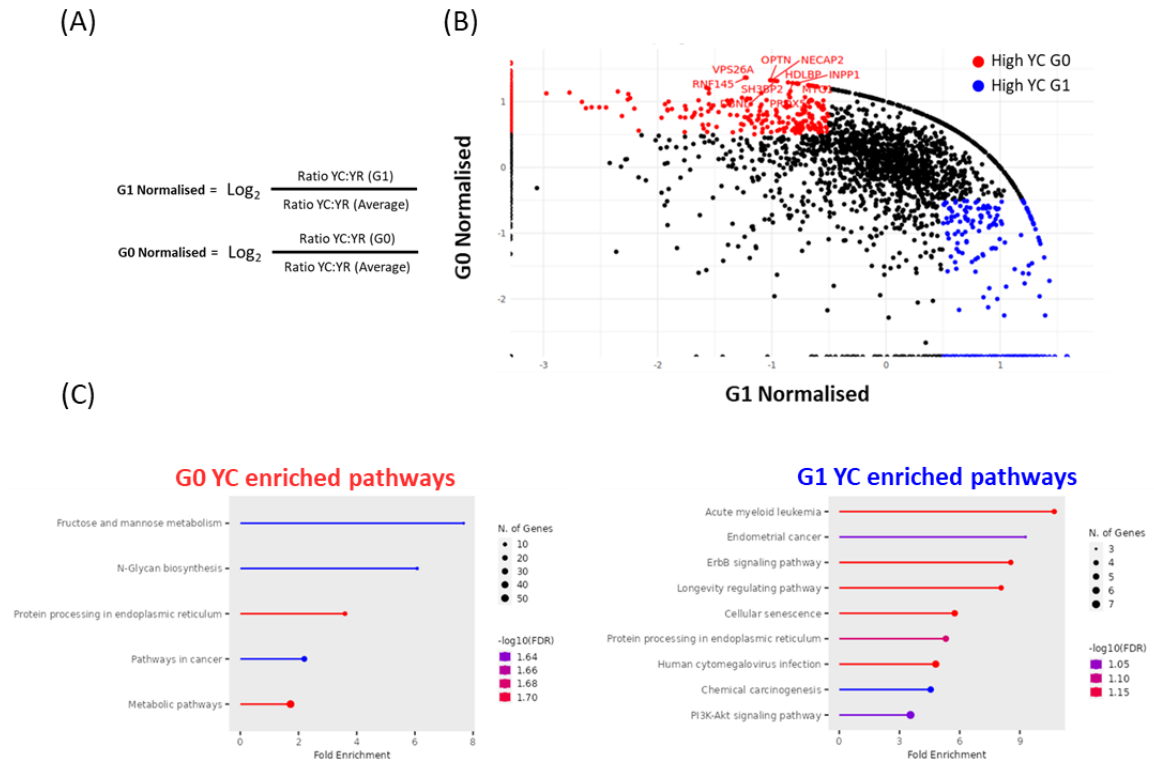


Figure 5.10. Discrimination of genes with YC enrichment in G1 and G0 cells. Per gene comparison of normalised YC:YR value for G1 vs G0. (A) normalisation equations per sample. (B) gene by gene scatter plot of normalised YC:YR value in G1 cells vs normalised YC:YR value in G0. Blue indicates genes where they show YC enrichment in G1 but YR enrichment in G0. Red indicates genes with G0 YC enrichment and G1 YR enrichment. (C) Kegg pathway analysis of enriched genes that show YC enrichment in G0 and YC enriched pathways in G1.

To better visualise the YC:YR distributions I converted the frequency distributions into box plots, as can be seen in **Figure 5.11** the median points of G1 and G0 are almost equal. Given that previous results show that high YC in G1 is attributable to translation associated genes many of which harbour the indicative TOP motif. I hypothesised that a higher degree of separation of G1 and G0 would be seen between genes not containing the motif. Therefore I repeated YC:YR frequency distribution analysis ranking genes by TOPness of the initiator motif one base at a time. Starting with genes with full TOP motif whereby the conventional 5' pyrimidine sequence is found, YC4 containing only 4 pyrimidines thus flanked by purine bases, YC3 with 3 pyrimidines and finally YC2 where only two pyrimidines with sequence ('CC' or 'TC') are found. Interestingly we found a greater degree of separation and hence YC shift in G0 in the YC2 genes than with the full TOP motif (**Figure 5.11**).

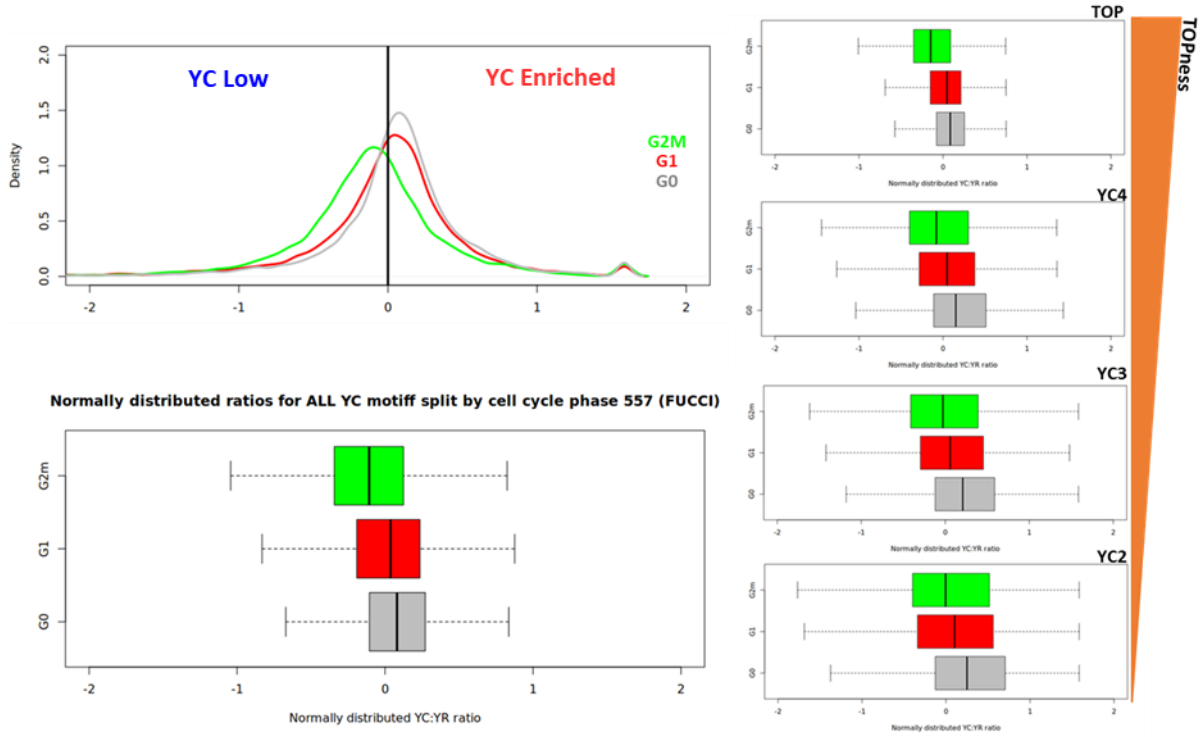


Figure 5.11. Decreasing TOP motif likeness separates YC enrichment of G1 and G0 cells. Frequency distribution bar chart plots for FUCCI separated cell cycle phases. (A) YC:YR including all YC signals. (B) Decreasing TOPness of YC initiator sequence. Top panel including YC signal only from genes with conventional TOP sequence. YC4 includes genes only with initiator sequence with 4 pyrimidines, YC3 with only and YC2 where only two pyrimidines are found or pure YC.

5.2.3 Selection of G0 meta gene for cell cycle phase stratification

To try and validate the G0 vs G1 findings and to be able to perform DGE analysis between G0 and G1 cells I aimed to identify the G0 cells in our single cell data by generating a G0 meta gene. The meta gene was composed of G0 vs G1 DEG that I could score the single cell data for. The initial TSS TPM frequency distribution graphs in **Figure 5.9**, show little difference between the two phases. I hypothesise that it is unlikely that these G0 cells represent senescent cells and given the nature of cancer organoid culturing, cells are likely continuously growing, explaining the significant similarity between G1 and G0. To identify any potentially differentially expressed genes between the G1 and G0 I performed clustering analysis using the self-organising maps (SOM) algorithm, as we currently do not have repeats to carry out statistically relevant *Deseq* analysis. I log transformed the totalled TPM per CC, meaning signal was not separated by initiator type and thus included all signal and performed SOM using gaussian neighbourhood functioning (**Figure 5.12a**).

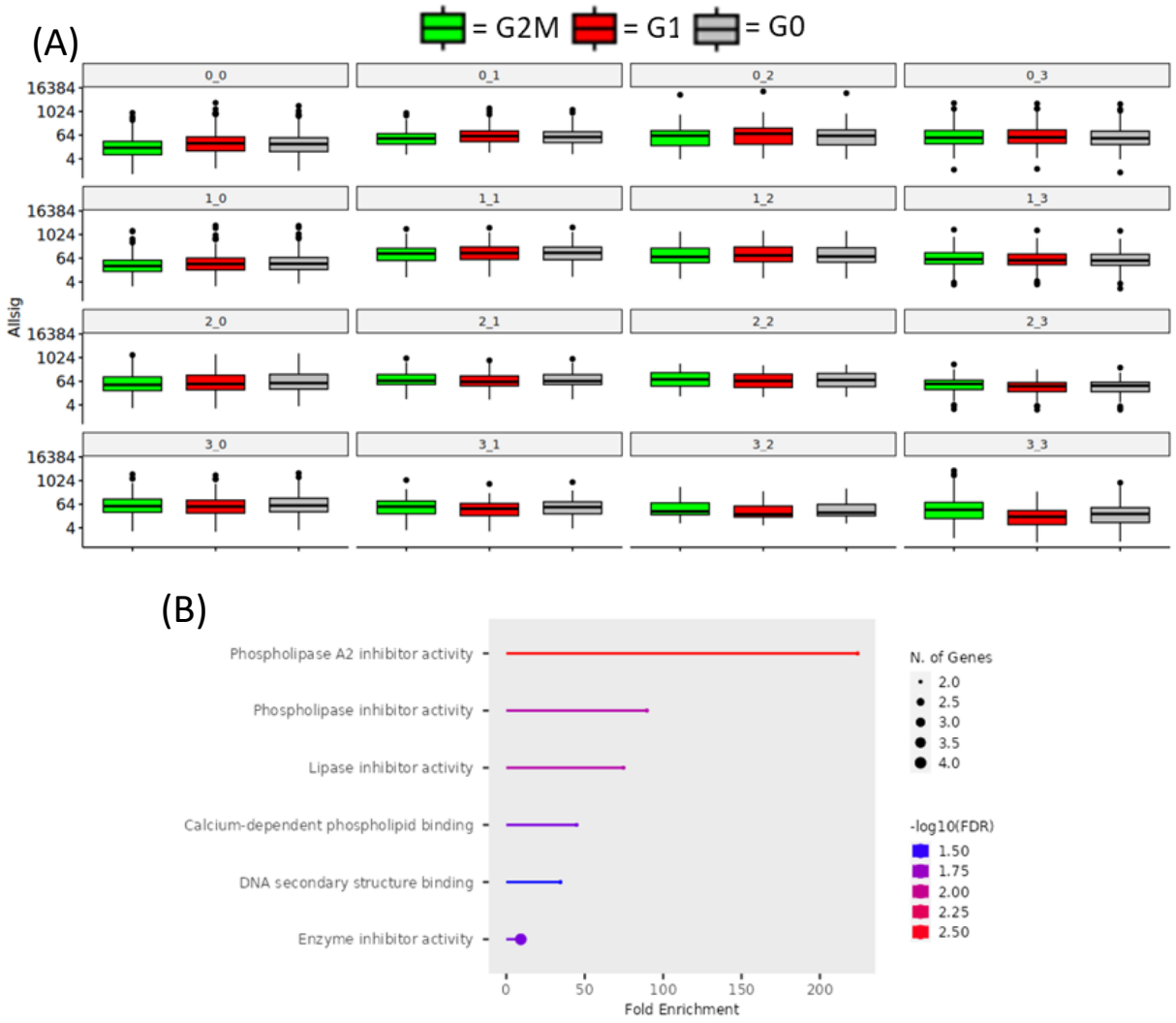


Figure 5.12. Identifying cycle phase differentially expressed genes using SOM cluster analysis. (A) 4 by 4 SOM cluster analysis of all TPM signal per cell cycle phase. (B) Gene ontology analysis of the differentially expressed genes between G0 and G1.

I extracted genes from clusters whereby the median of G0 was greater than G1, Clusters 3_0, 3_1 and 3_3, and defined these as the G0 meta gene. GO of the G0 meta gene shows lipid metabolism associated functions (**Figure 5.12b**), including the upregulation of genes associated with inhibition of phospholipase enzymes. Phospholipases are crucial enzymes that harbour varying functions ranging from production of lipid metabolic biproducts to signal transduction pathways all of which have shown significant involvement in cancer progression through proliferation dynamics (Salucci et al., 2023). Moreover, phospholipase has a significant role in maintenance of the WNT signalling pathway in CRC, a pathway known to mediate cellular proliferation (W. C. Hwang et al., 2022). Clinical inhibition of these

enzymes as a potential cancer treatment showed significant reduction in cell proliferation (W. C. Hwang et al., 2022; K. F. Scott et al., 2010). I concluded that the intrinsic upregulation of genes with such inhibition function seen between G1 and G0 populations are a sufficient G0 meta gene to identify potentially non-cycling cells in our single cell scored data.

To identify the potential G0 cells I performed module scoring using our G0 meta gene on single cell data of sample 557, as this was a sample match to our FUCCI CAGE sequencing. G0 identity was then assigned to cells in the upper quartile of prediction scores showing significant upregulation compared to average. Visualisation of the potential G0 cells seen in **Figure 5.13** showed that only cells that were previous G1/G0 assigned cells contained G0, consistent with expectation. The G0 cells appeared to also cluster together in a population distinctively separate from the main clusters and thus may represent cells with transcriptome associated with leaving the cell cycle whereas the other cluster highlights cells entering or within the cell cycle.

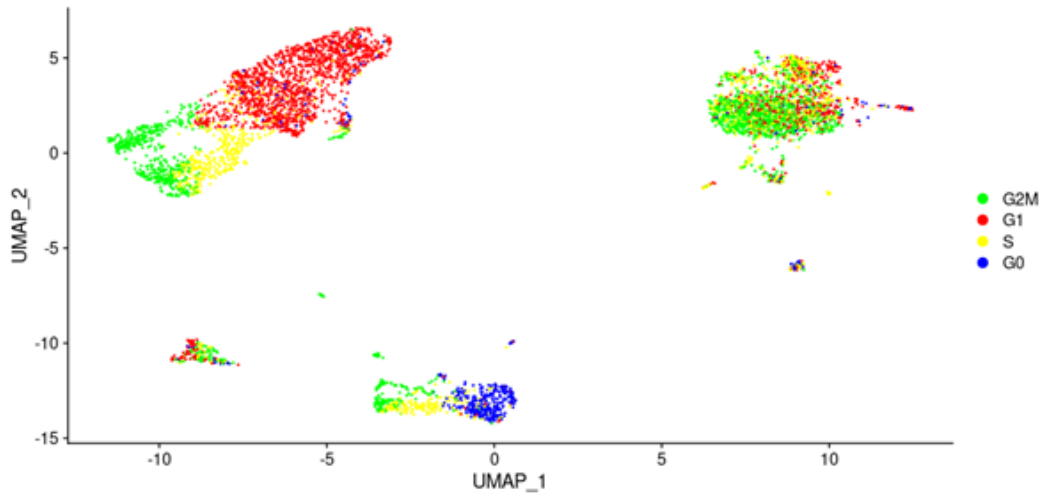


Figure 5.13. Clustering visualisation of G0 cells in single cell data. UMAP Clustering analysis of single cell sequenced Organoid 557 coloured by cell cycle phase identity achieved through cycle phase gene scoring including G1/G0 separation using G0 meta gene highlighted in blue.

To validate the observed G0 vs G1 TSS shifting in the FUCCI CAGE analysis, I pseudo bulked the sample 557 single cell data by cycle phase, now including the newly identified G0 cells. Then I extracted associated cellular barcodes and generated CTSS files using the filtered SCAFE reads and ran CAGEr workflow as previously. Frequency distribution analysis of dual

genes showed the same YC enrichment shifts in G1 and G0 seen in our FUCCI analysis in **Figure 5.14**. Furthermore, both analyses also showed a slightly greater YC shift in G0 compared to the other cycling phases represented by a higher median value (0.68) in the frequency distribution analysis. Although this analysis provides some degree of validation, experimental repeats are needed in addition to the same analysis on a separate line to identify whether this is universally found.

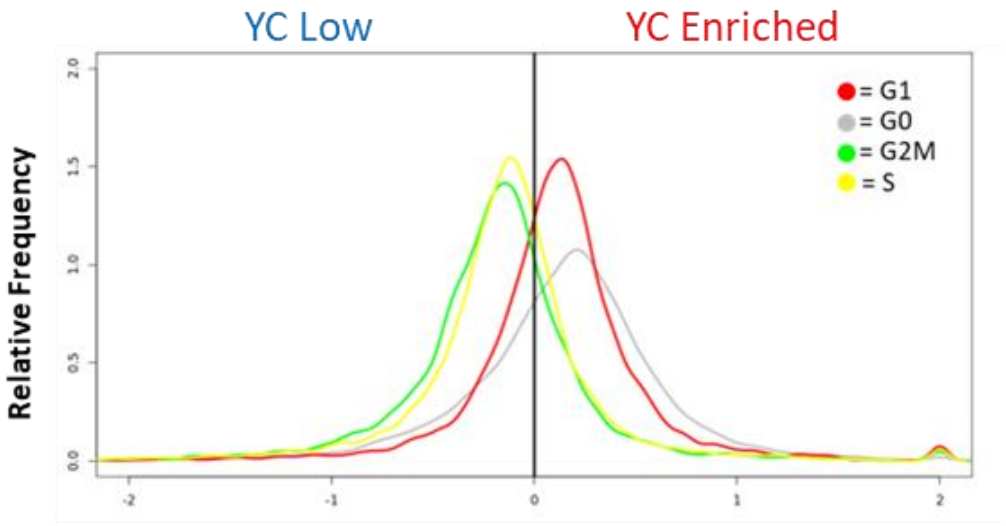


Figure 5.14. G1 and G0 scored cells retain YC enrichment. Frequency distribution analysis of YC:YR for single cells scored by cycle phase genes including G0 metagene to distinguish between G0 and G1 cells. YC:YR calculated per gene per cycle phase divided by average YC:YR for that gene.

5.2.4 Development of long-term high-resolution live imaging protocol

Evolutionary analysis of tumour development is becoming a popular approach to characterise the plasticity of cellular states and behaviours (McGranahan & Swanton, 2017; Shackleton et al., 2009). So far, our TSS usage associations to cell state are limited to single fixed time points. However, our results have shown that G1/G0 specific YC enrichment maybe a contributing factor in enabling proliferative capabilities. Research has highlighted the significant timing relevance of these associated cycling phases in enabling and sustaining high proliferative phenotypes (Fleifel & Cook, 2023; Viner-Breuer et al., 2019). For example, shortened G1 phase has been established in highly proliferative tumours and research has strived to understand the regulatory mechanisms that enable this change (Fleifel & Cook, 2023). Like many cellular states, cycling dynamics are believed to be highly

dependent on spatiotemporal factors (Gaglia et al., 2022). Therefore, I aimed to characterise the spatiotemporal patterning of cycling phases in our organoid development. This was to better understand the growth and hence cycling dynamics in our organoids and how they may differ between two differentially proliferating lines. Thus, exploring our hypothesis that the more proliferative line would be underpinned by more G2M populations and notably shorter cycling phase lengths. The complexity of 3D culturing conditions, however, makes high-resolution 4D imaging challenging. Furthermore, broad field microscopy fails to truly capture organoid 3D structures which results in low resolution of fluorescent reporter spatial patterning (**Figure 5.15a**). High resolution imaging such as that achieved with the Z1 light sheet microscope allows for better retention and analysis of 3D architecture (**Figure 5.15b**).

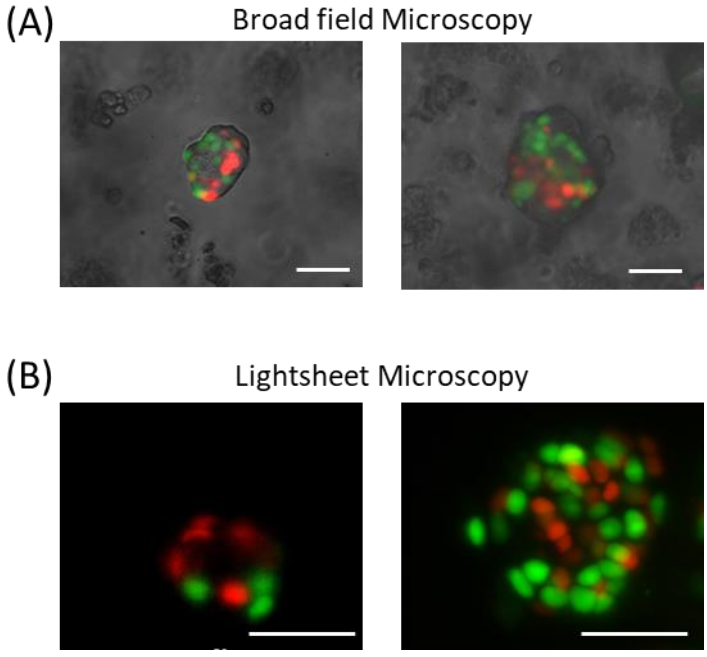


Figure 5.15. High resolution imaging captures organoid 3D architecture. (A) Broad field microscopy imaging of FAST-FUCCI organoids at day 10 (left) and day 14 (right). (B) high resolution lightsheet imaging of early 7-day 557 organoid (left) and late-stage day 14 (right). Scale bar 50µm.

I aimed to develop a protocol to enable high-resolution long-term imaging of organoids adapting culturing conditions to suit Z1 lightsheet imaging. The imaging set up for the Z1 lightsheet highlighted in **Figure 5.16** shows its complex hanging design that requires suspension of the sample into the imaging chamber, this methodology allows for 3 axis movement and hence visualisation. Therefore, my experimental design outlined in Chapter

2.3 **Figure 2.1** encompassed the conventional organoid culturing suspension in BME encased in a holding medium that would allow suspension from the lightsheet capillary. The cultured organoids would then be imaged in the lightsheet before being returned into an adapted vertical culturing approach utilising a modified T75 tissue culture lid on a falcon tube containing Intesticult organoid media. Although it would be desirable to fill the Z1 lightsheet imaging chamber with the organoid media, it contains phenol red and previous studies have shown it can increase background fluorescence (Stadtfeld et al., 2005). Furthermore, this media is highly expensive, and the volume required to sufficiently fill the chamber makes it extremely impractical. To enable the success of this protocol we needed to identify a holding medium that would have specific biochemical and biophysical properties that would not interfere with organoid growth and would be strong enough to hang in a suspension from the glass capillary. We sought out the collaboration of chemical engineers working with a novel fluid hydrogel called Gellan with malleable molecular characteristics to enable strategic manipulation to provide a suitable holding medium. Gellan is composed of a natural biopolymer containing polysaccharide repeating units (Cooke et al., 2018). This simple sugar-based structure forms distinct similarities to the extracellular matrix in cells (Albrecht et al., 2022; Cooke et al., 2018). The microstructure of Gellan can be seen in **Figure 2.1**, the fluidity of the gel giving it gel like properties is achieved using a sheering process in its production that creates Van der Waals forces between smaller fragments (Cooke et al., 2018; Jamshidi et al., 2016). The optimisation of Gellan and determining most suitable microstructure formation as our holding medium is outlined in **Chapter 2.3.3**.

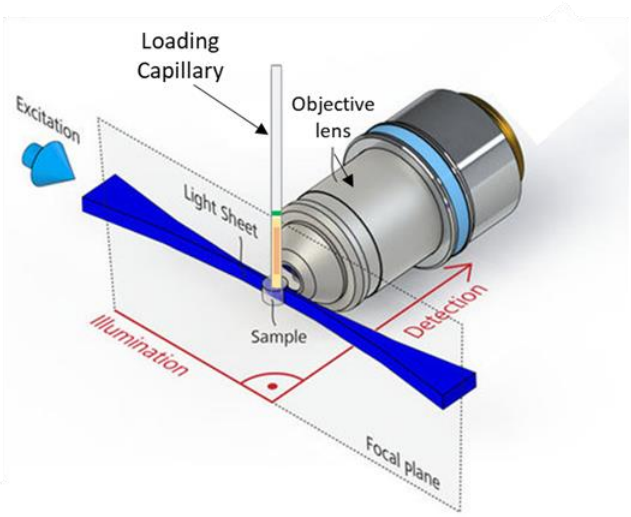


Figure 5.16. Longitudinal high resolution imaging protocol. Zeiss Z1 lightsheet sample set up showing suspension of holding capillary into imaging chamber. Loaded gel embedded sample is pushed into focal plane and hence hanging from capillary. Adapted from (Eberle et al., 2015)

I confirmed Gellan does not affect cellular viability through visualisation of cellular behaviour and appearance using Broadfield microscopy and comparing to conventional culturing conditions not embedded in Gellan (**Figure 5.17**). I found that no apparent affects to cell viability was seen and growth to 5 days showed retention of average organoid size in gels 1,2,4 and 5. However significant decrease in organoid growth and cell death was seen in the high salt containing gels 3 and 6. A likely cause was a disruption to osmotic balance and therefore causing cell stress and even bursting, these gels also lacked the optical clarity and therefore we did not proceed with these. This validated both permeability, viability, and optical clarity of gels. As outlined in **Chapter 2.3.3.1** the optimisation analysis highlighted gel 4 (**Table 2.2**) as the most appropriate microstructure for imaging and as shown in **Figure 5.17** cell viability is retained in this gel.

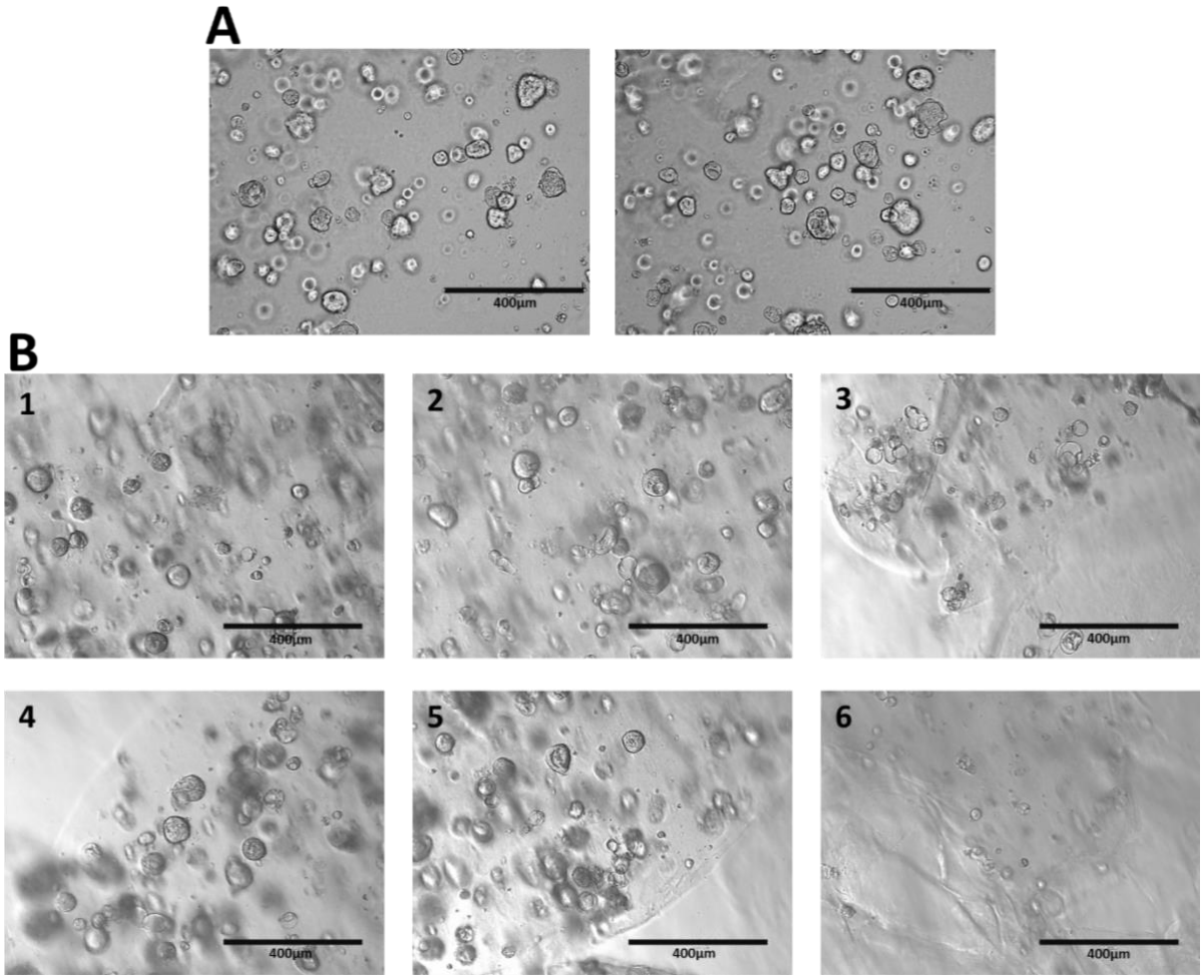


Figure 5.17. Gellan embedding does not impact organoid growth. Broadfield imaging of Day 5 organoids line 557 (A) normal culturing conditions including only BME embedding. (B) Gellan embedding of BME encased organoids using the different Gellan compositions corresponding to gel identity. Scale bar 400µm.

The imaging was set up as outlined in **Chapter 2.3.4**. with initial attempts imaging the FUCCI organoids every 12 hours and retained in the vertical culturing tube when not being imaged. As outlined in **Figure 5.18a** I was able to capture and hence map the initial cycling phase transitions in the first 7 days of organoid growth represented by fluorescent nuclei highlighting the presence of cycling phase. The continued expression of the Geminin reporter over multiple time points indicated cells were not undergoing stress induced G1 arrest. Therefore, determining the protocol thus far successful. Initial results showed shorter doubling times than calculated in the doubling time analysis. **Figure 5.8** showed doubling time was on average 3 days, however **Figure 5.18** shows it the be much shorter, with initial and most subsequent divisions occurring within 24 hours.

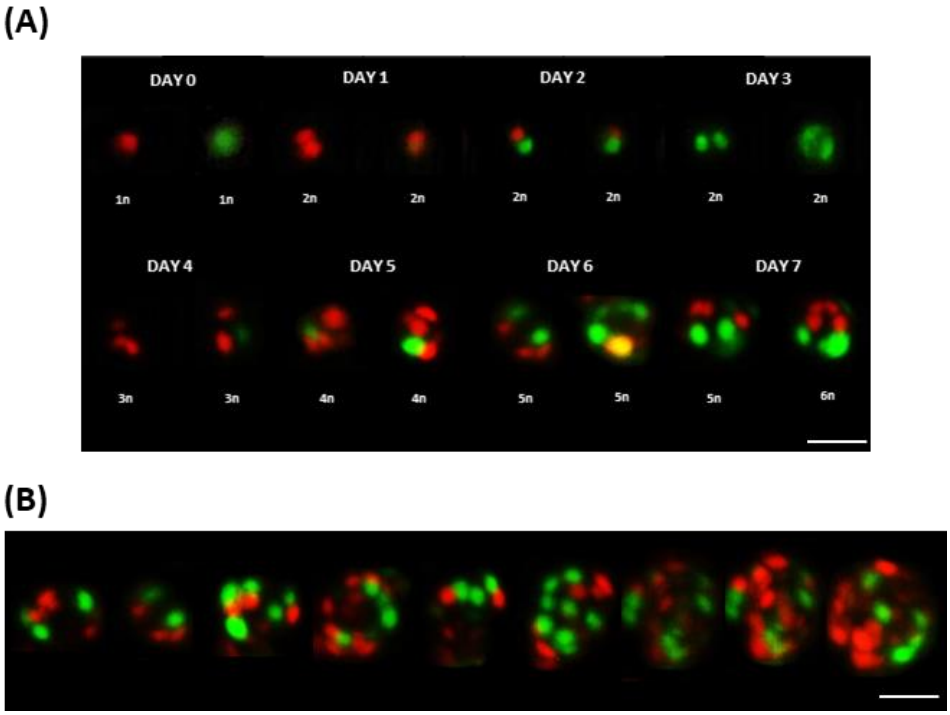


Figure 5.18. Cell cycle tracking of FUCCI organoid over 14 days. Fluorescent snapshots of nuclear expression of FUCCI fluorescent reporters in the same organoid over 14 days (Post seeding on day 0) using adapted long-term high-resolution imaging protocol in Zeiss Z1 lightsheet. Captured using selected lasers 488 and 561nm at 1.8 and 1.3 intensity and 219.71 millisecond (ms), scale 50µm. (A) Tracking of cell divisions using both fluorescent reporters to distinguish cycle phase and number of nuclei. Same organoid imaged every 12 hours. (B) Same organoid imaged for days 8-14 showing loss of synchronicity and resolution of individual cellular changes.

Although I was able to continue imaging post 7 days, the tracking of cellular changes became challenging as the time points were too far apart making resolution of each cellular behaviour insufficient (**Figure 5.18b**). To overcome this, future attempts would have to increase the time points after 5 days. Although it was an aim to complete this, such experiment is extremely time consuming and laborious. Therefore, to overcome this and hence make the approach more practical, we aim to adapt the protocol further to remove the need of removing of the sample, resulting in continuous imaging. Despite this set back, we have shown that Gellan works as a successful immobilisation medium to allow high resolution longitudinal imaging of organoids. To validate that the protocol and hence Gellan does not impact growth, **Figure 5.19** shows comparison of Day 14 organoids using conventional culturing and those cultured in the designed protocol. Both size and morphological characteristics were consistently retained across both.

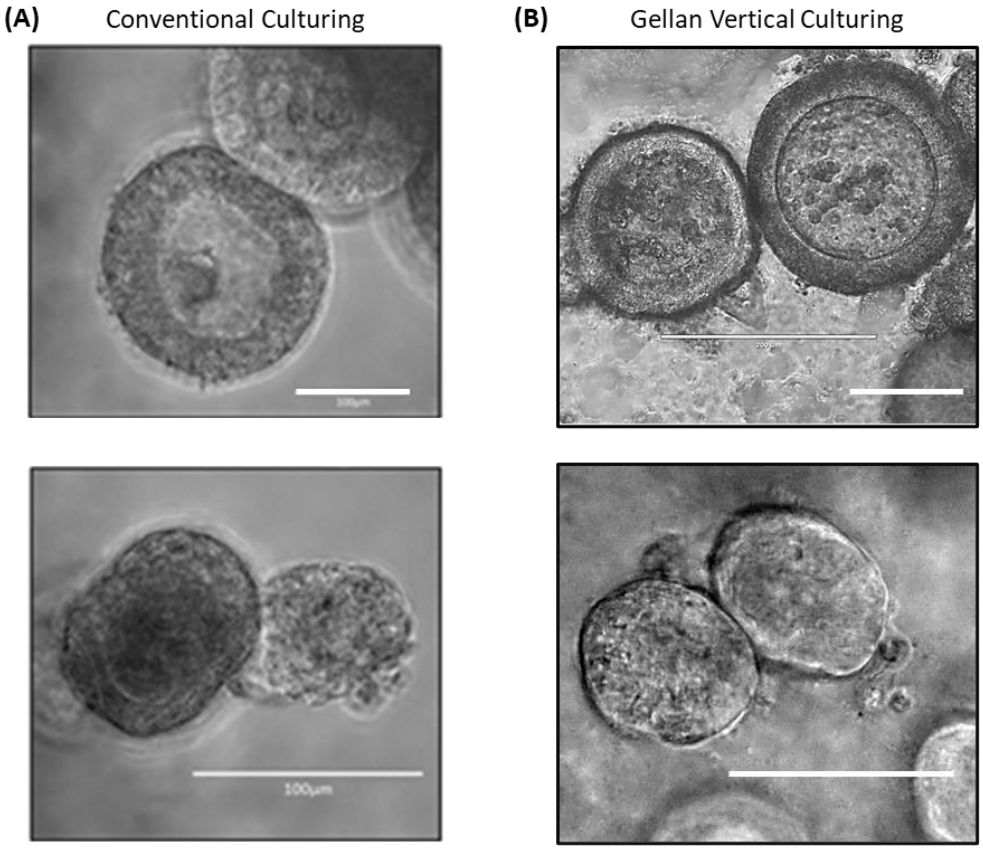


Figure 5.19. Morphological characteristics of organoids are retained using Gellan embedded vertical culturing. Day 14 snapshot image of organoids grown in different culturing protocols. (Top = organoid 389, Bottom = organoid 557). (A) Broadfield snapshot of organoids cultured as per standard culturing conditions embedded in BME and in a 24 well plate. (B) Lightsheet brightfield snapshot of matched organoids grown using Gellan embedded vertical culturing protocol designed in this thesis. Scale in all 100µm.

5.3 Discussion

The layman definition of cancer is defined as the uncontrolled growth of a cell (Brown et al., 2023). This highlights that even in its simplest terms dysregulation of proliferative dynamics is a core factor underpinning cancer aetiology. Understanding the regulatory processes that facilitate cell cycle dysregulation has been a cornerstone of cancer research and a key target for anti-cancer therapies (Feitelson et al., 2015; Sazonova et al., 2024). Many studies have already established the transcriptional influence on cell cycle dynamics in both healthy and cancer cells and noted the timed upregulation of specific genes vital for cycle progression (Bar-Joseph et al., 2008; Fischer et al., 2022). Furthermore, links to the transcriptional landscape and maintaining a distinctive proliferative phenotype have been identified. Gene expression patterns underpin both retention of plasticity and transitions down

differentiation lineages (Castelli et al., 2021; Ruiz et al., 2011; Tzamali et al., 2014). Although efforts have identified gene regulatory networks involved in controlling this, many of the findings have focused on conventional transcriptional regulation and often ignored the role of the promoter. Therefore, the aim of this chapter was to identify whether TSS usage differs between cell cycle phases. Building on the work in the previous chapters that identified the significance of core promoter TSS variation in cellular physiological states. This work suggests a transcriptional link to phenotypic determination not previously explored given that this dynamic would not be detectable in conventional RNA sequencing. We have already identified proliferation associated TSS usage relevance, in Chapter 3, I found a significant YC enrichment in faster growing organoids (**Figure 3.1**). Following this in Chapter 4 the ITH analysis highlighted cells enriched in genes associated with cell division, showed a significant depletion in YC usage compared to other cells (**Figure 4.11**). Therefore, in this chapter I aimed to see if cell cycle phases provided a greater association to TSS usage to elucidate a potential TSS usage role in cycling dynamics.

TSS usage shows cell cycle phase shifting.

Our primary observations indicate that, despite differing growth phenotypes, the initial single cell scoring analysis revealed similarly separated cell cycle phases across the samples, with nearly equal numbers of cells in G1 (**Figure 5.1**). However, given the single cell sample preparation process, and despite considerable efforts to minimize cellular stress, it is possible that cell cycle phase distribution may have been affected, particularly in the more sensitive 389 line, which could influence our results. Stressed cells tend to undergo a G1 arrest and therefore may result in a larger proportion of RFP cells (Fleifel & Cook, 2023; Foy et al., 2023; Kuczler et al., 2021). Therefore, the higher-than-expected G1 cells in sample 389 maybe resulting from this. Comparison of fixed day 14 organoids from both lines shows representative differences in both the presence and distribution of cycling populations (**Figure 5.7**). The imaging analysis further highlights the need for linking spatiotemporal dynamics with transcriptomic data to truly understand cellular behaviours.

Despite this, in this chapter, I show TSS usage shifting in a cell cycle phase dependent manner (**Figure 5.2**). Although previous studies have outlined differential gene expression underpinning cell cycle phase differences (Fischer et al., 2022; Riba et al., 2022; Whitfield et

al., 2002). I show that TSS usage also differs between the cycle phases and may also provide a previously unexplored level of differences in non-differentially expressed genes.

I further demonstrated our TSS shifting findings through our FUCCI sorted CAGE analysis and provide stratification between the G1 and G0 phase (**Figure 5.9**). Initial findings showed little differences between the G1 and G0 populations, we believe this is likely due to many of the cells continuously cycling. Given the nature of organoid culturing they are likely already beginning to re-enter the cell cycle. To identify terminally non dividing cells, we could score or sort for a marker of senescence such as the two cyclin dependent inhibitors p16 and p21 (W. Huang et al., 2022). Interestingly however, when utilising the generated G0 meta gene to identify G0 cells in our single cell data, they predominantly presented in a cluster distinctively distant from the rest of the cells (**Figure 5.13**). It could be predicted that this is the population of cells exiting the cell cycle and the larger group dominated by G1 cells are those entering the cell cycle. Our findings support the idea that these were continuously cycling as our SOM clustering showed similarity in gene expression between both G1 and G0 (**Figure 5.12**). However, disparities were identified and highlighted phospholipase inhibitor genes (W. C. Hwang et al., 2022; Reynisson et al., 2016). As previously stated, these inhibitors have been highly implicated in preventing cellular proliferation and felt an appropriate G0 meta gene.

Utilisation of YC TSS usage differs in gene preference between G0 and G1

Despite having little DEG between the G1 and G0, differential promoter usage was present. Although both showed a marked YC enrichment when compared to the later cycling phases, upon scrutinization of the genes responsible for YC enrichment in each phase we saw distinct differences (**Figure 5.10**). In G0, YC usage appears to preferentially shift toward metabolic-associated genes, while in G1, YC shifting and subsequent enrichment are more prominent in regulatory genes, which we have aptly named the sensors and translators. We hypothesise that this TSS usage dynamic may represent an adaptive process that plays a regulatory role of both progressing into the cell cycle but also the speed of the phasing. Previous studies have highlighted that cap dependent translation is likely an efficient, yet energy demanding process widely regulated by growth signals it is therefore the optimal mechanism in nutrient rich or unstressed cells (Borden & Volpon, 2020; Cormier et al., 2003; Merrick, 2004; Richter & Sonenberg, 2005). With regards to the cell cycle, particularly in cancer where cell turnover is

almost constant, G0 represents a phase to ensure sufficient metabolic intermediates to progress into the cell cycle (Kaplon et al., 2015; Mercadante & Kasi, 2021; Wiecek et al., 2023). The shift to YC dominance in the metabolic dual genes (**Figure 5.10c**) shows a potential increased need for this and that when a hypothesised threshold is met cells enter the G1 phase where translation has been shown to be abundant (An et al., 2018; Kronja & Orr-Weaver, 2011). Therefore, resulting in a YC enrichment in the translation associated machinery including ribosomal biogenesis genes (**Figure 5.10c**) we hypothesise to aid in the translation of the previously generated YC transcripts. Many of the regulatory genes seen in **Figure 5.10c,** that show YC enrichment in G1 are nutrient sensing genes. As highlighted in Chapter 4 (**Figure 4.23 and 4.24**) we observed many of the nutrient sensing genes that are dual, including *ATF4 and SREBF1* have very little distance between the relative YC and YR TSS. We hypothesise that the differential usage observed here is unlikely due to differences in transcriptional regulation. The TSS close proximity likely limits variations in transcription factor binding or conformational changes that could facilitate interactions with other CREs. Therefore, we believe at these nutrient sensing genes, differential usage could be resulting from nucleotide bioavailability a process highly dependent on the metabolic state of a cell, potentially determined in G0.

In this chapter, I also identify an increased YC enrichment in the G1/G0 of the more proliferative line compared to the G1/G0 of the slower growing line (**Figure 5.2b**). The abundance of YC transcripts in these phases in a high-growth signalling line may suggest a role for TSS usage in regulating cell cycle phase timing. Previous studies have identified a significant shortening of the G1 phase length in rapidly proliferating cells (Fleifel & Cook, 2023; Viner-Breuer et al., 2019; Zaveri & Dhawan, 2018). Efforts to establish the regulatory processes that aid this phase shortening have identified expression of oncogenes such as *CCNE1* critical to G1 length. Moreover, recent research suggests that core pluripotency reprogramming factors, particularly *c-MYC*, may play an integral role by targeting DNA licensing genes (Fleifel & Cook, 2023). We hypothesise that TSS usage could also influence cycling dynamics, with YC enrichment potentially facilitating the rapid upregulation of key genes to shorten such phases. However, further analysis is required to confirm this hypothesis. For instance, repeating FUCCI-CAGE experiments on both lines and comparing phase-specific TSS usage may provide insights. Additionally, it would be important to assess whether cell cycle phase

lengths differ between the lines and if these differences change as the organoid grows, highlighting the need for spatiotemporal image analysis to fully characterise the relevance of these phenotypes.

YC is lower in later stages of the cell cycle.

Our results consistently show YC usage is lowest during cellular states associated with cell division. Interestingly this process shows a significant transcriptional remodelling almost similar to that seen in stressed cells whereby gene expression is relatively conserved to specific subsets that aid in this function (Boström et al., 2017; Fischer et al., 2022; Ito et al., 2001). Additionally, studies have shown that cap-dependent translation is significantly depleted in the G2M phase of the cell cycle (Cormier et al., 2003). Given that previous research has shown YC transcripts, specifically those with a TOP motif, are targeted in a cap dependent manner, our findings of low YC in these phases supports the literature. Many of the genes that remain actively translated contain an internal ribosome entry site (IRES), which allows them to bypass the core translation machinery necessary for cap-dependent translation (Cormier et al., 2003; Walters & Thompson, 2016). This continued expression is often observed in DDR genes and represents an adaptive mechanism frequently exploited in cancer. This is particularly relevant in cases where a shortened G1 phase leads to reduced DNA licensing and increased replication stress (Fleifel & Cook, 2023). Understanding the regulatory mechanisms that enable the sustained expression and translation of these genes holds promise for developing novel anti-cancer therapies. Work so far has focused on the translation level role however we show a potential TSS and hence transcriptional role. Our identification of the shift to YR dominance in these phases potentially indicates a conservative TSS shifting potentially utilising less energy demanding cap independent translation. It is also possible that the low YC in these later cycling phases are a result of the translation of the increased YC transcripts produced in the former phases. This highlights an intricate regulatory network and highlights potentially previously unappreciated genes involved in cell cycle regulation, which could play a crucial role in cancer dysregulation. Many of these genes may have been overlooked as therapeutic targets because the dynamic nature of TSS usage is not typically captured in RNA sequencing studies.

Gellan represents a suitable immobilisation medium for organoid imaging.

All our findings highlight how dynamic the cycling process is and to fully understand its significance in governing tumour growth we must understand its relevance beyond time point bias restrictions and hence explore the spatiotemporal patterning of these states. We attempted to develop a longitudinal high-resolution imaging protocol that would enable us to map cycling dynamics through fluorescent reporter tracing whilst retaining true recapitulation of the complexity of 3D cultures like organoids. Other research has opted to develop an alternative lightsheet microscope the Bruker InVI SPIM Lattice Pro suited for accommodating the complex culturing conditions that facilitate organoid growth. Others have also resorted to using plate-based microscopes such as the Cytation (Ramm et al., 2022). The latter comes with significant limitations as imaging is only done from illumination upwards, whereas the lightsheet enables full 3 axis visualisation meaning imaging is not hindered by auto focusing limitations (Ramm et al., 2022; Rios & Clevers, 2018). As we do not have access to the newly developed organoid lightsheet, we aimed in this chapter to develop a protocol that would enable us to use our preexisting Z1 lightsheet an often widely present machine in many imaging facilities.

Although further testing and repeats are needed, we successfully utilised our protocol to image our organoid cultures up to 14 days (**Figure 5.18**). Furthermore, we show that our experimental conditions to not appear to impose growth defects on our lines (**Figure 5.19**), given that organoid size does not drastically differ between ours and the conventional culturing approach. Furthermore, distinctive morphological characteristics we previously identified in our lines were also retained, likely owing to our approach encompassing the use of BME essential for retaining cellular adhesions. In addition to further repeats incorporating more time points for further resolution of population changes between the two different samples, the approach can be further adapted to explore other questions. For example, as the Gellan is permeable to the media exploration of the impact of specific drug therapies on the cycling dynamics could be investigated. The adaptability of the Gellan microstructure represents an opportunity to alter its composition to aid in imaging other structures including other 3D cultures or even small organisms such as zebrafish embryos.

Chapter 6: General Discussion

Even small decisions can have a big impact.

The work in our lab shows that even at the smallest resolution of a cell, the nucleotide base choice governing the formation of the building blocks of a cell, can have an impact on its fundamental growth and survival function. Our understanding of the core regulatory elements controlling gene expression has extended to a newfound appreciation of the promoter. Previous work has often overlooked its significance, despite its vital integrative role for transcriptional signalling and its harbouring of initiation site (Haberle & Stark, 2018; Tirosh et al., 2009). Yet an interesting choice exists in 1000s of genes: the site of transcription initiation. This base resolution difference has led to the classification of promoters through alternative TSS defined simply by their initiator sequence, with potential differential pre and post transcriptional regulation (Nepal et al., 2020; Wragg et al., 2023). To better understand the functional relevance of these dual initiator genes in different contexts, our lab has begun identifying trends associated with differential TSS enrichment in both development and disease. This previous research identified an enrichment of the YC TSS usage in cancer which then raised questions as to whether there was a role of alternative TSS usage in cancer aetiology (Wragg et al., 2023). Such hypothesis was further supported by the identification of heterogeneity in the YC enrichment between cancers and even in vitro cultures such as organoids (Wragg et al., 2023). Despite providing a novel biomarker to irradiation responsiveness our understanding of the functional dependence and hence biological significance of TSS usage specifically YC enrichment in cancer was limited.

The core aim of the work in this thesis was to greaten our understanding of differential TSS usage in cancer cell behaviours, in hopes to identify potential regulatory mechanisms that may be novel therapeutic targets. The overarching finding in this work identified that TSS usage differs between cells in different states. This is beginning to paint a picture that TSS may have a more distinguished role in cellular functioning than previously understood. Although the work is still very preliminary, so far, we have shown correlative associations between global TSS usage and specific cellular states, specifically cell cycle phases. To better our understanding of the impact and consequence TSS usage has on cell function, more interrogative experiments must be undertaken. Despite this, the following discussion will highlight the current trends I have identified in this thesis, what I hypothesise this could

mean and the essential experiments that are needed to further our understanding of the fundamental biological role of TSS in gene regulation.

YC enrichment underpins growth and stem-like phenotypes.

In this thesis I attempted to identify a potential link between TSS usage and cellular physiological state and even fate through association to differentiation status. The resounding finding of the work presented here is that YC enrichment consistently correlates with fast growth dynamics, both on a single cell and global level. A fundamental principle of cancer evolution is the sustaining of continual proliferation (Feitelson et al., 2015; Hanahan & Weinberg, 2011; Keibler et al., 2016). Many efforts have attempted to identify the molecular underpinnings of dysregulated growth dynamics and consistently highlighted the significance of gene expression profiles (M. Li et al., 2017; J. Wang et al., 2015). This regulatory process is hypothesised to exist in an equilibrium, specifically in cancer where cell trajectories are highly dependent on balancing growth promotion and survival through metabolic adaptation (Faubert et al., 2020; Keibler et al., 2016; Zhu and Thompson, 2019; Schiliro and Firestein, 2021; Loftus et al., 2022). Therefore, the transcriptional landscape adapts through the upregulation of genes involved in these cellular functions (Faubert et al., 2020; Gallaher et al., 2019; Keibler et al., 2016; Morandi & Indraccolo, 2017; Nong et al., 2023; Phan et al., 2014). By understanding the mechanisms that control and support this adaptive process, especially in cancer, we can identify potential targets for intervention that might drive cancer cells to undergo apoptosis. As my results consistently show alternative TSS usage, in the form of YC enrichment, associating with fast growth dynamics, it raised the question as to whether TSS usage may play a role in this growth regulatory mechanism. To explore this further I interrogated the genes that had dual TSS and showed alternative TSS usage between the fast and slow growing populations. My results showed that many of these genes were linked to metabolic homeostasis (**Figure 3.14b, Figure 4.18b and Figure 5.22b**). The differential usage of TSS appears to be attributable to the state of the cell. In growth demanding conditions, a constitutive phenotype in many cancers, YC appears to be the preferred and hence dominant TSS in dual initiator genes (**Figure 3.1**). I show that this usage links to an increase in growth signals driving cellular proliferation and drives a global growth phenotype (**Figure 3.10**). ITH analysis showed that YC enrichment correlated with other features including less heterogeneity, represented by a lower number of clusters, and

absence of clear cellular differentiation (**Figure 3.12a, Figure 3.16 and Figure 3.17**). Contrastingly, YC usage is much lower in the slower growing, more differentiated populations. Given these findings, I hypothesise that enrichment of YC transcripts may confer a growth advantage. Such reasoning could be supported by evidence showing that these transcripts contain binding sites heavily targeted by mTOR regulated TIM, coupled with findings suggesting this specific translation mechanism maybe more efficient (Bouyahya et al., 2022; T. Tian et al., 2019)**.**

This choice may also come at a cost, YC enrichment also correlates with treatment responsiveness and an increased sensitivity to mTOR inhibition (**Figure 3.2**) (Wragg et al., 2023). Therefore, blocking of these signalling pathways directly or indirectly through stress, results in an inability to survive potentially resulting from the inability to translate the abundant YC transcripts. Further research is needed to identify whether the YC enrichment directly plays a role in treatment response, or if it is simply just a consequence of another regulatory factor. Although experimentally challenging, direct interference with YC TSS usage through alteration of YC motifs in genes of interest may answer this. Furthermore, an experiment to knock out key genes involved in the mTOR regulated TIM, would show the importance of the translational level of regulation. For example, many of these genes are believed to be targeted by LARP1 to enable mTOR associated translation. Knock out of LARP1 followed by CAGE analysis of these genes would be insightful in understanding whether their expression is continued but with a direct shift in TSS usage to YR.

In this work I successfully carried out single cell CAGE on cancer cells and explored cellular specific TSS usage even down to a single cell, a resolution yet to be achieved (**Figure 4.3 and Figure 4.14**). I found that despite high sequencing depth, characterisation of single cell TSS usage at a gene level is challenging and future work would strive for even deeper sequencing. Work in chapter 4 shows that TSS usage is heterogeneous even in a globally YC enriched sample (**Figure 4.3**). Characterising the cellular physiological state to identify expression profiles linked to YC enrichment was challenging due to the lack of a unified method for defining cell states (**Figure 3.12 and Figure 3.13**). Conventional single cell clustering may ignore underlying GO, however utilising archetypes provided reference states with core functions already defined in cancer aetiology (**Figure 4.19 and Figure 4.21**). Archetype analysis further supported our belief that TSS usage contributes to the

proliferation equilibrium. Work by Hausser et al defined key marker genes for the universal archetypes and emphasised the importance of overlapping genes, genes that are upregulated in multiple archetypes (Hart et al., 2015; Hausser et al., 2019). I show that these overlapping genes exhibit distinct dual initiator capacity, with their relative usage attributable to the cell's state or archetype. This analysis showed YC usage was highest in cells associated with metabolic archetypes and lowest in cells undergoing cell division (**Figure 4.20 and Figure 4.22**). This further supports our previous global analysis, highlighting the potential role of YC in meeting metabolic demands to facilitate cellular growth. Current analysis however is still correlative and direct links of the importance of TSS usage and maintaining metabolic balance is limited. Future experiments that directly disrupt cellular growth—such as limiting metabolic intermediates and assessing their impact on TSS dynamics—could offer invaluable insights into the relationship between TSS usage and cellular state adaptation. Potential approaches include altering core metabolites in the culture media, such as serum deprivation or limiting amino acid supplementation. To remove the complexity that accompanies organoid analysis, the use of 2D lines may also provide benefits to these interference experiments.

YC enrichment differs in a cell cycle dependent manner.

Although the two organoid lines analysed in the single-cell experiment exhibited distinctly different global phenotypes and TSS usage profiles, I identified shared ITH states with similar TSS usage dynamics (**Figure 4.11**). These shared states consistently correlated to cell cycle dynamics. Therefore, in Chapter 5 I characterised ITH by cycle phases and found YC usage was directly attributable to both G0 and G1 phases (**Figure 5.2, Figure 5.9 and Figure 5.14**). I hypothesise that this indicates a potential role of cell cycle dependent TSS usage to increase metabolic macromolecule production, enabling cells to meet the energy demand required for cell division. Highly proliferative cancers appear to have shortened growth phases in their cell cycle and is a hallmark of their stem-like plasticity (Coronado et al., 2013; Fleifel & Cook, 2023; Y. Hwang et al., 2020; L. Liu et al., 2019). Entry into and out of a cycle phase however is dependent on metabolic thresholds being met. Our findings that YC is not only enriched in faster growing cells but is also attributable specifically to the growth phases of the cell cycle, indicate that YC transcripts may directly or indirectly facilitate the meeting of this threshold faster. Previous research has shown that TOP containing transcripts harbour a significant

increase in translation efficiency through utilisation of cap-dependent machinery (Bouyahya et al., 2022; T. Tian et al., 2019). Given that our analysis shows YC usage dynamics behave similar to TOP usage they may share a potential unified regulatory mechanism, one that is more translationally efficient. Although speculative, these associations could suggest that cells utilise YC transcript production as a more efficient mechanism to facilitate fast growth dynamics. Combining polysome profiling with CAGE sequencing would provide evidence as to whether specific initiator motifs in transcripts are being differentially translated. Such experiments have already begun to be carried out, Wang D, et al found that translation efficiencies altered dependent on motifs in the 5' UTR of genes linked to synaptic activity (X. Wang et al., 2019). Using this experimental framework but comparing between fast and slow growing cells would allow for further insight into whether TSS directly links to growth dynamics.

The relative shift in YC usage in specific subsets of genes between G0 and G1 may further fuel this speculative functional dynamic that YC is favoured for efficiency. The observed high YC usage in metabolic genes in G0 could account for the need to provide sufficient metabolic and energy requirements to enter the cell cycle (**Figure 5.10**). The resultant shift to high YC in translation and regulatory genes seen in G1 could facilitate the translation of the previously upregulated metabolic transcripts, thereby promoting cell cycle progression. Finally, a YR usage shift in the remaining division phases likely represents a conservative transcriptional landscape, supporting previous literature that identified the decrease in cap dependent translation in S and G2M phases (Cormier et al., 2003; de la Parra et al., 2018). This dynamic shifting of TSS usage between cycling phases maybe a potential regulatory mechanism that contributes to cell cycle phase length and hence overall tumour growth phenotypes. To explore and validate this, future experiments could include comparing the TSS usage of cycling populations between differentially proliferating organoids. Performing CAGE on the FUCCI populations in conjunction with image based temporal mapping of cycling phase lengths would provide a more causative link.

TSS usage may shift in cellular stress.

While the focus so far has been on the role of the proliferation-equilibrium dynamic in driving growth and cellular division, this thesis also highlights the emerging recognition of the reverse trend. The shift from YC to YR in states of stress was initially outlined by preliminary

experiments in the lab that showed YC was depleted post irradiation (Wragg et al., 2023). As previously highlighted, many metabolism-associated genes possess dual TSS capacity, and other studies have shown their sustained transcription and translation in stressed cells (Ho et al., 2020; Ho & Lee, 2016; Tamarkin-Ben-Harush et al., 2017). The inactivation of the key cap-dependent translation pathway mTOR in these stressed environments, is the main driving force for global translation suppression (Tamarkin-Ben-Harush et al., 2017). Stressors including hypoxia and other nutrient deficiencies exert significant energy demands on a cell, and the expression of these metabolic reprogramming genes shifts the equilibrium from proliferation to survival (Tamarkin-Ben-Harush et al., 2017). The observed decrease in YC levels post-irradiation, a state of cellular stress, alongside the inactivation of mTOR—a pathway that targets YC-TOP-containing transcripts—suggests that YR usage may be favoured under these conditions. This shift could indicate that YR transcripts possess an alternative translation mechanism. Further experiments are required to explore this possibility. Future work could look to performing RNA immunoprecipitation targeting the mTOR-specific TIM, followed by RT-qPCR using specialised primers designed by Wragg et al. for candidate dual-function genes (Nepal et al., 2020; Wragg et al., 2023). This would help quantify the relative abundance of 5' motifs in the captured RNA subsequently targeted by TIM.

Furthermore, a similar stress like environmental condition is seen in the later stages of the cell cycle, where growth signals and translation are also lower and energy is redirected toward essential processes like DDR (Cormier et al., 2003). Here we find a significant YC depletion and hence YR TSS preference (**Figure 5.2, Figure 5.9 and Figure 5.14**). We hypothesise that YR usage in certain genes is favoured under energy-conserving conditions to enable cap-independent translation, bypassing the inactive mTOR-driven translation pathway. Recent studies have also identified ATM as a specific regulator of these genes (Ho et al., 2016; Ho & Lee, 2016; Tamarkin-Ben-Harush et al., 2017). However, the research has remained on the translational level, but our results indicate that a transcriptional element could interlink (**Figure 3.10, Figure 3.20, Figure 4.6 and Figure 4.18**). The shift toward YR dominance in metabolic reprogramming genes under conditions such as nutrient deprivation, hypoxia, and other energy-conserving states may represent a TSS-based adaptive mechanism that promotes cellular survival (**Figure 3.20**). Repeating the above proposed experiment but targeting the ATM in the immunoprecipitation would provide

evidence for this hypothesis. Furthermore, scrutinization of the TSS usage between whole and nascent RNA production would provide invaluable insight into TSS shifting dynamics in direct response to environmental stimuli. Such experiment could encompass the methodology outlined in slam-seq. The premise of this experimental procedure allows for the incorporation of a modified uracil nucleotide to establish the nascent RNA products (P. Bhat et al., 2023; Wissink et al., 2019). Therefore, the timed addition of the nucleotide following a stressor such as hypoxia would enable the identification of newly synthesised RNA transcripts and their preferential TSS usage. This would allow us to identify a direct shifting highlighting a transcriptional level of regulation to cellular survival.

So far, our TSS dynamics shows usage attributable to physiological states that indicates potential differential regulatory mechanisms possibly dependent on their 5’ motifs. Research has shown the role of specific translation initiation binding machinery and regulators such as LARP1 in controlling the translation of TOP containing transcripts (Hochstoeger et al., 2024; Jia et al., 2021; Mura et al., 2015; Ogami et al., 2022). Previous work in the lab, now further supported by this thesis, has shown that other YC transcripts exhibit similar behaviour (**Figure 5.11**), suggesting a potentially unified regulatory mechanism. Further characterisation is needed to determine whether these mechanisms affect all YC transcripts or only those containing a TOP motif. If the latter, the significance of the motif's distance from the 5' end in relation to differential regulation must also be established. Furthermore, we have yet to fully understand the impact of TSS usage on transcript diversity. Utilising a long-read sequencing approach to capture exon boundaries could reveal valuable splice variants and further illuminate differential post-transcriptional regulation (K. K. Huang et al., 2021; Marx, 2023; Reese et al., 2023).

Furthermore, the identification of functionally relevant genes involved in differential TSS usage opens the opportunity to further explore their CRE landscape. In this thesis I have identified an interesting feature present in nutrient sensing genes, whose resultant cascade can significantly rewire transcriptional outputs. The promoter architecture highlighted in **Figure 4.23** and **Figure 4.24** shows dominant TSS usages with little separation. It is unlikely that in these genes differential CRE can act on the different TSS, therefore it is hypothesised that their usage is dependent on nucleotide bioavailability, a metabolically regulated process. The data produced in this thesis represents an exciting resource to further

scrutinise. Firstly, on a gene level to identify potential CRE such as TF binding sites associated with specific TSS usage and through association with our characterised cellular states, their usage in an environmentally context dependent manner. Secondly on a single cell level to establish potential co-occurrence patterns of specific TSS usage.

Gellan represents a suitable holding medium for facilitating vertical culturing to enable long-term high-resolution imaging.

A major limitation of sequencing is its inherent bias toward a single time point, which restricts our ability to fully appreciate the significance of TSS usage in dynamic cellular states. Research has highlighted the importance of appreciating the plasticity of cancer cell behaviour (G. R. Bhat et al., 2024; Clairambault & Shen, 2020; Holton et al., 2024; S. Qin et al., 2020; J. shun Wu et al., 2021). Archetype analysis on cancer cells has emphasised the impact environmental factors have on cellular states and the plasticity of the transcriptome to adapt to these pressures (Combes et al., 2021; Groves et al., 2021; Hausser et al., 2019; Hausser & Alon, 2020). Spatiotemporal mapping of cellular states can therefore provide invaluable insight into regulatory dynamics in tumour evolution. Spatial transcriptomics is gaining significant traction with recent advances even enabling single cell resolution (Vahid et al., 2023; Zhou et al., 2023). Although we have characterised interesting ITH states in our organoids, we cannot conclude on their full relevance without understanding the SC population location and presence in all organoids. For example, do all the organoids present in a specific line harbour the same composition of cellular physiological states? How representative are our clusters to the whole line? Therefore, a spatial transcriptomic experiment would allow us to identify the distribution of our cellular states using transcriptomic markers we have established. This would allow us to see if the populations are pervasive throughout each organoid in that line or if specific organoids show different cluster dynamics highlighting a potential bias in our data. To achieve true resolution however, a long-term high-resolution mapping experiment will enable full appreciation. The vertical culturing protocol developed in this thesis represents a novel approach to visualising organoid dynamics, given its compatibility with Z1 lightsheet imaging set up. We have shown that the biosynthetic and natural polymer Gellan is an exciting malleable holding medium that can facilitate imaging set up demands without hindering culturing development (**Figure 5.18** and **Figure 5.19**). A significant challenge encountered was the loss of tracking individual

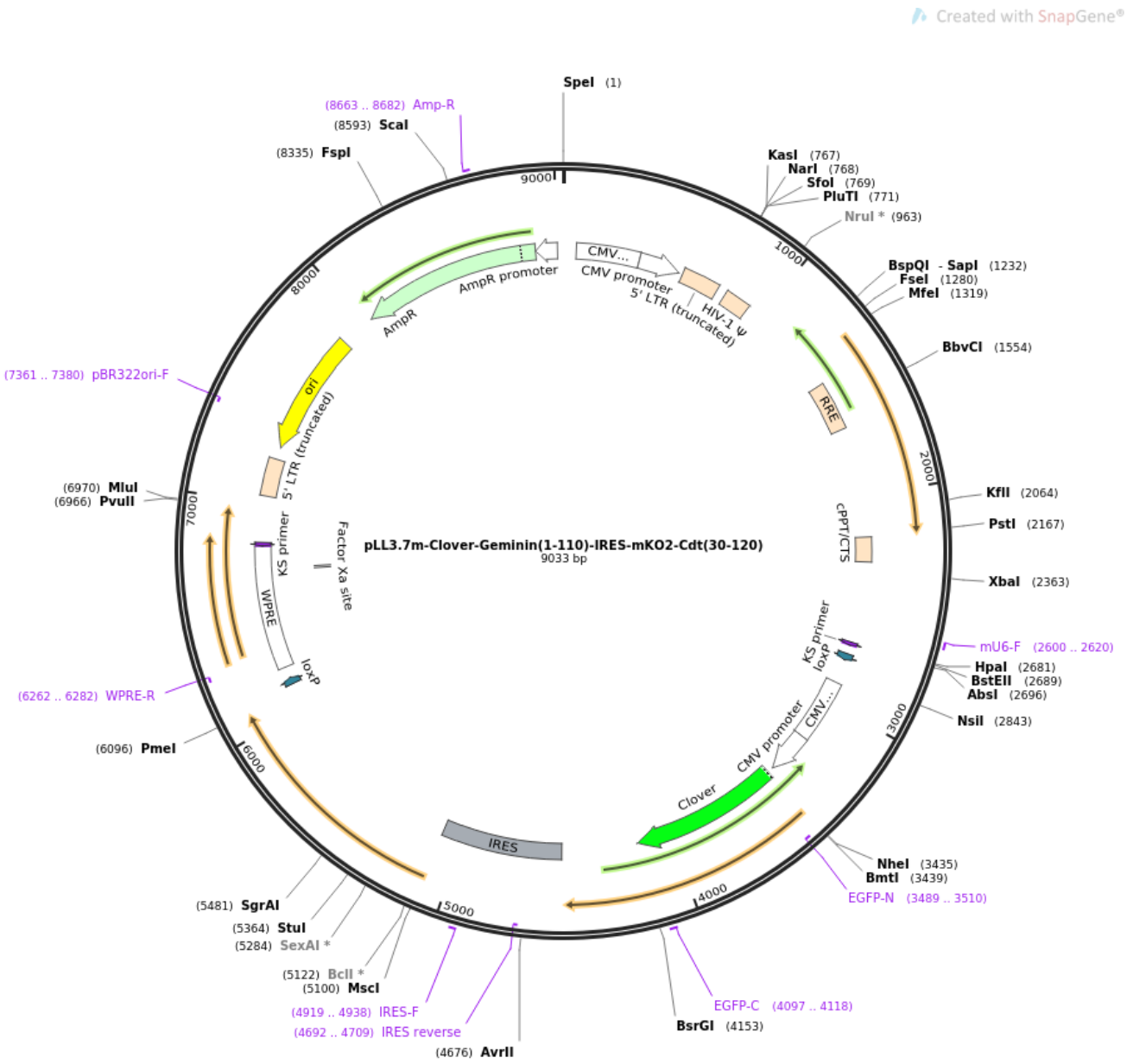
cells post 7 days given the loss of synchronicity in cycling dynamics meaning cycling changes were often missed. To overcome this, we could increase the imaging time points however this would require extensive experimental hours. We aim to further adapt the protocol to enable constant imaging, thus removing the need for the altered culturing in a falcon tube. This could be achieved through adapting the lightsheet imaging chamber to lower the volume needed thus reducing the amount of media making it feasible. Given the malleability of Gellan microstructure we show the adaptability of this resource to hopefully enable future use to other cellular structures both tissue culture based and whole organism such as zebrafish. Further adaptations on this method could also allow for redundancy in the use of the falcon tube and enable direct culturing in the lightsheet to achieve true 4D imaging.

Conclusions

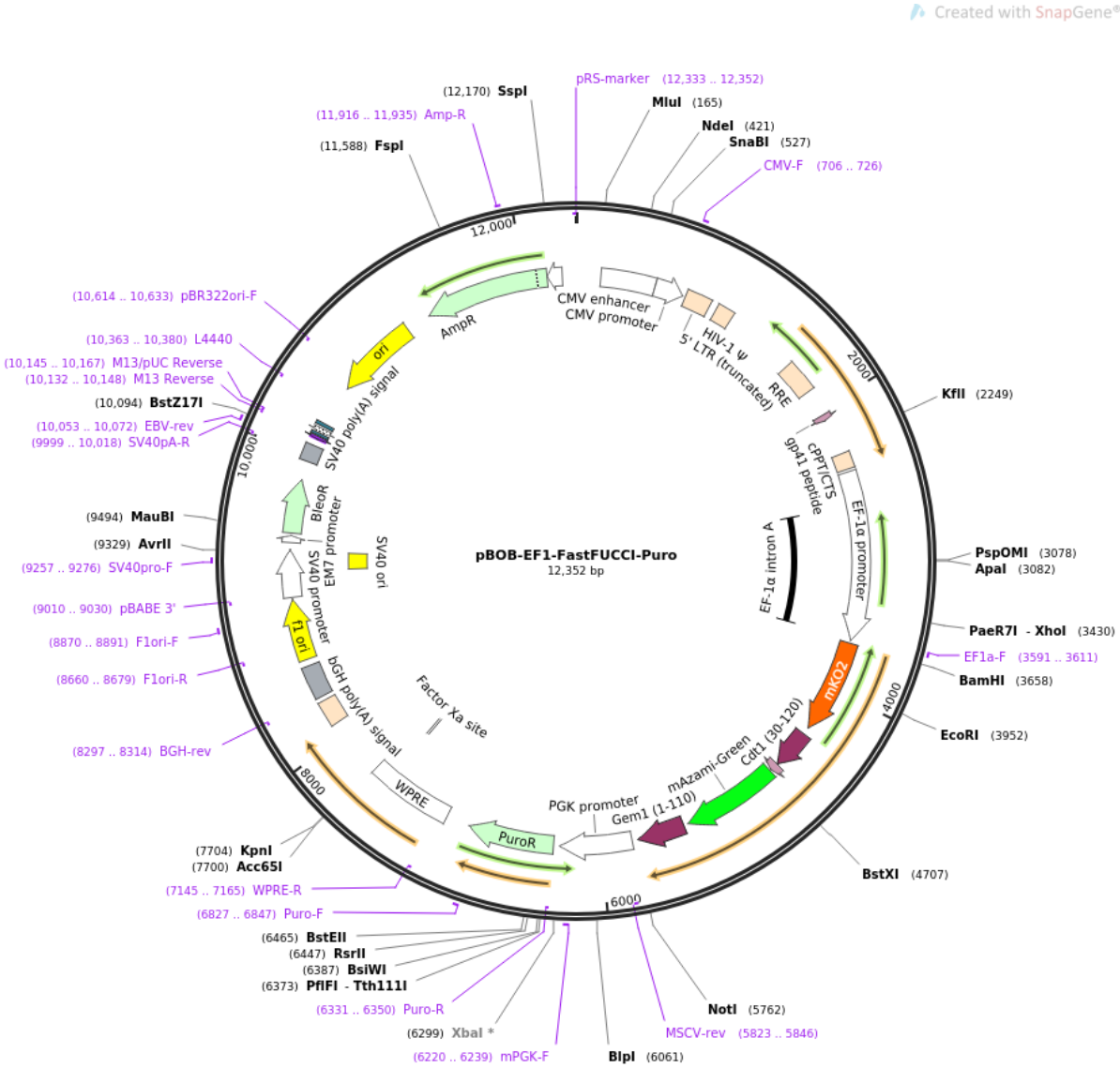
This thesis sheds new light on the previously underappreciated role of TSS usage dynamics in shaping transcriptional landscapes that define cellular physiological states, even at the resolution of a single cell. Our findings suggest the presence of a potentially dynamic mechanism driving the growth processes commonly seen in cancers, along with a shift that may occur to balance cellular growth and survival in a cell cycle-dependent manner. However, it is important to acknowledge that many of these insights remain associative, and moving beyond correlative analysis will require more direct experimental interventions to validate these observations. Despite the current limitations, this work lays the foundation for a deeper understanding of how promoter activity interconnects with other regulatory mechanisms, functioning intricately at the single-cell level. The results presented here are hopefully just the beginning, offering a new perspective on the significance of TSS usage in broader regulatory networks, and paving the way for future studies to unravel the complexities of transcriptional regulation in health and disease.

Appendix

Appendix A



Appendix A.1: pLJM1-FUCCI plasmid map. PLL3.7m-clover-geminin (1-110)-IRES-mKO2 -Cdt(30-120)) (Addgene 83841) truncated versions of the reporter genes were generated to facilitate the co expression with additional Plasmid that allows for full 4 phase resolution of the cell cycle outlined and generated in (Bajar et al., 2016).



Appendix A.2: Fast-FUCCI plasmid map. pBOB-EF1-FastFUCCI-Puro (Addgene plasmid, 86849) plasmid map generated and utilised by (Koh et al., 2017).

Appendix B

Table 1: Outline of patient cohort data for PDCO generation. Including Pre operation treatment (SCRT = Selective Combined Radiotherapy, LCCRT = Locally Combined Chemoradiotherapy), irradiation responsiveness of organoids in culture, relative TNM status of source tumour and CMS classification and mutational status of key driver genes.

Sample ID	Patient Age	Gender	CMS	PreOp Treatment	Irradiation Responsiveness	Pre-op TNM	Location	Mutation Count	Driver gene mutational status
S292**064**	79	Male	CMS2	None	Responsive 1 (R1)	T3N0M0	Sigmoid	11977	TP53 missense
									APC stop gained
									FBXW7 missense
									TGIF1 frameshift
S302**389**	81	Male	CMS2	None	Responsive 2 (R2)	T3N1M0	Colon	17275	TP53 missense
									APC frameshift
									SOX9 frameshift
									TCF7L2 frameshift
									SMAD4 missense
									MSH2 frameshift
S309**884**	74	Male	CMS2	SCRT	Moderately responsive	T3N1M0	Rectal	16028	TP53 missense
									NRAS missense
									APC stop gained
									MSH6 frameshift
									SOX9 frameshift
									GNAS missense
S345**653**	55	Female	CMS4	LCCRT	Non-responsive (NR1)	T2N0M0	Rectal	11993	TP53 inframe deletion
									KRAS missense
									APC frameshift
									SOX9 stop gained
									SMAD4 frameshift
									FBXW7 missense
S366**557**	75	Female	CMS1	None	Non-responsive (NR2)	T3N0M0	Colon	19511	KRAS missense
									APC missense
									MSH6 frameshift
									PIK3CA missense
									BMPR2 frameshift
									BCL9L frameshift
									PTEN missense
									RPL22 frameshift
									ACVR2A frameshift

Appendix C

Table 1: Full list of consensus dual initiator genes. Consensus cluster ID and HGNC gene name of genes found to have both YC and YR TSS across both organoid and pan cancer analysis.

ccID	Gene Name
671	UBAP2L
672	HAX1
681	ADAM15
682	EFNA4
684	EFNA1
686	DPM3
687	KRTCAP2
692	FAM189B
693	SCAMP3
698	DAP3
703	SSR2
704	LAMTOR2
705	RAB25
706	LMNA
710	PMF1-BGLAP,PMF1
712	GLMP
713	CCT3
725	HDGF
731	TAGLN2
734	DCAF8
743	PFDN2
746	UFC1
748	B4GALT3
750	SDHC
753	UHMK1
756	MGST3
758	TMCO1
763	CREG1
765	MPC2
780	PRDX6
787	CACYBP
789	KIAA0040
796	QSOX1
800	ACBD6
806	DHX9
812	C1orf21
815	RNF2
820	PTGS2
821	PLA2G4A
846	ELF3
858	TMEM183A
862	SNRPE
863	SOX13
865	PPP1R15B
871	CDK18
874	NUCKS1
876	EIF2D
881	CD55
885	LAMB3
900	NENF
913	EPRS1
915	IARS2
916	RAB3GAP2
924	CAPN2
933	WDR26
938	EPHX1
944	H3-3A
953	ARF1
956	GUK1
958	H2AW
959	H2BU1
961	RAB4A
963	GALNT2
969	GNPAT
974	TSNAX-DISC1,TSNAX
978	PCNX2
983	TOMM20
984	RBM34
990	GPR137B
996	FH
998	SDCCAG8
999	ADSS2
1002	COX20
1004	HNRNPU
1015	SH3YL1
1016	ACP1
1017	TMEM18
1022	RPS7
1032	IAH1
1040	HPCAL1
1051	NBAS
1063	SF3B6
1065	PTRHD1
1069	RAB10
1070	HADHA
1071	HADHB
1078	OST4
1089	PPM1G
1090	NRBP1
1091	KRTCAP3
1106	YPEL5
1110	YIPF4
1117	HEATR5B
1124	QPCT
1125	CDC42EP3
1137	COX7A2L
1141	LRPPRC
1157	ERLEC1
1159	SPTBN1
1161	RTN4
1163	MTIF2
1174	CCT4
1176	B3GNT2
1178	MDH1
1180	UGP2
1185	RAB1A
1187	SPRED2
1191	PPP3R1
1198	PCBP1
1203	FAM136A
1209	MPHOSPH10
1221	TPRKB
1233	INO80B
1234	WBP1
1236	MRPL53
1240	AUP1
1253	KCMF1
1256	RETSAT
1261	GGCX
1263	RNF181
1265	USP39
1267	IMMT
1270	CHMP3
1271	RMND5A
1280	CIAO1
1281	SNRNP200
1287	COX5B
1295	MRPL30
1297	EIF5B
1299	NPAS2
1303	RNF149
1307	MRPS9
1308	C2orf49
1310	FHL2
1321	MALL
1330	CHCHD5
1337	ACTR3
1343	DBI

ccID	Gene Name
1350	NIFK
1352	BIN1
1361	MZT2B
1367	MZT2A
1371	RAB3GAP1
1375	DARS1
1379	MMADHC
1387	ARL6IP6
1399	GCA
1400	GALNT3
1402	STK39
1407	PPIG
1411	SSB
1412	METTL5
1417	DYNC1I2
1418	DYNC1I2
1421	ITGA6
1423	MAP3K20
1424	CDCA7
1425	OLA1
1430	LNPK
1431	MTX2
1441	ITPRID2
1443	NCKAP1
1445	ZC3H15
1447	TFPI
1450	GULP1
1452	WDR75
1463	STAT1
1467	STK17B
1470	SF3B1
1474	HSPE1
1480	SPATS2L
1486	BZW1
1487	CLK1
1490	ORC2
1492	NDUFB3
1500	NOP58
1504	ABI2
1506	NDUFS1
1507	EEF1B2
1508	EEF1B2
1510	CREB1
1512	IDH1
1518	ATIC
1520	PECR

ccID	Gene Name
1523	RPL37A
1525	IGFBP2
1530	TMBIM1
1539	RETREG2
1557	MRPL44
1559	MFF
1566	ITM2C
1571	PSMD1
1573	B3GNT7
1575	NCL
1583	GIGYF2
1585	ATG16L1
1588	HJURP
1594	MLPH
1596	LRRFIP1
1605	NDUFA10
1606	COPS9
1611	GPR35
1612	PPP1R7
1615	HDLBP
1619	BOK
1621	THAP4
1631	ARL8B
1638	ARPC4
1644	BRK1
1645	VHL
1658	RPL32
1665	SLC6A6
1668	NR2C2
1672	SH3BP5
1674	EAF1
1678	TBC1D5
1683	UBE2E1
1690	SLC4A7
1696	OSBPL10
1698	CMTM8
1700	CMTM6
1701	DYNC1LI1
1703	GLB1,TMPPE
1706	PDCD6IP
1708	LRRFIP2
1711	ACAA1
1718	EIF1B
1719	RPL14
1722	TRAK1
1726	HIGD1A,ENSG0000 0280571

ccID	Gene Name
1728	TCAIM
1731	KIAA1143
1753	MAP4
1756	TMA7
1761	UQCRC1
1764	IP6K2
1767	ARIH2
1769	WDR6
1771	IMPDH2
1773	QARS1
1777	USP4
1780	RHOA
1787	APEH
1793	RBM6
1806	MANF
1810	TEX264
1813	ABHD14B
1820	WDR82
1830	GNL3
1832	SPCS1
1838	TKT
1849	ARF4
1860	PSMD6
1864	SUCLG2
1865	TMF1
1867	ARL6IP5
1876	GBE1
1883	CPOX
1884	DCBLD2
1888	NIT2
1889	TOMM70
1891	TFG
1893	PCNP
1894	ZBTB11
1896	RPL24
1912	NAA50
1913	ATP6V1A
1920	NDUFB4
1927	FAM162A
1934	CCDC14
1946	PODXL2
1949	SEC61A1
1952	RPN1
1953	RAB7A
1956	ISY1-RAB43,ISY1
1958	COPG1

ccID	Gene Name
1970	CDV3
1981	NCK1
1988	SLC25A36
1992	RNF7
2003	PLSCR1
2004	GYG1
2012	WWTR1
2016	SERP1
2017	EIF2A
2020	MBNL1
2026	SSR3
2038	NMD3
2050	EIF5A2
2051	TNIK
2055	TNFSF10
2057	ECT2
2060	MFN1
2063	NDUFB5,ENSG0000 0288698
2066	FXR1
2068	ATP11B
2069	DCUN1D1
2072	B3GNT5
2082	ALG3
2089	POLR2H
2102	DNAJB11
2103	EIF4A2
2105	ST6GAL1
2112	HES1
2119	TFRC
2126	PAK2
2131	MELTF
2134	FYTTD1
2136	RPL35A
2138	ZNF721
2140	ATP5ME
2142	GAK
2158	RNF4
2159	TNIP2
2163	NOP14
2170	STX18
2173	MRFAP1
2186	WDR1
2200	MED28
2205	DHX15
2207	SOD3
2212	SMIM20

ccID	Gene Name
2220	FAM114A1
2223	LIAS
2227	UBE2K
2237	SLC30A9
2244	OCIAD1
2258	PAICS
2259	SRP72
2261	NOA1
2264	UBA6
2265	YTHDC1
2267	UTP3
2268	GRSF1
2271	COX18
2276	CXCL3
2287	USO1
2299	ANXA3
2305	HNRNPD
2307	HNRNPDL
2309	ENOPH1
2310	SCD5
2314	PLAC8
2317	ABRAXAS1
2320	HSD17B11
2322	PYURF,PIGY
2329	EIF4E
2330	METAP1
2331	ADH5
2341	MANBA
2346	CISD2
2348	PPA2
2353	AIMP1
2354	PAPSS1
2355	HADH
2356	RPL34
2357	OSTC
2358	SEC24B
2362	GAR1
2372	CAMK2D
2379	USP53
2380	C4orf3
2383	ANXA5
2400	NAA15
2404	MGST2
2405	SCOC
2406	ELMOD2
2413	LSM6

ccID	Gene Name
2416	LRBA
2418	RPS3A
2424	MND1
2426	PLRG1
2427	PPID
2434	CBR4
2435	CLCN3
2443	HMGB2
2449	SPCS3
2452	DCTD
2457	RWDD4
2460	IRF2
2463	CENPU
2468	CFAP97
2482	BRD9
2487	LPCAT1
2489	MRPL36
2490	NDUFS6
2491	IRX2
2494	NSUN2
2495	SRD5A1
2501	CMBL
2502	MARCHF6
2504	DAP
2516	SUB1
2519	TARS1
2530	PRKAA1
2532	RPL37
2544	NDUFS4
2559	SMIM15
2565	SREK1IP1
2577	CENPH
2578	MRPS36
2588	MRPS27
2595	BTF3
2599	HEXB
2602	HMGCR
2611	TBCA
2613	SCAMP1
2622	RPS23
2623	TMEM167A
2625	COX7C
2626	RASA1
2630	TMEM161B
2642	GLRX
2644	CAST

ccID	Gene Name
2661	SRP19
2662	REEP5
2663	DCP2
2669	TMED7-TICAM2,TMED7
2670	ATG12
2676	TNFAIP8
2677	HSD17B4
2681	PPIC
2685	ALDH7A1
2694	HINT1
2698	P4HA2
2702	SEPTIN8
2703	UQCRQ
2704	AFF4
2706	HSPA4
2708	VDAC1
2715	CDKN2AIPNL
2717	SEC24A
2718	CAMLG
2719	DDX46
2721	TXNDC15
2725	MACROH2A1
2726	TGFBI
2728	HNRNPA0
2738	CTNNA1
2740	MATR3
2743	UBE2D2
2744	CYSTM1
2746	HBEGF
2755	WDR55
2756	HARS1
2759	ZMAT2
2760	TAF7
2766	DELE1
2769	NDFIP1
2772	YIPF5
2780	CSNK1A1
2784	CD74
2786	RPS14
2788	DCTN4
2789	SMIM3
2790	TNIP1
2791	ANXA6
2800	CNOT8
2805	THG1L
2809	TTC1

ccID	Gene Name
2813	PTTG1
2814	CCNG1
2820	RARS1
2825	NPM1
2831	RPL26L1
2832	ATP6V0E1
2840	HIGD2A
2845	UIMC1
2849	RAB24
2852	LMAN2
2855	PDLIM7
2871	HNRNPH1
2872	CANX
2877	MRNIP
2878	TBC1D9B
2879	RNF130
2894	RACK1
2898	GMDS
2902	SERPINB6
2907	PSMG4
2913	SSR1
2917	DSP
2920	TXNDC5
2926	TMEM14C
2928	NEDD9
2931	NOL7
2936	GMPR
2944	DEK
2964	H1-2
2965	H4C3
2971	H2BC6
2979	H2AC8
2985	H2BC9
2990	HMGN4
2996	H2BC12
3005	H2AC15
3006	H2BC15
3007	H2AC16
3012	H3C11
3014	H3C12
3015	H2AC17
3027	RPP21
3028	HLA-E
3030	GNL1
3031	PRR3
3033	PPP1R10

ccID	Gene Name
3034	MRPS18B
3036	C6orf136
3038	NRM
3040	TUBB
3047	DDR1
3050	TCF19
3055	MICA
3058	DDX39B
3065	CSNK2B
3072	CLIC1
3075	LSM2
3087	PSMB8
3092	SLC39A7
3095	RING1
3105	CUTA
3106	CUTA
3108	ITPR3
3111	HMGA1
3116	RPS10,RPS10-NUDT3
3120	TAF11
3137	PIM1
3144	GLO1
3151	TOMM6
3164	RRP36
3171	TJAP1
3173	YIPF3
3186	TMEM63B
3187	SLC29A1
3189	HSP90AB1
3201	TNFRSF21
3202	CD2AP
3226	LMBRD1
3234	COX7A2
3239	HMGN3
3244	IBTK
3251	NT5E
3264	UBE2J1
3274	ASCC3
3276	PREP
3287	SNX3
3289	FOXO3
3301	GTF3C6
3303	TRAF3IP2
3306	HDAC2
3307	NT5DC1
3309	RWDD1

ccID	Gene Name
3319	TPD52L1
3326	CENPW
3330	ARHGAP18
3332	EPB41L2
3337	TBPL1
3341	MYB
3346	PEX7
3350	PERP
3352	HEBP2
3355	ABRACL
3356	HECA
3359	ADGRG6
3360	AIG1
3361	ADAT2
3365	SF3B5
3368	GINM1
3381	MTRF1L
3395	EZR
3400	TCP1
3406	SFT2D1
3409	RNASET2
3418	PSMB1
3420	PDCD2
3456	ACTB
3462	RAC1
3464	KDELR2,DAGLB
3474	NDUFA4
3475	TMEM106B
3478	BZW2
3482	AGR2
3483	AGR3
3486	SNX13
3490	CDCA7L
3493	TOMM7
3508	CBX3
3520	GGCT
3521	GARS1
3526	FKBP9
3531	SEPTIN7
3533	ANLN
3540	MRPL32
3544	MRPS24
3552	OGDH
3554	ZMIZ2
3556	H2AZ2
3563	UPP1

ccID	Gene Name
3567	SEC61G
3571	MRPS17
3574	CCT6A
3577	CHCHD2
3589	SBDS
3602	BUD23
3606	CLDN3
3609	EIF4H
3611	CLIP2
3618	MDH2
3620	HSPB1
3631	TMEM60
3640	SLC25A40
3651	FAM133B
3653	CDK6
3658	PON3
3661	PON2
3664	SEM1
3669	BRI3
3670	BAIAP2L1
3674	ARPC1A
3675	ARPC1B
3676	PDAP1
3679	BUD31
3682	ATP5MF,ATP5MF-PTCD1
3687	TRIM4
3688	ZKSCAN1
3692	MCM7
3695	LAMTOR4
3696	TRAPPC14
3716	ZNHIT1
3717	FIS1
3719	CUX1
3725	ARMC10
3726	PMPCB
3733	SRPK2
3737	SYPL1
3738	NAMPT
3745	LAMB1
3749	DNAJB9
3750	IMMP2L
3752	ZNF277
3755	BMT2
3764	FAM3C
3769	ARF5
3770	SND1

ccID	Gene Name
3771	RBM28
3777	ATP6V1F
3789	COPG2
3794	CHCHD3
3799	STMP1
3800	MTPN
3802	CREB3L2
3810	TBXAS1
3818	MRPS33
3819	SSBP1
3824	EPHB6
3825	GSTK1
3828	ZYX
3829	EPHA1
3830	TCAF1
3833	PDIA4
3848	SLC4A2
3850	TMUB1
3866	UBE3C
3868	DNAJB6
3888	FDFT1
3899	MTUS1
3900	PCM1
3901	ASAH1
3910	FHIP2B
3911	NUDT18
3932	SLC25A37
3937	PPP2R2A
3940	DPYSL2
3943	PTK2B
3948	CCDC25
3957	DUSP4
3959	SARAF
3960	LEPROTL1
3961	DCTN6
3965	GSR
3976	BRF2
3982	LSM1
3985	DDHD2
3992	TM2D2
3993	ADAM9
4001	POLB
4002	VDAC3
4008	FNTA
4011	SPIDR
4016	PCMTD1

ccID	Gene Name
4018	ATP6V1H
4019	TCEA1
4021	MRPL15
4022	TMEM68
4024	LYN
4025	RPS20
4031	SDCBP
4039	ASPH
4043	MTFR1
4044	RRS1
4054	TRAM1
4055	LACTB2
4062	ELOC
4063	TMEM70
4066	PEX2
4074	FABP5
4076	ZFAND1
4078	CHMP4C
4079	RBIS
4081	RMDN1
4082	CPNE3
4084	OSGIN2
4086	DECR1
4090	TRIQK
4097	ESRP1
4105	UQCRB
4107	PTDSS1
4108	MTDH
4111	RPL30
4118	COX6C
4127	YWHAZ
4128	YWHAZ
4129	YWHAZ
4135	AZIN1
4142	OXR1
4143	EIF3E
4144	EMC2
4145	NUDCD1
4146	ENY2
4149	EIF3H
4151	RAD21
4157	TAF2
4158	DSCC1
4166	C8orf76
4167	ATAD2
4168	NTAQ1
4170	FAM91A1
4173	TATDN1
4174	NDUFB9
4178	NSMCE2
4181	LRATD2
4186	CYRIB
4188	ASAP1
4194	ST3GAL1
4197	CHRAC1
4198	AGO2
4199	PTK2
4205	LY6E
4208	TOP1MT
4212	GSDMD
4224	NRBP2
4229	PLEC
4231	GRINA
4237	SHARPIN
4240	BOP1
4241	HSF1
4248	CPSF1
4250	VPS28
4267	RPL8
4276	CDC37L1
4279	PLGRKT
4285	DMAC1
4295	HAUS6
4305	CDKN2B
4311	NDUFB6
4318	CHMP5
4332	DCTN3
4335	VCP
4338	UNC13B
4343	TPM2
4352	CLTA
4357	ZCCHC7
4358	GRHPR
4359	POLR1E
4360	ENSG00000256966,TOMM5
4372	CEMIP2
4379	ANXA1
4394	HNRNPK
4404	CTSL
4424	MFSD14B
4432	SLC35D2
4440	ANP32B
4446	SEC61B
4450	TEX10
4451	MRPL50
4455	SLC44A1
4457	RAD23B
4464	TXN
4466	ECPAS
4467	PTGR1
4470	GNG10
4472	PTBP3
4491	FBXW2
4496	GSN
4500	RBM18
4509	PSMB7
4511	RPL35
4512	ARPC5L
4526	STXBP1
4533	ST6GALNAC4
4540	LCN2
4541	BBLN
4543	GOLGA2
4544	SWI5
4547	COQ4
4551	SPTAN1
4553	SET
4554	ZDHHC12
4566	IER5L
4568	NTMT1
4573	C9orf78
4576	ASS1
4577	FUBP3
4582	AIF1L
4592	GTF3C5
4596	RPL7A
4599	SURF4
4605	MRPS2
4620	TMEM141
4622	RABL6
4625	EDF1
4628	PAXX
4630	NPDC1
4631	ENTPD2
4647	TOR4A
4652	MRPL41
4662	PFKP
4672	GDI2

ccID	Gene Name
4674	RBM17
4678	ATP5F1C
4685	NUDT5
4687	OPTN
4690	PRPF18
4700	STAM
4704	DNAJC1
4705	COMMD3,COMMD3-BMI1
4707	KIAA1217
4710	PRTFDC1
4724	ARHGAP12
4727	ITGB1
4734	ZNF33A
4738	CSGALNACT2
4739	FXYD4
4742	HNRNPF
4743	HNRNPF
4750	NCOA4
4762	CCDC6
4772	HERC4
4773	HNRNPH3
4777	DDX50
4778	DDX21
4786	PPA1
4787	EIF4EBP2
4789	SGPL1
4791	PCBD1
4795	PSAP
4796	ANAPC16
4797	ASCC1
4806	P4HA1
4809	MRPS16
4811	PPP3CB
4816	CHCHD1
4819	CAMK2G
4825	VDAC2
4828	RPS24
4830	ZMIZ1
4835	PRXL2A
4837	GHITM
4843	ADIRF
4844	GLUD1
4847	ATAD1
4848	PTEN
4849	ANKRD22
4851	FAS

ccID	Gene Name
4860	MYOF
4868	PDLIM1
4873	TM9SF3
4879	MMS19
4886	GOT1
4887	COX15
4895	SCD
4899	HIF1AN
4903	MRPL43
4910	NPM3
4924	ARL3
4928	ATP5MK
4930	STN1
4935	GSTO1
4937	GSTO2
4939	XPNPEP1
4943	SMNDC1
4945	PDCD4
4949	ACSL5
4954	TCF7L2
4967	EIF3A
4969	SFXN4
4970	PRDX3
4972	RGS10
4989	EEF1AKMT2
5000	GLRX3
5006	TUBGCP2
5008	PRAP1
5010	FUOM
5012	ECHS1
5014	MTG1
5023	IFITM2
5028	SIGIRR
5038	PHRF1
5044	TALDO1
5056	MUC5AC
5061	CTSD
5062	MRPL23
5066	CD81
5070	SLC22A18
5075	NAP1L4
5078	NUP98
5079	PGAP2
5081	RHOG
5082	RRM1
5084	TRIM5

ccID	Gene Name
5092	TAF10
5094	MRPL17
5096	CYB5R2
5105	IPO7
5108	SBF2
5109	ADM
5113	EIF4G2
5119	PARVA
5125	PSMA1
5126	C11orf58
5130	RPS13
5138	TSG101
5154	RCN1
5155	EIF3M
5156	PRRG4
5165	CAPRIN1
5166	CAPRIN1
5170	EHF
5174	CD44
5184	TTC17
5185	HSD17B12
5187	EXT2
5191	LARGE2
5195	MDK
5198	AMBRA1
5210	PSMC3
5215	NDUFS3
5216	MTCH2
5217	FNBP4
5221	SSRP1
5229	SELENOH
5230	CTNND1
5239	MRPL16
5246	DDB1
5251	SDHAF2
5253	TMEM258
5256	FTH1
5260	AHNAK
5262	EEF1G
5263	EEF1G
5265	MTA2
5267	GANAB
5271	UBXN1
5272	BSCL2
5277	TTC9C
5282	STX5

ccID	Gene Name
5288	PLAAT3
5289	ATL3
5290	RTN3
5294	COX8A,ENSG00000256100
5296	MACROD1
5300	DNAJC4
5301	FKBP2
5310	ESRRA
5311	TRMT112
5312	PRDX5
5322	ARL2,ARL2-SNX15
5324	CDCA5
5326	VPS51
5328	FAU
5332	CAPN1
5336	DPF2
5361	SCYL1
5369	RELA
5374	CFL1
5377	CCDC85B
5378	FOSL1
5379	DRAP1
5380	DRAP1
5386	SF3B2
5390	RIN1
5392	SLC29A2
5393	MRPL11
5400	RBM14
5408	SSH3
5410	POLD4
5412	PPP1CA
5414	CORO1B
5416	AIP
5418	CDK2AP2
5424	NDUFV1
5428	NDUFS8
5434	LRP5
5436	CPT1A
5439	MYEOV
5440	CCND1
5443	PPFIA1
5444	CTTN
5462	MRPL48
5463	COA4
5485	CLNS1A
5486	RSF1

ccID	Gene Name
5491	GAB2
5492	PRCP
5502	PICALM
5518	CWC15
5519	ENDOD1
5526	TMEM123
5531	DCUN1D5
5547	C11orf1
5550	SDHD
5551	IL18
5566	TMPRSS4
5570	ATP5MG
5572	IFT46
5579	RPS25
5582	SLC37A4
5600	SIAE
5605	EI24
5606	STT3A
5609	RPUSD4
5611	SRPRA
5616	ST3GAL4
5638	FKBP4
5648	CD9
5652	SCNN1A
5653	LTBR
5656	MRPL51
5659	NCAPD2
5660	GAPDH
5664	CHD4
5668	MLF2
5669	PTMS
5675	ENO2
5676	C12orf57
5679	PHB2
5685	M6PR
5692	YBX3
5695	BCL2L14
5701	CDKN1B
5704	HEBP1
5705	EMP1
5708	PLBD1
5712	EPS8
5715	MGST1
5716	PLEKHA5
5723	LDHB
5726	KRAS

ccID	Gene Name
5734	PPFIBP1
5735	MRPS35
5738	CAPRIN2
5742	SINHCAF
5746	DNM1L
5751	TWF1
5756	SLC38A2
5759	HDAC7
5762	PFKM
5766	ARF3
5767	PRKAG1
5771	TUBA1C
5781	CERS5
5788	LETMD1
5792	SMAGP
5795	NR4A1
5796	ATG101
5801	KRT8
5802	KRT18
5804	KRT8
5813	PFDN5
5815	MYG1
5819	PCBP2
5824	ATP5MC2
5828	CBX5
5830	COPZ1
5837	CD63
5838	SARNP
5839	ORMDL2
5843	PYM1
5850	PA2G4
5851	RPL41
5854	MYL6
5860	CS
5861	CNPY2
5864	TIMELESS
5872	PTGES3
5874	NACA
5875	PRIM1
5881	SHMT2
5886	DCTN2
5888	OS9
5901	GNS
5903	HMGA2
5904	LLPH
5906	CAND1

ccID	Gene Name
5915	LYZ
5920	YEATS4
5922	CCT2
5923	RAB3IP
5926	TSPAN8
5929	RAB21
5931	ATXN7L3B
5937	NAP1L1
5938	OSBPL8
5942	PAWR
5950	DUSP6
5955	BTG1
5957	UBE2N
5960	NDUFA12
5964	METAP2
5966	SNRPF
5967	LTA4H
5970	SLC25A3
5971	IKBIP
5976	UTP20
5977	ARL1
5986	C12orf73
5989	TXNRD1
5993	C12orf75
5994	CKAP4
6004	ALKBH2
6010	MMAB
6015	C12orf76
6016	ATP2A2
6018	ARPC3
6023	PPP1CC
6030	ERP29
6035	RPL6
6038	DDX54
6061	PXN
6062	COX6A1,ENSG0000 0111780
6065	SRSF9
6070	POP5
6073	UNC119B
6081	RHOF
6082	PSMD9
6085	DIABLO
6093	VPS37B
6095	ARL6IP4
6111	SCARB1
6116	AACS

ccID	Gene Name
6121	RAN
6122	SFSWAP
6137	PSPC1
6138	ZMYM2
6140	CRYL1
6158	NUP58
6163	RPL21
6168	POLR1D
6174	HSPH1
6193	VWA8
6194	DNAJC15
6198	TSC22D1
6202	GTF2F2
6207	TPT1
6209	SLC25A30
6213	ESD
6217	ITM2B
6220	RCBTB1
6221	EBPL
6223	TRIM13
6238	MZT1
6243	COMMD6
6251	NDFIP2
6257	DNAJC3
6263	IPO5
6264	IPO5
6265	FARP1
6266	STK24
6267	UBAC2
6268	TM9SF2
6277	ARGLU1
6284	ANKRD10
6287	TUBGCP3
6292	LAMP1
6295	TMCO3
6297	GAS6
6308	APEX1
6312	RNASE1
6313	METTL17
6316	SUPT16H
6320	DAD1
6322	OXA1L
6323	MRPL52
6325	LRP10
6328	HAUS4
6330	PSMB5

ccID	Gene Name
6335	BCL2L2
6339	DHRS4
6343	DCAF11
6344	PSME1
6349	IPO4
6353	NEDD8-MDP1,NEDD8
6360	STXBP6
6364	HECTD1
6369	EGLN3
6373	SNX6
6376	FAM177A1
6379	PSMA6
6380	NFKBIA
6389	PNN
6393	PRPF39
6396	RPS29
6399	RPL36AL
6404	ARF6
6409	NIN
6410	PYGL
6412	RTRAF
6414	ERO1A
6422	CDKN3
6423	CNIH1
6429	MAPK1IP1L
6436	KTN1
6442	ARMH4
6443	ACTR10
6444	PSMA3
6447	TIMM9
6450	DHRS7
6461	MTHFD1
6465	CHURC1
6467	GPX2
6470	GPHN
6477	VTI1B
6478	RDH11
6480	ACTN1
6483	SRSF5
6503	NPC2
6507	YLPM1
6508	DLST
6510	EIF2B2
6514	TMED10
6515	FOS
6522	GSTZ1

ccID	Gene Name
6525	AHSA1
6526	SPTLC2
6543	PPP4R3A
6547	NDUFB1
6551	ITPK1
6561	IFI27L2
6563	SERPINA1
6571	SETD3
6574	YY1
6589	EIF5
6600	ATP5MJ
6607	CDCA4
6610	MTA1
6612	ENSG00000257341
6615	CYFIP1
6626	SCG5
6630	NOP10
6638	SRP14
6652	RMDN3
6663	CHP1
6676	VPS39
6687	PDIA3
6690	SERF2
6692	MFAP1
6699	EIF3J
6701	B2M
6703	SORD
6710	EID1
6711	COPS2
6721	LYSMD2
6724	LEO1
6725	MAPK6
6728	ARPP19
6729	RSL24D1
6737	MYZAP,GCOM1
6740	SLTM
6745	GTF2A2
6753	VPS13C
6756	TPM1
6757	TPM1
6760	RAB8B
6762	USP3
6763	HERC1
6764	CIAO2A
6768	PPIB
6771	ENSG00000259316, PCLAF

ccID	Gene Name
6772	TRIP4
6774	SPG21
6781	RAB11A
6786	RPL4
6790	C15orf61
6793	CLN6
6798	RPLP1
6803	HEXA
6804	ARIH1
6810	PML
6820	COX5A
6832	FBXO22
6834	RCN2
6839	DNAJA4
6842	IREB2
6843	PSMA4
6845	MORF4L1
6846	CTSH
6848	TMED3
6851	ABHD17C
6856	RPS17
6864	SEC11A
6870	MRPS11
6883	SEMA4B
6912	SNRNP25
6914	MPG
6920	MRPL28
6922	NME4
6927	METTL26
6946	UBE2I
6965	NDUFB10
6970	GFER
6977	TRAF7
6983	RNPS1
6987	ATP6V0C
6992	SRRM2
7002	HCFC1R1
7006	IL32
7012	TRAP1
7014	PAM16
7015	DNAJA3
7022	SMIM22
7036	LITAF
7039	RSL1D1
7042	BFAR
7043	PDXDC1

ccID	Gene Name
7047	CEP20
7048	RPS15A
7053	GDE1
7062	LDAF1
7063	METTL9
7064	UQCRC2
7071	NDUFAB1
7078	LCMT1
7086	TUFM
7096	CDIPT
7097	SEZ6L2
7098	ASPHD1
7105	ALDOA
7106	ALDOA
7107	ALDOA
7112	PPP4C
7117	CD2BP2
7133	VKORC1
7135	KAT8
7136	PRSS8
7138	PYCARD
7140	ZNF720
7141	ZNF267
7149	DNAJA2
7156	BRD7
7162	NUDT21
7164	MT2A
7166	MT1E
7191	GOT2
7197	PDP2
7198	CIAO2B
7206	NOL3
7210	ELMO3
7222	NUTF2
7231	CDH3
7236	VPS4A
7242	NOB1
7249	SF3B3
7273	CFDP1
7285	HSBP1
7299	COX4I1
7314	CDT1
7315	APRT
7328	TCF25
7329	TUBB3
7344	YWHAE

ccID	Gene Name
7345	CRK
7346	MYO1C
7351	PRPF8
7354	RPA1
7366	EMC6
7389	SPAG7
7390	KIF1C
7393	RPAIN
7394	C1QBP
7400	TXNDC17
7403	RNASEK
7404	C17orf49
7405	ACADVL
7409	GABARAP
7410	CTDNEP1
7412	CLDN7
7415	EIF5A
7418	KCTD11
7420	PLSCR3
7421	TMEM256-PLSCR3,TMEM256
7431	CD68
7433	SAT2
7441	VAMP2
7456	ELAC2
7463	ZSWIM7
7481	SREBF1
7483	ATPAF2
7487	FLII
7490	SHMT1
7499	SPECC1
7504	WSB1
7507	TMEM97
7512	UNC119
7526	MYO18A
7527	NUFIP2
7539	NF1
7542	UTP6
7546	ZNF207
7548	MYO1D
7553	SLFN5
7558	MMP28
7561	ZNHIT3
7562	MYO19
7563	GGNBP2
7565	AATF
7577	RPL23

ccID	Gene Name
7589	STARD3
7590	ERBB2
7595	ORMDL3
7598	CASC3
7599	WIPF2
7600	CDC6
7606	KRT15
7611	KRT19
7613	EIF1
7618	DNAJC7
7631	MLX
7638	VPS25
7640	BECN1
7644	RPL27
7648	VAT1
7650	NBR1
7655	TMEM101
7656	LSM12
7661	SLC25A39
7678	KPNB1
7679	MRPL10
7682	PNPO
7683	PRR15L
7684	CDK5RAP3
7695	HOXB6
7703	CALCOCO2
7706	SNF8
7721	ABCC3
7727	NME2
7730	TOM1L1
7731	COX11
7740	MRPS23
7749	SKA2
7751	DHX40
7752	CLTC
7753	PTRH2
7754	VMP1
7766	DCAF7
7770	DDX42
7772	PSMC5
7787	GNA13
7792	PITPNC1
7794	KPNA2
7800	PRKAR1A
7812	RPL38
7820	ATP5PD

ccID	Gene Name
7821	NT5C
7824	SUMO2
7835	SAP30BP
7836	ITGB4
7842	WBP2
7852	UBALD2
7861	SRSF2
7862	SEC14L1
7863	SEPTIN9
7865	TMC6
7866	SYNGR2
7869	BIRC5
7876	LGALS3BP
7883	CBX4
7885	GAA
7893	NDUFAF8
7907	MRPL12
7910	MCRIP1
7911	P4HB
7916	SIRT7
7929	CSNK1D
7933	CYBC1
7935	WDR45B
7938	FN3K
7939	TBCD
7943	TYMS
7945	YES1
7953	MYL12B
7958	NDUFV2
7963	VAPA
7970	AFG3L2
7972	PSMG2
7973	PTPN2
7979	SNRPD1
7983	GATA6
7986	RIOK3
7988	NPC1
7990	TTC39C
8009	C18orf21
8012	ELP2
8014	TPGS2
8021	ATP5F1A
8022	HAUS1
8027	IER3IP1
8029	DYM
8032	RPL17

ccID	Gene Name
8036	MYO5B
8041	ME2
8046	TXNL1
8049	ATP8B1
8052	LMAN1
8063	CYB5A
8069	HSBP1L1
8070	TXNL4A
8071	ADNP2
8075	CDC34
8085	MISP
8089	R3HDM4
8092	CNN2
8096	GPX4
8102	ATP5F1D
8104	MIDN
8105	CIRBP
8106	FAM174C
8109	RPS15
8123	MKNK2
8127	PLEKHJ1
8128	SF3A2
8129	JSRP1
8132	LSM7
8134	TIMM13
8140	SGTA
8141	THOP1
8147	TLE5
8169	MAP2K2
8177	MYDGF
8184	SAFB2
8186	MICOS13
8187	RPL36
8190	FUT3
8191	ENSG00000267740, NDUFA11
8194	ALKBH7
8195	GTF2F1
8209	XAB2
8212	TRAPPC5
8215	SNAPC2
8218	CD320
8219	NDUFA7
8222	RAB11B
8224	HNRNPM
8230	ZNF561
8236	EIF3G

ccID	Gene Name
8244	CDC37
8246	ATG4D
8249	AP1M2
8251	ILF3
8262	RAB3D
8267	RGL3
8268	PRKCSH
8269	ECSIT
8273	ZNF440
8281	MAN2B1
8282	WDR83OS
8283	DHPS
8285	TRIR
8303	GADD45GIP1
8306	STX10
8312	C19orf53
8324	GIPC1
8325	DNAJB1
8326	TECR
8334	TPM4
8336	TPM4
8357	ANO8
8361	SLC27A1
8362	PGLS
8363	COLGALT1
8365	B3GNT3
8367	CCDC124
8370	RAB3A
8380	ISYNA1
8382	FKBP8
8383	KXD1
8384	UBA52
8393	MEF2B
8398	GATAD2A
8402	LPAR2
8409	ZNF431
8412	ZNF91
8415	UQCRFS1
8416	POP4
8420	PDCD5
8421	PDCD5
8435	UBA2
8436	GRAMD1A
8437	FXYD3
8439	FXYD5
8442	LSR

ccID	Gene Name
8443	DMKN
8444	TMEM147
8448	COX6B1
8463	SPINT2
8470	C19orf33
8474	PSMD8
8476	EIF3K
8491	RPS16
8492	SUPT5H
8495	FBL
8497	ZNF780A
8504	BLVRB
8508	C19orf54
8510	SNRPA
8516	HNRNPUL1
8524	RPS19
8535	PAFAH1B3
8541	ETHE1
8554	CLPTM1
8555	CLASRP
8562	PPP1R13L
8567	ERCC1
8571	EML2
8572	SNRPD2
8580	IGFL2
8583	CALM3
8588	AP2S1
8609	KDELR1
8612	CYTH2
8617	RPL18
8621	NUCB1
8622	BAX
8629	SNRNP70
8637	RPS11
8640	PRR12
8645	IRF3
8646	BCL2L12
8648	AP2A1
8660	MYH14
8664	EMC10
8670	KLK6
8675	KLK11
8694	MYADM
8698	CNOT3
8705	RPS9
8708	CDC42EP5

ccID	Gene Name
8712	TNNT1
8719	TMEM238
8730	EPN1
8744	TRIM28
8748	MZF1
8750	C20orf96
8752	RBCK1
8754	CSNK2A1
8756	PSMF1
8763	IDH3B
8768	DDRGK1
8771	C20orf27
8772	CENPB
8773	CDC25B
8774	MAVS
8794	SNRPB2
8795	DSTN
8802	KAT14,PET117
8805	SMIM26,ENSG0000 0284776
8807	NAA20
8808	CRNKL1
8810	XRN2
8816	CST3
8817	APMAP
8820	PYGB
8821	ABHD12
8824	HM13
8829	BCL2L1
8831	TPX2
8835	POFUT1
8836	KIF3B
8838	NOL4L
8842	NECAB3
8846	RALY
8849	EIF2S2
8850	AHCY
8855	ACSS2
8857	TRPC4AP
8859	PROCR
8861	MMP24OS
8865	ERGIC3
8869	CPNE1,RBM12
8872	RBM39
8873	PHF20
8882	RPN2
8885	SRC

ccID	Gene Name
8886	BLCAP
8910	SDC4
8911	DBNDD2
8913	WFDC2
8917	UBE2C
8921	CTSA
8924	SLC35C2
8932	STAU1
8950	PFDN4
8951	FAM210B
8962	STX16
8963	GNAS
8965	GNAS
8970	ATP5F1E
8971	PRELID3B
8972	LSM14B
8979	RPS21
8981	SLCO4A1
8989	MRGBP
8990	OGFR
8994	GID8
8998	PPDPF
9000	PTK6
9023	BTG3
9026	ATP5PF
9034	APP
9038	CCT8
9043	URB1
9056	ATP5PO
9059	CBR1
9066	PIGP
9070	ETS2
9072	BRWD1
9074	HMGN1
9078	BACE2
9079	FAM3B
9083	ABCG1
9090	TFF1
9098	RRP1B
9099	PDXK
9103	CSTB
9104	CSTB
9114	SLX9
9119	COL18A1
9144	GNB1L,RTL10
9161	UBE2L3

ccID	Gene Name
9163	SDF2L1
9165	MAPK1
9171	CHCHD10
9173	SMARCB1
9175	MIF
9180	SNRPD3
9192	KREMEN1
9207	SF3A1
9210	PES1
9217	SELENOM
9224	YWHAH
9227	HMGXB4
9232	RBFOX2
9238	EIF3D
9241	MPST
9250	LGALS1
9258	EIF3L
9262	MAFF
9263	TMEM184B
9267	DDX17
9269	TOMM22
9271	JOSD1
9273	SUN2
9284	ATF4
9289	ADSL
9293	ST13
9294	XPNPEP3
9298	RANGAP1
9300	TOB2
9301	PHF5A
9305	XRCC6
9307	SNU13
9309	SREBF2
9345	TBC1D22A
9350	TRABD
9353	HDAC10
9359	LMF2
9373	CD99
9382	ARHGAP6
9384	PRPS2
9393	SYAP1
9395	RBBP7
9399	EIF1AX
9403	SMS
9416	DYNLT3
9422	CXorf38

ccID	Gene Name
9431	NDUFB11
9432	RBM10
9434	UBA1
9438	TIMP1
9444	UXT
9448	RBM3
9449	WDR13
9453	PQBP1
9460	PLP2
9477	MAGED2
9490	PDZD11
9495	SNX12
9497	NONO
9499	OGT
9502	RPS4X
9507	ATRX
9509	COX7B
9510	PGK1
9514	TAF9B
9519	TSPAN6
9526	RPL36A,RPL36A-HNRNPH2
9530	ARMCX3
9533	TCEAL8
9534	TCEAL9
9536	TCEAL4
9537	TCEAL3
9558	PGRMC1
9559	SLC25A5
9561	UBE2A
9562	SEPTIN6
9565	RNF113A
9566	NDUFA1
9573	STAG2
9574	OCRL
9576	ZDHHC9
9581	RBMX2
9585	PHF6
9586	HPRT1
9592	RBMX
9599	HMGB3
9600	VMA21
9608	BCAP31
9611	SSR4
9612	NAA10
9616	FLNA
9617	FLNA

ccID	Gene Name
9619	EMD
9620	RPL10
9625	GDI1
8885	SRC
8886	BLCAP
8910	SDC4
8911	DBNDD2
8913	WFDC2
8917	UBE2C
8921	CTSA
8924	SLC35C2
8932	STAU1
8950	PFDN4
8951	FAM210B
8962	STX16
8963	GNAS
8965	GNAS
8970	ATP5F1E
8971	PRELID3B
8972	LSM14B
8979	RPS21
8981	SLCO4A1
8989	MRGBP
8990	OGFR
8994	GID8
8998	PPDPF
9000	PTK6
9023	BTG3
9026	ATP5PF
9034	APP
9038	CCT8
9043	URB1
9056	ATP5PO
9059	CBR1
9066	PIGP
9070	ETS2
9072	BRWD1
9074	HMGN1
9078	BACE2
9079	FAM3B
9083	ABCG1
9090	TFF1
9098	RRP1B
9099	PDXK
9103	CSTB
9114	SLX9

ccID	Gene Name
9119	COL18A1
9144	GNB1L,RTL10
9161	UBE2L3
9163	SDF2L1
9165	MAPK1
9171	CHCHD10
9173	SMARCB1
9175	MIF
9180	SNRPD3
9192	KREMEN1
9207	SF3A1
9210	PES1
9217	SELENOM
9224	YWHAH
9227	HMGXB4
9232	RBFOX2
9238	EIF3D
9241	MPST
9250	LGALS1
9258	EIF3L
9262	MAFF
9263	TMEM184B
9267	DDX17
9269	TOMM22
9271	JOSD1
9273	SUN2
9284	ATF4
9289	ADSL
9293	ST13
9294	XPNPEP3
9298	RANGAP1
9300	TOB2
9301	PHF5A
9305	XRCC6
9307	SNU13
9309	SREBF2
9345	TBC1D22A
9350	TRABD
9353	HDAC10
9359	LMF2
9373	CD99
9382	ARHGAP6
9384	PRPS2
9393	SYAP1
9395	RBBP7
9399	EIF1AX

ccID	Gene Name
9403	SMS
9416	DYNLT3
9422	CXorf38
9431	NDUFB11
9432	RBM10
9434	UBA1
9438	TIMP1
9444	UXT
9448	RBM3
9449	WDR13
9453	PQBP1
9460	PLP2
9477	MAGED2
9490	PDZD11
9495	SNX12
9497	NONO
9499	OGT
9502	RPS4X
9507	ATRX
9509	COX7B
9510	PGK1
9514	TAF9B
9519	TSPAN6
9526	RPL36A
9530	ARMCX3
9533	TCEAL8
9534	TCEAL9
9536	TCEAL4
9537	TCEAL3
9558	PGRMC1
9559	SLC25A5
9561	UBE2A
9562	SEPTIN6
9565	RNF113A
9566	NDUFA1
9573	STAG2
9574	OCRL
9576	ZDHHC9
9581	RBMX2
9585	PHF6
9586	HPRT1
9592	RBMX
9599	HMGB3
9600	VMA21
9608	BCAP31
9611	SSR4

ccID	Gene Name
9612	NAA10
9616	FLNA
9617	FLNA
9619	EMD
9620	RPL10
9625	GDI1

Appendix D

Table 1: Full list of Ribosomal associated genes removed in mentioned analysis. HGNC gene name of the 80 ribosomal genes removed to assess for ribosomal bias in analyses. All identified as having TOP TSS usage as their dominant promoter.

Gene Name	Gene Name
RPSA	RPL9
RPS2	RPL10
RPS3	RPL10A
RPS3A	RPL11
RPS4X	RPL12
RPS4Y	RPL13
RPS5	RPL13A
RPS6	RPL14
RPS7	RPL15
RPS8	RPL17
RPS9	RPL18
RPS10	RPL18A
RPS11	RPL19
RPS12	RPL21
RPS13	RPL22
RPS14	RPL23
RPS15	RPL23A
RPS15A	RPL24
RPS16	RPL26
RPS17	RPL27
RPS18	RPL27A
RPS19	RPL28
RPS20	RPL29
RPS21	RPL30
RPS23	RPL31
RPS24	RPL32
RPS25	RPL34
RPS26	RPL35
RPS27	RPL35A
RPS27A	RPL36
RPS28	RPL36A
RPS29	RPL37
RPS30	RPL37A
RPL3	RPL38
RPL4	RPL39
RPL5	RPL40
RPL6	RPL41
RPL7	RPLP0
RPL7A	RPLP1
RPL8	RPLP2

Bibliography

Ahmadi, S. E., Rahimi, S., Zarandi, B., Chegeni, R., & Safa, M. (2021). MYC: a multipurpose oncogene with prognostic and therapeutic implications in blood malignancies. In *Journal of Hematology and Oncology* (Vol. 14, Issue 1). https://doi.org/10.1186/s13045-021-01111-4

Albrecht, F. B., Dolderer, V., Nellinger, S., Schmidt, F. F., & Kluger, P. J. (2022). Gellan Gum Is a Suitable Biomaterial for Manual and Bioprinted Setup of Long-Term Stable, Functional 3D-Adipose Tissue Models. *Gels*, *8*(7). https://doi.org/10.3390/gels8070420

Alfonso-Gonzalez, C., & Hilgers, V. (2024). (Alternative) transcription start sites as regulators of RNA processing. In *Trends in Cell Biology*. https://doi.org/10.1016/j.tcb.2024.02.010

Al-Hamaly, M. A., Turner, L. T., Rivera-Martinez, A., Rodriguez, A., & Blackburn, J. S. (2023). Zebrafish Cancer Avatars: A Translational Platform for Analyzing Tumor Heterogeneity and Predicting Patient Outcomes. In *International Journal of Molecular Sciences* (Vol. 24, Issue 3). https://doi.org/10.3390/ijms24032288

Alladin, A., Chaible, L., del Valle, L. G., Sabine, R., Loeschinger, M., Wachsmuth, M., Hériché, J. K., Tischer, C., & Jechlinger, M. (2020). Tracking cells in epithelial acini by light sheet microscopy reveals proximity effects in breast cancer initiation. *ELife*. https://doi.org/10.7554/eLife.54066

An, T., Liu, Y., Gourguechon, S., Wang, C. C., & Li, Z. (2018). CDK Phosphorylation of Translation Initiation Factors Couples Protein Translation with Cell-Cycle Transition. *Cell Reports*, *25*(11). https://doi.org/10.1016/j.celrep.2018.11.063

Anderson, K. G., Braun, D. A., Buqué, A., Gitto, S. B., Guerriero, J. L., Horton, B., Keenan, B. P., Kim, T. S., Overacre-Delgoffe, A., Ruella, M., Triplett, T. A., Veeranki, O., Verma, V., & Zhang, F. (2023). Leveraging immune resistance archetypes in solid cancer to inform next-generation anticancer therapies. *Journal for ImmunoTherapy of Cancer*, *11*(6). https://doi.org/10.1136/jitc-2022-006533

Andersson, R., & Sandelin, A. (2020). Determinants of enhancer and promoter activities of regulatory elements. In *Nature Reviews Genetics*. https://doi.org/10.1038/s41576-019-0173-8

Ashley, N., Yeung, T. M., & Bodmer, W. F. (2013). Stem cell differentiation and lumen formation in colorectal cancer cell lines and primary tumors. *Cancer Research*, *73*(18). https://doi.org/10.1158/0008-5472.CAN-13-0454

Assi, S. A., Bonifer, C., & Cockerill, P. N. (2019). Rewiring of the Transcription Factor Network in Acute Myeloid Leukemia. In *Cancer Informatics*. https://doi.org/10.1177/1176935119859863

Atanasova, V. S., de Jesus Cardona, C., Hejret, V., Tiefenbacher, A., Mair, T., Tran, L., Pfneissl, J., Draganić, K., Binder, C., Kabiljo, J., Clement, J., Woeran, K., Neudert, B., Wohlhaupter, S., Haase, A., Domazet, S., Hengstschläger, M., Mitterhauser, M., Müllauer, L., … Egger, G. (2023). Mimicking Tumor Cell Heterogeneity of Colorectal Cancer in a Patient-derived Organoid-Fibroblast Model. *CMGH*, *15*(6). https://doi.org/10.1016/j.jcmgh.2023.02.014

Bajar, B. T., Lam, A. J., Badiee, R. K., Oh, Y. H., Chu, J., Zhou, X. X., Kim, N., Kim, B. B., Chung, M., Yablonovitch, A. L., Cruz, B. F., Kulalert, K., Tao, J. J., Meyer, T., Su, X. D., & Lin, M. Z. (2016). Fluorescent indicators for simultaneous reporting of all four cell cycle phases. In *Nature Methods*. https://doi.org/10.1038/nmeth.4045

Balwierz, P. J., Carninci, P., Daub, C. O., Kawai, J., Hayashizaki, Y., Van Belle, W., Beisel, C., & van Nimwegen, E. (2009). Methods for analyzing deep sequencing expression data: Constructing the human and mouse promoterome with deepCAGE data. *Genome Biology*, *10*(7). https://doi.org/10.1186/gb-2009-10-7-r79

Bar-Joseph, Z., Siegfried, Z., Brandeis, M., Brors, B., Lu, Y., Eils, R., Dynlacht, B. D., & Simon, I. (2008). Genome-wide transcriptional analysis of the human cell cycle identifies genes differentially regulated in normal and cancer cells. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(3). https://doi.org/10.1073/pnas.0704723105

Basu, S., Dong, Y., Kumar, R., Jeter, C., & Tang, D. G. (2022). Slow-cycling (dormant) cancer cells in therapy resistance, cancer relapse and metastasis. In *Seminars in Cancer Biology* (Vol. 78). https://doi.org/10.1016/j.semcancer.2021.04.021

Batchuluun, B., Pinkosky, S. L., & Steinberg, G. R. (2022). Lipogenesis inhibitors: therapeutic opportunities and challenges. In *Nature Reviews Drug Discovery* (Vol. 21, Issue 4). https://doi.org/10.1038/s41573-021-00367-2

Belinsky, S. A. (2005). Silencing of genes by promoter hypermethylation: Key event in rodent and human lung cancer. In *Carcinogenesis* (Vol. 26, Issue 9). https://doi.org/10.1093/carcin/bgi020

Beloribi-Djefaflia, S., Vasseur, S., & Guillaumond, F. (2016). Lipid metabolic reprogramming in cancer cells. In *Oncogenesis* (Vol. 5, Issue 1). https://doi.org/10.1038/ONCSIS.2015.49

Betge, J., Rindtorff, N., Sauer, J., Rauscher, B., Dingert, C., Gaitantzi, H., Herweck, F., Srour-Mhanna, K., Miersch, T., Valentini, E., Boonekamp, K. E., Hauber, V., Gutting, T., Frank, L., Belle, S., Gaiser, T., Buchholz, I., Jesenofsky, R., Härtel, N., … Boutros, M. (2022). The drug-induced phenotypic landscape of colorectal cancer organoids. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-30722-9

Bhat, G. R., Sethi, I., Sadida, H. Q., Rah, B., Mir, R., Algehainy, N., Albalawi, I. A., Masoodi, T., Subbaraj, G. K., Jamal, F., Singh, M., Kumar, R., Macha, M. A., Uddin, S., Akil, A. S. A. S., Haris, M., & Bhat, A. A. (2024). Cancer cell plasticity: from cellular, molecular, and genetic mechanisms to tumor heterogeneity and drug resistance. In *Cancer and Metastasis Reviews*. https://doi.org/10.1007/s10555-024-10172-z

Bhat, P., Cabrera-Quio, L. E., Herzog, V. A., Fasching, N., Pauli, A., & Ameres, S. L. (2023). SLAMseq resolves the kinetics of maternal and zygotic gene expression during early zebrafish embryogenesis. *Cell Reports*, *42*(2). https://doi.org/10.1016/j.celrep.2023.112070

Bhatlekar, S., Fields, J. Z., & Boman, B. M. (2018). Role of HOX genes in stem cell differentiation and cancer. In *Stem Cells International* (Vol. 2018). https://doi.org/10.1155/2018/3569493

Blagosklonny, M. V. (2013). Hypoxia, MTOR and autophagy. *Autophagy*. https://doi.org/10.4161/auto.22783

Böhm, R., Imseng, S., Jakob, R. P., Hall, M. N., Maier, T., & Hiller, S. (2021). The dynamic mechanism of 4E-BP1 recognition and phosphorylation by mTORC1. *Molecular Cell*, *81*(11). https://doi.org/10.1016/j.molcel.2021.03.031

Borden, K. L. B., & Volpon, L. (2020). The diversity, plasticity, and adaptability of cap-dependent translation initiation and the associated machinery. In *RNA Biology* (Vol. 17, Issue 9). https://doi.org/10.1080/15476286.2020.1766179

Boström, J., Sramkova, Z., Salašová, A., Johard, H., Mahdessian, D., Fedr, R., Marks, C., Medalová, J., Souček, K., Lundberg, E., Linnarsson, S., Bryja, V., Sekyrova, P., Altun, M., & Andäng, M. (2017). Comparative cell cycle transcriptomics reveals synchronization of developmental transcription factor networks in cancer cells. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0188772

Bouyahya, A., El Allam, A., Aboulaghras, S., Bakrim, S., El Menyiy, N., Alshahrani, M. M., Al Awadh, A. A., Benali, T., Lee, L. H., El Omari, N., Goh, K. W., Ming, L. C., & Mubarak, M. S. (2022). Targeting mTOR as a Cancer Therapy: Recent Advances in Natural Bioactive Compounds and Immunotherapy. In *Cancers* (Vol. 14, Issue 22). https://doi.org/10.3390/cancers14225520

Bowes, A. L., Tarabichi, M., Pillay, N., & Van Loo, P. (2022). Leveraging single-cell sequencing to unravel intratumour heterogeneity and tumour evolution in human cancers. In *Journal of Pathology* (Vol. 257, Issue 4). https://doi.org/10.1002/path.5914

Bowry, A., Kelly, R. D. W., & Petermann, E. (2021). Hypertranscription and replication stress in cancer. In *Trends in Cancer* (Vol. 7, Issue 9). https://doi.org/10.1016/j.trecan.2021.04.006

Brady, L., Kriner, M., Coleman, I., Morrissey, C., Roudier, M., True, L. D., Gulati, R., Plymate, S. R., Zhou, Z., Birditt, B., Meredith, R., Geiss, G., Hoang, M., Beechem, J., & Nelson, P. S. (2021). Inter- and intra-tumor heterogeneity of metastatic prostate cancer determined by digital spatial gene expression profiling. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-21615-4

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*. https://doi.org/10.3322/caac.21492

Brown, J. S., Amend, S. R., Austin, R. H., Gatenby, R. A., Hammarlund, E. U., & Pienta, K. J. (2023). Updating the Definition of Cancer. *Molecular Cancer Research*, *21*(11). https://doi.org/10.1158/1541-7786.MCR-23-0411

Burrell, R. A., McGranahan, N., Bartek, J., & Swanton, C. (2013). The causes and consequences of genetic heterogeneity in cancer evolution. In *Nature*. https://doi.org/10.1038/nature12625

Bushweller, J. H. (2019). Targeting transcription factors in cancer — from undruggable to reality. In *Nature Reviews Cancer* (Vol. 19, Issue 11). https://doi.org/10.1038/s41568-019-0196-7

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., & Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology*, *36*(5). https://doi.org/10.1038/nbt.4096

Caiado, F., Silva-Santos, B., & Norell, H. (2016). Intra-tumour heterogeneity – going beyond genetics. *FEBS Journal*. https://doi.org/10.1111/febs.13705

Carninci, P., Kvam, C., Kitamura, A., Ohsumi, T., Okazaki, Y., Itoh, M., Kamiya, M., Shibata, K., Sasaki, N., Izawa, M., Muramatsu, M., Hayashizaki, Y., & Schneider, C. (1996). High-efficiency full-length cDNA cloning by biotinylated CAP trapper. *Genomics*, *37*(3). https://doi.org/10.1006/geno.1996.0567

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C. A. M., Taylor, M. S., Engström, P. G., Frith, M. C., Forrest, A. R. R., Alkema, W. B., Tan, S. L., Plessy, C., Kodzius, R., Ravasi, T., Kasukawa, T., Fukuda, S., Kanamori-Katayama, M., … Hayashizaki, Y. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. *Nature Genetics*. https://doi.org/10.1038/ng1789

Carulli, A. J., Samuelson, L. C., & Schnell, S. (2014). Unraveling intestinal stem cell behavior with models of crypt dynamics. In *Integrative Biology (United Kingdom)* (Vol. 6, Issue 3). https://doi.org/10.1039/c3ib40163d

Casamassimi, A., & Ciccodicola, A. (2019). Transcriptional regulation: Molecules, involved mechanisms, and misregulation. In *International Journal of Molecular Sciences*. https://doi.org/10.3390/ijms20061281

Castelli, V., Giordano, A., Benedetti, E., Giansanti, F., Quintiliani, M., Cimini, A., & D'Angelo, M. (2021). The great escape: The power of cancer stem cells to evade programmed cell death. *Cancers*, *13*(2). https://doi.org/10.3390/cancers13020328

Cerezo, M., & Rocchi, S. (2020). Cancer cell metabolic reprogramming: a keystone for the response to immunotherapy. In *Cell Death and Disease*. https://doi.org/10.1038/s41419-020-03175-5

Chao, S., Zhang, F., Yan, H., Wang, L., Zhang, L., Wang, Z., Xue, R., Wang, L., Wu, Z., Jiang, B., Shi, G., Xue, Y., Du, J., & Bu, P. (2023). Targeting intratumor heterogeneity suppresses colorectal cancer chemoresistance and metastasis. *EMBO Reports*, *24*(8). https://doi.org/10.15252/embr.202256416

Chen, H., & Boutros, P. C. (2011). VennDiagram: A package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics*, *12*. https://doi.org/10.1186/1471-2105-12-35

Chen, R. J., Lu, M. Y., Williamson, D. F. K., Chen, T. Y., Lipkova, J., Noor, Z., Shaban, M., Shady, M., Williams, M., Joo, B., & Mahmood, F. (2022). Pan-cancer integrative histology-genomic analysis via multimodal deep learning. *Cancer Cell*, *40*(8). https://doi.org/10.1016/j.ccell.2022.07.004

Chen, X., & Xu, Y. (2022). Structural insights into assembly of transcription preinitiation complex. In *Current Opinion in Structural Biology* (Vol. 75). https://doi.org/10.1016/j.sbi.2022.102404

Cho, D. Y., & Przytycka, T. M. (2013). Dissecting cancer heterogeneity with a probabilistic genotype-phenotype model. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkt577

Choudhary, S., & Satija, R. (2022). Comparison and evaluation of statistical error models for scRNA-seq. *Genome Biology*, *23*(1). https://doi.org/10.1186/s13059-021-02584-9

Cibrián, D., & Sánchez-Madrid, F. (2017). CD69: from activation marker to metabolic gatekeeper. In *European Journal of Immunology* (Vol. 47, Issue 6). https://doi.org/10.1002/eji.201646837

Clairambault, J., & Shen, S. (2020). Cell plasticity in cancer cell populations. In *F1000Research* (Vol. 9). https://doi.org/10.12688/f1000research.24803.1

Clevers, H. (2011). The cancer stem cell: Premises, promises and challenges. *Nature Medicine*. https://doi.org/10.1038/nm.2304

Cline, E., Wisittipanit, N., Boongoen, T., Chukeatirote, E., Struss, D., & Eungwanichayapant, A. (2020). Recalibration of mapping quality scores in Illumina short-read alignments improves SNP detection results in low-coverage sequencing data. *PeerJ*, *8*. https://doi.org/10.7717/peerj.10501

Cockman, E., Anderson, P., & Ivanov, P. (2020). Top mrnps: Molecular mechanisms and principles of regulation. In *Biomolecules*. https://doi.org/10.3390/biom10070969

Combes, A. J., Samad, B., & Krummel, M. F. (2023). Defining and using immune archetypes to classify and treat cancer. *Nature Reviews Cancer*, *23*(7). https://doi.org/10.1038/s41568-023-00578-2

Combes, A. J., Samad, B., Tsui, J., Chew, N. W., Yan, P., Reeder, G. C., Kushnoor, D., Shen, A., Davidson, B., Barczak, A. J., Adkisson, M., Edwards, A., Naser, M., Barry, K. C., Courau, T., Hammoudi, T., Argüello, R. J., Rao, A. A., Olshen, A. B., … Krummel, M. F. (2022). Discovering dominant tumor immune archetypes in a pan-cancer census. *Cell*, *185*(1). https://doi.org/10.1016/j.cell.2021.12.004

Combes, A. J., Samad, B., Tsui, J., Chew, N. W., Yan, P., Reeder, G. C., Kushnoor, D., Shen, A., Davidson, B., Barczak, A. J., Adkisson, M., Edwards, A., Naser, M., Barry, K. C., Courau, T., Hammoudi, T., Arguello, R., Rao, A. A., Olshen, A. B., … Krummel, M. F. (2021). A Pan-Cancer Census of Dominant Tumor Immune Archetypes. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3833003

Cooke, M. E., Jones, S. W., ter Horst, B., Moiemen, N., Snow, M., Chouhan, G., Hill, L. J., Esmaeli, M., Moakes, R. J. A., Holton, J., Nandra, R., Williams, R. L., Smith, A. M., & Grover, L. M. (2018). Structuring of Hydrogels across Multiple Length Scales for Biomedical Applications. *Advanced Materials*, *30*(14). https://doi.org/10.1002/adma.201705013

Corbett, A. H. (2018). Post-transcriptional regulation of gene expression and human disease. In *Current Opinion in Cell Biology* (Vol. 52). https://doi.org/10.1016/j.ceb.2018.02.011

Cormier, P., Pyronnet, S., Salaün, P., Mulner-Lorillon, O., & Sonenberg, N. (2003). Cap-dependent translation and control of the cell cycle. In *Progress in cell cycle research* (Vol. 5).

Coronado, D., Godet, M., Bourillot, P. Y., Tapponnier, Y., Bernat, A., Petit, M., Afanassieff, M., Markossian, S., Malashicheva, A., Iacone, R., Anastassiadis, K., & Savatier, P. (2013). A short G1 phase is an intrinsic determinant of naïve embryonic stem cell pluripotency. *Stem Cell Research*, *10*(1). https://doi.org/10.1016/j.scr.2012.10.004

Couvillion, M., Harlen, K. M., Lachance, K. C., Trotta, K. L., Smith, E., Brion, C., Smalec, B. M., & Churchman, L. S. (2022). Transcription elongation is finely tuned by dozens of regulatory factors. *ELife*, *11*. https://doi.org/10.7554/eLife.78944

Cowling, V. H. (2010). Regulation of mRNA cap methylation. In *Biochemical Journal* (Vol. 425, Issue 2). https://doi.org/10.1042/BJ20091352

Csibi, A., Lee, G., Yoon, S. O., Tong, H., Ilter, D., Elia, I., Fendt, S. M., Roberts, T. M., & Blenis, J. (2014). The mTORC1/S6K1 pathway regulates glutamine metabolism through the eif4b-dependent control of c-Myc translation. *Current Biology*, *24*(19). https://doi.org/10.1016/j.cub.2014.08.007

Cvetesic, N., Leitch, H. G., Borkowska, M., Müller, F., Carninci, P., Hajkova, P., & Lenhard, B. (2018). SLIC-CAGE: High-resolution transcription start site mapping using nanogram-levels of total RNA. *Genome Research*, *28*(12). https://doi.org/10.1101/gr.235937.118

Cyll, K., Ersvær, E., Vlatkovic, L., Pradhan, M., Kildal, W., Avranden Kjær, M., Kleppe, A., Hveem, T. S., Carlsen, B., Gill, S., Löffeler, S., Haug, E. S., Wæhre, H., Sooriakumaran, P., & Danielsen, H. E. (2017). Tumour heterogeneity poses a significant challenge to cancer biomarker research. *British Journal of Cancer*, *117*(3). https://doi.org/10.1038/bjc.2017.171

Dagogo-Jack, I., & Shaw, A. T. (2018). Tumour heterogeneity and resistance to cancer therapies. In *Nature Reviews Clinical Oncology*. https://doi.org/10.1038/nrclinonc.2017.166

Dalton, S. (2015). Linking the Cell Cycle to Cell Fate Decisions. In *Trends in Cell Biology* (Vol. 25, Issue 10). https://doi.org/10.1016/j.tcb.2015.07.007

Dang, C. V. (2012). MYC on the path to cancer. In *Cell* (Vol. 149, Issue 1). https://doi.org/10.1016/j.cell.2012.03.003

Danks, G. B., Navratilova, P., Lenhard, B., & Thompson, E. M. (2018). Distinct core promoter codes drive transcription initiation at key developmental transitions in a marine chordate. *BMC Genomics*, *19*(1), 164. https://doi.org/10.1186/s12864-018-4504-5

Das, S. K., Lewis, B. A., & Levens, D. (2023). MYC: a complex problem. In *Trends in Cell Biology* (Vol. 33, Issue 3). https://doi.org/10.1016/j.tcb.2022.07.006

de la Parra, C., Ernlund, A., Alard, A., Ruggles, K., Ueberheide, B., & Schneider, R. J. (2018). A widespread alternate form of cap-dependent mRNA translation initiation. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-05539-0

de Medeiros, G., Ortiz, R., Strnad, P., Boni, A., Moos, F., Repina, N., Challet Meylan, L., Maurer, F., & Liberali, P. (2022). Multiscale light-sheet organoid imaging framework. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-32465-z

De Rosa, M., Pace, U., Rega, D., Costabile, V., Duraturo, F., Izzo, P., & Delrio, P. (2015). Genetics, diagnosis and management of colorectal cancer (Review). *Oncology Reports*. https://doi.org/10.3892/or.2015.4108

De Turris, V., Di Leva, G., Caldarola, S., Loreni, F., Amaldi, F., & Bozzoni, I. (2004). TOP promoter elements control the relative ratio of intron-encoded snoRNA versus spliced mRNA biosynthesis. *Journal of Molecular Biology*. https://doi.org/10.1016/j.jmb.2004.09.049

DeBerardinis, R. J., Lum, J. J., Hatzivassiliou, G., & Thompson, C. B. (2008). The Biology of Cancer: Metabolic Reprogramming Fuels Cell Growth and Proliferation. In *Cell Metabolism* (Vol. 7, Issue 1). https://doi.org/10.1016/j.cmet.2007.10.002

Demircioğlu, D., Cukuroglu, E., Kindermans, M., Nandi, T., Calabrese, C., Fonseca, N. A., Kahles, A., Lehmann, K. Van, Stegle, O., Brazma, A., Brooks, A. N., Rätsch, G., Tan, P., & Göke, J. (2019). A Pan-cancer Transcriptome Analysis Reveals Pervasive Regulation through Alternative Promoters. *Cell*. https://doi.org/10.1016/j.cell.2019.08.018

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1). https://doi.org/10.1093/bioinformatics/bts635

Eales, K. L., Hollinshead, K. E. R., & Tennant, D. A. (2016). Hypoxia and metabolic adaptation of cancer cells. *Oncogenesis*. https://doi.org/10.1038/oncsis.2015.50

Eberle, A. L., Selchow, O., Thaler, M., Zeidler, D., & Kirmse, R. (2015). Mission (im)possible – Mapping the brain becomes a reality. In *Microscopy* (Vol. 64, Issue 1). https://doi.org/10.1093/jmicro/dfu104

Ebert, S. M., Rasmussen, B. B., Judge, A. R., Judge, S. M., Larsson, L., Wek, R. C., Anthony, T. G., Marcotte, G. R., Miller, M. J., Yorek, M. A., Vella, A., Volpi, E., Stern, J. I., Strub, M. D., Ryan, Z., Talley, J. J., & Adams, C. M. (2022). Biology of Activating Transcription Factor 4 (ATF4) and Its Role in Skeletal Muscle Atrophy. *Journal of Nutrition*, *152*(4). https://doi.org/10.1093/jn/nxab440

Endo, H., Owada, S., Inagaki, Y., Shida, Y., & Tatemichi, M. (2020). Metabolic reprogramming sustains cancer cell survival following extracellular matrix detachment. *Redox Biology*, *36*. https://doi.org/10.1016/j.redox.2020.101643

Farooq, Z., Kusuma, F., Burke, P., Dufour, C. R., Lee, D., Tabatabaei, N., Toboz, P., Radovani, E., Greenblatt, J. F., Rehman, J., Class, J., Khoutorsky, A., Fonseca, B. D., Richner, J. M., Mercier, E., Bourque, G., Giguère, V., Subramaniam, A. R., Han, J., & Tahmasebi, S. (2022). The amino acid sensor GCN2 suppresses terminal oligopyrimidine (TOP) mRNA translation via La-related

protein 1 (LARP1). *Journal of Biological Chemistry*, *298*(9). https://doi.org/10.1016/j.jbc.2022.102277

Faubert, B., Solmonson, A., & DeBerardinis, R. J. (2020). Metabolic reprogramming and cancer progression. In *Science*. https://doi.org/10.1126/science.aaw5473

Fedi, A., Vitale, C., Ponschin, G., Ayehunie, S., Fato, M., & Scaglione, S. (2021). In vitro models replicating the human intestinal epithelium for absorption and metabolism studies: A systematic review. In *Journal of Controlled Release* (Vol. 335). https://doi.org/10.1016/j.jconrel.2021.05.028

Feitelson, M. A., Arzumanyan, A., Kulathinal, R. J., Blain, S. W., Holcombe, R. F., Mahajna, J., Marino, M., Martinez-Chantar, M. L., Nawroth, R., Sanchez-Garcia, I., Sharma, D., Saxena, N. K., Singh, N., Vlachostergios, P. J., Guo, S., Honoki, K., Fujii, H., Georgakilas, A. G., Bilsland, A., … Nowsheen, S. (2015). Sustained proliferation in cancer: Mechanisms and novel therapeutic targets. In *Seminars in Cancer Biology* (Vol. 35). https://doi.org/10.1016/j.semcancer.2015.02.006

Fischer, M., Schade, A. E., Branigan, T. B., Müller, G. A., & DeCaprio, J. A. (2022). Coordinating gene expression during the cell cycle. In *Trends in Biochemical Sciences* (Vol. 47, Issue 12). https://doi.org/10.1016/j.tibs.2022.06.007

Fisher, R., Pusztai, L., & Swanton, C. (2013). Cancer heterogeneity: Implications for targeted therapeutics. In *British Journal of Cancer*. https://doi.org/10.1038/bjc.2012.581

Fleifel, D., & Cook, J. G. (2023). G1 Dynamics at the Crossroads of Pluripotency and Cancer. In *Cancers* (Vol. 15, Issue 18). https://doi.org/10.3390/cancers15184559

Floriddia, E. (2022). The impact of ambient RNA. In *Nature Neuroscience* (Vol. 25, Issue 12). https://doi.org/10.1038/s41593-022-01232-0

Foerster, E. G., Mukherjee, T., Cabral-Fernandes, L., Rocha, J. D. B., Girardin, S. E., & Philpott, D. J. (2022). How autophagy controls the intestinal epithelial barrier. *Autophagy*, *18*(1). https://doi.org/10.1080/15548627.2021.1909406

Forrest, A. R. R., Kawaji, H., Rehli, M., Baillie, J. K., De Hoon, M. J. L., Haberle, V., Lassmann, T., Kulakovskiy, I. V., Lizio, M., Itoh, M., Andersson, R., Mungall, C. J., Meehan, T. F., Schmeier, S., Bertin, N., Jørgensen, M., Dimont, E., Arner, E., Schmidl, C., … Hayashizaki, Y. (2014). A promoter-level mammalian expression atlas. *Nature*, *507*(7493). https://doi.org/10.1038/nature13182

Foy, R., Crozier, L., Pareri, A. U., Valverde, J. M., Park, B. H., Ly, T., & Saurin, A. T. (2023). Oncogenic signals prime cancer cells for toxic cell overgrowth during a G1 cell cycle arrest. *Molecular Cell*, *83*(22). https://doi.org/10.1016/j.molcel.2023.10.020

Fujii, M., Shimokawa, M., Date, S., Takano, A., Matano, M., Nanki, K., Ohta, Y., Toshimitsu, K., Nakazato, Y., Kawasaki, K., Uraoka, T., Watanabe, T., Kanai, T., & Sato, T. (2016). A Colorectal Tumor Organoid Library Demonstrates Progressive Loss of Niche Factor Requirements during Tumorigenesis. *Cell Stem Cell*, *18*(6). https://doi.org/10.1016/j.stem.2016.04.003

Furuichi, Y. (2015). Discovery of m7G-cap in eukaryotic mRNAs. In *Proceedings of the Japan Academy Series B: Physical and Biological Sciences* (Vol. 91, Issue 8). https://doi.org/10.2183/pjab.91.394

Gaglia, G., Kabraji, S., Rammos, D., Dai, Y., Verma, A., Wang, S., Mills, C. E., Chung, M., Bergholz, J. S., Coy, S., Lin, J. R., Jeselsohn, R., Metzger, O., Winer, E. P., Dillon, D. A., Zhao, J. J., Sorger, P. K., & Santagata, S. (2022). Temporal and spatial topography of cell proliferation in cancer. *Nature Cell Biology*, *24*(3). https://doi.org/10.1038/s41556-022-00860-9

Gallaher, J. A., Brown, J. S., & Anderson, A. R. A. (2019). The impact of proliferation-migration tradeoffs on phenotypic evolution in cancer. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-39636-x

García-Gutiérrez, L., Bretones, G., Molina, E., Arechaga, I., Symonds, C., Acosta, J. C., Blanco, R., Fernández, A., Alonso, L., Sicinski, P., Barbacid, M., Santamaría, D., & León, J. (2019). Myc stimulates cell cycle progression through the activation of Cdk1 and phosphorylation of p27. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-54917-1

Ge, S. X., Jung, D., Jung, D., & Yao, R. (2020). ShinyGO: A graphical gene-set enrichment tool for animals and plants. *Bioinformatics*, *36*(8). https://doi.org/10.1093/bioinformatics/btz931

Gonzalgo, M. L., & Jones, P. A. (1997). Mutagenic and epigenetic effects of DNA methylation. In *Mutation Research - Reviews in Mutation Research*. https://doi.org/10.1016/S1383-5742(96)00047-6

Groves, S. M., Ildefonso, G. V., McAtee, C. O., Ozawa, P. M. M., Ireland, A. S., Stauffer, P. E., Wasdin, P. T., Huang, X., Qiao, Y., Lim, J. S., Bader, J., Liu, Q., Simmons, A. J., Lau, K. S., Iams, W. T., Hardin, D. P., Saff, E. B., Holmes, W. R., Tyson, D. R., … Quaranta, V. (2022). Archetype tasks link intratumoral heterogeneity to plasticity and cancer hallmarks in small cell lung cancer. *Cell Systems*, *13*(9). https://doi.org/10.1016/j.cels.2022.07.006

Groves, S. M., Ireland, A., Liu, Q., Simmons, A. J., Lau, K., Iams, W. T., Tyson, D., Lovly, C. M., Oliver, T. G., & Quaranta, V. (2021). Cancer Hallmarks Define a Continuum of Plastic Cell States between Small Cell Lung Cancer Archetypes. *BioRxiv*.

Guinney, J., Dienstmann, R., Wang, X., De Reyniès, A., Schlicker, A., Soneson, C., Marisa, L., Roepman, P., Nyamundanda, G., Angelino, P., Bot, B. M., Morris, J. S., Simon, I. M., Gerster, S., Fessler, E., De Sousa .E Melo, F., Missiaglia, E., Ramay, H., Barras, D., … Tejpar, S. (2015). The consensus molecular subtypes of colorectal cancer. *Nature Medicine*, *21*(11). https://doi.org/10.1038/nm.3967

Guo, D., Bell, E., Mischel, P., & Chakravarti, A. (2015). Targeting SREBP-1-driven Lipid Metabolism to Treat Cancer. *Current Pharmaceutical Design*, *20*(15). https://doi.org/10.2174/1381612811319999048 6

Haberle, V., Forrest, A. R. R., Hayashizaki, Y., Carninci, P., & Lenhard, B. (2015). CAGEr: Precise TSS data retrieval and high-resolution promoterome mining for integrative analyses. *Nucleic Acids Research*, *43*(8). https://doi.org/10.1093/nar/gkv054

Haberle, V., & Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. In *Nature Reviews Molecular Cell Biology* (Vol. 19, Issue 10). https://doi.org/10.1038/s41580-018-0028-8

Hanahan, D., & Weinberg, R. A. (2000). The Hallmarks of Cancer Review Douglas. *Cell*, *100*(7).

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. In *Cell*. https://doi.org/10.1016/j.cell.2011.02.013

Hart, Y., Sheftel, H., Hausser, J., Szekely, P., Ben-Moshe, N. B., Korem, Y., Tendler, A., Mayo, A. E., & Alon, U. (2015). Inferring biological tasks using Pareto analysis of high-dimensional data. *Nature Methods*, *12*(3). https://doi.org/10.1038/nmeth.3254

Hausser, J., & Alon, U. (2020). Tumour heterogeneity and the evolutionary trade-offs of cancer. In *Nature Reviews Cancer*. https://doi.org/10.1038/s41568-020-0241-6

Hausser, J., Szekely, P., Bar, N., Zimmer, A., Sheftel, H., Caldas, C., & Alon, U. (2019). Tumor diversity and the trade-off between universal cancer tasks. *Nature Communications*. https://doi.org/10.1038/s41467-019-13195-1

Heiser, L. M., Sadanandam, A., Kuo, W. L., Benz, S. C., Goldstein, T. C., Ng, S., Gibb, W. J., Wang, N. J., Ziyad, S., Tong, F., Bayani, N., Hu, Z., Billig, J. I., Dueregger, A., Lewis, S., Jakkula, L., Korkola, J. E., Durinck, S., Pepin, F., … Spellman, P. T. (2012). Subtype and pathway specific responses to anticancer compounds in breast cancer. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1018854108

Herman, J. G., & Baylin, S. B. (2003). Gene Silencing in Cancer in Association with Promoter Hypermethylation. *New England Journal of Medicine*, *349*(21). https://doi.org/10.1056/nejmra023075

Hinnebusch, A. G., & Lorsch, J. R. (2012). The mechanism of eukaryotic translation initiation: New insights and challenges. *Cold Spring Harbor Perspectives in Biology*, *4*(10). https://doi.org/10.1101/cshperspect.a011544

Ho, J. J. D., Balukoff, N. C., Cervantes, G., Malcolm, P. D., Krieger, J. R., & Lee, S. (2018). Oxygen-Sensitive Remodeling of Central Carbon Metabolism by Archaic eIF5B. *Cell Reports*. https://doi.org/10.1016/j.celrep.2017.12.031

Ho, J. J. D., Balukoff, N. C., Theodoridis, P. R., Wang, M., Krieger, J. R., Schatz, J. H., & Lee, S. (2020). A network of RNA-binding proteins controls translation efficiency to activate anaerobic metabolism. *Nature Communications*. https://doi.org/10.1038/s41467-020-16504-1

Ho, J. J. D., & Lee, S. (2016). A Cap for Every Occasion: Alternative eIF4F Complexes. In *Trends in Biochemical Sciences*. https://doi.org/10.1016/j.tibs.2016.05.009

Ho, J. J. D., Schatz, J. H., Uniacke, J., & Lee, S. (2021). Jekyll and Hyde: Activating the Hypoxic Translational Machinery. In *Trends in Biochemical Sciences* (Vol. 46, Issue 3). https://doi.org/10.1016/j.tibs.2020.11.006

Ho, J. J. D., Wang, M., Audas, T. E., Kwon, D., Carlsson, S. K., Timpano, S., Evagelou, S. L., Brothers, S., Gonzalgo, M. L., Krieger, J. R., Chen, S., Uniacke, J., & Lee, S. (2016). Systemic Reprogramming of Translation Efficiencies on Oxygen Stimulus. *Cell Reports*, *14*(6), 1293–1300. https://doi.org/10.1016/j.celrep.2016.01.036

Hochstoeger, T., & Chao, J. A. (2024). Towards a molecular understanding of the 5'TOP motif in regulating translation of ribosomal mRNAs. In *Seminars in Cell and Developmental Biology* (Vol. 154). https://doi.org/10.1016/j.semcdb.2023.06.001

Hochstoeger, T., Papasaikas, P., Piskadlo, E., & Chao, J. A. (2024). Distinct roles of LARP1 and 4EBP1/2 in regulating translation and stability of 5'TOP mRNAs. *Science Advances*, *10*(7). https://doi.org/10.1126/sciadv.adi7830

Hof, L., Moreth, T., Koch, M., Liebisch, T., Kurtz, M., Tarnick, J., Lissek, S. M., Verstegen, M. M. A., van der Laan, L. J. W., Huch, M., Matthäus, F., Stelzer, E. H. K., & Pampaloni, F. (2021). Long-term live imaging and multiscale analysis identify heterogeneity and core principles of epithelial organoid morphogenesis. *BMC Biology*, *19*(1). https://doi.org/10.1186/s12915-021-00958-w

Holton, E., Muskovic, W., & Powell, J. E. (2024). Deciphering cancer cell state plasticity with single-cell genomics and artificial intelligence. In *Genome Medicine* (Vol. 16, Issue 1). https://doi.org/10.1186/s13073-024-01309-4

Hoorn, S. Ten, De Back, T. R., Sommeijer, D. W., & Vermeulen, L. (2022). Clinical Value of Consensus Molecular Subtypes in Colorectal Cancer: A Systematic Review and Meta-Analysis. In *Journal of the National Cancer Institute* (Vol. 114, Issue 4). https://doi.org/10.1093/jnci/djab106

Hopkins, T. G., Mura, M., Al-Ashtal, H. A., Lahr, R. M., Abd-Latip, N., Sweeney, K., Lu, H., Weir, J., El-Bahrawy, M., Steel, J. H., Ghaem-Maghami, S., Aboagye, E. O., Berman, A. J., & Blagden, S. P. (2016). The RNA-binding protein LARP1 is a post-transcriptional regulator of survival and tumorigenesis in ovarian cancer. *Nucleic Acids Research*, *44*(3). https://doi.org/10.1093/nar/gkv1515

Houseley, J., & Tollervey, D. (2009). The Many Pathways of RNA Degradation. In *Cell* (Vol. 136, Issue 4). https://doi.org/10.1016/j.cell.2009.01.019

Hsieh, A. C., Liu, Y., Edlind, M. P., Ingolia, N. T., Janes, M. R., Sher, A., Shi, E. Y., Stumpf, C. R., Christensen, C., Bonham, M. J., Wang, S., Ren, P., Martin, M., Jessen, K., Feldman, M. E., Weissman, J. S., Shokat, K. M., Rommel, C., & Ruggero, D. (2012). The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature*. https://doi.org/10.1038/nature10912

Huang, J., & Manning, B. D. (2008). The TSC1-TSC2 complex: A molecular switchboard controlling cell growth. In *Biochemical Journal* (Vol. 412, Issue 2). https://doi.org/10.1042/BJ20080281

Huang, K. K., Huang, J., Wu, J. K. L., Lee, M., Tay, S. T., Kumar, V., Ramnarayanan, K., Padmanabhan, N., Xu, C., Tan, A. L. K., Chan, C., Kappei, D., Göke, J., & Tan, P. (2021). Long-read transcriptome sequencing reveals abundant promoter diversity in distinct molecular subtypes of gastric cancer. *Genome Biology*, *22*(1). https://doi.org/10.1186/s13059-021-02261-x

Huang, W., Hickson, L. T. J., Eirin, A., Kirkland, J. L., & Lerman, L. O. (2022). Cellular senescence: the good, the bad and the unknown. In *Nature Reviews Nephrology* (Vol. 18, Issue 10). https://doi.org/10.1038/s41581-022-00601-z

Huin, V., Buée, L., Behal, H., Labreuche, J., Sablonnière, B., & Dhaenens, C. M. (2017). Alternative promoter usage generates novel shorter MAPT mRNA transcripts in Alzheimer's disease and progressive supranuclear palsy brains. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-12955-7

Hwang, W. C., Song, D., Lee, H., Oh, C., Lim, S. H., Bae, H. J., Kim, N. D., Han, G., & Min, D. S. (2022). Inhibition of phospholipase D1 induces immunogenic cell death and potentiates cancer immunotherapy in colorectal cancer. *Experimental and Molecular Medicine*, *54*(9). https://doi.org/10.1038/s12276-022-00853-6

Hwang, Y., Hidalgo, D., & Socolovsky, M. (2020). The shifting shape and functional specializations of the cell cycle during lineage development. In *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*. https://doi.org/10.1002/wsbm.1504

Iacona, J. R., & Lutz, C. S. (2019). miR-146a-5p: Expression, regulation, and functions in cancer. In *Wiley Interdisciplinary Reviews: RNA*. https://doi.org/10.1002/wrna.1533

Ilicic, T., Kim, J. K., Kolodziejczyk, A. A., Bagger, F. O., McCarthy, D. J., Marioni, J. C., & Teichmann, S. A. (2016). Classification of low quality cells from single-cell RNA-seq data. *Genome Biology*, *17*(1). https://doi.org/10.1186/s13059-016-0888-1

Islam, Z., Ali, A. M., Naik, A., Eldaw, M., Decock, J., & Kolatkar, P. R. (2021). Transcription Factors: The Fulcrum Between Cell Development and Carcinogenesis. In *Frontiers in Oncology* (Vol. 11). https://doi.org/10.3389/fonc.2021.681377

Ito, M., Araki, S., Matsunaga, S., Itoh, T., Nishihama, R., Machida, Y., Doonan, J. H., & Watanabe, A. (2001). G2/M-Phase–Specific Transcription during the Plant Cell Cycle Is Mediated by c-Myb–Like Transcription Factors. *The Plant Cell*, *13*(8). https://doi.org/10.1105/tpc.010102

Jamshidi, P., Chouhan, G., Williams, R. L., Cox, S. C., & Grover, L. M. (2016). Modification of gellan gum with nanocrystalline hydroxyapatite facilitates cell expansion and spontaneous osteogenesis. *Biotechnology and Bioengineering*, *113*(7). https://doi.org/10.1002/bit.25915

Jeon, S. M. (2016). Regulation and function of AMPK in physiology and diseases. In *Experimental and Molecular Medicine* (Vol. 48, Issue 7). https://doi.org/10.1038/EMM.2016.81

Jeong, N., Kim, S. C., Park, J. W., Park, S. G., Nam, K. H., Lee, J. O., Shin, Y. K., Bae, J. M., Jeong, S. Y., Kim, M. J., & Ku, J. L. (2022). Multifocal organoids reveal clonal associations between synchronous intestinal tumors with pervasive heterogeneous drug responses. *Npj Genomic Medicine*, *7*(1). https://doi.org/10.1038/s41525-022-00313-0

Jia, J. J., Lahr, R. M., Solgaard, M. T., Moraes, B. J., Pointet, R., Yang, A. D., Celucci, G., Graber, T. E., Hoang, H. D., Niklaus, M. R., Pena, I. A., Hollensen, A. K., Smith, E. M., Chaker-Margot, M., Anton, L., Dajadian, C., Livingstone, M., Hearnden, J., Wang, X. D., … Fonseca, B. D. (2021). MTORC1 promotes TOP mRNA translation through site-specific phosphorylation of LARP1. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkaa1239

Jiang, J., & Liu, L. Y. (2015). Zinc finger protein X-linked is overexpressed in colorectal cancer and is associated with poor prognosis. *Oncology Letters*, *10*(2). https://doi.org/10.3892/ol.2015.3353

Jin, H. R., Wang, J., Wang, Z. J., Xi, M. J., Xia, B. H., Deng, K., & Yang, J. L. (2023). Lipid metabolic reprogramming in tumor microenvironment: from mechanisms to therapeutics. In *Journal of Hematology and Oncology* (Vol. 16, Issue 1). https://doi.org/10.1186/s13045-023-01498-2

Jin, H., Zhang, C., Zwahlen, M., von Feilitzen, K., Karlsson, M., Shi, M., Yuan, M., Song, X., Li, X., Yang, H., Turkez, H., Fagerberg, L., Uhlén, M., & Mardinoglu, A. (2023). Systematic transcriptional analysis of human cell lines for gene expression landscape and tumor representation. *Nature Communications*, *14*(1). https://doi.org/10.1038/s41467-023-41132-w

Jingwen, B., Yaochen, L., & Guojun, Z. (2017). Cell cycle regulation and anticancer drug discovery. *Cancer Biology & Medicine*, *14*(4). https://doi.org/10.20892/j.issn.2095-3941.2017.0033

Jögi, A., Vaapil, M., Johansson, M., & Påhlman, S. (2012). Cancer cell differentiation heterogeneity and aggressive behavior in solid tumors. In *Upsala Journal of Medical Sciences* (Vol. 117, Issue 2). https://doi.org/10.3109/03009734.2012.659294

Jones, M. F., Hara, T., Francis, P., Li, X. L., Bilke, S., Zhu, Y., Pineda, M., Subramanian, M., Bodmer, W. F., & Lal, A. (2015). The CDX1-microRNA-215 axis regulates colorectal cancer stem cell differentiation. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(13). https://doi.org/10.1073/pnas.1503370112

Kaplon, J., van Dam, L., & Peeper, D. (2015). Two-way communication between the metabolic and cell cycle machineries: the molecular basis. In *Cell Cycle* (Vol. 14, Issue 13). https://doi.org/10.1080/15384101.2015.1044172

Kashfi, S. M. H., Almozyan, S., Jinks, N., Koo, B. K., & Nateri, A. S. (2018). Morphological alterations of cultured human colorectal matched tumour and healthy organoids. *Oncotarget*, *9*(12). https://doi.org/10.18632/oncotarget.24279

Kaushik, G., Ponnusamy, M. P., & Batra, S. K. (2018). Concise Review: Current Status of Three-Dimensional Organoids as Preclinical Models. In *Stem Cells*. https://doi.org/10.1002/stem.2852

Kazanets, A., Shorstova, T., Hilmi, K., Marques, M., & Witcher, M. (2016). Epigenetic silencing of tumor suppressor genes: Paradigms, puzzles, and potential. In *Biochimica et Biophysica Acta - Reviews on Cancer* (Vol. 1865, Issue 2). https://doi.org/10.1016/j.bbcan.2016.04.001

Keibler, M. A., Wasylenko, T. M., Kelleher, J. K., Iliopoulos, O., Vander Heiden, M. G., & Stephanopoulos, G. (2016). Metabolic requirements for cancer cell proliferation. *Cancer & Metabolism*, *4*(1). https://doi.org/10.1186/s40170-016-0156-6

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, and D. (2002). The Human Genome Browser at UCSC. *Genome Research*, *12*(6). https://doi.org/10.1101/gr.229102

Khaliq, A. M., Erdogan, C., Kurt, Z., Turgut, S. S., Grunvald, M. W., Rand, T., Khare, S., Borgia, J. A., Hayden, D. M., Pappas, S. G., Govekar, H. R., Kam, A. E., Reiser, J., Turaga, K., Radovich, M., Zang, Y., Qiu, Y., Liu, Y., Fishel, M. L., … Masood, A. (2022). Refining colorectal cancer classification and clinical stratification through a single-cell atlas. *Genome Biology*, *23*(1). https://doi.org/10.1186/s13059-022-02677-z

Kodzius, R., Kojima, M., Nishiyori, H., Nakamura, M., Fukuda, S., Tagami, M., Sasaki, D., Imamura, K., Kai, C., Harbers, M., Hayashizaki, Y., & Carninci, P. (2006). Cage: Cap analysis of gene expression. *Nature Methods*. https://doi.org/10.1038/nmeth0306-211

Koh, S. B., Mascalchi, P., Rodriguez, E., Lin, Y., Jodrell, D. I., Richards, F. M., & Lyons, S. K. (2017). A quantitative FastFUCCI assay defines cell cycle dynamics at a single-cell level. *Journal of Cell Science*, *130*(2). https://doi.org/10.1242/jcs.195164

Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P. ru, & Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods*, *16*(12). https://doi.org/10.1038/s41592-019-0619-0

Kotliar, D., Veres, A., Nagy, M. A., Tabrizi, S., Hodis, E., Melton, D. A., & Sabeti, P. C. (2019). Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. *ELife*, *8*. https://doi.org/10.7554/eLife.43803

Koundouros, N., & Poulogiannis, G. (2020). Reprogramming of fatty acid metabolism in cancer. In *British Journal of Cancer* (Vol. 122, Issue 1). https://doi.org/10.1038/s41416-019-0650-z

Kouno, T., Moody, J., Kwon, A. T. J., Shibayama, Y., Kato, S., Huang, Y., Böttcher, M., Motakis, E., Mendez, M., Severin, J., Luginbühl, J., Abugessaisa, I., Hasegawa, A., Takizawa, S., Arakawa, T., Furuno, M., Ramalingam, N., West, J., Suzuki, H., … Shin, J. W. (2019). C1 CAGE detects transcription start sites and enhancer activity at single-cell resolution. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-018-08126-5

Kowalczyk, M. S., Tirosh, I., Heckl, D., Rao, T. N., Dixit, A., Haas, B. J., Schneider, R. K., Wagers, A. J., Ebert, B. L., & Regev, A. (2015). Single-cell RNA-seq reveals changes in cell cycle and

differentiation programs upon aging of hematopoietic stem cells. *Genome Research*, *25*(12). https://doi.org/10.1101/gr.192237.115

Kozlowski, M. T., Crook, C. J., & Ku, H. T. (2021). Towards organoid culture without Matrigel. In *Communications Biology* (Vol. 4, Issue 1). https://doi.org/10.1038/s42003-021-02910-8

Krausova, M., & Korinek, V. (2014). Wnt signaling in adult intestinal stem cells and cancer. In *Cellular Signalling* (Vol. 26, Issue 3). https://doi.org/10.1016/j.cellsig.2013.11.032

Kreß, J. K. C., Jessen, C., Hufnagel, A., Schmitz, W., Xavier da Silva, T. N., Ferreira dos Santos, A., Mosteo, L., Goding, C. R., Friedmann Angeli, J. P., & Meierjohann, S. (2023). The integrated stress response effector ATF4 is an obligatory metabolic activator of NRF2. *Cell Reports*, *42*(7). https://doi.org/10.1016/j.celrep.2023.112724

Kronja, I., & Orr-Weaver, T. L. (2011). Translational regulation of the cell cycle: When, where, how and why? In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 366, Issue 1584). https://doi.org/10.1098/rstb.2011.0084

Kuczler, M. D., Olseen, A. M., Pienta, K. J., & Amend, S. R. (2021). ROS-induced cell cycle arrest as a mechanism of resistance in polyaneuploid cancer cells (PACCs). *Progress in Biophysics and Molecular Biology*, *165*. https://doi.org/10.1016/j.pbiomolbio.2021.05.002

Kumari, R., & Jat, P. (2021). Mechanisms of Cellular Senescence: Cell Cycle Arrest and Senescence Associated Secretory Phenotype. In *Frontiers in Cell and Developmental Biology* (Vol. 9). https://doi.org/10.3389/fcell.2021.645593

Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T. R., & Weirauch, M. T. (2018). The Human Transcription Factors. In *Cell*. https://doi.org/10.1016/j.cell.2018.01.029

Lancaster, M. A., & Huch, M. (2019). Disease modelling in human organoids. *DMM Disease Models and Mechanisms*. https://doi.org/10.1242/dmm.039347

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3). https://doi.org/10.1186/gb-2009-10-3-r25

Laplante, M., & Sabatini, D. M. (2013). Regulation of mTORC1 and its impact on gene expression at a glance. *Journal of Cell Science*, *126*(8). https://doi.org/10.1242/jcs.125773

Lause, J., Berens, P., & Kobak, D. (2021). Analytic Pearson residuals for normalization of single-cell RNA-seq UMI data. *Genome Biology*, *22*(1). https://doi.org/10.1186/s13059-021-02451-7

Lawrence, M., Gentleman, R., & Carey, V. (2009). rtracklayer: An R package for interfacing with genome browsers. *Bioinformatics*, *25*(14). https://doi.org/10.1093/bioinformatics/btp328

Lawson, D. A., Kessenbrock, K., Davis, R. T., Pervolarakis, N., & Werb, Z. (2018). Tumour heterogeneity and metastasis at single-cell resolution. In *Nature Cell Biology* (Vol. 20, Issue 12). https://doi.org/10.1038/s41556-018-0236-7

Leal-Esteban, L. C., & Fajas, L. (2020). Cell cycle regulators in cancer cell metabolism. In *Biochimica et Biophysica Acta - Molecular Basis of Disease* (Vol. 1866, Issue 5). https://doi.org/10.1016/j.bbadis.2020.165715

Lee, H., Woo, S. M., Jang, H., Kang, M., & Kim, S. Y. (2022). Cancer depends on fatty acids for ATP production: A possible link between cancer and obesity. In *Seminars in Cancer Biology* (Vol. 86). https://doi.org/10.1016/j.semcancer.2022.07.005

Lenz, G., Onzi, G. R., Lenz, L. S., Buss, J. H., dos Santos, J. A., & Begnini, K. R. (2021). The Origins of Phenotypic Heterogeneity in Cancer. *Cancer Research*, *82*(1). https://doi.org/10.1158/0008-5472.CAN-21-1940

Li, H., Hou, J., Bai, L., Hu, C., Tong, P., Kang, Y., Zhao, X., & Shao, Z. (2015). Genome-wide analysis of core promoter structures in Schizosaccharomyces pombe with DeepCAGE. *RNA Biology*, *12*(5), 525–537. https://doi.org/10.1080/15476286.2015.1022704

Li, M., Sun, Q., & Wang, X. (2017). Transcriptional landscape of human cancers. *Oncotarget*, *8*(21). https://doi.org/10.18632/oncotarget.15837

Li, X., Yang, Y., Zhang, B., Lin, X., Fu, X., An, Y., Zou, Y., Wang, J. X., Wang, Z., & Yu, T. (2022). Lactate metabolism in human health and disease. In *Signal Transduction and Targeted Therapy* (Vol. 7, Issue 1). https://doi.org/10.1038/s41392-022-01151-3

Li, Y., Wang, Z., Ajani, J. A., & Song, S. (2021). Drug resistance and Cancer stem cells. In *Cell Communication and Signaling* (Vol. 19, Issue 1). https://doi.org/10.1186/s12964-020-00627-5

Li, Z., Zhang, Y. Y., Zhang, H., Yang, J., Chen, Y., & Lu, H. (2022). Asymmetric Cell Division and Tumor Heterogeneity. In *Frontiers in Cell and Developmental Biology* (Vol. 10). https://doi.org/10.3389/fcell.2022.938685

Liang, J., Wen, J., Huang, Z., Chen, X., Zhang, B., & Chu, L. (2019). Small Nucleolar RNAs: Insight Into Their Function in Cancer. *Frontiers in Oncology*. https://doi.org/10.3389/fonc.2019.00587

Liberti, M. V., & Locasale, J. W. (2016). The Warburg Effect: How Does it Benefit Cancer Cells? In *Trends in Biochemical Sciences*. https://doi.org/10.1016/j.tibs.2015.12.001

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., & Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Systems*, *1*(6). https://doi.org/10.1016/j.cels.2015.12.004

Linares, J. F., Cordes, T., Duran, A., Reina-Campos, M., Valencia, T., Ahn, C. S., Castilla, E. A., Moscat, J., Metallo, C. M., & Diaz-Meco, M. T. (2017). ATF4-Induced Metabolic Reprograming Is a Synthetic Vulnerability of the p62-Deficient Tumor Stroma. *Cell Metabolism*, *26*(6). https://doi.org/10.1016/j.cmet.2017.09.001

Lindell, E., Zhong, L., & Zhang, X. (2023). Quiescent Cancer Cells—A Potential Therapeutic Target to Overcome Tumor Resistance and Relapse. In *International Journal of Molecular Sciences* (Vol. 24, Issue 4). https://doi.org/10.3390/ijms24043762

Ling, S., Hu, Z., Yang, Z., Yang, F., Li, Y., Lin, P., Chen, K., Dong, L., Cao, L., Tao, Y., Hao, L., Chen, Q., Gong, Q., Wu, D., Li, W., Zhao, W., Tian, X., Hao, C., Hungate, E. A., … Wu, C. I. (2015). Extremely high genetic diversity in a single tumor points to prevalence of non-Darwinian cell evolution. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1519556112

Liu, J., Peng, Y., & Wei, W. (2022). Cell cycle on the crossroad of tumorigenesis and cancer therapy. In *Trends in Cell Biology* (Vol. 32, Issue 1). https://doi.org/10.1016/j.tcb.2021.07.001

Liu, L., Michowski, W., Kolodziejczyk, A., & Sicinski, P. (2019). The cell cycle in stem cell proliferation, pluripotency and differentiation. In *Nature Cell Biology* (Vol. 21, Issue 9). https://doi.org/10.1038/s41556-019-0384-4

Liu, Y., Chen, S., Wang, S., Soares, F., Fischer, M., Meng, F., Du, Z., Lin, C., Meyer, C., DeCaprio, J. A., Brown, M., Liu, X. S., & He, H. H. (2017). Transcriptional landscape of the human cell cycle. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1617636114

Loftus, L. V., Amend, S. R., & Pienta, K. J. (2022). Interplay between Cell Death and Cell Proliferation Reveals New Strategies for Cancer Therapy. In *International Journal of Molecular Sciences* (Vol. 23, Issue 9). https://doi.org/10.3390/ijms23094723

Lötsch, J., Malkusch, S., & Ultsch, A. (2021). Optimal distribution-preserving downsampling of large biomedical data sets (opdisDownsampling). *PLoS ONE*, *16*(8 August). https://doi.org/10.1371/journal.pone.0255838

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12). https://doi.org/10.1186/s13059-014-0550-8

Ma, Y. S., Li, W., Liu, Y., Shi, Y., Lin, Q. L., & Fu, D. (2020). Targeting colorectal cancer stem cells as an effective treatment for colorectal cancer. In *Technology in Cancer Research and Treatment* (Vol. 19, Issue 1). https://doi.org/10.1177/1533033819892261

Ma, Y., Temkin, S. M., Hawkridge, A. M., Guo, C., Wang, W., Wang, X. Y., & Fang, X. (2018). Fatty acid oxidation: An emerging facet of metabolic transformation in cancer. In *Cancer Letters* (Vol. 435). https://doi.org/10.1016/j.canlet.2018.08.006

Ma, Y., Zhang, P., Wang, F., Yang, J., Yang, Z., & Qin, H. (2010). The relationship between early embryo development and tumourigenesis. *Journal of Cellular and Molecular Medicine*. https://doi.org/10.1111/j.1582-4934.2010.01191.x

Magee, J. A., Piskounova, E., & Morrison, S. J. (2012). Cancer Stem Cells: Impact, Heterogeneity, and Uncertainty. In *Cancer Cell* (Vol. 21, Issue 3). https://doi.org/10.1016/j.ccr.2012.03.003

Mah, A. T., Yan, K. S., & Kuo, C. J. (2016). Wnt pathway regulation of intestinal stem cells. In *Journal of Physiology* (Vol. 594, Issue 17). https://doi.org/10.1113/JP271754

Makena, M. R., Ranjan, A., Thirumala, V., & Reddy, A. P. (2020). Cancer stem cells: Road to therapeutic resistance and strategies to overcome resistance. In *Biochimica et Biophysica Acta - Molecular Basis of Disease* (Vol. 1866, Issue 4). https://doi.org/10.1016/j.bbadis.2018.11.015

Maleki, E. H., Bahrami, A. R., & Matin, M. M. (2024a). Cancer cell cycle heterogeneity as a critical determinant of therapeutic resistance. In *Genes and Diseases* (Vol. 11, Issue 1). https://doi.org/10.1016/j.gendis.2022.11.025

Maleki, E. H., Bahrami, A. R., & Matin, M. M. (2024b). Cancer cell cycle heterogeneity as a critical determinant of therapeutic resistance. In *Genes and Diseases* (Vol. 11, Issue 1). https://doi.org/10.1016/j.gendis.2022.11.025

Mao, H., Jia, J., Sheng, J., Zhang, S., Huang, K., Li, H., & He, F. (2021). Protective and anti-inflammatory role of REG1A in inflammatory bowel disease induced by JAK/STAT3 signaling axis. *International Immunopharmacology*, *92*. https://doi.org/10.1016/j.intimp.2020.107304

Marques-Ramos, A., Candeias, M. M., Menezes, J., Lacerda, R., Willcocks, M., Teixeira, A., Locker, N., & Romão, L. (2017). Cap-independent translation ensures mTOR expression and function upon protein synthesis inhibition. *RNA*, *23*(11). https://doi.org/10.1261/rna.063040.117

Martin, F. J., Amode, M. R., Aneja, A., Austine-Orimoloye, O., Azov, A. G., Barnes, I., Becker, A., Bennett, R., Berry, A., Bhai, J., Bhurji, S. K., Bignell, A., Boddu, S., Branco Lins, P. R., Brooks, L., Ramaraju, S. B., Charkhchi, M., Cockburn, A., Da Rin Fiorretto, L., … Flicek, P. (2023). Ensembl 2023. *Nucleic Acids Research*, *51*(1 D). https://doi.org/10.1093/nar/gkac958

Marusyk, A., Almendro, V., & Polyak, K. (2012). Intra-tumour heterogeneity: A looking glass for cancer? In *Nature Reviews Cancer*. https://doi.org/10.1038/nrc3261

Marusyk, A., Janiszewska, M., & Polyak, K. (2020). Intratumor Heterogeneity: The Rosetta Stone of Therapy Resistance. In *Cancer Cell*. https://doi.org/10.1016/j.ccell.2020.03.007

Marusyk, A., & Polyak, K. (2010). Tumor heterogeneity: Causes and consequences. In *Biochimica et Biophysica Acta - Reviews on Cancer*. https://doi.org/10.1016/j.bbcan.2009.11.002

Marx, V. (2023). Method of the year: long-read sequencing. In *Nature Methods* (Vol. 20, Issue 1). https://doi.org/10.1038/s41592-022-01730-w

McGranahan, N., & Swanton, C. (2017). Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. In *Cell*. https://doi.org/10.1016/j.cell.2017.01.018

McInnes, L., Healy, J., Saul, N., & Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software*. https://doi.org/10.21105/joss.00861

Meacham, C. E., & Morrison, S. J. (2013). Tumour heterogeneity and cancer cell plasticity. In *Nature*. https://doi.org/10.1038/nature12624

Mendoza, M. C., Er, E. E., & Blenis, J. (2011). The Ras-ERK and PI3K-mTOR pathways: Cross-talk and compensation. In *Trends in Biochemical Sciences* (Vol. 36, Issue 6). https://doi.org/10.1016/j.tibs.2011.03.006

Mercadante, A. A., & Kasi, A. (2021). Genetics, Cancer Cell Cycle Phases. In *StatPearls*.

Merrick, W. C. (2004). Cap-dependent and cap-independent translation in eukaryotic systems. In *Gene* (Vol. 332, Issues 1–2). https://doi.org/10.1016/j.gene.2004.02.051

Meyer, K. D., Patil, D. P., Zhou, J., Zinoviev, A., Skabkin, M. A., Elemento, O., Pestova, T. V., Qian, S. B., & Jaffrey, S. R. (2015). 5′ UTR m6A Promotes Cap-Independent Translation. *Cell*, *163*(4). https://doi.org/10.1016/j.cell.2015.10.012

Meyuhas, O., & Kahan, T. (2015). The race to decipher the top secrets of TOP mRNAs. In *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* (Vol. 1849, Issue 7). https://doi.org/10.1016/j.bbagrm.2014.08.015

Miller, K. D., Nogueira, L., Mariotto, A. B., Rowland, J. H., Yabroff, K. R., Alfano, C. M., Jemal, A., Kramer, J. L., & Siegel, R. L. (2019). Cancer treatment and survivorship statistics, 2019. *CA: A Cancer Journal for Clinicians*. https://doi.org/10.3322/caac.21565

Moody, J., Kouno, T., Chang, J. C., Ando, Y., Carninci, P., Shin, J. W., & Hon, C. C. (2022). SCAFE: a software suite for analysis of transcribed cis-regulatory elements in single cells. *Bioinformatics*, *38*(22). https://doi.org/10.1093/bioinformatics/btac644

Morandi, A., & Indraccolo, S. (2017). Linking metabolic reprogramming to therapy resistance in cancer. In *Biochimica et Biophysica Acta - Reviews on Cancer*. https://doi.org/10.1016/j.bbcan.2016.12.004

Morgan, E., Arnold, M., Gini, A., Lorenzoni, V., Cabasag, C. J., Laversanne, M., Vignat, J., Ferlay, J., Murphy, N., & Bray, F. (2023). Global burden of colorectal cancer in 2020 and 2040: Incidence and mortality estimates from GLOBOCAN. *Gut*, *72*(2). https://doi.org/10.1136/gutjnl-2022-327736

Mori, S., Nada, S., Kimura, H., Tajima, S., Takahashi, Y., Kitamura, A., Oneyama, C., & Okada, M. (2014). The mTOR pathway controls cell proliferation by regulating the FoxO3a transcription factor via SGK1 kinase. *PLoS ONE*, *9*(2). https://doi.org/10.1371/journal.pone.0088891

Mounier, C., Bouraoui, L., & Rassart, E. (2014). Lipogenesis in cancer progression (review). In *International Journal of Oncology* (Vol. 45, Issue 2). https://doi.org/10.3892/ijo.2014.2441

Mourksi, N.-E.-H., Morin, C., Fenouil, T., Diaz, J.-J., & Marcel, V. (2020). snoRNAs Offer Novel Insight and Promising Perspectives for Lung Cancer Understanding and Management. *Cells*. https://doi.org/10.3390/cells9030541

Müller, F., & Tora, L. (2014). Chromatin and DNA sequences in defining promoters for transcription initiation. In *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. https://doi.org/10.1016/j.bbagrm.2013.11.003

Mura, M., Hopkins, T. G., Michael, T., Abd-Latip, N., Weir, J., Aboagye, E., Mauri, F., Jameson, C., Sturge, J., Gabra, H., Bushell, M., Willis, A. E., Curry, E., & Blagden, S. P. (2015). LARP1 post-transcriptionally regulates mTOR and contributes to cancer progression. *Oncogene*, *34*(39). https://doi.org/10.1038/onc.2014.428

Murata, M., Nishiyori-Sueki, H., Kojima-Ishiyama, M., Carninci, P., Hayashizaki, Y., & Itoh, M. (2014). Detecting expressed genes using CAGE. *Methods in Molecular Biology*, *1164*. https://doi.org/10.1007/978-1-4939-0805-9_7

Nairuz, T., Mahmud, Z., Manik, R. K., & Kabir, Y. (2023). Cancer stem cells: an insight into the development of metastatic tumors and therapy resistance. In *Stem Cell Reviews and Reports* (Vol. 19, Issue 6). https://doi.org/10.1007/s12015-023-10529-x

Najafi, M., Farhood, B., Mortezaee, K., Kharazinejad, E., Majidpoor, J., & Ahadi, R. (2020). Hypoxia in solid tumors: a key promoter of cancer stem cell (CSC) resistance. In *Journal of Cancer Research and Clinical Oncology* (Vol. 146, Issue 1). https://doi.org/10.1007/s00432-019-03080-1

Nandagopal, N., & Roux, P. P. (2015). Regulation of global and specific mRNA translation by the mTOR signaling pathway. *Translation*, *3*(1). https://doi.org/10.4161/21690731.2014.983402

Nath, A., Cosgrove, P. A., Mirsafian, H., Christie, E. L., Pflieger, L., Copeland, B., Majumdar, S., Cristea, M. C., Han, E. S., Lee, S. J., Wang, E. W., Fereday, S., Traficante, N., Salgia, R., Werner, T., Cohen, A. L., Moos, P., Chang, J. T., Bowtell, D. D. L., & Bild, A. H. (2021). Evolution of core archetypal phenotypes in progressive high grade serous ovarian cancer. *Nature Communications*. https://doi.org/10.1038/s41467-021-23171-3

Navarro, C., Ortega, Á., Santeliz, R., Garrido, B., Chacín, M., Galban, N., Vera, I., De Sanctis, J. B., & Bermúdez, V. (2022). Metabolic Reprogramming in Cancer Cells: Emerging Molecular Mechanisms and Novel Therapeutic Approaches. In *Pharmaceutics* (Vol. 14, Issue 6). https://doi.org/10.3390/pharmaceutics14061303

Naz, F., Shi, M., Sajid, S., Yang, Z., & Yu, C. (2021). Cancer stem cells: a major culprit of intra-tumor heterogeneity. *American Journal of Cancer Research*, *11*(12).

Nenkov, M., Ma, Y., Gaßler, N., & Chen, Y. (2021). Metabolic reprogramming of colorectal cancer cells and the microenvironment: Implication for therapy. In *International Journal of Molecular Sciences* (Vol. 22, Issue 12). https://doi.org/10.3390/ijms22126262

Nepal, C., & Andersen, J. B. (2023). Alternative promoters in CpG depleted regions are prevalently associated with epigenetic misregulation of liver cancer transcriptomes. *Nature Communications*, *14*(1). https://doi.org/10.1038/s41467-023-38272-4

Nepal, C., Hadzhiev, Y., Balwierz, P., Tarifeño-Saldivia, E., Cardenas, R., Wragg, J. W., Suzuki, A.-M., Carninci, P., Peers, B., Lenhard, B., Andersen, J. B., & Müller, F. (2020). Dual-initiation promoters with intertwined canonical and TCT/TOP transcription start sites diversify transcript processing. *Nature Communications*, *11*(1), 168. https://doi.org/10.1038/s41467-019-13687-0

Nepal, C., Hadzhiev, Y., Previti, C., Haberle, V., Li, N., Takahashi, H., Suzuki, A. M. M., Sheng, Y., Abdelhamid, R. F., Anand, S., Gehrig, J., Akalin, A., Kockx, C. E. M., Van Der Sloot, A. A. J., Van IJcken, W. F. J., Armant, O., Rastegar, S., Watson, C., Strahle, U., … Muller, F. (2013). Dynamic regulation of the transcription initiation landscape at single nucleotide resolution during vertebrate embryogenesis. *Genome Research*. https://doi.org/10.1101/gr.153692.112

Newman, R. H., & Zhang, J. (2008). Fucci: Street Lights on the Road to Mitosis. In *Chemistry and Biology*. https://doi.org/10.1016/j.chembiol.2008.02.003

Nong, S., Han, X., Xiang, Y., Qian, Y., Wei, Y., Zhang, T., Tian, K., Shen, K., Yang, J., & Ma, X. (2023). Metabolic reprogramming in cancer: Mechanisms and therapeutics. In *MedComm* (Vol. 4, Issue 2). https://doi.org/10.1002/mco2.218

Nosrati, N., Kapoor, N. R., & Kumar, V. (2014). Combinatorial action of transcription factors orchestrates cell cycle-dependent expression of the ribosomal protein genes and ribosome biogenesis. *FEBS Journal*, *281*(10). https://doi.org/10.1111/febs.12786

Ogami, K., Oishi, Y., Sakamoto, K., Okumura, M., Yamagishi, R., Inoue, T., Hibino, M., Nogimori, T., Yamaguchi, N., Furutachi, K., Hosoda, N., Inagaki, H., & Hoshino, S. ichi. (2022). mTOR- and LARP1-dependent regulation of TOP mRNA poly(A) tail and ribosome loading. *Cell Reports*, *41*(4). https://doi.org/10.1016/j.celrep.2022.111548

Ohshima, K., & Morii, E. (2021). Metabolic reprogramming of cancer cells during tumor progression and metastasis. In *Metabolites*. https://doi.org/10.3390/metabo11010028

Oksdath Mansilla, M., Salazar-Hernandez, C., Perrin, S. L., Scheer, K. G., Cildir, G., Toubia, J., Sedivakova, K., Tea, M. N., Lenin, S., Ponthier, E., Yeo, E. C. F., Tergaonkar, V., Poonnoose, S., Ormsby, R. J., Pitson, S. M., Brown, M. P., Ebert, L. M., & Gomez, G. A. (2021). 3D-printed microplate inserts for long term high-resolution imaging of live brain organoids. *BMC Biomedical Engineering*. https://doi.org/10.1186/s42490-021-00049-5

Osorio, D., & Cai, J. J. (2021). Systematic determination of the mitochondrial proportion in human and mice tissues for single-cell RNA-sequencing data quality control. *Bioinformatics*, *37*(7). https://doi.org/10.1093/bioinformatics/btaa751

Panigrahi, A., & O'Malley, B. W. (2021). Mechanisms of enhancer action: the known and the unknown. In *Genome Biology* (Vol. 22, Issue 1). https://doi.org/10.1186/s13059-021-02322-1

Papadaki, S., & Magklara, A. (2022). Regulation of Metabolic Plasticity in Cancer Stem Cells and Implications in Cancer Therapy. In *Cancers* (Vol. 14, Issue 23). https://doi.org/10.3390/cancers14235912

Parry, T. J., Theisen, J. W. M., Hsu, J. Y., Wang, Y. L., Corcoran, D. L., Eustice, M., Ohler, U., & Kadonaga, J. T. (2010). The TCT motif, a key component of an RNA polymerase II transcription system for the translational machinery. *Genes and Development*, *24*(18). https://doi.org/10.1101/gad.1951110

Patel, A. B., Greber, B. J., & Nogales, E. (2020). Recent insights into the structure of TFIID, its assembly, and its binding to core promoter. In *Current Opinion in Structural Biology* (Vol. 61). https://doi.org/10.1016/j.sbi.2019.10.001

Patra, D., Bhavya, K., Ramprasad, P., Kalia, M., & Pal, D. (2023). *Anti-cancer drug molecules targeting cancer cell cycle and proliferation* (pp. 343–395). https://doi.org/10.1016/bs.apcsb.2022.11.011

Pérez-González, A., Bévant, K., & Blanpain, C. (2023). Cancer cell plasticity during tumor progression, metastasis and response to therapy. In *Nature Cancer* (Vol. 4, Issue 8). https://doi.org/10.1038/s43018-023-00595-y

Phan, L. M., Yeung, S. C. J., & Lee, M. H. (2014). Cancer metabolic reprogramming: importance, main features, and potentials for precise targeted anti-cancer therapies. In *Cancer Biology and Medicine* (Vol. 11, Issue 1). https://doi.org/10.7497/j.issn.2095-3941.2014.01.001

Philippe, L., van den Elzen, A. M. G., Watson, M. J., & Thoreen, C. C. (2020). Global analysis of LARP1 translation targets reveals tunable and dynamic features of 5′ TOP motifs. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1912864117

Philippe, L., Vasseur, J. J., Debart, F., & Thoreen, C. C. (2018). La-related protein 1 (LARP1) repression of TOP mRNA translation is mediated through its cap-binding domain and controlled by an adjacent regulatory region. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkx1237

Plutynski, A. (2021). Testing Multi-Task Cancer Evolution: How Do We Test Ecological Hypotheses in Cancer? *Frontiers in Ecology and Evolution*. https://doi.org/10.3389/fevo.2021.666262

Poulain, S., Kato, S., Arnaud, O., Morlighem, J. É., Suzuki, M., Plessy, C., & Harbers, M. (2017). NanoCAGE: A method for the analysis of coding and noncoding 5′-capped transcriptomes. In *Methods in Molecular Biology* (Vol. 1543). https://doi.org/10.1007/978-1-4939-6716-2_4

Pountain, A. W., Jiang, P., Yao, T., Homaee, E., Guan, Y., McDonald, K. J. C., Podkowik, M., Shopsin, B., Torres, V. J., Golding, I., & Yanai, I. (2024). Transcription–replication interactions reveal bacterial genome regulation. *Nature*, *626*(7999), 661–669. https://doi.org/10.1038/s41586-023-06974-w

Pranzini, E., Pardella, E., Paoli, P., Fendt, S. M., & Taddei, M. L. (2021). Metabolic Reprogramming in Anticancer Drug Resistance: A Focus on Amino Acids. In *Trends in Cancer* (Vol. 7, Issue 8). https://doi.org/10.1016/j.trecan.2021.02.004

Prasetyanti, P. R., & Medema, J. P. (2017). Intra-tumor heterogeneity from a cancer stem cell perspective. In *Molecular Cancer*. https://doi.org/10.1186/s12943-017-0600-4

Prieto-Vila, M., Takahashi, R. U., Usuba, W., Kohama, I., & Ochiya, T. (2017). Drug resistance driven by cancer stem cells and their niche. In *International Journal of Molecular Sciences* (Vol. 18, Issue 12). https://doi.org/10.3390/ijms18122574

Punt, C. J. A., Koopman, M., & Vermeulen, L. (2017). From tumour heterogeneity to advances in precision treatment of colorectal cancer. In *Nature Reviews Clinical Oncology*. https://doi.org/10.1038/nrclinonc.2016.171

Qin, S., Jiang, J., Lu, Y., Nice, E. C., Huang, C., Zhang, J., & He, W. (2020). Emerging role of tumor cell plasticity in modifying therapeutic response. In *Signal Transduction and Targeted Therapy* (Vol. 5, Issue 1). https://doi.org/10.1038/s41392-020-00313-5

Qin, X., Jiang, B., & Zhang, Y. (2016). 4E-BP1, a multifactor regulated multifunctional protein. In *Cell Cycle* (Vol. 15, Issue 6). https://doi.org/10.1080/15384101.2016.1151581

Ramm, S., Vary, R., Gulati, T., Luu, J., Cowley, K. J., Janes, M. S., Radio, N., & Simpson, K. J. (2022). High-Throughput Live and Fixed Cell Imaging Method to Screen Matrigel-Embedded Organoids. *Organoids*, *2*(1). https://doi.org/10.3390/organoids2010001

Ramón y Cajal, S., Sesé, M., Capdevila, C., Aasen, T., De Mattos-Arruda, L., Diaz-Cano, S. J., Hernández-Losa, J., & Castellví, J. (2020). Clinical implications of intratumor heterogeneity: challenges and opportunities. In *Journal of Molecular Medicine*. https://doi.org/10.1007/s00109-020-01874-2

Reese, F., Williams, B., Balderrama-Gutierrez, G., Wyman, D., Hasan Çelik, M., Rebboah, E., Rezaie, N., Trout, D., Razavi-Mohseni, M., Jiang, Y., Borsari, B., Morabito, S., Liang, H. Y., Mcgill, C. J., Rahmanian, S., Sakr, J., Jiang, S., Zeng, W., Carvalho, K.-B., … Mortazavi, A. (2023). The ENCODE4 long-read RNA-seq collection reveals distinct classes of transcript structure diversity. *Laura Reinholdt*, *18*.

Rejali, L., Seifollahi Asl, R., Sanjabi, F., Fatemi, N., Asadzadeh Aghdaei, H., Saeedi Niasar, M., Ketabi Moghadam, P., Nazemalhosseini Mojarad, E., Mini, E., & Nobili, S. (2023). Principles of Molecular Utility for CMS Classification in Colorectal Cancer Management. In *Cancers* (Vol. 15, Issue 10). https://doi.org/10.3390/cancers15102746

Reynisson, J., Jaiswal, J. K., Barker, D., D'mello, S. A. N., Denny, W. A., Baguley, B. C., & Leung, E. Y. (2016). Evidence that phospholipase C is involved in the antitumour action of NSC768313, a new thieno[2,3-b]pyridine derivative. *Cancer Cell International*, *16*(1). https://doi.org/10.1186/s12935-016-0293-6

Riba, A., Oravecz, A., Durik, M., Jiménez, S., Alunni, V., Cerciat, M., Jung, M., Keime, C., Keyes, W. M., & Molina, N. (2022). Cell cycle gene regulation dynamics revealed by RNA velocity and deep-learning. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-30545-8

Richter, J. D., & Sonenberg, N. (2005). Regulation of cap-dependent translation by eIF4E inhibitory proteins. *Nature*, *433*(7025). https://doi.org/10.1038/nature03205

Ricoult, S. J. H., Yecies, J. L., Ben-Sahra, I., & Manning, B. D. (2016). Oncogenic PI3K and K-Ras stimulate de novo lipid synthesis through mTORC1 and SREBP. *Oncogene*, *35*(10). https://doi.org/10.1038/onc.2015.179

Rios, A. C., & Clevers, H. (2018). Imaging organoids: A bright future ahead. In *Nature Methods*. https://doi.org/10.1038/nmeth.4537

Rodriguez-Esteban, R., & Jiang, X. (2017). Differential gene expression in disease: A comparison between high-throughput studies and the literature. *BMC Medical Genomics*. https://doi.org/10.1186/s12920-017-0293-y

Roerink, S. F., Sasaki, N., Lee-Six, H., Young, M. D., Alexandrov, L. B., Behjati, S., Mitchell, T. J., Grossmann, S., Lightfoot, H., Egan, D. A., Pronk, A., Smakman, N., Van Gorp, J., Anderson, E., Gamble, S. J., Alder, C., Van De Wetering, M., Campbell, P. J., Stratton, M. R., & Clevers, H. (2018). Intra-tumour diversification in colorectal cancer at the single-cell level. *Nature*. https://doi.org/10.1038/s41586-018-0024-3

Roux, P. P., & Topisirovic, I. (2012). Regulation of mRNA translation by signaling pathways. *Cold Spring Harbor Perspectives in Biology*, *4*(11). https://doi.org/10.1101/cshperspect.a012252

Ruiz, S., Panopoulos, A. D., Herrerías, A., Bissig, K. D., Lutz, M., Berggren, W. T., Verma, I. M., & Izpisua Belmonte, J. C. (2011). A high proliferation rate is required for cell reprogramming and maintenance of human embryonic stem cell identity. *Current Biology*, *21*(1). https://doi.org/10.1016/j.cub.2010.11.049

Ryl, T., Kuchen, E. E., Bell, E., Shao, C., Flórez, A. F., Mönke, G., Gogolin, S., Friedrich, M., Lamprecht, F., Westermann, F., & Höfer, T. (2017). Cell-Cycle Position of Single MYC-Driven Cancer Cells Dictates Their Susceptibility to a Chemotherapeutic Drug. *Cell Systems*, *5*(3). https://doi.org/10.1016/j.cels.2017.07.005

Rysman, E., Brusselmans, K., Scheys, K., Timmermans, L., Derua, R., Munck, S., Van Veldhoven, P. P., Waltregny, D., Daniëls, V. W., Machiels, J., Vanderhoydonc, F., Smans, K., Waelkens, E., Verhoeven, G., & Swinnen, J. V. (2010). De novo lipogenesis protects cancer cells from free radicals and chemotherapeutics by promoting membrane lipid saturation. *Cancer Research*, *70*(20). https://doi.org/10.1158/0008-5472.CAN-09-3871

Salimullah, M., Mizuho, S., Plessy, C., & Carninci, P. (2011). NanoCAGE: A high-resolution technique to discover and interrogate cell transcriptomes. *Cold Spring Harbor Protocols*, *6*(1). https://doi.org/10.1101/pdb.prot5559

Salucci, S., Aramini, B., Bartoletti-Stella, A., Versari, I., Martinelli, G., Blalock, W., Stella, F., & Faenza, I. (2023). Phospholipase Family Enzymes in Lung Cancer: Looking for Novel Therapeutic Approaches. In *Cancers* (Vol. 15, Issue 12). https://doi.org/10.3390/cancers15123245

Sasaki, N., & Clevers, H. (2018). Studying cellular heterogeneity and drug sensitivity in colorectal cancer using organoid technology. In *Current Opinion in Genetics and Development*. https://doi.org/10.1016/j.gde.2018.09.001

Sayed, I. M., El-Hafeez, A. A. A., Maity, P. P., Das, S., & Ghosh, P. (2021). Modeling colorectal cancers using multidimensional organoids. In *Advances in Cancer Research* (Vol. 151). https://doi.org/10.1016/bs.acr.2021.02.005

Sazonova, E. V., Yapryntseva, M. A., Pervushin, N. V., Tsvetcov, R. I., Zhivotovsky, B., & Kopeina, G. S. (2024). Cancer Drug Resistance: Targeting Proliferation or Programmed Cell Death. *Cells*, *13*(5), 388. https://doi.org/10.3390/cells13050388

Schaefer, C. F., Anthony, K., Krupa, S., Buchoff, J., Day, M., Hannay, T., & Buetow, K. H. (2009). PID: The pathway interaction database. *Nucleic Acids Research*, *37*(SUPPL. 1). https://doi.org/10.1093/nar/gkn653

Schiliro, C., & Firestein, B. L. (2021). Mechanisms of metabolic reprogramming in cancer cells supporting enhanced growth and proliferation. In *Cells* (Vol. 10, Issue 5). https://doi.org/10.3390/cells10051056

Schmitt, C. A., Wang, B., & Demaria, M. (2022). Senescence and cancer — role and therapeutic opportunities. In *Nature Reviews Clinical Oncology* (Vol. 19, Issue 10). https://doi.org/10.1038/s41571-022-00668-4

Schneider, C., Erhard, F., Binotti, B., Buchberger, A., Vogel, J., & Fischer, U. (2022). An unusual mode of baseline translation adjusts cellular protein synthesis capacity to metabolic needs. *Cell Reports*, *41*(2). https://doi.org/10.1016/j.celrep.2022.111467

Schoenfelder, S., & Fraser, P. (2019). Long-range enhancer–promoter contacts in gene expression control. In *Nature Reviews Genetics* (Vol. 20, Issue 8). https://doi.org/10.1038/s41576-019-0128-0

Schwenzer, H., Abdel Mouti, M., Neubert, P., Morris, J., Stockton, J., Bonham, S., Fellermeyer, M., Chettle, J., Fischer, R., Beggs, A. D., & Blagden, S. P. (2021). LARP1 isoform expression in human cancer cell lines. *RNA Biology*, *18*(2). https://doi.org/10.1080/15476286.2020.1744320

Scott, K. F., Sajinovic, M., Hein, J., Nixdorf, S., Galettis, P., Liauw, W., de Souza, P., Dong, Q., Graham, G. G., & Russell, P. J. (2010). Emerging roles for phospholipase A2 enzymes in cancer. In *Biochimie* (Vol. 92, Issue 6). https://doi.org/10.1016/j.biochi.2010.03.019

Scott, M. S., & Ono, M. (2011). From snoRNA to miRNA: Dual function regulatory non-coding RNAs. In *Biochimie*. https://doi.org/10.1016/j.biochi.2011.05.026

Serra, D., Mayr, U., Boni, A., Lukonin, I., Rempfler, M., Challet Meylan, L., Stadler, M. B., Strnad, P., Papasaikas, P., Vischi, D., Waldt, A., Roma, G., & Liberali, P. (2019). Self-organization and symmetry breaking in intestinal organoid development. *Nature*. https://doi.org/10.1038/s41586-019-1146-y

Sever, R., & Brugge, J. S. (2015). Signal transduction in cancer. *Cold Spring Harbor Perspectives in Medicine*, *5*(4). https://doi.org/10.1101/cshperspect.a006098

Shackleton, M., Quintana, E., Fearon, E. R., & Morrison, S. J. (2009). Heterogeneity in Cancer: Cancer Stem Cells versus Clonal Evolution. In *Cell*. https://doi.org/10.1016/j.cell.2009.08.017

Shen, L. (2009). Functional morphology of the gastrointestinal tract. In *Current Topics in Microbiology and Immunology* (Vol. 337, Issue 1). https://doi.org/10.1007/978-3-642-01846-6_1

Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T., Fukuda, S., Sasaki, D., Podhajska, A., Harbers, M., Kawai, J., Carninci, P., & Hayashizaki, Y. (2003). Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.2136655100

Shyh-Chang, N., Daley, G. Q., & Cantley, L. C. (2013). Stem cell metabolism in tissue development and aging. In *Development (Cambridge)* (Vol. 140, Issue 12). https://doi.org/10.1242/dev.091777

Sievert, C. (2020). Interactive Web-Based Data Visualization with R, plotly, and shiny. In *Interactive Web-Based Data Visualization with R, plotly, and shiny*. https://doi.org/10.1201/9780429447273

Singer, G. A. C., Wu, J., Yan, P., Plass, C., Huang, T. H. M., & Davuluri, R. V. (2008). Genome-wide analysis of alternative promoters of human genes using a custom promoter tiling array. *BMC Genomics*, *9*. https://doi.org/10.1186/1471-2164-9-349

Singh, A. M., Trost, R., Boward, B., & Dalton, S. (2016). Utilizing FUCCI reporters to understand pluripotent stem cell biology. *Methods*. https://doi.org/10.1016/j.ymeth.2015.09.020

Singh, K. P., Miaskowski, C., Dhruva, A. A., Flowers, E., & Kober, K. M. (2018). Mechanisms and Measurement of Changes in Gene Expression. *Biological Research for Nursing*, *20*(4). https://doi.org/10.1177/1099800418772161

Singh, M. P., Rai, S., Pandey, A., Singh, N. K., & Srivastava, S. (2021). Molecular subtypes of colorectal cancer: An emerging therapeutic opportunity for personalized medicine. In *Genes and Diseases* (Vol. 8, Issue 2). https://doi.org/10.1016/j.gendis.2019.10.013

Slomovic, S., Laufer, D., Geiger, D., & Schuster, G. (2006). Polyadenylation of ribosomal RNA in human cells. *Nucleic Acids Research*, *34*(10). https://doi.org/10.1093/nar/gkl357

Sperling, S. (2007). Transcriptional regulation at a glance. In *BMC Bioinformatics* (Vol. 8, Issue SUPPL. 6). https://doi.org/10.1186/1471-2105-8-S6-S2

Stadtfeld, M., Varas, F., & Graf, T. (2005). Fluorescent protein-cell labeling and its application in time-lapse analysis of hematopoietic differentiation. In *Methods in molecular medicine* (Vol. 105). https://doi.org/10.1385/1-59259-826-9:395

Stanta, G., & Bonin, S. (2018a). Overview on clinical relevance of intra-tumor heterogeneity. In *Frontiers in Medicine*. https://doi.org/10.3389/fmed.2018.00085

Stanta, G., & Bonin, S. (2018b). Overview on clinical relevance of intra-tumor heterogeneity. In *Frontiers in Medicine*. https://doi.org/10.3389/fmed.2018.00085

Stavraka, C., & Blagden, S. (2015). The La-related proteins, a family with connections to cancer. In *Biomolecules* (Vol. 5, Issue 4). https://doi.org/10.3390/biom5042701

Strausberg, R. L., & Levy, S. (2007). Promoting transcriptome diversity. In *Genome Research* (Vol. 17, Issue 7). https://doi.org/10.1101/gr.6499807

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P., & Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell*, *177*(7). https://doi.org/10.1016/j.cell.2019.05.031

Su, H., Xu, T., Ganapathy, S., Shadfan, M., Long, M., Huang, T. H. M., Thompson, I., & Yuan, Z. M. (2014). Elevated snoRNA biogenesis is essential in breast cancer. *Oncogene*. https://doi.org/10.1038/onc.2013.89

Subramanian, A., Alperovich, M., Yang, Y., & Li, B. (2022). Biology-inspired data-driven quality control for scientific discovery in single-cell transcriptomics. *Genome Biology*, *23*(1). https://doi.org/10.1186/s13059-022-02820-w

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(43). https://doi.org/10.1073/pnas.0506580102

Sumigray, K. D., Terwilliger, M., & Lechler, T. (2018). Morphogenesis and Compartmentalization of the Intestinal Crypt. *Developmental Cell*, *45*(2). https://doi.org/10.1016/j.devcel.2018.03.024

Sun, X. X., & Yu, Q. (2015). Intra-tumor heterogeneity of cancer cells and its implications for cancer treatment. In *Acta Pharmacologica Sinica* (Vol. 36, Issue 10). https://doi.org/10.1038/aps.2015.92

Sun, Z., Fan, J., Dang, Y., & Zhao, Y. (2023). Enhancer in cancer pathogenesis and treatment. In *Genetics and Molecular Biology* (Vol. 46, Issue 3). https://doi.org/10.1590/1678-4685-GMB-2022-0313

Takahashi, H., Kato, S., Murata, M., & Carninci, P. (2012). CAGE (Cap analysis of gene expression): A protocol for the detection of promoter and transcriptional networks. *Methods in Molecular Biology*. https://doi.org/10.1007/978-1-61779-292-2_11

Takahashi, H., Lassmann, T., Murata, M., & Carninci, P. (2012). 5' end-centered expression profiling using cap-analysis gene expression and next-generation sequencing. *Nature Protocols*. https://doi.org/10.1038/nprot.2012.005

Tamarkin-Ben-Harush, A., Vasseur, J. J., Debart, F., Ulitsky, I., & Dikstein, R. (2017). Cap-proximal nucleotides via differential eIF4E binding and alternative promoter usage mediate translational response to energy stress. *ELife*. https://doi.org/10.7554/eLife.21907

Tang, D. T. P., Plessy, C., Salimullah, M., Suzuki, A. M., Calligaris, R., Gustincich, S., & Carninci, P. (2013). Suppression of artifacts and barcode bias in high-throughput transcriptome analyses utilizing template switching. *Nucleic Acids Research*, *41*(3). https://doi.org/10.1093/nar/gks1128

Thoreen, C. C., Chantranupong, L., Keys, H. R., Wang, T., Gray, N. S., & Sabatini, D. M. (2012). A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature*. https://doi.org/10.1038/nature11083

Tian, T., Li, X., & Zhang, J. (2019). mTOR signaling in cancer and mtor inhibitors in solid tumor targeting therapy. In *International Journal of Molecular Sciences* (Vol. 20, Issue 3). https://doi.org/10.3390/ijms20030755

Tian, X., Zhang, S., Zhou, L., Seyhan, A. A., Hernandez Borrero, L., Zhang, Y., & El-Deiry, W. S. (2021). Targeting the Integrated Stress Response in Cancer Therapy. In *Frontiers in Pharmacology* (Vol. 12). https://doi.org/10.3389/fphar.2021.747837

Tirosh, I., Barkai, N., & Verstrepen, K. J. (2009). Promoter architecture and the evolvability of gene expression. In *Journal of Biology* (Vol. 8, Issue 11). https://doi.org/10.1186/jbiol204

Torborg, S. R., Li, Z., Chan, J. E., & Tammela, T. (2022). Cellular and molecular mechanisms of plasticity in cancer. In *Trends in Cancer* (Vol. 8, Issue 9). https://doi.org/10.1016/j.trecan.2022.04.007

Torrence, M. E., Macarthur, M. R., Hosios, A. M., Valvezan, A. J., Asara, J. M., Mitchell, J. R., & Manning, B. D. (2021). The mtorc1-mediated activation of atf4 promotes protein and glutathione synthesis downstream of growth signals. *ELife*, *10*. https://doi.org/10.7554/eLife.63326

Trapnell, C. (2015). Defining cell types and states with single-cell genomics. In *Genome Research* (Vol. 25, Issue 10). https://doi.org/10.1101/gr.190595.115

Tzamali, E., Grekas, G., Marias, K., & Sakkalis, V. (2014). Exploring the competition between proliferative and invasive cancer phenotypes in a continuous spatial model. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0103191

Uniacke, J., Holterman, C. E., Lachance, G., Franovic, A., Jacob, M. D., Fabian, M. R., Payette, J., Holcik, M., Pause, A., & Lee, S. (2012). An oxygen-regulated switch in the protein synthesis machinery. *Nature*. https://doi.org/10.1038/nature11055

Vahid, M. R., Brown, E. L., Steen, C. B., Zhang, W., Jeon, H. S., Kang, M., Gentles, A. J., & Newman, A. M. (2023). High-resolution alignment of single-cell and spatial transcriptomes with CytoSPACE. *Nature Biotechnology*, *41*(11). https://doi.org/10.1038/s41587-023-01697-9

Valcárcel, L. V., Amundarain, A., Kulis, M., Charalampopoulou, S., Melnick, A., San Miguel, J., Martín-Subero, J. I., Planes, F. J., Agirre, X., & Prosper, F. (2021). Gene expression derived from alternative promoters improves prognostic stratification in multiple myeloma. *Leukemia*, *35*(10). https://doi.org/10.1038/s41375-021-01263-9

Van Den Beucken, T., Magagnin, M. G., Jutten, B., Seigneuric, R., Lambin, P., Koritzinsky, M., & Wouters, B. G. (2011). Translational control is a major contributor to hypoxia induced gene expression. *Radiotherapy and Oncology*. https://doi.org/10.1016/j.radonc.2011.05.058

Van Den Brink, S. C., Sage, F., Vértesy, Á., Spanjaard, B., Peterson-Maduro, J., Baron, C. S., Robin, C., & Van Oudenaarden, A. (2017). Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. In *Nature Methods* (Vol. 14, Issue 10). https://doi.org/10.1038/nmeth.4437

Varghese, E., Samuel, S. M., Líšková, A., Samec, M., Kubatka, P., & Büsselberg, D. (2020). Targeting glucose metabolism to overcome resistance to anticancer chemotherapy in breast cancer. In *Cancers* (Vol. 12, Issue 8). https://doi.org/10.3390/cancers12082252

Vaysse, C., Philippe, C., Martineau, Y., Quelen, C., Hieblot, C., Renaud, C., Nicaise, Y., Desquesnes, A., Pannese, M., Filleron, T., Escourrou, G., Lawson, M., Rintoul, R. C., Delisle, M. B., Pyronnet, S., Brousset, P., Prats, H., & Touriol, C. (2015). Key contribution of eIF4H-mediated translational control in tumor promotion. *Oncotarget*, *6*(37). https://doi.org/10.18632/oncotarget.5442

Vermeulen, L., De Sousa E Melo, F., Van Der Heijden, M., Cameron, K., De Jong, J. H., Borovski, T., Tuynman, J. B., Todaro, M., Merz, C., Rodermond, H., Sprick, M. R., Kemper, K., Richel, D. J., Stassi, G., & Medema, J. P. (2010). Wnt activity defines colon cancer stem cells and is regulated by the microenvironment. *Nature Cell Biology*. https://doi.org/10.1038/ncb2048

Vettore, L., Westbrook, R. L., & Tennant, D. A. (2020). New aspects of amino acid metabolism in cancer. In *British Journal of Cancer* (Vol. 122, Issue 2). https://doi.org/10.1038/s41416-019-0620-5

Vihervaara, A., Mahat, D. B., Guertin, M. J., Chu, T., Danko, C. G., Lis, J. T., & Sistonen, L. (2017). Transcriptional response to stress is pre-wired by promoter and enhancer architecture. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-00151-0

Villard, J. (2004). Transcription regulation and human diseases. In *Swiss Medical Weekly* (Vol. 134, Issues 39–40). https://doi.org/10.4414/smw.2004.10191

Viner-Breuer, R., Yilmaz, A., Benvenisty, N., & Goldberg, M. (2019). The essentiality landscape of cell cycle related genes in human pluripotent and cancer cells. *Cell Division*, *14*(1). https://doi.org/10.1186/s13008-019-0058-4

Waclaw, B., Bozic, I., Pittman, M. E., Hruban, R. H., Vogelstein, B., & Nowak, M. A. (2015). A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. *Nature*. https://doi.org/10.1038/nature14971

Wahl, G. M., & Spike, B. T. (2017). Cell state plasticity, stem cells, EMT, and the generation of intra-tumoral heterogeneity. In *npj Breast Cancer* (Vol. 3, Issue 1). https://doi.org/10.1038/s41523-017-0012-z

Walter, R. J., Sonnentag, S. J., Munoz-Sagredo, L., Merkel, M., Richert, L., Bunert, F., Heneka, Y. M., Loustau, T., Hodder, M., Ridgway, R. A., Sansom, O. J., Mely, Y., Rothbauer, U., Schmitt, M., & Orian-Rousseau, V. (2022). Wnt signaling is boosted during intestinal regeneration by a CD44-positive feedback loop. *Cell Death and Disease*, *13*(2). https://doi.org/10.1038/s41419-022-04607-0

Walters, B., & Thompson, S. R. (2016). Cap-independent translational control of carcinogenesis. In *Frontiers in Oncology* (Vol. 6, Issue MAY). https://doi.org/10.3389/fonc.2016.00128

Wang, J., Liu, Q., & Shyr, Y. (2015). Dysregulated transcription across diverse cancer types reveals the importance of RNA-binding protein in carcinogenesis. *BMC Genomics*, *16*(7). https://doi.org/10.1186/1471-2164-16-S7-S5

Wang, J., Rojas, P., Mao, J., Mustè Sadurnì, M., Garnier, O., Xiao, S., Higgs, M. R., Garcia, P., & Saponaro, M. (2021). Persistence of RNA transcription during DNA replication delays duplication of transcription start sites until G2/M. *Cell Reports*, *34*(7). https://doi.org/10.1016/j.celrep.2021.108759

Wang, X., & Proud, C. G. (2006). The mTOR pathway in the control of protein synthesis. In *Physiology* (Vol. 21, Issue 5). https://doi.org/10.1152/physiol.00024.2006

Wang, X., You, X., Langer, J. D., Hou, J., Rupprecht, F., Vlatkovic, I., Quedenau, C., Tushev, G., Epstein, I., Schaefke, B., Sun, W., Fang, L., Li, G., Hu, Y., Schuman, E. M., & Chen, W. (2019). Full-length transcriptome reconstruction reveals a large diversity of RNA and protein isoforms in rat hippocampus. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-13037-0

Wang, Y. L., Duttke, S. H. C., Chen, K., Johnston, J., Kassavetis, G. A., Zeitlinger, J., & Kadonaga, J. T. (2014). TRF2, but not TBP, mediates the transcription of ribosomal protein genes. *Genes and Development*, *28*(14). https://doi.org/10.1101/gad.245662.114

Wang, Y., Yang, W., Liu, T., Bai, G., Liu, M., & Wang, W. (2019). Over-expression of SOX8 predicts poor prognosis in colorectal cancer: A retrospective study. *Medicine*, *98*(27). https://doi.org/10.1097/MD.0000000000016237

Wanigasooriya, K., Barros-Silva, J. D., Tee, L., El-asrag, M. E., Stodolna, A., Pickles, O. J., Stockton, J., Bryer, C., Hoare, R., Whalley, C. M., Tyler, R., Sillo, T., Yau, C., Ismail, T., & Beggs, A. D. (2022). Patient Derived Organoids Confirm That PI3K/AKT Signalling Is an Escape Pathway for Radioresistance and a Target for Therapy in Rectal Cancer. *Frontiers in Oncology*, *12*. https://doi.org/10.3389/fonc.2022.920444

Weber, R., Ghoshdastider, U., Spies, D., Duré, C., Valdivia-Francia, F., Forny, M., Ormiston, M., Renz, P. F., Taborsky, D., Yigit, M., Bernasconi, M., Yamahachi, H., & Sendoel, A. (2023). Monitoring the 5'UTR landscape reveals isoform switches to drive translational efficiencies in cancer. *Oncogene*, *42*(9). https://doi.org/10.1038/s41388-022-02578-2

Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y. J., & Zemla, J. (2017). Visualization of a Correlation Matrix. R package "corrplot". *Statistician*, *56*.

Wen, G. M., Xu, X. Y., & Xia, P. (2022). Metabolism in Cancer Stem Cells: Targets for Clinical Treatment. In *Cells* (Vol. 11, Issue 23). https://doi.org/10.3390/cells11233790

Wen, Y. A., Xiong, X., Zaytseva, Y. Y., Napier, D. L., Vallee, E., Li, A. T., Wang, C., Weiss, H. L., Evers, B. M., & Gao, T. (2018). Downregulation of SREBP inhibits tumor growth and initiation by altering cellular metabolism in colon cancer article. *Cell Death and Disease*, *9*(3). https://doi.org/10.1038/s41419-018-0330-6

Whitfield, M. L., Sherlock, G., Saldanha, A. J., Murray, J. I., Ball, C. A., Alexander, K. E., Matese, J. C., Perou, C. M., Hurt, M. M., Brown, P. O., & Botstein, D. (2002). Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Molecular Biology of the Cell*, *13*(6). https://doi.org/10.1091/mbc.02-02-0030.

Wiecek, A. J., Cutty, S. J., Kornai, D., Parreno-Centeno, M., Gourmet, L. E., Tagliazucchi, G. M., Jacobson, D. H., Zhang, P., Xiong, L., Bond, G. L., Barr, A. R., & Secrier, M. (2023). Genomic hallmarks and therapeutic implications of G0 cell cycle arrest in cancer. *Genome Biology*, *24*(1). https://doi.org/10.1186/s13059-023-02963-4

Wilbertz, J. H., Voigt, F., Horvathova, I., Roth, G., Zhan, Y., & Chao, J. A. (2019). Single-Molecule Imaging of mRNA Localization and Regulation during the Integrated Stress Response. *Molecular Cell*, *73*(5). https://doi.org/10.1016/j.molcel.2018.12.006

Wilkinson, L. (2011). ggplot2: Elegant Graphics for Data Analysis by WICKHAM, H. *Biometrics*, *67*(2). https://doi.org/10.1111/j.1541-0420.2011.01616.x

Williams, G. T., & Farzaneh, F. (2012). Are snoRNAs and snoRNA host genes new players in cancer? *Nature Reviews Cancer*. https://doi.org/10.1038/nrc3195

Wissink, E. M., Vihervaara, A., Tippens, N. D., & Lis, J. T. (2019). Nascent RNA analyses: tracking transcription and its regulation. In *Nature Reviews Genetics* (Vol. 20, Issue 12). https://doi.org/10.1038/s41576-019-0159-6

Wölfl, B., te Rietmole, H., Salvioli, M., Kaznatcheev, A., Thuijsman, F., Brown, J. S., Burgering, B., & Staňková, K. (2022). The Contribution of Evolutionary Game Theory to Understanding and Treating Cancer. In *Dynamic Games and Applications* (Vol. 12, Issue 2). https://doi.org/10.1007/s13235-021-00397-w

Wolock, S. L., Lopez, R., & Klein, A. M. (2019). Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Systems*, *8*(4). https://doi.org/10.1016/j.cels.2018.11.005

Wortel, I. M. N., van der Meer, L. T., Kilberg, M. S., & van Leeuwen, F. N. (2017). Surviving Stress: Modulation of ATF4-Mediated Stress Responses in Normal and Malignant Cells. In *Trends in Endocrinology and Metabolism* (Vol. 28, Issue 11). https://doi.org/10.1016/j.tem.2017.07.003

Wragg, J. W., Roos, L., Vucenovic, D., Cvetesic, N., Lenhard, B., & Müller, F. (2020). Embryonic tissue differentiation is characterized by transitions in cell cycle dynamic-associated core promoter regulation. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkaa563

Wragg, J. W., White, P.-L., Hadzhiev, Y., Wanigasooriya, K., Stodolna, A., Tee, L., Barros-Silva, J. D., Beggs, A. D., & Müller, F. (2023). Intra-promoter switch of transcription initiation sites in proliferation signaling-dependent RNA metabolism. *Nature Structural & Molecular Biology*, *30*(12), 1970–1984. https://doi.org/10.1038/s41594-023-01156-8

Wu, F., Fan, J., He, Y., Xiong, A., Yu, J., Li, Y., Zhang, Y., Zhao, W., Zhou, F., Li, W., Zhang, J., Zhang, X., Qiao, M., Gao, G., Chen, S., Chen, X., Li, X., Hou, L., Wu, C., … Zhou, C. (2021). Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-22801-0

Wu, J. shun, Jiang, J., Chen, B. jun, Wang, K., Tang, Y. ling, & Liang, X. hua. (2021). Plasticity of cancer cell invasion: Patterns and mechanisms. In *Translational Oncology* (Vol. 14, Issue 1). https://doi.org/10.1016/j.tranon.2020.100899

Xiaolei, Y., Mead, B. E., Helia, S., Robert, L., M, K. J., & Oren, L. (2017). Stem Cell Organoid Engineering. *Cell Stem Cell*. https://doi.org/10.1016/j.stem.2015.12.005.Stem

Yamashita, R., Suzuki, Y., Takeuchi, N., Wakaguri, H., Ueda, T., Sugano, S., & Nakai, K. (2008). Comprehensive detection of human terminal oligo-pyrimidine (TOP) genes and analysis of their characteristics. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkn248

Yang, C., Jiang, L., Zhang, H., Shimoda, L. A., Deberardinis, R. J., & Semenza, G. L. (2014). Analysis of hypoxia-induced metabolic reprogramming. In *Methods in Enzymology* (Vol. 542). https://doi.org/10.1016/B978-0-12-416618-9.00022-4

Yang, M., Lu, Y., Piao, W., & Jin, H. (2022). The Translational Regulation in mTOR Pathway. In *Biomolecules* (Vol. 12, Issue 6). https://doi.org/10.3390/biom12060802

Yang, S., Corbett, S. E., Koga, Y., Wang, Z., Johnson, W. E., Yajima, M., & Campbell, J. D. (2020). Decontamination of ambient RNA in single-cell RNA-seq with DecontX. *Genome Biology*, *21*(1). https://doi.org/10.1186/s13059-020-1950-6

Yano, S., Tazawa, H., Kagawa, S., Fujiwara, T., & Hoffman, R. M. (2020). FUCCI real-time cell-cycle imaging as a guide for designing improved cancer therapy: A review of innovative strategies to target quiescent chemo-resistant cancer cells. In *Cancers*. https://doi.org/10.3390/cancers12092655

Yano, S., Zhang, Y., Miwa, S., Tome, Y., Hiroshima, Y., Uehara, F., Yamamoto, M., Suetsugu, A., Kishimoto, H., Tazawa, H., Zhao, M., Bouvet, M., Fujiwara, T., & Hoffman, R. M. (2014). Spatial-temporal FUCCI imaging of each cell in a tumor demonstrates locational dependence of cell cycle dynamics and chemoresponsiveness. *Cell Cycle*. https://doi.org/10.4161/cc.29156

Yin, X., Mead, B. E., Safaee, H., Langer, R., Karp, J. M., & Levy, O. (2016). Engineering Stem Cell Organoids. In *Cell Stem Cell*. https://doi.org/10.1016/j.stem.2015.12.005

Yu, J., Chen, X., Zhao, S., Jing, J., Wang, Q., & Dang, Y. (2022). HOXC10 Promotes Metastasis in Colorectal Cancer by Recruiting Myeloid-derived Suppressor Cells. *Journal of Cancer*, *13*(12). https://doi.org/10.7150/jca.76945

Yu, T., Gao, X., Zheng, Z., Zhao, X., Zhang, S., Li, C., & Liu, G. (2021). Intratumor Heterogeneity as a Prognostic Factor in Solid Tumors: A Systematic Review and Meta-Analysis. In *Frontiers in Oncology* (Vol. 11). https://doi.org/10.3389/fonc.2021.744064

Zaborowski, A. M., Adamina, A. A. M., Aigner, F., D'Allens, L., Allmer, C., Álvarez, A., Anula, R., Andric, M., Bach, S. A. S., Bala, M., Barussaud, M., Bausys, A., Beggs, A., Bellolio, F., Bennett, M. R., Berdinskikh, A., Bevan, V., Biondo, S., Bislenghi, G., … Winter, D. C. (2022). Impact of microsatellite status in early-onset colonic cancer. *British Journal of Surgery*, *109*(7). https://doi.org/10.1093/bjs/znac108

Zaveri, L., & Dhawan, J. (2018). Cycling to meet fate: Connecting pluripotency to the cell cycle. In *Frontiers in Cell and Developmental Biology*. https://doi.org/10.3389/fcell.2018.00057

Zhai, Z., Lei, Y. L., Wang, R., & Xie, Y. (2022). Supervised capacity preserving mapping: A clustering guided visualization method for scRNA-seq data. *Bioinformatics*, *38*(9). https://doi.org/10.1093/bioinformatics/btac131

Zhao, R., Bodnar, M. S., & Spector, D. L. (2009). Nuclear neighborhoods and gene expression. In *Current Opinion in Genetics and Development* (Vol. 19, Issue 2). https://doi.org/10.1016/j.gde.2009.02.007

Zhao, X. H., Zhao, P., Deng, Z., Yang, T., Qi, Y. X., An, L. Y., Sun, D. L., & He, H. Y. (2023). Integrative analysis reveals marker genes for intestinal mucosa barrier repairing in clinical patients. *IScience*, *26*(6). https://doi.org/10.1016/j.isci.2023.106831

Zhao, Z., Chen, X., Dowbaj, A. M., Sljukic, A., Bratlie, K., Lin, L., Fong, E. L. S., Balachander, G. M., Chen, Z., Soragni, A., Huch, M., Zeng, Y. A., Wang, Q., & Yu, H. (2022). Organoids. *Nature Reviews Methods Primers*, *2*(1), 94. https://doi.org/10.1038/s43586-022-00174-y

Zheng, J. (2012). Energy metabolism of cancer: Glycolysis versus oxidative phosphorylation (review). In *Oncology Letters* (Vol. 4, Issue 6). https://doi.org/10.3892/ol.2012.928

Zhou, Z., Zhong, Y., Zhang, Z., & Ren, X. (2023). Spatial transcriptomics deconvolution at single-cell resolution using Redeconve. *Nature Communications*, *14*(1). https://doi.org/10.1038/s41467-023-43600-9

Zhu, J., & Thompson, C. B. (2019). Metabolic regulation of cell growth and proliferation. In *Nature Reviews Molecular Cell Biology* (Vol. 20, Issue 7). https://doi.org/10.1038/s41580-019-0123-5

Zhu, X., Chen, H. H., Gao, C. Y., Zhang, X. X., Jiang, J. X., Zhang, Y., Fang, J., Zhao, F., & Chen, Z. G. (2020). Energy metabolism in cancer stem cells. *World Journal of Stem Cells*, *12*(6). https://doi.org/10.4252/WJSC.V12.I6.448

Zielke, N., & Edgar, B. A. (2015). FUCCI sensors: Powerful new tools for analysis of cell proliferation. In *Wiley Interdisciplinary Reviews: Developmental Biology*. https://doi.org/10.1002/wdev.189

Zilbauer, M., James, K. R., Kaur, M., Pott, S., Li, Z., Burger, A., Thiagarajah, J. R., Burclaff, J., Jahnsen, F. L., Perrone, F., Ross, A. D., Matteoli, G., Stakenborg, N., Sujino, T., Moor, A., Bartolome-Casado, R., Bækkevold, E. S., Zhou, R., Xie, B., … Wilson, K. T. (2023). A Roadmap for the Human Gut Cell Atlas. *Nature Reviews Gastroenterology and Hepatology*, *20*(9). https://doi.org/10.1038/s41575-023-00784-1

Zou, Z., Tao, T., Li, H., & Zhu, X. (2020). MTOR signaling pathway and mTOR inhibitors in cancer: Progress and challenges. In *Cell and Bioscience* (Vol. 10, Issue 1). https://doi.org/10.1186/s13578-020-00396-1